

# Anzo<sup>®</sup> 5.4 Administration Guide

Last Updated: 3/22/2024

---

Online documentation is available at [docs.cambridgesemantics.com](https://docs.cambridgesemantics.com)

# Table of Contents

<b>About This Doc</b> .....	<b>7</b>
<b>Getting Started with Admin after a New Installation</b> .....	<b>9</b>
Accessing the Admin Application .....	10
Getting Started Checklist .....	11
<b>Anzo Server Administration</b> .....	<b>15</b>
Starting and Stopping Anzo .....	16
Changing Anzo Server Settings .....	18
Managing Certificates .....	30
Replacing the Self-Signed Certificate .....	31
Adding a Certificate to the Trust Store .....	36
Updating the Server License .....	37
Managing Volumes .....	39
Creating a New Volume .....	40
Mounting an Existing Volume .....	42
Unmounting a Volume .....	44
Uploading a Plugin .....	46
Advanced Semantic Service Configuration .....	47
Setting the Default File Upload Path .....	48
Enabling the System Monitor Service .....	50
Routing Hi-Res Analytics to a Custom URL .....	53
Separating Audit Logs by Event Type .....	56
Limiting the Age/Size of Audit Logs .....	57
Limiting the Size/Number of anzo_full Logs .....	59

Configuring a User Inactivity Timeout .....	61
Reporting on Binary Store Access Events .....	63
Setting the Max Page Size for OData Feeds .....	65
Scanning Whole CSV Files on Import .....	67
Including Views as Database Schemas .....	68
Limiting the Number of Unstructured Status Journals .....	69
Disabling Cloud Location Pricing Information .....	71
Setting a Heartbeat for LDAP Connections .....	72
<b>Connection Administration .....</b>	<b>74</b>
Connecting to a File Store .....	75
Connecting to File Storage .....	76
Connection Settings Reference .....	77
Creating an Anzo Data Store .....	85
Connecting to AnzoGraph .....	89
Advanced Settings .....	92
Connecting to Elasticsearch .....	98
Connecting to a Distributed Unstructured Cluster .....	101
Connecting to a Cloud Location .....	104
Importing the NFS Configuration .....	105
Creating a Cloud Location .....	108
<b>Administration Tools .....</b>	<b>110</b>
Workflow Manager .....	111
Adding a Workflow .....	112
Adding a Task to a Workflow .....	115
Adding a Task that Runs an Unstructured Pipeline .....	115

Adding a Task that Refreshes or Reloads a Graphmart .....	117
Adding a Task that Pauses the Workflow .....	121
Running a Workflow .....	125
<b>Migration Packages .....</b>	<b>127</b>
Creating a Migration Package .....	128
Exporting a Migration Package .....	133
Export Configuration Settings Reference .....	137
Editing Migration Package Template Files .....	142
Importing a Migration Package .....	148
<b>User Management .....</b>	<b>153</b>
User Management and Access Control Concepts .....	154
User Management Concepts .....	155
Artifact Access Control Concepts .....	159
Connecting to a Directory Server .....	164
Adding Directory Users and Groups to Anzo .....	173
Enabling Self-Authorization for Directory Users .....	177
Configuring Single Sign-On Authentication .....	179
Adding a Basic Provider .....	180
Adding a JWT Provider .....	186
Adding a Kerberos Provider .....	192
Adding an Oauth 2 Provider .....	199
Adding an Open ID Connect Provider .....	204
Adding a SAML Provider .....	211
Creating and Managing Roles .....	216
Creating an Internal Anzo User .....	222

Predefined Anzo Roles and Permissions .....	225
Role Permissions Reference .....	233
Managing Default Access Policies .....	241
Default Access Policy Permissions Reference .....	242
Default Access Policy Reference .....	244
Configuring Default Access Policies .....	249
<b>Monitoring and Diagnostics .....</b>	<b>252</b>
Managing Anzo Logging .....	253
Logging Concepts and Configuration .....	254
Adding the Recommended Log Packages .....	261
Enabling AnzoGraph Query Logs .....	261
Enabling the Audit Logs .....	267
Viewing Log Files .....	273
Retrieving AnzoGraph Diagnostic Files .....	277
Monitoring AnzoGraph Statistics .....	280
System Query Audit .....	287
<b>AnzoGraph Administration .....</b>	<b>291</b>
Starting and Stopping AnzoGraph .....	292
Configuring AnzoGraph for Kerberos Authentication .....	294
AnzoGraph CLI .....	295
AnzoGraph Settings Reference .....	300
Changing AnzoGraph Settings .....	314
Configuration File Overview .....	315
Managing AnzoGraph File Access Policies .....	316

Relocating AnzoGraph Directories .....	320
Enabling Persistence (Preview) .....	322
Ignoring Missing Graphs .....	324
Changing the Default FROM Clause Behavior .....	326
Managing Automatic Restarts .....	327
<b>Anzo Admin CLI .....</b>	<b>331</b>
Getting Started with the Admin CLI .....	332
Admin CLI Basics .....	336
<b>Ongoing Operations .....</b>	<b>339</b>
Platform Usage and Performance Monitoring .....	340
Data Quality Monitoring .....	347
<b>Troubleshoot .....</b>	<b>350</b>
Error Message Reference .....	351
Investigating when Anzo is Unresponsive .....	358
Updating an Expired License .....	363
Restoring the Server ID .....	364
Viewing the Current Stack in a Browser .....	366
Taking AnzoGraph X-Rays from the Command Line .....	368
Generating Diagnostic Files in AnzoGraph 2.5 .....	369
Generating Diagnostic Files in AnzoGraph 3.1 .....	371

# About This Doc

This document provides guidance for Anzo administrators on the configuration, administration, and troubleshooting of Anzo components.

## Tip

You can view the contents of this guide as well as release notes, end-user, and deployment documentation online at [docs.cambridgesemantics.com](https://docs.cambridgesemantics.com). You can also find PDF versions of the end-user and deployment documentation [here](#).

The following list introduces the sections in this guide.

- [Getting Started with Admin after a New Installation](#): Gives guidance on accessing the Administration application and where to start if you are an administrator configuring Anzo after the initial deployment.
- [Anzo Server Administration](#): Provides information on server settings, managing certificates, updating the license, managing volumes, uploading plugins, and configuring semantic services.
- [Connection Administration](#): Includes instructions on connecting to file stores, creating Anzo data stores, and connecting to AnzoGraph, Elasticsearch, Anzo Distributed Unstructured clusters, and Cloud Locations.
- [Administration Tools](#): Provides information on automating tasks with Workflow Manager and assembling migration packages for bulk export and import of artifacts.
- [User Management](#): Includes information on user management and access control and instructions on connecting to a directory server, SSO provider, and managing roles, users, and groups.
- [Monitoring and Diagnostics](#): Provides information on monitoring events and managing Anzo and AnzoGraph diagnostic files.
- [AnzoGraph Administration](#): Includes reference information and instructions on performing administrative tasks on an AnzoGraph server.

- [Anzo Admin CLI](#): Includes basic information about the advanced Anzo admin command line interface.
- [Ongoing Operations](#): Gives guidance on tasks to perform on a regular basis to monitor the usage, performance, and quality of the data in the system.
- [Troubleshoot](#): Includes an error message reference as well as instructions on retrieving diagnostic files and resolving certain issues.



# Getting Started with Admin after a New Installation

This topic gives guidance on where to start if you are an administrator configuring Anzo after the initial deployment.

- [Accessing the Admin Application](#)
- [Getting Started Checklist](#)

# Accessing the Admin Application


To open the Administration application, go to the following URL:

```
https://hostname:port/sdl/index.html#/admin
```

Where `hostname` is the Anzo server DNS name or IP address and `port` is the HTTP/S port for the Administration application. The default HTTPS port is 8946, and HTTP is 8945.

## Tip

You can change the URL for the Administration application by configuring the **Admin Home Page** value in server settings. For more information, see [Change the Application Home Page URLs](#).

To access the Administration application from the Anzo application, click the administration icon () on the right side of the top menu bar. Clicking the icon opens the Administration menu, and selecting a menu item opens the application.

# Getting Started Checklist

This section describes the tasks that are important to consider and complete before users start to access Anzo and onboard data. You can complete these tasks in any order.

- [Review the Server Settings](#)
- [Replace the Self-Signed Certificate](#)
- [Connect the Platform Applications](#)
- [Upload JDBC Drivers for Custom Database Connections](#)
- [Set the Default File Upload Path](#)
- [Create a Data Store](#)
- [Add Users and Configure Access Control Policies](#)
- [Configure Logging Options](#)

## Review the Server Settings

It is a good idea to review the Anzo server settings to verify that options such as the application ports and access URLs, binary store location, SPARQL endpoint, and versioning environment are configured as desired. See [Changing Anzo Server Settings](#) for information.

## Replace the Self-Signed Certificate

Anzo installations include a self-signed certificate. To strengthen the security of the environment, Cambridge Semantics recommends that you replace the default certificate with a trusted one. For instructions, see [Replacing the Self-Signed Certificate](#).

## Connect the Platform Applications

The other applications that were installed in your environment need to be connected to Anzo. The list below provides links to the instructions:

- If a Kubernetes cluster was set up so that AnzoGraph, Elasticsearch, and Anzo Unstructured applications can be dynamically deployed as needed, create a Cloud Location. See [Connecting to a Cloud Location](#) for instructions.
- If a static AnzoGraph cluster was deployed, see [Connecting to AnzoGraph](#) for instructions on configuring the connection to Anzo.
- If a static Elasticsearch instance was deployed, see [Connecting to Elasticsearch](#) for instructions on configuring the connection to Anzo and AnzoGraph.
- If a static Anzo Distributed Unstructured cluster was deployed, see [Connecting to a Distributed Unstructured Cluster](#) for instructions on configuring the connection to Anzo.

## Upload JDBC Drivers for Custom Database Connections

Anzo and AnzoGraph include JDBC drivers for connecting to the following databases:

- Databricks
- H2
- IBM DB2
- Microsoft SQL Server
- MariaDB
- Oracle
- PostgreSQL
- SAP Sybase (jTDS)
- Snowflake

If your organization plans to onboard data from other databases, the drivers for those sources need to be added to both Anzo and AnzoGraph. For instructions on adding drivers to Anzo, see [Uploading a Plugin](#). For instructions on adding drivers to AnzoGraph, see [Deploy Drivers for Custom Database Sources](#) in the Deployment Guide.

## Set the Default File Upload Path

By default, if a user imports a file (such as a CSV, XML, or JSON file) to Anzo from their computer, Anzo is configured to copy the file to the data directory in the installation path. When the file is in the installation path instead of the shared file store, it is not accessible by the other applications in the platform. In addition, other users cannot onboard that data because they typically do not have access to the files in that location. It is important to configure a new default file upload path so that source files can be shared with other applications and users. For instructions, see [Setting the Default File Upload Path](#).

## Create a Data Store

An Anzo Data Store is a directory on the shared file store where Anzo can write file-based linked data sets (FLDS). A data store is required when setting up unstructured pipelines. A data store is also required when structured data is onboarded with the automated direct load workflow and the workflow is configured to export the data to a dataset. For instructions on creating a data store, see [Creating an Anzo Data Store](#).

## Add Users and Configure Access Control Policies

Before you get started with adding users to Anzo, it may be beneficial to review the [User Management and Access Control Concepts](#) topics to learn about how users, groups, and roles are used in Anzo and how data access policies are implemented. Once you are familiar with the concepts, you can connect to your organization's directory server, add users and groups to Anzo, create and configure roles, and set up single-sign on access if desired. For links to all of the information about user management, see [User Management](#).

As part of user management, it is also important to review the Default Access Policies for your deployment. These are the security policies that are applied by default to the artifacts that are stored in Anzo. By default, most access policies give the creator of an artifact "admin" rights to that artifact, meaning the creator can view, modify, and delete that artifact. In addition, the Everyone role (i.e. all authenticated users) is given "view" permissions for the artifacts, meaning all authenticated users can see that an artifact exists but they cannot modify or delete it. For more information, see [Managing Default Access Policies](#).

## Configure Logging Options

By default, Anzo is configured to log information about core server operations and services to ensure that diagnostics are generated when errors occur. Additional logging can be enabled, however, to provide information for auditing or monitoring purposes. To learn the basics about logging in Anzo, it might be helpful to review [Logging Concepts and Configuration](#). Then see [Adding the Recommended Log Packages](#) for information about enabling additional logging. Also see [Enabling the System Monitor Service](#) for information about monitoring the state of the Java virtual machine and capturing stack and heap dumps for troubleshooting.

# Anzo Server Administration

The topics in this section provide information about managing the Anzo server configuration.

**In this section:**

- Starting and Stopping Anzo ..... 16
- Changing Anzo Server Settings ..... 18
- Managing Certificates ..... 30
- Updating the Server License ..... 37
- Managing Volumes ..... 39
- Uploading a Plugin ..... 46
- Advanced Semantic Service Configuration ..... 47

## Starting and Stopping Anzo

If Anzo is run via a systemd service, as described in [Configure and Start the Anzo Service](#) in the Deployment Guide, use `systemctl` to start and stop Anzo.

To start Anzo, run the following command:

```
sudo systemctl start anzo-server
```

To stop Anzo, run the following command:

```
sudo systemctl stop anzo-server
```

To start Anzo using the `AnzoServer` utility, run the following command. Make sure that you are logged in as the Anzo service user before stopping or starting Anzo:

```
<install_path>/Server/AnzoServer start
```

To stop Anzo, run the following command:

```
<install_path>/Server/AnzoServer stop
```

You can also start and stop Anzo from the symbolic links if they were created for your installation. For example, `/etc/init.d/AnzoServer start` or `/etc/init.d/AnzoServer stop`.

## Monitoring Startup Status

It can take a few minutes for Anzo to complete the startup process. You can monitor the status by viewing the Status page. To see the Status page, go to the following URL in your browser:

```
http://<hostname_or_IP_address>:8945/status
```

Where `<hostname_or_IP_address>` is the name or IP address of the server that hosts Anzo.

For example, the following image shows the Status page message displayed while Anzo is starting:



[DISABLE AUTO REFRESH](#)

**Overall Status: Not All Started**

[Show JVM Details](#)

**CREATED**

[Show Details](#)

**STARTING**

[Hide Details](#)

**org.openanzo.activemq.EmbeddedActiveMQServer - STARTING**

[Show Details](#)

---

**STARTED**

[Show Details](#)

**NOT\_ENABLED**

[Show Details](#)

The image below shows the Status page message when startup is complete:

**Overall Status: OK**

[Show JVM Details](#)

**STARTED**

[Show Details](#)

**NOT\_ENABLED**

[Show Details](#)

# Changing Anzo Server Settings

This topic provides instructions for changing the Anzo server settings. To access the settings, expand the **Servers** menu in the Administration application and click **Server Settings**.

## Important

After changing any of the server configuration settings, you must restart Anzo to apply the change.

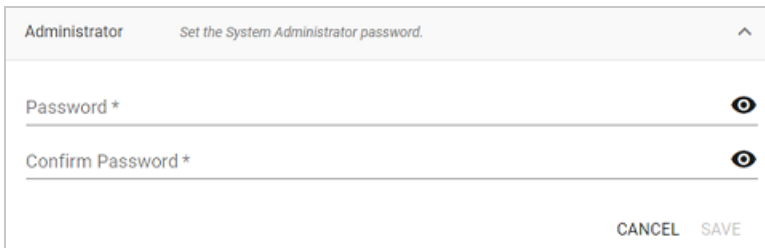
## Note

You can have one option open for editing at a time. If you are in the process of modifying an option and have not saved the changes, all other Edit buttons are disabled until you save or cancel the changes.

- [Change the sysadmin Password](#)
- [Regenerate the Server Secret](#)
- [Configure the Server Ports](#)
- [Configure the Binary Store](#)
- [Connect an Email Server](#)
- [Change the Application Home Page URLs](#)
- [Change the HTTP Session Timeout](#)
- [Configure Anonymous User Access](#)
- [Configure the URI Prefix and SPARQL Endpoint](#)
- [Configure Global Prefixes](#)
- [Set the Version Environment Variable](#)
- [Configure DU Cluster Network Connections](#)
- [Set the Default Anzo Data Store](#)

## Change the sysadmin Password

To change the system administrator (**sysadmin**) password, expand the **Administrator** option and click **Edit**.



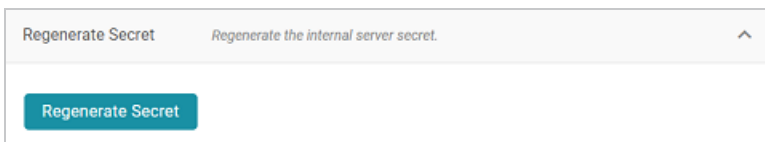
A dialog box titled "Administrator" with the subtitle "Set the System Administrator password." It contains two password input fields: "Password \*" and "Confirm Password \*", each with a toggle icon to the right. At the bottom right, there are "CANCEL" and "SAVE" buttons.

Type the new password in the **Password** and **Confirm Password** fields. Then click **Save**.

## Regenerate the Server Secret

Cambridge Semantics recommends that you back up the current Anzo installation before regenerating the secret. Regenerating the secret requires a restart of Anzo.

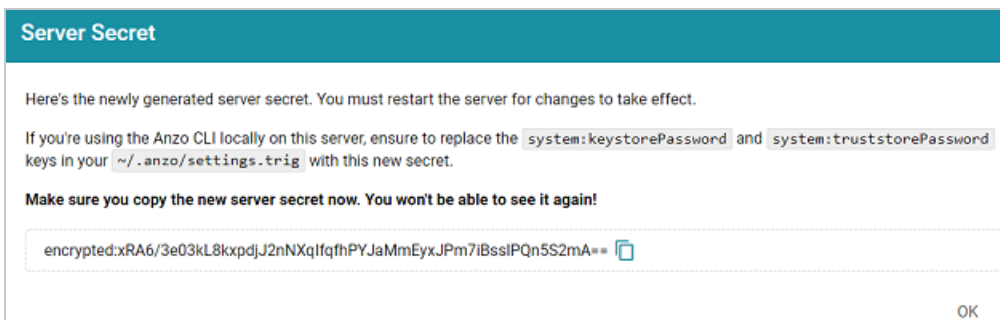
1. To change the password for the Anzo key and trust stores, expand the **Regenerate Secret** option.



A dialog box titled "Regenerate Secret" with the subtitle "Regenerate the internal server secret." It features a single blue button labeled "Regenerate Secret".

2. Click the **Regenerate Secret** button. Review the confirmation message that is displayed and click **Yes** to generate a new secret.

Anzo generates the new secret and presents a dialog box that displays the encrypted secret to copy. For example:



A dialog box titled "Server Secret" with a teal header. The text inside reads: "Here's the newly generated server secret. You must restart the server for changes to take effect. If you're using the Anzo CLI locally on this server, ensure to replace the `system:keystorePassword` and `system:truststorePassword` keys in your `~/ .anzo/settings.trig` with this new secret. **Make sure you copy the new server secret now. You won't be able to see it again!**" Below the text is a dashed border containing the encrypted secret: `encrypted:xRA6/3e03kL8kxpdjJ2nNXqfIqfhPYJaMmEyxJPm7iBssIPQn5S2mA==` with a copy icon to its right. An "OK" button is located at the bottom right.

3. Make sure you copy the secret because it is not possible to view again.
4. If you regenerated the secret on a server where the Anzo Admin CLI is used, the new secret also needs to be changed in the `~/.anzo/settings.trig` file for the Anzo service user. To replace the secret in `settings.trig`, follow these steps:
  - a. Open `~/.anzo/settings.trig` for editing.
  - b. Locate the `system:keystorePassword` and `system:truststorePassword` properties.
  - c. Replace both the object values for both properties with the secret that was copied in step 3. Replace only the content between the quotation marks as shown below:
 

```
system:keystorePassword "<new_secret>"^^anzo:password ;
system:truststorePassword "<new_secret>"^^anzo:password ;
```
  - d. Save and close `settings.trig`.
5. Restart Anzo to apply the new secret.

## Configure the Server Ports

To change, enable, or disable the Anzo server ports, expand the **Ports** option and click **Edit**.

Enabled	Port	SSL Port	Certificates
<b>Anzo Port and Anzo SSL Port</b>			
Application	8080		
Application SSL	8443		anzo
Auxiliary	8945		
Auxiliary SSL	8946		anzo

CANCEL SAVE

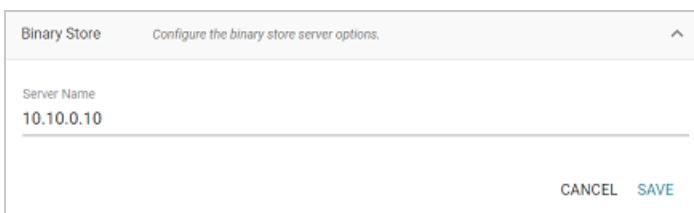
Change the values in the Port fields to specify alternate port numbers. To enable or disable a port, move the slider next to the application name to the left or right. The list below describes the settings:

- The fields at the top of the screen specify the Anzo server ports. By default, the Anzo and Anzo SSL ports are enabled. If you want to disable one of the ports, click the **Enabled** drop down list and select the option that you want to leave enabled. To change port numbers, click in the **Port** field and specify the port.
- The **Application** and **Application SSL** ports are the HTTP and HTTPS client application ports.
- The **Auxiliary** and **Auxiliary SSL** ports are the HTTP and HTTPS Administration client ports.

For information about managing the certificates to use for the SSL ports, see [Replacing the Self-Signed Certificate](#).

## Configure the Binary Store

To change the host server for the binary (blob) store, expand **Binary Store** and click **Edit**.



The Server Name defaults to the host name or IP address for the Anzo server. To specify a different host for the binary store, type the new host name or IP address in the **Server Name** field, and then click **Save**.

## Connect an Email Server

To configure an SMTP server for sending email, expand **Email Server Configuration** and click **Edit**.

Email Server Configuration Configure the SMTP server used to send email.

Host Name  
smtp.example.com

Port  
25

Use SSL

Username

Password

CANCEL SAVE

- **Host Name** is the host name or IP address for the SMTP server.
- **Port** is the port for the connection.
- If the email server is configured for SSL authentication, select the **Use SSL** checkbox to enable SSL authentication.
- Specify the **Username** and **Password** to use for authentication.

Click **Save** to save the changes.

## Change the Application Home Page URLs

To change the home page path for the Anzo and Administration application URLs, expand **Home Pages** and click **Edit**.

Home Pages Configure the default root page served.

Admin Home Page  
sdl/index.html#/admin/server-settings

Application Home Page  
sdl

CANCEL SAVE

- The **Admin Home Page** is the home page path for the Administration application.
- The **Application Home Page** is the home page path for the Anzo application.

Click **Save** to save the changes.

## Change the HTTP Session Timeout

To configure the HTTP session timeout value, expand **HTTP Session Management** and click **Edit**.



The screenshot shows a configuration dialog box for HTTP Session Management. The title bar reads "HTTP Session Management" and "Configure HTTP session options." with an expand/collapse arrow. The main content area has a label "Session Timeout" above a dropdown menu currently set to "7 days". At the bottom right, there are two buttons: "CANCEL" and "SAVE".

Click the **Session Timeout** drop-down list and select the timeout value. Then click **Save** to save the change.

## Configure Anonymous User Access

Before enabling anonymous access, consider the following security implications:

- [Anonymous User Permissions](#)
- [Anonymous User Limitations](#)
- [Important Considerations](#)

### Anonymous User Permissions

When anonymous access is enabled:

- The server allows any user to connect to the Hi-Res Analytics application without a username and password. A user can connect to without having an account in Anzo.
- Anonymous users are considered members of the Everyone role. Anonymous users can read data in Anzo that is tagged as readable by Everyone.

### Anonymous User Limitations

Anonymous users cannot:

- Add, delete, or modify data. Anonymous users cannot write or delete data even if the Everyone role has write or delete access.

- Change permissions on the artifacts in Anzo. Anonymous users cannot change the Sharing or Security tab settings for any data on the server even if the Everyone role has write or delete access to an artifact's metadata.

## Important Considerations

This section lists important ideas to consider before enabling anonymous access.

- **Consider Existing Access Control:** User permissions might have been previously configured without anticipating that other users could have anonymous access. Before enabling anonymous access, consider that data that is viewable by the **Everyone** role becomes visible to anonymous users. You might need to change the permissions for existing data, such as by granting read access to the **Authenticated Users** role instead of the Everyone role. For more information about permissions, see [Predefined Anzo Roles and Permissions](#).
- **Consider Server Network Protections:** Consider that anyone who can reach the server via the network will be able to use it as an anonymous user. Evaluate firewalls and other network protection mechanisms to limit access to the Anzo server as desired. For example, you might want to allow anonymous access to anyone inside your organization's internal network but disable access to the server from the public internet.
- **Consider that anonymous access can be useful:** Allowing anonymous access makes it easy to share data and views of data with others. For example, it means that you can share your Hi-Res Analytics dashboards with people who do not have a user account. It also lets you embed read-only interactive Hi-Res Analytic views inside other websites.

## Configuring Anonymous Access

To enable or disable anonymous user access, expand **Anonymous User Access** and click **Edit**.



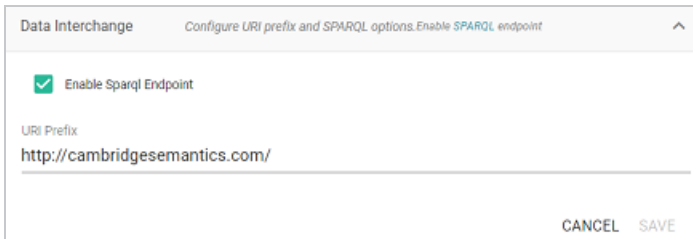
The screenshot shows a dialog box titled "Anonymous User Access" with the subtitle "Configure anonymous user access settings." and an upward-pointing arrow icon. Inside the dialog, there is a checkbox labeled "Allow Anonymous Access" which is currently unchecked. At the bottom right of the dialog, there are two buttons: "CANCEL" and "SAVE".



To enable anonymous access, select the **Allow Anonymous Access** checkbox. To disable anonymous access if it is enabled, clear the checkbox. Then click **Save**.

## Configure the URI Prefix and SPARQL Endpoint

To enable or disable the Anzo SPARQL endpoint or customize the URI prefix that Anzo generates for data identifiers, expand **Data Interchange** and click **Edit**.



The screenshot shows a configuration window titled "Data Interchange" with a subtitle "Configure URI prefix and SPARQL options.Enable SPARQL endpoint". It features a checked checkbox for "Enable Sparql Endpoint" and a text input field for "URI Prefix" with the value "http://cambridgesemantics.com/". "CANCEL" and "SAVE" buttons are located at the bottom right.

- If you want to enable or disable the Anzo SPARQL endpoint, select or clear the **Enable SPARQL Endpoint** checkbox.
- To change the prefix that Anzo uses when generating URIs, type the new value in the **URI Prefix** field. The URI Prefix is mostly used for consistency in internal data, but it is also used by default for data model URI prefixes when the model does not define the URI template to use. When changing the URI Prefix, make sure that the value is a valid prefix. See [Relative IRIs](#) in the SPARQL Query Language specification for more information.

## Configure Global Prefixes

The Global Prefix Manager stores standard prefixes and any custom prefixes that you want Anzo to recognize globally. Defining global prefixes creates shortcuts for inserting the prefixes in Query Builder and data layer queries. To manage global prefixes, expand **Global Prefix Manager**.

Prefix	Uri	EDIT	DELETE
dcterms	http://purl.org/dc/terms/	EDIT	DELETE
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#	EDIT	DELETE
owl	http://www.w3.org/2002/07/owl#	EDIT	DELETE
dc	http://purl.org/dc/elements/1.1/	EDIT	DELETE
rdfs	http://www.w3.org/2000/01/rdf-schema#	EDIT	DELETE
foaf	http://xmlns.com/foaf/0.1/	EDIT	DELETE
xsd	http://www.w3.org/2001/XMLSchema#	EDIT	DELETE

To add a prefix, click **Add Prefix**. Anzo opens the Create Prefix dialog box. In the **Prefix** field, specify the abbreviation that you want to use to represent the URI. In the **Prefix URI** field, specify the full, valid URI. For example:

**Create Prefix**

Prefix \*

Prefix URI \*

CANCEL SAVE

Click **Save** to save the definition. To use global prefix shortcuts in the Anzo application, type "prefix" followed by a space in the Query Builder or a Query Step to open a tooltip that lists the global prefixes. For example:



Clicking a prefix inserts a PREFIX statement into the query. In addition, typing the abbreviation for a global prefix followed by a colon (:) automatically inserts the PREFIX statement into the query without opening the tooltip. For example, typing **ex:** inserts a statement for the prefix that was defined in the example above.

## Set the Version Environment Variable

To change the variable value for the Version Environment tag that is displayed at the top of the Anzo application and that Anzo adds to archived versions of entities, expand **Versioning** and click **Edit**.

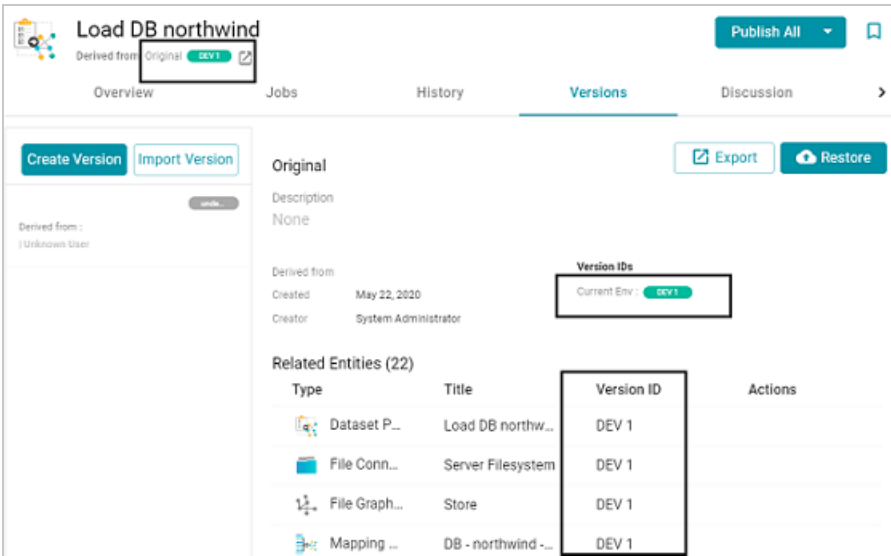


The image shows a dialog box titled "Versioning" with the subtitle "Configure the versioning environment". Inside the dialog, there is a field labeled "Versioning Environment" containing the text "DEV". At the bottom right of the dialog, there are two buttons: "CANCEL" and "SAVE".

Edit the value in the **Versioning Environment** field and click **Save**. The images below show examples of the version tags that are controlled by the Versioning Environment setting. This image shows the version at the top of the Anzo application:



For artifact versions, the black rectangles in the image below highlight the areas where the environment version variable value is displayed:



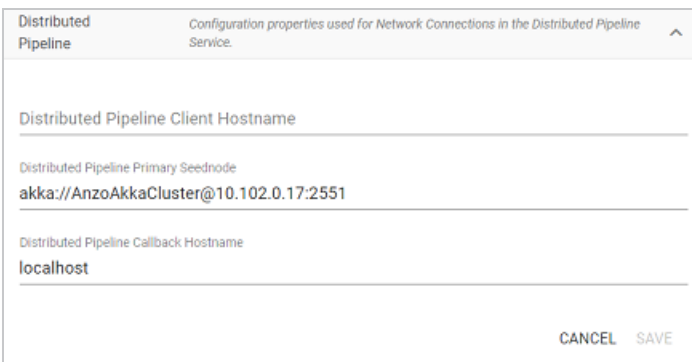
## Configure DU Cluster Network Connections

These settings configure the connection from the worker nodes back to the leader node. For instructions on connecting the leader to Anzo, see [Connecting to a Distributed Unstructured Cluster](#).

To change the network settings, expand **Distributed Pipeline** and click **Edit**.

### Note

If the Kubernetes infrastructure is set up to deploy Anzo Unstructured clusters on-demand, you do not need to configure these settings. For information about Kubernetes-based deployments, see [Configure K8s for Dynamic Deployments](#) in the Deployment Guide.



Modify the settings as needed:

- **Distributed Pipeline Client Hostname:** The hostname or IP address for the leader instance.

### Important

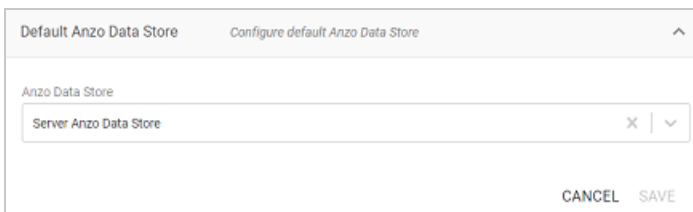
The value must be a routable IP address or hostname. If the leader node is installed on the Anzo host server, specify the IP address or hostname of the server. Do not use `127.0.0.1` or `localhost`.

- **Distributed Pipeline Primary Seednode:** The IP address and port for the leader instance. By default the leader port is **2551**.
- **Distributed Pipeline Callback Hostname:** The hostname or IP address for the instance. Typically this is the same value as the **Distributed Pipeline Client Hostname**.

Click **Save** to save the changes.

## Set the Default Anzo Data Store

To set the default Anzo Data Store so that so that it is automatically selected when users set up unstructured pipelines or export datasets from the automated direct load workflow, expand **Default Anzo Data Store** and click **Edit**.



The screenshot shows a configuration window titled "Default Anzo Data Store" with a subtitle "Configure default Anzo Data Store". Inside the window, there is a section labeled "Anzo Data Store" which contains a text input field with "Server Anzo Data Store" and a drop-down arrow. At the bottom right of the window, there are two buttons: "CANCEL" and "SAVE".

Click the **Anzo Data Store** drop-down list and select the data store to make the default store. Then click **Save**.

# Managing Certificates

The topics in this section provide information about managing server certificates.

**In this section:**

<a href="#">Replacing the Self-Signed Certificate .....</a>	<a href="#">31</a>
<a href="#">Adding a Certificate to the Trust Store .....</a>	<a href="#">36</a>

## Replacing the Self-Signed Certificate

By default, Anzo installations include a self-signed certificate. Follow the instructions below if you want to replace the default certificate with a trusted one. The steps guide you through using OpenSSL to generate an SSL certificate and signing request and then uploading the signed certificate to Anzo.

- [Generate an SSL Certificate and Signing Request](#)
- [Upload the Trusted Certificate to Anzo](#)

### Generate an SSL Certificate and Signing Request

1. If necessary, install OpenSSL.
2. Create a request configuration file. For example, create a file called **certificate.cnf**. Then add the following contents to the file. These contents include parameters for creating a multi-domain certificate:

```
# certificate.cnf

[req]
default_bits = 2048
prompt = no
default_md = rsa
req_extensions = req_ext
distinguished_name = dn

[ dn ]
C = <country>
ST = <state>
L = <locality>
O = <organization-or-company-name>
OU = <organizational-unit>
emailAddress = <email-address>
CN = <common-name-or-server-fqdm>

[ req_ext ]
subjectAltName = @alt_names
```

```
[ alt_names ]
DNS.1 = <domain1-name-or-ip>
DNS.2 = <domain2-name-or-ip>
DNS.3 = <domain3-name-or-ip>
```

3. Replace the placeholders in the file with the appropriate values. For example:

```
# certificate.cnf

[req]
default_bits = 2048
prompt = no
default_md = rsa
req_extensions = req_ext
distinguished_name = dn

[ dn ]
C = US
ST = MA
L = Boston
O = Cambridge Semantics
OU = IT
emailAddress = webmaster@cambridgesemantics.com
CN = sample.cambridgesemantics.com

[ req_ext ]
subjectAltName = @alt_names

[ alt_names ]
DNS.1 = sample1.domain.com
DNS.2 = 10.0.33.103
DNS.3 = sample3.domain.com
```

4. Run the following command to generate the signing request and private key using the configuration file:

```
openssl req -new -sha256 -nodes -out <csr_file_name>.csr -newkey rsa:2048
-keyout <key_name>.pem -config <config_file_name>.cnf
```

For example:

```
openssl req -new -sha256 -nodes -out anzo-csr.csr -newkey rsa:2048
-keyout anzo-key.pem -config certificate.cnf
```



5. Send the resulting CSR to a certificate authority for signing.

## Upload the Trusted Certificate to Anzo

1. When you receive the signed certificate from the certificate authority, rename the certificate to **anzo-crt.crt**.
2. Then follow the steps below to create a PKCS12 key:
  - a. Run the following command to concatenate the signed certificate and private key file that you generated into an `anzo.pem` file:

```
cat <key_name>.pem anzo-crt.crt > anzo.pem
```

For example:

```
cat anzo-key.pem anzo-crt.crt > anzo.pem
```

- b. Run the following command to convert the resulting `anzo.pem` file to PKCS12, choose a name for the certificate, and set an export password:

```
openssl pkcs12 -export -in anzo.pem -out anzo.pkcs12 -name "<destination_ alias>"
```

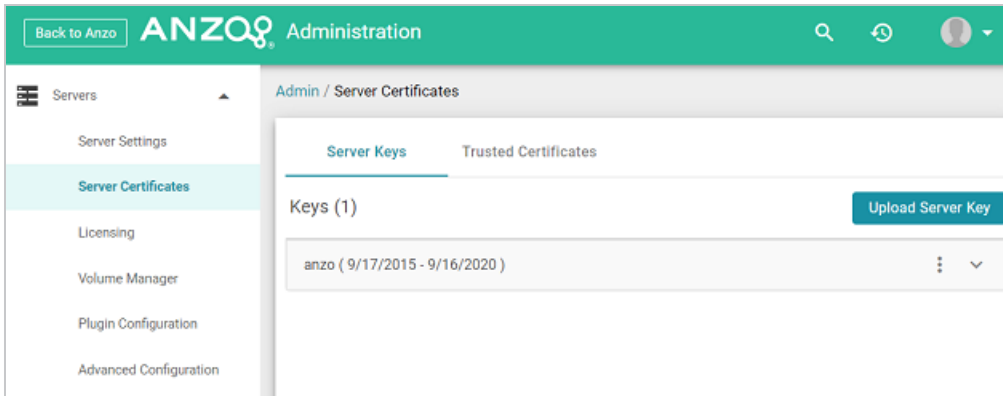
### Note

If you have installed OpenSSL version 3 or later, include the `--legacy` flag in the command (shown below):

```
openssl pkcs12 -export -in anzo.pem -out anzo.pkcs12 -name  
"<destination_alias>" --legacy
```

```
Enter Export Password:  
Verifying - Enter Export Password:
```

3. Copy the `anzo.pkcs12` certificate to your computer if necessary.
4. In the Administration application, expand the **Servers** menu and click **Server Certificates**. Anzo displays the Server Certificates screen. For example:



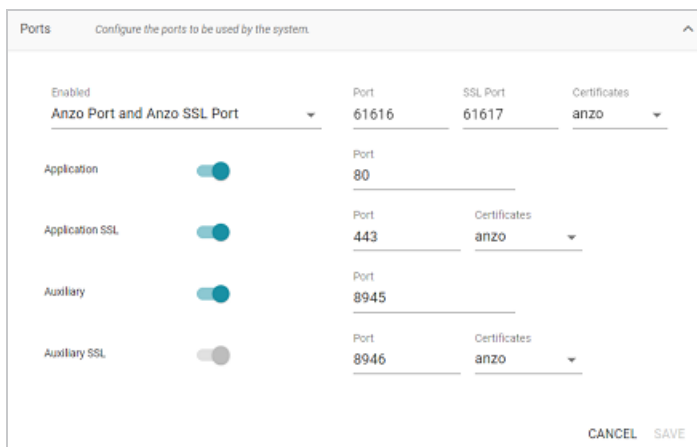
5. Click **Upload Server Key**. Anzo displays the Upload Server Key dialog box.

6. Supply the required values:

- In the **Destination Alias** field, specify the destination alias that you chose when you created the PKCS12 certificate.
- In the **Password** field, specify the Export Password that you set when you created the PKCS12 certificate.
- Click the **Choose File** button and select the **anzo.pkcs12** file.
- Click the **Keystore type** field and select **PKCS12** from the drop-down list.

7. Click **Upload** to upload the certificate.

8. Finally, follow these steps to apply the new certificate to the Anzo server SSL ports:
  - a. In the Servers menu, click **Server Settings**.
  - b. On the Server Settings screen, expand **Ports** and click **Edit**. For example:

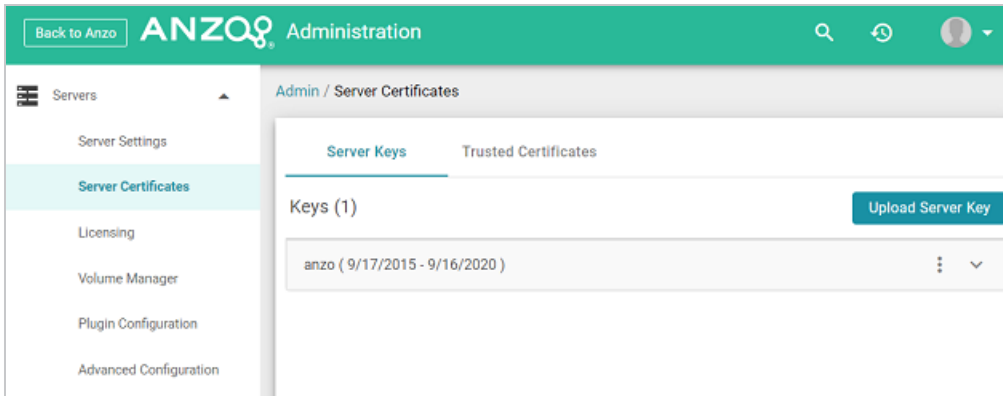


- c. Click the **Certificates** drop-down list for each of the enabled SSL ports and select the new certificate. Then click **Save**.
9. Restart Anzo to apply the configuration change.

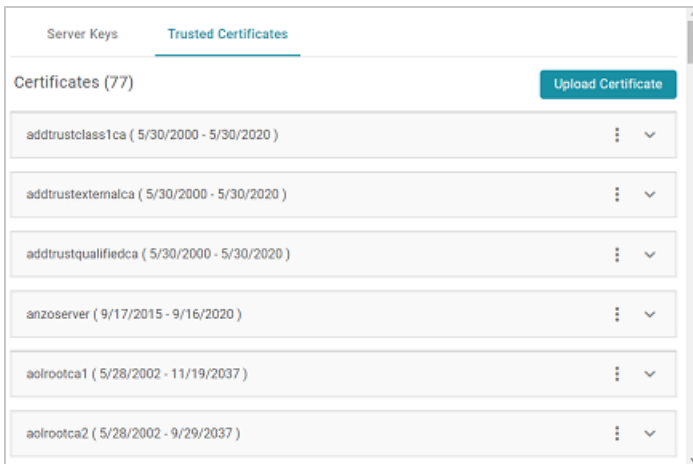
# Adding a Certificate to the Trust Store

To add a certificate to the Anzo trust store, follow the steps below.

1. In the Administration application, expand the **Servers** menu and click **Server Certificates**. Anzo displays the Server Certificates screen. For example:



2. On the Server Certificates screen, click the **Trusted Certificates** tab. Anzo displays the list of existing certificates. For example:



3. To upload a new certificate, click the **Upload Certificate** button. Browse to the certificate file, and double-click the file to upload it to Anzo.
4. Once the file is uploaded, restart Anzo to apply the change.

# Updating the Server License

Follow the instructions below to update the Anzo server license key.

## Important

If your license is expired, do not follow the steps below. The Server Licensing screen (shown in step 1 below) will be blank except for an `Access Denied/Forbidden License is invalid` error message. To update an expired license, follow the instructions in [Updating an Expired License](#).

1. In the Administration application, expand the **Servers** menu and click **Licensing**. Anzo displays the Server Licensing Information screen. For example:

Server Licensing Information Restart Server

License Details	
Product Family	ENTERPRISE
Purpose	Evaluation
Usage	Non Production
License Contact	CSI
License Company	CSI
Max Named Users	Unlimited
License Expiration:	31 August 2021
Server ID:	14BF-24AB-7402-46E0-B936-397F

Update Licensed Features

Licensed Features

2. Click **Update Licensed Features** to expand that section of the screen.

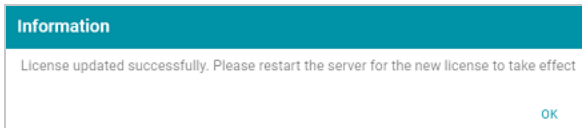
Update Licensed Features

Copy and paste the license key into the textbox below.  
You can access your license key from the homepage of your [Cambridge Semantics support account](#).

License Key

Update License

3. Paste the new license key into the **License Key** field, and then click the **Update License** button. The license is updated but does not take effect until Anzo is restarted. The following dialog box is displayed:



4. Click **OK** to close the dialog box. Then restart Anzo to apply the license updates. You can click the **Restart Server** button at the top of the screen. For information about other ways to stop and start Anzo, see [Starting and Stopping Anzo](#).

### Note

It may take Anzo noticeably longer to start for the first time after the license is updated. Subsequent starts will return to the usual startup time.

# Managing Volumes

The topics in this section provide information about creating new volumes (also known as journals or database instances) and mounting or unmounting existing volumes.

**In this section:**

Creating a New Volume .....	40
Mounting an Existing Volume .....	42
Unmounting a Volume .....	44

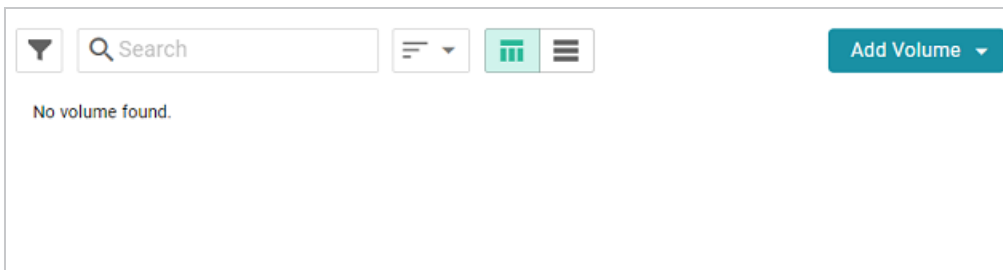
# Creating a New Volume

This topic provides instructions for creating new volumes or journals.

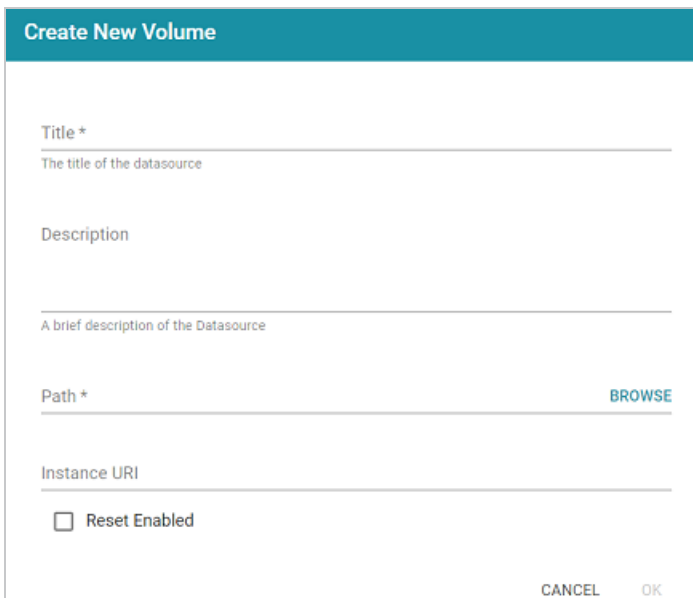
## Note

The number of volumes that you can create depends on your software license. For more information, contact Cambridge Semantics Support.

1. In the Administration application, expand the **Servers** menu and click **Volume Manager**. Anzo displays the Volume Manager screen, which lists any existing user-defined volumes (system volumes can be displayed by selecting the system data filter). For example:

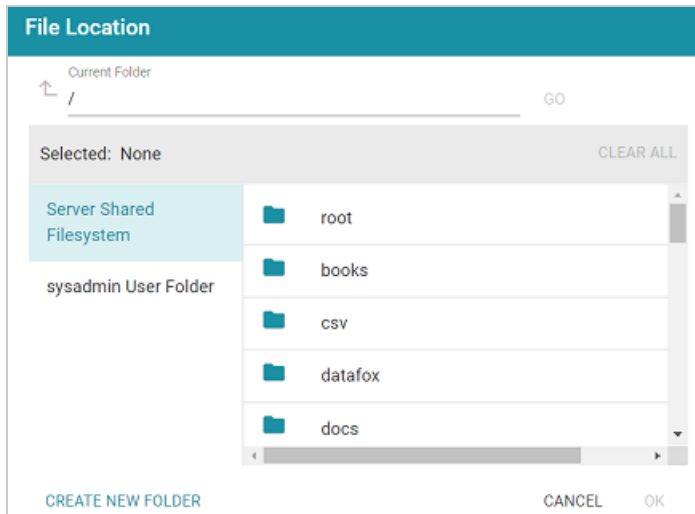


2. Click the **Add Volume** button and select **Add Volume**. Anzo displays the Create New Volume dialog box.

A screenshot of the 'Create New Volume' dialog box. The title bar is teal and contains the text 'Create New Volume'. The dialog has several input fields: 'Title \*' with a subtitle 'The title of the datasource', 'Description' with a subtitle 'A brief description of the Datasource', 'Path \*' with a 'BROWSE' link to its right, and 'Instance URI'. At the bottom left, there is a checkbox labeled 'Reset Enabled'. At the bottom right, there are 'CANCEL' and 'OK' buttons.



3. In the **Title** field, type a name for the new volume, and type an optional description in the **Description** field.
4. Click the **Path** field to open the File Location dialog box. For example:



5. On the left side of the screen, select the file store where you want to create this volume. On the right side of the screen, select the directory where you want Anzo to save the volume. If needed, you can click **Create New Folder** to create a new directory. Then click **OK** to close the File Location dialog box.
6. On the Create New Volume screen, complete the remaining fields:
  - **Instance URI**: Anzo automatically assigns an instance URI to this volume. If you want to specify a custom URI, type the URI in this field.
  - **Reset Enabled**: Controls whether volume resets are enabled. When reset is enabled, you can use an Anzo Admin CLI call to reset the entire contents of the volume without having to delete and recreate it. To enable resets for this volume, select the **Reset Enabled** checkbox.
7. Click **Save** to create the new volume in the location that you specified.

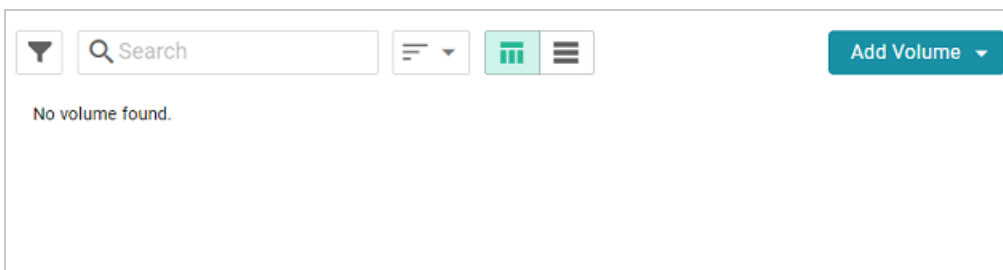
## Mounting an Existing Volume

This topic provides instructions for mounting an existing volume or journal.

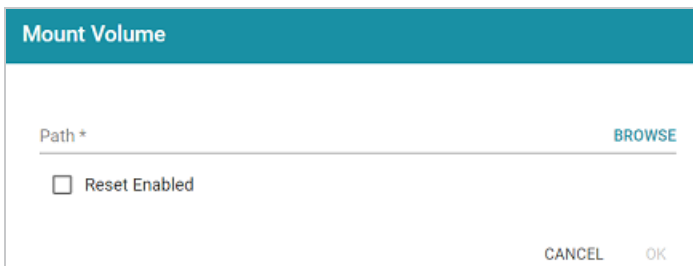
### Note

The number of volumes that you can mount depends on your software license. For more information, contact Cambridge Semantics Support.

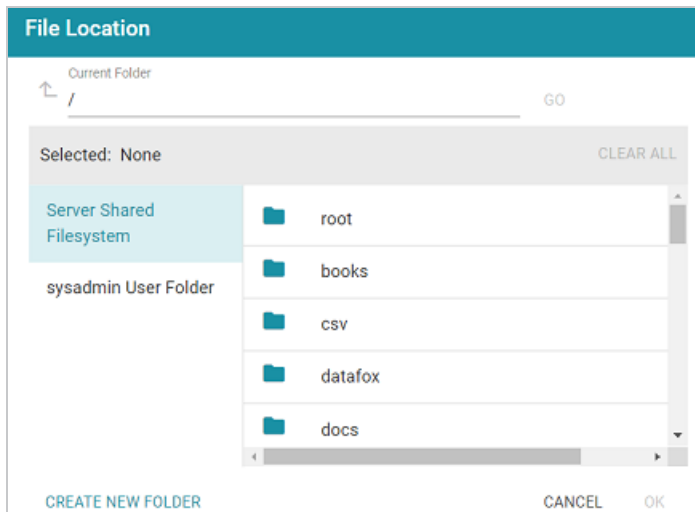
1. In the Administration application, expand the **Servers** menu and click **Volume Manager**. Anzo displays the Volume Manager screen, which lists any existing user-defined volumes (system volumes can be displayed by selecting the system data filter). For example:



2. Click the **Add Volume** button and select **Mount Volume**. Anzo displays the Mount Volume screen.



3. Click the **Path** field to open the File Location dialog box. For example:



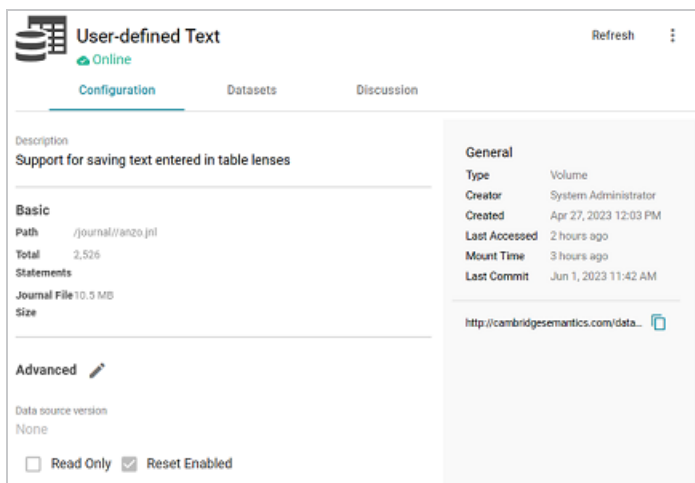
4. On the left side of the screen, select the file store that hosts the volume (.jnl file) that you want to mount. On the right side of the screen, navigate to the .jnl file and select it. Then click **OK**. Anzo mounts the new volume.

# Unmounting a Volume

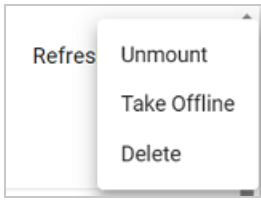
One way to free up system resources when necessary is to unmount unused volumes or take them offline. For example, if users run unstructured pipelines without limiting the creation of status journals (as described in [Limiting the Number of Unstructured Status Journals](#)), there may be dozens of volumes online that are not being queried but are using system threads. If a volume might be needed in the future, you can take it offline to preserve system resources. If a volume will not be used in the future, you can unmount or delete it. If you unmount a volume, you can remount it later if needed (see [Mounting an Existing Volume](#)). If you delete a volume, the volume is removed from disk and cannot be recovered.

Follow the steps below to take offline, unmount, or delete a volume.

1. In the Administration application, expand the **Servers** menu and click **Volume Manager**. The Volume Manager screen displays the list of existing user-defined volumes. To access system volumes, such as unstructured pipeline status journals, click the filter icon and select **Only show system data** in the Filters panel.
2. In the list of volumes, you can delete a volume by clicking the trashcan icon (🗑️) in the row for that volume. Note that **deleting a volume deletes it from disk**. To unmount a volume or take it offline, click the volume that you want to configure. The Configuration screen for that volume is displayed. For example:



3. On the top right of the screen, click the menu icon (⋮) to display the options:



4. Select **Unmount** to unmount the volume but leave it on disk, select **Take Offline** to take the volume offline but leave it mounted, or select **Delete** to delete the volume from disk.

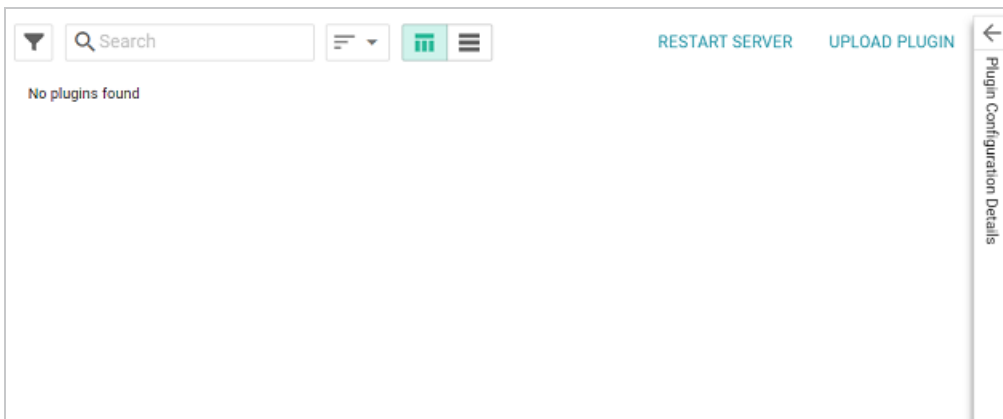
# Uploading a Plugin

When connecting to a relational database to import data, you may need to upload a JDBC driver to Anzo. You may also need to import custom bundles or other bundles received from Cambridge Semantics. This topic provides instructions for uploading executable .jar files from your computer to Anzo.

## Note

Not all .jar files are compatible with Anzo. Custom drivers need to be converted to an OSGI bundle before they are uploaded. For more information and instructions on creating an OSGI bundle from a .jar file, contact your Cambridge Semantics Customer Success manager.

1. In the Administration application, expand the **Servers** menu and click **Plugin Configuration**. Anzo displays the Plugin Configuration screen. For example:



2. In the top right corner, click **Upload Plugin**. The application opens the file browser on your computer.
3. In the file browser, navigate to the .jar file to upload, and then double-click the file to upload it. Anzo uploads the file and displays a "Completed" message. You do not need to restart Anzo to apply the new executable.

# Advanced Semantic Service Configuration

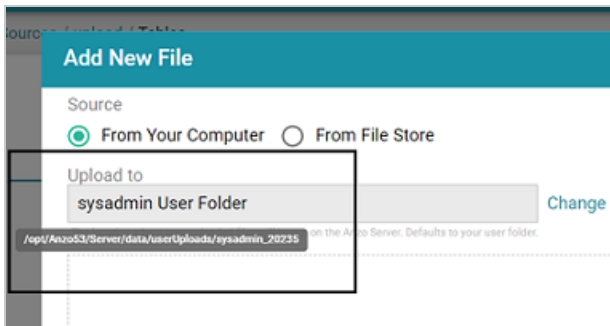
The topics in the section provide instructions for making the types of semantic service or application configuration changes that are commonly desired.

## In this section:

Setting the Default File Upload Path .....	48
Enabling the System Monitor Service .....	50
Routing Hi-Res Analytics to a Custom URL .....	53
Separating Audit Logs by Event Type .....	56
Limiting the Age/Size of Audit Logs .....	57
Limiting the Size/Number of anzo_full Logs .....	59
Configuring a User Inactivity Timeout .....	61
Reporting on Binary Store Access Events .....	63
Setting the Max Page Size for OData Feeds .....	65
Scanning Whole CSV Files on Import .....	67
Including Views as Database Schemas .....	68
Limiting the Number of Unstructured Status Journals .....	69
Disabling Cloud Location Pricing Information .....	71
Setting a Heartbeat for LDAP Connections .....	72

## Setting the Default File Upload Path

By default, if a user uploads a file (such as a CSV, XML, or JSON file) to a data source from their computer, Anzo is configured to copy the file to the server's data directory, `<install_path>/Anzo/Server/data/userUploads`. This is the path that is selected by default in the **Upload To** field on the Add New File screen. For example, the image below shows the default upload path for the sysadmin user:



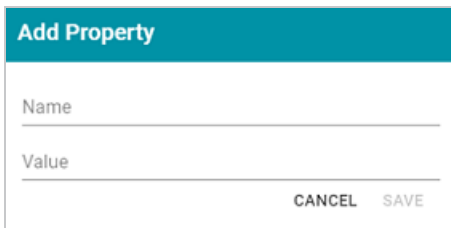
When the file is in the server installation path and not the shared File Store it is not accessible by applications like AnzoGraph. In addition, other users cannot onboard data from that source because they typically do not have access to the file. Source files that are routinely updated and re-ingested should be hosted on the shared File Store.

Follow the instructions below to configure the base upload path so that it points to a location on the File Store by default.

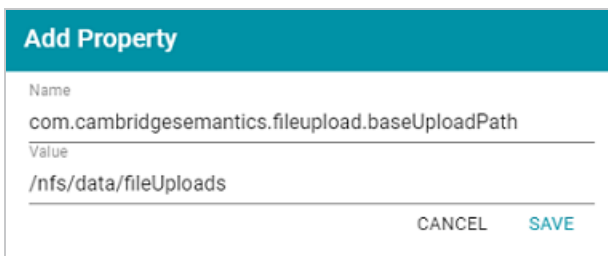
1. If necessary, create a directory on the shared File Store that you can designate as the base location for saving uploaded files.
2. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
3. Search for the **Anzo File Upload** bundle and view its details.
4. Click the **Services** tab and expand the **com.cambridgesemantics.anzo.fileupload.FileUploadServlet** service.



5. Click **Add Property** next to the service name. The Add Property dialog box is displayed.

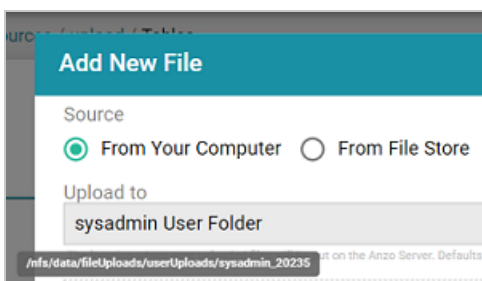


6. In the **Name** field, specify **com.cambridgesemantics.fileupload.baseUploadPath**, and then set the **Value** to the location on the file store where uploaded files should be saved. The base directory that you specify must exist on the file store. For example:



7. Click **Save** to add the new property. And restart Anzo to apply the configuration changes.

When the base upload path is configured, the location that you specified becomes the default path in the Upload To field on the Add New File dialog box. For example, the image below shows the Add New File screen for the sysadmin user when baseUploadPath is set to **/nfs/data/fileUploads**.

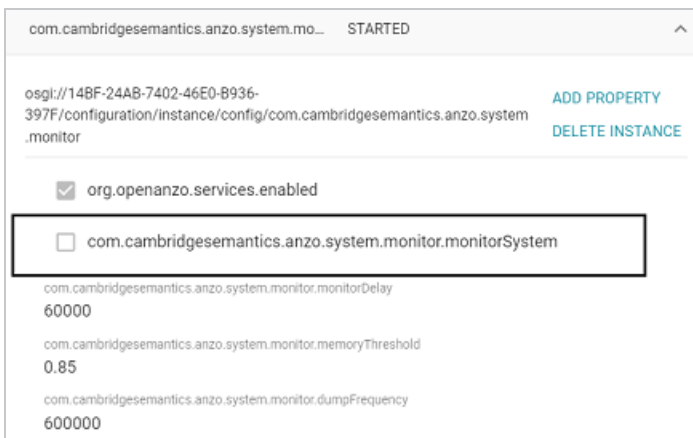


## Enabling the System Monitor Service

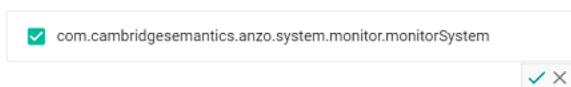
The System Monitor service, which monitors the state of the Java virtual machine (JVM), is disabled by default. You can enable the service to poll the state of the JVM at a certain interval and capture stack and heap dumps when memory utilization increases beyond a specified threshold. This topic provides instructions for enabling the service and configuring its options.

Follow the steps below to enable the System Monitor.

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Anzo System Monitor** bundle and view its details.
3. Click the **Services** tab and expand **System Monitor Activator**.
4. Locate the **com.cambridgesemantics.anzo.system.monitor.monitorSystem** property (shown in the image below).



5. Click the property to make it editable, and then select the checkbox to enable it.



6. Click the checkmark icon (✓) for that property to save the change.
7. Next, configure the service to dump the stack and/or heap logs to disk by enabling the properties under the **monitorSystem** property:

```

com.cambridgesemantics.anzo.system.monitor.dumpFrequency
600000
 com.cambridgesemantics.anzo.system.monitor.produceHeap
com.cambridgesemantics.anzo.system.monitor.heapLocation
${system.ANZO_SERVER_HOME}/logs/system_monitor/heap
com.cambridgesemantics.anzo.system.monitor.maxHeapFiles
-1
 com.cambridgesemantics.anzo.system.monitor.deleteOldHeapFiles
 com.cambridgesemantics.anzo.system.monitor.checkHeapSpace
 com.cambridgesemantics.anzo.system.monitor.produceStack
com.cambridgesemantics.anzo.system.monitor.stackLocation
${system.ANZO_SERVER_HOME}/logs/system_monitor/stack
com.cambridgesemantics.anzo.system.monitor.maxStackFiles
-1
 com.cambridgesemantics.anzo.system.monitor.deleteOldStackFiles

```

To create heap dumps, enable **com.cambridgesemantics.anzo.system.monitor.produceHeap**. To create stack dumps, enable **com.cambridgesemantics.anzo.system.monitor.produceStack**.

8. You can restart Anzo to enable the service without performing additional configuration. Or see [Configure the System Monitor Service](#) below for information about the configuration options.

## Configure the System Monitor Service

By default, the System Monitor Service is configured to monitor memory usage and take the following actions:

- Every **60 seconds** (60000 milliseconds), evaluate whether a stack or thread dump should be written.
- Write stack and/or heap dumps if the memory threshold reaches **85%** (0.85).
- Continue to write stack and/or heap dumps at an interval of every **10 minutes** (600000 milliseconds) as long as memory usage remains at or above the threshold.
- Save heap and stack dumps in the `<install_path>/Server/logs/system_monitor/heap` and `stack` directories.

To modify the characteristics described above, you can change the values for the following properties:

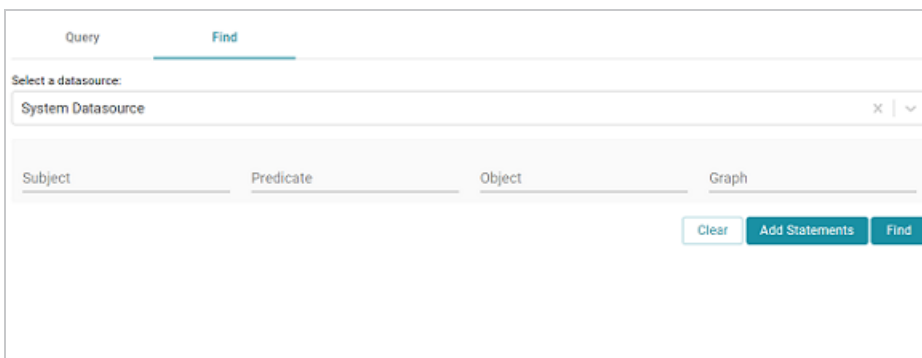
- To change the frequency with which memory usage is evaluated to see if it has reached the threshold, update the **com.cambridgesemantics.anzo.system.monitor.monitorDelay** property. Specify the number of milliseconds to wait between checks.
- To change the memory threshold, update the **com.cambridgesemantics.anzo.system.monitor.memoryThreshold** property. Specify the percent of total memory as a decimal value.
- To change how often stack and/or heap dumps are written when memory usage is above the threshold, update the **com.cambridgesemantics.anzo.system.monitor.dumpFrequency** property. Specify the number of milliseconds to wait between dumps.
- To change the location where heap and/or stack dumps are saved, update the **com.cambridgesemantics.anzo.system.monitor.heapLocation** and/or **com.cambridgesemantics.anzo.system.monitor.stackLocation** property to specify an alternate path and directory.

After changing any of the properties, make sure that you restart Anzo to apply the configuration change.

## Routing Hi-Res Analytics to a Custom URL

If you have a custom skin or personality for the Hi-Res Analytics application, and you want those customizations to be loaded automatically when users access the application, you can configure the Anzo application to re-route users to the preferred URL. Follow the instructions below to change the entry points to the Hi-Res application in the Anzo application. The instructions use the Find feature in the Query Builder to find and modify the object of the Hi-Res Analytics routing property.

1. In the Anzo application, expand the **Access** menu and click **Query Builder**.
2. In the Query Builder, click the **Find** tab. The Find screen is displayed with the **System Datasource** selected as the Source.



The screenshot shows the 'Find' tab in the Query Builder. At the top, there are two tabs: 'Query' and 'Find', with 'Find' being the active tab. Below the tabs, there is a section labeled 'Select a datasource:' with a dropdown menu showing 'System Datasource'. Underneath, there are four input fields: 'Subject', 'Predicate', 'Object', and 'Graph'. At the bottom right, there are three buttons: 'Clear', 'Add Statements', and 'Find'.

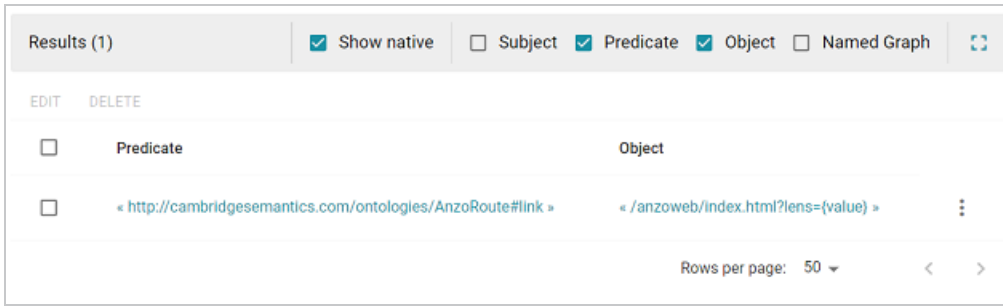
3. In the **Subject** field, specify the following URI:

```
http://cambridgesemantics.com/Routes/sdi/hi-res-analytics-urn
```

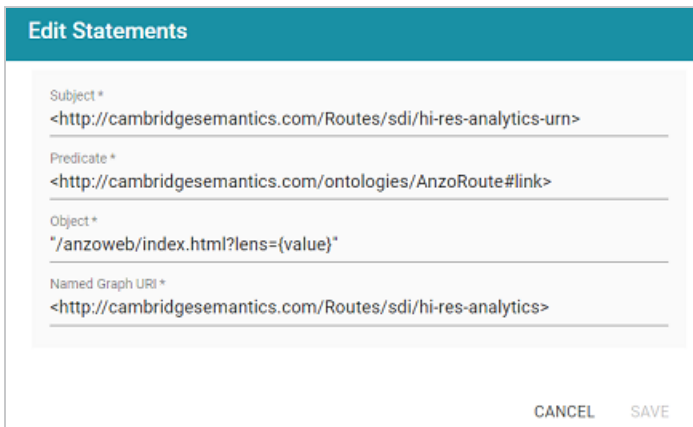
4. In the **Predicate** field, specify this URI:

```
http://cambridgesemantics.com/ontologies/AnzoRoute#link
```

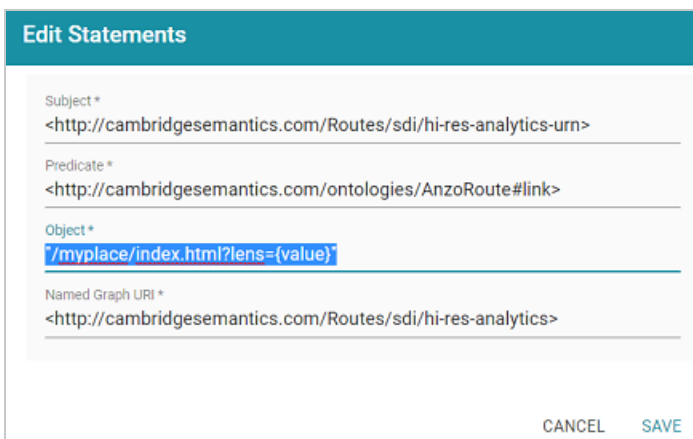
5. Click **Find** to display the quads with the specified subject and predicate. You can clear the **Subject** and **Named Graph** Quick Filter checkboxes to make the results easier to read. For example:



- Click the menu icon (⋮) for the quad and select **Edit**. Anzo opens the Edit Statements dialog box.



- In the Edit Statement dialog box, replace the **Object** value ("`/anzoweb/index.html?lens={value}`") with the URL that you want to route users to. For example: "`/myplace/index.html?lens={value}`".



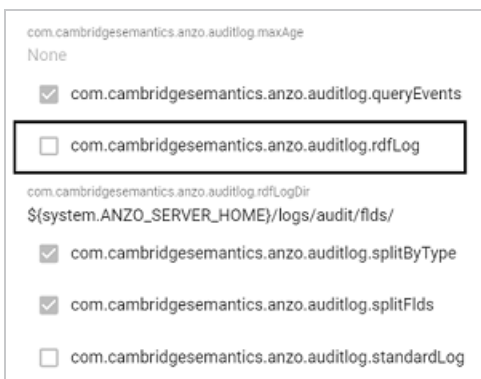
- Click **Save** to apply the change and return to the Find screen.

The Anzo application is now configured to route users to the custom URL if they open the Hi-Res Analytics application from the Home page, open a dashboard from the Hi-Res Analytics screen, or click **Create Dashboard** from a Graphmart screen.

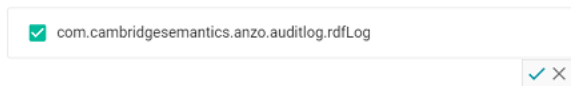
## Separating Audit Logs by Event Type

By default, when Audit Log Packages, such as UserAudit, are enabled and set to Log Level **Info**, all types of audit events are logged to a single file: **anzo\_audit\_info.log**. You have the option, however, to configure Anzo to create and store smaller audit logs by generating separate files in subdirectories that are sorted by event type, such as `userEvents`, `queryEvents`, `accessEvents`, etc. Follow the instructions below to enable this option:

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Anzo Audit Logging Framework** bundle and view its details.
3. Click the **Services** tab and expand **com.cambridgesemantics.anzo.AuditLog**.
4. Find the **com.cambridgesemantics.anzo.auditlog.rdfLog** property (shown below).



5. Click the property to make it editable, and then select the checkbox to enable it.



6. Click the checkmark icon (✓) to save the change.
7. Restart Anzo to apply the configuration changes.

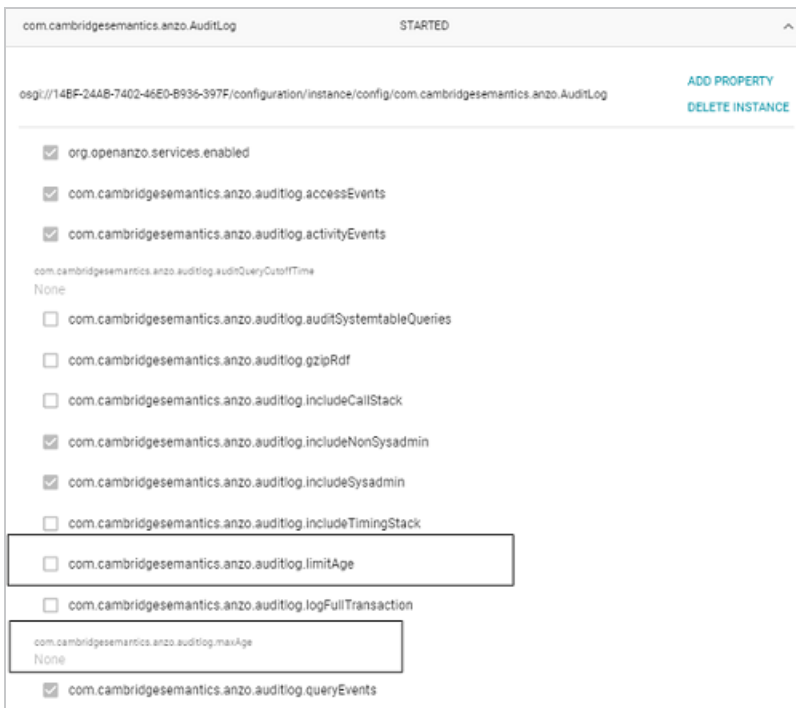
Once new audit events are triggered, an `audit/audit-flds` subdirectory is created in the `<install_path>/Server/logs` directory. And audit logs will be created in the `userEvents`, `queryEvents`, `accessEvents`, etc. subdirectories.



## Limiting the Age/Size of Audit Logs

If you want to retain all of the audit log data but work with smaller data sets when loading and analyzing the log, you can configure Anzo to add an age limit (in days) to audit log data sets. Once an audit log data set reaches that age, Anzo stops writing to it and a new audit log data set is started. Follow the instructions below to configure the audit log service to add an age limit.

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Anzo Audit Logging Framework** bundle and view its details.
3. Click the **Services** tab and expand **com.cambridgesemantics.anzo.AuditLog**.
4. Find the **limitAge** and **maxAge** properties (shown below).



5. Select the **com.cambridgesemantics.anzo.auditlog.limitAge** checkbox to enable the age limit feature.

6. Edit the **com.cambridgesemantics.anzo.auditlog.maxAge** property to specify the maximum number of days to log in each data set. When the current audit log reaches that age, Anzo starts writing to a new data set.
7. Restart Anzo to apply the configuration changes.

## Limiting the Size/Number of anzo\_full Logs

Follow the instructions below if you want to configure the Pax Logging SLF4j Listener Service to limit the size and number of anzo\_full logs that are retained on disk. You can also set a limit on the total size of all anzo\_\* logs.

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Pax Logging SLF4j Listener** bundle and view its details.
3. Click the **Services** tab and expand the **SLF4j Log Listener** service.
4. Find the **maxLogFileSize**, **maxHistory**, and **totalSizeCap** properties (shown below).

```
pax.logging.logback.defaultLevel
ERROR
pax.logging.logback.defaultPattern
%-23.23date %-18.18(%-5level [%-0.10marker]) %22([%-0.20thread]) - %mdco(op
eration,OpName)%mdco(operationId,OpId)%mdco(userUri,OpUser)%mdco(runA
sUser,RunAs)%mdco(userDescription,UserDesc)%logger{32}- %m%n%xEx(full)
pax.logging.logback.logFileNamePattern
anzo_$(errorTag)
pax.logging.logback.maxLogFileSize
50MB
pax.logging.logback.maxHistory
None
pax.logging.logback.totalSizeCap
None
pax.logging.logback.logger.AccessAudit
INFO
pax.logging.logback.logger.ActivityAudit
INFO
```

5. Edit any of the following properties to set them to the desired values:
  - **pax.logging.logback.maxLogFileSize**: This property sets the maximum file size for anzo\_full.log. When the maximum size is reached, Anzo stops writing to that file and creates a new one.
  - **pax.logging.logback.maxHistory**: This property specifies the maximum number of historical anzo\_full.log files to keep. When this limit is reached, Anzo deletes the oldest file.
  - **pax.logging.logback.totalSizeCap**: This property sets the total size limit for all anzo\_\* log files combined.

6. After editing a property, click the checkmark icon (✓) for that property to save the change.
7. Restart Anzo to apply the configuration changes.

## Configuring a User Inactivity Timeout

By default, the user inactivity timeout setting in the Anzo Java Script Runtime Assembler service is set to **unlimited**, meaning Anzo will not automatically log out users who have a session open but remain inactive. If you want to configure Anzo to log users out if they are inactive for a period of time, follow the instructions below.

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Anzo Java Script Runtime Assembler** bundle and view its details.
3. Click the **Services** tab and expand the **Anzo Java Script Runtime Assembler** service.
4. Edit the `com.cambridgesemantics.anzowt.runtimeassembler.inactivityLogoutTimeout` property (shown in the image below) to specify the number of **milliseconds** that a user can remain inactive before being logged out.



For example, setting the value to **900000** milliseconds means that a user who is inactive for more than 15 minutes is automatically logged out.



5. After specifying the value, click the checkmark icon (✓) for that property to save the change.
6. Restart Anzo to apply the configuration change.

**Note**

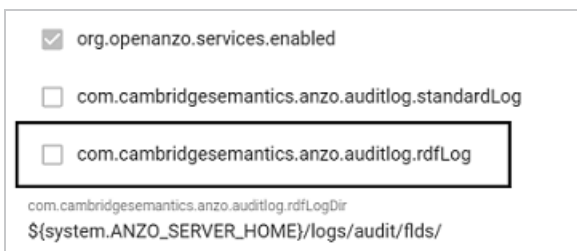
By default, Anzo is not configured to log an event when the user inactivity value is changed. If you would like this event to be noted in the Audit log when the setting is changed, see

[Enabling the Audit Logs](#) for instructions.

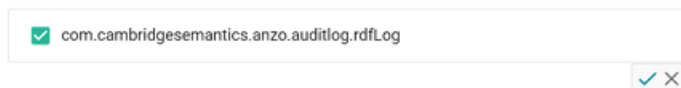
## Reporting on Binary Store Access Events

By default, binary store access events are not captured in the Audit log. You can configure the audit logging framework to capture information about binary store requests, however. Data such as the time of the request, the user who made the request, and the document that was accessed will be captured. Follow the instructions below to configure the log to report on binary store events.

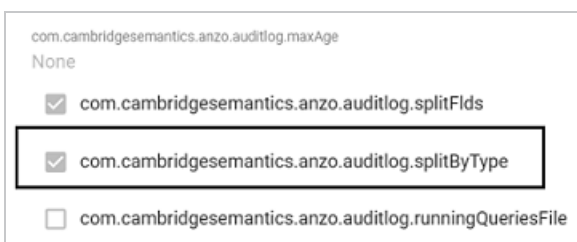
1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Anzo Audit Logging Framework** bundle and view its details.
3. Click the **Services** tab and expand **com.cambridgesemantics.anzo.AuditLog**.
4. Locate the **com.cambridgesemantics.anzo.auditlog.rdfLog** property (shown in the image below).



5. Click the property to make it editable, and then select the checkbox to enable it.



6. Click the checkmark icon (✓) for that property to save the change.
7. Scroll down and make sure that the **com.cambridgesemantics.anzo.auditlog.splitByType** property is selected/enabled (it is enabled by default).



8. Restart Anzo to apply the configuration change.

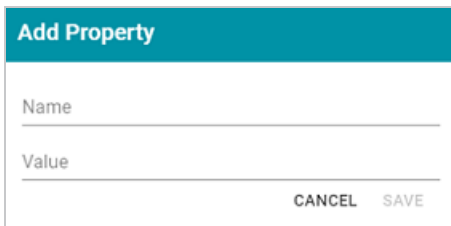
New binary store access audit events will be added to the logs in the subdirectories under `<install_path>/Server/logs/audit/audit-flds`.



## Setting the Max Page Size for OData Feeds

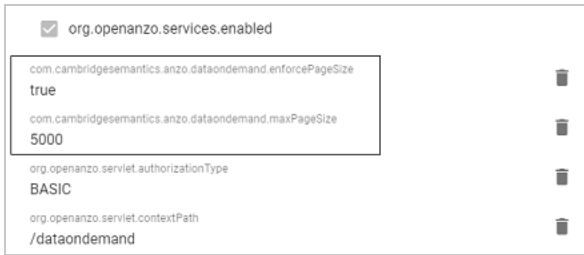
When a user sends a request to an Anzo Data on Demand endpoint, they do not necessarily know the total number of results that will be returned. In some cases, the result set can be hundreds of millions of values, and the request times out before the results can be returned. You can configure the Data on Demand service to specify a maximum limit on the number of results that can be returned for a single OData feed request. If a user sends a request and the result set is larger than the maximum value, Anzo will limit the results to the configured maximum value. Follow the instructions below to configure the Data on Demand service to enforce a maximum page size.

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Anzo DataOnDemand** bundle and view its details.
3. Click the **Services** tab and expand **DataOnDemandServiceActivator**.
4. Click **Add Property** next to the service name. The Add Property dialog box is displayed.



The image shows a dialog box titled "Add Property". It contains two text input fields, one labeled "Name" and one labeled "Value". At the bottom right of the dialog, there are two buttons: "CANCEL" and "SAVE".

5. In the **Name** field, specify **com.cambridgesemantics.anzo.dataondemand.enforcePageSize**, and set the **Value** to **true**. Then click **Save**.
6. Click **Add Property** again. In the **Name** field, specify **com.cambridgesemantics.anzo.dataondemand.maxPageSize**, and set the **Value** to the maximum number of results that to return per request. Then click **Save**. The two settings are displayed on the Services screen. For example:



7. Restart Anzo to apply the configuration changes.

## Scanning Whole CSV Files on Import

To help improve accuracy of data type assignment when importing CSV files, you have the option to configure the system so that any time a CSV file is imported, Anzo scans the entire file before inferring the data types for each column. Follow the instructions below if you want to configure the system to scan entire CSV files.

### Important

This change affects all CSV file imports. Users cannot opt-out of a complete scan at import time. This configuration is not related to the **Use Extended Sample** setting in the edit file options. Choosing to scan entire files will significantly increase the time it takes to import files. However, scanning the complete file is the best way to ensure that data type assignments are accurate.

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Anzo Utilityservices VFS** bundle and view its details.
3. Click the **Services** tab and expand **UtilityServices VFS Activator**.
4. Find the **com.cambridgesemantics.anzo.utilityservices.vfs.isSampleEntireFile** property, and select the checkbox to enable the option.

### Note

When **SampleEntireFile** is enabled, the values in the **maxSampleSize** and **sampleSize** properties are ignored and Anzo always scans entire CSV files on import.

5. Restart Anzo to apply the configuration changes.

## Including Views as Database Schemas

By default, when you create a database data source and import a predefined schema, views are excluded from the list of schemas that are available to import. However, you can configure the Anzo Database DataSource Provider Service to include views as schemas. Follow the steps below to remove views from the list of table types that are excluded from import.

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Anzo Database DataSource Provider** bundle and view its details.
3. Click the **Services** tab and expand **com.cambridgesemantics.anzo.database.IDbConnectionService**.
4. Locate the **com.cambridgesemantics.anzo.database.excludeTableTypes** property (shown in the image below).



5. Click the property to make it editable, and then delete the word **VIEW**.



6. Click the checkmark icon (✓) for that property to save the change.
7. Restart Anzo to apply the configuration change.

The service is now configured to display views in the Import Schemas dialog box.

## Limiting the Number of Unstructured Status Journals

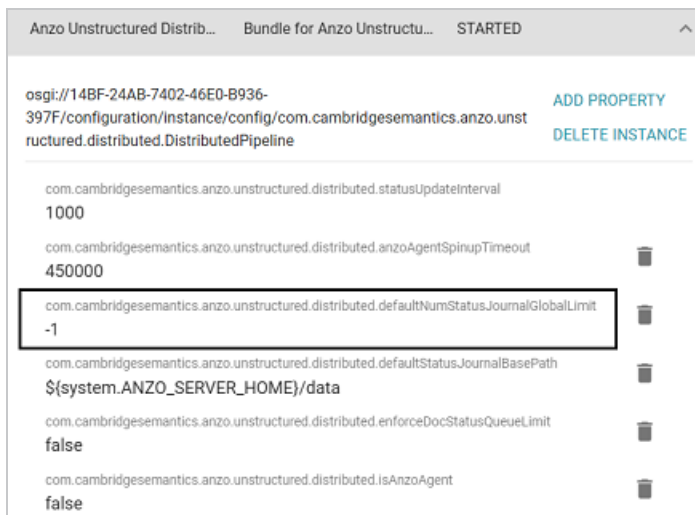
To limit the disk space used by Anzo Unstructured pipelines, you have the option to configure the Anzo Unstructured Distributed service to limit the number of status journals that are preserved on disk. When the specified limit is reached and a pipeline generates a new journal, the oldest journal is deleted.

### Note

Journals are removed based on their timestamps alone. The pipeline they are associated with is not a factor in determining the journals to delete.

Follow the instructions below to configure the Unstructured Distributed service to limit the number of status journals on disk.

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Anzo Unstructured Distributed** bundle and view its details.
3. Click the **Services** tab and expand **Anzo Unstructured Distributed**.
4. Locate the **com.cambridgesemantics.anzo.unstructured.distributed.defaultNumStatusJournalGlobalLimit** property (shown in the image below).



5. Click the property to make it editable. Then replace the current value with the maximum number of status journals to keep on disk. The default value is -1 (unlimited). For example, in the image below, the value is set to keep 10 status journals.



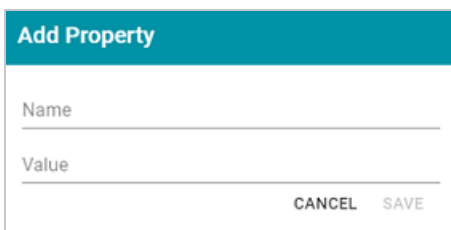
The image shows a configuration field with the property name `com.cambridgesemantics.anzo.unstructured.distributed.defaultNumStatusJournalGlobalLimit` and the value `10`. A checkmark icon and an 'x' icon are visible at the bottom right of the field.

6. After changing the value, click the checkmark icon (✓) for that property to save the change.
7. Restart Anzo to apply the configuration change.

## Disabling Cloud Location Pricing Information

By default, pricing information from the cloud service providers is retrieved and displayed for the Cloud Locations that you configure in Anzo. To disable the retrieval of pricing information, follow the steps below.

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Cloud Infrastructure** bundle and view its details.
3. Click the **Services** tab and expand **com.cambridgesemantics.anzo.cloud.k8s.deployment.KubernetesDeploymentService**.
4. Click **Add Property** next to the service name. The Add Property dialog box is displayed.



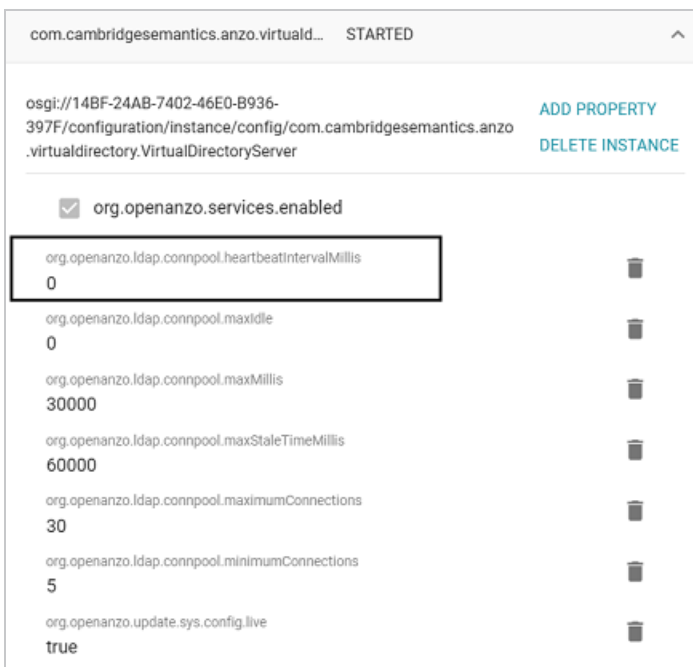
The image shows a dialog box titled "Add Property". It contains two text input fields, one labeled "Name" and one labeled "Value". At the bottom right of the dialog, there are two buttons: "CANCEL" and "SAVE".

5. In the **Name** field, specify **com.cambridgesemantics.anzo.cloud.k8s.deployment.pricinginfo.disable**, and set the **Value** to **true**. Then click **Save**.
6. Restart Anzo to apply the configuration changes.

## Setting a Heartbeat for LDAP Connections

To keep connections to a directory server alive and avoid timing out while authenticating users, you can configure a heartbeat to maintain the connection from Anzo. Follow the instructions below to set up the heartbeat.

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Anzo Enterprise Directory Connect** bundle and view its details.
3. Click the **Services** tab and expand **com.cambridgesemantics.anzo.virtualdirectory.VirtualDirectoryServer**.
4. Locate the **org.openanzo.ldap.connpool.heartbeatIntervalMillis** property (shown in the image below).



5. Click the property to make it editable. Then replace the current value (0 by default) with the number of milliseconds to set the heartbeat interval. Depending on the network, Cambridge Semantics recommends that you set the interval to between 1 and 5 minutes (60000 to 300000 milliseconds). For example, the image below sets the interval to 3 minutes.



org.openanzo.ldap.connpool.heartbeatIntervalMillis  
180000



6. Click the checkmark icon (✓) for that property to save the change.
7. You can edit other properties as desired. When you have finished making changes, restart Anzo to apply the configuration changes.

# Connection Administration

The topics in this section provide information about managing connections to the Anzo server.

**In this section:**

Connecting to a File Store .....	75
Creating an Anzo Data Store .....	85
Connecting to AnzoGraph .....	89
Connecting to Elasticsearch .....	98
Connecting to a Distributed Unstructured Cluster .....	101
Connecting to a Cloud Location .....	104

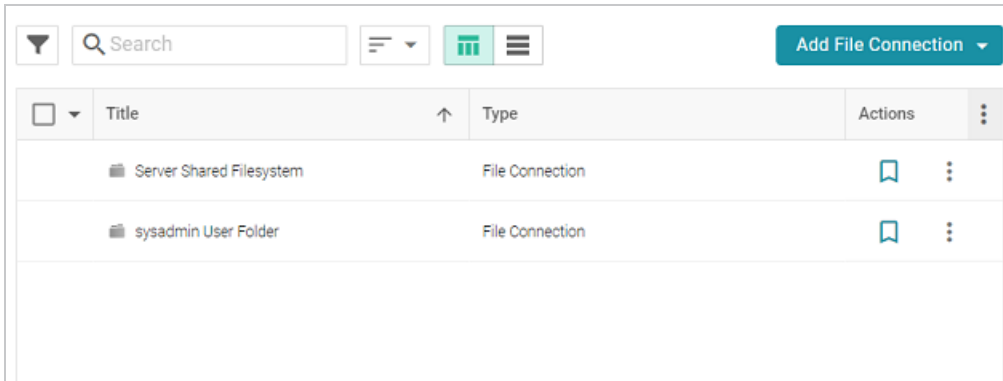
## Connecting to a File Store

When the Anzo Server is initially installed, the default file store (typically an NFS mount) is configured. This topic provides instructions for connecting to an additional file storage system. The supported systems are local file systems (such as a mounted NFS), Hadoop Distributed File Systems (HDFS), File Transfer Protocol (FTP or FTPS) systems, Google Cloud Platform (GCP) storage, and Amazon Simple Cloud Storage Service (S3).

- [Connecting to File Storage](#)
- [Connection Settings Reference](#)

## Connecting to File Storage

1. In the Administration application, expand the **Connections** menu and click **File Store**. Anzo displays the File Store screen, which lists existing file store connections. For example:

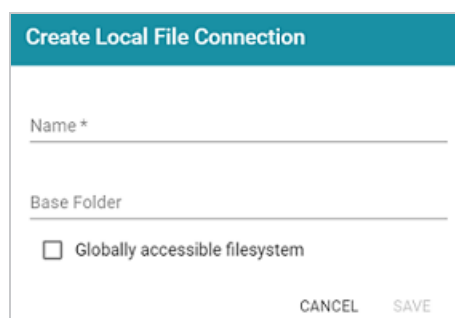


2. Click the **Add File Connection** button and select the type of file connection that you want to create. For the local disk or mounted NFS, choose **Local File Connection**. Anzo displays the create connection screen for the type of connection you chose.
3. On the connection screen, provide the file system details. The settings that display depend on the type of file connection that you chose. See [Connection Settings Reference](#) below for details.
4. Click **Save** to save the configuration. The file store connection that you specified becomes available as a choice when you create Anzo Data Stores or select source files to onboard, etc.

## Connection Settings Reference

- [Local File Connection](#)
- [HDFS File Connection](#)
- [FTP or FTPS File Connection](#)
- [Google Cloud Platform File Connection](#)
- [S3 File Connection](#)

### Local File Connection



Create Local File Connection

Name\*

Base Folder

Globally accessible filesystem

CANCEL SAVE

Setting	Description
<b>Name</b>	The name to use for this file connection.
<b>Base Folder</b>	The base or root folder on the file system where you want Anzo to read or write files. Each time Anzo generates new files it creates a new subdirectory under this base location.
<b>Globally accessible filesystem</b>	Select this option if this file store is accessible by all of the servers in an AnzoGraph cluster. If only the leader server can access this system, leave this option blank.

## HDFS File Connection

### Important

If you use Kerberos Authentication with HDFS, you must also configure the AnzoGraph cluster to authenticate with Kerberos. For instructions, see [Configuring AnzoGraph for Kerberos Authentication](#).

The screenshot shows a dialog box titled "Create HDFS File Connection". It contains the following fields and controls:

- Name \***: A text input field.
- Nameservice IP or Name \***: A text input field.
- Port**: A text input field.
- Base Folder**: A text input field.
- HDFS Configuration Path**: A text input field with a **BROWSE** button to its right.
- Keytab Path**: A text input field with a **BROWSE** button to its right.
- CANCEL** and **SAVE** buttons are located at the bottom right of the dialog.

Setting	Description
<b>Name</b>	The name to use for this file connection.
<b>Nameservice IP or Name</b>	The IP address or host name for the storage system.
<b>Port</b>	The RPC port to access the server on. The default RPC port is 8020.
<b>Base Folder</b>	The base or root folder on the file system where you want Anzo to read or write files. Each time Anzo generates new files it creates a new subdirectory under this base location.

Setting	Description
<b>HDFS Configuration Path</b>	The full path to the configuration files.
<b>Keytab Path</b>	The full path to the keytab file.
<b>Password Confirm Password</b>	The password for the account used to access the server.
<b>Nameservice Rest IP or Name</b>	The HTTP REST IP address or host name. Typically this value is the same as the <code>Nameservice IP or Name</code> .
<b>Nameservice Rest Port</b>	The HTTP port. AnzoGraph uses this port to access HDFS and load the FLDS. The default HTTP port for the namenode is 9870.
<b>Nameservice Rest Protocol</b>	The protocol to use for requests. Choose one of the following values: <ul style="list-style-type: none"> <li>• <b>hdfs</b>: Specify <code>hdfs</code> for non-secure HTTP protocol.</li> <li>• <b>shdfs</b>: Specify <code>shdfs</code> for secure HTTPS protocol.</li> <li>• <b>khdfs</b>: Specify <code>khdfs</code> for non-secure HTTP protocol with Kerberos authentication.</li> <li>• <b>kshdfs</b>: Specify <code>kshdfs</code> for secure HTTPS protocol with Kerberos authentication.</li> </ul>
<b>Globally accessible filesystem</b>	Select this option if this file store is accessible by all of the servers in an AnzoGraph cluster. If only the leader server can access this system, leave this option blank.

## FTP or FTPS File Connection

### Create FTPS File Connection

Name \*

Server IP or Name \*

Port

Base Folder

Username

Password

Confirm Password

Keystore Path [BROWSE](#)

Globally accessible filesystem

CANCEL SAVE

Setting	Description
<b>Name</b>	The name to use for this file connection.
<b>Server IP or Name</b>	The IP address or host name for the storage system.
<b>Port</b>	The port to access the server on.
<b>Base Folder</b>	The base or root folder on the file system where you want Anzo to read or write files. Each time Anzo generates new files it creates a new subdirectory under this base location.
<b>Username</b>	The user name for the account used to access the server.
<b>Password</b>	The password for the account used to access the server.



Setting	Description
<b>Confirm Password</b>	
<b>Keystore Path</b>	For FTPS connections, the full path to the keystore file.
<b>Globally accessible filesystem</b>	Select this option if this file store is accessible by all of the servers in an AnzoGraph cluster. If only the leader server can access this system, leave this option blank.

## Google Cloud Platform File Connection

Create Google Cloud Platform File Connection

Name \*

---

Bucket Name \*

---

Base Folder

---

Account Email

---

Key File Location BROWSE

---

Globally accessible filesystem

HTTP Proxy Url

---

HTTPs Proxy Url

---

CANCEL
SAVE

Setting	Description
<b>Name</b>	The name to use for this file connection.
<b>Bucket Name</b>	The name of the bucket to store files in.

Setting	Description
<b>Base Folder</b>	The base or root folder on the file system where you want Anzo to read or write files. Each time Anzo generates new files it creates a new subdirectory under this base location.
<b>Account Email</b>	The email address for the account used to access the storage.
<b>Key File Location</b>	The full path to the keystore password file.
<b>Globally accessible filesystem</b>	Select this option if this file store is accessible by all of the servers in an AnzoGraph cluster. If only the leader server can access this system, leave this option blank.

## S3 File Connection

### Important

When using Amazon S3 for file storage, do not use client-side encryption, where data is encrypted before it is sent to S3. Anzo cannot read files on S3 if the object store uses client-side encryption.

**Create S3 File Connection**

Name \*

---

Bucket Name \*

---

Base Folder

---

Access Key

---

Secret Key 👁

---

Confirm Secret Key 👁

---

S3 URI Scheme | v

---

Globally accessible filesystem

CANCEL   SAVE

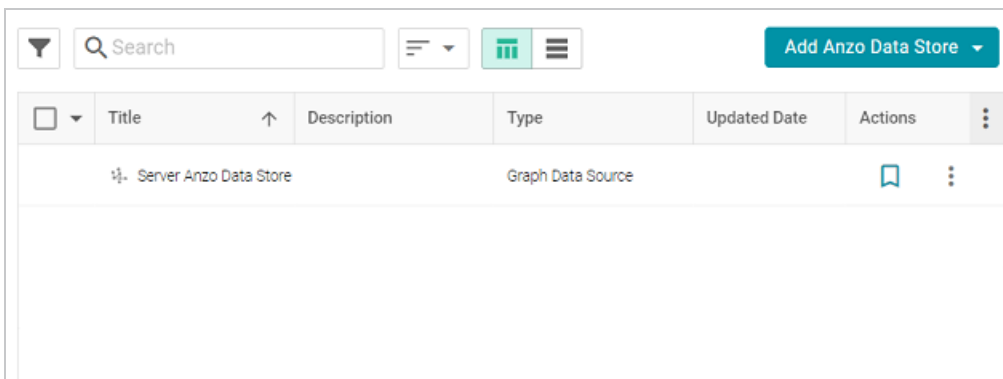
Setting	Description
<b>Name</b>	The name to use for this file connection.
<b>Bucket Name</b>	The name of the bucket to store files in.
<b>Base Folder</b>	The base or root folder on the file system where you want Anzo to read or write files. Each time Anzo generates new files it creates a new subdirectory under this base location.
<b>Access Key</b>	The Access Key ID to use for accessing the S3 location.
<b>Secret Key Confirm Secret Key</b>	The Secret Key ID for the Access Key.
<b>S3 URI Scheme</b>	Specifies whether the URI scheme is S3, S3 Native, or S3A.

Setting	Description
<b>Globally accessible filesystem</b>	Select this option if this file store is accessible by all of the servers in an AnzoGraph cluster. If only the leader server can access this system, leave this option blank.

# Creating an Anzo Data Store

This topic provides instructions for creating an Anzo Data Store. A data store is a directory on a shared file store where file-based linked data sets can be written by Anzo. If you onboard unstructured data, a data store is required. In addition, a data store is required if you use the automated direct load workflow and configure the workflow to export the data to a dataset. You can create one data store and configure all pipelines and workflows to write to that store or you can create multiple data stores to use for different datasets.

1. In the Administration application, expand the **Connections** menu and click **Anzo Data Store**. Anzo displays the Anzo Data Store screen, which lists any existing data stores. For example:



## Important

The **Server Anzo Data Store** is a default data store that points to the local Anzo file system. This store exists so that first-time users can quickly test the onboarding process. It is not meant to be used in production. Do not change the Data Location to a shared file store; reconfiguring this Data Store can cause unexpected consequences when upgrading or migrating the system. It is safe to delete this store so that it is not presented as an option when users configure ingestion pipelines.

2. On the Anzo Data Store screen, click the **Add Anzo Data Store** button and select **Add Anzo Data Store**. Anzo opens the Create Anzo Data Store screen.

3. Type a **Title** and optional **Description** for the data store.
4. Click in the **Data Location** field. Anzo opens the File Location dialog box.

5. On the left side of the screen, select the File Store on which to create this data store. On the right side of the screen, navigate to the directory that you want to designate as the data location. Select a directory, and then click **OK**. Or click **Create New Folder** to create a new directory. Each time a pipeline is run for this data store, a new subdirectory is created under the specified data location.

### Note

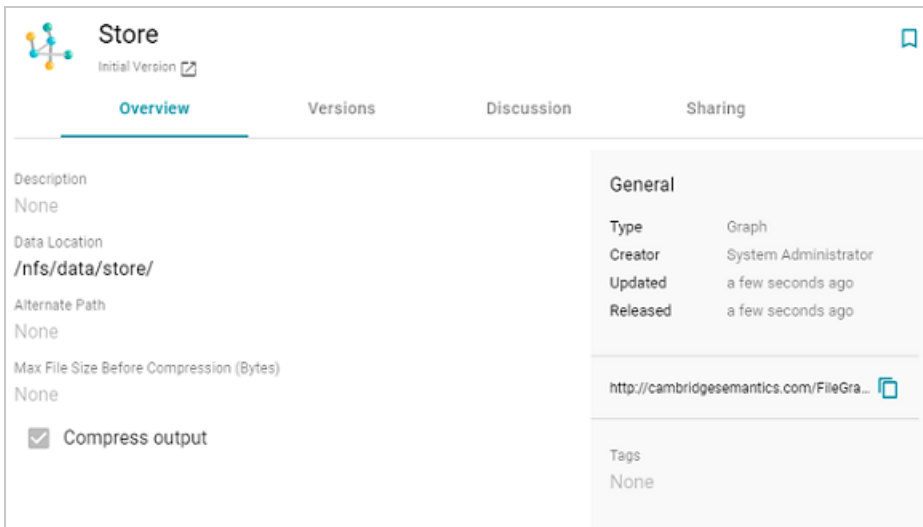
The Data Location needs to be a directory on the file store that is shared between Anzo, AnzoGraph, and any Anzo Unstructured, or Elasticsearch servers. If you want Anzo to generate files for this data store in one location and then load the files into AnzoGraph from another location, specify the file generation location in this field, and then specify the AnzoGraph load location in the **Alternate Data Location** field that is displayed on the Details screen after you save the data store.

6. If necessary, you can modify the maximum limit for the size of the files that are created by pipelines that write to this data store by specifying the size (in bytes) in the **Max File Size Before Compression (Bytes)** field. The value applies to files before they are compressed.

### Note

Cambridge Semantics recommends that you do not set this value unless instructed to do so by Cambridge Semantics Support.

7. Specify whether to compress the generated load files. By default, the **Compress output** checkbox is selected, indicating that Anzo generates .ttl.gz files when writing to this graph data source. If you clear the checkbox, Anzo generates uncompressed .ttl files. To preserve disk space and reduce read times when loading data into memory, Cambridge Semantics recommends that you accept the default configuration and compress load files.
8. Click **Save** to create the data store. Anzo saves the configuration and displays the details view. For example:



You can click the Edit icon (✎) to modify any of the options. Click the check mark icon (✓) to save changes to an option, or click the X icon (✕) to clear the value for an option.

9. If you plan to load files into AnzoGraph from a location that is different than the **Data Location** that you specified, edit the **Alternate Data Location** field and select the location for AnzoGraph load files.

Once you have create the new data store, you can designate it as the default store so that it is automatically selected when users set up data onboarding workflows. See [Set the Default Anzo Data Store](#) for instructions.



# Connecting to AnzoGraph

This topic provides instructions for configuring the connection to a static AnzoGraph instance. It also includes reference information for the Advanced connection settings. For information about installing AnzoGraph, see [Deploying a Static AnzoGraph Cluster](#) in the Deployment Guide.

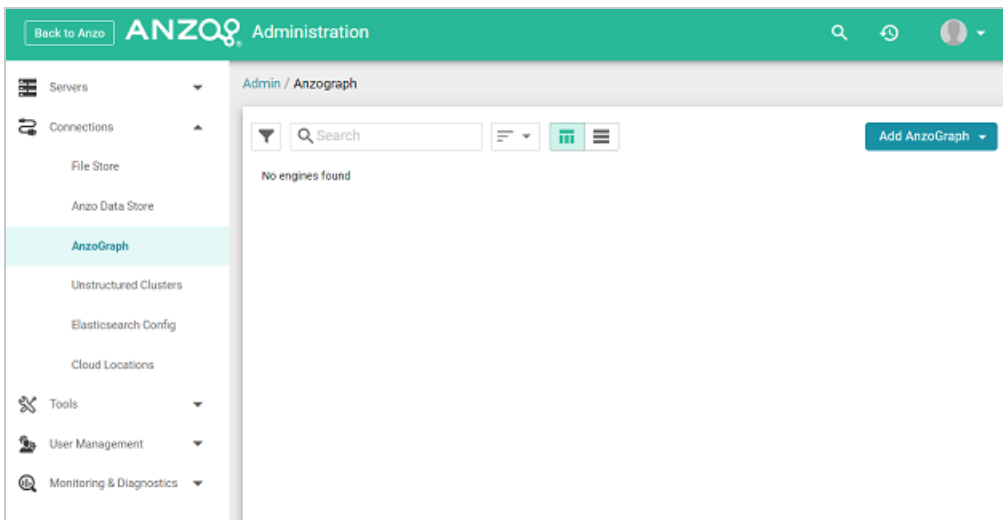
## Important

Do not connect multiple Anzo instances to the same AnzoGraph instance. Since AnzoGraph is stateless and Anzo manages all of the data, connecting more than one Anzo instance to the same AnzoGraph instance causes severe data management conflicts that result in unexpected behavior. This type of configuration is not supported.

- [Creating the Connection](#)
- [Advanced Settings](#)

## Creating the Connection

1. In the Administration application, expand the **Connections** menu and click **AnzoGraph**. Anzo opens the AnzoGraph connection overview screen, which lists any existing connections. For example:



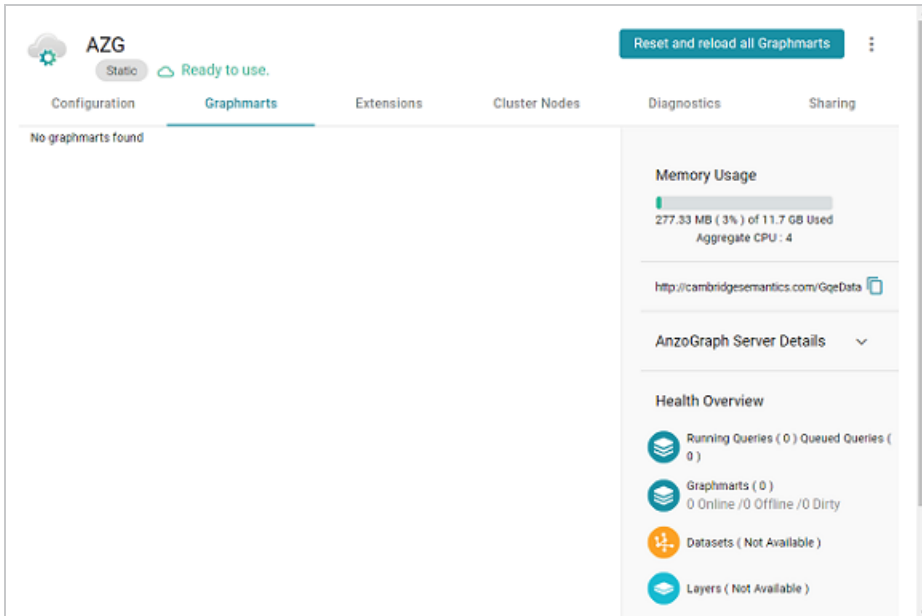
2. On the AnzoGraph screen, click **Add AnzoGraph** and select **Add AnzoGraph** from the drop-down list. Anzo displays the Create AnzoGraph dialog box.

The screenshot shows the 'Create AnzoGraph' dialog box with the following fields and controls:

- Header: Create AnzoGraph
- Tabs: Basic (selected), Advanced
- Fields: Title \*, Description, Host \*, AnzoGraph User \*, AnzoGraph Password \*, Confirm AnzoGraph Password \*
- Dropdown: Elasticsearch Configuration
- Buttons: TEST CONNECTION, CANCEL, SAVE

3. On the Basic tab, type a name for the instance in the **Title** field.
4. In the optional **Description** field, type a description for the instance. If you leave this field blank, Anzo creates a description when you save the configuration.
5. In the **Host** field, type the AnzoGraph server host name or IP address. If you have a cluster, type the name or IP address of the leader server.
6. In the **AnzoGraph User** field, type the Admin username that was created when AnzoGraph was installed.
7. Type the password for the AnzoGraph user in the **AnzoGraph Password** and **Confirm Password** fields.
8. If this AnzoGraph instance will host data associated with Elasticsearch, click the **Elasticsearch Configuration** drop-down list and select the Elasticsearch instance to use with this AnzoGraph connection. For information about configuring an Elasticsearch connection, see [Connecting to Elasticsearch](#).
9. Click **Test Connection** to check if Anzo can connect to AnzoGraph. If the connection fails, make sure that AnzoGraph is running and that you typed the correct username and password.

10. **Optional:** Click the **Advanced** tab and configure any of the optional advanced settings. For details about the Advanced settings, see [Advanced Settings](#) below.
11. Click **Save** to save the configuration. Anzo connects to AnzoGraph and opens the Graphmarts tab. For example:



To change configuration details, click the **Configuration** tab and adjust values as needed. The right side of the screen shows connection status as well as memory usage details, overall data statistics, and graphmart details.

# Advanced Settings

This section describes the connection settings that are available on the Advanced tab when adding a static AnzoGraph connection or the Configuration tab when editing an existing connection.

**Create AnzoGraph**

Basic    **Advanced**

Instance URI

Trust All TLS Certificates

AnzoGraph Concurrent Queries  
10

AnzoGraph connection timeout (seconds)  
60

Use AnzoGraph persistence if available

Force reload of Graphmart data during AnzoGraph activation or reconnection.

Keep AnzoGraph Datasource enabled on Anzo startup.

Port  
5700

AnzoGraph Management Port  
5600

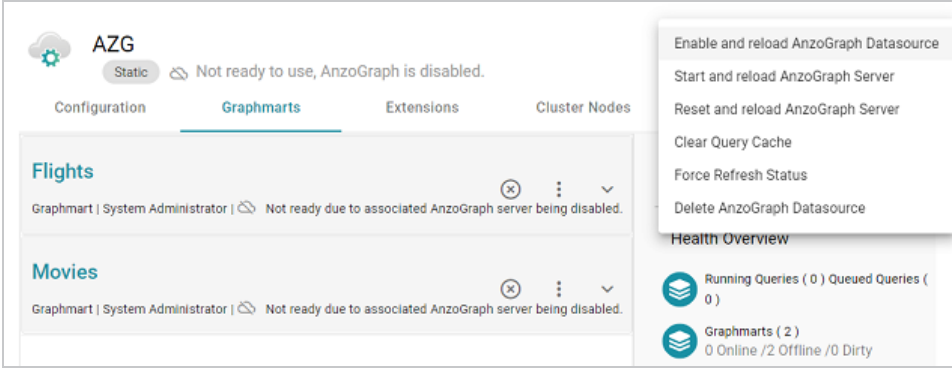
Callback HostName

Readonly Replica     Vacuum     Gather Statistics on Load

TEST CONNECTION    CANCEL    SAVE

Setting	Description
<b>Instance URI</b>	Defines the URI for this instance. When this setting is empty Anzo automatically assigns an instance URI. If you specify a custom URI, make sure that the URI is valid and unique.
<b>Trust All TLS Certificates</b>	Indicates whether Anzo should trust the AnzoGraph certificates for this connection. Cambridge Semantics recommends that you accept the default value of enabled.
<b>AnzoGraph Concurrent Queries</b>	Specifies the maximum number of queries that Anzo can send to AnzoGraph concurrently. The default value is <b>10</b> queries. Cambridge Semantics recommends that you accept the default value. If you want to increase the number of concurrent queries, Cambridge Semantics recommends that you

Setting	Description
	choose a value between 10 and 20.
<b>AnzoGraph Connection Timeout</b>	Controls how often (in seconds) Anzo checks the status of the connection to this instance. The connection is tested every <i>N</i> seconds, where <i>N</i> is the value of this setting. The default value is <b>60</b> . If the test fails, Anzo re-tests the connection every 15 seconds for 2 minutes to rule out a brief network glitch. If the connection continues to fail after 2 minutes, the status is changed to "Offline." If the connection is re-established within the 2-minute window, Anzo determines whether the connection came back automatically or whether AnzoGraph was restarted.
<b>Use AnzoGraph Persistence if Available</b>	Controls how Anzo manages graphmart data if persistence is enabled and AnzoGraph is restarted. The <b>Use AnzoGraph Persistence if Available</b> setting is enabled by default but persistence is disabled for AnzoGraph by default. For information about how Anzo manages the data when persistence is enabled and for instructions on enabling persistence, see <a href="#">Enabling Persistence (Preview)</a> .
<b>Force Reload of Graphmart Data During AnzoGraph Activation or Reconnection</b>	<p>This option is enabled by default and means that Anzo forces a reload of active graphmarts when one of the following actions occur:</p> <ol style="list-style-type: none"> <li>1. Anzo restarts and reconnects to AnzoGraph.</li> <li>2. Anzo restarts and a user manually re-enables this data source by selecting <b>Enable and reload AnzoGraph Datasource</b> from the menu on the AnzoGraph administration screen.</li> </ol> <p>When this option is disabled and AnzoGraph persistence is also disabled, graphmarts must be reloaded by clicking the <b>Reset and Reload all Graphmarts</b> button on the AnzoGraph screen after the connection is re-established. If persistence is enabled and <b>Force reload of Graphmart data...</b> is disabled, Anzo may force a reload if the last updated timestamp in AnzoGraph does not match the last updated value in Anzo.</p>

Setting	Description
<b>Keep AnzoGraph Datasource Enabled on Anzo Startup</b>	<p>This option is enabled by default and means that Anzo leaves this instance in a "Ready to use" state if Anzo is restarted (if this data source is online at the time Anzo is restarted). When this option is disabled, Anzo disables this data source when Anzo is restarted. When Anzo comes online, this source must be manually enabled by selecting <b>Enable and reload AnzoGraph Datasource</b> from the menu on the AnzoGraph administration screen. For example:</p>  <p>The screenshot shows the AnzoGraph administration interface. At the top, there's a header with 'AZG' and a status indicator 'Static' and a message 'Not ready to use, AnzoGraph is disabled.' Below this are tabs for 'Configuration', 'Graphmarts', 'Extensions', and 'Cluster Nodes'. The 'Graphmarts' tab is active, showing two graphmarts: 'Flights' and 'Movies'. Both are in a 'Not ready' state due to the associated AnzoGraph server being disabled. A context menu is open over the 'Flights' graphmart, listing several actions: 'Enable and reload AnzoGraph Datasource', 'Start and reload AnzoGraph Server', 'Reset and reload AnzoGraph Server', 'Clear Query Cache', 'Force Refresh Status', and 'Delete AnzoGraph Datasource'. At the bottom right, there's a 'Health Overview' section showing 'Running Queries ( 0 )', 'Queued Queries ( 0 )', and 'Graphmarts ( 2 )' with a sub-status of '0 Online / 2 Offline / 0 Dirty'.</p>
<b>Port</b>	<p>The (gRPC) port for secure communication between AnzoGraph and Anzo. The default value is <b>5700</b>. Do not change the value unless instructed to do so by Cambridge Semantics Support.</p>
<b>AnzoGraph Management Port</b>	<p>The SSL system management port for AnzoGraph. It is the port that Anzo uses to connect to the system manager. The default value is <b>5600</b>. Do not change the value unless instructed to do so by Cambridge Semantics Support.</p>
<b>Callback HostName</b>	<p>The Anzo instance to call when AnzoGraph makes service callbacks. If you have multiple Anzo servers and one or more of them are not routable by the AnzoGraph server, the Callback HostName is the Anzo host that AnzoGraph can target when making service calls.</p>
<b>Readonly Replica</b>	<p>This option is for use if you have multiple Anzo servers and only one of those servers loads data to AnzoGraph. When <b>Readonly Replica</b> is selected, Anzo treats this instance as a read-only source so that the data can be read but not</p>

Setting	Description
	changed.
<b>Vacuum</b>	<p>Controls whether Anzo initiates an AnzoGraph vacuum process after each load, reload, or refresh operation. The vacuum process improves data organization in memory, deduplicates data, and reclaims memory after data is deleted. Completing a vacuum after update operations is extremely important for maintaining overall query performance and memory allocation accuracy.</p> <p><b>Do not disable vacuum unless you are instructed to do so by Cambridge Semantics Support.</b></p>
<b>Gather Statistics on Load</b>	<p>Controls whether Anzo initiates AnzoGraph's internal statistics gathering queries after loading data. Gathering statistics helps the query planner generate ideal query execution plans when queries are run. When this option is enabled, the statistics queries are run immediately after a graphmart is loaded. It increases graphmart load time but reduces execution time for the first analytic queries, such as when a dashboard is created. When this option is disabled (the checkbox is clear), AnzoGraph automatically performs statistics gathering when the first queries are run, increasing the execution time for the initial queries.</p> <div data-bbox="423 1171 1474 1539" style="background-color: #e6f2ff; padding: 10px; border-radius: 5px;"> <p><b>Note</b></p> <p>Cambridge Semantics recommends that you leave Gather Statistics on Load enabled so that AnzoGraph gathers statistics at the end of a load rather than during query execution. Since loads take longer than queries, adding more time to the load is less noticeable than waiting for statistics to be generated during initial query execution.</p> </div>
<b>Use Priority Queue Query Manager</b>	<p>Controls whether Anzo provides a view of the queries that are in the queue waiting to be run. The queued queries are displayed in the System Query Audit log.</p>

Setting	Description
	<p><b>Important</b></p> <p>Enabling or disabling this option after saving the initial configuration requires a restart of Anzo.</p>
<p><b>Enable Detailed Query Timing</b></p>	<p>When the Priority Queue Query Manager is enabled, this option controls whether Anzo obtains detailed timing statistics for every AnzoGraph query. If this option is enabled, Anzo sends additional statistics gathering queries to AnzoGraph for each user query. The extra query timing details, such as query compilation time, compilation statistics, and a query summary, are displayed in the System Query Audit log. For more information about this setting, see <a href="#">AnzoGraph Detailed Query Timing</a>.</p> <p><b>Important</b></p> <p>Enabling detailed query timing increases the AnzoGraph workload and may decrease overall query performance.</p>
<p><b>Max Allowed Duration for System Operations</b></p>	<p>Sets a limit on the duration of time that Anzo waits for AnzoGraph to complete system operation related queries, such as queries for CPU and memory usage statistics. The default value is not set. When not set, Anzo waits for 2 minutes. If Anzo is waiting on system information from AnzoGraph and AnzoGraph does not respond within the specified time, Anzo cancels the request.</p>
<p><b>Max Allowed Duration for Queries</b></p>	<p>Sets a limit on the amount of time that Anzo waits for AnzoGraph to complete a user query (such as dashboard, data layer, or Query Builder queries). By default, Anzo waits indefinitely. To set a maximum duration, specify the amount of time in any combination of days, hours, and minutes. For example, specifying <b>1d</b> sets the maximum duration to one day. Specifying <b>10h</b>, sets the maximum duration to 10 hours, and specifying <b>1d12h30m</b> sets the duration to 1 day, 12 hours, and 30 minutes. If <b>Max Allowed Duration for Queries</b> is set and a query does not complete in the specified time, Anzo cancels the request</p>

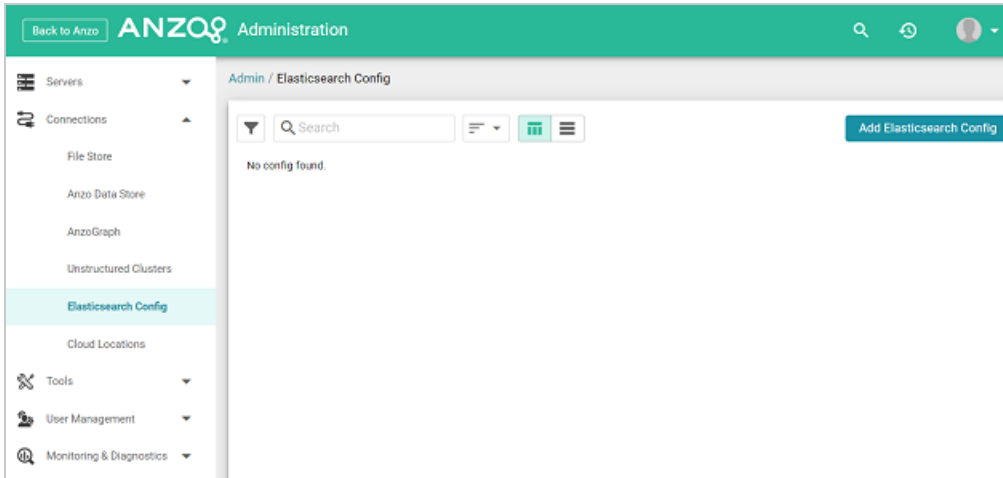


Setting	Description
	regardless of whether AnzoGraph has returned partial results.
<b>Use Minimal Number of SPARQL Rewriters</b>	When Anzo processes SPARQL queries before sending them to AnzoGraph, there is a set of rewrites it makes to try to optimize the query execution. This setting controls whether Anzo performs the full set of rewrites to optimize the query or whether it performs only the minimal required modifications. When this setting is disabled (the default value) Anzo performs the full set of rewrites. When this setting is enabled, Anzo performs only a minimal set of rewrites. <b>Do not enable this setting unless you are instructed to do so by Cambridge Semantics Support.</b>

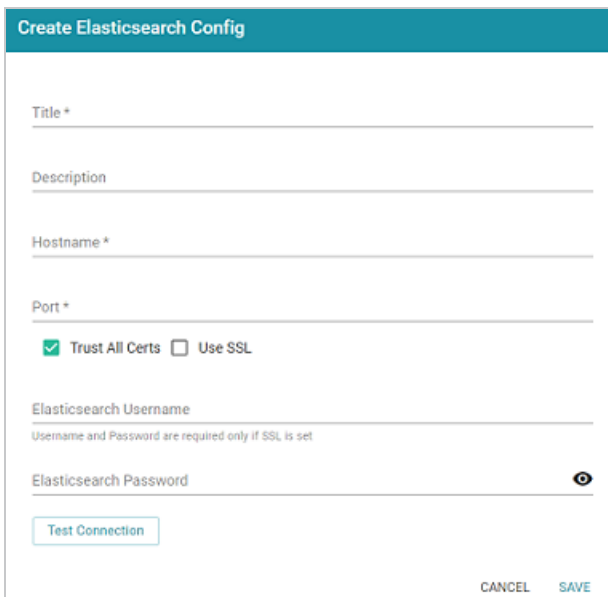
# Connecting to Elasticsearch

This topic provides instructions for configuring a connection to a static Elasticsearch instance. For information about installing Elasticsearch, see [Installing Elasticsearch](#) in the Deployment Guide.

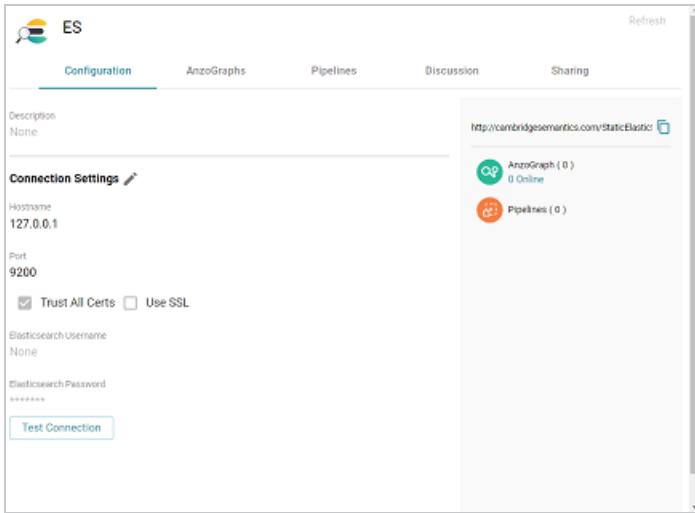
1. In the Administration application, expand the **Connections** menu and click **Elasticsearch Config**. Anzo displays the Elasticsearch Config screen, which lists any existing Elasticsearch connections. For example:



2. On the Elasticsearch Config screen, click the **Add Elasticsearch Config** button. Anzo opens the Create Elasticsearch Config dialog box.

The screenshot shows a "Create Elasticsearch Config" dialog box with a teal header. It contains several input fields: "Title \*" (required), "Description", "Hostname \*" (required), "Port \*" (required), and "Elasticsearch Username". Below the "Elasticsearch Username" field is a note: "Username and Password are required only if SSL is set". There are two checkboxes: "Trust All Certs" (checked) and "Use SSL" (unchecked). Below these is the "Elasticsearch Password" field with a toggle icon. At the bottom left is a "Test Connection" button, and at the bottom right are "CANCEL" and "SAVE" buttons.

3. On the Create Elasticsearch Config screen, provide the following details about the Elasticsearch instance:
  - **Title:** Type a name for this Elasticsearch connection.
  - **Description:** Optional description for this connection.
  - **Hostname:** Specify the IP address or hostname of the Elasticsearch server.
  - **Port:** Specify the port to use for the Elasticsearch connection. The default Elasticsearch port is **9200**.
  - **Trust All Certs:** Indicates whether Anzo should trust the Elasticsearch certificates for this connection. Cambridge Semantics recommends that you accept the default value of enabled.
  - **Use SSL:** If this Elasticsearch instance is configured for SSL authentication, select the **Use SSL** checkbox.
  - **Elasticsearch Username:** If Use SSL is specified, type the user name to use to connect to Elasticsearch.
  - **Elasticsearch Password:** If Use SSL is specified, type the password for the user name that you specified.
4. Click **Test Connection** to check if Anzo can connect to Elasticsearch. If the connection fails, make sure that Elasticsearch is running and that you entered the correct connection details.
5. Anzo displays a Connection Successful dialog box. Click **OK** to close the dialog, and then click **Save** to save the new connection. Anzo saves the connection and displays the Configuration overview screen. You can adjust configuration details as needed. For example:

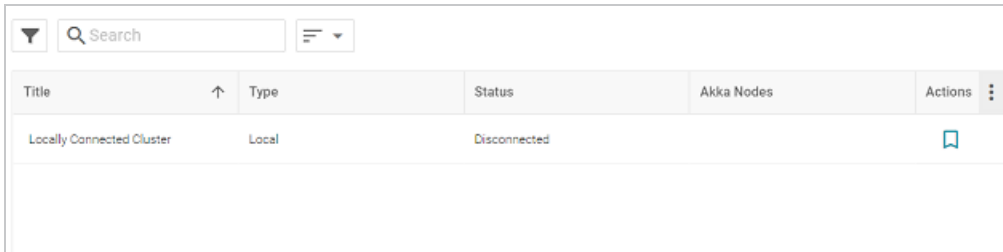


To connect this Elasticsearch instance to an AnzoGraph instance, view the configuration details for the AnzoGraph instance and choose this Elasticsearch connection in the **Elasticsearch Configuration** field. See [Connecting to AnzoGraph](#) for more information.

# Connecting to a Distributed Unstructured Cluster

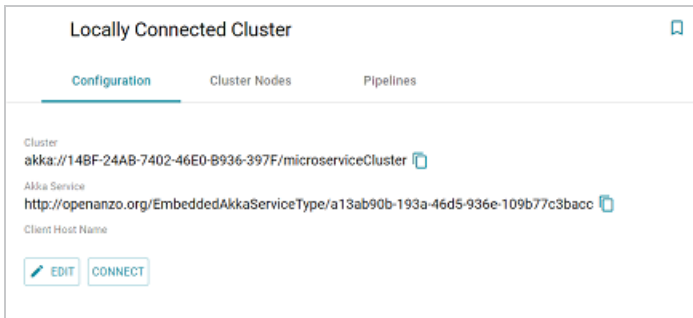
This topic provides instructions for connecting to a static Distributed Unstructured (DU) cluster. For information about installing DU, see [Installing the Distributed Unstructured Cluster](#) in the Deployment Guide.

1. In the Administration application, expand the **Connections** menu and click **Unstructured Clusters**. The Unstructured Clusters screen lists the available clusters. For example, the image below shows a cluster that was just installed. Note that the Status is **Disconnected**:



Title	Type	Status	Actions
Locally Connected Cluster	Local	Disconnected	

2. Click the name of the cluster to open the Configuration screen. For example:



Locally Connected Cluster

**Configuration** Cluster Nodes Pipelines

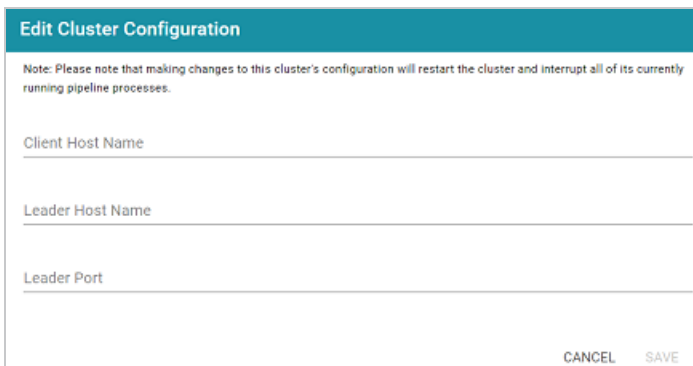
Cluster  
akka://148F-24A8-7402-46E0-8936-397F/microserviceCluster

Akka Service  
<http://openanzo.org/EmbeddedAkkaServiceType/a13ab90b-193a-46d5-936e-109b77c3bacc>

Client Host Name

EDIT CONNECT

3. Click the **Edit** button to open the Edit Cluster Configuration dialog box.



**Edit Cluster Configuration**

Note: Please note that making changes to this cluster's configuration will restart the cluster and interrupt all of its currently running pipeline processes.

Client Host Name  
\_\_\_\_\_  
Leader Host Name  
\_\_\_\_\_  
Leader Port  
\_\_\_\_\_

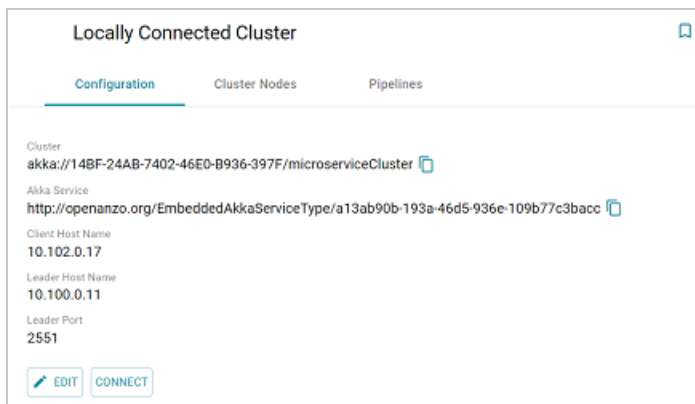
CANCEL SAVE

- On the Edit Cluster Configuration dialog box, complete the Client and Leader Host Name fields. You do not need to specify the Leader Port as Anzo automatically populates the port once the connection is established.
  - Client Host Name:** Specify the hostname or IP address of the Anzo server.
  - Leader Host Name:** Specify the hostname or IP address of the leader server. This is the value specified during the installation of the leader software as described in [Deploy the Leader Node](#) in the Deployment Guide.

### Important

The value must be a routable IP address or hostname. If the leader instance is installed on the Anzo host server, specify the IP address or hostname of the server. Do not use `127.0.0.1` or `localhost`.

- Click **Save** to save the connection configuration. Anzo connects to the cluster, adds the Leader Port value, and returns to the Configuration screen. For example:



The cluster is now connected to Anzo and ready to process unstructured pipelines. If you return to the Unstructured Clusters screen, the status of the cluster is now **Connected** and the number of Akka Nodes is displayed. For example:

Title	↑	Type	Status	Akka Nodes	Actions
Locally Connected Cluster		Local	Connected	3	

If you changed the IP address or hostname of the leader node, review the network settings for the connection between the worker nodes and the leader to ensure that the workers can reach the new leader. See [Configure DU Cluster Network Connections](#) for more information.

# Connecting to a Cloud Location

A Cloud Location is a connection between Anzo and the Kubernetes (K8s) cluster that will host the dynamic Anzo Agent and Anzo Unstructured, AnzoGraph, and Elasticsearch applications. When you create a Cloud Location, Anzo discovers the K8s cluster and any internal container registries, authenticates the K8s API services, obtains the node pool or group specifications and retrieves pricing information from the Cloud Service Provider for the configured compute instances, and maps the node pool specifications to Launch Configurations in Anzo.

### Tip

For instructions on deploying the K8s infrastructure to support Cloud Locations, see [Configure K8s for Dynamic Deployments](#) in the Deployment Guide.

The topics in this section provide instructions on setting up the NFS configuration for the dynamically deployed applications and creating a Cloud Location.

#### In this section:

<a href="#">Importing the NFS Configuration</a> .....	105
<a href="#">Creating a Cloud Location</a> .....	108



# Importing the NFS Configuration

Before creating a Cloud Location, the configuration details for the NFS server or servers need to be imported into Anzo. This is a one-time procedure; the configuration that you import is used for all Cloud Locations. Anzo will automatically mount the NFS server to any nodes that are provisioned when applications are deployed.

## Tip

For information about the NFS requirements, see [NFS Guidelines](#) in the Deployment Guide.

- [Create the NFS Configuration File](#)
- [Import the File](#)

## Create the NFS Configuration File

The NFS configuration details need to be specified in TriG format. The TriG file is imported to Anzo using the Anzo Admin CLI. Use the following contents as a template to create a .trig file on the Anzo server. If you have multiple NFS servers for different regions, you can configure each server in the same configuration file. The objects to supply values for are described below:

```
@prefix : <http://cambridgesemantics.com/ontologies/cloud/deployment/config#> .
@prefix nfsmountconfig:
<http://cambridgesemantics.com/ontologies/CloudDeployment/NFSMountConfiguration/> .
@prefix deployment: <http://cambridgesemantics.com/ontologies/CloudDeployment/> .
@prefix anzo: <http://openanzo.org/ontologies/2008/07/Anzo#> .
@prefix int: <http://openanzo.org/system/internal/> .
@prefix role: <http://openanzo.org/Role/> .
@prefix reg: <http://cambridgesemantics.com/registries/> .

#Mode:REPLACE
:nfsMountConfig1
{
  :nfsMountConfig1 a deployment:NFSMountConfiguration, deployment:MountConfiguration;
  nfsmountconfig:NFSfqdn "NFSfqdn" ;
  nfsmountconfig:NFSMountDir "NFSMountDir" ;
  nfsmountconfig:NFSMountOptions "NFSMountOptions" ;
  nfsmountconfig:NFSSharedDir "NFSSharedDir" .
}
```

```

#Mode:ADD
reg:CloudLocation {
  reg:CloudLocation anzo:defaultNamedGraph :nfsMountConfig1
}
# :nfsMountConfig2
# {
#   :nfsMountConfig2 a deployment:NFSMountConfiguration, deployment:MountConfiguration;
#   nfsmountconfig:NFSfqdn "NFSfqdn2" ;
#   nfsmountconfig:NFSMountDir "NFSMountDir2" ;
#   nfsmountconfig:NFSMountOptions "NFSMountOptions2" ;
#   nfsmountconfig:NFSSharedDir "NFSSharedDir2" .
# }
# Mode:ADD
# reg:CloudLocation {
#   reg:CloudLocation anzo:defaultNamedGraph :nfsMountConfig2
# }

```

Object	Description
<b>NFSfqdn</b>	The IP address for the NFS server.
<b>NFSMountDir</b>	The NFS mount location on the Anzo server. The same mount location will be used to mount the NFS when dynamic resources are provisioned.
<b>NFSMountOptions</b>	The mount options to use when mounting the NFS.
<b>NFSSharedDir</b>	The NFS directory to share between Anzo and the dynamic resources.

For example:

```

# nfs-config.trig
@prefix : <http://cambridgesemantics.com/ontologies/cloud/deployment/config#> .
@prefix nfsmountconfig:
<http://cambridgesemantics.com/ontologies/CloudDeployment/NFSMountConfiguration/> .
@prefix deployment: <http://cambridgesemantics.com/ontologies/CloudDeployment/> .
@prefix anzo: <http://openanzo.org/ontologies/2008/07/Anzo#> .
@prefix int: <http://openanzo.org/system/internal/> .
@prefix role: <http://openanzo.org/Role/> .

#Mode:REPLACE

```

```

:nfsMountConfig1
{
  :nfsMountConfig1 a deployment:NFSMountConfiguration, deployment:MountConfiguration;
  nfsmountconfig:isTransferFiles false ;
  nfsmountconfig:NFSfqdn "10.104.0.6" ;
  nfsmountconfig:NFSMountDir "/private/var/nfsshare_dev" ;
  nfsmountconfig:NFSMountOptions "hard,nfsvers=4.1" ;
  nfsmountconfig:NFSSharedDir "/global/nfs/data" .
}
#Mode:ADD
reg:CloudLocation {
  reg:CloudLocation anzo:defaultNamedGraph :nfsMountConfig1
}

```

## Import the File

Once the NFS configuration file is created, run the following command to import the file to Anzo with the Anzo Admin CLI:

```
<install_path>/Client/anzo <file_path>/<filename>.trig -u sysadmin --useModes
```

For example:

```
/opt/Anzo/Client/anzo import nfs-config.trig -u sysadmin --useModes
```

When the NFS configuration details have been imported to Anzo, see [Creating a Cloud Location](#) for next steps.

# Creating a Cloud Location

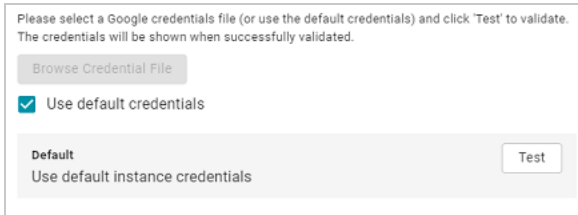
Follow the instructions below to create a Cloud Location. Note that the steps below are in progress and more details are forthcoming.

1. In the Administration application, expand the **Connections** menu and click **Cloud Locations**.
2. On the Cloud Locations screen, click the **Add Cloud Location** button and select the Cloud Service Provider that hosts your Kubernetes (K8s) cluster. The Create Cloud Location dialog box is displayed. For example, the image below shows the Create Cloud Location screen for Google:

The screenshot shows a 'Create Cloud Location' dialog box. At the top, it says 'You're creating a Google cloud location.' Below this are input fields for 'Title \*' and 'Description'. A note states: 'Please select a Google credentials file (or use the default credentials) and click "Test" to validate. The credentials will be shown when successfully validated.' There is a 'Browse Credential File' button and a checkbox for 'Use default credentials'. Below these are three dropdown menus: 'Kubernetes Cluster', 'Kubernetes Namespace', and 'Kubernetes Docker Registry'. A note under the namespace dropdown says: 'Note: You can choose from the list or key in a custom namespace (Must consist of lower case alphanumeric characters or ". Must start and end with an alphanumeric character. Ex: "myname" or "123-abc").' At the bottom, there is an 'Operators Support' section with a list of operators: Anzo Operator, AnzoGraph Operator, Elasticsearch Operator, Unstructured Operator, and Spark Engine Operator (Livy Deployment for Spark Config). The dialog ends with 'CANCEL' and 'DEPLOY' buttons.

3. At the top of the screen, specify a **Title** for this Cloud Location and type an optional **Description**.
4. Next, specify the credentials that have permission to connect to the Cloud Service Provider (CSP) API and deploy resources in the Kubernetes cluster. There are two options, depending on the user that is running Anzo and the Service Account, Principal, or Group that was assigned the K8s Cluster Developer IAM policy when the K8s infrastructure was set up:
  - Since the Anzo Service Account, Principal, or Group is typically running Anzo, and the K8s Cluster Developer IAM policy was assigned to that account when the K8s

infrastructure was set up, the appropriate credentials are already applied to this Anzo instance. In this case, select the **Use Default Credentials** checkbox. The dialog box indicates that the default instance credentials will be used and presents a **Test** button (shown in the image below).



Click **Test** to retrieve the credentials and test that they are valid.

- If another user is running Anzo, and that account does not have the Cluster Developer IAM permissions, retrieve from your CSP the JSON configuration file for the account that is assigned the Cluster Developer IAM policy. Then click the **Browse Credential File** button and upload the JSON credentials file that you downloaded.

# Administration Tools

Anzo's Administration Tools aid administrators in performing repetitive, bulk operations.

**In this section:**

- [Workflow Manager](#) ..... 111
- [Migration Packages](#) ..... 127

# Workflow Manager

The Workflow Manager is used to automate tasks such as unstructured pipeline runs and graphmart loads. Workflows can be triggered from the Anzo Admin CLI, and the CLI call can be automated by setting up cron jobs. The topics in this section provide instructions for creating Workflows, adding Tasks, and configuring cron jobs.

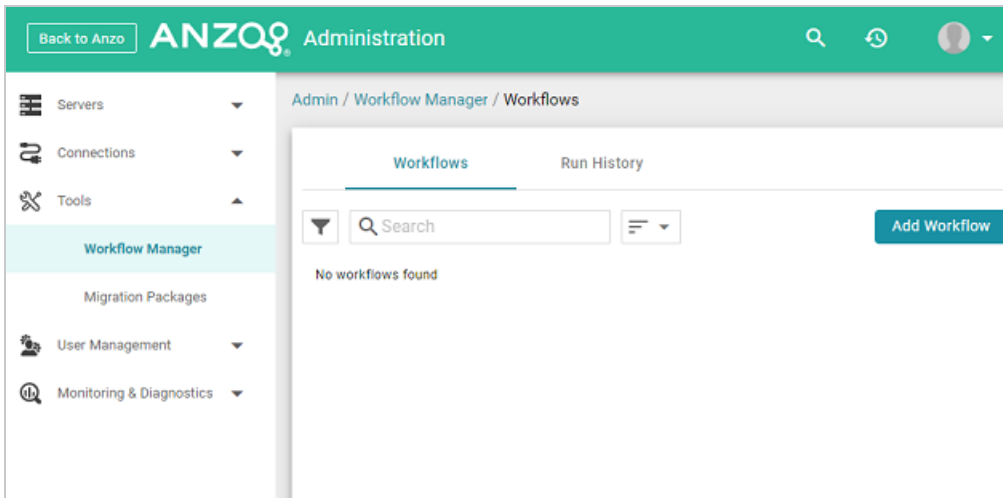
**In this section:**

- [Adding a Workflow](#) ..... 112
- [Adding a Task to a Workflow](#) ..... 115
- [Running a Workflow](#) ..... 125

# Adding a Workflow

A workflow is a container for tasks. Running a workflow runs all of the tasks in that workflow. Consider whether any tasks have dependencies when determining the number and type of tasks to group in one workflow.

1. In the Administration application, expand the **Tools** menu and click **Workflow Manager**. Anzo displays the Workflows screen, which lists any existing workflows. The image below shows the Workflows screen on an environment without any existing workflows.

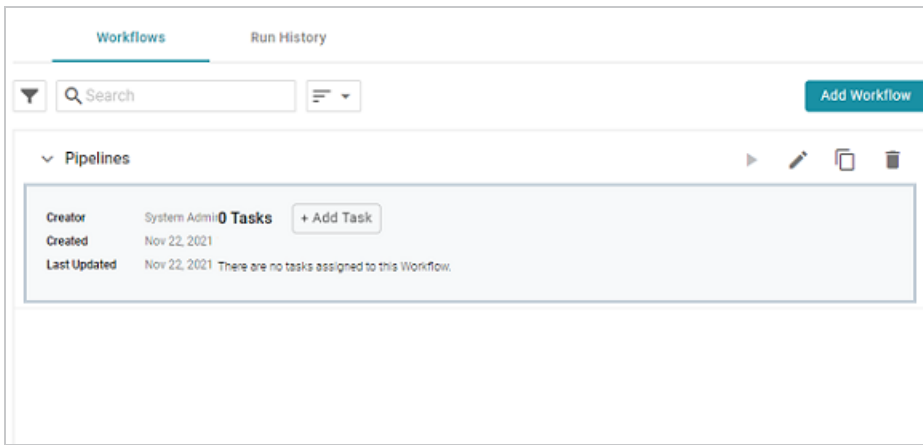


2. Click **Add Workflow**. The Create Workflow dialog box is displayed:

The 'Create Workflow' dialog box has a teal header. It contains several input fields: 'Label \*' (required), 'Description', and 'Load Timeout (ms)' with a value of '7200000' and a tooltip 'Time in milliseconds to wait for a load service to return from execution.' There is a checkbox for 'Stop on Failure' which is currently unchecked. Below are two dropdown menus for 'Ingest Manager Service Connection' and 'Orchestration Service Connection', both set to 'Local Orchestration Services Connection'. Each dropdown has a tooltip: 'The connection to use to talk to the Ingest Manager Service.' and 'The connection to use to talk to the Orchestration Service.' respectively. At the bottom are 'CANCEL' and 'CREATE' buttons.



3. Configure the Workflow by completing the following fields as needed. Only **Label** is a required field.
  - **Label:** This field specifies the name of the workflow.
  - **Description:** This field specifies an optional description for the workflow.
  - **Load Timeout (ms):** This field specifies the time limit (in milliseconds) for the workflow to complete. The default value is 7200000 milliseconds (120 minutes). If all of the tasks in the workflow are not finished before the load timeout, the workflow will be stopped.
  - **Stop on Failure:** This option controls whether the workflow is stopped if one of the tasks fails or whether the workflow continues to process the rest of the tasks if there is a failure.
  - **Ingest Manager Service Connection:** This field specifies the Ingest Manager connection to use for this workflow. The field defaults to the local Orchestration Services connection. If you have registered additional connections, you can select an alternate connection.
  - **Orchestration Service Connection:** This field specifies the Orchestration Service connection to use for this workflow. The field defaults to the local Orchestration Services connection. If you have registered additional connections, you can select an alternate connection.
4. Click **Create** to add the workflow. The new workflow is added to the list of workflows on the Workflows screen. For example, the image below shows that there is one new workflow without any tasks.



Once the workflow is configured, you can add any number of tasks that the workflow should run. For instructions, see [Adding a Task to a Workflow](#).

## Adding a Task to a Workflow

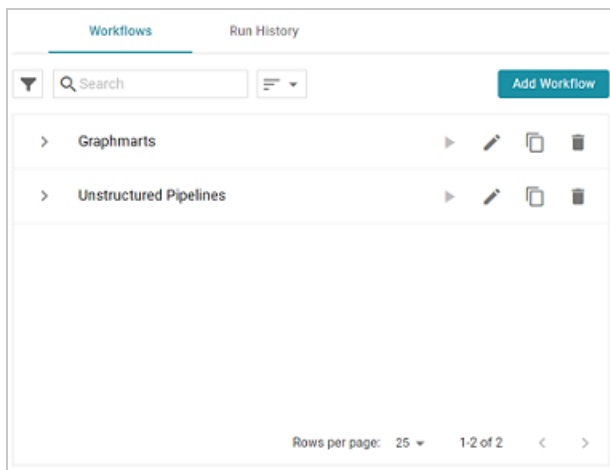
This topic provides instructions for configuring each type of task that is available for adding to a workflow.

- [Adding a Task that Runs an Unstructured Pipeline](#)
- [Adding a Task that Refreshes or Reloads a Graphmart](#)
- [Adding a Task that Pauses the Workflow](#)

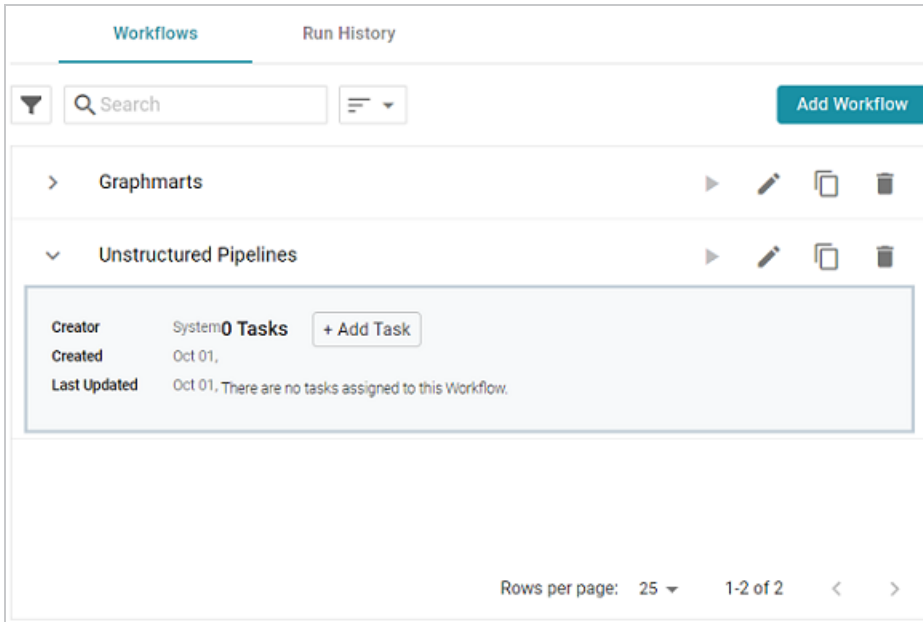
### Adding a Task that Runs an Unstructured Pipeline

Follow the instructions below to add a task that runs an unstructured pipeline.

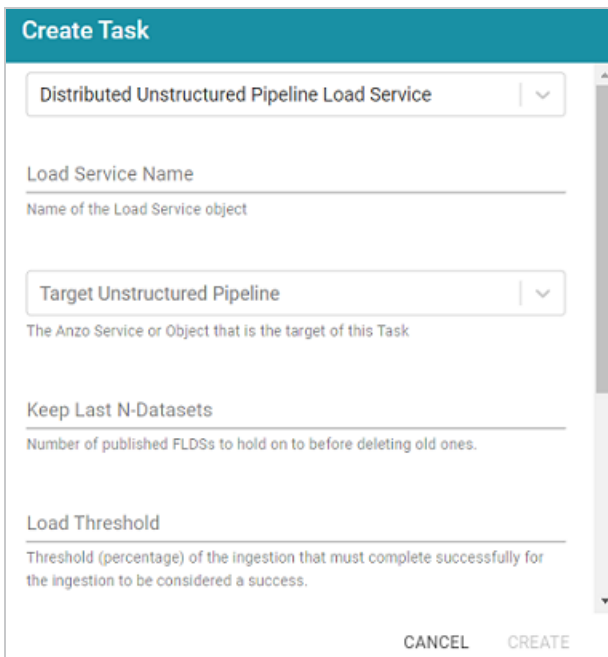
1. In the Administration application, expand the **Tools** menu and click **Workflow Manager**. Anzo displays the Workflows screen, which lists any existing workflows. For example:



2. Expand the workflow that you want to add a task to. For example:



3. Click **Add Task**. The Create Task dialog box is displayed:



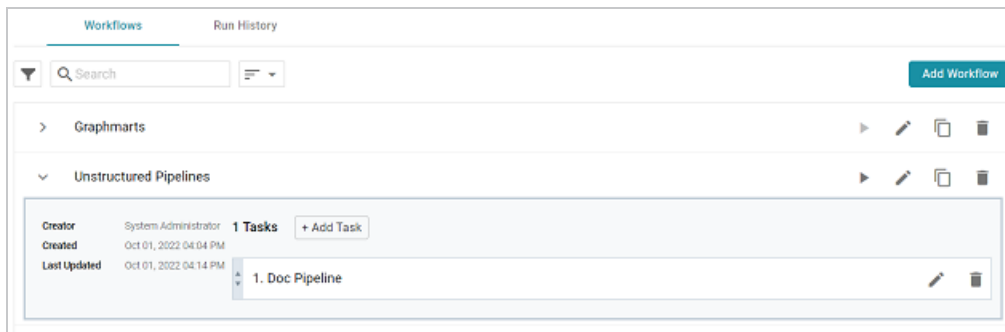
4. Configure the Task by completing the following fields as needed:

- **Task Type:** The drop-down list at the top of the dialog box specifies the type of task to create. **Distributed Unstructured Pipeline Load Service** is selected by default.

Accept the default value.

- **Load Service Name:** This field specifies the name for the task.
- **Target Unstructured Pipeline:** This field specifies the unstructured pipeline that this task should run. Click the drop-down list and select the desired pipeline.
- **Keep Last N-Datasets:** This field specifies the number of file-based linked data sets (FLDS) from this pipeline to retain on disk before deleting the oldest ones.
- **Load Threshold:** This field specifies the percentage of the pipeline that must complete successfully for the ingestion to be considered a success.
- **Distributed Unstructured Pipeline Stop Timeout:** This field specifies the number of milliseconds to wait for a pipeline to stop.
- **Distributed Unstructured Pipeline Percent Timeout:** This field specifies the number of milliseconds to wait before timing out if there is no change in the percentage of documents processed.
- **Index:** This field specifies a numeric value that represents the order in which this task should run in the workflow.

5. Click **Create** to add the task to the workflow. For example, the image below shows a workflow with one task.

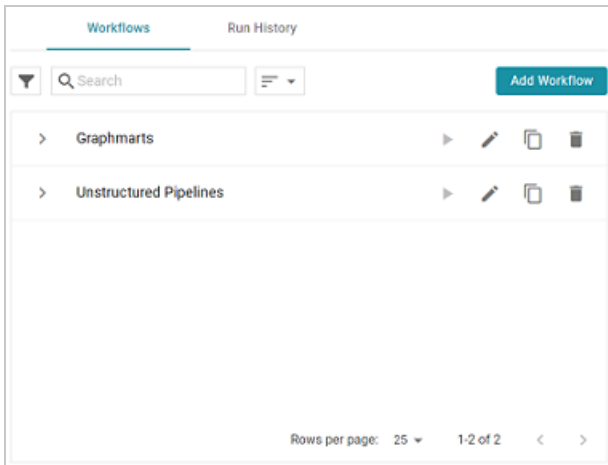


You can repeat this process to add tasks that run additional unstructured pipelines.

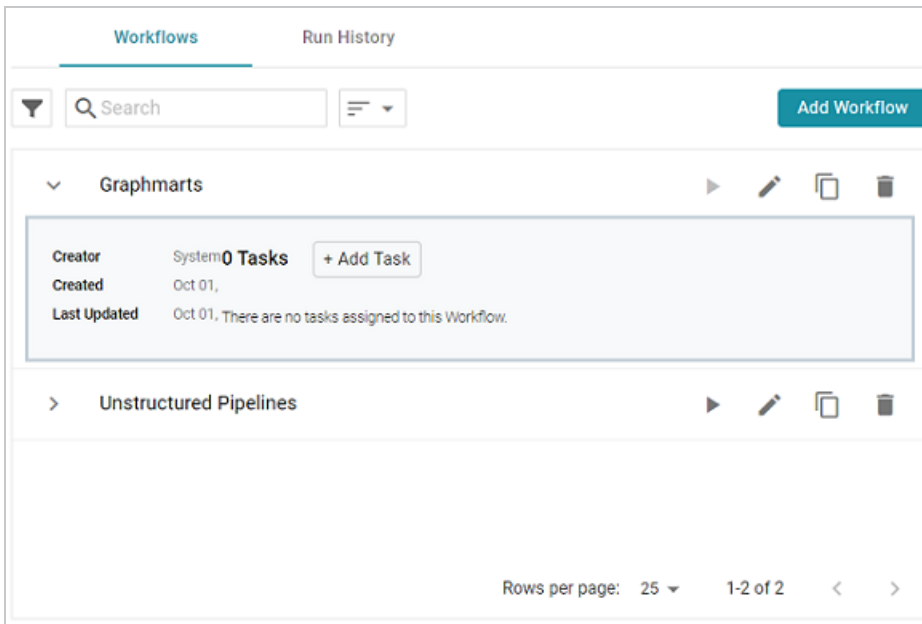
## Adding a Task that Refreshes or Reloads a Graphmart

Follow the instructions below to add a task that refreshes or reloads a graphmart.

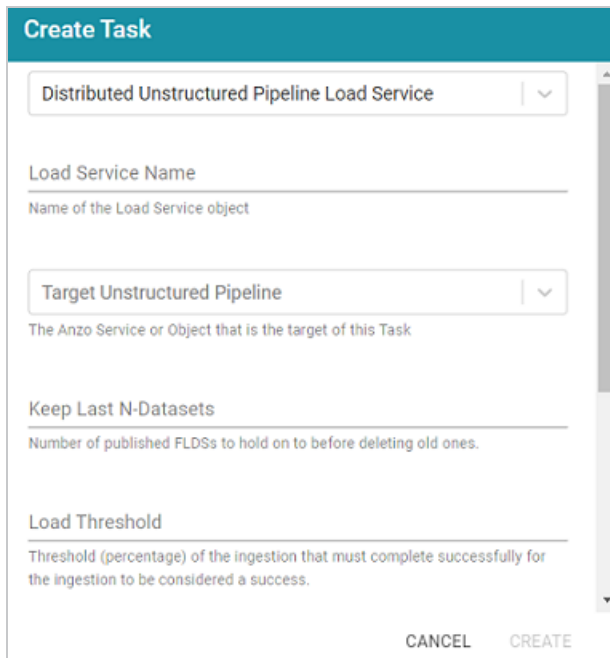
1. In the Administration application, expand the **Tools** menu and click **Workflow Manager**. Anzo displays the Workflows screen, which lists any existing workflows. For example:



2. Expand the workflow that you want to add a task to. For example:

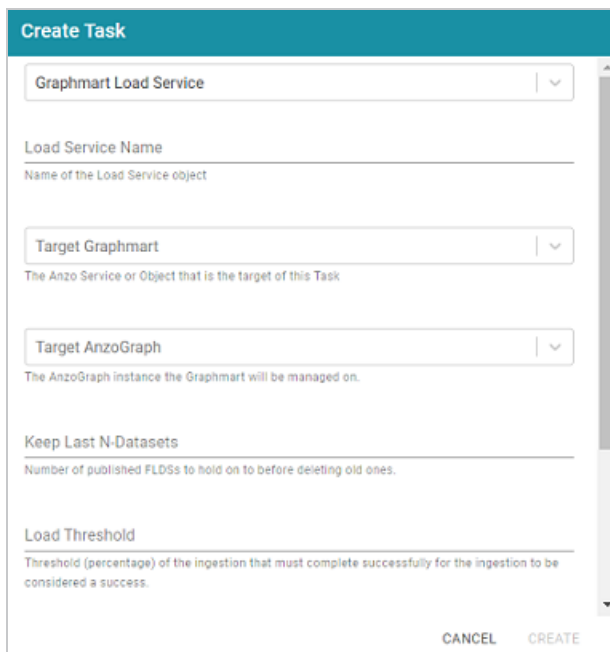


3. Click **Add Task**. The Create Task dialog box is displayed:



The screenshot shows the 'Create Task' dialog box with a teal header. The top dropdown menu is set to 'Distributed Unstructured Pipeline Load Service'. Below it is a text input field for 'Load Service Name' with the placeholder text 'Name of the Load Service object'. The second dropdown menu is set to 'Target Unstructured Pipeline' with the placeholder text 'The Anzo Service or Object that is the target of this Task'. Below that is another text input field for 'Keep Last N-Datasets' with the placeholder text 'Number of published FLDSs to hold on to before deleting old ones.'. The final text input field is for 'Load Threshold' with the placeholder text 'Threshold (percentage) of the ingestion that must complete successfully for the ingestion to be considered a success.'. At the bottom right, there are two buttons: 'CANCEL' and 'CREATE'.

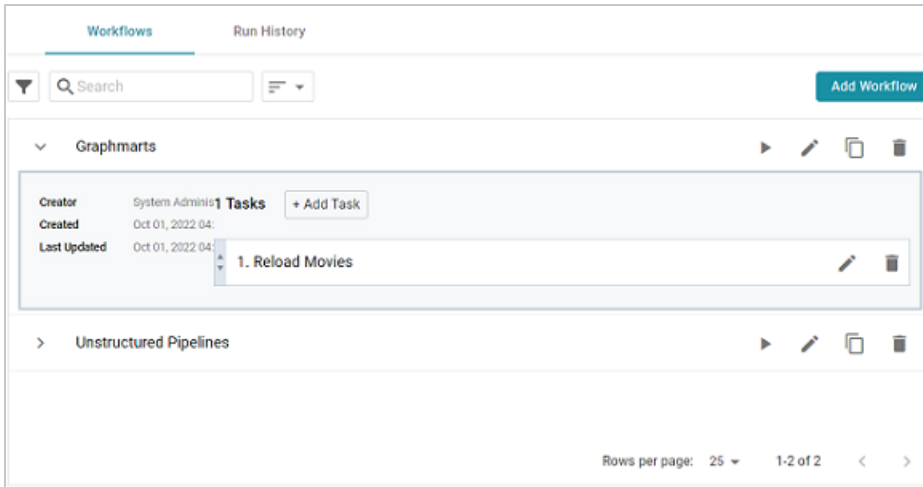
4. At the top of the dialog box, click the drop-down list and select **Graphmart Load Service** to set up a task that reloads or refreshes a graphmart. The dialog box presents the options that are valid for graphmart Load Service Tasks:



The screenshot shows the 'Create Task' dialog box with a teal header. The top dropdown menu is now set to 'Graphmart Load Service'. Below it is a text input field for 'Load Service Name' with the placeholder text 'Name of the Load Service object'. The second dropdown menu is set to 'Target Graphmart' with the placeholder text 'The Anzo Service or Object that is the target of this Task'. A new dropdown menu has appeared below that, set to 'Target AnzoGraph' with the placeholder text 'The AnzoGraph instance the Graphmart will be managed on.'. Below that is a text input field for 'Keep Last N-Datasets' with the placeholder text 'Number of published FLDSs to hold on to before deleting old ones.'. The final text input field is for 'Load Threshold' with the placeholder text 'Threshold (percentage) of the ingestion that must complete successfully for the ingestion to be considered a success.'. At the bottom right, there are two buttons: 'CANCEL' and 'CREATE'.

5. Configure the Task by completing the following fields as needed:
  - **Load Service Name:** This field specifies the name for the task.
  - **Target Graphmart:** This field specifies the graphmart that this task should reload or refresh. Click the drop-down list and select the desired graphmart.
  - **Target AnzoGraph:** This field specifies the AnzoGraph instance that hosts this graphmart.
  - **Keep Last N-Datasets:** This field is not relevant for Graphmart Load Service tasks.
  - **Load Threshold:** This field is not relevant for Graphmart Load Service tasks.
  - **Graphmart Action:** This field specifies whether to refresh or reload the target graphmart. For refresh, click the drop-down list and select **Refresh Target Graphmart**. To perform a reload, click the drop-down list and select **Reload Target Graphmart**.
  - **Activate:** This option indicates whether the target graphmart needs to be activated before the refresh or reload is attempted. If the target graphmart is offline when the workflow is run, this task will fail unless **Activate** is enabled.
  - **Deactivate:** This option indicates whether to deactivate the graphmart after the task is complete. If you want Anzo to deactivate the target graphmart after the reload or refresh is complete, select the **Deactivate** checkbox.
  - **Index:** This field specifies a numeric value that represents the order in which this task should run in the workflow.
6. Click **Create** to add the task to the workflow. For example, the image below shows a workflow with one Task.



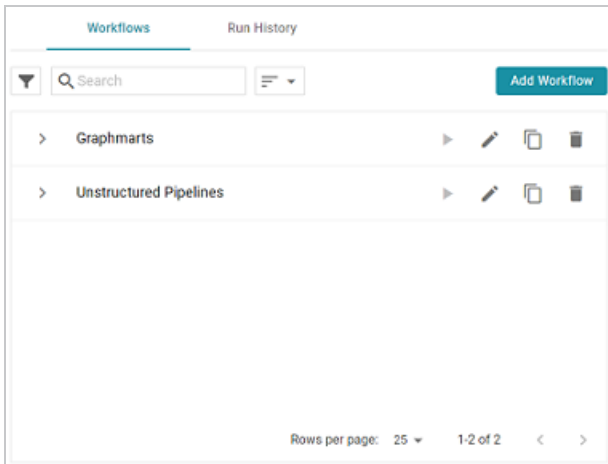


You can repeat this process to add tasks that refresh or reload additional graphmarts.

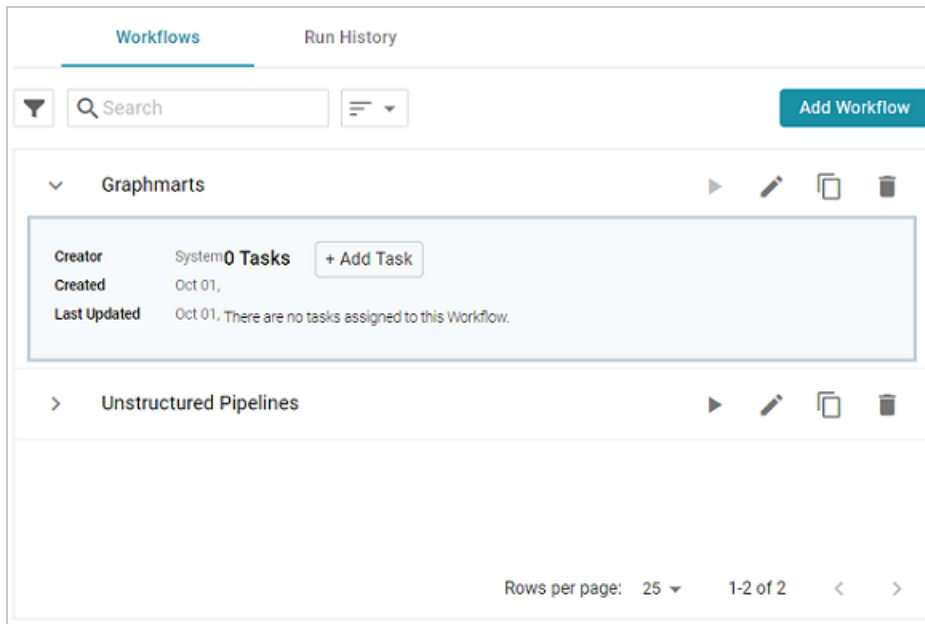
## Adding a Task that Pauses the Workflow

Follow the instructions below to create a task that adds a pause between tasks in a workflow. For example, you may want to add a pause between one task that reloads a graphmart and another task that refreshes a graphmart that depends on the updated data from the reloaded graphmart.

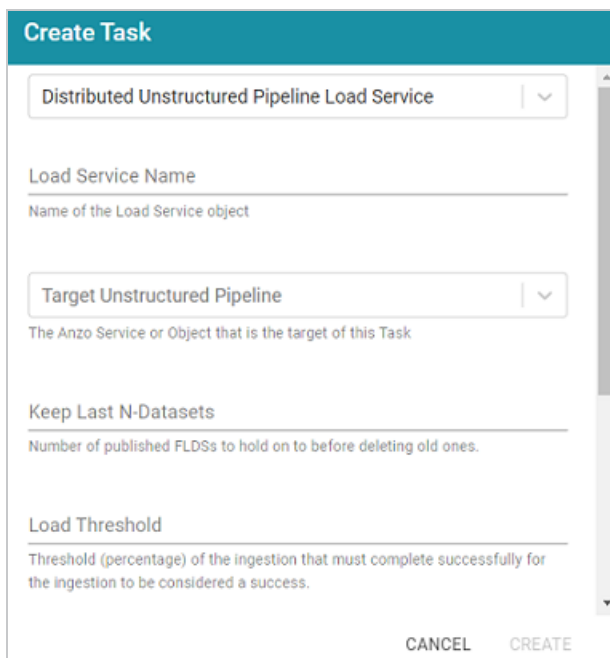
1. In the Administration application, expand the **Tools** menu and click **Workflow Manager**. Anzo displays the Workflows screen, which lists any existing workflows. For example:



- Expand the workflow that you want to add a task to. For example:



- Click **Add Task**. The Create Task dialog box is displayed:



- At the top of the dialog box, click the drop-down list and select **Pause Load Service**. The dialog box presents the options that are valid for Pause Load Service Tasks:

**Create Task**

Pause Load Service | v

**Load Service Name**  
Name of the Load Service object

**Pause Time**  
Milliseconds to pause.

**Keep Last N-Datasets**  
Number of published FLDSs to hold on to before deleting old ones.

**Load Threshold**  
Threshold (percentage) of the ingestion that must complete successfully for the ingestion to be considered a success.

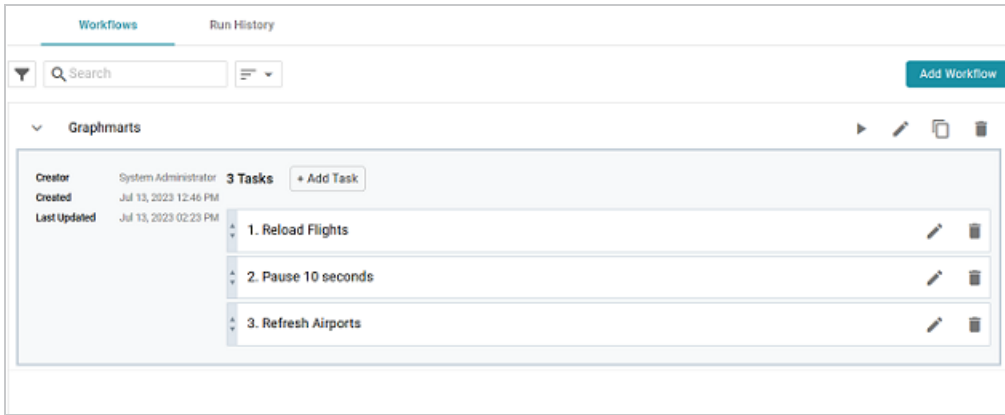
**Index**  
Index of the Load Service to determine the order in which it is invoked

CANCEL CREATE

5. Configure the Task by completing the following fields as needed:

- **Load Service Name:** This field specifies the name for the task.
- **Pause Time:** This field defines the number of milliseconds to pause between tasks. For example, **10000** is 10 seconds.
- **Keep Last N-Datasets:** This field is not relevant for Pause Load Service tasks.
- **Load Threshold:** This field is not relevant for Pause Load Service tasks.
- **Index:** This field specifies a numeric value that represents the order in which this task should run in the workflow.

6. When you have finished configuring the task, click **Create** to add the task to the workflow. For example, the image below shows a workflow with a pause between reloading one graphmart and refreshing another.



You can repeat this process to create additional pause tasks.

## Running a Workflow

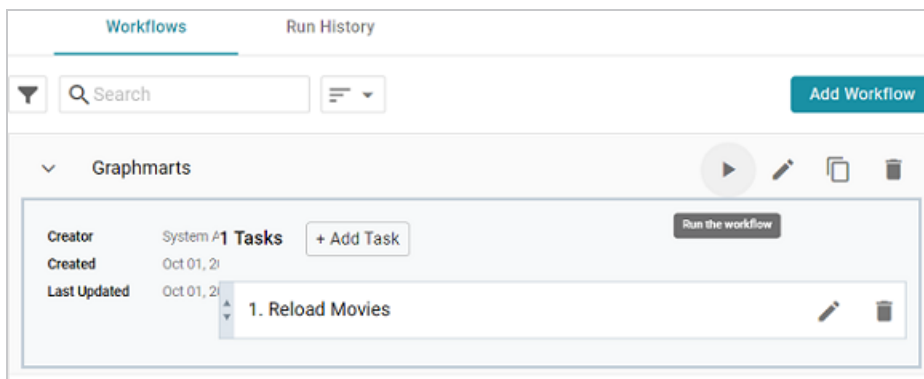
There are multiple ways to run workflows. You can initiate a workflow manually from the Administration application or the Anzo Admin CLI. You can also automate workflows by using the Linux Cron utility or a similar application to schedule them. This topic provides instructions for running a workflow manually and gives an example of a cron job that runs a workflow on a schedule.

- [Running a Workflow Manually](#)
- [Scheduling a Workflow to Run Automatically](#)

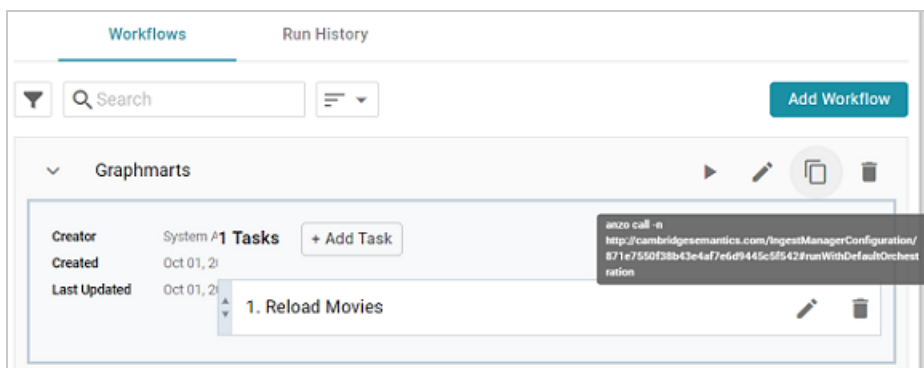
### Running a Workflow Manually

There are two ways to run a workflow manually:

1. You can click the run icon (▶) for the Workflow in the Administration application.



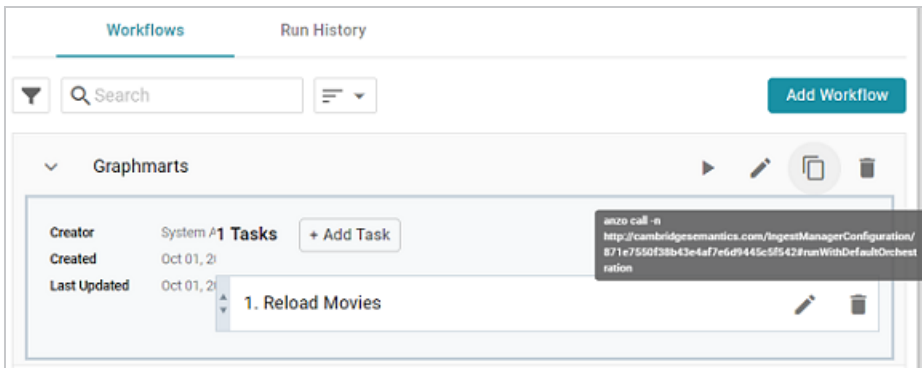
2. You can click the copy icon (📄) to copy the `anzo call` statement for the Workflow and run it with the Admin CLI.



## Scheduling a Workflow to Run Automatically

To automate the running of a workflow, you can set up a cron job that runs the `anzo call` statement on a schedule. This section gives example steps to follow to set up a cron job that schedules a single workflow.

1. First, find the `anzo call` statement for the workflow that you want to schedule. As shown in the image below, you can click the copy icon (📄) for the workflow to copy the statement.



2. On the Anzo server, run the following command to open a crontab:

```
sudo crontab -e -u <user_name>
```

For example, the following command opens a crontab as the Anzo service user:

```
sudo crontab -e -u anzo
```

3. Add contents to the file using the following syntax. Use an asterisk in place of options that you do not want to set:

```
<minute> <hour> <day_of_month> <month> <day_of_week> <absolute_path_to_client/anzo_call_statement>
```

For example, the following contents run the workflow every day at 8:00 AM:

```
0 8 * * * /opt/Anzo/Client/anzo call -n
http://cambridgesemantics.com/IngestManagerConfiguration/ff3b3e313f634535b49e7
1167ca56096#runWithDefaultOrchestration
```

4. Save and close the crontab.

# Migration Packages

When migrating artifacts between environments, administrators can perform a bulk export (and import) by assembling a migration package that includes any number and type of artifacts and their related entities. The export configuration is maintained at the package level and applied to all of the contained artifacts, which means the configuration can be reused as artifacts are added to or removed from the package. The topics in this section provide instructions on creating and configuring migration packages as well as exporting and importing packages.

**In this section:**

<a href="#">Creating a Migration Package</a> .....	128
<a href="#">Exporting a Migration Package</a> .....	133
<a href="#">Export Configuration Settings Reference</a> .....	137
<a href="#">Editing Migration Package Template Files</a> .....	142
<a href="#">Importing a Migration Package</a> .....	148

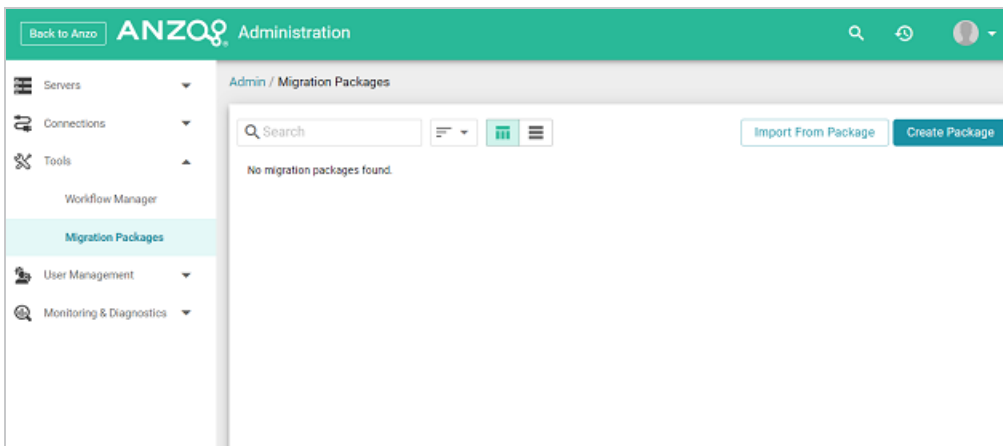
## Creating a Migration Package

Follow the instructions below to create a new migration package, add artifacts to the package, and configure the export options.

### Note

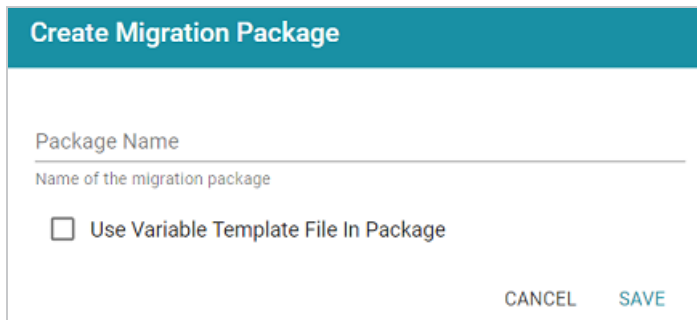
There are two permissions that control access to export, modify, and import migration packages: **Manage Migration Packages** and **Perform Migration Package Operations As Sysadmin**. If a user has only the Manage Migration Packages permission, they cannot modify, export, or import artifacts in packages unless they have the appropriate permissions on the artifacts. If a user has Perform Migration Package Operations As Sysadmin, that means they can modify, export, and import migration packages that include artifacts they may not otherwise have permission to operate on.

1. In the Administration application, expand the **Tools** menu and click **Migration Packages**. Anzo displays the Migration Packages screen, which lists any existing packages. The image below shows the Migration Packages screen on an environment without any existing packages.





2. Click the **Create Package** button. The Create Migration Package dialog box is displayed.



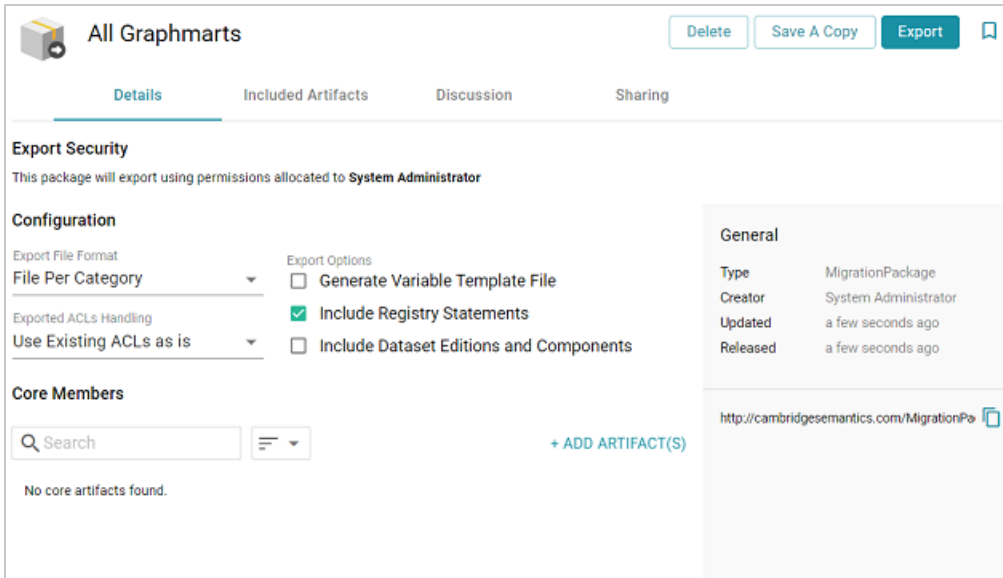
**Create Migration Package**

Package Name  
Name of the migration package

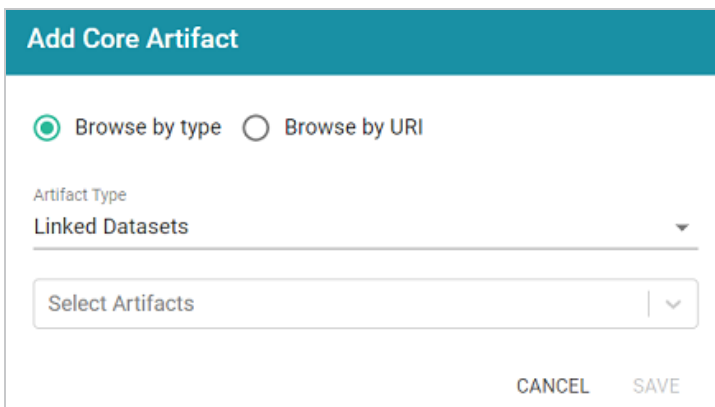
Use Variable Template File In Package

CANCEL SAVE

3. On the Create Migration Package screen, type a name for the package in the **Package Name** field.
4. Next, determine whether you want to add a Variable Template File to the package. A Variable Template File is a TriG file that contains statements for all of the properties that have replaceable values for each artifact included in the package. Properties with replaceable values are objects such as file paths and Anzo Data Store locations, which might differ on the source and target Anzo servers. The template that is generated has placeholder text that you replace with the desired values for the target server. To generate a Variable Template File with the package, select the **Use Variable Template File In Package** checkbox. Leave the checkbox blank if you do not want to generate the file.
5. Click **Save** to save the package. Anzo creates the package and displays the Details tab where you can add artifacts and configure additional options. For example, the image below shows a new package called All Graphmarts.



- First, determine the artifacts to add to the package. Click **Add Artifacts** under Core Members. The Add Core Artifact dialog box is displayed:



- By default, the dialog box is set to **Browse by Type** of artifact, such as Linked Dataset, Data Source, or Graphmart. To choose a type, click the **Artifact Type** drop-down list and select a type to filter by. The **Select Artifacts** list is filtered to show only the selected type of artifact.

### Tip

If you have a list of artifacts that you want to find by URI, you can select the **Browse by URI** radio button and then specify a URI.

- Click **Select Artifacts** and select an artifact from the list. Repeat this step to select multiple artifacts of the same type.

### Tip

All related entities for the selected artifact are automatically added to the package. You do not need to find and select each related artifact individually. If you change the Artifact Type, any selections will be cleared from the Select Artifacts field.

- When you have finished selecting artifacts, click **Save** to add the artifacts to the package. The artifacts are added to the list at the bottom of the screen. For example, the package shown in the image below contains three graphmarts. The **Included Artifacts** column shows the total number of artifacts that are related to the core member and are also included in the package.

The screenshot shows the 'All Graphmarts' interface. At the top, there are buttons for 'Delete', 'Save A Copy', and 'Export'. Below these are tabs for 'Details', 'Included Artifacts', 'Discussion', and 'Sharing'. The 'Details' tab is active, showing 'Export Security' information (permissions for System Administrator), 'Configuration' options (Export File Format, File Per Category, Exported ACLs Handling, Use Existing ACLs as is, and Export Options like Generate Variable Template File, Include Registry Statements, and Include Dataset Editions and Components), and 'Core Members' with a search bar and a '+ ADD ARTIFACT(S)' button. A table at the bottom lists included artifacts:

Title	Type	Available	# Included Artifacts	Actions
Movies	Graphmart	✓	9	✕
Northwind	Graphmart	✓	8	✕
Tickets	Graphmart	✓	9	✕

### Tip

You can view specifics about the included artifacts on the **Included Artifacts** tab. More information about the tab is included in [Editing Migration Package Template Files](#).

10. Once the desired artifacts have been added to the package, review the export options at the top of the screen and make adjustments as needed. For details about each of the settings, see [Export Configuration Settings Reference](#).

Once the migration package includes the desired artifacts and the export options are configured, the package can be exported. For instructions, see [Exporting a Migration Package](#).

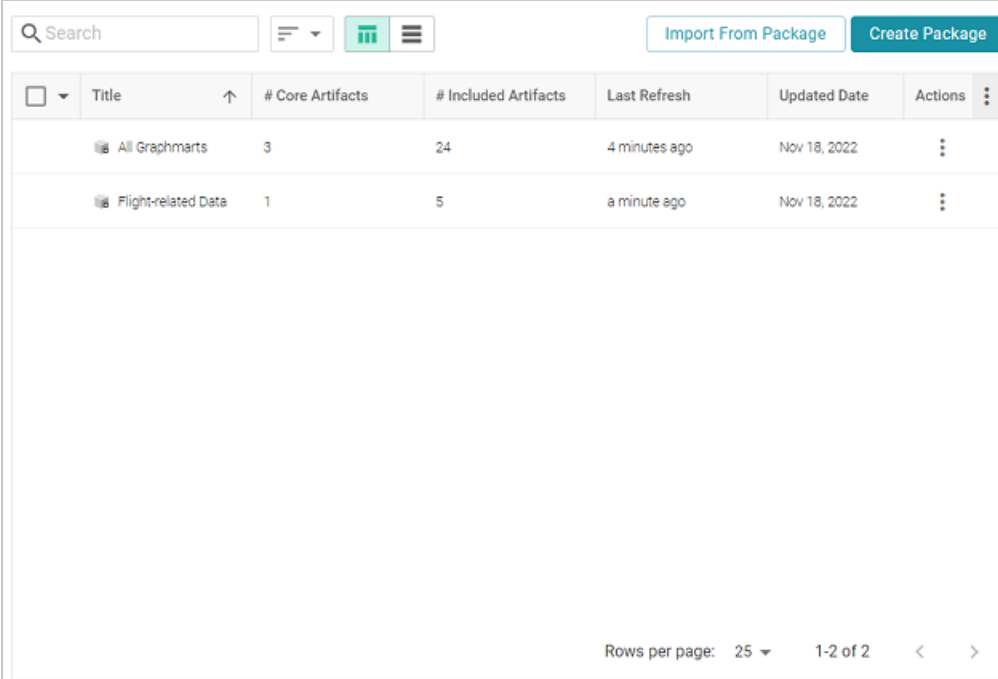
## Exporting a Migration Package

Follow the steps below to export a migration package.

### Note

There are two permissions that control access to export, modify, and import migration packages: **Manage Migration Packages** and **Perform Migration Package Operations As Sysadmin**. If a user has only the Manage Migration Packages permission, they cannot modify, export, or import artifacts in packages unless they have the appropriate permissions on the artifacts. If a user has Perform Migration Package Operations As Sysadmin, that means they can modify, export, and import migration packages that include artifacts they may not otherwise have permission to operate on.

1. In the Administration application, expand the **Tools** menu and click **Migration Packages**. Anzo displays the Migration Packages screen, which lists any existing packages. For example:



<input type="checkbox"/>	Title	# Core Artifacts	# Included Artifacts	Last Refresh	Updated Date	Actions
<input type="checkbox"/>	All Graphmarts	3	24	4 minutes ago	Nov 18, 2022	⋮
<input type="checkbox"/>	Flight-related Data	1	5	a minute ago	Nov 18, 2022	⋮

Rows per page: 25 1-2 of 2 < >

2. Click the name of the migration package that you want to export. The Details tab for the package is displayed. For example:

The screenshot shows the 'All Graphmarts' interface. At the top, there are buttons for 'Delete', 'Save A Copy', and 'Export'. Below these are tabs for 'Details', 'Included Artifacts', 'Discussion', and 'Sharing'. The 'Details' tab is active, showing 'Export Security' information and 'Configuration' options. The 'Configuration' section includes 'Export File Format', 'File Per Category', 'Exported ACLs Handling', and 'Use Existing ACLs as is'. Under 'Export Options', there are checkboxes for 'Generate Variable Template File', 'Include Registry Statements', and 'Include Dataset Editions and Components'. A 'General' sidebar on the right shows package details like 'Type', 'Creator', 'Updated', and 'Released'. At the bottom, there is a 'Core Members' section with a search bar and a table of artifacts.

Title	Type	Available	# Included Artifacts	Actions
Movies	Graphmart	✓	9	✕
Northwind	Graphmart	✓	8	✕
Tickets	Graphmart	✓	9	✕

- If desired, you can change the export configuration by adjusting the Configuration settings at the top of the screen. For details about the options, refer to [Export Configuration Settings Reference](#).
- If **Generate Variable Template File** is enabled and you want to change replaceable property values before performing the export and generating the template, you can click the **Included Artifacts** tab. The properties with editable values can be expanded by clicking the > character next to the artifact. To filter the list to show only the rows that have replaceable values, you can select the **Only show rows containing Replaceable Statements** checkbox at the top of the screen. For example:

Type	Title	Included For	Included In Export
File Backed Linked Data Set	Movies	Movies	<input checked="" type="checkbox"/>
Subject	Statement	Current Value	Export Replacement Value
http://openanzo.org/DataLocatio...	http://cambridgesemantics.com/...	/store/movies/	None
File Backed Linked Data Set	Northwind	Northwind	<input checked="" type="checkbox"/>
Subject	Statement	Current Value	Export Replacement Value
http://openanzo.org/DataLocatio...	http://cambridgesemantics.com/...	/store/northwind/	None

Rows per page: 25 1-4 of 4

- To edit a statement, click the value in the **Template Value** column and replace the placeholder text with the desired value. Then click the checkmark icon (✓) to save the change. Any changes you make on the Included Artifacts tab will be included in the variable template that is generated during the export. For example, in the image below the File Connection placeholder is replaced with the path on the target server.

Subject	Statement	Current Value	Export Replacement Value
Server Shared Filesystem	http://cambridgesemantics.com/...	/nfs/data	<input type="text" value="/opt/anzoshare/data"/>

- When you are ready to export the package, click the **Export** button. Anzo exports each of the included artifacts as TriG files and packages the TriG files into a .zip file. The contents of the .zip file are laid out according to the specified Export File Format. Once the package is assembled, the .zip file is automatically downloaded to your computer.

## Note

If changes were made to the artifacts since they were added to the package and the package was not refreshed before the export, Anzo automatically creates a version of the changed artifacts.

Once the package is exported, you can extract the file to access any generated templates and to place the artifacts in source control if that is part of your organization's process. For information about working with the generated templates, see [Editing Migration Package Template Files](#). When you are ready to import the package to the target server, see [Importing a Migration Package](#).



# Export Configuration Settings Reference

This topic describes the Export Configuration options that are available on the Details tab when creating or configuring a Migration Package.

Details	Included Artifacts	Discussion	Sharing
<b>Export Security</b> This package will export using permissions allocated to <b>System Administrator</b>			
<b>Configuration</b>			
Export File Format		Export Options	
File Per Category	▼	<input type="checkbox"/>	Generate Variable Template File
Exported ACLs Handling		<input checked="" type="checkbox"/>	Include Registry Statements
Use Existing ACLs as is	▼	<input type="checkbox"/>	Include Dataset Editions and Components

- [Export File Format](#)
- [Exported ACLs Handling](#)
- [Generate Variable Template File](#)
- [Include Registry Statements](#)
- [Include Dataset Editions and Components](#)

## Export File Format

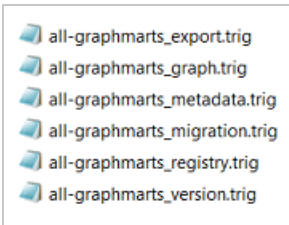
This option configures the file structure of the exported .zip package. There are three options to choose from:

- [File Per Category](#)
- [File Per Graph](#)
- [Folder Per Type](#)

### File Per Category

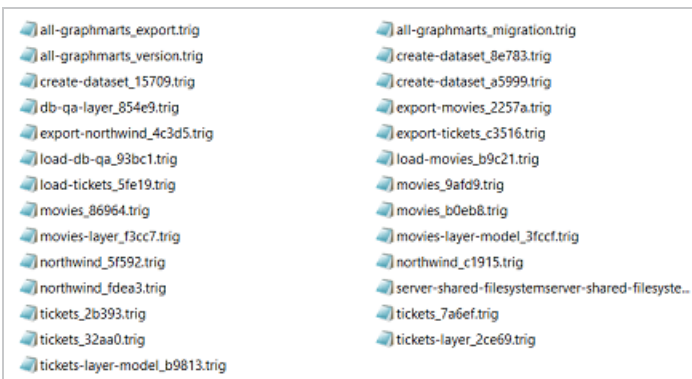
This option (the default setting) creates one TriG file per type or category of information that is included in the export. This is the same as the layout of files that results when you export an artifact from the Versions tab in the Anzo application. If **File Per Category** is selected, the exported

package contains one file per each of the following categories: Export, Migration, Versions, Metadata, Registries, and Graph. The files that are generated depend on the chosen Export Options. The relevant information for all of the included artifacts is written to the same category file. For example, the image below shows the contents of a package that was exported with Export File Format set to **File Per Category**. The Migration Package name is "All Graphmarts."



## File Per Graph

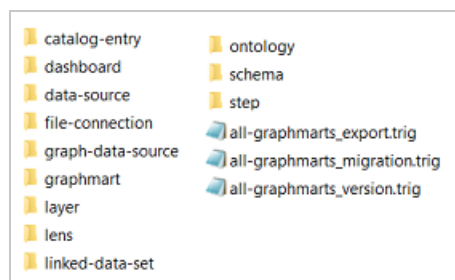
This option creates one TriG file per graph. Unlike the Files Per Category option, where data is separated by type of information, each graph file contains all of the data that is related to that graph, such as the metadata and registry information. For example, the image below shows the contents of a package that was exported with Export File Format set to **File Per Graph**. There is a TriG file for each data source graph, layer graph, model graph, dashboard graph, etc. The Migration Package name is "All Graphmarts."



## Folder Per Type

Like the File Per Graph option, this option creates one TriG file per graph, where each graph file contains all of the data that is related to the graph, such as the metadata and registry information. However, the graph files are organized into subdirectories by base graph type, such as data source,

graphmart, layer, schema, etc. The folders contain all graphs of that type for all of the included artifacts. For example, the image below shows the contents of a package that was exported with Export File Format set to **Folder Per Type**. The Migration Package name is "All Graphmarts."



## Exported ACLs Handling

This option determines how to handle the ACL configuration for the artifacts in the package. There are three options to choose from:

- [Use Existing ACLs as is](#)
- [Generate Access Control Template File](#)
- [Use the Destination Server's Default Access Policy on Import](#)

### Use Existing ACLs as is

This option exports the ACL metadata for all of the artifacts as-is. No template file will be generated and the artifacts will be imported into the target system with the same permissions as the artifacts on the source system.

### Generate Access Control Template File

This option generates a template file that contains access control statements with placeholder values in the objects. You replace the placeholder values with the Group or User URIs that should have permission to access all of the artifacts in the Migration Package. For more information about the template, see [Editing Migration Package Template Files](#).

## Use the Destination Server's Default Access Policy on Import

This option excludes the ACL metadata when exporting the artifacts in the package. That means the artifacts will inherit the access policies from the target server when they are imported. If the artifacts already exist in the target system, the existing ACLs on those artifacts remain in place. If the artifacts do not exist in the target system, they will inherit the policies that are configured in the target system's Default Access Policies.

## Generate Variable Template File

This option indicates whether to generate a Variable Template File in the export package. A Variable Template File is a TriG file that contains statements for all of the properties that have replaceable values. Properties with replaceable values are objects such as file paths and Anzo Data Store locations, which might differ on the source and target Anzo servers. The template that is generated has placeholder text that you replace with the desired values for the target server. If you want a template to be generated, select the **Generate Variable Template File** checkbox. If you do not want to make changes to artifacts before they are imported to the target system, clear the **Generate Variable Template File** checkbox. For more information about the template, see [Editing Migration Package Template Files](#).

## Include Registry Statements

This option is selected by default and indicates whether to export the registry statements for the artifacts in the package. A registry is like a container for all artifacts of a certain type. For example, the Data Sources Registry stores information about all of the Data Sources. Registry statements should be included in exports except in rare cases when you do not intend to import the Migration Package back into Anzo. When registry statements are not included in an export, the imported artifacts are not displayed in Anzo. For example, if a data source artifact is imported without registry statements, it would not be added to the Data Sources registry and therefore not be displayed in the list of Data Sources in the Anzo application.

## **Include Dataset Editions and Components**

This option specifies whether the export includes all of the editions and components for each dataset in the package. When **Include Dataset Editions and Components** is selected, the exported package includes the Managed and Saved Editions and all of their components for each dataset.

## Editing Migration Package Template Files

If a Variable Template File and/or Access Control Template File is included in a Migration Package export, the files must be edited to replace all of the placeholder values before the package can be imported to the target server. This topic describes the template files and provides guidance on editing the templates.

- [Editing a Variable Template File](#)
- [Editing an Access Control Template File](#)

### Editing a Variable Template File

A Variable Template File is a TriG file that is used to define the values to use in Replaceable Statements. Properties with replaceable values are objects such as File Connection paths, Anzo Data Store locations, File-Backed Linked Data Set locations, and file locations for file-based Data Sources, which might differ on the source and target Anzo servers. The template that is generated has placeholder text that you replace with the desired values for the target server. **The values that you specify are applied to all artifacts included the Migration Package.**

When you open a Variable Template File, the placeholder values are denoted by three hash characters (###) and all capital letters, for example, ###FILEPATH-1###. The example below shows a snippet of a Variable Template File. The placeholder text is shown in **bold**:

```
<http://openanzo.org/ReplacementObject/10441bd7-03fb-494a-b56a-cc0eea32aed2> {
  <http://cambridgesemantics.com/PathConnection/6bd218a15c644045ba43f007c824d830>
  <http://cambridgesemantics.com/ontologies/DataSources#filePath> "###FILEPATH-2###" .

  <http://openanzo.org/ReplacementObject/10441bd7-03fb-494a-b56a-cc0eea32aed2> a
  <http://cambridgesemantics.com/ontologies/2021/06/Migration#ReplacementObject> ;
  <http://cambridgesemantics.com/ontologies/2021/06/Migration#forGraph>
  <http://cambridgesemantics.com/CSVDataSource/f9a54e23d83549999536782b5de1981c> .
}

<http://openanzo.org/ReplacementObject/1c79ae8a-c570-4170-a9f8-f1a5dd967c6d> {
  <http://csi.com/DataLocation/157ae35ecab30f803c754d314be18e44>
  <http://cambridgesemantics.com/ontologies/DataSources#filePath> "###FILEPATH-9###" .

  <http://openanzo.org/ReplacementObject/1c79ae8a-c570-4170-a9f8-f1a5dd967c6d> a
```

```

<http://cambridgesemantics.com/ontologies/2021/06/Migration#ReplacementObject> ;
  <http://cambridgesemantics.com/ontologies/2021/06/Migration#forGraph>
<http://csi.com/FileBasedLinkedDataSet/157ae35ecab30f803c754d314be18e44> .
}

<http://openanzo.org/ReplacementObject/21f24dda-fba6-47d7-bb6d-90e1fafec623> {
  <http://csi.com/DataLocation/7214e9ec270347dabecbfc7328b4bed>
<http://cambridgesemantics.com/ontologies/DataSources#filePath> ###FILEPATH-7### .

  <http://openanzo.org/ReplacementObject/21f24dda-fba6-47d7-bb6d-90e1fafec623> a
<http://cambridgesemantics.com/ontologies/2021/06/Migration#ReplacementObject> ;
  <http://cambridgesemantics.com/ontologies/2021/06/Migration#forGraph>
<http://csi.com/FileBasedLinkedDataSet/7214e9ec270347dabecbfc7328b4bed> .
}

```

When replacing the placeholder text, edit the text inside the quotation marks. All objects should retain the quotes. If the replacement value is a URI, place the URI inside the quotation marks.

## Editing an Access Control Template File

The Access Control Template is a TriG file that is used to define the permissions to be assigned on **all artifacts included in the Migration Package**. The template contains two sets of statements, one set for the artifact graphs

(`<http://openanzo.org/namedGraphs/reserved/graphs/defaultGraphTemplate>`) and one for the artifact metadata graphs (

`<http://openanzo.org/namedGraphs/reserved/graphs/defaultMetadataGraphTemplate>`). The objects in the template are placeholder URIs that must be replaced with the Group and/or User URIs on the target server. A copy of the template is shown below. The placeholder URIs are shown in **bold**:

```

<http://openanzo.org/namedGraphs/AclTemplate> {
  <http://openanzo.org/namedGraphs/reserved/graphs/defaultGraphTemplate>
<http://openanzo.org/ontologies/2008/07/Anzo#canBeAddedToBy> <urn://ACL-ADD-ROLE-PLACEHOLDER> ;
  <http://openanzo.org/ontologies/2008/07/Anzo#canBeReadBy> <urn://ACL-READ-ROLE-PLACEHOLDER> ;
  <http://openanzo.org/ontologies/2008/07/Anzo#canBeRemovedFromBy> <urn://ACL-REMOVE-ROLE-PLACEHOLDER> .
}

```

```

<http://openanzo.org/namedGraphs/reserved/graphs/defaultMetadataGraphTemplate>
<http://openanzo.org/ontologies/2008/07/Anzo#canBeAddedToBy> <urn://ACL-METAADD-ROLE-
PLACEHOLDER> ;
    <http://openanzo.org/ontologies/2008/07/Anzo#canBeReadBy> <urn://ACL-METAREAD-ROLE-
PLACEHOLDER> ;
    <http://openanzo.org/ontologies/2008/07/Anzo#canBeRemovedFromBy> <urn://ACL-
METAREMOVE-ROLE-PLACEHOLDER> .
}

```

The **defaultGraphTemplate** statements configure who can view, modify, and delete the artifact. The **defaultMetadataGraphTemplate** statements configure who can view, modify, and delete artifact metadata, such as an artifact's permissions. The list below describes how the template properties map to permissions:

- **canBeReadBy**: This property assigns **View** and **Meta View** permissions. On the **defaultGraphTemplate**, this property assigns **View**, which grants access to see the artifact but not change it. On the **defaultMetadataGraphTemplate**, this property assigns **Meta View**, which grants access to see the artifact's permissions but not change them.
- **canBeAddedToBy**: This property assigns **Add/Edit** and **Meta Add/Edit** permissions. On the **defaultGraphTemplate**, this property assigns **Add/Edit**, which grants permission to change the artifact or add an entity to it, such as to add a Schema to a Data Source. On the **defaultMetadataGraphTemplate**, this property assigns **Meta Add/Edit**, which grants permission to change the artifact's permissions.
- **canBeRemovedFromBy**: This property assigns **Delete** and **Meta Delete** permissions. On the **defaultGraphTemplate**, this property assigns **Delete**, which grants permission to delete an entity from an artifact, such as to delete a Data Layer from a Graphmart. On the **defaultMetadataGraphTemplate**, this property assigns **Meta Delete**, which grants permission to delete the parent artifact and change the artifact's permissions.

### Tip

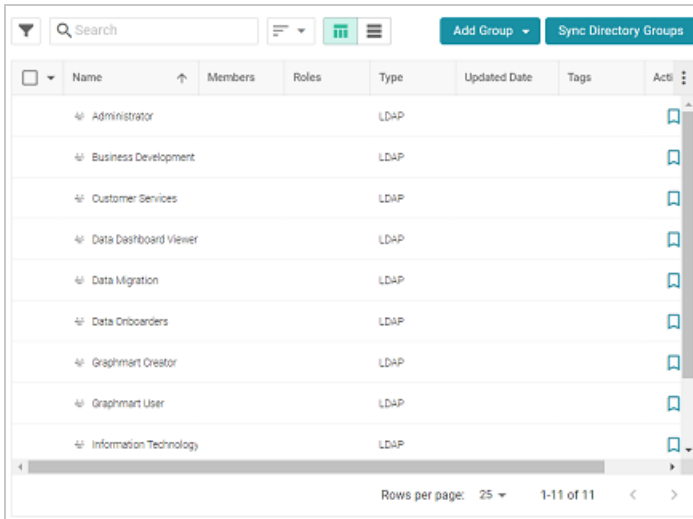
For more information about artifact permissions, see [Permission Settings](#) in the User Guide.



## Finding Group and User URIs

In order to complete the Access Control Template and give groups access to the artifacts in the package, you need to find the Group and/or User URIs on the target server to add as objects to the template properties (`canBeReadBy`, `canBeAddedToBy`, and `canBeRemovedFromBy`).

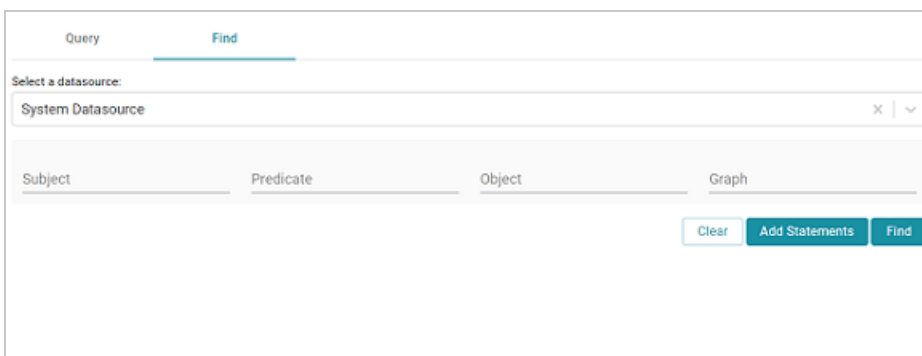
1. If you need to review a list of the Groups that are available on the target system, open the Administration application on that server. To access the Group names, expand the **User Management** menu and click **Groups**. For example:



The screenshot shows a web application interface for group management. At the top, there is a search bar and two buttons: "Add Group" and "Sync Directory Groups". Below this is a table with columns: Name, Members, Roles, Type, Updated Date, Tags, and Actions. The table lists several groups, all of which are of type "LDAP". The groups listed are: Administrator, Business Development, Customer Services, Data Dashboard Viewer, Data Migration, Data Onboarders, Graphmart Creator, Graphmart User, and Information Technology. At the bottom of the table, there is a pagination control showing "Rows per page: 25" and "1-11 of 11".

Name	Members	Roles	Type	Updated Date	Tags	Acti
Administrator			LDAP			
Business Development			LDAP			
Customer Services			LDAP			
Data Dashboard Viewer			LDAP			
Data Migration			LDAP			
Data Onboarders			LDAP			
Graphmart Creator			LDAP			
Graphmart User			LDAP			
Information Technology			LDAP			

2. Note the names of the groups whose URIs you want to add to the template.
3. Next, find the URIs for the group names. In the Anzo application on the target server, expand the **Access** menu and click **Query Builder**. Anzo displays the Query tab. Click the **Find** tab.



The screenshot shows the Anzo Query Builder interface. At the top, there are two tabs: "Query" and "Find", with "Find" being the active tab. Below the tabs, there is a section labeled "Select a datasource:" with a dropdown menu showing "System Datasource". Below this, there are four input fields: "Subject", "Predicate", "Object", and "Graph". At the bottom right, there are three buttons: "Clear", "Add Statements", and "Find".

4. On the Find tab, leave the datasource set to **System Datasource** and then type a Group Name in the **Object** field. For example:

- Next, click the **Find** button. The result is a statement that defines that Group. The value in the **Subject** position is the URI for the Group. For example:

Subject	Predicate	Object
<ldap:///cn=data%20onboarders,ou=groups,dc=acme,dc=com>	<http://xmlns.com/foaf/0.1/name>	<Data Onboarders>

- Click the URI to add it to the **Subject** field at the top of the screen, and then copy the URI from that field. For example, the URI copied from the image above is  
<ldap:///cn=data%20onboarders,ou=groups,dc=acme,dc=com>.

Repeat the steps above to find all of the URIs that you want to add to the template. To add URIs to the file, replace each of the placeholder URIs. You can add multiple URIs to a property in a comma-separated list. For example:

```
<http://openanzo.org/namedGraphs/AclTemplate> {
  <http://openanzo.org/namedGraphs/reserved/graphs/defaultGraphTemplate>
<http://openanzo.org/ontologies/2008/07/Anzo#canBeAddedToBy>
  <ldap:///cn=data%20onboarders,ou=groups,dc=acme,dc=com>,
<ldap:///cn=graphmart%20creator,ou=groups,dc=acme,dc=com>,
<ldap:///cn=graphmart%20user,ou=groups,dc=acme,dc=com> ;
  <http://openanzo.org/ontologies/2008/07/Anzo#canBeReadBy>
<ldap:///cn=graphmart%20user,ou=groups,dc=acme,dc=com> ;
  <http://openanzo.org/ontologies/2008/07/Anzo#canBeRemovedFromBy>
<ldap:///cn=administrator,ou=groups,dc=acme,dc=com> .

  <http://openanzo.org/namedGraphs/reserved/graphs/defaultMetadataGraphTemplate>
<http://openanzo.org/ontologies/2008/07/Anzo#canBeAddedToBy>
<ldap:///cn=data%20onboarders,ou=groups,dc=acme,dc=com> ;
  <http://openanzo.org/ontologies/2008/07/Anzo#canBeReadBy>
```

```
<ldap:///cn=graphmart%20user,ou=groups,dc=acme,dc=com> ;  
  <http://openanzo.org/ontologies/2008/07/Anzo#canBeRemovedFromBy>  
<ldap:///cn=administrator,ou=groups,dc=acme,dc=com> .  
}
```

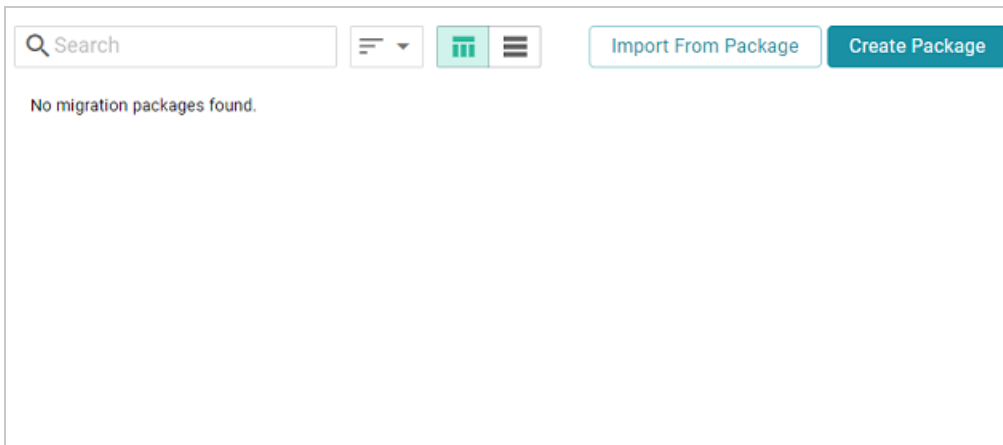
## Importing a Migration Package

Follow the instructions below to import a Migration Package.

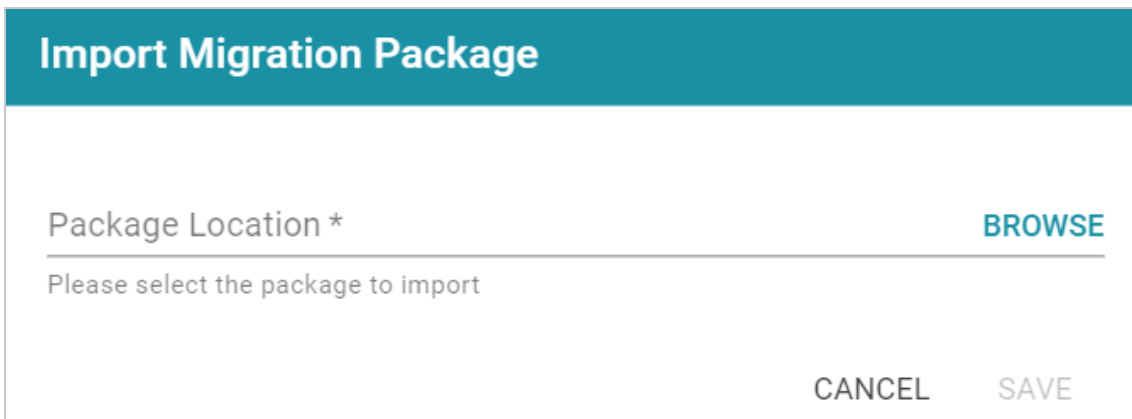
### Note

There are two permissions that control access to export, modify, and import migration packages: **Manage Migration Packages** and **Perform Migration Package Operations As Sysadmin**. If a user has only the Manage Migration Packages permission, they cannot modify, export, or import artifacts in packages unless they have the appropriate permissions on the artifacts. If a user has Perform Migration Package Operations As Sysadmin, that means they can modify, export, and import migration packages that include artifacts they may not otherwise have permission to operate on.

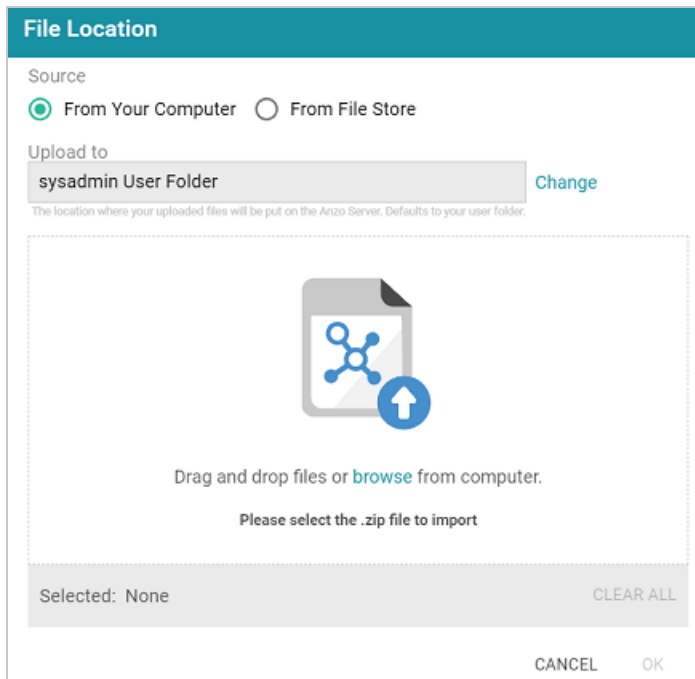
1. First, if the package contains ACL and/or Variable Template files that were exported from the source Anzo server, make sure the files have been completed; all of the placeholder values are replaced with the desired values for the target server. For information about the templates, see [Editing Migration Package Template Files](#).
2. If the package is unpacked, compress the directory to a .zip file so that it can be imported. Any template files should be included inside the .zip file. The package can be imported from your computer or a location on the target server's File Store.
3. In the Administration application on the target server, expand the **Tools** menu and click **Migration Packages**. Anzo displays the Migration Packages screen, which lists any packages that were created on this server. For example, the image below shows a target system where packages have been imported but not created:



4. Click the **Import From Package** button at the top of the screen. The Import Migration Package dialog box is displayed:



5. Click **Browse** or the **Package Location** field to open the File Location dialog box.



6. Depending on the location of the package to import, follow the appropriate instructions below:

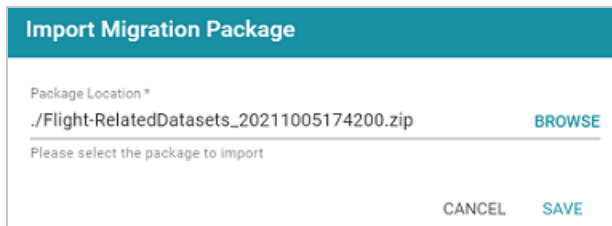
- If the package is on your computer, leave **From Your Computer** selected and drag and drop the file to the dialog box or click **browse** and select the file.

### Tip

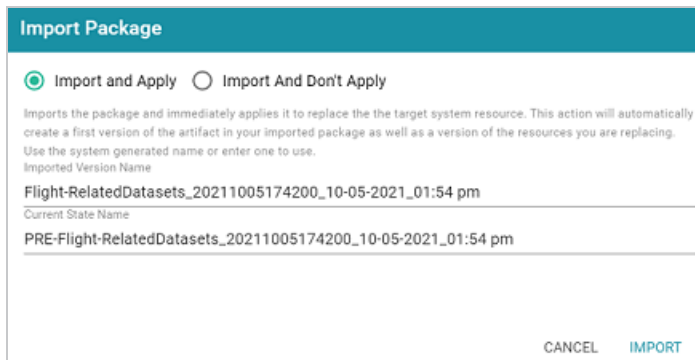
As a best practice when uploading files from your computer, check the upload location that is listed in the **Upload To** field by hovering your pointer over the value to view the full path as a tooltip. Make sure the upload location is set to the desired directory. If necessary, you can click **Change** and select a different upload path.

- If the package is on the File Store, select **From File Store**. Navigate to the location of the .zip file on the store and select it.

7. Click **OK** to add the file location to the Package Location field. For example:



8. Click **Save** to save the import configuration. Anzo validates the import by checking whether any included template files are completed. If the import is valid, the Import Package screen is displayed. For example:



9. On the Import Package screen, specify how you want the artifacts to be applied to the target server, either **Import And Apply** or **Import And Don't Apply**:

### **Import And Apply**

Selecting this option means the artifacts included in the package should be applied to the target system as the current, working versions of the artifacts. When **Import And Apply** is selected, Anzo follows the procedure below:

- a. If the artifacts to be imported also exist on the target system, Anzo compares the existing version with the import version. If the artifacts differ, Anzo creates a backup version of the existing artifacts. If the artifacts match, Anzo does not create backup versions of the existing artifacts.
- b. Next, Anzo imports the artifacts from the package as versions. This ensures that the target server includes a copy of the artifacts exactly as they were originally imported.

- c. The imported version of the artifacts are applied as the current, working version. In other words, the current version is now derived from the imported version and is given a **Derived from: <imported\_version\_name>** label. For example, the image below shows the label for a Dataset that was derived from an imported version.



## Import And Don't Apply

Selecting this option means the artifacts included in the package should not be applied to the target system as the current, working versions. When **Import And Don't Apply** is selected, Anzo imports the artifacts as backup versions and does not replace the current versions of any existing artifacts.

- Next, you have the option to modify the auto-generated names for the versions that are created during the import:
  - If you selected **Import And Apply**, you can edit the following values:
    - Imported Version Name:** This is the name of the new version that is created by the import.
    - Current State Name:** If the existing version of an artifact differs from the imported version, this is the name to give the backup version of the current state before the imported version is applied.
  - If you selected **Import And Don't Apply**, you can edit the **Imported Version Name** value to specify the name to give the imported version.
- When you are ready to import the package, click **Import**. Anzo imports all of the artifacts according to your import configuration.

### Note

When the import is complete, the imported package is not displayed in the list of Migration Packages. The Migration Packages screen displays only the packages that are created on this server.



# User Management

Anzo offers granular artifact and data access control as well as role-based security for controlling access to the Anzo applications and features. This section provides setup and administration information for role-based access control. The topics include instructions for connecting to your central directory server, connecting to an identity provider for SSO access, and configuring users, groups, roles, and permissions in Anzo.

## Tip

When planning the user and access management solution for your system, Cambridge Semantics recommends that you refer to [User Management and Access Control Concepts](#) to learn about the fundamental concepts behind Anzo's access control implementation.

### In this section:

<a href="#">User Management and Access Control Concepts</a> .....	154
<a href="#">Connecting to a Directory Server</a> .....	164
<a href="#">Adding Directory Users and Groups to Anzo</a> .....	173
<a href="#">Enabling Self-Authorization for Directory Users</a> .....	177
<a href="#">Configuring Single Sign-On Authentication</a> .....	179
<a href="#">Creating and Managing Roles</a> .....	216
<a href="#">Creating an Internal Anzo User</a> .....	222
<a href="#">Predefined Anzo Roles and Permissions</a> .....	225
<a href="#">Role Permissions Reference</a> .....	233
<a href="#">Managing Default Access Policies</a> .....	241

# User Management and Access Control Concepts

The topics in this section provide an overview of user management and access control in Anzo and introduce the key concepts to consider when planning and implementing user and data access management for your system.

**In this section:**

User Management Concepts .....	155
Artifact Access Control Concepts .....	159

# User Management Concepts

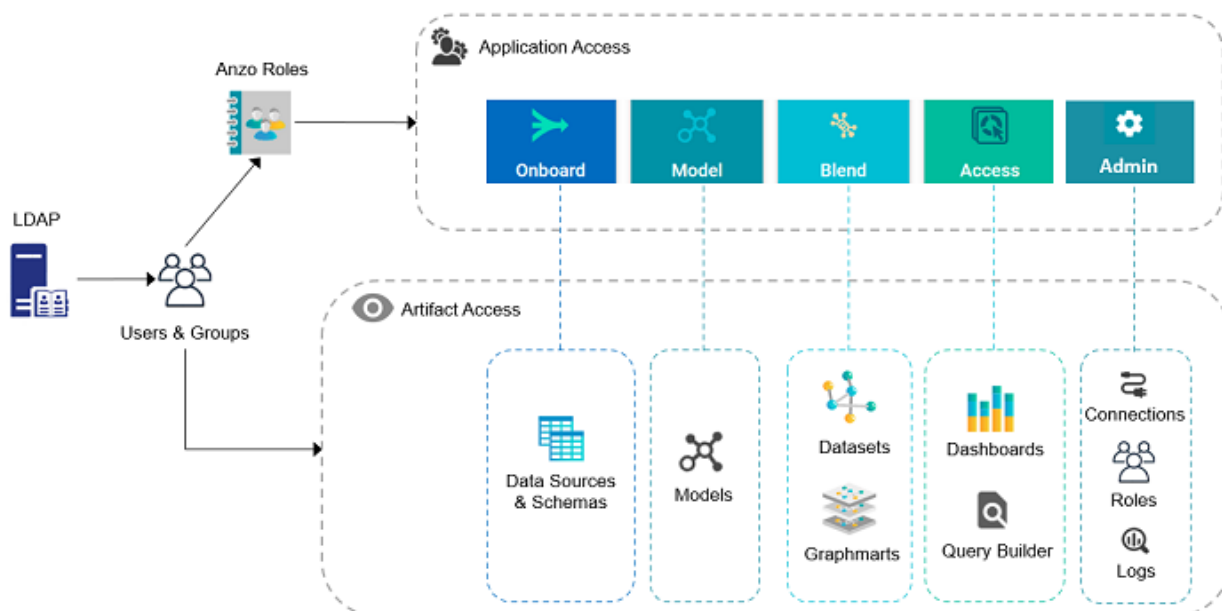
Typically organizations connect Anzo to their central directory server and then add users and groups from the server to Anzo. Once the accounts are added to Anzo, access control is managed in two ways:

1. Groups (or users) are added to **Roles** and the roles are configured to grant access to *functionality* in Anzo. Role permissions grant access to menus and screens in the Anzo and Administration applications. Access to functionality cannot be assigned to groups or users, only to roles.
2. Groups and users are used to control access to individual artifacts—data sources, models, graphmarts, etc.—and your data that is stored in Anzo.

## Note

Though Anzo is flexible and allows you to assign artifact access to roles, the recommendation is to control access to artifacts with users and groups and reserve roles for granting access to functions in the applications.

The following diagram illustrates the concepts of roles and groups in Anzo:

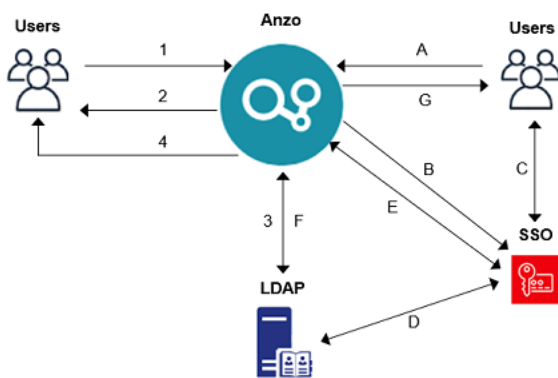


A user's role determines whether they can access the **Onboard** menu and create a new data source or see the **Blend** menu and create a new graphmart. But their group assignment determines whether they can view, modify, or delete data source and graphmart artifacts that are created by other users.

For more information about leveraging a directory server and details about users, groups, and roles see the sections below.

## Leveraging a Directory Server (LDAP)

Anzo can be configured to access your directory server via Direct Authorization or Single Sign-On (SSO). The diagram below shows the procedures that are followed for both methods. The left side of the diagram (the numbered steps) shows the direct authorization method. The right side of the diagram (the lettered steps) shows the SSO method. The table below the diagram describes the processes for each method.



Direct Authorization	Single Sign-On
1. A new (unknown) user navigates to the Anzo application.	A. A new (unknown) user navigates to the Anzo application.
2. Anzo redirects the user to a login form. The user supplies credentials and submits the form.	B. Anzo redirects the user to the SSO provider. The SSO provider controls authentication validation.
3. Anzo queries the LDAP for the user and group membership.	C. Depending on the policy, the SSO provider presents a login screen for the user to complete and submit.

Direct Authorization	Single Sign-On
<p>4. Anzo redirects the user to the application with the appropriate roles applied.</p>	<p>D. As needed, the SSO provider validates the credentials with the LDAP server.</p> <p>E. The SSO provider authenticates the Anzo session with a callback.</p> <p>F. Anzo fetches group information from the LDAP server.</p> <div data-bbox="760 583 1448 905" style="background-color: #e6f2ff; padding: 10px; border: 1px solid #add8e6;"> <p><b>Note</b> For SSO-configured systems, Anzo currently requires direct access to the LDAP directory (and a bind user) to look up groups.</p> </div> <p>G. Anzo redirects the user to the application with the appropriate roles applied.</p>

For more information on connecting to a directory server, see the following topics:

- [Connecting to a Directory Server](#)
- [Configuring Single Sign-On Authentication](#)

## Users and Groups

Groups typically originate in a directory server and are synced to Anzo. However, you can also create custom groups that are internal to Anzo. Typically users also originate from the directory server, but you can create user accounts in Anzo. Any users and groups that are created in Anzo are stored in Anzo's internal LDAP server.

For information about retrieving user and groups from the directory server or creating internal Anzo users, see the following topics:

- [Adding Directory Users and Groups to Anzo](#)
- [Creating an Internal Anzo User](#)

## Roles

Anzo is configured with predefined roles. You can create new roles and disregard the predefined roles, remove the predefined roles, or add your groups to the predefined roles and modify the assigned permissions as needed.

For details about the default roles and instructions on creating new roles, see the following topics:

- [Predefined Anzo Roles and Permissions](#)
- [Creating and Managing Roles](#)

## Permissions

The way you give a role access to the Anzo applications and particular functions in those applications is to assign permissions to the role. All permissions are predefined in Anzo. Custom permissions cannot be created, and the predefined permissions cannot be deleted.

For details about all of the permissions, see the following topic:

- [Role Permissions Reference](#)

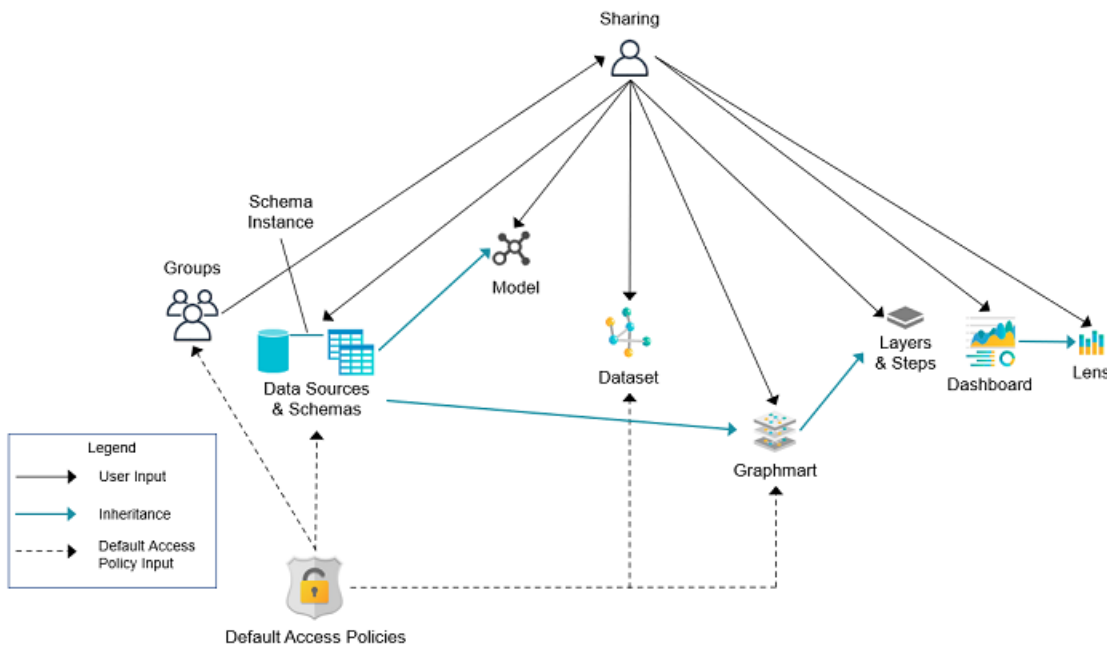
For an overview of the data access management concepts, see [Artifact Access Control Concepts](#).

# Artifact Access Control Concepts

The implementation of artifact and data access control in Anzo is an aggregation of three mechanisms:

1. **Default Access Policies:** These are the base permissions that are applied to artifacts by default when they are created. For most types of artifacts, the access control that is supplied by a Default Access Policy is augmented by the other two access control mechanisms.
2. **Permission Inheritance:** To facilitate common workflows, the Anzo application applies logic so that artifacts in the same workflow inherit the same permissions. For example, when a user creates a data source and adds a schema, the schema inherits permissions from the data source. This permission inheritance is applied in addition to the applicable Default Access Policy.
3. **Sharing:** An artifact's creator can also share access to their artifact with other users or groups. When an artifact is shared, those user-configured permissions are applied in addition to any permissions that were inherited.

The following diagram illustrates the above concepts. Details about the processes and components depicted in the diagram are provided in the sections below.



## Default Access Policies

Default Access Policies are the security policies that are applied by default to the artifacts that belong to a particular system **registry** (see [Registries](#) below). Default Access Policies are the base permissions that get assigned when an artifact is created—before any other access control logic (e.g., [Permission Inheritance](#)) is applied. Any artifact-level logic that is applied by Anzo or configured from the **Sharing** tab in the Anzo application augments the permissions that were supplied by the Default Access Policy.

For more information about Default Access Policies, see the following topic:

- [Managing Default Access Policies](#)

## Registries

A registry is a system-level graph that stores metadata about artifacts of the same type. For example, a Data Sources Registry stores metadata about all of the data source and schema artifacts, and an Ontology Registry stores metadata about all of the model artifacts. Like onboarded data, registries are stored and managed as RDF named graphs according to system ontologies.

### Important

Aside from changing the Default Access Policy for a registry, do not make additional modifications to registries. Changing or removing a registry can irreparably damage your Anzo server.

## Permission Inheritance

The concept of inheritance is fundamental to the implementation of access control in Anzo. Inheritance allows related entities to share permissions with each other, making access easier to manage collectively, and ensuring that users have the appropriate access to each of the dependent artifacts that are crucial to their workflow. The following subsections describe the relationships and inheritance rules for each type of artifact.



- [Data Sources & Schemas](#)
- [Graphmarts](#)
- [Unstructured Pipelines](#)
- [Users and Roles](#)
- [Role Permissions and Registries](#)

## Data Sources & Schemas

Data sources and schemas have a fundamental relationship since schemas are imported from data sources and, in a sense, belong to them. Because a data source can have more than one schema and the schemas can be managed independently, data sources and schemas exist as separate artifacts in Anzo. However, because of their implicit relationship, Anzo uses inheritance to facilitate users' interaction with data sources and the schemas created from them.

If Anzo did not apply inheritance, a user who shares a data source would have to remember to add the new user to the data source *and* navigate to each related schema and add the new user there as well. Keeping permissions in sync manually presents a big challenge that is curtailed by applying inheritance.

To summarize the inheritance rules for data sources and schemas:

- Schemas inherit from the data source from which they were imported.
- Schema instances, which link schemas to their data source, inherit from both the schema and the data source.

## Graphmarts

When a user creates a graphmart from scratch, the graphmart is assigned permissions according to the [Graphmarts Registry](#) Default Access Policy. When a user creates a graphmart from a data source, the graphmart inherits permissions from the source schema.

Graphmarts contain data layers that describe and group the transformations that take place as the knowledge graph is generated. Since layers are created in the context of a graphmart, they inherit their permissions from the graphmart by default.

If Anzo did not apply this inheritance, a user who wanted to share a graphmart would have to remember to configure each newly created layer to assign permissions that match the graphmart's permissions. Otherwise someone who had access to the graphmart would not be able to view or edit its layers and steps.

To summarize the inheritance rules for graphmarts:

- Graphmarts created from scratch via the Add Graphmart button inherit permissions from the Graphmarts Registry Default Access Policy.
- Graphmarts created from a data source inherit permissions from the source.
- Data layers and steps created in a graphmart inherit from the graphmart.

For more information about graphmart permissions, see [Sharing Access to Graphmarts](#) in the User Guide.

## Unstructured Pipelines

Running an unstructured pipeline produces a dataset, which inherits its permissions from the pipeline. Additionally, each pipeline run produces a status dataset that is specific to the pipeline's execution. Since these status datasets are implicitly related to the unstructured pipeline, they inherit permissions from the pipeline.

To summarize the inheritance rules for unstructured pipelines:

- Datasets created from unstructured pipeline runs inherit from the corresponding pipeline.
- Pipeline status datasets inherit from the related unstructured pipeline. From an end user's perspective, this relates to the status information that is displayed in the Unstructured Pipeline user interface.

## Users and Roles

Users and roles are typically managed by administrators as a collective group. There are not clear use cases for a given user to manage some user and role accounts but not others. The expectation is that users who have the **Manage Users, Groups, and Roles** permission should be able to manage all users and roles, not just a subset of them.

To accomplish the above expectation, all users inherit permissions from one system registry, the **Role and Permissions Registry**. If user and role permissions were not centralized, there could be circumstances where one user creates a new user or role in Anzo and other users cannot see or edit that account even if they belong to a role that has the Manage Users, Groups, and Roles permission. Also if the original user or role creator had the Manage Users, Groups, and Roles permission revoked, they may retain control over the accounts they created when they had the ability to do so.

To summarize the inheritance rules for users and roles:

- Anyone who has the **Manage Users, Groups, and Roles** permission has the **Admin** level of access to all users, groups, and roles.
- The **Everyone** role has **View** access to all users, groups, and roles so that they can share artifacts with other users and groups.

## Role Permissions and Registries

Access to certain registries is mapped to specific Anzo permissions. This is helpful when artifacts that are added to a registry inherit their permissions from the registry itself rather than another artifact, such as with [Users and Roles](#). When users have a permission that grants them access to a registry, that means they can see all artifacts that belong to that registry.

The list below describes the registry access that is controlled by a permission.

- Access to the Role and Permissions Registry is granted by the **Manage Users, Groups, and Role** permission.

For more information about the Anzo permissions, see [Role Permissions Reference](#).

## Sharing

Artifacts can be shared with other users and groups from the artifact's **Sharing** tab in the Anzo application. When an artifact is shared, those user-defined permissions are added to the set of permissions that came from the Default Access Policy for the related registry as well as the permission inheritance that is applied by Anzo.

For details about artifact sharing, see [Share Access to Artifacts](#) in the User Guide.

# Connecting to a Directory Server

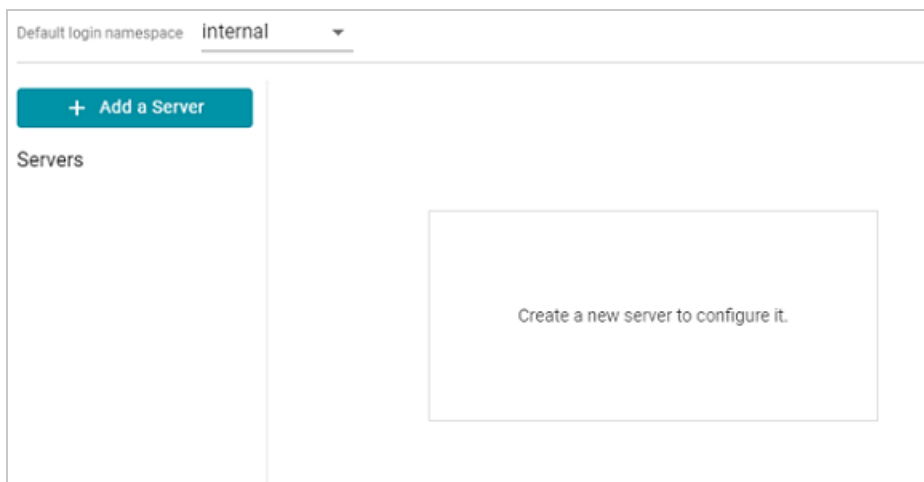
This section provides instructions for connecting to a directory server and mapping the user and group configuration to Anzo so that Anzo can leverage the users and groups from the server.

- [Connect to the Server](#)
- [Map Users to Anzo](#)
- [Map Groups to Anzo](#)

## Connect to the Server

Follow the steps below to create a connection between Anzo and your directory server.

1. In the Administration application, expand the **User Management** menu and click **Directory**. Anzo displays the Directory screen. For example:



2. On the Directory screen, click the **Add a Server** button. Anzo displays the Create New Server Configuration screen.

The screenshot shows a 'Create New Server Configuration' dialog box with the following fields and options:

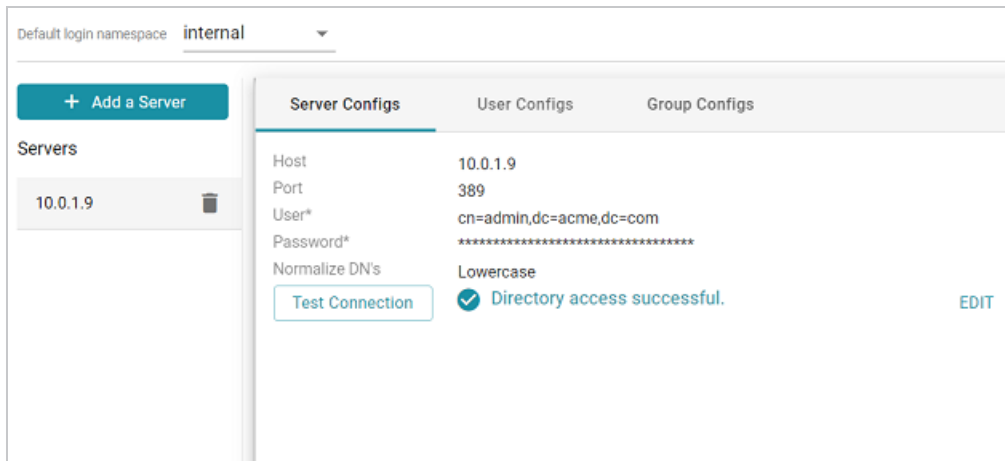
- Host \*
- Port \*
- SSL Connection  Anonymous Bind
- User DN \*
- Password \* (with an eye icon for visibility)
- Confirm Password \* (with an eye icon for visibility)
- Normalize LDAP DN's (dropdown menu)
- Test Connection (button)
- Not connected (status indicator)
- CANCEL SAVE (bottom right)

3. Enter the connection details for the server:

- **Host:** The host name or IP address for the directory server.
- **Port:** The port to use to connect to the directory server.
- **SSL Connection:** Indicates whether the directory server uses an SSL connection. Select the **SSL Connection** checkbox to enable the SSL connection. If you use SSL, make sure that you load the directory server's certificate to the Anzo trust store. See [Adding a Certificate to the Trust Store](#) for instructions.
- **Anonymous Bind:** This option indicates whether you want Anzo to connect to the directory server anonymously. To avoid Anzo login problems when enabling this option, make sure the directory server allows anonymous binding and searches when bound anonymously. Select the **Anonymous Bind** checkbox to enable anonymous binding.
- **User DN:** The full distinguished name of the account that Anzo will bind against to perform searches on the directory server.
- **Password and Confirm Password:** The password for the User DN.
- **Normalize LDAP DN's:** To ensure that duplicate user accounts are not created in Anzo if an LDAP distinguished name has both a lowercase and uppercase version, you can configure the system to normalize distinguished name strings so that values that differ only in capitalization are treated as the same value. If you do not want distinguished names to be normalized, leave the field blank or select **None**. To normalize

distinguished names to lowercase, select **Lowercase**, or select **Uppercase** if you want names to be normalized to uppercase.

4. Anzo attempts to connect to the server automatically. If the connection fails, make sure that you entered the correct connection details. You can also click **Test Connection** to check if Anzo can connect to the server.
5. Click **Save** to save the server configuration and return to the Directory screen. The new server configuration is selected on the screen. For example:



Once the connection to the server is established, create a user configuration for mapping directory users to Anzo. See [Map Users to Anzo](#) below for instructions.

## Map Users to Anzo

Follow the steps below to create a user configuration by supplying the mapping the attributes to use to sync users with Anzo.

1. On the Directory screen, click the **User Configs** tab. Then click the **Create New User Config** button. Anzo displays the Create New Config dialog box.

The screenshot shows a 'Create New Config' dialog box with the following fields and values:

- ID \*
- User Base DN \*
- Ldap Filter
- http://www.w3.org/1999/02/22-rdf-syntax-ns#type\* (value: person)
- http://openanzo.org/ontologies/2008/07/System#user\*
- http://xmlns.com/foaf/0.1/surname\*

Buttons: CANCEL, SAVE

2. Complete the following required fields and specify the optional values as desired. Each time you map an attribute, Anzo displays some samples of the values it retrieves for that attribute. If the specified attribute does not match an attribute in the system, Anzo displays an "LDAP Attribute unavailable" message.

- **ID: Required** setting that defines the unique name for this user configuration. Anzo uses this value as a namespace for usernames in case you connect to multiple directories with conflicting names.
- **User Base DN: Required** setting that specifies the LDAP distinguished name.
- **LDAP Filter:** The optional LDAP filter to apply when searching for users (usually left blank).
- **http://www.w3.org/1999/02/22-rdf-syntax-ns#type: Required** setting that specifies the LDAP class of the type of accounts that should be logged on. Typically **person**.
- **http://openanzo.org/ontologies/2008/07/System#user: Required** setting that specifies the LDAP attribute that contains user login information. Typically **uid**.
- **http://xmlns.com/foaf/0.1/surname: Required** setting that specifies the LDAP attribute that contains users' surnames. Typically **sn**.
- **http://xmlns.com/foaf/0.1/name: Required** setting that specifies the LDAP attribute that contains users' full names. Typically **cn**.

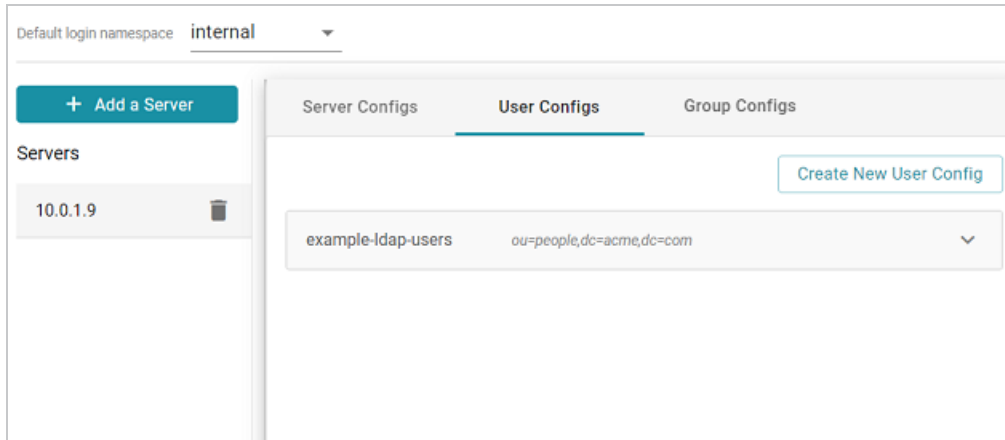
- <http://xmlns.com/foaf/0.1/givenname>: **Required** setting that specifies the LDAP attribute that contains users' first names. Typically **givenName**.
- <http://xmlns.com/foaf/0.1/title>: Optional value that specifies the LDAP attribute that contains users' job titles. Typically **title**.
- <http://www.w3.org/2003/06/sw-vocab-status/ns#term-status>: Optional value that specifies the status at the level of terms.
- [http://xmlns.com/wot/0.1/src\\_assurance](http://xmlns.com/wot/0.1/src_assurance): Optional value that specifies the source for Assured Replication.
- <http://xmlns.com/foaf/0.1/phone>: Optional value that specifies the LDAP attribute that contains user phone numbers. Typically **telephoneNumber**.
- <http://xmlns.com/foaf/0.1/mbox>: Optional value that specifies the LDAP attribute that contains users' email addresses. Typically **mail**.
- <http://openanzo.org/ontologies/2008/07/Anzo#location>: Optional value that specifies the LDAP attribute that contains user location information.
- <http://openanzo.org/ontologies/2008/07/Anzo#isInternalUser>: Optional boolean value that indicates whether users are Anzo internally managed users.
- <http://xmlns.com/foaf/0.1/img>: Optional value that specifies the LDAP attribute that contains images for users.
- <http://purl.org/dc/elements/1.1/description>: Optional value that specifies the LDAP attribute that contains user descriptions. Typically **description**.
- <http://openanzo.org/ontologies/2008/07/Anzo#defaultGroup>: Optional value that specifies the LDAP attribute that contains the value of users' Anzo Default Group assignment.
- <http://openanzo.org/ontologies/2008/07/Anzo#companyDepartment>: Optional value that specifies the LDAP attribute that contains user department information. Typically **department**.



- **http://xmlns.com/wot/0.1/assurance**: Optional boolean value that indicates whether Assured Replication is enabled.

3. When you have finished mapping attributes, click **Save** to save the user configuration.

The new user configuration is added to the system and Anzo returns to the Directory screen, which shows the newly created configuration. For example:



Once the user configuration is complete, create a group configuration for mapping directory groups to Anzo. See [Map Groups to Anzo](#) below for instructions.

## Map Groups to Anzo

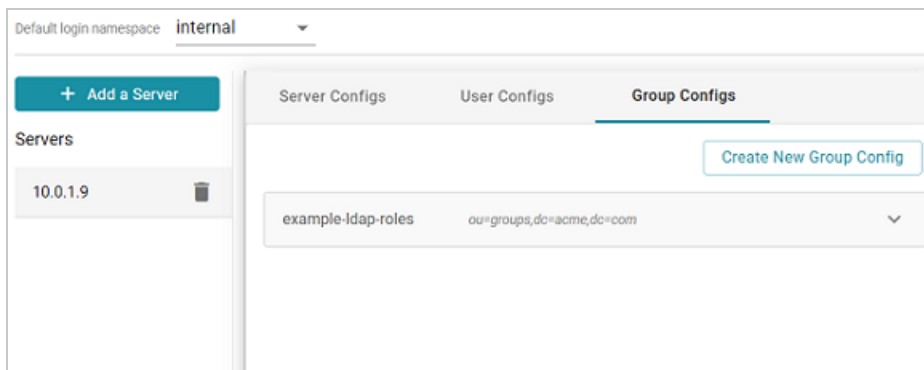
Follow the steps below to create a role configuration by supplying the mapping the attributes to use to sync groups with Anzo.

1. On the Directory screen, click the **Group Configs** tab. Then click the **Create New Group Config** button. Anzo displays the Create New Config dialog box.

2. Complete the following required fields and specify the optional values as desired. Each time you map an attribute, Anzo displays some samples of the values it retrieves for that attribute. If the specified attribute does not match an attribute in the system, Anzo displays an "LDAP Attribute unavailable" message.

- **ID: Required** setting that defines the unique name for this role configuration.
- **Base DN: Required** setting that specifies the LDAP distinguished name that contains all of the system roles.
- **LDAP Filter:** The optional LDAP filter to apply when searching for roles (usually left blank).
- **http://www.w3.org/1999/02/22-rdf-syntax-ns#type: Required** setting that specifies the group object class of the type of roles. Typically **groupOfNames**.
- **http://xmlns.com/foaf/0.1/name: Required** setting that specifies the LDAP attribute that contains the names of the roles.
- **http://xmlns.com/foaf/0.1/member: Required** setting that specifies the LDAP attribute that contains common member attributes. Typically **member** or **uniqueMember**.
- **http://openanzo.org/ontologies/2008/07/Anzo#usedBy:** Optional value that specifies how the role is used by Anzo.

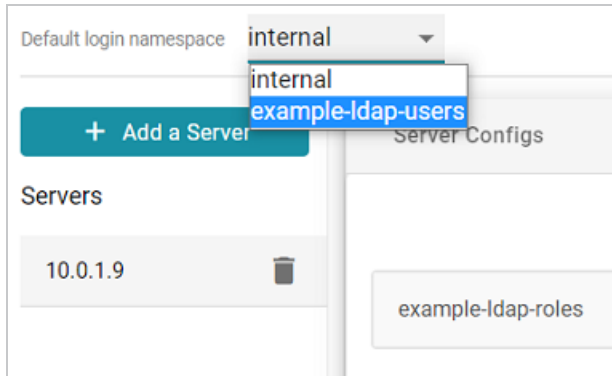
- <http://www.w3.org/2003/06/sw-vocab-status/ns#term-status>: Optional value that specifies the status at the level of terms.
  - [http://xmlns.com/wot/0.1/src\\_assurance](http://xmlns.com/wot/0.1/src_assurance): Optional value that specifies the source for Assured Replication.
  - <http://openanzo.org/ontologies/2008/07/Anzo#permission>: Optional value that specifies the LDAP attribute that contains the Anzo permissions to assign to the roles.
  - <http://purl.org/dc/elements/1.1/description>: Optional value that specifies the LDAP attribute that contains role descriptions.
  - <http://purl.org/dc/elements/1.1/date>: Optional value that specifies the LDAP attribute that contains role dates.
  - <http://xmlns.com/wot/0.1/assurance>: Optional boolean value that indicates whether Assured Replication is enabled.
3. Click **Save** to save the role configuration. The new role configuration is added to the system and Anzo returns to the Directory screen, which shows the newly created configuration. For example:



4. The last step in configuring the connection is to designate the default login namespace. This is the namespace to default to if a user does not fully qualify their username with the `@namespace` suffix when they log in to Anzo. To set the default namespace, click the **Default login namespace** drop-down list at the top of the screen and select the desired default namespace. Typically users select the newly added connection. It is displayed as the ID value from the user configuration. The "internal" namespace is the embedded Anzo LDAP

server for local users.

For example, the image below shows the available namespaces based on the example configuration in the steps above. If the `example-ldap-users` namespace is selected as the default, users who have only local Anzo accounts need to specify `username@internal` when they log in. But users with accounts in the `example-ldap-users` server can log in with `username`. The `@namespace` is not needed.



Once you have connected the directory server to Anzo and created user and group mappings, the next step is to add the directory users and groups to Anzo. See [Adding Directory Users and Groups to Anzo](#) for instructions.

### Tip

You can also set up single-sign on access to Anzo. See [Configuring Single Sign-On Authentication](#) for instructions.

# Adding Directory Users and Groups to Anzo

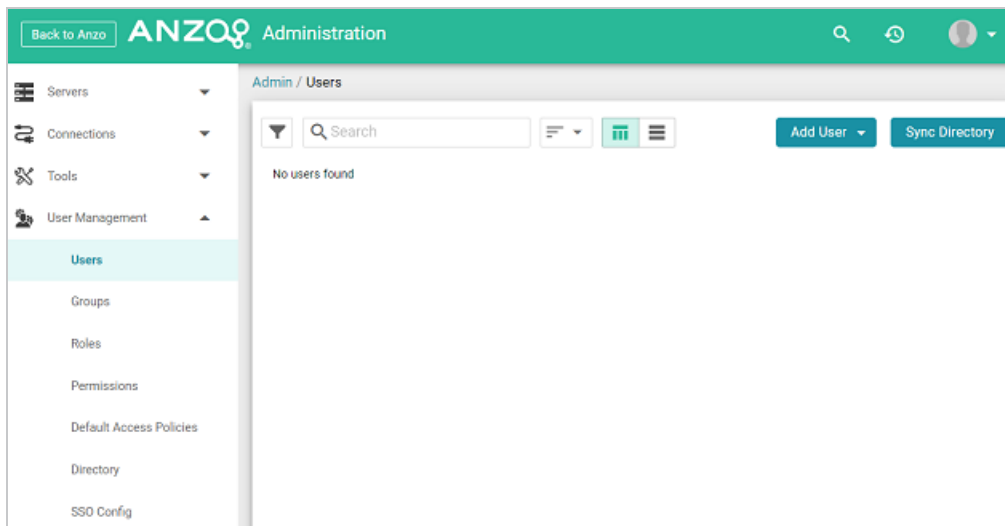
After you connect to a central directory server, you have multiple options for how LDAP users gain access to Anzo. Some organizations retrieve the LDAP users and groups from the server and add them to Anzo. An Anzo administrator then manages role and license assignment in Anzo. Other organizations pre-define LDAP-to-Anzo role configurations and mappings so that users are automatically assigned an Anzo license and can log in to Anzo as soon as the LDAP administrator adds them to the appropriate LDAP role. With this option, no action needs to be taken in Anzo once the directory server is connected and user and role mappings are configured.

This topic provides instructions for adding directory users and groups to Anzo. For instructions on setting up self-authorization for directory users so that they can log into Anzo and automatically become licensed after being added to the appropriate LDAP group, see [Enabling Self-Authorization for Directory Users](#).

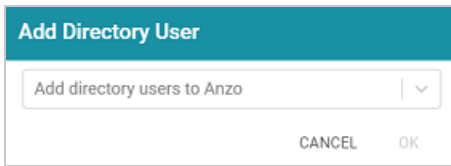
- [Add Directory Users to Anzo](#)
- [Add Directory Groups to Anzo](#)

## Add Directory Users to Anzo

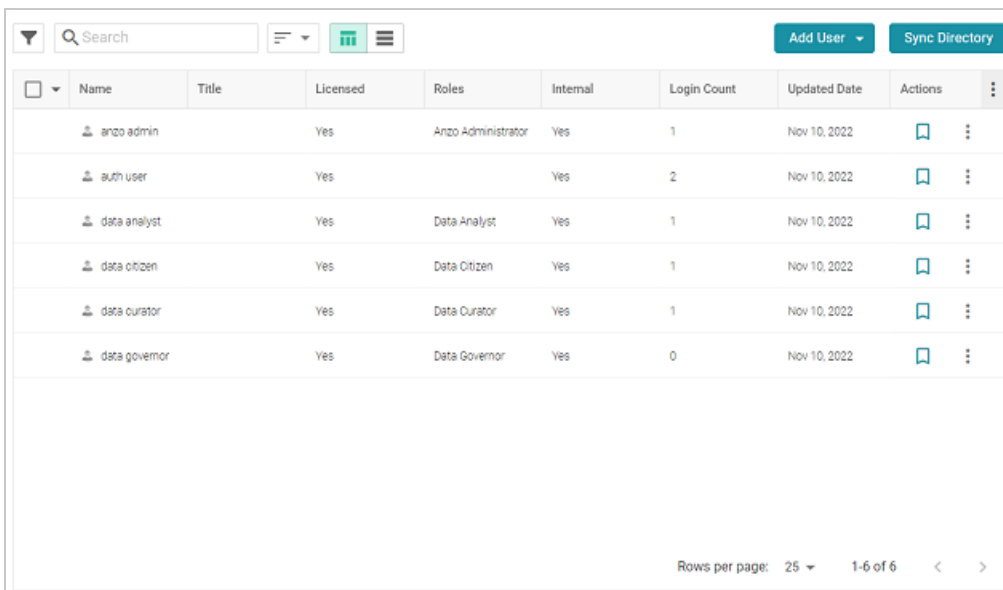
1. To add directory users to Anzo, select **Users** from the **User Management** menu in the Administration application. The Users screen is displayed. For example:



2. Click the **Add User** button and select **Add Directory Users**. The Add Directory User dialog box is displayed:



3. Click the **Add directory users to Anzo** drop-down list, and select each user to add to Anzo. Repeat this step for all of the users that you want to add.
4. When you have finished adding users, click **OK** to return to the Users screen. For example:



<input type="checkbox"/>	Name	Title	Licensed	Roles	Internal	Login Count	Updated Date	Actions
<input type="checkbox"/>	anzo admin		Yes	Anzo Administrator	Yes	1	Nov 10, 2022	
<input type="checkbox"/>	auth user		Yes		Yes	2	Nov 10, 2022	
<input type="checkbox"/>	data analyst		Yes	Data Analyst	Yes	1	Nov 10, 2022	
<input type="checkbox"/>	data citizen		Yes	Data Citizen	Yes	1	Nov 10, 2022	
<input type="checkbox"/>	data curator		Yes	Data Curator	Yes	1	Nov 10, 2022	
<input type="checkbox"/>	data governor		Yes	Data Governor	Yes	0	Nov 10, 2022	

### Note

In order for the new users to be able to log in to Anzo, they must be **Licensed** users. Complete the next step to designate licensed users.

5. The last step in the process is to configure the **Licensed** users. If you want a user to be able to log in to Anzo, they must be specified as a licensed user. To designate a user as licensed, open the Edit User dialog box by clicking a user's name in the Users list. In the dialog box, select the **Licensed** checkbox and click **Save**. For example:

**Edit User**

Username  
aaron.heart

First Name  
Aaron

Last Name  
Heart

Licensed

Position / Title  
Supreme Accounting Mascot

Email  
aaron.heart@acme.com

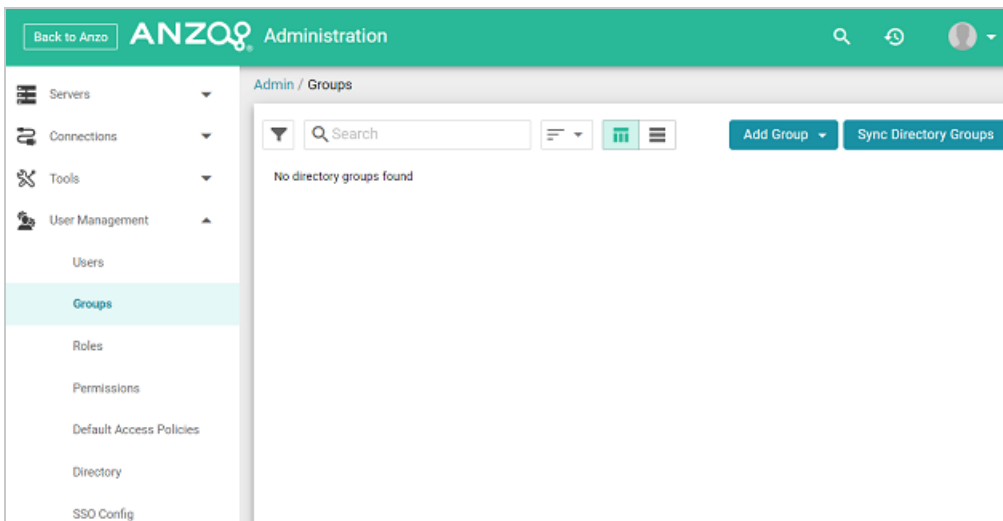
CANCEL SAVE

Repeat this step for all of the users who should be licensed.

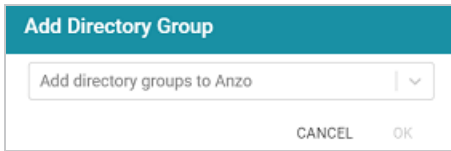
For instructions on adding groups to Anzo, proceed to [Add Directory Groups to Anzo](#) below.

## Add Directory Groups to Anzo

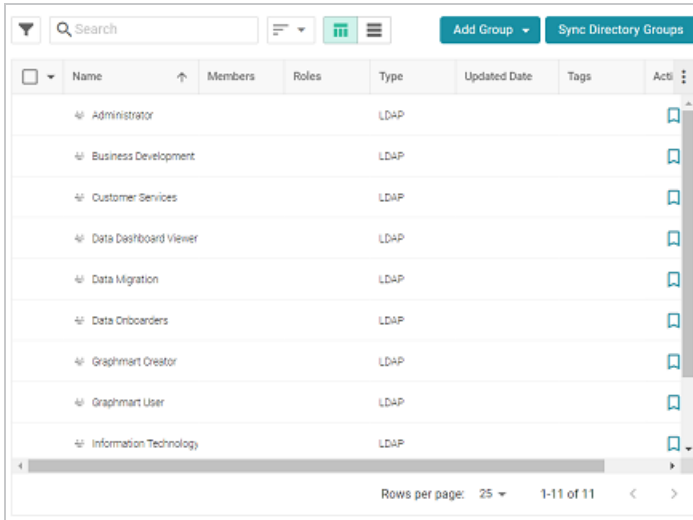
1. To add directory groups to Anzo, select **Groups** from the **User Management** menu in the Administration application. The Groups screen is displayed. For example:



2. Click the **Add Group** button and select **LDAP Directory Group**. The Add Directory Group dialog box is displayed:



3. Click the **Add directory groups to Anzo** drop-down list, and select each group to add to Anzo. Repeat this step for all of the groups that you want to add.
4. When you have finished adding groups, click **OK** to return to the Groups screen. For example:



Now that the users and groups from the directory server are available in Anzo, the next step is to associate the groups with Anzo roles. Roles are used to grant access to the Anzo applications and the functionality in those applications. See [Creating and Managing Roles](#) for instructions.



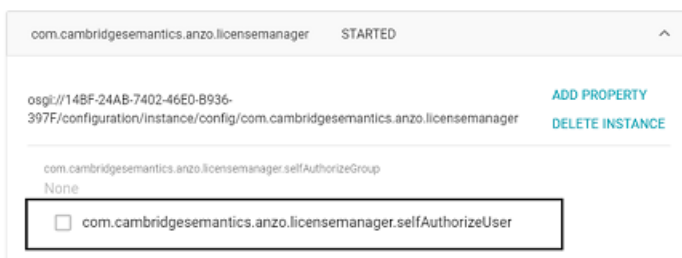
# Enabling Self-Authorization for Directory Users

In order to log in to Anzo, a user must be a **Licensed** user. If you defined LDAP-to-Anzo role configurations and mappings so that you can manage all permissions in the directory server without retrieving the user and group accounts and adding them to Anzo, you can configure Anzo to automatically license those users as they log in. Follow the instructions below to enable self-authorization.

## Note

When deciding whether to enable self-authorization at all or whether to limit it to certain LDAP groups, consider the number of users who will access Anzo and the number of users allowed by your Anzo license. Your Cambridge Semantics Customer Success manager can help determine whether to enable the feature if you have questions.

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Anzo License and Entitlement Manager** bundle and view its details.
3. Click the **Services** tab and expand **com.cambridgesemantics.anzo.licensemanager**.
4. Find the **com.cambridgesemantics.anzo.licensemanager.selfAuthorizeUser** property (shown below).



5. Click the property to make it editable, and then select the checkbox to enable it.



6. Click the checkmark icon (✓) to save the change.
7. If you want to limit the ability to self-authorize to a certain LDAP group, click the **com.cambridgesemantics.anzo.licensemanager.selfAuthorizeGroup** property to make it editable, and then specify the group name to include.
8. Click the checkmark icon (✓) to save the change.

Changes to the Anzo License and Entitlement Manager service take effect immediately. You do not need to restart Anzo or the service to apply the change.

# Configuring Single Sign-On Authentication

The topics in this section provide instructions for configuring single sign-on (SSO) access for each of the supported SSO providers.

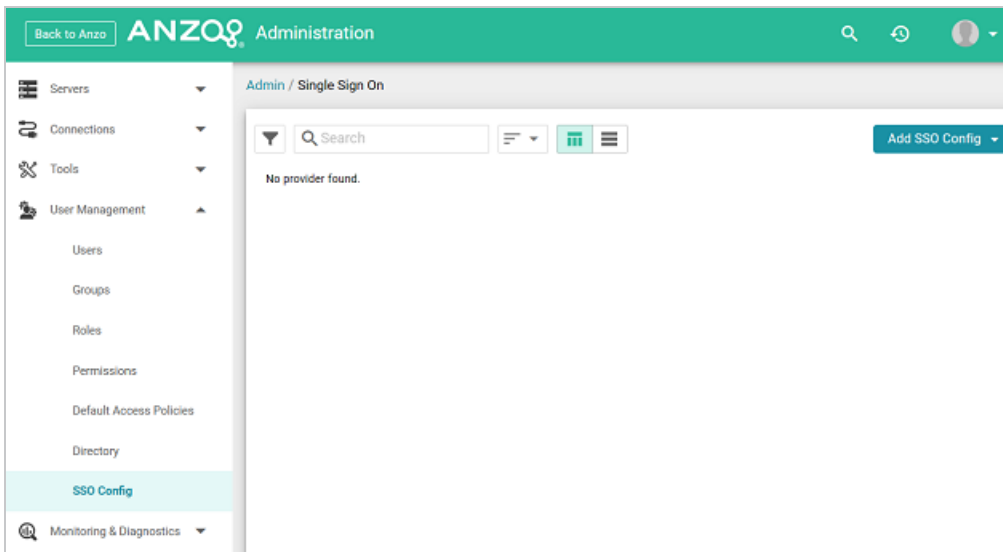
## In this section:

<a href="#">Adding a Basic Provider</a>	180
<a href="#">Adding a JWT Provider</a>	186
<a href="#">Adding a Kerberos Provider</a>	192
<a href="#">Adding an Oauth 2 Provider</a>	199
<a href="#">Adding an Open ID Connect Provider</a>	204
<a href="#">Adding a SAML Provider</a>	211

## Adding a Basic Provider

Follow the steps below to add a Direct or Indirect Basic SSO Provider.

1. In the Administration application, expand **User Management** and click **SSO Config**. Anzo displays the Single Sign On screen, which lists any existing providers. For example:



2. Click the **Add SSO Config** button and select **Basic Provider**. Then choose **Direct Basic Provider** or **Indirect Basic Provider**, depending on the type of authentication that is used. The Create screen for that type of provider is displayed. For example:

**Create Direct Basic Provider**

Title \*

Description

Enable on matched container ID \*

This provider will be active if the request container ID matches one of the supplied container IDs.

Realm Name

**authentication required**

The text that is displayed in the dialog box that appears when the browser prompts the user for login data.

Enable on match regex ADD

This provider will be active if the request uri matches the supplied regex. It will be active by default if no value is supplied.

Disable on match regex ADD

This provider will be inactive if the request uri matches the supplied regex. It will be active by default if no value is supplied.

CANCEL SAVE

- Configure the required properties and any optional settings as needed. The lists below describe the properties for [Direct](#) and [Indirect](#) providers.

## Direct

- **Title:** This property sets the name for the connection that you are creating.
- **Description:** This property can be used to provide a brief description of the provider configuration.
- **Enable on matched container ID:** This property sets the list of container IDs to match. This provider will be active if the request container ID matches one of the listed container IDs. Click the field and select a container ID from the drop-down list. To specify multiple IDs, click the field again and select another value. To remove a container from the list, click the X on the right of the container name.
- **Realm Name:** This property can be used to define the name of the security realm. The text appears in the dialog box that is displayed when the browser prompts a user for their credentials.

- **Enable on match regex:** This property can be used to define regular expression rules for matching request URLs to enable. To add a rule, type an expression in the field and click **Add**. This provider will be active if the request URL matches any of the supplied expressions. If this field is blank, the provider will be active by default.
- **Disable on match regex:** This property can be used to define regular expression rules for matching request URLs to disable. To add a rule, type an expression in the field and click **Add**. This provider will be inactive if the request URL matches any of the supplied expressions. If this field is blank, the provider will be active by default.
- **User Identifier:** This property specifies the SSO provider attribute, such as `email` or `username`, to use for looking up users in the directory server.
- **Email Template regex:** If an email attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between email addresses stored by the SSO provider and email addresses returned by the directory server.
- **Email Template Replacement:** This property can be used to define a replacement email template to use if there are variations found by `Email Template regex`.
- **User Template regex:** If a username attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between user names stored by the SSO provider and names returned by the directory server.
- **User Template Replacement:** This property can be used to define a replacement user template to use if there are variations found by `User Template regex`.
- **Use username directly:** This property controls whether the identity provider directly authenticates a user by validating a username and password or by validating an assertion about the user's identity as defined by a separate identity provider.
- **Skip CSRF check:** This property controls whether to perform or skip a cross-site request forgery (CSRF) check.

- **LDAP domain:** This property identifies the LDAP domain to use for user lookup.
- **LDAP email property:** This property defines the LDAP email property to use to find the associated user's dn. For example,  
<http://openanzo.org/ontologies/2008/07/Anzo#ldapEmailInfo>.
- **Principal Template:** This property can be used to define the template to use for populating roles and returning user URIs.

## Indirect

- **Title:** This property sets the name for the connection that you are creating.
- **Description:** This property can be used to provide a brief description of the provider configuration.
- **Enable on matched container ID:** This property sets the list of container IDs to match. This provider will be active if the request container ID matches one of the listed container IDs. Click the field and select a container ID from the drop-down list. To specify multiple IDs, click the field again and select another value. To remove a container from the list, click the X on the right of the container name.
- **Realm Name:** This property can be used to define the name of the security realm. The text appears in the dialog box that is displayed when the browser prompts a user for their credentials.
- **Enable on login page:** This property controls whether to display a link for this provider on the Anzo login screen.
- **Callback URL:** This property specifies the URL that the provider should use to redirect users back to the Anzo application after a successful login. Include the full URL to the Anzo instance, through the proxy if one exists. Specify the URL in quotes and append the value with `/anzo_authenticate`, i.e., `"hostname:port/anzo_authenticate"`.
- **Callback URL port replacement:** This property can be used to define the port to use if the one specified in the `Callback URL` field is unavailable.

- **User Identifier:** This property specifies the SSO provider attribute, such as `email` or `username`, to use for looking up users in the directory server.
- **Default to IDP Logout:** This property controls whether to log a user out of the IDP by default when they log out of Anzo.
- **Email Template regex:** If an email attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between email addresses stored by the SSO provider and email addresses returned by the directory server.
- **Email Template Replacement:** This property can be used to define a replacement email template to use if there are variations found by `Email Template regex`.
- **User Template regex:** If a username attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between user names stored by the SSO provider and names returned by the directory server.
- **User Template Replacement:** This property can be used to define a replacement user template to use if there are variations found by `User Template regex`.
- **Use username directly:** This property controls whether the identity provider directly authenticates a user by validating a username and password or by validating an assertion about the user's identity as defined by a separate identity provider.
- **Skip CSRF check:** This property controls whether to perform or skip a cross-site request forgery (CSRF) check.
- **LDAP domain:** This property identifies the LDAP domain to use for user lookup.
- **LDAP email property:** This property defines the LDAP email property to use to find the associated user's dn. For example,  
`http://openanzo.org/ontologies/2008/07/Anzo#ldapEmailInfo`.
- **Principal Template:** This property can be used to define the template to use for populating roles and returning user URIs.

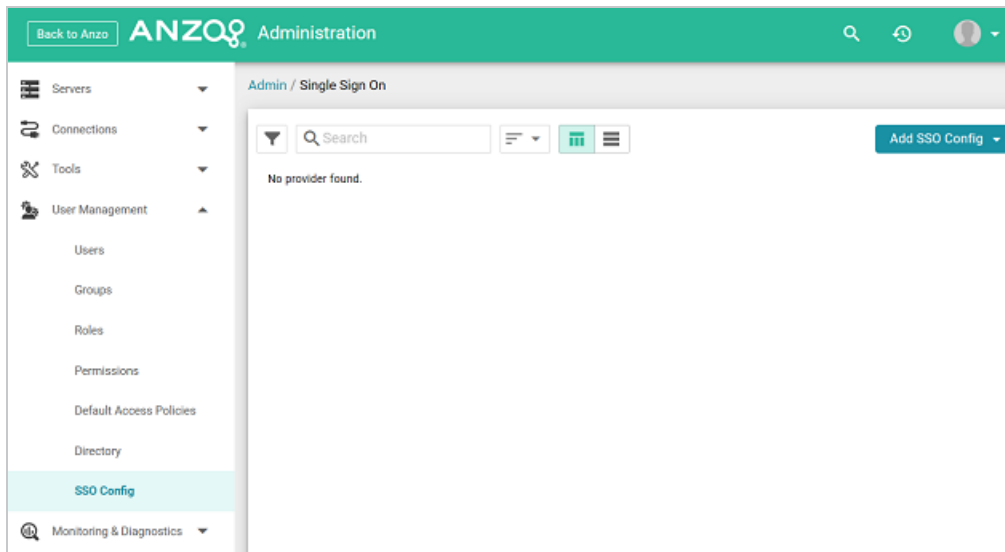


- **Icon:** This property can be used to include an SSO icon on the Anzo login screen. To select an image, click the **Icon** field and select **Add File**.
  - **With State:** This property controls whether information about the application's state is included in authentication requests.
4. When you have finished configuring properties, click **Save** to save the provider setup.

## Adding a JWT Provider

Follow the steps below to add a JWT Header or Parameter SSO Provider.

1. In the Administration application, expand **User Management** and click **SSO Config**. Anzo displays the Single Sign On screen, which lists any existing providers. For example:



2. Click the **Add SSO Config** button and select **JWT Provider**. Then choose **JWT Header Provider** or **JWT Parameter Provider**, depending on the type of authentication that is used. The Create screen for that type of provider is displayed. For example:

**Create JWT Header Provider**

Title \*

Description

Enable on matched container ID \*

This provider will be active if the request container ID matches one of the supplied container IDs.

Header Prefix

Header Name

Signing Secret \*

Private key, private key passcode and/or shared secret depending on algorithm

CANCEL SAVE

3. Configure the required properties and any optional settings as needed. The lists below describe the properties for JWT [Header](#) and [Parameter](#) providers.

## Header

- **Title:** This property sets the name for the connection that you are creating.
- **Description:** This property can be used to provide a brief description of the provider configuration.
- **Enable on matched container ID:** This property sets the list of container IDs to match. This provider will be active if the request container ID matches one of the listed container IDs. Click the field and select a container ID from the drop-down list. To specify multiple IDs, click the field again and select another value. To remove a container from the list, click the X on the right of the container name.
- **Header Prefix:** This property can be used to specify the header prefix if one is used.
- **Header Name:** This property can be used to specify the header name.
- **Signing Secret:** This property specifies the secret with which the token is signed.

- **Key Algorithm:** This property can be used to specify the signing algorithm that is used for the key.
- **Encryption Algorithm:** This property can be used to specify the encryption algorithm that is used to sign or encrypt the tokens.
- **Encryption Method:** This property can be used to specify encryption method that is used for encrypted tokens.
- **Encryption Secret:** This property can be used to specify the secret that is used for encrypted tokens.
- **Enable on match regex:** This property can be used to define regular expression rules for matching request URLs to enable. To add a rule, type an expression in the field and click **Add**. This provider will be active if the request URL matches any of the supplied expressions. If this field is blank, the provider will be active by default.
- **Disable on match regex:** This property can be used to define regular expression rules for matching request URLs to disable. To add a rule, type an expression in the field and click **Add**. This provider will be inactive if the request URL matches any of the supplied expressions. If this field is blank, the provider will be active by default.
- **User Identifier:** This property specifies the SSO provider attribute, such as `email` or `username`, to use for looking up users in the directory server.
- **Email Template regex:** If an email attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between email addresses stored by the SSO provider and email addresses returned by the directory server.
- **Email Template Replacement:** This property can be used to define a replacement email template to use if there are variations found by `Email Template regex`.
- **User Template regex:** If a username attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations

between user names stored by the SSO provider and names returned by the directory server.

- **User Template Replacement:** This property can be used to define a replacement user template to use if there are variations found by `User Template regex`.
- **Use username directly:** This property controls whether the identity provider directly authenticates a user by validating a username and password or by validating an assertion about the user's identity as defined by a separate identity provider.
- **Skip CSRF check:** This property controls whether to perform or skip a cross-site request forgery (CSRF) check.
- **LDAP domain:** This property identifies the LDAP domain to use for user lookup.
- **LDAP email property:** This property defines the LDAP email property to use to find the associated user's dn. For example,  
`http://openanzo.org/ontologies/2008/07/Anzo#ldapEmailInfo`.
- **Principal Template:** This property can be used to define the template to use for populating roles and returning user URIs.

## Parameter

- **Title:** This property sets the name for the connection that you are creating.
- **Description:** This property can be used to provide a brief description of the provider configuration.
- **Enable on matched container ID:** This property sets the list of container IDs to match. This provider will be active if the request container ID matches one of the listed container IDs. Click the field and select a container ID from the drop-down list. To specify multiple IDs, click the field again and select another value. To remove a container from the list, click the X on the right of the container name.
- **Parameter Name:** This property specifies the header parameter name.

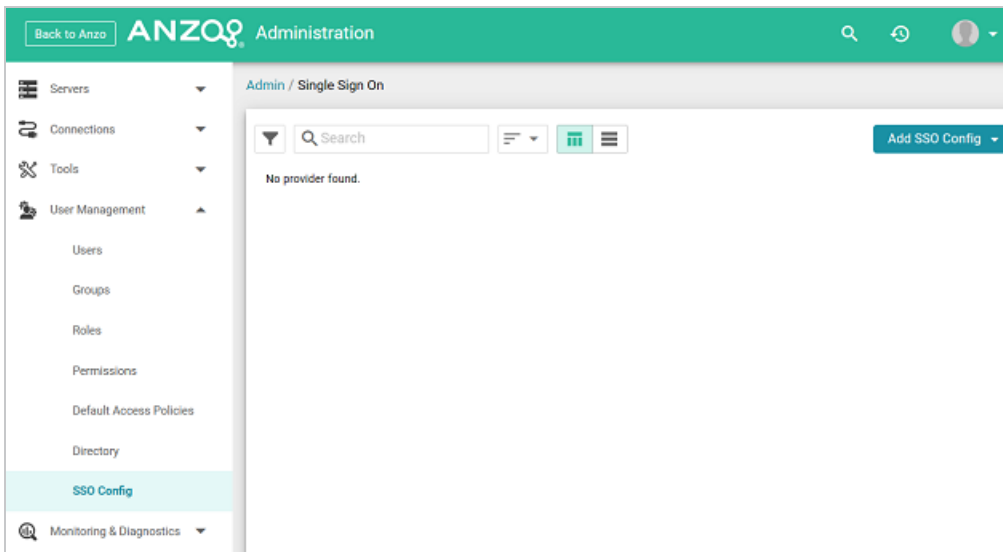
- **Supports GET request:** This property controls whether GET requests are supported using the token.
- **Supports POST request:** This property controls whether POST requests are supported using the token.
- **Signing Secret:** This property specifies the secret with which the token is signed.
- **Key Algorithm:** This property can be used to specify the signing algorithm that is used for the key.
- **Encryption Algorithm:** This property can be used to specify the encryption algorithm that is used to sign or encrypt the tokens.
- **Encryption Method:** This property can be used to specify encryption method that is used for encrypted tokens.
- **Encryption Secret:** This property can be used to specify the secret that is used for encrypted tokens.
- **Enable on match regex:** This property can be used to define regular expression rules for matching request URLs to enable. To add a rule, type an expression in the field and click **Add**. This provider will be active if the request URL matches any of the supplied expressions. If this field is blank, the provider will be active by default.
- **Disable on match regex:** This property can be used to define regular expression rules for matching request URLs to disable. To add a rule, type an expression in the field and click **Add**. This provider will be inactive if the request URL matches any of the supplied expressions. If this field is blank, the provider will be active by default.
- **User Identifier:** This property specifies the SSO provider attribute, such as `email` or `username`, to use for looking up users in the directory server.
- **Email Template regex:** If an email attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between email addresses stored by the SSO provider and email addresses returned by the directory server.

- **Email Template Replacement:** This property can be used to define a replacement email template to use if there are variations found by `Email Template regex`.
  - **User Template regex:** If a username attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between user names stored by the SSO provider and names returned by the directory server.
  - **User Template Replacement:** This property can be used to define a replacement user template to use if there are variations found by `User Template regex`.
  - **Use username directly:** This property controls whether the identity provider directly authenticates a user by validating a username and password or by validating an assertion about the user's identity as defined by a separate identity provider.
  - **Skip CSRF check:** This property controls whether to perform or skip a cross-site request forgery (CSRF) check.
  - **LDAP domain:** This property identifies the LDAP domain to use for user lookup.
  - **LDAP email property:** This property defines the LDAP email property to use to find the associated user's dn. For example,  
`http://openanzo.org/ontologies/2008/07/Anzo#ldapEmailInfo`.
  - **Principal Template:** This property can be used to define the template to use for populating roles and returning user URIs.
4. When you have finished configuring properties, click **Save** to save the provider setup.

## Adding a Kerberos Provider

Follow the steps below to add a Direct or Indirect Kerberos SSO Provider.

1. In the Administration application, expand **User Management** and click **SSO Config**. Anzo displays the Single Sign On screen, which lists any existing providers. For example:



2. Click the **Add SSO Config** button and select **Kerberos Provider**. Then choose **Direct Kerberos Provider** or **Indirect Kerberos Provider**, depending on the type of authentication that is used. The Create screen for that type of provider is displayed. For example:



3. Configure the required properties and any optional settings as needed. The lists below describe the properties for [Direct](#) and [Indirect](#) providers.

### Direct

- **Title:** This property sets the name for the connection that you are creating.
- **Description:** This property can be used to provide a brief description of the provider configuration.
- **Enable on matched container ID:** This property sets the list of container IDs to match. This provider will be active if the request container ID matches one of the listed container IDs. Click the field and select a container ID from the drop-down list. To specify multiple IDs, click the field again and select another value. To remove a container from the list, click the X on the right of the container name.
- **Service Principal:** This property lists the service and DNS name for the Kerberos application. For authentication through the web browser, specify the service principal value in the following format: `HTTP/FQDN@domain`. For example, `HTTP/server.example.com@example.com`.

## Note

The keytab file must contain the key for this service principal.

- **Keytab:** This property specifies the `.keytab` file that lists the Kerberos principals and encrypted keys. Click the **Keytab** field to open the File Location dialog box and select the keytab file.
- **Realm:** This property can be used to specify the Kerberos realm that the service principal maps to.
- **KRB Configuration:** This property can be used to specify the path and file name for the `krb5.conf` file on the Kerberos instance. When not specified, the default location is `/etc/krb5.conf`.
- **KDC:** This field can be used to specify the domain name for the Key Distribution Center.
- **Debug mode:** This property controls whether Kerberos debug logging is enabled.
- **Enable on match regex:** This property can be used to define regular expression rules for matching request URLs to enable. To add a rule, type an expression in the field and click **Add**. This provider will be active if the request URL matches any of the supplied expressions. If this field is blank, the provider will be active by default.
- **Disable on match regex:** This property can be used to define regular expression rules for matching request URLs to disable. To add a rule, type an expression in the field and click **Add**. This provider will be inactive if the request URL matches any of the supplied expressions. If this field is blank, the provider will be active by default.
- **User Identifier:** This property specifies the SSO provider attribute, such as `email` or `username`, to use for looking up users in the directory server.
- **Email Template regex:** If an email attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between email addresses stored by the SSO provider and email addresses returned by the directory server.

- **Email Template Replacement:** This property can be used to define a replacement email template to use if there are variations found by `Email Template regex`.
- **User Template regex:** If a username attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between user names stored by the SSO provider and names returned by the directory server.
- **User Template Replacement:** This property can be used to define a replacement user template to use if there are variations found by `User Template regex`.
- **Use username directly:** This property controls whether the identity provider directly authenticates a user by validating a username and password or by validating an assertion about the user's identity as defined by a separate identity provider.
- **Skip CSRF check:** This property controls whether to perform or skip a cross-site request forgery (CSRF) check.
- **LDAP domain:** This property identifies the LDAP domain to use for user lookup.
- **LDAP email property:** This property defines the LDAP email property to use to find the associated user's dn. For example,  
`http://openanzo.org/ontologies/2008/07/Anzo#ldapEmailInfo`.
- **Principal Template:** This property can be used to define the template to use for populating roles and returning user URIs.

## Indirect

- **Title:** This property sets the name for the connection that you are creating.
- **Description:** This property can be used to provide a brief description of the provider configuration.
- **Enable on matched container ID:** This property sets the list of container IDs to match. This provider will be active if the request container ID matches one of the listed container IDs. Click the field and select a container ID from the drop-down list. To

specify multiple IDs, click the field again and select another value. To remove a container from the list, click the X on the right of the container name.

- **Service Principal:** This property lists the service and DNS name for the Kerberos application. For authentication through the web browser, specify the service principal value in the following format: `HTTP/FQDN@domain`. For example, `HTTP/server.example.com@example.com`.

#### Note

The keytab file must contain the key for this service principal.

- **Keytab:** This property specifies the `.keytab` file that lists the Kerberos principals and encrypted keys. Click the **Keytab** field to open the File Location dialog box and select the keytab file.
- **Realm:** This property can be used to specify the Kerberos realm that the service principal maps to.
- **KRB Configuration:** This property can be used to specify the path and file name for the `krb5.conf` file on the Kerberos instance. When not specified, the default location is `/etc/krb5.conf`.
- **KDC:** This field can be used to specify the domain name for the Key Distribution Center.
- **Debug mode:** This property controls whether Kerberos debug logging is enabled.
- **Enable on login page:** This property controls whether to display a link for this provider on the Anzo login screen.
- **Callback URL:** This property specifies the URL that the provider should use to redirect users back to the Anzo application after a successful login. Include the full URL to the Anzo instance, through the proxy if one exists. Specify the URL in quotes and append the value with `/anzo_authenticate`, i.e., `"hostname:port/anzo_authenticate"`.

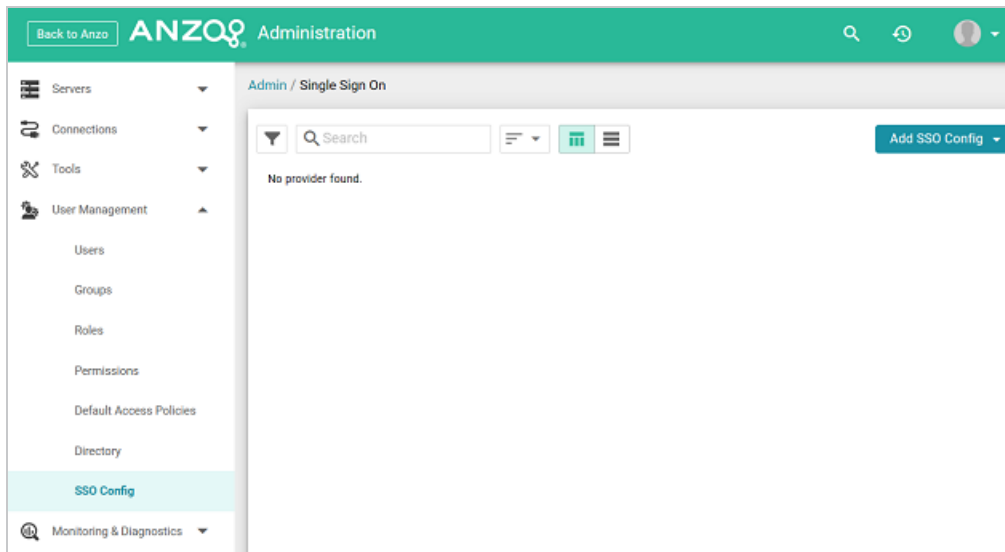
- **Callback URL port replacement:** This property can be used to define the port to use if the one specified in the `Callback URL` field is unavailable.
- **User Identifier:** This property specifies the SSO provider attribute, such as `email` or `username`, to use for looking up users in the directory server.
- **IDP Logout Capable:** This property can be used to indicate whether the SSO provider supports logging the user out of the IDP when they log out of Anzo.
- **Default to IDP Logout:** This property controls whether to log a user out of the IDP by default when they log out of Anzo.
- **Logout URL Suffix:** When `Default to IDP Logout` is enabled, this property can be used to specify the logout URL for the SSO provider. The `[urlAfterLogout]` placeholder is replaced with the SSO provider server URL.
- **Email Template regex:** If an email attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between email addresses stored by the SSO provider and email addresses returned by the directory server.
- **Email Template Replacement:** This property can be used to define a replacement email template to use if there are variations found by `Email Template regex`.
- **User Template regex:** If a username attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between user names stored by the SSO provider and names returned by the directory server.
- **User Template Replacement:** This property can be used to define a replacement user template to use if there are variations found by `User Template regex`.
- **Use username directly:** This property controls whether the identity provider directly authenticates a user by validating a username and password or by validating an assertion about the user's identity as defined by a separate identity provider.

- **Skip CSRF check:** This property controls whether to perform or skip a cross-site request forgery (CSRF) check.
  - **LDAP domain:** This property identifies the LDAP domain to use for user lookup.
  - **LDAP email property:** This property defines the LDAP email property to use to find the associated user's dn. For example,  
`http://openanzo.org/ontologies/2008/07/Anzo#ldapEmailInfo.`
  - **Principal Template:** This property can be used to define the template to use for populating roles and returning user URIs.
  - **Icon:** This property can be used to include an SSO icon on the Anzo login screen. To select an image, click the **Icon** field and select **Add File**.
  - **With State:** This property controls whether information about the application's state is included in authentication requests.
4. When you have finished configuring properties, click **Save** to save the provider setup.

## Adding an Oauth 2 Provider

Follow the steps below to add a generic, Facebook, or Windows Live Oauth 2 SSO Provider.

1. In the Administration application, expand **User Management** and click **SSO Config**. Anzo displays the Single Sign On screen, which lists any existing providers. For example:



2. Click the **Add SSO Config** button and select **Oauth 2 Provider**. Then choose **Facebook Provider** or **Generic Oauth 2 Provider**, or **Windows Live Provider**, depending on the type of API that is used. The Create screen for that type of provider is displayed. For example:

**Create Generic OAuth 2 Provider**

Title \*

Description

Enable on matched container ID \*  
This provider will be active if the request container ID matches one of the supplied container IDs.

Client ID \*  
Client Identifier

Secret \*

Confirm Secret \*   
OAuth secret

Authorization URL  
Authorization URL

Token Endpoint URL  
Token Endpoint URL

Profile URL

CANCEL SAVE

3. Configure the required properties and any optional settings as needed. The list below describes the properties for Facebook, Generic, and Windows Live provider configurations.
- **Title:** This property sets the name for the connection that you are creating.
  - **Description:** This property can be used to provide a brief description of the provider configuration.
  - **Enable on matched container ID:** This property sets the list of container IDs to match. This provider will be active if the request container ID matches one of the listed container IDs. Click the field and select a container ID from the drop-down list. To specify multiple IDs, click the field again and select another value. To remove a container from the list, click the X on the right of the container name.
  - **Client ID:** This property specifies the unique App ID for the client application.
  - **Secret: Password and Confirm Password:** These properties list the secret for the specified `Client ID`.



- **Authorization URL:** This property can be used to specify the URL for the authentication endpoint (`/authorize`).
- **Token Endpoint URL:** This property can be used to specify the URL for the token endpoint (`/oauth/token`).
- **Profile URL:** This property can be used to specify the URL from which to retrieve user profiles.
- **Logout URL:** This property can be used to specify the logout URL for the SSO provider if you plan to configure the connection to log users out of the IDP when they log out of Anzo.
- **With State:** This property controls whether information about the application's state is included in authentication requests.
- **Profile Attributes:** This property can be used to specify the user profile attributes to return.
- **Scope:** This property can be used to specify the parameters to send to the authorization endpoint with the request.
- **Enable on login page:** This property controls whether to display a link for this provider on the Anzo login screen.
- **Callback URL:** This property specifies the URL that the provider should use to redirect users back to the Anzo application after a successful login. Include the full URL to the Anzo instance, through the proxy if one exists. Specify the URL in quotes and append the value with `/anzo_authenticate`, i.e., `"hostname:port/anzo_authenticate"`.
- **Callback URL port replacement:** This property can be used to define the port to use if the one specified in the `Callback URL` field is unavailable.
- **User Identifier:** This property specifies the SSO provider attribute, such as `email` or `username`, to use for looking up users in the directory server.

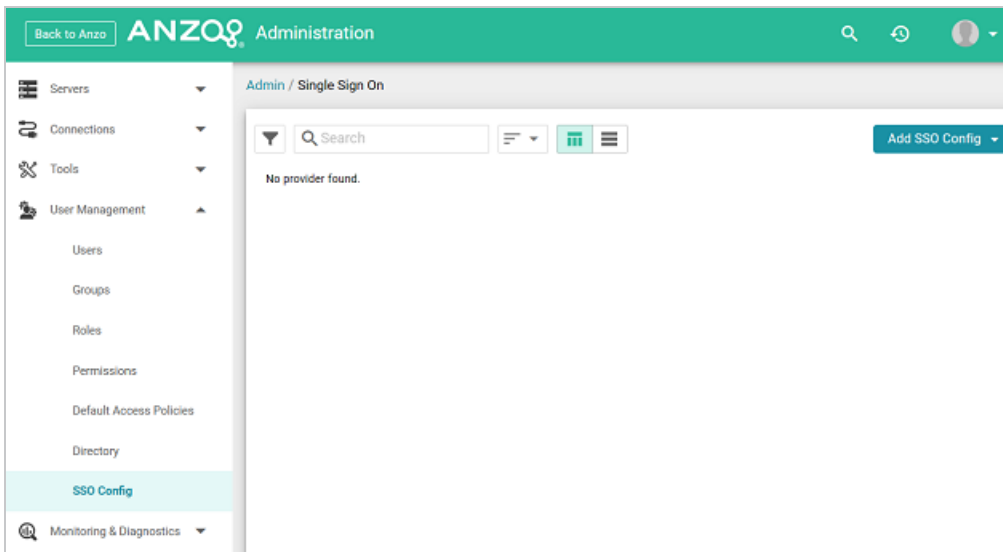
- **IDP Logout Capable:** This property can be used to indicate whether the SSO provider supports logging the user out of the IDP when they log out of Anzo.
- **Default to IDP Logout:** This property controls whether to log a user out of the IDP by default when they log out of Anzo.
- **Logout URL Suffix:** When `Default to IDP Logout` is enabled, this property can be used to specify the logout URL for the SSO provider. The `[urlAfterLogout]` placeholder is replaced with the SSO provider server URL.
- **Email Template regex:** If an email attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between email addresses stored by the SSO provider and email addresses returned by the directory server.
- **Email Template Replacement:** This property can be used to define a replacement email template to use if there are variations found by `Email Template regex`.
- **User Template regex:** If a username attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between user names stored by the SSO provider and names returned by the directory server.
- **User Template Replacement:** This property can be used to define a replacement user template to use if there are variations found by `User Template regex`.
- **Use username directly:** This property controls whether the identity provider directly authenticates a user by validating a username and password or by validating an assertion about the user's identity as defined by a separate identity provider.
- **Skip CSRF check:** This property controls whether to perform or skip a cross-site request forgery (CSRF) check.
- **LDAP domain:** This property identifies the LDAP domain to use for user lookup.

- **LDAP email property:** This property defines the LDAP email property to use to find the associated user's dn. For example,  
`http://openanzo.org/ontologies/2008/07/Anzo#ldapEmailInfo.`
  - **Principal Template:** This property can be used to define the template to use for populating roles and returning user URIs.
  - **Icon:** This property can be used to include an SSO icon on the Anzo login screen. To select an image, click the **Icon** field and select **Add File**.
4. When you have finished configuring properties, click **Save** to save the provider setup.

## Adding an Open ID Connect Provider

Follow the steps below to add a generic or Google OIDC SSO Provider.

1. In the Administration application, expand **User Management** and click **SSO Config**. Anzo displays the Single Sign On screen, which lists any existing providers. For example:



2. Click the **Add SSO Config** button and select **Open ID Connect Provider**. Then choose **Generic OIDC Provider**, or **Google OIDC Provider**, or **Keycloak OIDC Provider**, depending on the type of API that is used. The Create screen for that type of provider is displayed. For example:

**Create Open ID Connect Provider**

Title \*

Description

Enable on matched container ID \*

This provider will be active if the request container ID matches one of the supplied container IDs.

Client ID \*

Client Identifier

**Secret \***

**Confirm Password \***

OAuth secret

Discovery URI \*

discovery URI for fetching OP metadata (http://openid.net/specs/openid-connect-discovery-1\_0.html)

Scope

openid profile email

CANCEL SAVE

- Configure the required properties and any optional settings as needed. The list below describes the properties for [Generic](#) and [Google](#) OIDC provider configurations.

## Generic

- **Title:** This property sets the name for the connection that you are creating.
- **Description:** This property can be used to provide a brief description of the provider configuration.
- **Enable on matched container ID:** This property sets the list of container IDs to match. This provider will be active if the request container ID matches one of the listed container IDs. Click the field and select a container ID from the drop-down list. To specify multiple IDs, click the field again and select another value. To remove a container from the list, click the X on the right of the container name.
- **Client ID:** This property specifies the client ID or consumer key value from the provider application.
- **Secret: Password and Confirm Password:** These properties list the secret for the specified `Client ID`.

- **Discovery URI:** This property specifies the discovery URI to use for fetching OP metadata.
- **Logout URL:** This property can be used to specify the logout URL for the SSO provider if you plan to configure the connection to log users out of the IDP when they log out of Anzo.
- **Scope:** This property can be used to specify the parameters to send to the authorization endpoint with the request.
- **Preferred JWS Algorithm:** This property can be used to specify the preferred signing algorithm.
- **Enable on login page:** This property controls whether to display a link for this provider on the Anzo login screen.
- **Callback URL:** This property specifies the URL that the provider should use to redirect users back to the Anzo application after a successful login. Include the full URL to the Anzo instance, through the proxy if one exists. Specify the URL in quotes and append the value with `/anzo_authenticate`, i.e., `"hostname:port/anzo_authenticate"`.
- **Callback URL port replacement:** This property can be used to define the port to use if the one specified in the `Callback URL` field is unavailable.
- **User Identifier:** This property specifies the SSO provider attribute, such as `email` or `username`, to use for looking up users in the directory server.
- **IDP Logout Capable:** This property can be used to indicate whether the SSO provider supports logging the user out of the IDP when they log out of Anzo.
- **Default to IDP Logout:** This property controls whether to log a user out of the IDP by default when they log out of Anzo.
- **Logout URL Suffix:** When `Default to IDP Logout` is enabled, this property can be used to specify the logout URL for the SSO provider. The `[urlAfterLogout]` placeholder is replaced with the SSO provider server URL.

- **Email Template regex:** If an email attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between email addresses stored by the SSO provider and email addresses returned by the directory server.
- **Email Template Replacement:** This property can be used to define a replacement email template to use if there are variations found by `Email Template regex`.
- **User Template regex:** If a username attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between user names stored by the SSO provider and names returned by the directory server.
- **User Template Replacement:** This property can be used to define a replacement user template to use if there are variations found by `User Template regex`.
- **Use username directly:** This property controls whether the identity provider directly authenticates a user by validating a username and password or by validating an assertion about the user's identity as defined by a separate identity provider.
- **Skip CSRF check:** This property controls whether to perform or skip a cross-site request forgery (CSRF) check.
- **LDAP domain:** This property identifies the LDAP domain to use for user lookup.
- **LDAP email property:** This property defines the LDAP email property to use to find the associated user's dn. For example,  
`http://openanzo.org/ontologies/2008/07/Anzo#ldapEmailInfo`.
- **Principal Template:** This property can be used to define the template to use for populating roles and returning user URIs.
- **Icon:** This property can be used to include an SSO icon on the Anzo login screen. To select an image, click the **Icon** field and select **Add File**.
- **With State:** This property controls whether information about the application's state is included in authentication requests.

- **Trust All:** This property controls whether all responses from the identity provider are trusted.

## Google

- **Title:** This property sets the name for the connection that you are creating.
- **Description:** This property can be used to provide a brief description of the provider configuration.
- **Enable on matched container ID:** This property sets the list of container IDs to match. This provider will be active if the request container ID matches one of the listed container IDs. Click the field and select a container ID from the drop-down list. To specify multiple IDs, click the field again and select another value. To remove a container from the list, click the X on the right of the container name.
- **Client ID:** This property specifies the client ID or consumer key value from the provider application.
- **Secret: Password and Confirm Password:** These properties list the secret for the specified `Client ID`.
- **Scope:** This property can be used to specify the parameters to send to the authorization endpoint with the request.
- **Preferred JWS Algorithm:** This property can be used to specify the preferred signing algorithm.
- **Enable on login page:** This property controls whether to display a link for this provider on the Anzo login screen.
- **Callback URL:** This property specifies the URL that the provider should use to redirect users back to the Anzo application after a successful login. Include the full URL to the Anzo instance, through the proxy if one exists. Specify the URL in quotes and append the value with `/anzo_authenticate`, i.e., `"hostname:port/anzo_authenticate"`.



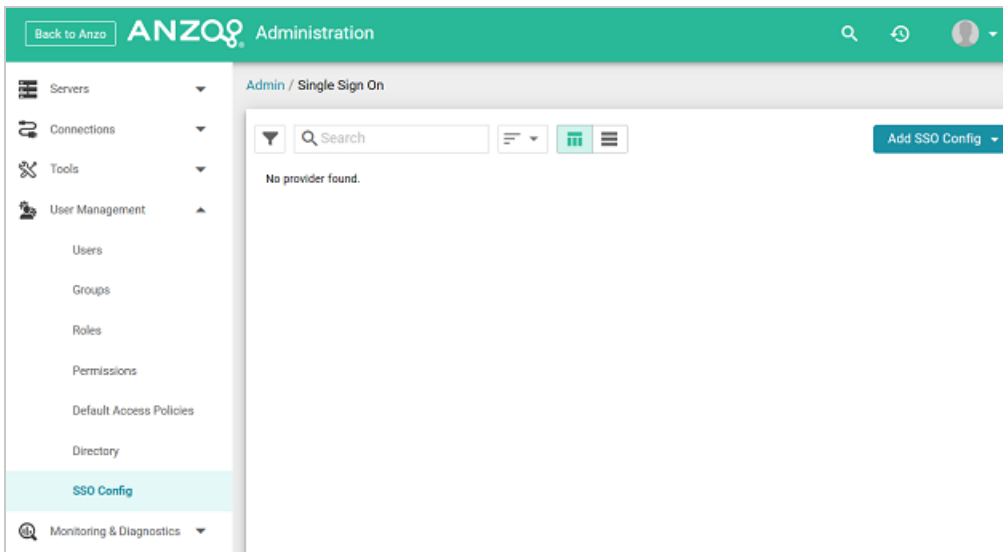
- **Callback URL port replacement:** This property can be used to define the port to use if the one specified in the `Callback URL` field is unavailable.
- **User Identifier:** This property specifies the SSO provider attribute, such as `email` or `username`, to use for looking up users in the directory server.
- **IDP Logout Capable:** This property can be used to indicate whether the SSO provider supports logging the user out of the IDP when they log out of Anzo.
- **Default to IDP Logout:** This property controls whether to log a user out of the IDP by default when they log out of Anzo.
- **Logout URL Suffix:** When `Default to IDP Logout` is enabled, this property can be used to specify the logout URL for the SSO provider. The `[urlAfterLogout]` placeholder is replaced with the SSO provider server URL.
- **Email Template regex:** If an email attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between email addresses stored by the SSO provider and email addresses returned by the directory server.
- **Email Template Replacement:** This property can be used to define a replacement email template to use if there are variations found by `Email Template regex`.
- **User Template regex:** If a username attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between user names stored by the SSO provider and names returned by the directory server.
- **User Template Replacement:** This property can be used to define a replacement user template to use if there are variations found by `User Template regex`.
- **Use username directly:** This property controls whether the identity provider directly authenticates a user by validating a username and password or by validating an assertion about the user's identity as defined by a separate identity provider.

- **Skip CSRF check:** This property controls whether to perform or skip a cross-site request forgery (CSRF) check.
  - **LDAP domain:** This property identifies the LDAP domain to use for user lookup.
  - **LDAP email property:** This property defines the LDAP email property to use to find the associated user's dn. For example,  
`http://openanzo.org/ontologies/2008/07/Anzo#ldapEmailInfo.`
  - **Principal Template:** This property can be used to define the template to use for populating roles and returning user URIs.
  - **Icon:** This property can be used to include an SSO icon on the Anzo login screen. To select an image, click the **Icon** field and select **Add File**.
  - **With State:** This property controls whether information about the application's state is included in authentication requests.
  - **Trust All:** This property controls whether all responses from the identity provider are trusted.
4. When you have finished configuring properties, click **Save** to save the provider setup.

## Adding a SAML Provider

Follow the steps below to add a SAML SSO Provider.

1. In the Administration application, expand **User Management** and click **SSO Config**. Anzo displays the Single Sign On screen, which lists any existing providers. For example:



2. Click the **Add SSO Config** button and select **SAML Provider**. The Create SAML Provider screen is displayed.

3. Configure the required properties and any optional settings as needed. The list below describes the properties.

- **Title:** This property sets the name for the connection that you are creating.
- **Description:** This property can be used to provide a brief description of the provider configuration.
- **Enable on matched container ID:** This property sets the list of container IDs to match. This provider will be active if the request container ID matches one of the listed container IDs. Click the field and select a container ID from the drop-down list. To specify multiple IDs, click the field again and select another value. To remove a container from the list, click the X on the right of the container name.
- **Identity Provider Metadata:** This property can be used to include the identity provider metadata .xml file. To add the file, click the field and then click **Add File** to select the file.

- **Service Provider Entity ID:** This property can be used to identify the service provider to the identity provider during the SSO process. The entity ID is a unique identifier for the service provider, typically a URL that points to the service provider's metadata.
- **Service Provider Metadata:** This property can be used to include the server provider metadata .xml file. To add the file, click the field and then click **Add File** to select the file.
- **Authentication Request Binding:** This property can be used to specify the redirect binding that the service provider should use to pass an authentication request to the identity provider.
- **Maximum Authentication Lifetime (seconds):** This property can be used to adjust the amount of time the authentication spans. By default, the SAML client accepts assertions for one hour based on a previous authentication. To change the lifetime, you can set this property to the desired number of seconds.
- **Enable on login page:** This property controls whether to display a link for this provider on the Anzo login screen.
- **Callback URL:** This property specifies the URL that the provider should use to redirect users back to the Anzo application after a successful login. Include the full URL to the Anzo instance, through the proxy if one exists. Specify the URL in quotes and append the value with `/anzo_authenticate`, i.e., `"hostname:port/anzo_authenticate"`.
- **Callback URL port replacement:** This property can be used to define the port to use if the one specified in the `Callback URL` field is unavailable.
- **User Identifier:** This property specifies the SSO provider attribute, such as `email` or `username`, to use for looking up users in the directory server.
- **IDP Logout Capable:** This property can be used to indicate whether the SSO provider supports logging the user out of the IDP when they log out of Anzo.

- **Default to IDP Logout:** This property controls whether to log a user out of the IDP by default when they log out of Anzo.
- **Logout URL Suffix:** When `Default to IDP Logout` is enabled, this property can be used to specify the logout URL for the SSO provider. The `[urlAfterLogout]` placeholder is replaced with the SSO provider server URL.
- **Email Template regex:** If an email attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between email addresses stored by the SSO provider and email addresses returned by the directory server.
- **Email Template Replacement:** This property can be used to define a replacement email template to use if there are variations found by `Email Template regex`.
- **User Template regex:** If a username attribute was specified as the User Identifier, this property can be used to specify a regular expression to use for identifying variations between user names stored by the SSO provider and names returned by the directory server.
- **User Template Replacement:** This property can be used to define a replacement user template to use if there are variations found by `User Template regex`.
- **Use username directly:** This property controls whether the identity provider directly authenticates a user by validating a username and password or by validating an assertion about the user's identity as defined by a separate identity provider.
- **Skip CSRF check:** This property controls whether to perform or skip a cross-site request forgery (CSRF) check.
- **LDAP domain:** This property identifies the LDAP domain to use for user lookup.
- **LDAP email property:** This property defines the LDAP email property to use to find the associated user's dn. For example,  
`http://openanzo.org/ontologies/2008/07/Anzo#ldapEmailInfo`.

- **Principal Template:** This property can be used to define the template to use for populating roles and returning user URIs.
  - **Icon:** This property can be used to include an SSO icon on the Anzo login screen. To select an image, click the **Icon** field and select **Add File**.
  - **With State:** This property controls whether information about the application's state is included in authentication requests.
4. When you have finished configuring properties, click **Save** to save the provider setup.

# Creating and Managing Roles

In Anzo, groups (or users) are added to roles and the roles are configured to grant access to *functionality*. Role permissions control access to menus and screens in the Anzo and Administration applications. Access to functionality cannot be assigned to groups or users, only to roles.

## Tip

For more information about role, user, and group management, see [User Management Concepts](#).

This topic provides instructions for creating or changing the roles to use for controlling access to Anzo functionality. For information about the predefined Anzo roles, see [Predefined Anzo Roles and Permissions](#).

- [Creating a New Role](#)
- [Adding Users or Groups to a Role](#)
- [Configuring Role Permissions](#)

## Creating a New Role

1. In the Administration application, expand the **User Management** menu and click **Roles**. Anzo displays the Roles screen, which lists the existing roles. For example:



<input type="checkbox"/>	Name	Members	Updated Date	Tags	Actions
	Anzo Administrator	anzo admin	Nov 10, 2022		
	Data Analyst	data analyst	Nov 10, 2022		
	Data Citizen	data citizen	Nov 10, 2022		
	Data Curator	data curator	Nov 10, 2022		
	Data Governor	data governor	Nov 10, 2022		
	Data Scientist				

Rows per page: 25 1-6 of 6

2. On the Roles screen, click the **Add Role** button. Anzo displays the Create Role dialog box.

**Create Role**

Name\*

Description

Members

The members of the role

Permissions

The roles permissions

Copy URI

CANCEL SAVE

3. Complete the required fields and enter any optional group details:

- **Name:** The name for the new role.
- **Description:** An optional description of the role.
- **Members:** The users or groups who are members of the role. Click the **Members** field to select a member. Click the field again to select additional members.

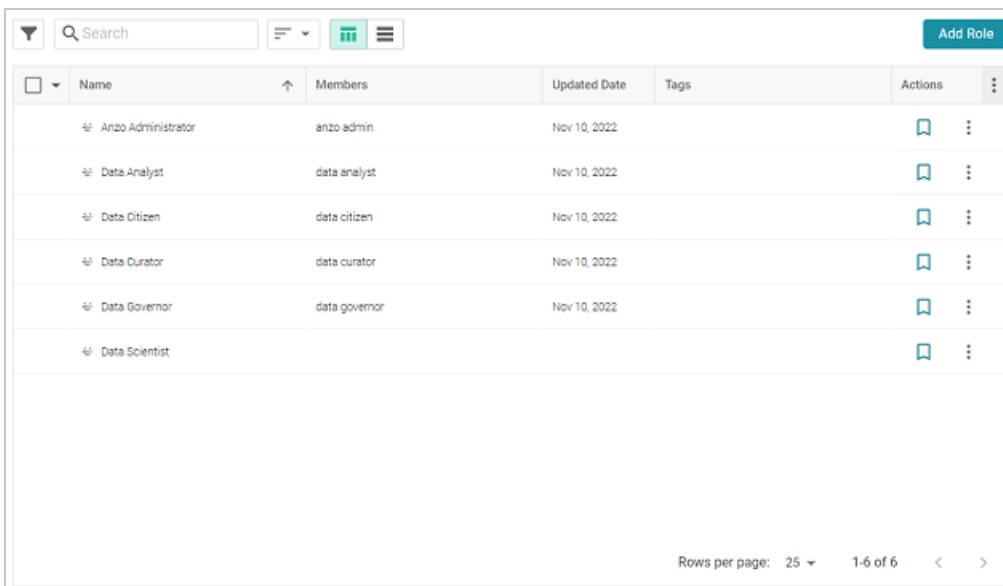
- **Permissions:** The list of Anzo features that this role has permission to access. Click the **Permissions** field and select a permission to add it to the list. Click the field again to select additional permissions. For details about each of the permissions, see the [Role Permissions Reference](#).

4. Click **Save** to add the role to the system. Anzo adds the new role to the list of roles on the Roles screen.

## Adding Users or Groups to a Role

Follow the instructions below to add users and/or groups to a role.

1. In the Administration application, expand the **User Management** menu and click **Roles**. Anzo displays the Roles screen, which lists the existing roles. For example:



<input type="checkbox"/>	Name	Members	Updated Date	Tags	Actions
<input type="checkbox"/>	Anzo Administrator	anzo admin	Nov 10, 2022		<a href="#">Bookmark</a> <a href="#">More</a>
<input type="checkbox"/>	Data Analyst	data analyst	Nov 10, 2022		<a href="#">Bookmark</a> <a href="#">More</a>
<input type="checkbox"/>	Data Citizen	data citizen	Nov 10, 2022		<a href="#">Bookmark</a> <a href="#">More</a>
<input type="checkbox"/>	Data Curator	data curator	Nov 10, 2022		<a href="#">Bookmark</a> <a href="#">More</a>
<input type="checkbox"/>	Data Governor	data governor	Nov 10, 2022		<a href="#">Bookmark</a> <a href="#">More</a>
<input type="checkbox"/>	Data Scientist				<a href="#">Bookmark</a> <a href="#">More</a>

Rows per page: 25 1-6 of 6

2. Click the name of the role that you want to add users or groups to. Anzo opens the Edit Role dialog box. For example:

**Edit Role**

Name\*  
Data Analyst

Description  
Members of this role can create graphmarts and view dashboard analytics.

data analyst x

The members of the role

Permissions

Data On Demand x Create Dashboards x View Graphmarts x View Datasets x  
Anzo Application x View Activity Logs x Hi-Res Analytics x Create Graphmarts x  
Browse Dashboards x Show Query Builder x

The roles permissions

Copy URI

CANCEL SAVE

3. Click the **Members** drop-down list to display the list of all available users and groups. You can also search for a user or group by typing a name in the Members field. Click a name to add that user or group to the role. Click the field again to select additional members. To remove a member from the role, click the X to the right of the name.

### Note

If you do not see users or groups that you expect to see, it is possible that Anzo is out of sync with the directory server. If groups or users have been modified on the directory server, and a user has not logged in to Anzo for an extended time, the data may need to be refreshed in Anzo. The **Users** and **Groups** screens in the User Management menu have **Sync Directory** buttons that you can click to synchronize with the directory server and update the data in Anzo.

4. When you have finished adding members, click **Save** to save the changes to the role.

### Note

When modifying an existing user's access by adding or removing roles from their account, Cambridge Semantics recommends that the user logs out of Anzo and clears their browser cache to ensure that the access changes are reflected in the user interface.

# Configuring Role Permissions

Follow the instructions below to add or remove permissions from a role. For details about each of the permissions, see the [Role Permissions Reference](#).

1. In the Administration application, expand the **User Management** menu and click **Roles**. Anzo displays the Roles screen, which lists the existing roles. For example:

Name	Members	Updated Date	Tags	Actions
Anzo Administrator	anzo admin	Nov 10, 2022		
Data Analyst	data analyst	Nov 10, 2022		
Data Citizen	data citizen	Nov 10, 2022		
Data Curator	data curator	Nov 10, 2022		
Data Governor	data governor	Nov 10, 2022		
Data Scientist				

2. Click the name of the role for which you want to configure permissions. Anzo opens the Edit Role dialog box. For example:

**Edit Role**

Name\*  
Data Analyst

Description  
Members of this role can create graphmarts and view dashboard analytics.

Members of the role  
data analyst

Permissions  
Data On Demand, Create Dashboards, View Graphmarts, View Datasets, Anzo Application, View Activity Logs, Hi-Res Analytics, Create Graphmarts, Browse Dashboards, Show Query Builder

The roles permissions

Copy URI

CANCEL SAVE

3. The **Permissions** field lists all of the permissions that are applied to the role. To remove a permission, click the X to the right of the permission name. To add a permission click the field to open the Permissions drop-down list. Click a name to add that permission to the role. Click the field again to select additional permissions.
4. When you have finished changing permissions, click **Save** to save the changes to the role.

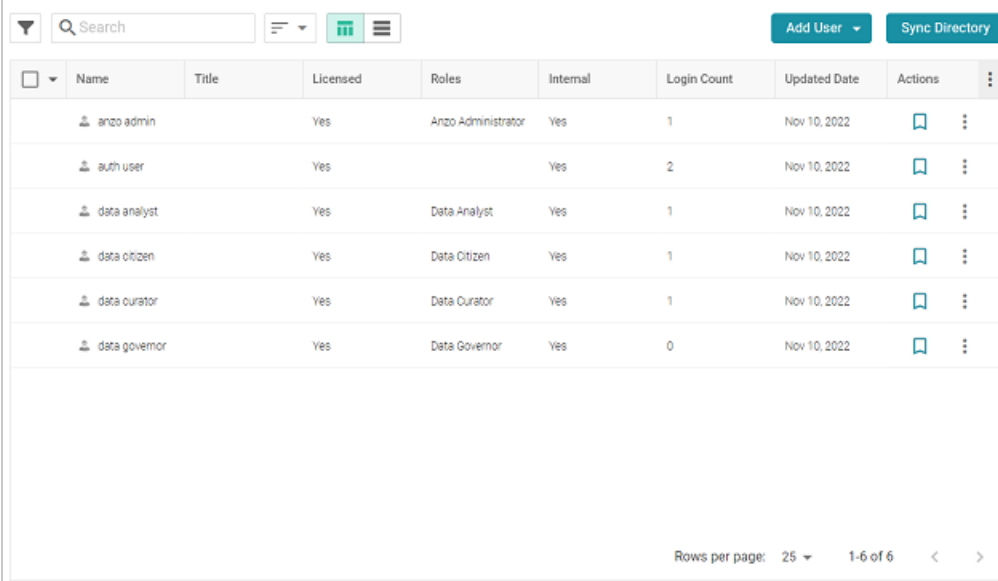
# Creating an Internal Anzo User

User and group accounts are typically managed in a central directory server that is connected to Anzo. The groups from the directory server are added to Anzo roles, and access to Anzo applications and features is configured for the roles. However, you can create a user account directly in Anzo. Accounts that are created in Anzo are stored in Anzo's internal LDAP server. Follow the instructions below to create a new internal Anzo user account.

## Tip

For instructions on adding directory users to Anzo, see [Adding Directory Users and Groups to Anzo](#).

1. In the Administration application, expand the **User Management** menu and click **Users**. Anzo displays the Users screen, which lists the existing users. For example:



The screenshot shows the 'Users' screen in the Anzo Administration application. At the top, there is a search bar, a filter icon, and two buttons: 'Add User' and 'Sync Directory'. Below this is a table with the following columns: Name, Title, Licensed, Roles, Internal, Login Count, Updated Date, and Actions. The table lists six users: 'anzo admin', 'auth user', 'data analyst', 'data citizen', 'data curator', and 'data governor'. Each user row includes a checkmark in the 'Licensed' column, a role name in the 'Roles' column, a 'Yes' or 'No' in the 'Internal' column, a login count, an updated date of 'Nov 10, 2022', and a set of action icons (edit, delete, etc.). At the bottom right, there is a pagination control showing 'Rows per page: 25' and '1-6 of 6'.

<input type="checkbox"/>	Name	Title	Licensed	Roles	Internal	Login Count	Updated Date	Actions
<input type="checkbox"/>	anzo admin		Yes	Anzo Administrator	Yes	1	Nov 10, 2022	
<input type="checkbox"/>	auth user		Yes		Yes	2	Nov 10, 2022	
<input type="checkbox"/>	data analyst		Yes	Data Analyst	Yes	1	Nov 10, 2022	
<input type="checkbox"/>	data citizen		Yes	Data Citizen	Yes	1	Nov 10, 2022	
<input type="checkbox"/>	data curator		Yes	Data Curator	Yes	1	Nov 10, 2022	
<input type="checkbox"/>	data governor		Yes	Data Governor	Yes	0	Nov 10, 2022	

2. On the Users screen, click the **Add User** button and select **Add User**. Anzo opens the Create User dialog box.

The screenshot shows a 'Create User' form with the following fields and controls:

- Username \*
- First Name \*
- Last Name \*
- Password \* (with an eye icon for visibility toggle)
- Confirm Password \* (with an eye icon for visibility toggle)
- Licensed
- Position / Title
- Email
- CANCEL and SAVE buttons at the bottom right.

- Complete the required fields and enter any optional user details:
  - Username:** The user name that the user will use to log in to Anzo.
  - First Name:** The user's first name.
  - Last Name:** The user's last name.
  - Password and Confirm Password:** Type a password for the user.
  - Licensed:** Select the **Licensed** checkbox if you want this user to be able to log in to the Anzo applications. If you want to add this user to the system but do not want to give him or her access to Anzo applications at this time, clear the Licensed checkbox.
  - Position/Title:** The user's job title or position.
  - Email:** The user's email address.
  - Phone:** The user's phone number.
  - Roles:** The role or roles that the user is a member of. Roles define the user's level of access to Anzo applications and features. Click the **Roles** field and select a role from the drop-down list. Click the field again to select additional roles.
- When you have finished configuring the user account, click **Save** to add the user to the system.

For more information about roles, see [Creating and Managing Roles](#). For a description of the default Anzo roles, see [Predefined Anzo Roles and Permissions](#).



# Predefined Anzo Roles and Permissions

This topic describes the roles that are predefined in Anzo and lists the permissions that are assigned to each role by default. The predefined roles can be removed or modified as desired. For instructions on changing roles, see [Creating and Managing Roles](#).

- [System Administrator](#)
- [Base Permissions \(Everyone and Authenticated User Roles\)](#)
- [Anzo Administrator](#)
- [Data Analyst](#)
- [Data Citizen](#)
- [Data Curator](#)
- [Data Governor](#)
- [Data Scientist](#)

## System Administrator

The System Administrator account, usually named **sysadmin**, is created during the Anzo installation. This account has permission to access all Anzo features in the main Anzo application as well as administrative features in the Administration application. In addition, the sysadmin user has read and write access to all of the artifacts (such as data sources, models, and pipelines) that are created by all Anzo users. The sysadmin user permissions cannot be changed, and the account cannot be deleted. In addition, artifacts cannot be configured to restrict sysadmin access. For information about changing the system administrator password, see [Change the sysadmin Password](#).

## Base Permissions (Everyone and Authenticated User Roles)

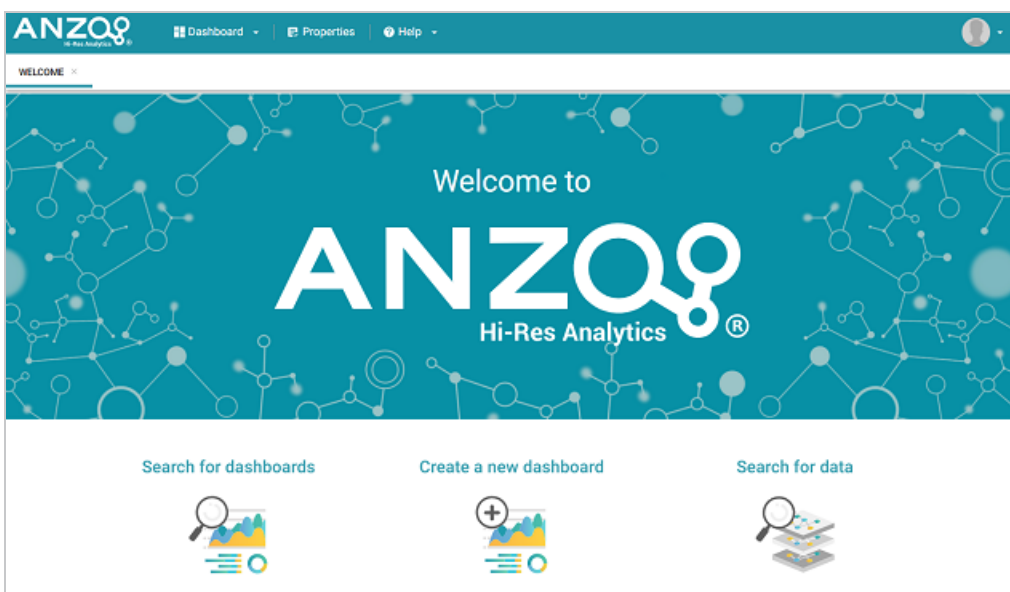
There is a set of base permissions that are applied to all user accounts by default. If a user account is created in Anzo but no roles are assigned, that user has the permissions of the **Authenticated User** role. By default, authenticated users cannot access the Anzo application but can access the

Hi-Res Analytics application where they can browse for and create dashboards. They can also view data that is shared from Data on Demand endpoints.

### Note

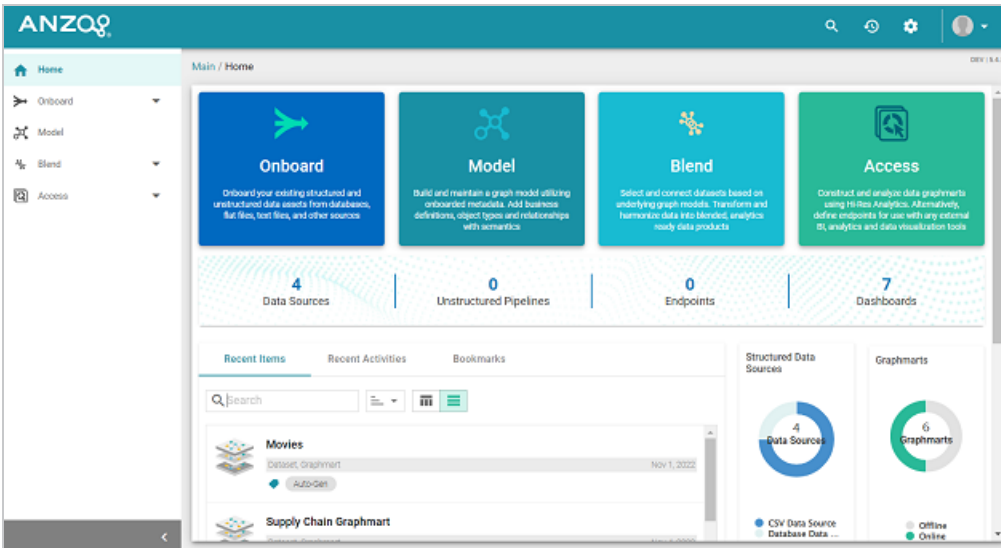
If **Anonymous User Access** is enabled on the system, unauthenticated users (users who do not have a user account in Anzo) have the permissions that are included in the **Everyone** role. The Everyone role is only used to apply permissions for unauthenticated users when anonymous access is allowed. For information about anonymous access, see [Configure Anonymous User Access](#).

The image below shows an example of the view an authenticated user has in the Hi-Res Analytics application.



## Anzo Administrator

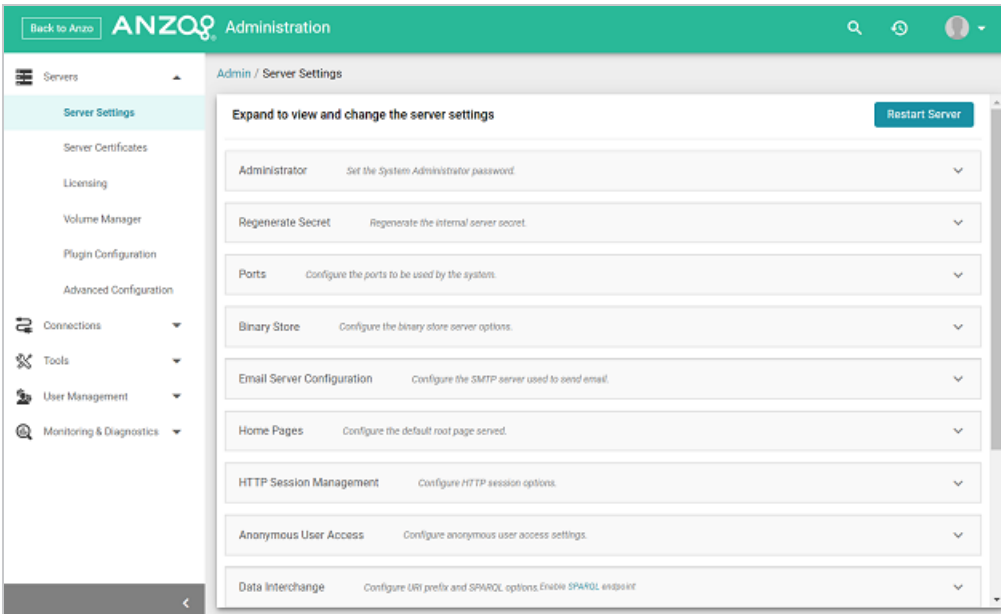
By default the Anzo Administrator role has access to all menus and features in the Anzo application as well as the Administration application. The image below shows an example of the view a user with the Anzo Administrator role has in the Anzo application.



## Note

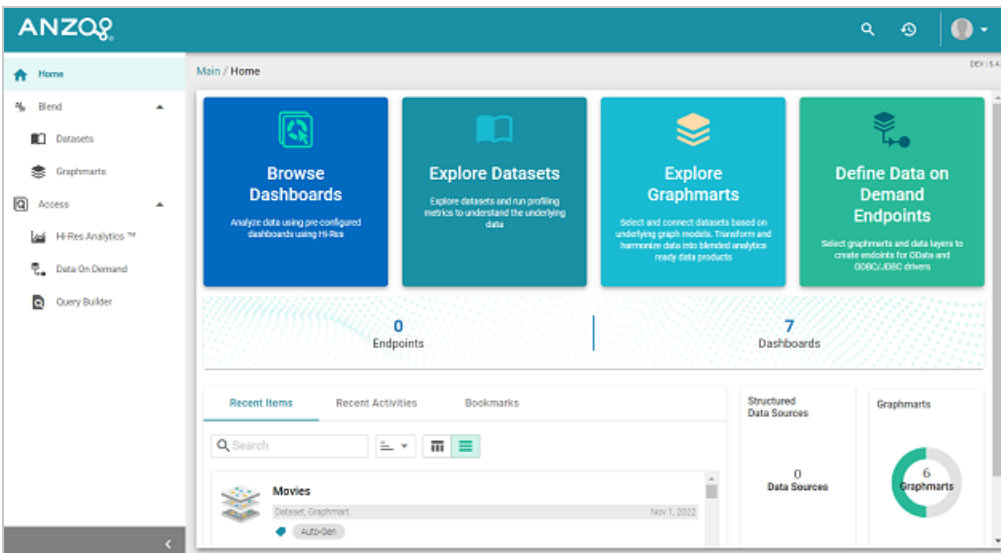
Having full access to all features does not mean the Anzo Administrator has full access to all of the data in the system. Unlike the **sysadmin** user, Anzo Administrators must still be granted data access to specific artifacts. For example, if the Anzo Administrator role is assigned the Manage Graphmarts permission, a user with that role can activate and deactivate all graphmarts, but they cannot view the data contained in all of the graphmarts unless that role is granted data access to each graphmart.

The following image shows an example of the Anzo Administrator view of the Administration application.



## Data Analyst

By default the Data Analyst role has access to the Blend menu, Access menu, and Activity Log in the Anzo application. The image below shows an example of the view a user with the Data Analyst role has in the Anzo application.



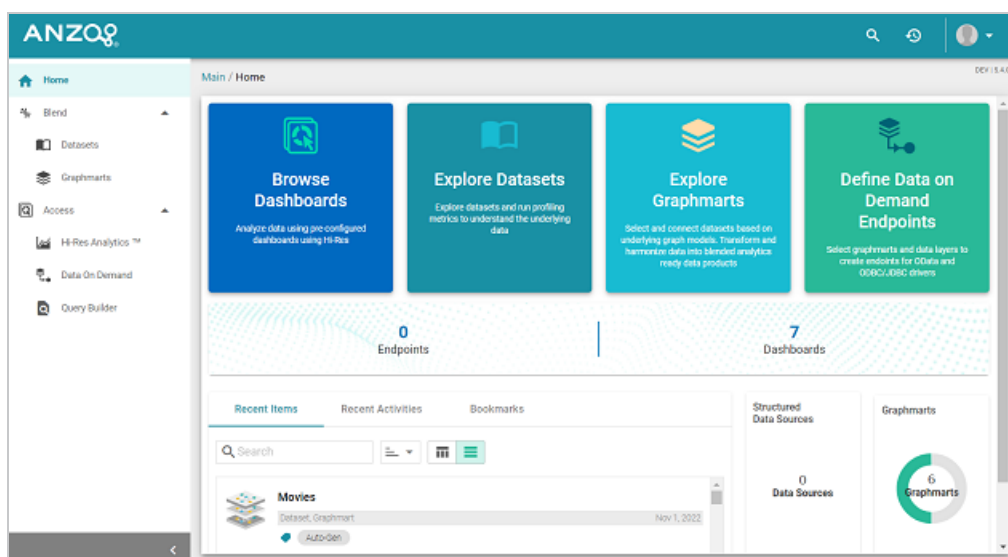
Members of the Data Analyst role can:

- View the Datasets catalog
- View and create graphmarts

- View and create Hi-Res Analytics dashboards
- View the Activity Log
- Access data with the Query Builder
- Create and access Data on Demand endpoints

## Data Citizen

By default the Data Citizen role has access to the Blend menu, Access menu, and Activity Log in the Anzo application. The image below shows an example of the view a user with the Data Citizen role has in the Anzo application.

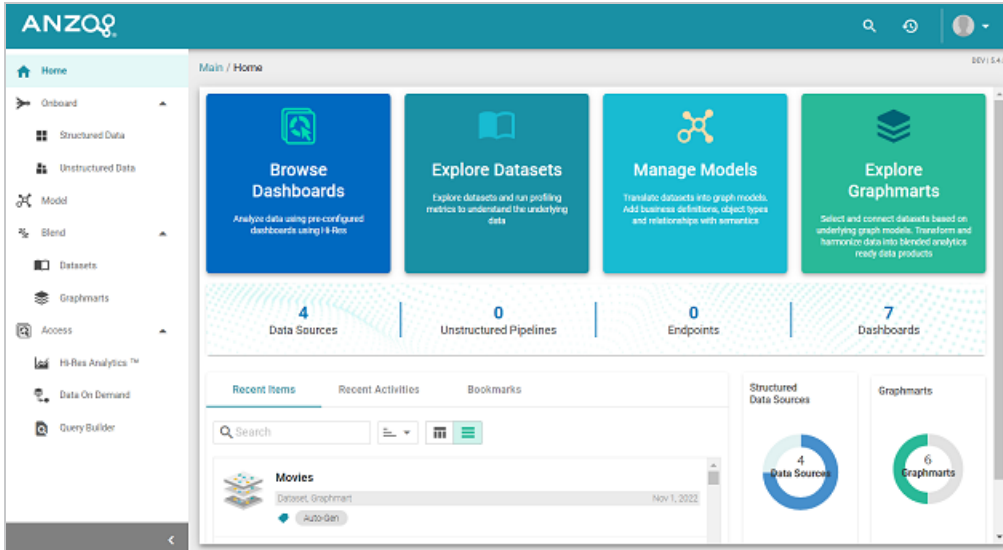


Members of the Data Citizen role can:

- View the Datasets catalog
- View graphmarts
- View and create Hi-Res Analytics dashboards
- View the Activity Log
- Access data with the Query Builder
- Create and access Data on Demand endpoints

# Data Curator

By default the Data Curator role has access to the Onboard menu, Model manager, Blend menu, Access menu, and Activity Log in the Anzo application. The image below shows an example of the view a user with the Data Curator role has in the Anzo application.

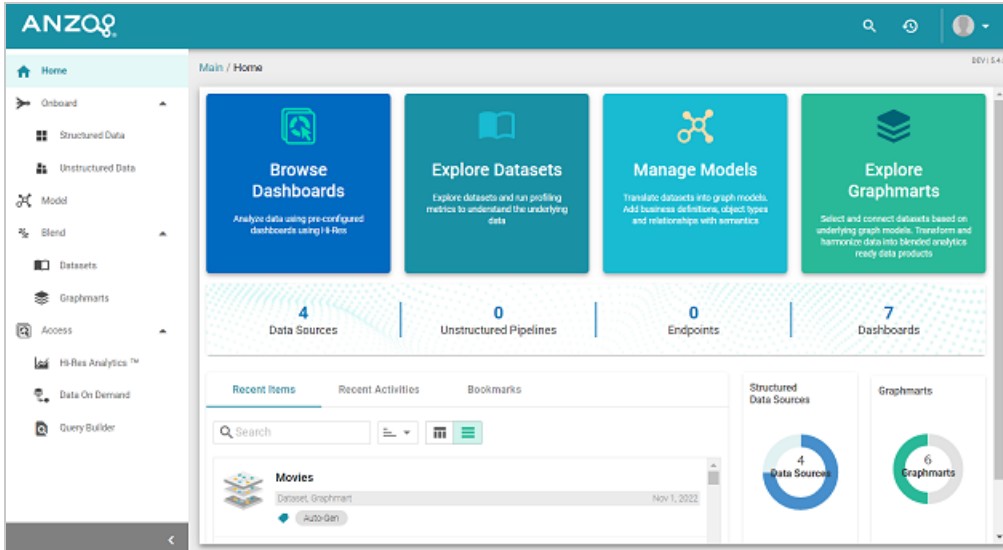


Members of the Data Curator role can:

- Connect to data sources and onboard structured and unstructured data
- View and create data models
- View the Datasets catalog
- View and create graphmarts
- View and create Hi-Res Analytics dashboards
- View the Activity Log
- Access data with the Query Builder
- Create and access Data on Demand endpoints

# Data Governor

By default the Data Governor role has access to the Onboard menu, Model manager, Blend menu, Access menu, and Activity Log in the Anzo application. The image below shows an example of the view a user with the Data Governor role has in the Anzo application.

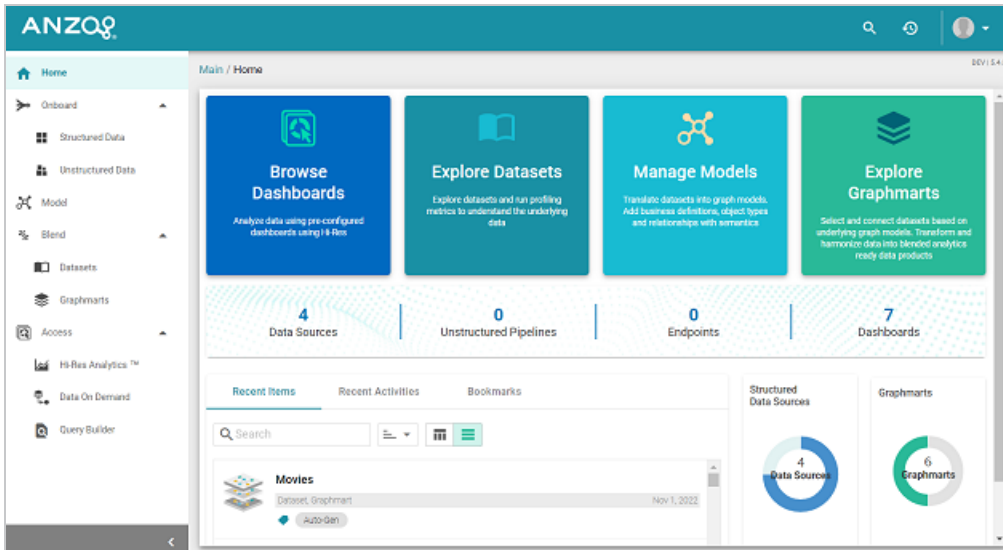


Members of the Data Governor role can:

- Connect to data sources and onboard structured and unstructured data
- View and create data models
- View the Datasets catalog
- View and create graphmarts
- View and create Hi-Res Analytics dashboards
- View the Activity Log
- Access data with the Query Builder
- Create and access Data on Demand endpoints

# Data Scientist

By default the Data Scientist role has access to the Onboard menu, Model manager, Blend menu, Access menu, and Activity Log in the Anzo application. The image below shows an example of the view a user with the Data Scientist role has in the Anzo application.



Members of the Data Scientist role can:

- Connect to data sources and onboard structured and unstructured data
- View and create data models
- View the Dataset catalog
- View and create graphmarts
- View and create Hi-Res Analytics dashboards
- View the Activity Log
- Access data with the Query Builder
- Create and access Data on Demand endpoints

To review the specific permissions for each role, select **Roles** in the **User Management** menu in the Admin application. Click a role to open the Edit dialog box and review the permissions. For more information about the permissions, see [Role Permissions Reference](#).



# Role Permissions Reference

This topic provides details about each of the permissions that can be applied to roles. These permissions grant access to functionality, i.e., the menus and screens in the Anzo and Administration applications. For example, role permissions determine whether a member of a role can access the **Onboard** menu and create a new data source or see the **Blend** menu and create a new graphmart. Whether a member can view, modify, or delete a data source or graphmart artifact that is created by someone else, however, is controlled by the user or group permissions that are applied at the artifact level.

## Tip

For more information about artifact-level permissions, see [Artifact Access Control Concepts](#). And for more information about roles versus users and groups, see [User Management Concepts](#).

- [Permissions Screen](#)
- [Permission Descriptions](#)

## Permissions Screen

To view an overview of the configured permissions for all Anzo roles, you can view the **Permissions** page under the **User Management** menu in the Administration application. The screen displays a table; the heading row lists each role, and the first column lists each permission. The permissions are grouped into categories, such as Application or Data Onboarding. For example:

	Everyone	Authenticated Users	Anzo Administrator	Data Analyst	Data Citizen	Data Curator	Data Governor	Data Scientist
<b>Default</b>								
Activate Graphmarts	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Browse Dashboards	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Browse Models	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Create Dashboards	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Create Graphmarts	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Data On Demand	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Import Artifacts	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Manage Graphmarts	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Manage Models	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Rest API	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Show Query Builder	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
View Datasets	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
View Graphmarts	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>Administration</b>								

The rows for each role column include checkboxes that control permissions. You can select or clear checkboxes to enable or disable permissions for a role.

## Permission Descriptions

The tables below list the permissions in each category and describe the pages and menus that are enabled for members of a role where that permission is applied.

### Note

The permissions described below give access to functionality in the Anzo and Administration applications. Whether members of the role have view or edit access to certain datasets, models, dashboards, graphmarts, etc. depends on the permissions that are granted at the artifact level.

### Default

Permission	Description
<b>Activate Graphmarts</b>	If the user has the appropriate permissions at the graphmart level, this permission allows them to activate and deactivate graphmarts and import

Permission	Description
	<p>graphmarts into Anzo. Does not give permission to create new graphmarts or delete graphmarts.</p> <p>To be able to access a Graphmart screen in the Anzo application and move the <b>Inactive</b> → <b>Active</b> slider, the <b>Anzo Application</b> permission also needs to be applied.</p>
<b>Browse Dashboards</b>	Gives permission to view existing dashboards in the Hi-Res Analytics application. Does not give permission to create new dashboards.
<b>Browse Models</b>	Gives permission to view existing data models. Applying this permission exposes the <b>Models</b> menu item in the Anzo application. Must also have the <b>Anzo Application</b> permission to access the Anzo application.
<b>Create Dashboards</b>	Gives permission to create dashboards in the Hi-Res Analytics application. Applying this permission also exposes the <b>Create Dashboard</b> button on the Graphmart screens in the Anzo application when the user has the <b>Anzo Application</b> permission.
<b>Create Graphmarts</b>	Gives permission to create new graphmarts. Applying this permission exposes the <b>Add Graphmart</b> button on the Graphmarts screen. Must also have the <b>Anzo Application</b> permission to create graphmarts in the application.
<b>Data on Demand</b>	If the user has the appropriate permissions at the graphmart level, this permission enables the user to create Data on Demand endpoints. Applying this permission enables the <b>Create New Endpoint</b> button on the Data on Demand tab for graphmarts. Must also have the <b>Anzo Application</b> permission to access the application.
<b>Import Artifacts</b>	Gives permission to perform Import operations from the Anzo application. If a user is a member of a role that has <b>Import Artifact</b> assigned, they will see the <b>Import</b> option in the menu when they click the <b>Add</b> button to add a data source,


Permission	Description
	dataset, model, etc. Must also have the <b>Anzo Application</b> permission.
<b>Manage Graphmarts</b>	Gives permission to manage permissions for graphmarts. Must also have the <b>Anzo Application</b> permission to access the graphmart screens.
<b>Manage Models</b>	Gives permission to create and import models. Must also have the <b>Anzo Application</b> permission to access the Model screen.
<b>Rest API</b>	Gives permission to send requests via the Anzo REST API.
<b>Show Query Builder</b>	Gives permission to find data and run SPARQL queries using the Query Builder. Applying this permission exposes the <b>Query Builder</b> option in the <b>Access</b> menu. Must also have the <b>Anzo Application</b> permission.
<b>View Datasets</b>	Gives permission to view the Datasets catalog. Applying this permission exposes the <b>Datasets</b> option in the <b>Blend</b> menu in the Anzo application. Must also have the <b>Anzo Application</b> permission.
<b>View Graphmarts</b>	Gives permission to view the list of existing graphmarts. Must also have the <b>Anzo Application</b> permission to view the Graphmarts screen in the Anzo application.

## Administration

Permission	Description
<b>Administer System Setup</b>	<p>Gives permission to access the options in the Administration application that are related to system setup, such as <b>Server Settings</b>, <b>Licensing</b>, <b>Anzo Data Store</b>, and <b>Directory</b> server configuration.</p> <p>The image below shows the view of the Administration menu that users have if <b>Administer System Setup</b> and <b>Anzo Application</b> are the only two</p>

Permission	Description																								
	<p>applied permissions:</p> <table border="1" data-bbox="409 264 1468 550"> <thead> <tr> <th data-bbox="409 264 665 310">SERVERS</th> <th data-bbox="665 264 899 310">CONNECTIONS</th> <th data-bbox="899 264 1159 310">USER MANAGEMENT</th> <th data-bbox="1159 264 1468 310">MONITORING &amp; DIAGNOSTICS</th> </tr> </thead> <tbody> <tr> <td data-bbox="409 310 665 357">Server Settings</td> <td data-bbox="665 310 899 357">Anzo Data Store</td> <td data-bbox="899 310 1159 357">Default Access Policies</td> <td data-bbox="1159 310 1468 357">System Query Audit</td> </tr> <tr> <td data-bbox="409 357 665 403">Licensing</td> <td data-bbox="665 357 899 403">Elasticsearch Config</td> <td data-bbox="899 357 1159 403">Directory</td> <td data-bbox="1159 357 1468 403">Semantic Services</td> </tr> <tr> <td data-bbox="409 403 665 449">Volume Manager</td> <td data-bbox="665 403 899 449">Cloud Locations</td> <td data-bbox="899 403 1159 449">SSO Config</td> <td data-bbox="1159 403 1468 449">System Information</td> </tr> <tr> <td data-bbox="409 449 665 495">Plugin Configuration</td> <td></td> <td></td> <td></td> </tr> <tr> <td data-bbox="409 495 665 550">Advanced Configuration</td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <div data-bbox="409 600 1474 1083" style="background-color: #e6f2ff; padding: 10px; margin-top: 10px;"> <p><b>Note</b></p> <p>Some menu items in the above image, such as <b>Semantic Services</b>, <b>AnzoGraph</b>, and <b>Anzo Data Store</b>, are also controlled by more granular permissions: <b>Manage Semantic Services</b>, <b>Manage AnzoGraph</b>, and <b>Create Anzo Data Stores</b>. To give an administrator full create, modify, and delete access to those functions, the granular permissions need to be enabled in addition to <b>Administer System Setup</b>.</p> </div>	SERVERS	CONNECTIONS	USER MANAGEMENT	MONITORING & DIAGNOSTICS	Server Settings	Anzo Data Store	Default Access Policies	System Query Audit	Licensing	Elasticsearch Config	Directory	Semantic Services	Volume Manager	Cloud Locations	SSO Config	System Information	Plugin Configuration				Advanced Configuration			
SERVERS	CONNECTIONS	USER MANAGEMENT	MONITORING & DIAGNOSTICS																						
Server Settings	Anzo Data Store	Default Access Policies	System Query Audit																						
Licensing	Elasticsearch Config	Directory	Semantic Services																						
Volume Manager	Cloud Locations	SSO Config	System Information																						
Plugin Configuration																									
Advanced Configuration																									
<p><b>Batch Direct Data Loading</b></p>	<p>Gives permission to create a graphmart from multiple data sources at once when ingesting sources via graphmarts. For more information, see <a href="#">Loading Data Sources via the Automated Workflow</a> in the User Guide.</p>																								
<p><b>Manage AnzoGraph</b></p>	<p>Gives permission to view and create AnzoGraph connections. Does not give permission to delete connections or change the configuration of an existing connection. <b>Administer System Setup</b> is required to grant permission to delete and change existing AnzoGraph connections.</p>																								
<p><b>Manage Certificates</b></p>	<p>Gives permission to upload and delete server certificates.</p>																								
<p><b>Manage File</b></p>	<p>Gives permission to create new File Store connections and view existing</p>																								

Permission	Description
<b>Stores</b>	connections. Does not grant permission to delete or change existing file store connections. The <b>Administer System Setup</b> permission is required in conjunction with <b>Manage File Stores</b> to be able to delete or edit existing file stores.
<b>Manage Query Blocklists</b>	<p>Gives permission to create and remove queries from the <b>Query Blocklist</b> tab in the <b>System Query Audit Log</b>.</p> <p><b>Note</b> If a user only has the <b>Manage Query Blocklist</b> permission, the Administration menu is not available. Use this permission in conjunction with <b>Administer System Setup</b> to grant access to System Query Audit and the Query Blocklist.</p>
<b>Manage Semantic Services</b>	<p>Gives permission to stop and start semantic services from the Semantic Services screen as well as view details about the services and use the Service Builder to generate and run semantic service requests.</p> <p><b>Note</b> If a user only has the <b>Manage Semantic Services</b> permission, the Administration menu is not available. Use this permission in conjunction with <b>Administer System Setup</b> to grant access to the Semantic Services screen.</p>
<b>Manage Users, Groups, and Roles</b>	Gives permission to create, change, and delete users, groups, and roles. A user who has this permission has Admin level access to all users, groups, and roles.
<b>Profile Data</b>	Gives permission to profile datasets and graphmarts. Applying this permission exposes the <b>Profile Data</b> button on the Dataset and Graphmart screens.

Permission	Description
<b>Use Experimental Anzo Features</b>	Grants permission use experimental Anzo features. Experimental features are recently implemented and may not be reliable for production use.
<b>View Activity Logs</b>	Gives permission to view the Activity Log. Applying this permission exposes the Activity Log icon (  ) in the top menu bar of the Anzo and Administration applications. The <b>Anzo Application</b> permission is needed to give access to the Anzo application.
<b>View Log Files</b>	Gives permission to view and download log files from the Log Files tab. Does not grant permission to change logging levels or add new log packages. Use this permission in conjunction with <b>Administer System Setup</b> to grant access to configure log levels and packages.

## Application

Permission	Description
<b>Anzo Application</b>	Grants access to the main Anzo application.
<b>Anzo CLI</b>	Gives permission to use the administration command line interface.
<b>Hi-Res Analytics</b>	Grants access to the Hi-Res Analytics application.

## Data Onboarding

Permission	Description
<b>Create Anzo Data Stores</b>	Gives permission to create Anzo Data Stores. Must also have the <b>Administer System Setup</b> permission to make the <b>Anzo Data Store</b> option available in the Administration application.

Permission	Description
<b>Create Data Sources</b>	Gives permission to add new data sources. Does not give permission to delete existing sources. Must also have the <b>Anzo Application</b> and <b>Onboard Structured Data</b> permissions to access the Data Sources screen and add new sources.
<b>Onboard Structured Data</b>	Gives permission to access the <b>Onboard &gt; Structured Data</b> menu. Must also have the <b>Anzo Application</b> permission.
<b>Onboard Unstructured Data</b>	Gives permission to create pipelines to onboard unstructured data. Applying this permission exposes the <b>Onboard &gt; Unstructured Data</b> menu. Must also have the <b>Anzo Application</b> permission.

## Migration

Permission	Description
<b>Manage Migration Packages</b>	Gives permission to create, export, and import migration packages that include artifacts the user has access to.
<b>Perform Migration Package Operations As Sysadmin</b>	Gives permission to create, export, and import migration packages with sysadmin privileges. That means the package can include artifacts the user may not otherwise have permission to access.



## Managing Default Access Policies

Default Access Policies are the security policies that are applied by default to the artifacts that are stored in Anzo. Artifacts are all of the objects that are created when connections to data sources and applications are made and when data is onboarded to Anzo. For example, when users connect to a database or a file source, those connections are stored as artifacts, and when the data from a data source is ingested, the resulting schema, model, graphmart, and any generated datasets are also artifacts. All artifacts of the same type are stored in a particular **registry**, and each registry has an access policy associated with it. A registry is a system-level graph that stores metadata about artifacts of the same type. For example, metadata about all of your data source artifacts is stored in a Data Sources Registry, and metadata about all of your data model artifacts is stored in an Ontology Registry. A Default Access Policy defines the base permissions to assign to a type of artifact when it is created—before permission inheritance and user-configured sharing is applied.

### Note

Any **Permission Inheritance** that is applied by Anzo and artifact-level **Sharing** that is configured by users is applied to artifacts in addition to the permissions supplied by the Default Access Policy. For more information about permission inheritance and artifact sharing, see [Artifact Access Control Concepts](#).

This topic provides information about the permission sets that can be assigned to users and groups and describes the default access policies for each registry. This topic also includes instructions for changing access policies.

- [Default Access Policy Permissions Reference](#)
- [Default Access Policy Reference](#)
- [Configuring Default Access Policies](#)

## Default Access Policy Permissions Reference

Default access policies use the same predefined permission sets and mechanism for assigning permissions as other artifacts in the Anzo application (see [Share Access to Artifacts](#) in the User Guide for more information).

There are three predefined permission sets that include a combination of six permissions that can be assigned to the creator of an artifact and other users and groups. The tables below list the predefined sets and describe the privileges that are granted for each permission that is part of the set.

### View

The following table describes the permissions in the **View** set.

Permission	Description
<b>View</b>	This permission allows a user to: <ul style="list-style-type: none"><li>• See an artifact in the Anzo application.</li><li>• Create versions of the artifact.</li></ul>
<b>Meta View</b>	Relates only to an artifact's permissions. A user with Meta View can see the permissions on the artifact's <b>Sharing</b> tab but they cannot change permissions.

### Modify

In addition to the **View** and **Meta View** permissions described above, the **Modify** set includes the **Add/Edit** and **Delete** permissions described below.

Permission	Description
<b>Add/Edit</b>	This permission allows a user to: <ul style="list-style-type: none"><li>• Change an artifact, such as to rename it or edit its description.</li></ul>

Permission	Description
	<ul style="list-style-type: none"> <li>• Add a related entity to an artifact. For example, add a schema to a data source or a layer to a graphmart.</li> </ul>
<b>Delete</b>	<p>This permission allows a user to:</p> <ul style="list-style-type: none"> <li>• Remove a related entity from the artifact. For example, delete a layer from a graphmart or a schema from a data source.</li> <li>• Does not give permission to remove the parent artifact. For example, a user can remove a schema from a source but cannot delete the data source.</li> </ul>

## Admin

In addition to the **View**, **Meta View**, **Add/Edit**, and **Delete** permissions described above, the **Admin** set includes the **Meta Add/Edit** and **Meta Delete** permissions described below.

Permission	Allows a user to:
<b>Meta Add/Edit</b>	Relates only to an artifact's permissions. A user with Meta Add/Edit can add permissions to a user or group. They cannot remove permissions from any user or group.
<b>Meta Delete</b>	<p>This permission allows a user to:</p> <ul style="list-style-type: none"> <li>• Remove permissions from a user or group.</li> <li>• Delete the parent artifact and its related entities.</li> </ul>

## Default Access Policy Reference

There is a configurable Default Access Policy for several of the Anzo registries. To see and manage the Default Access Policies, go to the Administration application, expand the **User Management** menu, and click **Default Access Policies**.

### Important

Never modify any of the registries. Changing or removing a registry can irreparably damage your Anzo server.

The sections below provide details about each of the registries for which you can configure Default Access Policies:

- [Data Sources Registry](#)
- [Elastic Search Configuration Registry](#)
- [Global Linked Data Configuration](#)
- [Graphmarts Registry](#)
- [Linked Data Set Registry](#)
- [Ontology Registry](#)
- [Orchestration Configuration Registry](#)
- [Query Builder Registry](#)
- [Role and Permissions Registry](#)
- [SDI Registry](#)

### Data Sources Registry

The **Data Sources Registry** is the system graph that stores metadata about all of the **File Store**, **Anzo Data Store**, **Data Source**, and **Schema** artifacts that have been created in Anzo. Since data sources and schemas have a fundamental relationship in that schemas are derived or imported

from data sources, one registry stores metadata about both types of artifacts. The Data Sources Registry access policy is applied by default when a user creates a data source or an Anzo Data Store.

### Default Permissions Configuration

- The **Creator** of a source is assigned the [Admin](#) permission set for that source and the associated schemas. In addition, the Creator of an Anzo Data Store is also assigned the Admin permission set for that data store.
- The **Everyone** role is assigned the [View](#) permission set for a new source and its schemas. The Everyone role is also assigned the View permission set for any Anzo Data Stores.
- The **Creator Default Group** is assigned the [Modify](#) permission set for new source, schema, and Anzo Data Store artifacts.

### Elastic Search Configuration Registry

The **Elastic Search Configuration Registry** is the system graph that stores metadata about all of the **Elasticsearch** connection artifacts in Anzo. This access policy is applied by default when an Elasticsearch connection is created.

### Default Permissions Configuration

- The **Creator** of an Elasticsearch connection is assigned the [Admin](#) permission set for that artifact.
- The **Everyone** role is assigned the [View](#) permission set for that Elasticsearch connection artifact.
- The **Creator Default Group** is assigned the [Modify](#) permission set for that artifact.

### Global Linked Data Configuration

The **Global Linked Data Configuration Registry** is a global policy that applies to all artifacts created in Anzo—unless another Default Access Policy (such as the Data Sources Registry, Graphmarts Registry, or Ontology Registry) applies.

## Example

If a user created a model and the Ontology Registry Default Access Policy was removed or unset, the Global Linked Data Configuration access policy would be applied to that model artifact.

### Default Permissions Configuration

- The **Creator** of an artifact that follows this policy is assigned the [Admin](#) permission set for that artifact.
- The **Creator Default Group** is assigned the [Modify](#) permission set for that artifact.

### Graphmarts Registry

The **Graphmarts Registry** is a system graph that stores metadata about all of the **Graphmart** artifacts in Anzo. All graphmarts inherit permissions from the Graphmarts Registry Default Access Policy. In addition, since data layers and steps are created in the context of a graphmart, they inherit permissions from the graphmart by default.

### Default Permissions Configuration

- The **Creator** of a graphmart is assigned the [Admin](#) permission set for that artifact.
- The **Everyone** role is assigned the [View](#) permission set for that graphmart.
- The **Creator Default Group** is assigned the [Modify](#) permission set for the graphmart.

### Linked Data Set Registry

The **Linked Data Set Registry** is a system graph that stores metadata about all of the linked data sets, notably the File-Based Linked Data Sets (FLDS) that are listed in the Datasets catalog. This includes datasets that are generated from unstructured pipelines as well as datasets that are created by users, such as empty datasets, dataset from Export Steps, and Existing RDF imports directly to the Datasets catalog.

## Default Permissions Configuration

FLDS artifacts inherit from the workflow that created it. If raw RDF files are imported to the catalog or an empty dataset is created, the Linked Data Set Registry Default Access policy is applied to the resulting FLDS artifact.

## Ontology Registry

The **Ontology Registry** is the system graph that stores metadata about all of the model artifacts in Anzo. This access policy is applied by default if a model is imported or manually created by a user. When a model is generated from an unstructured pipeline or the automated Direct Data Load workflow, however, the model inherits the permissions from the related data source.

### Default Permissions Configuration

- The **Creator** of a model is assigned the [Admin](#) permission set for that artifact.
- The **Everyone** role is assigned the [View](#) permission set for that model.
- The **Creator Default Group** is assigned the [Modify](#) permission set for that artifact.

## Orchestration Configuration Registry

The **Orchestration Configuration Registry** is a system graph that stores metadata about workflows. This access policy is applied by default when a workflow is created.

### Default Permissions Configuration

- The **Anzo Administrator** is assigned the [Admin](#) permission set for the artifact.
- The **Creator** of a workflow that follows this policy is assigned the [Admin](#) permission set for that artifact.
- The **Creator Default Group** is assigned the [Modify](#) permission set for that artifact.

## Query Builder Registry

The **Query Builder Registry** is a system graph that stores metadata about saved Query Builder queries. This access policy is applied by default when a new query is saved.

## Default Permissions Configuration

The user who saves a query is assigned the [Admin](#) permission set. By default, saved queries are unique to each creator, and other users do not see the creator's queries.

## Role and Permissions Registry

The **Role and Permissions Registry** is a system graph that stores metadata about roles and permissions. Roles are not treated like other artifacts in Anzo. Unlike a data source, model, or graphmart artifact, for example, the permissions for a single role or subset of roles cannot be configured separately. Access to create and edit roles is controlled by the **Manage Users, Groups, and Roles** permission. For more information, see [Role Permissions and Registries](#).

## Default Permissions Configuration

- The **System Administrator** is assigned the [Admin](#) permission set for all role artifacts.
- The **Everyone** role is assigned the [View](#) permission set for all role artifacts.
- A member of a role that is assigned the **Manage Users, Groups, and Roles** permission has the [Admin](#) permission set for all role artifacts.

## SDI Registry

The **SDI Registry** is a legacy system graph that stored metadata about the mapping, pipeline, and job artifacts that were manually created by a user.

## Default Permissions Configuration

- The **Creator** of a mapping, pipeline, or job is assigned the [Admin](#) permission set for that artifact.
- The **Everyone** role is assigned the [View](#) permission set for the new artifact.
- The **Creator Default Group** is assigned the [Modify](#) permission set for that artifact.



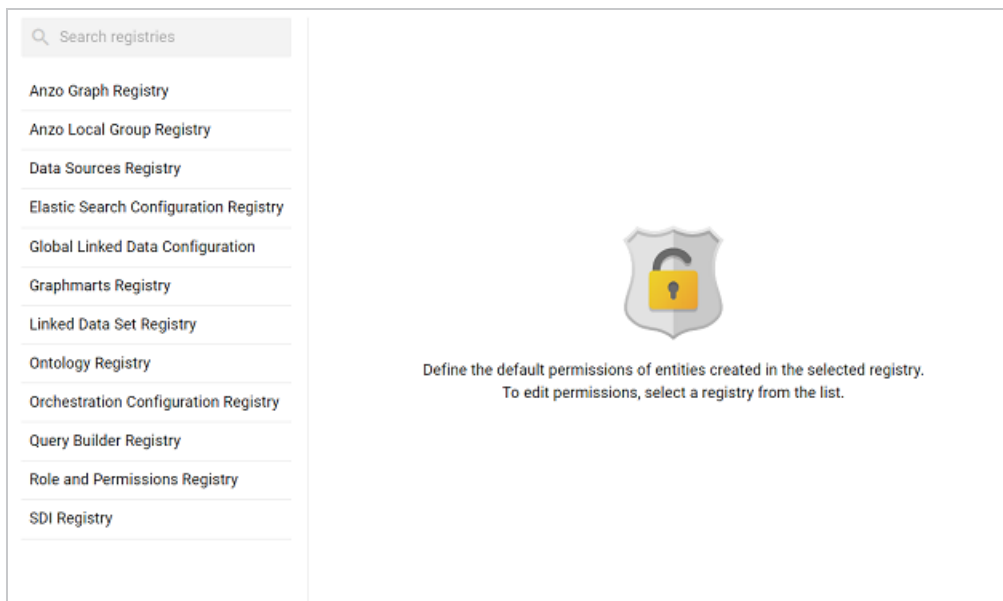
# Configuring Default Access Policies

Follow the instructions below to change the default access policy for a registry.

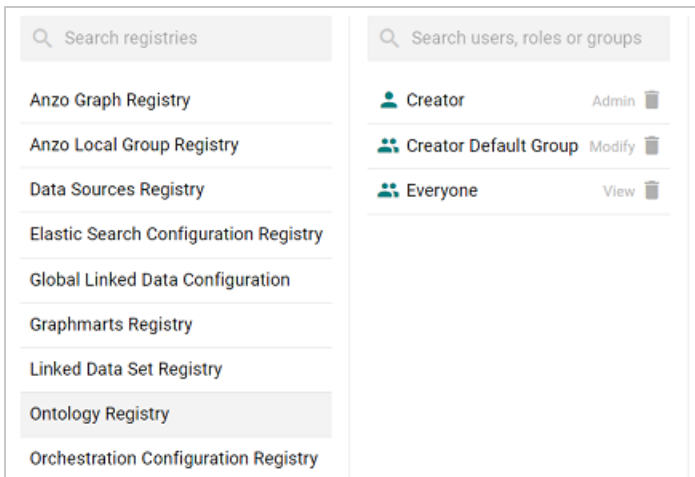
## Important

Changing default access control policies does not change permissions on any existing artifacts. The changes affect only new artifacts that are created after the change.

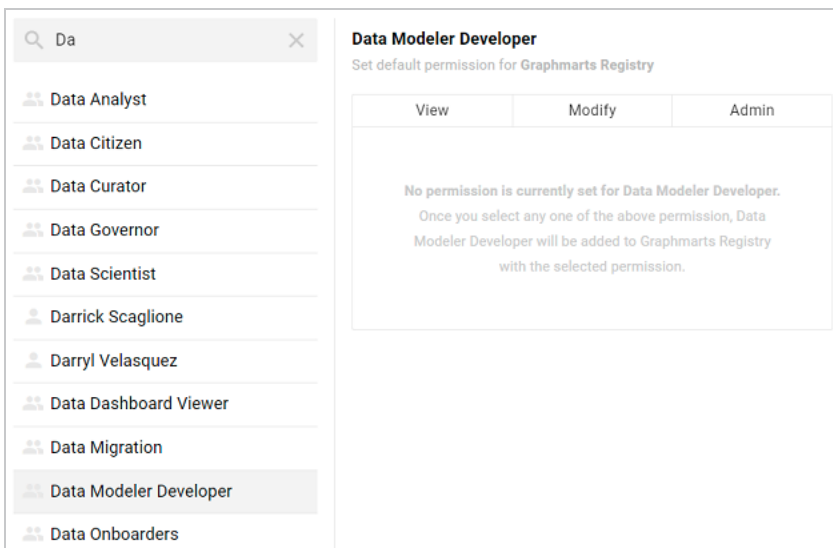
1. In the Administration application, expand the **User Management** menu and click **Default Access Policies**. The Default Access Policies screen is displayed.



2. On the left side of the screen, select the access policy that you want to configure. The current configuration for that policy is shown on the right side of the screen. For example, the image below shows the Ontology Registry. The model creator has **Admin** permissions, the Everyone role has **View** permissions, and the Creator Default Group has **Modify** permissions.



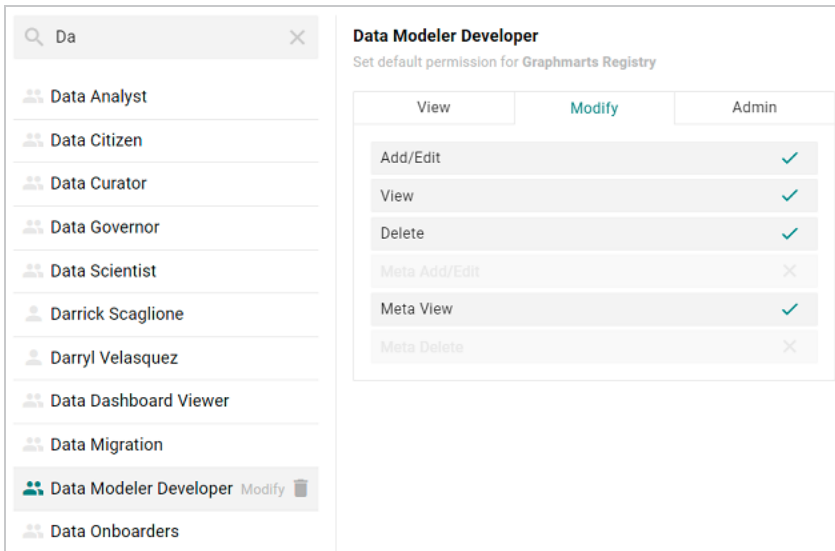
- To change a configured user or group, select a name in the list to view the permissions on the right side of the screen. To add a user or group, type a term in the **Search** field. Then select a name in the result list to view the permissions details. For example, the image below shows the search results for additional groups and selects the Data Modeler Developer group:



## Note

Though Anzo is flexible and allows you to assign default access policies to roles, the recommendation is to control access to artifacts in a registry with users and groups. For more information, see [User Management Concepts](#).

- On the right side of the screen, click the tab for the predefined permission set that you want to assign to the selected user or group. For information about the permission sets, see [Default Access Policy Permissions Reference](#) above. For example, the image below assigns the **Modify** permission set to the Data Modeler Developer group.



### Tip

To clear permissions for a user or group, click the trashcan icon (🗑️) next to the user, role, or group name.

- To configure additional users or groups, select the name and then repeat the step above to apply a permission set. Changes to access control policies are automatically saved.

# Monitoring and Diagnostics

The topics in this section provide information about monitoring events and managing Anzo and AnzoGraph diagnostic files.

## Tip

For information about enabling the System Monitor service, see [Enabling the System Monitor Service](#). For information about view the current stack, see [Viewing the Current Stack in a Browser](#).

### In this section:

<a href="#">Managing Anzo Logging</a> .....	253
<a href="#">Retrieving AnzoGraph Diagnostic Files</a> .....	277
<a href="#">Monitoring AnzoGraph Statistics</a> .....	280
<a href="#">System Query Audit</a> .....	287

# Managing Anzo Logging

The topics in this section provide general information about logging in Anzo, instructions for adding logging for new components, changing the level or type of information that is logged, and reviewing log files. This section also provides guidance on enabling the recommended log packages.

**In this section:**

Logging Concepts and Configuration .....	254
Adding the Recommended Log Packages .....	261
Viewing Log Files .....	273

# Logging Concepts and Configuration

This topic provides an introduction to Anzo logging concepts, an overview of the Logging user interface, and information about the type of logging that is enabled by default. It also gives a high-level overview about turning on additional logging, adjusting the level of information that is logged, and reviewing log files.

- [Logging Concepts](#)
- [Default Logging Configuration](#)
- [Adding Log Packages](#)
- [Log Level Definitions](#)

## Logging Concepts

In order to give users granular control and flexibility over the type and breadth of information that is captured, the concept of **Log Packages** is integral to logging in Anzo. A Log Package is a listener for events that are related to a particular Semantic Service or component, such as core system, LDAP, Anzo Unstructured, or AnzoGraph events. To give users flexibility over the depth of information that is logged, each Log Package can be configured to capture events at a certain **Log Levels**, from all events to fatal events only.

## Default Logging Configuration

Logging is managed in the Administration application. To view the Log Packages that are enabled for your server, expand the **Monitoring & Diagnostics** menu in the Administration application and click **Logging**. Then click the **Log Levels** tab to show the enabled Log Packages and their Log Level configuration. For example, the image below shows the default configuration for a new installation:

Log Files	Log Levels
Configure the log level of a package or add an additional package to log. <a href="#">Edit</a>	
AccessAudit	INFO
ActivityAudit	INFO
AuditLog	ERROR
com.cambridgesemantics	ERROR
InstallUpdateLog	INFO
org.apache.directory	OFF
org.openanzo	ERROR
org.openanzo.client.registry.RegistryManifestLoader	INFO
org.openanzo.combus.endpoint.BaseServiceListener	ERROR
org.openanzo.osgi.bootstrap.BootstrapActivator	INFO
org.pac4j.http.client.direct.DirectBasicAuthClient	OFF
org.pac4j.http.client.direct.HeaderClient	OFF
QueryAudit	INFO
SystemAudit	INFO
TimingStack	ERROR
UserAudit	INFO

## Default Log Packages

The table below describes Log Packages that are enabled by default as well as their default Log Level. Log Levels are defined in [Log Level Definitions](#) below.

Package	Level	Description
<b>AccessAudit</b>	Info	Listener for access audit events such as user login attempts.
<b>ActivityAudit</b>	Info	Listener for activity audit events.
<b>AuditLog</b>	Error	Logger for audit events when the appropriate packages are enabled. For more information, see <a href="#">Enabling the Audit Logs</a> .

Package	Level	Description
com.cambridgesemantics	Error	Like the org.openanzo package, this base package listens for core system events. Changing the Log Level of this package affects logs across Anzo components and services.
InstallUpdateLog	Info	Listener for installation and upgrade events. Captures information about bundle imports and updates.
org.apache.directory	Off	Listener for events related to the underlying internal LDAP server. <b>Do not modify the Log Level for this package.</b>
org.openanzo	Error	Like the com.cambridgesemantics package, this base package listens for core system events. Changing the Log Level of this package affects logs across Anzo components and services.
org.openanzo.client.registry.RegistryManifestLoader	Info	Listener for installation and upgrade events. Captures



Package	Level	Description
		information about bundle imports and updates.
<b>org.openanzo.combus.endpoint.BaseServiceListener</b>	Error	Core server listener for requests sent from clients to the server.
<b>org.openanzo.osgi.bootstrap.BootstrapActivator</b>	Info	Listener for installation and upgrade events. Captures information about bundle imports and updates.
<b>org.openanzo.services.PublicLog</b>	Off	Listener for internal Anzo events. <b>Do not modify the Log Level for this package.</b>
<b>org.pac4j.http.client.direct.DirectBasicAuthClient</b>	Off	Low-level listener for user login events.
<b>org.pac4j.http.client.direct.HeaderClient</b>	Off	Low-level listener for user login events.
<b>QueryAudit</b>	Info	Listener for query audit events.
<b>SystemAudit</b>	Info	Listener for system audit events such as changes to bundle properties.
<b>TimingStack</b>	Error	Listener for events related to internal system journal


Package	Level	Description
		queries.
UserAudit	Info	Listener for user administration related events, such as changes to roles.

## Adding Log Packages

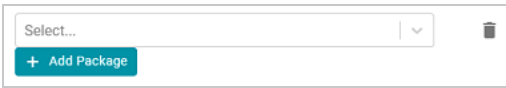
### Tip

For guidance on adding the recommended Log Packages, see [Adding the Recommended Log Packages](#).

To enable additional Log Packages, click the **Edit** button on the Log Levels screen.

Log Files	Log Levels
Configure the log level of a package or add an additional package to log. 	
AccessAudit	INFO
ActivityAudit	INFO
AuditLog	ERROR
com.cambridgesemantics	ERROR
InstallUpdateLog	INFO
org.apache.directory	OFF
org.openanzo	ERROR
org.openanzo.client.registry.RegistryManifestLoader	INFO
org.openanzo.combus.endpoint.BaseServiceListener	ERROR
org.openanzo.osgi.bootstrap.BootstrapActivator	INFO
org.pac4j.http.client.direct.DirectBasicAuthClient	OFF
org.pac4j.http.client.direct.HeaderClient	OFF
QueryAudit	INFO
SystemAudit	INFO
TimingStack	ERROR
UserAudit	INFO

Then click **Add Package** at the bottom of the screen.



Clicking the **Select** field opens the package drop-down list. You can browse through the options, or you can start typing a keyword to search for a package. Click a package to add it to the list of packages on the Edit Log Packages screen. Adjust the Log Level as needed and then click **Save** to save the change. See [Log Level Definitions](#) below for more information about Log Levels.

## Log Level Definitions

This section defines the Log Levels that are available to apply to a Log Package:

- **Off**: Turns logging off for the Log Package.
- **Debug**: Logs fine-grained error messages that are intended to help debug a problem with an application or the server.
- **Trace**: Logs finer-grained error information than Debug.
- **Info**: The highest level of logging. The Log Package captures all events or queries.
- **Warn**: Logs information about potentially problematic situations.
- **Error**: Logs errors that usually allow the application to continue running.
- **Fatal**: Logs severe errors that prevent the application from running.

To change the Log Level for a package, click the **Log Level** field for the Log Package that you want to change and select a level from the drop-down list. Click **Save** when you are finished making changes.

### Edit Log Packages

org.openanzo.client.registry.RegistryManifestLoader	Info	
org.pac4j.http.client.direct.HeaderClient	Off	
AuditLog	Error	
InstallUpdateLog	Info	
org.apache.directory	Off	
TimingStack	Error	
org.openanzo	Error	
org.pac4j.http.client.direct.DirectBasicAuthClient	Off	
com.cambridgesemantics	Error	
org.openanzo.osgi.bootstrap.BootstrapActivator	Info	
org.openanzo.combus.endpoint.BaseServiceListener	Error	
org.openanzo.services.PublicLog	Off	

[+ Add Package](#)

CANCEL [SAVE](#)

## Adding the Recommended Log Packages

The Log Packages that are enabled by default cover the core Anzo server operations and services to ensure that diagnostics are generated when errors occur. Anzo includes several additional Log Packages, however, that are disabled by default but can be configured to provide valuable information for auditing purposes, such as information about user logins, user administration events, and AnzoGraph queries. This section describes the packages that Cambridge Semantics recommends that you enable and provides information about reading the resulting log files.

- [Enabling AnzoGraph Query Logs](#)
- [Enabling the Audit Logs](#)

### Enabling AnzoGraph Query Logs

The GqeQueries Log Package listens for AnzoGraph events like connection errors, restarts, and successful and unsuccessful queries. GqeQueries is Off by default but can be enabled to monitor and log all of the queries that are sent to AnzoGraph by users through dashboards, the Query Builder, data layers, etc., or sent by Anzo, such as when requesting the total number of statements in a graph.

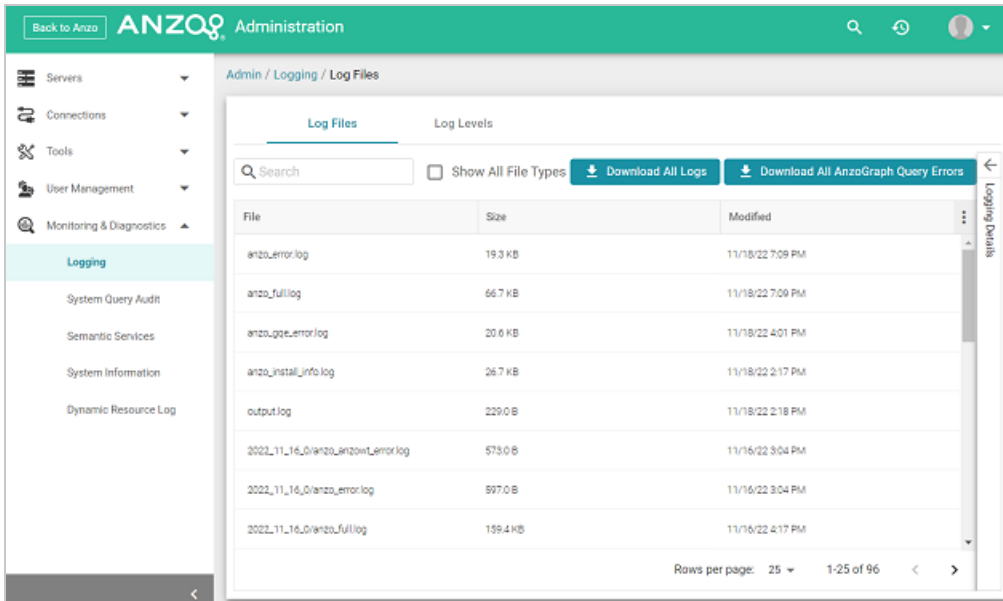
#### Note

Though GqeQueries is Off by default, AnzoGraph query errors are still captured automatically in the `<install_path>/Server/logs/gqe/queriesError` directory, and connection-related errors are captured in `anzo_gqe_error.log`.

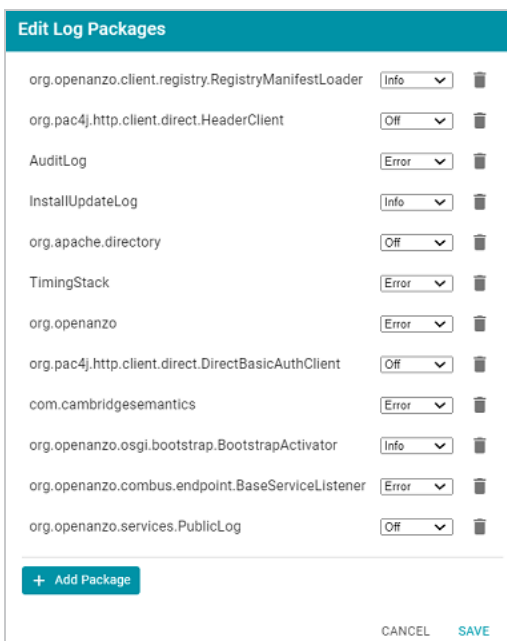
### Enabling the GqeQueries Log Package

Follow the steps below to enable the GqeQueries package.

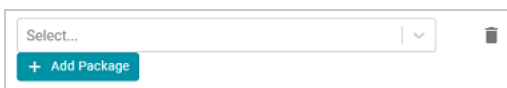
1. In the Administration application, expand the **Monitoring & Diagnostics** menu and select **Logging**. The Log Files tab is displayed. For example:



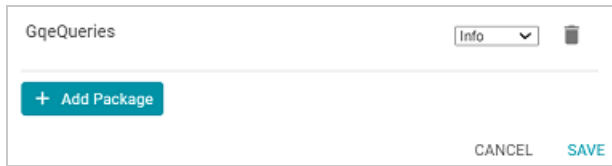
2. Click the **Log Levels** tab. Then click the Edit button at the top of the screen. The Edit Log Packages dialog box is displayed.



3. Click **Add Package** at the bottom of the screen. The Select field is displayed:



4. Click the **Select** field and type **GqeQueries**. Then press **Enter** to add GqeQueries to the list of Log Packages. The package is added to the list with the default Log Level of **Off**.
5. Click the Log Level drop-down list and select **Info**. Then click **Save** to save the change.



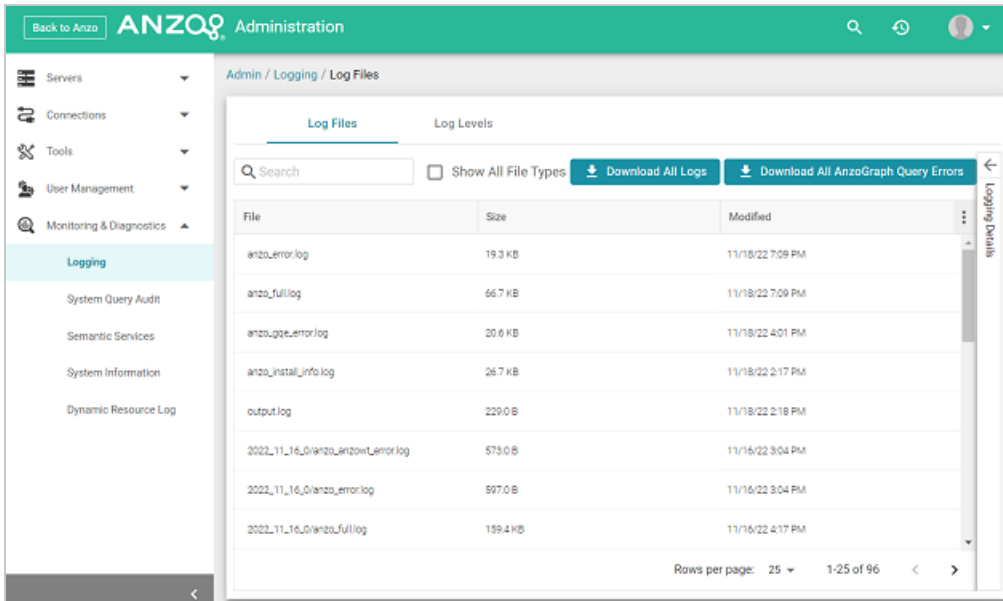
The screenshot shows a configuration dialog box for a log package named "GqeQueries". At the top left, the package name "GqeQueries" is displayed. To its right is a dropdown menu currently set to "Info", with a trash icon to its right. Below this is a blue button with a plus sign and the text "+ Add Package". At the bottom right of the dialog are two buttons: "CANCEL" and "SAVE".

The GqeQueries Log Package is now enabled and will start to log the events described above. The log messages for successful queries are captured in a new **anzo\_gqe\_info.log** file as well as in the `<install_path>/Server/logs/gqe/queriesInfo` directory on the server. Details about each request is logged to a separate file in that directory. The `anzo_gqe_info.log` and the files in `logs/gqe/queriesInfo` can be viewed and downloaded from the Administration application.

### Viewing the AnzoGraph Query Logs

Follow the steps below to view the AnzoGraph log files in the application. For information about viewing logs on the server, see [Viewing Logs on the Server](#).

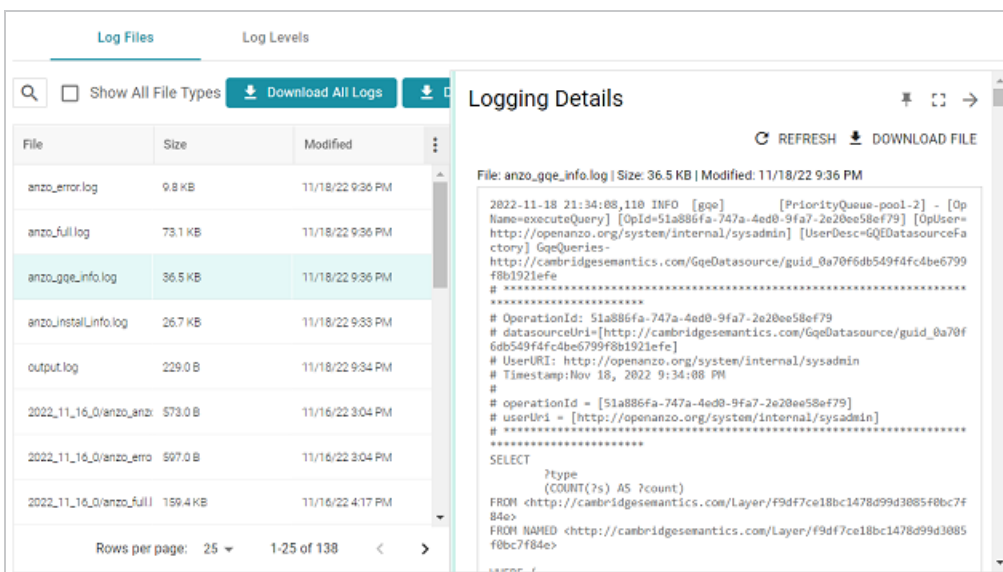
1. In the Administration application, expand the **Monitoring & Diagnostics** menu and select **Logging**. The Log Files tab is displayed. Log Packages that have the Log Level set to **Error** log events to files with the suffix **\_error**. Operational information that is logged by packages that are set to **Info** is captured in files with the suffix **\_info**.



## Note

The current version of **anzo\_gqe\_info.log** is shown toward the top of the list. Earlier versions of that log are prefixed with the name of the `<date>_<part>` subdirectory they are saved in. And individual query files are named as `/gqe/queriesInfo/<operation_ID><epoch_timestamp>`.

2. Select the **anzo\_gqe\_info.log** file. The contents of the file are displayed in the Logging Details section of the screen. For example:





You can expand the details view by clicking the Expand icon (⌵) in the top right corner.

The messages in **anzo\_gqe\_info.log** vary by the query source, such as whether the query originated in a dashboard lens or the Query Builder. In general, GqeQueries Info messages contain the following information:

- Date and time the event was logged. For example, `2021-04-28 01:06:48`.
- The type of message, i.e., the Log Level, such as `INFO`.
- The type of log. For example, `[gqe]`.
- The area of the system or service that processed the event. For example, `[PriorityQueue-pool-2]`.
- The Log Package that was listening for the event, i.e., `GqeQueries`.
- The Data Source URI. For example, `http://cambridgesemantics.com/GqeDatasource/guid_e1f38b640fe04bf8fee71bdf5184bf41`.
- The Operation ID assigned to the query. This value can be used to track the query, such as to find the individual log file in the `logs/gqe/queriesInfo` directory. For example, `OperationId: 7b0op0wbzqeqls2d482xudkez-83`. The corresponding log file is named `query_7b0op0wbzqeqls2d482xudkez-83.log`.
- The User URI for the user who submitted the query. For example, `UserURI: ldap:///cn=Jay.Blue,ou=groups,dc=com`.
- If the query was submitted from the Hi-Res Analytics application, the message also includes details for identifying the dashboard and lens that submitted the request. For example:

```
# ex_requestDashboard = [http://cambridgesemantics.com/354db630-02b6-46b2-82d0-ef4a7543ebca]
# ex_requestSource = [http://cambridgesemantics.com/4a039bdb-bdcb-4117-830b-cb29190ce18f]
# ex_requestSourceId = [com_cambridgesemantics_application_anzoweb_lens_grid_GridLens_7]
```

- The text of the query that was sent by Anzo. Note that the text is the query as rewritten by Anzo and sent to AnzoGraph. It may not be the exact text that was written by the user.
- When a query returns, a result message is also added to anzo\_gqe\_info.log below the query text. The QueryResults message includes the Operation ID (which matches the ID from the query that was sent), and it returns the AnzoGraph and Anzo query execution time as well as the number of results returned. In the following example, the QueryResults message is shown in bold. The first value (**2631**) is the number of milliseconds AnzoGraph spent executing the query. The value in brackets (**[13155]**) is the number of milliseconds Anzo spent executing the query. And the last value (**20**) is the number of results that were returned.

```
2021-04-28 22:53:57,134 INFO [gqe] [PriorityQueue-pool-7] - [OpName=query]
[OpId=8tt1rrc29y31z1ga30srk6t2xx-212]
[OpUser=http://openanzo.org/system/internal/sysadmin]
GqeQueries- QueryResults:2631 [13155]: 20
```

### Note

A QueryResults message is not logged if the query uses the Anzo cache or returns an error.

A complete example message is shown below:

```
2021-04-27 19:54:25,648 INFO [gqe] [PriorityQueue-pool-2] - GqeQueries-
http://cambridgesemantics.com/GqeDatasource/guid_elf38b640fe04bf8fee71bdf5184bf41
# *****
# OperationId: 66ed1f10-5aae-45b0-861c-3a851022d294
# datasourceUri=[http://cambridgesemantics.com/GqeDatasource/guid_
elf38b640fe04bf8fee71bdf5184bf41]
# UserURI: http://openanzo.org/system/internal/sysadmin
# Timestamp:Apr 27, 2021 7:54:25 PM
#
# operationId = [66ed1f10-5aae-45b0-861c-3a851022d294]
# userUri = [http://openanzo.org/system/internal/sysadmin]
# *****
SELECT
  ?type
  (COUNT(?s) AS ?count)
FROM <http://cambridgesemantics.com/Layer/f44db5d106ca4186b953a591e873a5f0>
FROM NAMED <http://cambridgesemantics.com/Layer/f44db5d106ca4186b953a591e873a5f0>
```

```
WHERE {  
    ?s <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> ?type .  
}  
GROUP BY ?type  
2021-04-27 19:54:25,670 INFO [gqe] [PriorityQueue-pool-2] - GqeQueries-  
QueryResults:16 [100]: 11
```

## Enabling the Audit Logs

The Audit Log Packages listen for user- or security-related events such as access attempts and user administration-related events such as modifications to users, groups, and roles. The Audit Log packages are disabled by default but can be enabled to monitor and log the following types of events:

- The inactivity timeout is changed.
- A bundle's properties are changed or a bundle is restarted.
- A user successfully logs in or out or there are failed login attempts.
- A user account is created or deleted or a user's password is changed.
- A user or group is synchronized with the directory server.
- A role is created or deleted.
- A user is added to or removed from a role or group.
- A permission is added to or removed from a role.
- Data access permissions are changed on artifacts.

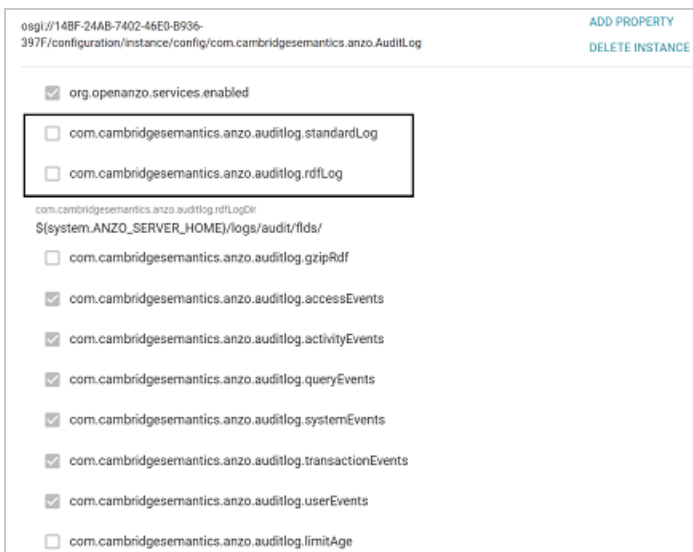
## Enabling the Audit Log Packages

By default, the Audit Log packages (UserAudit, AccessAudit, QueryAudit, ActivityAudit, and SystemAudit) are set to the Log Level **Info**, which means they are configured to capture all audit events. However, logging the audit events are disabled by default in the Anzo Audit Logging Framework service. Follow the instructions below to configure the service to enable audit logging.

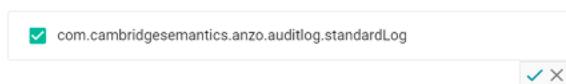
## Important

Before enabling logging, consider the file storage implications. Depending on the number and activity of Anzo users, the logs can grow to several GB in size. Make sure that you have the disk space to retain the logs, and consider limiting the size of the logs. See [Limiting the Age/Size of Audit Logs](#) for information.

1. In the Administration application, expand the **Servers** menu and click **Advanced Configuration**. Click **I understand and accept the risk**.
2. Search for the **Anzo Audit Logging Framework** bundle and view its details.
3. Click the **Services** tab and expand **com.cambridgesemantics.anzo.AuditLog**.
4. Find the **com.cambridgesemantics.anzo.auditlog.standardLog** and **com.cambridgesemantics.anzo.auditlog.rdfLog** properties (shown below).



5. Click the **standardLog** property to make it editable, and then select the checkbox to enable it.



6. Click the checkmark icon (✓) to save the change.

7. If you would also like Anzo to generate RDF files for the audit logs so that the logs can be loaded to a graphmart for analysis, click the **rdfLog** property to make it editable, and then select the checkbox to enable it. Then click the checkmark icon (✓) to save the change.
8. Restart Anzo to apply the configuration changes.

The Audit Log Packages are now enabled and will start to log the events described above. The log messages are captured in **anzo\_full.log** as well as a new file called **anzo\_audit\_info.log**. All Anzo log files are generated in the `<install_path>/Server/logs` directory on the server. Files in that directory can be viewed and downloaded from the Administration application.

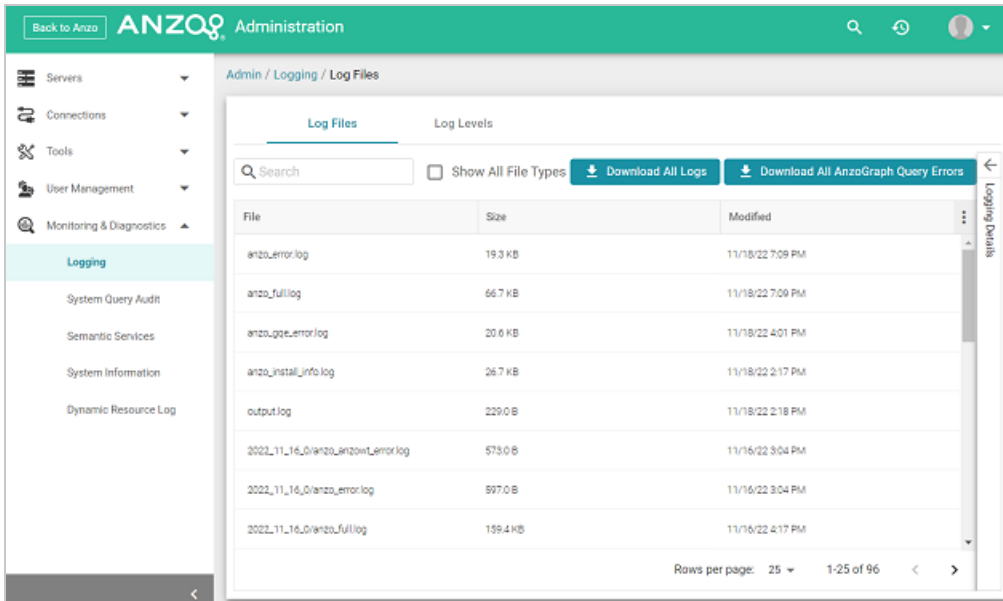
## Viewing the Audit Log

Follow the steps below to view the Audit log file in the application. For information about viewing logs on the server, see [Viewing Logs on the Server](#).

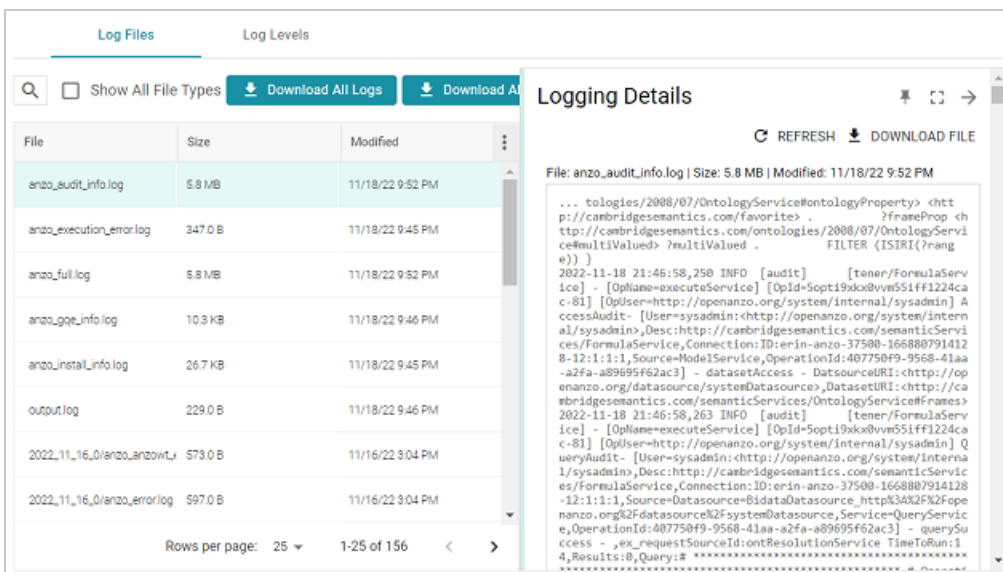
### Tip

You have the option to split the Audit log into separate files based on the type of event that is being logged, such a user event or a query event. See [Separating Audit Logs by Event Type](#) for information. The steps below refer to the default Audit Log where all types of audit events are recorded in a single file.

1. In the Administration application, expand the **Monitoring & Diagnostics** menu and select **Logging**. The Log Files tab is displayed on the Logging screen. Log Packages that have the Log Level set to **Error** log events to files with the suffix **\_error**. Operational information that is logged by packages that are set to **Info** is captured in files with the suffix **\_info**. The current versions of the log files are shown at the top of the list. Earlier versions of the logs are prefixed with the name of the `<date>_<part>` subdirectory they are saved in. For example:



2. Select the **anzo\_audit\_info.log** file. The contents of the file are displayed in the Logging Details section of the screen. For example:



You can expand the details view by clicking the Expand icon (⌵) in the top right corner.

The elements included in each message vary by message type. In general, UserAudit Info messages contain the following information:

- Date and time the event was logged. For example, 2021-04-28 01:06:48.
- The type of message, i.e., the Log Level, such as INFO.

- The type of log. For example, [audit].
- The area of the system or service that processed the event. For example, [UniformSaveService].
- The Log Package that was listening for the event, i.e., UserAudit.
- The message text, such as `User Connected` or `Authentication Failed`.
- The unique Operation ID assigned for the operation. For example, [OpId=518ombnsruyvu8k6pf0a76y4fc-1414].
- The name of the service that performed the operation. For example, [OpName=executeService].
- The user who performed the operation. For example, [OpUser=http://openanzo.org/system/internal/sysadmin].

Below are examples of the types of messages that are logged (line breaks added for readability):

## Successful User Login

```
2021-04-27 16:12:28,754 INFO [audit] [persistent=false#1-1] - UserAudit-
User Connected:sysadmin:<http://openanzo.org/system/internal/sysadmin>,
ConnectionId:ID:anzo-36673-1619539948446-4:1,
RemoteAddress:vm://localhost?broker.persistent=false#0
```

## Failed User Login

```
2021-04-28 01:06:48,341 INFO [audit] [erverThreadPool-3323] -
[OpName=ServerRealm.Authenticate]
[OpId=a876f781-5ddf-424d-8d54-c2ea07c87561]
UserAudit-
Authentication Failed:test,
Message:ErrorCode[3844] User test not found.
```

## Inactivity Timeout Value Changed

```
2021-04-27 19:50:17,316 INFO [audit] [Service Update Queue] -
[OpName=executeService]
[OpId=518ombnsruyvu8k6pf0a76y4fc-1802]
```

```
[OpUser=http://openanzo.org/system/internal/sysadmin]
UserAudit- Inactivity Logout Timeout Changed: Old=-1 New=900000
```

## New Role Created

```
2021-04-27 18:58:38,276 INFO [audit] [r/UniformSaveService] -
[OpName=executeService]
[OpId=518ombnsruyvu8k6pf0a76y4fc-1414]
[OpUser=http://openanzo.org/system/internal/sysadmin]
UserAudit-
Role Created: <http://cambridgesemantics.com/Role/952810ffb74a42f8b502adc422608e64>
```

## Permission Added to a Role

```
2021-04-28 20:41:10,926 INFO [audit] [r/UniformSaveService] -
[OpName=executeService]
[OpId=5q6p7zmp9xn2xujksz4l7pzzl-1808]
[OpUser=http://openanzo.org/system/internal/sysadmin]
UserAudit-
Permission <http://cambridgesemantics.com/permissions/feature/e5c11e5b-afb2-4af0-b1d7-
0e4b620a0378>
added to Role <http://cambridgesemantics.com/Role/952810ffb74a42f8b502adc422608e64>
```



## Viewing Log Files

All Anzo log files are generated in the `<install_path>/Server/logs` directory on the server. Files in that directory can be viewed and downloaded from the Administration application on the **Log Files** tab on the Logging screen.

- [Viewing Logs on the Server](#)
- [Viewing Logs in the Administration Application](#)

### Viewing Logs on the Server

To avoid generating large log files that are difficult to manage (especially for Log Packages set to **Info**), Anzo starts logging to a new version of a file when any of the following events occur:

- A file size reaches 50 MB.
- Log settings are changed.
- Anzo is restarted.

The current, most recent version of a file is stored directly in the `<install_path>/Server/logs` directory. Earlier versions of the files are saved in `<year>_<month>_<day>_<part>` subdirectories in `Server/logs`. For example:

```
logs
├── 2021_04_27_0
│   ├── anzo_audit_info.log
│   ├── anzo_error.log
│   ├── anzo_full.log
│   ├── anzo_gqe_info.log
│   └── anzo_internal_error.log
├── 2021_04_27_1
│   ├── anzo_audit_info.log
│   ├── anzo_datasource_error.log
│   ├── anzo_error.log
│   ├── anzo_full.log
│   ├── anzo_gqe_error.log
│   ├── anzo_gqe_info.log
│   ├── anzo_install_error.log
│   └── anzo_install_info.log
```

```

├─ 2021_04_28_0
|   ├─ anzo_audit_info.log
|   ├─ anzo_error.log
|   ├─ anzo_full.log
|   ├─ anzo_gqe_info.log
|   ├─ anzo_install_error.log
|   └─ anzo_install_info.log
├─ 2021_04_28_1
|   ├─ anzo_error.log
|   └─ anzo_full.log
├─ 2021_04_28_2
|   ├─ anzo_audit_info.log
|   ├─ anzo_error.log
|   └─ anzo_full.log
├─ anzo_audit_info.log
├─ anzo_error.log
├─ anzo_full.log
├─ anzo_gqe_info.log
├─ anzo_install_error.log
├─ anzo_install_info.log
├─ anzo_internal_error.log

```

AnzoGraph query log files are stored in a directory named **gqe** in the `<install_path>/Server/logs` directory. By default all queries that are unsuccessful are captured in the **queriesError** directory. When the AnzoGraph queries Log Package is enabled, successful queries are also captured in the **queriesInfo** directory. For example:

```

logs
├─ gqe
|   ├─ queriesError
|   └─ queriesInfo
|       ├─ query_1a5548ac-6404-4321-b36b-d5eda4ca45a7_1619540406734.log
|       ├─ query_1a5548ac-6404-4321-b36b-d5eda4ca45a7.log
|       ├─ query_292f102e-d222-4261-a069-d7d0c8ceb823_1619469563646.log
|       ├─ query_292f102e-d222-4261-a069-d7d0c8ceb823.log
|       ├─ query_2ddc5f96-758d-4133-80d7-21de5f23134f_1619627154151.log
|       ├─ query_2ddc5f96-758d-4133-80d7-21de5f23134f.log
|       └─ query_518ombnsruyvu8k6pf0a76y4fc-674.log

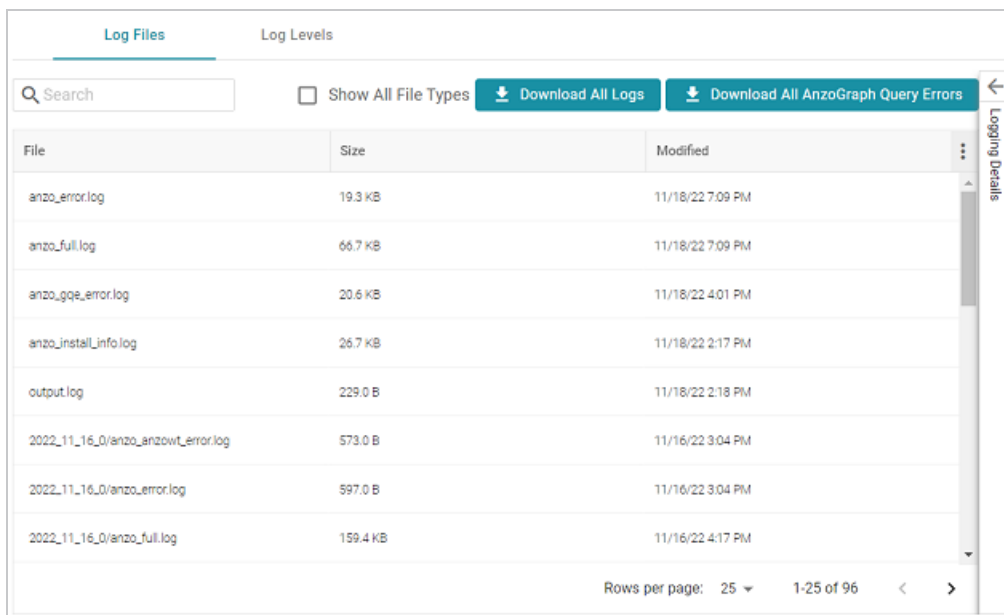
```

## Tip

For instructions on enabling the AnzoGraph query Log Package, see [Enabling AnzoGraph Query Logs](#).

## Viewing Logs in the Administration Application

Logs in the `<install_path>/Server/logs` directory can be viewed and downloaded from the Administration application on the **Log Files** tab on the Logging screen. The Log Files tab lists the logs that are available to view. For example:



File	Size	Modified
anzo_error.log	19.3 KB	11/18/22 7:09 PM
anzo_full.log	66.7 KB	11/18/22 7:09 PM
anzo_gqe_error.log	20.6 KB	11/18/22 4:01 PM
anzo_install_info.log	26.7 KB	11/18/22 2:17 PM
output.log	229.0 B	11/18/22 2:18 PM
2022_11_16_0/anzo_anzow_Error.log	573.0 B	11/16/22 3:04 PM
2022_11_16_0/anzo_error.log	597.0 B	11/16/22 3:04 PM
2022_11_16_0/anzo_full.log	159.4 KB	11/16/22 4:17 PM

Log Packages that have the Log Level set to **Error** log events to files with the suffix **\_error**.

Operational information that is logged by packages that are set to **Info** is captured in files with the suffix **\_info**. The current versions the log files are shown at the top of the list. Earlier versions of the logs are prefixed with the name of the `<date>_<part>` subdirectory they are saved in.

Selecting a log from the list displays its contents in the Logging Details section of the screen. For example:

The screenshot shows a web interface for viewing log files. At the top, there are tabs for 'Log Files' and 'Log Levels'. Below the tabs, there is a search icon, a checkbox for 'Show All File Types', and a 'Download All Logs' button. The main area is divided into two sections: a table of log files on the left and a 'Logging Details' panel on the right.

File	Size	Modified
anzo_errorlog	19.3 KB	11/18/22 7:09 PM
anzo_full.log	66.7 KB	11/18/22 7:09 PM
anzo_gqe_errorlog	20.6 KB	11/18/22 4:01 PM
anzo_install_info.log	26.7 KB	11/18/22 2:17 PM
output.log	229.0 B	11/18/22 2:18 PM
2022_11_16_0/anzo_an	573.0 B	11/16/22 3:04 PM
2022_11_16_0/anzo_en	597.0 B	11/16/22 3:04 PM
2022_11_16_0/anzo_full	159.4 KB	11/16/22 4:17 PM

The 'Logging Details' panel shows the selected file 'anzo\_full.log' with a size of 66.7 KB and a modification time of 11/18/22 7:09 PM. It contains a list of log entries, including timestamps, log levels (INFO), and messages related to registry initialization and bundle processing.

The following options are available for viewing and downloading log files:

- To download a .zip file that contains all of the listed logs, click the **Download All Logs** button at the top of the screen.
- To download just the query error logs for AnzoGraph, click the **Download All AnzoGraph Query Errors** button at the top of the screen.
- To re-load the display with the latest version of the selected file, click the **Refresh** button at the top of the details.
- To download the file so you can view it in another editor, click **Download File** at the top of the details.

# Retrieving AnzoGraph Diagnostic Files

When Cambridge Semantics Support requests AnzoGraph diagnostic files for troubleshooting an issue, you can quickly retrieve the files from the Diagnostics tab on the AnzoGraph page in the Anzo Administration application. This topic provides information about the AnzoGraph diagnostics and instructions for retrieving the files.

## Note

If the Administration application is unavailable, you can retrieve the diagnostic files with the AnzoGraph system manager. See [Taking AnzoGraph X-Rays from the Command Line](#) for instructions.

## Diagnostic File Details

There are two types of AnzoGraph diagnostic files:

- **XRay:** X Rays are generated on-demand. If you encounter an error and the database remains running, you generate an XRay to produce the diagnostic files.
- **Crash Dump:** If you encounter an error that crashes the database, AnzoGraph automatically generates a crash dump that contains diagnostic information about the crash.

Xrays and crash dumps are valuable tools that enable Cambridge Semantics to diagnose and fix issues without access or any other visibility into a customer's data or database system. They can also be used to report on overall and detailed system performance, resulting in improved query performance for future releases of AnzoGraph.

Xrays and crash dumps harvest the diagnostic data that is stored in AnzoGraph's system tables. They include information such as:

- A low level, de-identified log of the requests that were sent to the database.
- Statistics like query operation step execution times, number of rows processed, and amount of memory used.
- Detailed but de-identified trace information for errors that were encountered.

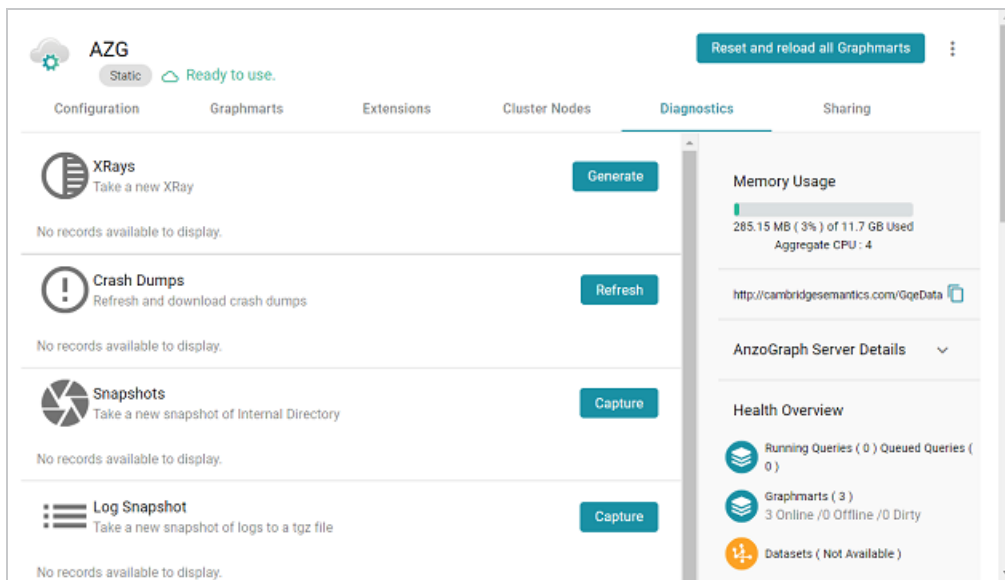
- Configuration information such as the number of nodes in the cluster and AnzoGraph system settings values.

Xrays and crash dumps are designed to be anonymous and can be safely shared with Cambridge Semantics Support. They do NOT capture user information or any of the data that is loaded into memory by a user, nor do they expose details that could be used to reveal the nature of the data being queried.

## Retrieving the Files

Follow the instructions below to download an xray or crash dump to send to Cambridge Semantics Support.

1. In the Administration application, expand the **Connections** menu and select **AnzoGraph**. The AnzoGraph screen is displayed and lists the connected AnzoGraph instances.
2. Click the name of the AnzoGraph instance for which you want to download an xray or crash dump. Anzo displays the Graphmarts screen for the instance.
3. Click the **Diagnostics** tab. Anzo displays the available options. For example:



4. If you want to retrieve an xray, click the **Generate** button for Xrays. Anzo creates the xray and produces a tarball with a .xray extension. For example:

The screenshot shows a user interface for XRay. At the top left is the XRay logo with the text 'XRay' and 'Take a new XRay' below it. To the right is a blue 'Generate' button. Below the logo, it says 'Last Generated : Nov 18, 2022 4:55 PM' and 'xray\_202211182155.xray'. On the right side, there is a link 'View All XRays' with a dropdown arrow.

Click the xray file name to download the tarball to your computer for sending to Cambridge Semantics.

### Note

The files in the tarball are compressed. Do not compress the .xray file before sending it to Cambridge Semantics.

5. If you want to retrieve a crash dump, click the **Refresh** button next to Crash Dumps to refresh the list of available crash dump files. Click the file name that you want to download. Anzo downloads the file to your computer.

# Monitoring AnzoGraph Statistics

This topic provides information about viewing AnzoGraph's memory usage, query performance statistics, and network bandwidth.

- [Viewing Current Memory Usage](#)
- [Reviewing Query Performance Statistics](#)
- [Evaluating Network Performance on Clusters](#)

## Viewing Current Memory Usage

Follow the steps below to view AnzoGraph's current memory usage.

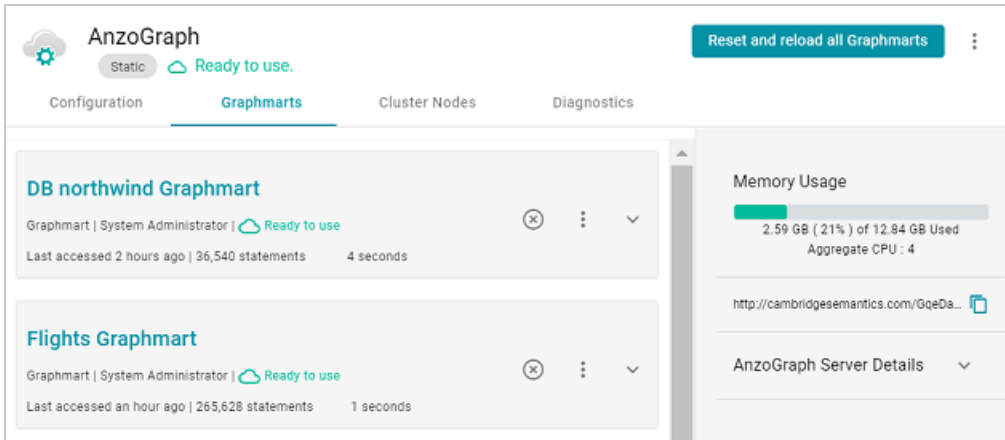
### Note

The memory usage metric is only an estimate of the actual usage. It does not differentiate between the memory that is in use and the memory that is reserved by AnzoGraph but not in use. The value shows the amount of memory that is in use plus the amount that is reserved for future use. Future versions of AnzoGraph will have more exact, granular memory usage reporting.

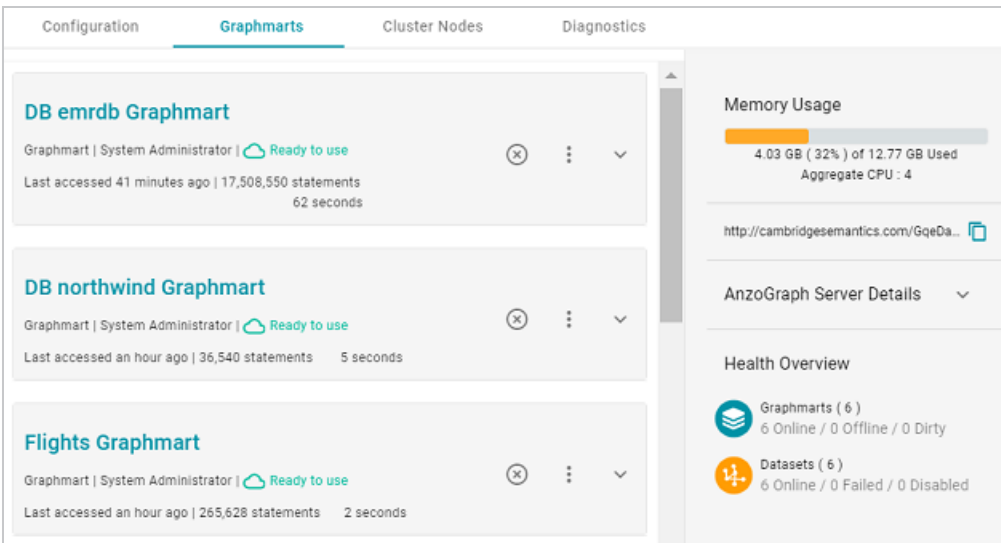
1. In the Administration application, expand the **Connections** menu and select **AnzoGraph**. The AnzoGraph screen is displayed and lists the connected AnzoGraph instances.
2. Click the name of the instance that you want to evaluate. Anzo displays the Graphmarts screen for that instance. The memory usage details are displayed in the top right corner on all of the tabs. For example, the test instance below shows that 21% of the available memory is



in use:



Ideally, the data at rest should use only 25%-30% of the available memory because query execution and intermediate result storage can temporarily consume a very large amount of RAM, especially when multiple users run queries concurrently. When memory usage increases so that the data uses more than 25% - 30% of the available memory, the status bar changes color to orange as a warning. For example:

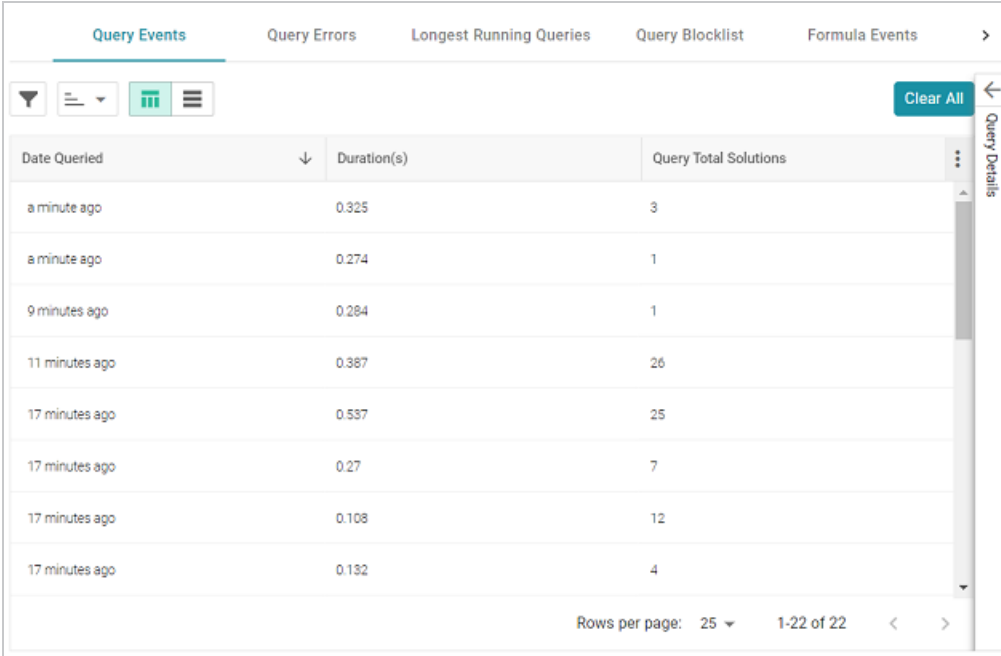


If memory usage for the data at rest remains above 50%, Cambridge Semantics recommends that you increase the amount of RAM available. For more information about memory usage, see [Sizing Guidelines for In-Memory Storage](#) in the Deployment Guide.

# Reviewing Query Performance Statistics

The System Query Audit log provides details about all system events. Users can filter the log to view query execution times for AnzoGraph queries. Follow the steps below to filter and view the log.

1. In the Administration application, expand the **Monitoring & Diagnostics** menu and select **System Query Audit**. Anzo displays the Query Events log. For example:



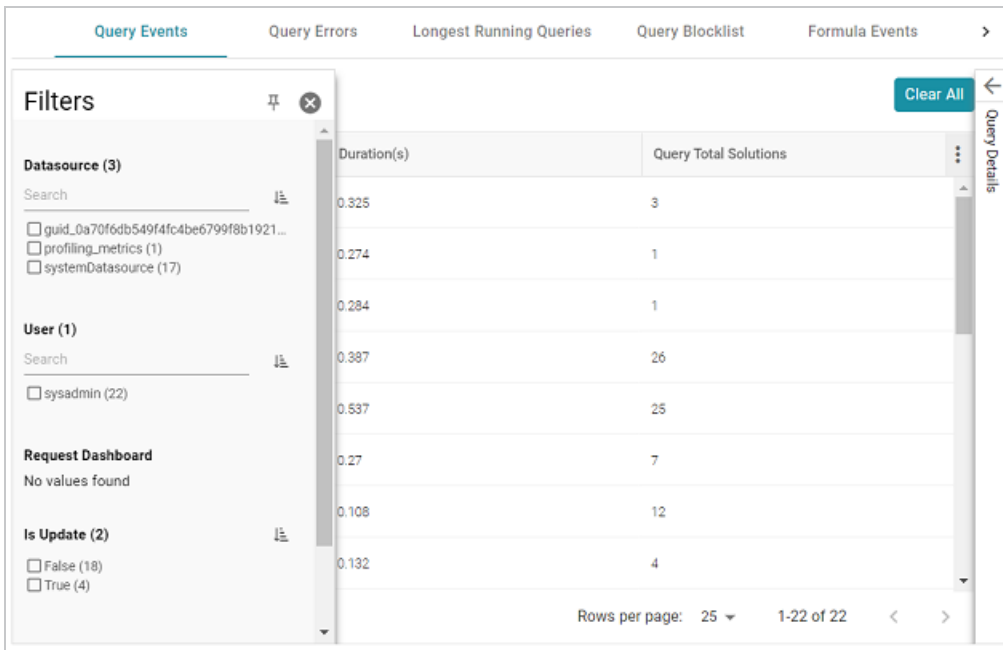
The screenshot shows the 'Query Events' tab in the Anzo Administration application. The interface includes a navigation bar with tabs for 'Query Events', 'Query Errors', 'Longest Running Queries', 'Query Blocklist', and 'Formula Events'. Below the navigation bar are filter and view controls, including a filter icon, a view icon, and a 'Clear All' button. The main content is a table with the following data:

Date Queried	Duration(s)	Query Total Solutions
a minute ago	0.325	3
a minute ago	0.274	1
9 minutes ago	0.284	1
11 minutes ago	0.387	26
17 minutes ago	0.537	25
17 minutes ago	0.27	7
17 minutes ago	0.108	12
17 minutes ago	0.132	4

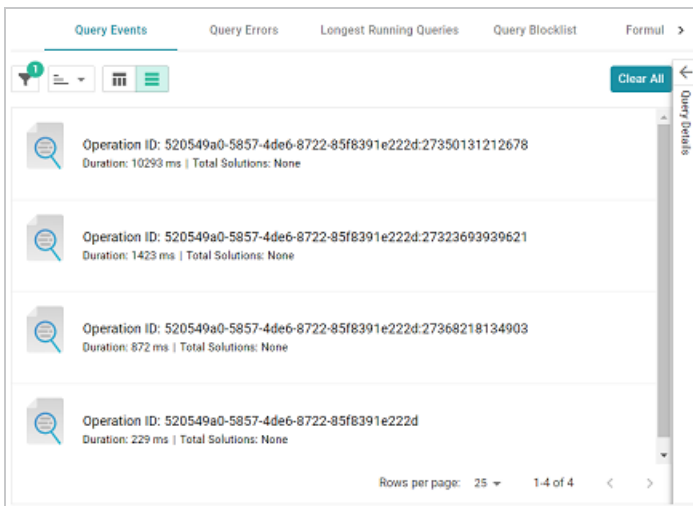
At the bottom of the table, there is a pagination control showing 'Rows per page: 25' and '1-22 of 22'.

By default, the log shows an overview of all query events for all data sources. The table lists the date queried, the duration in milliseconds, and total number of solutions returned for each query event. You can select an event in the table to view details about that event, such as the target data source and query text, on the right side of the screen.

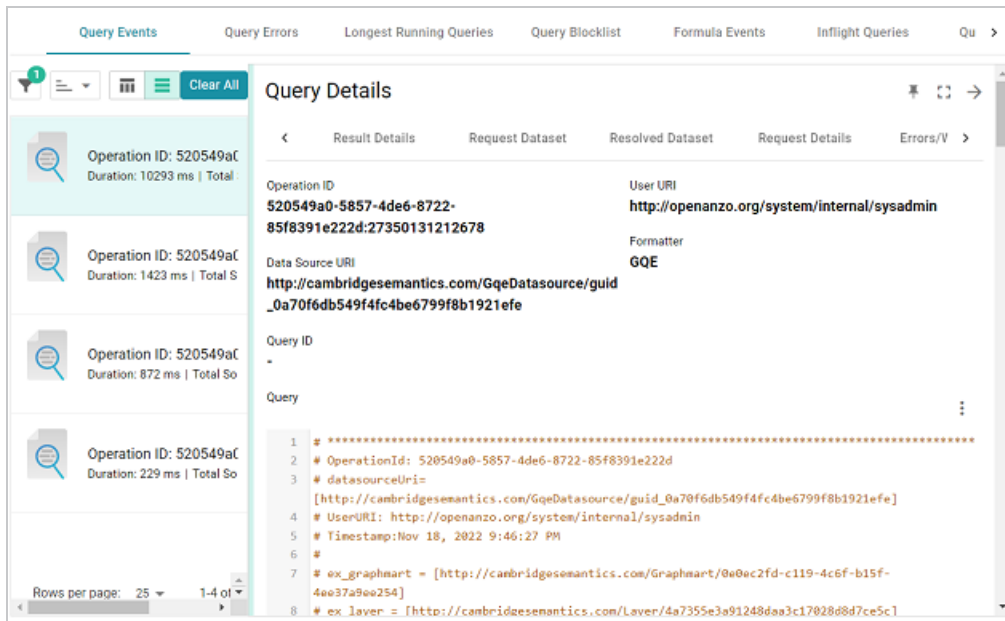
2. To filter the events to display only AnzoGraph queries, open the Filters panel by clicking the filter icon (🔍) in the top left corner of the screen. For example:



- In the Filters panel under **Datasource**, select the checkbox for the AnzoGraph data source. Typically the name starts with **guid\_**. The table of events is filtered to display AnzoGraph events. At the top of the screen, you can choose between a table view (☰) or list view (☰), and you can sort by date, duration, or total solutions. For example, the image below shows a list view of AnzoGraph query events sorted by duration:



- Select any query in the list to view the query details on the right side of the screen. For example:



To view more details about the query event, click the additional tabs in the Query Details panel.

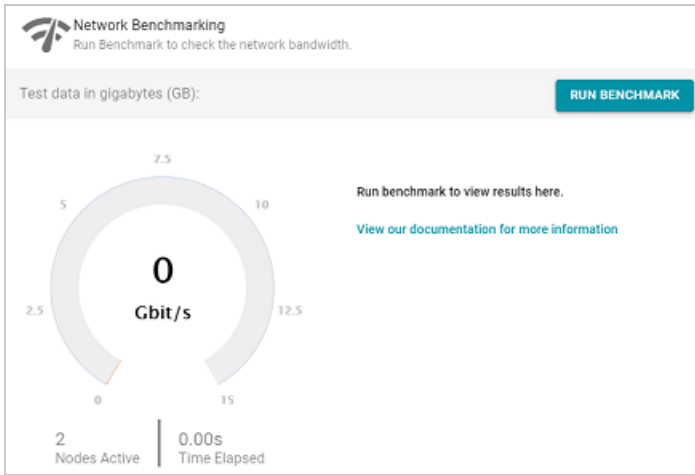
## Evaluating Network Performance on Clusters

The AnzoGraph Diagnostics screen provides a network benchmark that you can run to evaluate the network bandwidth of a cluster. Follow the steps below to run the benchmark.

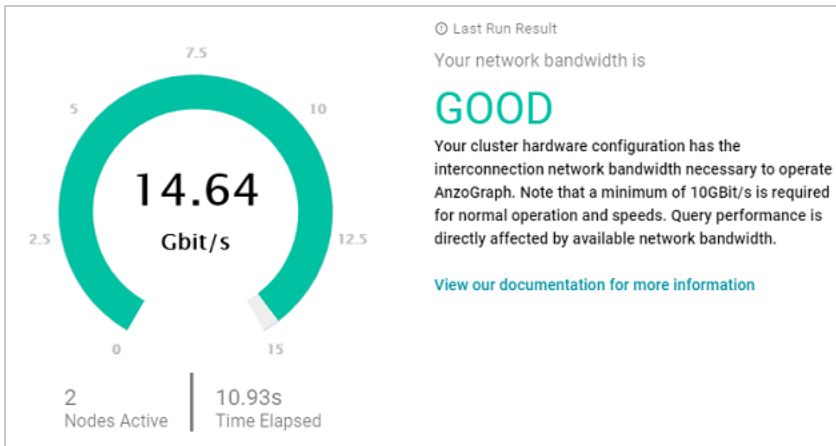
### Note

Network performance is not applicable for single servers. The benchmark described below is not available for single-server AnzoGraph deployments.

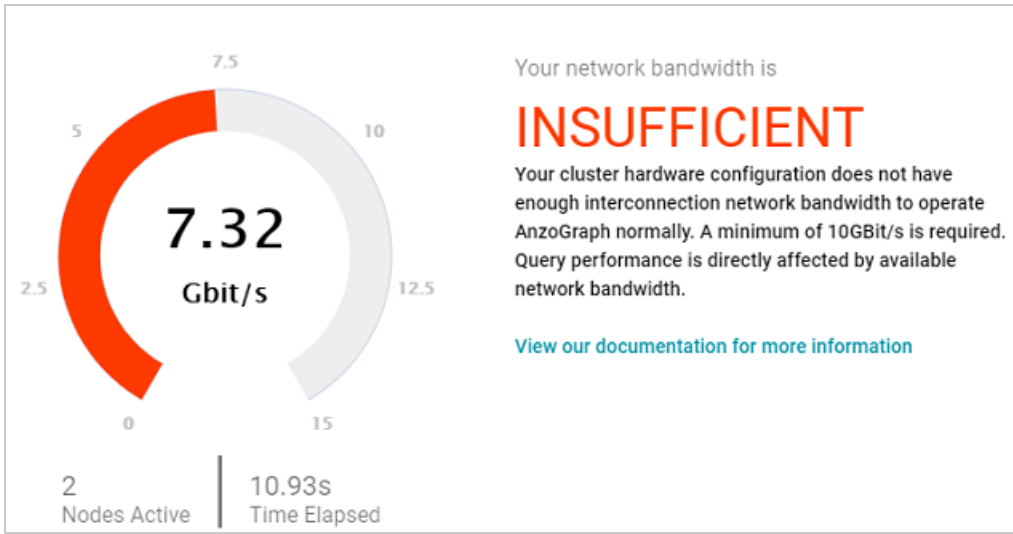
1. In the Administration application, expand the **Connections** menu and select **AnzoGraph**. The AnzoGraph screen is displayed and lists the connected AnzoGraph instances.
2. Click the name of the cluster that you want to evaluate. Anzo displays the Graphmarts screen for the cluster.
3. Click the **Diagnostics** tab and find the Network Benchmarking option at the bottom the screen.



4. By default, the benchmark is set to distribute 20 GB of data per node over the network. Each node in the cluster sends 20 GB to every other node. You can specify a different size if necessary. Note that increasing the value also increases the time to run the benchmark.
5. To run the test, click the **Run Benchmark** button. Anzo runs the benchmark and displays the results. For example:



If the bandwidth is less than 10 Gbit/s, Anzo displays an "Insufficient" result. For example:



When the results are insufficient, Cambridge Semantics recommends that you increase the network bandwidth. You can continue to use the cluster with the expectation of slower performance for network-bound operations.

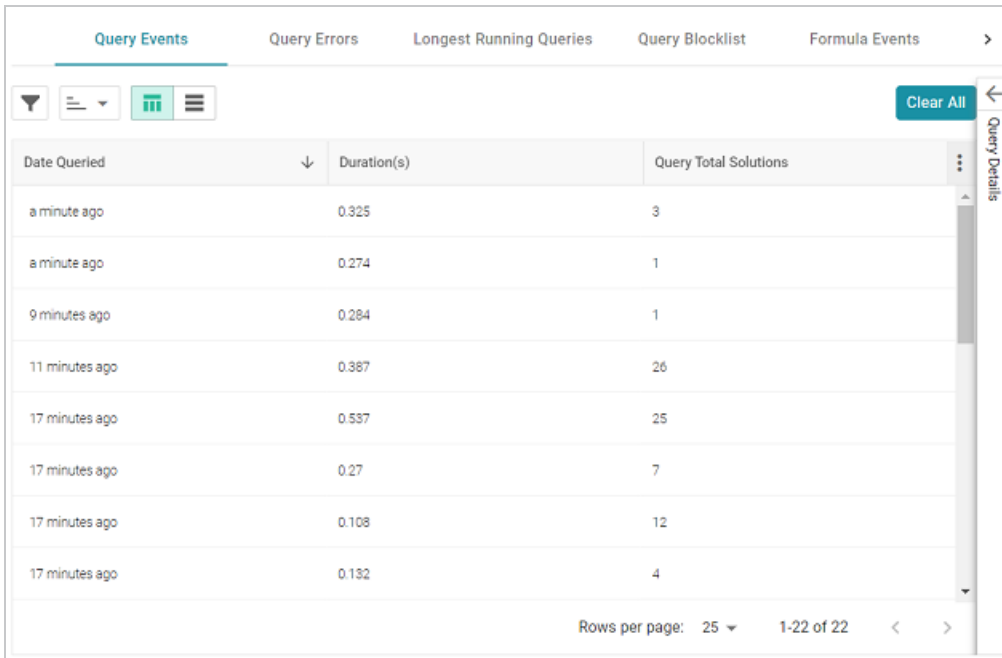
# System Query Audit

The System Query Audit screen enables administrators to quickly view a log of query events, query errors, the duration time for the longest running queries, and a list of any queries that have been blacklisted. The audit log also includes a Queued Queries tab that displays a list of the queries that are queued behind currently running queries. Administrators can cancel queries from the list and remove them from the queue. This topic provides information about using the System Query Audit log.

- [Viewing the System Query Audit Log](#)
- [AnzoGraph Detailed Query Timing](#)

## Viewing the System Query Audit Log

In the Administration application, expand the **Monitoring & Diagnostics** menu and select **System Query Audit**. Anzo displays the Query Events log. For example:



The screenshot shows the 'Query Events' tab in the System Query Audit interface. The table displays the following data:

Date Queried	Duration(s)	Query Total Solutions
a minute ago	0.325	3
a minute ago	0.274	1
9 minutes ago	0.284	1
11 minutes ago	0.387	26
17 minutes ago	0.537	25
17 minutes ago	0.27	7
17 minutes ago	0.108	12
17 minutes ago	0.132	4

At the bottom of the table, it shows 'Rows per page: 25' and '1-22 of 22'.

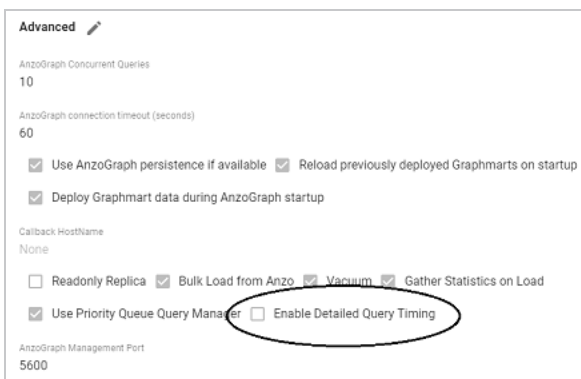
By default, the log shows an overview of all query events for all data sources. The table lists the date queried, the duration in milliseconds, and total number of solutions returned for each query event. You can select an event in the table to view details about that event, such as the target data source and query text, on the right side of the screen.

### Note

The System Query Audit log does not report on queries that complete in less than 100 milliseconds. In addition, queries that reuse the query cache from a previous run are not captured in the log. However, if a query takes less than 100 ms and uses cache, the original entry for the query is updated to increase the Cache Hit count.

## AnzoGraph Detailed Query Timing

In the Advanced settings for the AnzoGraph connection configuration, there is an **Enable Detailed Query Timing** setting (shown in the image below) that controls the level of information that is displayed for AnzoGraph queries in the System Query Audit log. This section describes the differences in logging when the setting is enabled and disabled.



### Important

Enabling detailed query timing increases the AnzoGraph workload and may decrease overall query performance.

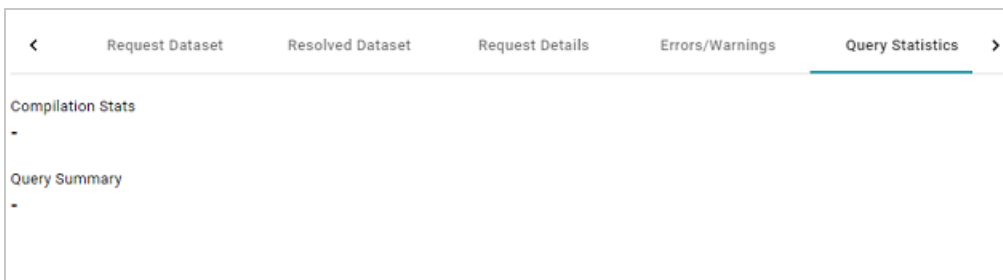
Enable Detailed Query Timing is disabled by default, meaning that Anzo will not run the additional statistics gathering queries unless you enable the setting. When Enable Detailed Query Timing is disabled, the System Query Audit log displays fewer query timing details. For example, the images



in the table below show a comparison between the **Result Details** tab when Enable Detailed Query Timing is disabled versus enabled. When the setting is disabled, details such as query Compilation Time are not recorded.

Enable Detailed Query Timing Disabled		Enable Detailed Query Timing Enabled	
Query Duration (ms) <b>10431</b>	Cache Hits -	Date Queried <b>a minute ago</b>	Original Query Date -
Query Total Solutions <b>14</b>	Query Results Cached <b>true</b>	Query Duration (ms) <b>13396</b>	Cache Hits -
Is Update <b>false</b>	Cache Hit <b>false</b>	Query Total Solutions <b>1</b>	Query Results Cached <b>true</b>
Is Error <b>false</b>	Dataset Cache Hit -	Is Update <b>false</b>	Cache Hit <b>false</b>
Query Canceled <b>false</b>	Query Results Valid <b>true</b>	Is Error <b>false</b>	Dataset Cache Hit -
Query Queued Time <b>0</b>	Query Already Compiled -	Query Canceled <b>false</b>	Query Results Valid <b>true</b>
	Compilation Time (ms) -	Query Queued Time <b>3</b>	Query Already Compiled <b>false</b>
	Query Execution Time (ms) <b>10424.964</b>		Compilation Time (ms) <b>17025.002</b>
			Query Execution Time (ms) <b>13388.833</b>

In addition, the images below show a comparison between the **Query Statistics** tab when Enable Detailed Query Timing is disabled versus enabled. When the setting is disabled, the Compilation Stats and Query Summary tables are empty:



When the setting is enabled, the Query Statistics tab is populated:

Overview	Result Details	Request Dataset	Resolved Dataset	Request Details	Errors/Warnings	Query Statistics						
Compilation Stats												
?query	?segment	?compile	?optimized	?secondpass	?duration	?bytes	?codeid	?path	?usertime	?systemtime	?rss	?starttime
1843	0	1	0	0	115108	31349	292	"code/292/0.dylib"	40	81	482564	2020-04-24T20:39
1843	0	1	1	1	165937	31349	292	"code/292/0.dylib"	30	138	482564	2020-04-24T20:39
1843	1	1	0	0	135463	30557	293	"code/293/0.dylib"	42	101	482564	2020-04-24T20:39
1843	1	1	1	1	1852621	30557	293	"code/293/0.dylib"	82	1788	482564	2020-04-24T20:39
1843	2	1	0	0	131465	28153	294	"code/294/0.dylib"	36	102	482564	2020-04-24T20:39
1843	2	1	1	1	179733	28153	294	"code/294/0.dylib"	43	140	482564	2020-04-24T20:39
1843	3	1	0	0	134088	29009	295	"code/295/0.dylib"	41	101	482564	2020-04-24T20:39
1843	3	1	1	1	1106167	29009	295	"code/295/0.dylib"	62	1002	482564	2020-04-24T20:39
1843	4	1	0	0	161024	40411	296	"code/296/0.dylib"	37	131	482564	2020-04-24T20:39
1843	4	1	1	1	261154	40411	296	"code/296/0.dylib"	35	232	482564	2020-04-24T20:39
1843	5	1	0	0	153252	34179	297	"code/297/0.dylib"	35	126	482564	2020-04-24T20:39
1843	5	1	1	1	240542	34179	297	"code/297/0.dylib"	46	201	482564	2020-04-24T20:39
1843	6	1	0	0	133239	24909	298	"code/298/0.dylib"	35	105	482564	2020-04-24T20:39
1843	6	1	1	1	197318	24909	298	"code/298/0.dylib"	37	164	482564	2020-04-24T20:39
1843	7	1	0	0	196144	40500	299	"code/299/0.dylib"	44	161	482564	2020-04-24T20:39

To enable detailed query timing, edit the AnzoGraph connection and select the **Enable Detailed Query Timing** checkbox. You do not need to restart Anzo or AnzoGraph after changing the setting.

# AnzoGraph Administration

The topics in this section provide reference information and instructions for performing administrative tasks on an AnzoGraph server. Some tasks, such as modifying server configuration settings, cannot be done via the Anzo Administration application. Other tasks, such as starting and stopping AnzoGraph using the system manager, are documented as alternate methods of managing AnzoGraph if the Administration application is unavailable or you prefer to use the AnzoGraph command line interface.

## In this section:

<a href="#">Starting and Stopping AnzoGraph</a> .....	292
<a href="#">Configuring AnzoGraph for Kerberos Authentication</a> .....	294
<a href="#">AnzoGraph CLI</a> .....	295
<a href="#">AnzoGraph Settings Reference</a> .....	300
<a href="#">Changing AnzoGraph Settings</a> .....	314

# Starting and Stopping AnzoGraph

This topic provides instructions for starting and stopping AnzoGraph.

## Note

The system management daemon, **azgmgrd**, should remain running at all times. When you restart the database, do not stop and start the daemon. There are two circumstances that require you to restart **azgmgrd**:

1. When uninstalling AnzoGraph.
2. When making changes to the `<install_path>/config/ip_addrs.conf` file if you add or remove servers from a cluster.

Follow the appropriate instructions below, depending on the current state of AnzoGraph and your use case:

- [Stop the Database \(Leave the System Management Daemon Running\)](#)
- [Start the Database \(the Daemon is Running\)](#)
- [Stop the Database and Daemon](#)
- [Start the Daemon and Database](#)
- [Reinitialize the Database](#)

## Stop the Database (Leave the System Management Daemon Running)

To stop the database, run the following command from the **leader server**:

```
sudo systemctl stop anzograph
```

If queries are running, the system manager waits the number of seconds in `stop_timeout` (the default value is 30 seconds) for any outstanding queries to complete and then stops the database.

## Start the Database (the Daemon is Running)

To start the database, run the following command from the **leader server**:

```
sudo systemctl start anzograph
```

## Stop the Database and Daemon

To stop the database and system management daemon, run the following commands from the **leader server**:

```
sudo systemctl stop anzograph
```

```
sudo systemctl stop azgmgrd
```

## Start the Daemon and Database

To start the system management daemon, run the following command. On clusters, run the command on **each server in the cluster**:

```
sudo systemctl start azgmgrd
```

To start the database after the system management daemon is running, run the following command on the **leader node**:

```
sudo systemctl start anzograph
```

## Reinitialize the Database

If you need to reinitialize the database to remove the generated code and any persisted data, run the following command. The system management daemon (azgmgrd) should be running.

### Important

Make sure that you are logged in as the Anzo service user any time you reinitialize the database.

```
<install_path>/bin/azgctl -start -init
```

# Configuring AnzoGraph for Kerberos Authentication

If you plan to load data to AnzoGraph from an HDFS file store that uses Kerberos authentication, follow the steps below to configure AnzoGraph for Kerberos authentication.

1. In order to be able to generate an authentication token for requesting encrypted ticket-granting tickets (TGT) from the key distribution center (KDC), each AnzoGraph host server must include the Kerberos workstation package, **krb5-workstation**. On each server in the cluster, run the following command to install the package:

```
sudo yum install -y krb5-workstation
```

2. In order to establish a connection to the KDC, AnzoGraph must have a copy of the KDC's **krb5.conf** file. Place a copy of **krb5.conf** in the **/etc** directory on each AnzoGraph host server.
3. In addition to **krb5.conf**, each AnzoGraph server needs a copy of the **.keytab** file from the principal node. The keytab file and principal name are used to generate an authentication token.

## Note

To find the location of the **.keytab** file and the principal name, you can look up the `dfs.web.authentication.kerberos.keytab` and `dfs.web.authentication.kerberos.principal` values in **hdfs-site.xml** on the HDFS master node.

Copy the **.keytab** file to any location on each AnzoGraph host server, and then run the following command to generate the authentication token:

```
kinit -p <principal_name> -k -t <path>/<keytab_file>
```

Where `<principal_name>` is the Kerberos principal name and `<path>/<keytab_file>` is the location and name of the **.keytab** file.

# AnzoGraph CLI

You can use the **azgi** command line interface (CLI) in the `<install_path>/bin` directory to issue commands directly to the database.

## Important

The azgi CLI works on the SPARQL HTTPS port and is enabled only when SSL protocol is enabled. SSL access is controlled by the `enable_ssl_protocol` setting. If HTTPS access is disabled and you want to enable it so that you can use the CLI, see [Changing AnzoGraph Settings](#) for instructions.

## AZGI Usage

This section describes the available azgi commands. To view the help, run `azgi -help`.

```
azgi [-f <filename>] [-c "<command>"] [-set <param>=<value>] [-h <host_url>] [-p <port>]
      [-u <username>:<password>] [-v] [-timer] [-raw] [-csv] [-json] [-xml] [-silent]
      [-nohead] [-noprogress] [-maxwid <width>] [-wide]
      [-nossl] [-o <file>] [-certs <directory>] [-context <json_file>]
```

### -f <filename>

Runs the specified SPARQL query file. For example, the following command runs the query or queries in the `query.rq` file:

```
azgi -f /home/user/query.rq
```

### -c "<command>"

Runs the command in quotation marks. For example, this command runs a query:

```
azgi -c "select distinct ?eventname from <http://cambridgesemantics.com/tickit>
where {?event <http://cambridgesemantics.com/tickit/eventname> ?eventname} limit
100"
```

You can include multiple `-c` options to run multiple commands. For example, this command runs two queries:

```
azgi -c "select * from <http://cambridgesemantics.com/tickit> where {?s ?p ?o} limit 100"
-c "select distinct ?likes from <http://cambridgesemantics.com/tickit> where {?person <http://cambridgesemantics.com/like> ?likes}"
```

And this command sets the `query_label` configuration setting to "events" before running the query:

```
azgi -c "set query_label to 'events'" -c "select distinct ?eventname from <http://cambridgesemantics.com/tickit> where {?event <http://cambridgesemantics.com/eventname> ?eventname} limit 100"
```

### **-set <param>=<value>**

Sets or changes parameter values in query files. For example this command runs the query in the `query_summary.rq` file with the `$query` parameter set to 2:

```
azgi -set query=2 -f query_summary.rq
```

### **-h <host\_url>**

Connects to a remote AnzoGraph server. For example, the following statement runs a query against AnzoGraph on host 10.104.55.27:

```
azgi -h 10.104.55.27 -c "select * from <http://cambridgesemantics.com/tickit> where {?s ?p ?o} limit 100"
```

### **-p <port>**

Used to connect to AnzoGraph on a non-default port. The default azgi port is 8256.

### **-u <username>:<password>**

Connects to the database with credentials (basic authentication). If you type `-u <username>` and exclude the password, the client prompts for the password. For example, the following command uses basic authentication to run a query:

```
azgi -u admin:Passw0rd1 -c "select ?g where {graph ?g {?s ?p ?o}} limit 100"
```

### **-v**

Displays verbose output such as client connection details. For example:



```
azgi -v -c "select distinct ?p from <http://cambridgesemantics.com/ticket>
where {<http://cambridgesemantics.com/person1> ?p ?o}"
```

```
Connecting to host=localhost port=8256
IPv4: connected
POST /sparql HTTP/1.1
Host: Anon
Accept: application/sparql-results+xml
User-Agent: azgi
Connection: keep-alive
Content-Length: 38
Content-Type: application/sparql-query
select distinct ?p from <http://cambridgesemantics.com/ticket> where
{<http://cambridgesemantics.com/person1> ?p ?o}
HTTP/1.1 200 OK
Date: Tue, 30 Jun 2020 00:24:42 GMT
Server: AnzoGraph
Access-Control-Allow-Origin: *
X-AnzoGraph-QueryExecution-Time: 20
Connection: close
Content-Type: application/sparql-results+xml; charset=utf-8
...
```

### **-timer**

Reports query execution time in milliseconds.

### **-raw**

Displays query results in raw XML, JSON, or CSV format, depending on what format you request.

### **-csv**

Displays results in CSV format.

### **-json**

Displays results in JSON format.

### **-xml**

Displays results in XML format.

### **-silent**

Suppresses the query output.

### **-nohead**

Suppresses headings in query results.

### **-noprogess**

Suppresses the progress messages that are displayed for queries that are in flight.

### **-maxwid <width>**

Overrides the default maximum column width of 50 characters for tabular query results. Using the [-wide](#) option described below is equivalent to `maxwid 60000`.

### **-wide**

Increases the column width for tabular query results from the default 50 characters to 60,000 characters. Equivalent to `-maxwid 60000`.

### **-noss1**

Instructs the client to make a non-SSL (HTTP) connection to the database. When using AZGI to send a request to a remote AnzoGraph server, include the `-h <host_url>` and `-p <port>` options when using `-noss1`. The default HTTP port is 7070. For example:

```
azgi -noss1 -h 10.100.0.20 -p 7070 -c "select (count(*) as ?cnt) where {?s ?p ?o}"
```

### **-o <file>**

Writes the response to the specified file. If the file exists, it is overwritten.

#### **Note**

When you specify this option to redirect output to a file, all progress messages will also be written to the file unless you also specify the [-noprogess](#) option. Cambridge Semantics recommends that you include `-noprogess` any time you output results to a file.

## **-certs <directory>**

Instructs the client to make a certified secure connection to the database. The AnzoGraph certificates are **ca.crt**, **serv.crt** (public key), and **serv.key** (private key) in the `<install_path>/config` directory. When sending requests to a remote AnzoGraph server, you can copy the AnzoGraph certificates to the server where you are using AZGI. For example, the following command runs a query on a remote AnzoGraph server. The command makes a certified connection using the AnzoGraph certificates, which were copied to the `/home/user/certs` directory:

```
azgi -h 10.10.10.01 -certs /home/user/certs
-c "select ?g where {graph ?g {?s ?p ?o}} limit 100"
```

This command runs the same query from the AnzoGraph server.

```
azgi -certs /opt/cambridgesemantics/anzograph/config -c "select ?g where {graph ?g
{?s ?p ?o}} limit 100"
```

## **-context <json\_file>**

Specifies the query context file on the AnzoGraph server file system to use with the request. Context files are JSON-formatted files with key-value pairs that provide connection details, such as user credentials, keys, and tokens, for authentication against data sources. For example:

```
{
  "url": "jdbc:mysql://10.111.4.9:3306/NORTHWIND",
  "username": "sysadmin",
  "password": "admin123"
}
```

# AnzoGraph Settings Reference

This topic provides reference information for each of the AnzoGraph system configuration settings. The configuration file, `<install_path>/config/settings.conf`, categorizes the settings as either **Basic** or **Advanced**. The advanced-level settings should only be configured by system administrators or users with an advanced level of knowledge about AnzoGraph or databases in general. For instructions on changing settings, see [Changing AnzoGraph Settings](#).

- [Basic Settings](#)
- [Advanced Settings](#)

## Basic Settings

This section describes the settings in the Basic section of `settings.conf`.

- `enable_persistence`
- `enable_sparql_protocol`
- `enable_ssl_protocol`
- `internal_directory`
- `max_memory`
- `output_format`
- `persistence_directory`
- `sparql_protocol_port`
- `sparql_spec_default_graph`
- `spill_directory`
- `ssl_protocol_port`
- `startup_info`
- `stop_timeout`

- [truncate\\_clob](#)
- [use\\_custom\\_ssl\\_files](#)
- [user\\_queues](#)
- [xray\\_sth\\_portion](#)
- [xray\\_sth\\_spool\\_duration](#)
- [xray\\_sth\\_spool\\_maxgb](#)

Setting	Default Value (type)	Description
<b>enable_persistence</b>	false (boolean)	Controls whether AnzoGraph saves a copy of the data in memory to disk. For more information, see <a href="#">Enabling Persistence (Preview)</a> .
<b>enable_sparql_protocol</b>	false (boolean)	Controls whether to enable the HTTP SPARQL endpoint. The <a href="#">sparql_protocol_port</a> setting controls the port to use to access the endpoint.  <b>Note</b> Enabling the SPARQL HTTP protocol opens the standard SPARQL-compliant HTTP endpoint. Unlike the Anzo protocol endpoint, the SPARQL HTTP endpoint is not secured.
<b>enable_ssl_protocol</b>	false (boolean)	Controls whether to enable the secure HTTPS SPARQL endpoint. The <a href="#">ssl_protocol_port</a> setting controls the port to use.  <b>Note</b> Enabling the SPARQL HTTPS protocol opens the

Setting	Default Value (type)	Description
		<p>standard SPARQL-compliant HTTPS endpoint. Unlike the Anzo protocol endpoint, the SPARQL HTTPS endpoint is encrypted but not authenticated.</p>
<b>internal_directory</b>	Not set (char)	The directory where AnzoGraph should save internal database-related files such as generated code, logs, and query plans. When not set, the default is <code>&lt;install_path&gt;/internal</code> . For more information, see <a href="#">Relocating AnzoGraph Directories</a> .
<b>max_memory</b>	System-based (int)	Specifies the amount of memory (in MB) to make available for AnzoGraph. The default is system-based; at startup, AnzoGraph determines the amount of RAM that is available and sets <code>max_memory</code> . In test environments where AnzoGraph may be co-located with other programs, you can set the <code>max_memory</code> value to put a limit on the amount of memory AnzoGraph can use. However, Cambridge Semantics recommends that you do not set <code>max_memory</code> unless instructed to do so by Support.
<b>output_format</b>	xml (char)	Specifies the default output format for AnzoGraph responses. Valid values are <code>xml</code> , <code>json</code> , or <code>csv</code> .
<b>persistence_directory</b>	Not set (char)	The directory where AnzoGraph should save data when <a href="#">enable_persistence</a> is <code>true</code> and data is persisted to disk. When not set, the default is <code>&lt;install_path&gt;/persistence</code> . For more information, see <a href="#">Relocating AnzoGraph Directories</a> .

Setting	Default Value (type)	Description
<code>sparql_protocol_port</code>	7070 (int)	Specifies the port to use to access the SPARQL HTTP endpoint when <code>enable_sparql_protocol</code> is <code>true</code> .
<code>sparql_spec_default_graph</code>	false (boolean)	Controls the default scope of SPARQL queries when FROM clauses are excluded from a query. When <code>false</code> , queries without FROM clauses target the default graph (DEFAULTSET) only. Triples in named graphs will not be included in the scope of the query. When <code>true</code> , AnzoGraph conforms to the SPARQL specification and includes the default graph and all named graphs in the scope of a query that omits the FROM clause. For more information, see <a href="#">Changing the Default FROM Clause Behavior</a> .
<code>spill_directory</code>	Not set (char)	The directory where AnzoGraph should save temporary query files that spill to disk. When not set, the default is <code>&lt;install_path&gt;/spill</code> . For more information, see <a href="#">Relocating AnzoGraph Directories</a> . <div style="background-color: #fff9c4; padding: 10px; margin-top: 10px;"> <p><b>Important</b></p> <p>AnzoGraph uses <code>O_DIRECT</code> to read the spill files into the database. If you relocate the spill directory, make sure to place it on an ext4 file system that supports <code>O_DIRECT</code>.</p> </div>
<code>ssl_protocol_port</code>	8256 (int)	This setting specifies the port to use to access the SPARQL HTTPS endpoint when <code>enable_ssl_protocol</code> is <code>true</code> .
<code>startup_info</code>	1 (int)	Specifies how verbose the database startup message is: -

Setting	Default Value (type)	Description
		<b>0</b> -quiet, <b>1</b> -ready, <b>2</b> -ports, <b>3</b> -more.
<b>stop_timeout</b>	30 (int)	When the database stop command is issued, this setting specifies the number of seconds to wait for queries to finish before stopping the database.
<b>truncate_clob</b>	false (boolean)	Controls whether to automatically truncate large strings to the maximum string size (2 MB).
<b>use_custom_ssl_files</b>	false (boolean)	Indicates whether you are replacing AnzoGraph's self-signed certificates with your own custom certificates. To configure AnzoGraph to use your certificates, follow the instructions in <a href="#">Replace the Default Self-Signed Certificates with Trusted Certificates</a> in the Deployment Guide. <div data-bbox="699 1026 1474 1451" style="background-color: #fff9c4; padding: 10px; margin-top: 10px;"> <p><b>Important</b></p> <p>Anzo also needs to trust the new certificates. Make sure you have <a href="#">Trust All TLS Certificates</a> enabled on the AnzoGraph connection or make sure Anzo's trust store has either the certificate for the CA that signed the certificate or the certificate itself.</p> </div>
<b>user_queues</b>	40 (int)	Sets the limit on the number of queries that can run concurrently.
<b>xray_sth_portion</b>	0.001 (float)	<b>In 3.1 releases</b> , this setting configures the percentage of total memory to use for storing historical system table information in memory before spilling to disk. The default



Setting	Default Value (type)	Description
		value 0.001 = 0.1% of memory.
<b>xray_sth_spool_duration</b>	7days (char)	<b>In 3.1 releases</b> , this setting controls the length of time to accumulate historical system table information on disk for xrays.
<b>xray_sth_spool_maxgb</b>	20 (int)	<b>In 3.1 releases</b> , this setting controls the maximum size (in GB) per node of historical system table information to keep on disk for xrays. When the limit is reached, AnzoGraph deletes the oldest N records, where N depends on the server workload but is typically about 5 to 6 minutes worth of system table data.

## Advanced Settings

This section describes the settings in the Advanced section of settings.conf.

- [anzo\\_protocol\\_port](#)
- [auto\\_restart\\_directory](#)
- [auto\\_restart\\_max\\_attempts](#)
- [auto\\_restart\\_time](#)
- [aws\\_log\\_level](#)
- [aws\\_search\\_regions](#)
- [azgmgrd\\_client\\_auth](#)
- [azgmgrd\\_password](#)
- [bits\\_per\\_pred\\_index](#)
- [bits\\_per\\_uri\\_index](#)

- blank\_node\_name
- call\_home\_for\_updates
- comm\_port\_base
- compile\_concurrent
- compile\_max\_memory
- compile\_max\_seconds
- compile\_optimized
- copy\_file\_size
- enable\_owlstats
- enable\_refresh\_stats\_on\_update
- enable\_root\_user
- enable\_unbound\_variables
- float\_decimals
- float\_format
- grpc\_token\_expiry
- ignore\_deniedlist\_queries
- jvm\_max\_memory
- jvm\_options
- log\_directory
- policy\_file\_enabled

Setting	Default Value (type)	Description
anzo_protocol_port	5700 (int)	The Anzo protocol (gRPC) port for secure communication between AnzoGraph and Anzo.

Setting	Default Value (type)	Description
<b>auto_restart_directory</b>	Not set (char)	Specifies the base location of the <b>auto_restart</b> directory, which contains the <code>denied_list</code> , <code>warned_list</code> , and <code>unanalyzed_list</code> directories. When not set, the default is <code>&lt;install_path&gt;/internal</code> . For more information about the auto-restart feature, see <a href="#">Managing Automatic Restarts</a> .
<b>auto_restart_max_attempts</b>	5 (int)	Specifies the number of times the system manager should attempt to start the database after a crash. The default value is <b>5</b> , which means the system manager will attempt to restart the database a maximum of 5 times. Changing <code>auto_restart_max_attempts</code> to <b>0</b> disables the auto-restart feature. For more information about the auto-restart feature, see <a href="#">Managing Automatic Restarts</a> .
<b>auto_restart_time</b>	600 (int)	Specifies the number of seconds to spend attempting to restart the database. If all attempts fail and this time limit is reached, the system manager stops trying to restart the database. The default value is <b>600</b> , which means that the system manager will attempt to restart the database for a maximum of 600 seconds (10 minutes). For more information about the auto-restart feature, see <a href="#">Managing Automatic Restarts</a> .
<b>aws_log_level</b>	2 (int)	AnzoGraph uses an AWS C++ SDK for loading data from S3. This setting controls the logging level for the AWS SDK. The default value is <b>2</b> , which is error level logging. Valid values are <b>0</b> (off), <b>1</b> (fatal), <b>2</b> (error), <b>3</b> (warn), <b>4</b> (info), <b>5</b> (debug), and <b>6</b> (trace).
<b>aws_search_regions</b>	Not set (char)	Lists the regions to search for AWS S3 buckets that are listed as file locations for LOAD queries.

Setting	Default Value (type)	Description
<b>azgmgrd_client_auth</b>	false (boolean)	Controls whether the system management daemon (azgmgrd) and system manager (azgctl) use authentication in addition to encryption when connecting to other system managers over the system management gRPC port (5600). The default value is <b>false</b> , which means the system management connections are encrypted but not authenticated. For more information about azgmgrd authentication, see <a href="#">Enable System Manager Authentication</a> in the Deployment Guide.
<b>azgmgrd_password</b>	N/A	This is the password that the system management daemon (azgmgrd) uses for gRPC access to the database. Typically this value is not changed as it is only used internally for authentication between the system manager and the database. If you do want to change the password, you cannot change it directly in the settings.conf file. See <a href="#">Change the System Manager Password</a> in the Deployment Guide for instructions.
<b>bits_per_pred_index</b>	16 (int)	Specifies the maximum number of unique graph and predicate URIs that can be stored in AnzoGraph. The maximum number is two to the power of this value. The default value (16) for bits_per_pred_index is set to the maximum value and should not be changed. $2^{16} = 64k$ unique predicate and graph URIs.
<b>bits_per_uri_index</b>	32 (int)	Specifies the maximum number of unique subject URIs that can be stored in AnzoGraph. The maximum number is two to the power of this value. The default value (32) for bits_per_uri_index is set to the maximum value and should not be changed. $2^{32} = 4+$ trillion unique subject URIs.

Setting	Default Value (type)	Description
<b>blank_node_name</b>	genid (char)	Specifies the default name basis for blank nodes. By default, AnzoGraph generates a number ID for the node. For example, inserting <code>_:a</code> results in a URI such as <code>bnode:a__63</code> .
<b>call_home_for_updates</b>	false (boolean)	Controls whether AnzoGraph checks for updates over the internet.
<b>comm_port_base</b>	9100 (int)	Specifies the port to use for internal cluster communication.
<b>compile_concurrent</b>	8 (int)	Specifies the maximum number of generated code compilations to perform concurrently.
<b>compile_max_memory</b>	500 (int)	Sets the limit on the amount of memory (in MB) that AnzoGraph can allocate for compiling generated code before switching from optimized compile to non-optimized compile.
<b>compile_max_seconds</b>	30 (int)	Sets the limit on the number of seconds to spend compiling generated code before switching from optimized compile to non-optimized compile.
<b>compile_optimized</b>	background (char)	Specifies the type of optimized compile to perform.
<b>copy_file_size</b>	5 (int)	Controls the size (in MB) of the Turtle files that are generated when graphmart contents are exported to files.
<b>enable_owlstats</b>	true (boolean)	In order to generate query execution plans, AnzoGraph

Setting	Default Value (type)	Description
		needs to gather statistics about the data, such as the number of triples per graph and number of distinct subjects and predicates. This setting controls whether advanced statistics gathering, called OWL stats, is enabled. OWL stats use the metadata from data models to generate statistics. Cambridge Semantics recommends that you leave <code>enable_owlstats</code> enabled unless otherwise instructed.
<code>enable_refresh_stats_on_update</code>	true (boolean)	Controls whether the statistics in AnzoGraph are flagged as outdated when a graph is updated.
<code>enable_root_user</code>	false (boolean)	Controls whether to allow a user running with root privileges to start AnzoGraph.
<code>enable_unbound_variables</code>	false (boolean)	Controls whether AnzoGraph returns an empty result or an error if a query references a missing graph or includes unbound variables. This value is set to <b>false</b> by default, which means AnzoGraph returns an error. For more information, see <a href="#">Ignoring Missing Graphs</a> .
<code>float_decimals</code>	6 (int)	<p><b>This setting does not apply to results that are returned from AnzoGraph to Anzo over gRPC protocol.</b> Anzo converts floating point values to Java native float objects with 6 – 7 total digits of precision. <b>This setting would only affect results that are returned directly from AnzoGraph to another application over HTTP/S protocol.</b></p> <p>AnzoGraph formats floating point types using the printf</p>

Setting	Default Value (type)	Description
		<p>format string <code>%.precision format</code>, where <b>precision</b> is the value of the <code>float_decimals</code>, and <b>format</b> is the value of <code>float_format</code>.</p> <div style="border: 1px solid #ccc; background-color: #f0f0f0; padding: 10px; margin: 10px 0;"> <p><b>Note</b></p> <p>The interpretation of <code>float_decimals</code> differs depending on the value in <code>float_format</code>. For fixed point formats (f and F), <code>float_decimals</code> specifies the number of digits to include after the decimal point, padded with zeros if necessary. For floating point formats (e, E, g, and G), <code>float_decimals</code> specifies the number of significant digits to round the result to.</p> </div>
<code>float_format</code>	g (char)	<p><b>This setting does not apply to results that are returned from AnzoGraph to Anzo over gRPC protocol.</b> Anzo converts floating point values to Java native float objects with 6 – 7 total digits of precision. <b>This setting would only affect results that are returned directly from AnzoGraph to another application over HTTP/S protocol.</b></p> <p>AnzoGraph formats floating point types using the printf format string <code>%.precision format</code>, where <b>format</b> is the value of the <code>float_format</code>, and <b>precision</b> is the value of <code>float_decimals</code>. Valid values for <code>float_format</code> are e, E, f, F, g, or G. In the default configuration, a value of 10000000000.123 is returned as 1e+10.</p>
<code>grpc_token_</code>	0 (int)	Controls how often (in seconds) the gRPC token expires. A

Setting	Default Value (type)	Description
<code>expiry</code>		value of 0 means the token never expires.
<code>ignore_deniedlist_queries</code>	true (boolean)	Controls whether denied list queries are blocked from running or are allowed to be run when the database is returned to normal operation. The default value is <b>true</b> , which means denied list queries are ignored. Incoming queries are not compared with the denied list and are permitted to run. If <code>ignore_deniedlist_queries</code> is <b>false</b> , denied list queries are not ignored and are therefore blocked from running until they are removed from the denied list. For more information about the auto-restart feature, see <a href="#">Managing Automatic Restarts</a> .
<code>jvm_max_memory</code>	Not set (char)	Specifies the maximum size of the heap that can be used by the embedded Java virtual machine (JVM). This setting affects memory used for queries that employ AnzoGraph Java extensions, such as the Graph Data Interface.  Use <b>k</b> , <b>m</b> , or <b>g</b> (case insensitive) for KiB, MiB, or GiB. You can also specify % to indicate a percentage of the total memory that is available to AnzoGraph. By default, this value is not set, which means <code>jvm_max_memory</code> defaults to either <b>5%</b> of the total memory or <b>4g</b> , whichever value is smaller.
<code>jvm_options</code>	Not set (char)	Lists any optional parameters to use for configuring the embedded JVM. Use a semicolon-delimited (;) list to specify multiple parameters. For information about JVM options, see <a href="#">Options</a> in the Java Documentation.
<code>log_directory</code>	Not set (char)	Specifies where to write system management daemon



Setting	Default Value (type)	Description
		<p>(azgmgrd) log files. These types of logs (<code>azgmgrd.log</code>, <code>azgctl-&lt;user&gt;.log</code>, <code>azgpidmgr.log</code>, and <code>azgpids.log</code>) are created before the system is initialized and may be written before the <code>&lt;install_path&gt;/internal/log</code> directory exists. Therefore, they are located outside of the AnzoGraph file system, <code>/tmp</code> by default. If you change the <code>log_directory</code> value, Cambridge Semantics recommends that you choose another location that is outside the internal AnzoGraph directories.</p>
<p><b>policy_file_enabled</b></p>	<p>false (boolean)</p>	<p>Enables or disables file system access control policies. When <code>policy_file_enabled</code> is <b>false</b> (the default value), AnzoGraph does not perform file path access checks when a query reads or writes files or directories on the file system. When <code>policy_file_enabled</code> is <b>true</b> and a query attempts to access a file or directory on the file system, AnzoGraph performs the file path access checks that are configured in the <code>file_policy_*</code> settings and returns an access denied error message if the path is not accessible. For instructions on configuring file access policies and the <code>file_policy_read</code>, <code>write</code>, <code>delete</code>, and <code>deny</code> settings, see <a href="#">Managing AnzoGraph File Access Policies</a>.</p>

# Changing AnzoGraph Settings

The topics in the section give an overview of the AnzoGraph system configuration file and provide general rules for changing the configuration as well as detailed instructions on making common types of changes.

## In this section:

Configuration File Overview .....	315
Managing AnzoGraph File Access Policies .....	316
Relocating AnzoGraph Directories .....	320
Enabling Persistence (Preview) .....	322
Ignoring Missing Graphs .....	324
Changing the Default FROM Clause Behavior .....	326
Managing Automatic Restarts .....	327

## Configuration File Overview

The default AnzoGraph system configuration is optimized for most AnzoGraph installations. If Cambridge Semantics Support recommends that you change the configuration, you can edit the configuration file, `<install_path>/config/settings.conf`, to modify or add settings. Each time you start the database, AnzoGraph reads this file and stores the configuration in memory. **On a cluster, change settings.conf on the leader server only.** See [AnzoGraph Settings Reference](#) for information about the units of measurement for the settings as well as any special instructions.

- The commented lines in the file show the default configuration values. To customize the value for a setting that is commented out, uncomment the line and edit the value portion of `setting_name=value`.
- To add settings to settings.conf, add the setting and new value in the format below. Type each setting and value pair on a new line.

```
setting_name=value
```

### Note

AnzoGraph applies settings from the top to the bottom of the file. If the same setting appears more than once, AnzoGraph applies the value for the last instance of the setting. The last instance overrides any previous instances.

- To revert AnzoGraph to a previous configuration from a backup file, rename the existing settings.conf file and then change the name of the desired backup file to **settings.conf**.

### Important

After you change settings.conf, you must restart AnzoGraph for the settings to take effect. See [Starting and Stopping AnzoGraph](#) for instructions.

# Managing AnzoGraph File Access Policies

In AnzoGraph Version 2.5.6 and later, you can configure file system access control policies to ensure that only certain files or directories are accessible to AnzoGraph during the execution of a query. This topic describes the configuration settings that define the file access policies and provides instructions for setting up policies.

- [File Access Policy Settings Reference](#)
- [File Access Control Behavior](#)
- [Setting Up File Access Policies](#)

## File Access Policy Settings Reference

### `policy_file_enabled`

The `policy_file_enabled` setting is the parent setting that controls whether or not file system access policies are enabled and followed. When `policy_file_enabled` is **false** (the default value), AnzoGraph does not perform file path access checks when a query references files or directories on the file system. When `policy_file_enabled` is **true** and a query attempts to access a file or directory on the file system, AnzoGraph performs the file path access checks that are configured in the [policy\\_file\\_read](#), [write](#), [delete](#), and [deny](#) settings described below.

### `policy_file_read, write, delete, and deny`

The `policy_file_read`, `write`, `delete`, and `deny` settings specify the paths to directories and/or files on the file system that AnzoGraph requests are allowed to read from, write to, or delete from. For each of the "allowed" read, write, and delete settings, there is a corresponding **deny** setting that configures the paths for which requests are denied read, write, and delete access. This enables you to allow broad access to parent directories, if desired, and then use the deny settings to restrict access to certain subdirectories under them if needed.

The values for the settings are wildcard patterns that AnzoGraph uses to match directories and/or file names. Patterns are specified using basic file globbing syntax as described in the [glob](#)

(7) [Linux manual page](#). Each `policy_file_*` setting accepts one or more patterns. Separate multiple patterns with a semicolon (;). For readability, you can also include spaces between patterns.

### Important

Prior to matching paths in an incoming request to the configured access policy patterns, AnzoGraph resolves the paths in the request to canonical paths (using the `std::filesystem::weakly_canonical` function described [here](#) at [cppreference.com](http://cppreference.com)). That means segments such as `./` or `../` are fully expanded prior to being compared to patterns. If a segment in the request path is a symlink, that segment is also expanded prior to checking for a match. **Make sure that all access policy patterns match absolute paths.** Otherwise, expanded relative path or symlink segments in a request will not match any patterns. For example, if users normally include a path like `/source-files/` in a request but `/source-files/` is a symlink to `/mnt/anzoshare/data/source-files/`, include the path to `/mnt/anzoshare/data/source-files/` in the pattern.

The following list describes the settings and provides sample pattern values. The [File Access Control Behavior](#) section below includes specifics about pattern matching and access checks.

- **policy\_file\_read:** Specifies the pattern(s) to match for paths that queries have permission to read from. For example, a value such as the following gives AnzoGraph requests read-only access to all files and directories under the `/opt/anzoshare` and `/mnt/data` directories:

```
policy_file_read=/opt/anzoshare/* ; /mnt/data/*
```

- **policy\_file\_read\_deny:** Specifies the pattern(s) to match for paths that queries should not be allowed to read. For example, the following value means requests will not be allowed to read any files or directories under `/etc` or `/root`:

```
policy_file_read_deny=/etc/* ; /root/*
```

- **policy\_file\_write:** Specifies the pattern(s) to match for paths that queries have permission to write to. For example, the following value gives requests write access to the `/tmp` and `/home`

directories in addition to the `/opt/anzoshare/store` and `/mnt/data/store` directories.

```
policy_file_write=/tmp/* ; /home/* ; /opt/anzoshare/store/* ;  
/mnt/data/store/*
```

### Important

If you have Graphmarts with Export Steps, make sure the write policy gives AnzoGraph write access to the appropriate Anzo Data Store.

- **policy\_file\_write\_deny:** Specifies the pattern(s) to match for paths that queries are denied write access to.
- **policy\_file\_delete:** Specifies the pattern(s) to match for paths that queries have permission to delete.
- **policy\_file\_delete\_deny:** Specifies the pattern(s) to match for paths that queries are denied delete access to.

### Note

The AnzoGraph installation path (`<install_path>/*`) is automatically added to each of the `*_deny` policies.

## File Access Control Behavior

When a query that includes a path to a file or directory is run (such as in a GDI query with `s:url "/opt/anzoshare/data/csv"` or in a `LOAD <dir:/mnt/data/rdf.ttl.gz>` statement), AnzoGraph resolves that path (for example, if the path includes `./` or `../` segments) to a canonical path prior to checking whether it matches a `policy_file` pattern. If any segment of the path is a symlink, that segment is also expanded prior to being matched to a pattern. If the specified file or directory matches one of the allowed access patterns and it is not matched to a deny pattern, the query is executed. If the specified path is matched to a denied pattern or is not matched to any of the allowed patterns, the query is aborted and AnzoGraph returns an access denied error message.

## Setting Up File Access Policies

1. Stop the database. See [Stop the Database \(Leave the System Management Daemon Running\)](#) for instructions.
2. **On the leader node**, open the AnzoGraph settings file, **settings.conf**, in a text editor. The file is in the `<install_path>/config` directory.
3. In **settings.conf**, uncomment the `policy_file_enabled=false` line and change the value to **true**:

```
policy_file_enabled=true
```

4. Locate the additional `policy_file_*` settings:

```
# File system paths that may be deleted (';' delimited) ()
# policy_file_delete=

# File system paths that may not be deleted (';' delimited) ()
# policy_file_delete_deny=

# File system paths that may be read from (';' delimited) ()
# policy_file_read=

# File system paths that may not be read from (';' delimited) ()
# policy_file_read_deny=

# File system paths that may be written to (';' delimited) ()
# policy_file_write=

# File system paths that may not be written to (';' delimited) ()
# policy_file_write_deny=
```

5. Uncomment each of the `policy_file_*=` lines that you want to set, and add the wildcard pattern or patterns that you want to match for each of the policies.
6. Save and close **settings.conf**.
7. Restart the database to apply the configuration change. See [Start the Database \(the Daemon is Running\)](#) for instructions.

## Relocating AnzoGraph Directories

Follow the instructions in this section to designate alternate locations for certain directories included in the AnzoGraph installation. You have the option to relocate the **persistence** directory where the system saves the data in memory to the file system, the **internal** directory where the system saves database-related files such as logs and generated code, and the **spill** directory where the system saves any temporary query files that spill to disk.

You can change the settings described in this section at any time. Once you restart the database, AnzoGraph starts saving any new files in the directory locations that you specify.

### Note

The system does not relocate any existing directories or files. You can move the existing files manually if needed.

1. Stop the database. See [Stop the Database \(Leave the System Management Daemon Running\)](#) for instructions.
2. **On the leader node**, open the AnzoGraph settings file, **settings.conf**, in a text editor. The file is in the `<install_path>/config` directory.
3. Uncomment the lines for any of the following settings in `settings.conf`. Then edit the value portion of `setting=value` to specify the desired directory.
  - **internal\_directory**: The directory where you want AnzoGraph to save internal database-related files such as generated code, logs, and query plans. The default value is `<install_path>/internal`.
  - **persistence\_directory**: The directory where you want AnzoGraph to save data when writing data to disk. The default value is `<install_path>/persistence`.
  - **spill\_directory**: The directory where you want the AnzoGraph to save any temporary query files that spill to disk. The default value is `<install_path>/spill`.



### Important

AnzoGraph uses O\_DIRECT to read the spill files into the database. If you relocate the spill directory, make sure to place it on an ext4 file system that supports O\_DIRECT.

4. Save and close settings.conf.
5. Restart the database to apply the configuration change. See [Start the Database \(the Daemon is Running\)](#) for instructions.

## Enabling Persistence (Preview)

By default, Anzo manages the data in AnzoGraph by automatically reloading Graphmart data into memory when AnzoGraph is restarted. You also have the option to enable persistence on the AnzoGraph instance. When persistence is enabled, AnzoGraph saves the data in memory to disk after every transaction. Each time AnzoGraph is restarted, the persisted data is automatically loaded back into memory. Once the data is loaded into memory, rather than automatically reloading active Graphmarts, Anzo checks to see if the last updated timestamp in AnzoGraph matches the last updated value in Anzo. If the timestamps match, Anzo does not initiate a reload. If there is a mismatch, Anzo reloads the active Graphmarts to update the data in memory to the latest version.

### Note

The AnzoGraph persistence feature is available as a **Preview** release, which means the implementation has recently been completed but is not yet thoroughly tested with Anzo and could be unstable. The feature is available for trial usage, but Cambridge Semantics recommends that you do not rely on Preview features in production environments.

This topic lists important information to consider before enabling persistence and provides instructions for enabling persistence in the AnzoGraph configuration file.

## Important Considerations

Before enabling persistence, consider the following important notes:

- In general, each AnzoGraph server needs access to about twice as much disk space as RAM on the server. By default, AnzoGraph saves data to the `<install_path>/persistence` directory on the local file system. You can also configure AnzoGraph to save data to a mounted file system. For more information, see [Relocating AnzoGraph Directories](#).
- Persisted data is unique to each AnzoGraph version and cannot be re-used after an upgrade. If you upgrade AnzoGraph and persistence is enabled, the database will not start until it is reinitialized to remove the persisted data. See [Reinitialize the Database](#) for instructions.

- When persistence is enabled, transactional workloads that perform many concurrent write operations may experience a performance degradation due to the overhead of writing the data from each transaction to disk.

## Enabling Persistence

Follow the steps below to enable the AnzoGraph save to disk option.

1. Stop the database. See [Stop the Database \(Leave the System Management Daemon Running\)](#) for instructions.
2. **On the leader node**, open the AnzoGraph settings file, **settings.conf**, in a text editor. The file is in the `<install_path>/config` directory.
3. In `settings.conf`, find the following line in the file:

```
enable_persistence=false
```

4. Change the `enable_persistence` value to **true**:

```
enable_persistence=true
```

5. Save and close `settings.conf`.
6. Restart the database to apply the configuration change. See [Start the Database \(the Daemon is Running\)](#) for instructions.

After each transaction, AnzoGraph saves the data in memory to disk in the location specified in the `persistence_directory` setting. Each time AnzoGraph is restarted, the persisted data is automatically loaded back into memory.

### Note

To avoid unnecessary reloads, make sure that the AnzoGraph connection in Anzo is configured to enable the **Use AnzoGraph persistence if available** option. See [Connecting to AnzoGraph](#) for more information.

## Ignoring Missing Graphs

By default, AnzoGraph returns a "No such graph or view" error and aborts the query if a query references a graph that does not exist. You can configure AnzoGraph to conform to the SPARQL specification and return an empty result instead of an error, however, if a query references a missing graph. Follow the instructions below to configure the system to return empty results instead of an error when a referenced graph does not exist.

1. Stop the database. See [Stop the Database \(Leave the System Management Daemon Running\)](#) for instructions.
2. **On the leader node**, open the AnzoGraph settings file, **settings.conf**, in a text editor. The file is in the `<install_path>/config` directory.
3. In **settings.conf**, uncomment the `enable_unbound_variables=false` line and change the value to true:

```
enable_unbound_variables=true
```

4. Save and close **settings.conf**.
5. Restart the database to apply the configuration change. See [Start the Database \(the Daemon is Running\)](#) for instructions.

### Note

In addition to allowing queries that reference non-existent graphs to succeed, setting `enable_unbound_variables` to true also configures AnzoGraph to ignore unbound variables elsewhere in queries. For example, by default (when `enable_unbound_variables=false`), if a query includes a variable in the SELECT list that is not referenced in a WHERE clause pattern, AnzoGraph aborts the query and returns a "Named variable not in contained WHERE clause" error. When `enable_unbound_variables=true`, AnzoGraph does not warn the user about unbound variables. Instead, the results are empty for the unbound variable. For example:

```
SELECT ?unbound ?person ?name
FROM <http://cambridgesemantics.com/people>
```

```
WHERE {?person <http://cambridgesemantics.com/people#firstname> ?name}
LIMIT 5
```

unbound	person	name
	person35632	Ross
	person20216	Quin
	person35859	Kellie
	person2551	Maris
	person24963	Madonna

5 rows

## Changing the Default FROM Clause Behavior

By default, if a query omits FROM clauses, the scope of the query is limited to the default graph (DEFAULTSET). Triples in named graphs will not be included in the scope of the query. The default behavior is controlled by the `sparql_spec_default_graph` configuration setting. To configure AnzoGraph to conform to the SPARQL specification and include the default graph and all named graphs in the scope of a query that omits the FROM clause, follow the instructions below.

1. Stop the database. See [Stop the Database \(Leave the System Management Daemon Running\)](#) for instructions.
2. **On the leader node**, open the AnzoGraph settings file, `settings.conf`, in a text editor. The file is in the `<install_path>/config` directory.
3. In `settings.conf`, uncomment the `sparql_spec_default_graph=false` line and change the value to true:

```
sparql_spec_default_graph=true
```

4. Save and close `settings.conf`.
5. Restart the database to apply the configuration change. See [Start the Database \(the Daemon is Running\)](#) for instructions.

## Managing Automatic Restarts

AnzoGraph can be configured so that the system manager automatically restarts the database and evaluates the queries that were running if AnzoGraph shuts down unexpectedly. This topic describes the process that occurs when AnzoGraph automatically restarts and provides information about the configuration settings that control the functionality as well as administrative information for managing the evaluated queries.

- [Automated Restart Procedure](#)
- [Automated Restart System Settings](#)
- [Removing a Query from the Block List](#)

### Automated Restart Procedure

The steps below describe what occurs during the automatic restart process after AnzoGraph has crashed:

1. The system manager restarts the database in **safe mode**. In safe mode, AnzoGraph is locked to users and returns the following message if a user runs a query: "AnzoGraph is running in safe-mode. Cannot execute query." In addition, running `azgctl -status` to check the status of the database returns the message "AnzoGraph is running in safe-mode." If persistence is enabled, the data that was in memory at the time of the crash is reloaded into memory.
2. While in safe mode, AnzoGraph runs any queries that were inflight at the time of the crash. By executing the queries that were running, AnzoGraph tries to determine if the crash was directly caused by one of the inflight queries.
3. Depending on the outcome of running the inflight queries, AnzoGraph does the following:
  - If all inflight queries run to completion in safe mode, they are all added to the **warned\_list**. In addition, each query is copied to a file named `<query_ID>.txt` in the `<install_path>/internal/auto_restart/<timestamp>/warned_list` directory.

### Note

When all inflight queries complete successfully, that means it is unlikely that any one of the queries on its own is the culprit for the crash. However, all of the queries are added to the warned list because it is possible that the combination of queries run concurrently could have caused the crash.

- If any of the inflight queries fail or crash the database in safe mode, those queries are added to the **denied\_list**. In addition, each query is copied to a file named `<query_ID>.txt` in the `<install_path>/internal/auto_restart/<timestamp>/denied_list` directory.

### Note

If an inflight query fails, none of the inflight queries are added to the warned list. Instead, the failed queries are added to the denied list.

- If AnzoGraph runs a query in safe mode and cannot determine if it should be added to the denied or warned list, those queries are copied to a file named `<query_ID>.txt` in the `<install_path>/internal/auto_restart/<timestamp>/unanalyzed_list` directory.
- Metadata about the `warned_list`, `denied_list`, and `unanalyzed_list` queries is captured in the **stc\_blocklist** system table.

### Note

The **auto\_restart\_directory** setting in the system configuration file, `<install_path>/config/settings.conf`, controls the location of the `auto_restart` directories listed above. For more information about the setting, see the [Automated Restart System Settings](#) section below.

4. After the inflight queries have been run, AnzoGraph restarts the database, loads the persisted data back into memory, and returns the system to normal operation.



To help prevent the circumstance that caused the database to crash, any queries that were added to the **denied** list are blocked from being executed when the system returns to normal operation. When a user runs a query, AnzoGraph compares that query with the denied list. If the query is on the list, the query is terminated and AnzoGraph returns an "Attempting to execute a denied-listed query" error message. Queries on the warned list are not blocked. A denied list query cannot be run unless it is removed from the denied list. This behavior is controlled by the **ignore\_deniedlist\_queries** setting. For more information about the setting, see the [Automated Restart System Settings](#) section below. For information about removing queries from the denied list, see [Removing a Query from the Block List](#) below.

## Automated Restart System Settings

The automatic restart feature is controlled by the following four settings in `<install_path>/config/settings.conf`:

- **auto\_restart\_max\_attempts**: This setting specifies the number of times the system manager should attempt to start the database after a crash. The default value is **5**, which means the system manager will attempt to restart the database a maximum of 5 times. Changing `auto_restart_max_attempts` to **0** disables the auto-restart feature.
- **auto\_restart\_time**: This setting specifies the number of seconds to spend attempting to restart the database. If all attempts fail and this time limit is reached, the system manager stops trying to restart the database. The default value is **600**, which means that the system manager will attempt to restart the database for a maximum of 600 seconds (10 minutes).
- **auto\_restart\_directory**: This setting specifies the base location of the **auto\_restart** directory, which contains the `denied_list`, `warned_list`, and `unanalyzed_list` directories. The default value is `<install_path>/internal`.
- **ignore\_deniedlist\_queries**: This setting controls whether denied list queries are blocked from running or are allowed to be run when the database is returned to normal operation. The default value is **false**, which means denied list queries are not ignored and are therefore blocked from running. If `ignore_deniedlist_queries` is **true**, incoming queries are not compared with the denied list and are run.

## Important

Changing the `auto_restart_max_attempts`, `auto_restart_time`, or `auto_restart_directory` values requires a restart of the system management daemon, `azgmgrd`, as well as the database. See [Starting and Stopping AnzoGraph](#) for instructions.

## Removing a Query from the Block List

AnzoGraph stores metadata about the denied and warned list queries in the `stc_blocklist` system table. To remove a query from either list, you remove the entry from the `stc_blocklist` table by running the `REMOVE_FROM_BLOCKLIST` command.

```
REMOVE_FROM_BLOCKLIST '<list_name>' <query_ID>
```

Where `<list_name>` is the name of the list that the query is on and `<query_ID>` is the ID number for the query. To retrieve the list name and query ID values, run the following query to return the `stc_blocklist` contents:

```
SELECT * WHERE { TABLE 'stc_blocklist' } ORDER BY ?blocklist
```

For example:

```
/opt/anzograph/bin/azgi -c "select * where {table 'stc_blocklist'} order by ?blocklist"
```

query	blocklist	updated	query_text	part
3587	denied_list	2020-08-25 14:29:27	select * from <http://an..	0
3592	denied_list	2020-08-25 14:29:32	select * where {?s ?p ?o}	0
3612	warned_list	2020-08-25 14:32:15	select * from <http://an..	0

In the results, the `<list_name>` is the value in the `blocklist` column, and `<query_id>` is the value in the `query` column. Running the following command removes the first entry from the `stc_blocklist` table, which removes that query from the denied list.

```
REMOVE_FROM_BLOCKLIST 'denied_list' 3587
```

# Anzo Admin CLI

The Anzo command line interface (CLI) utility, called **anzo**, is an advanced administration tool for managing Anzo. It is primarily used for debugging and inspecting system configurations, executing batch operations, and aiding in migrations and deployments. The topics in this section provide information about the CLI.

## Note

The CLI does not replace the user interface and does not support all operations available in the UI. To script user interface operations or control Anzo with the CLI, contact Cambridge Semantics.

### In this section:

<a href="#">Getting Started with the Admin CLI .....</a>	<a href="#">332</a>
<a href="#">Admin CLI Basics .....</a>	<a href="#">336</a>

# Getting Started with the Admin CLI

This topic provides instructions for configuring the admin command line interface, **anzo**, and viewing the help menu. The anzo client is in the `<install_path>/Client` directory.

## Important

The anzo CLI is an advanced administration tool for managing Anzo. It is primarily used for debugging and inspecting system configurations, executing batch operations, and aiding in migrations and deployments. To script user interface operations or control Anzo with the CLI, please contact Cambridge Semantics.

- [Add the CLI to the Anzo Service User PATH](#)
- [Configure the CLI](#)
- [View the CLI Help Menu](#)

## Add the CLI to the Anzo Service User PATH

Follow the instructions below to configure the PATH environment variable to include the Client directory so that the anzo CLI can be called from any directory.

1. If necessary, run the following command to become the Anzo service user:

```
sudo su - <anzo_user_name>
```

For example:

```
sudo su - anzo
```

2. Open `~/.bash_profile` in a text editor.
3. Change the PATH to the following value:

```
PATH=$PATH:$HOME/.local/bin:$HOME/bin:<install_path>/Client
```

For example:

```
PATH=$PATH:$HOME/.local/bin:$HOME/bin:/opt/Anzo/Client
```

4. Save and close the file, and then run the following command:

```
source ~/.bash_profile
```

5. Type **anzo** to verify that you can access the CLI. For example:

```
[anzo@anzo-server ~]$ anzo
Anzo Command Line Client.
Copyright (c) 2017 - 2023 Cambridge Semantics Inc and others.
All rights reserved.
Version: 5.4.3.r202307131505
Type anzo help for usage
```

## Configure the CLI

Follow the instructions below to configure a settings file that specifies the default CLI configuration values for parameters such as host, port, user, and password. Specifying these details in the settings file eliminates the need to include those options in subsequent commands.

To create and populate the settings file, **settings.trig**, in your home directory, run the following command:

```
anzo setup <options>
```

Where <options> include the following choices:

-beep , --beep	beep when command is completed
-ds , --datasource <datasource>	URI of the datasource to query, if other than primary datasource.
	Option not available for dataset queries.
-h , --host <hostname>	anzo server hostname
-http , --http	Use http connection to server.
-p , --port <int>	anzo server port
-pause , --pause-exit	Wait for a user key entry before an abnormal exit.
-ssl , --use-ssl	Use SSL for connection.
-t , --timeout <timeout>	override the default 30 second timeout for operations
-timer , --timer	Print out the total operation time
-trace , --show-trace	Show stack trace for errors.
-trust , --trust-all	Trust all certificates including invalid ones
-u , --user <string>	username to connect with

<code>-w , --password &lt;string&gt;</code>	user's password
<code>-x , --exclude-prefixes</code>	Do not use prefixes defined in user settings to expand options,
	arguments, or to write RDF.
<code>-z , --settings &lt;file&gt;</code>	override the default settings file location

For example:

```
anzo setup -h localhost -p 61616 -u sysadmin -w @nz0
```

Anzo creates the `settings.trig` file in the `~/user/.anzo` directory. You can edit the file as needed.

The installation also includes a sample settings file, `settings_example.trig`, in the `Client` directory. You can view the sample file for reference. For example:

```
### standard prefixes
@prefix foaf      : <http://xmlns.com/foaf/0.1/> .
@prefix rdfs     : <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc       : <http://purl.org/dc/elements/1.1/> .
@prefix xsd      : <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf      : <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
#### anzo prefixes:
@prefix cli      : <http://openanzo.org/cli/> .
@prefix system   : <http://openanzo.org/ontologies/2008/07/System#> .
@prefix anzo     : <http://openanzo.org/ontologies/2008/07/Anzo#> .
@prefix ld       : <http://cambridgesemantics.com/ontologies/2009/05/LinkedData#> .
@prefix anzowt   : <http://cambridgesemantics.com/ontologies/2009/05/AnzoWebToolkit#> .
@prefix reg      : <http://cambridgesemantics.com/registries/> .
@prefix ontserv  : <http://cambridgesemantics.com/semanticServices/OntologyService#> .
@prefix ldserv   : <http://cambridgesemantics.com/semanticServices/LinkedData#> .
cli:config {
  cli:config
#   system:user "" ;
#   system:password "" ;
  system:timeout "0";
  system:useSsl "false";
  system:port "61616";
  system:keystoreFile "${ANZO_CLI_HOME}/../Common/ssl/client.ks";
  system:keystoreType "JCEKS";
  system:keystorePassword "p@ssw0rd";
  system:truststoreFile "${ANZO_CLI_HOME}/../Common/ssl/client.ts";
  system:truststoreType "JCEKS";
  system:truststorePassword "p@ssw0rd";
}
```

```
}
```

## View the CLI Help Menu

The CLI help menu lists all of the available subcommands. To view the subcommands, run `anzo help`.

```
usage: anzo <subcommand> [options] [args]
Anzo Command Line Client.
Type 'anzo help <subcommand>' for help with a specific subcommand.
Available subcommands:
acls          Ensure the graphs in a dataset inherit their ACLs from the dataset
analyze       Provides several flavors of analysis for Anzo request/response logs
call          Calls an anzo semantic service and prints the service response to the
console
...
```

To view the help for a specific subcommand, run `anzo help <command>`. For example, the following command displays help for the `find` command:

```
[user@anzo Client]# ./anzo help find
usage: anzo find [options] [NAMED-GRAPH-URI...]
Retrieves statements from the server via simple pattern find.
-beep , --beep                beep when command is completed
-ds , --datasource <datasource>  URI of the datasource to query, if other than
primary datasource.
                                Option not available for dataset queries.
-f , --output-file <file>        write the find results to a file
-h , --host <hostname>          anzo server hostname
...
```

# Admin CLI Basics

This topic introduces the basic commands that are useful for viewing data in named graphs, running queries, and finding particular named graphs, subjects, predicates, or objects.

## Important

The anzo CLI is an advanced administration tool for managing Anzo. It is primarily used for debugging and inspecting system configurations, executing batch operations, and aiding in migrations and deployments. To script user interface operations or control Anzo with the CLI, please contact Cambridge Semantics.

- [Retrieving Named Graphs](#)
- [Finding Statements](#)
- [Running Queries](#)
- [Changing the Output Format](#)

## Retrieving Named Graphs

To retrieve named graphs from Anzo, use the `get` subcommand. This section lists example usage of `anzo get`. To view the help for `get`, run `anzo help get`.

Get a named graph:

```
anzo get http://film.com/films/PulpFiction
```

Get a named graph and save its contents to a file:

```
anzo get -f pulp.trig http://film.com/films/PulpFiction
```

Get a named graph and its metadata graph:

```
anzo get -m http://film.com/films/PulpFiction
```

Get only the metadata graph:

```
anzo get -M http://film.com/films/PulpFiction
```



## Finding Statements

To retrieve particular statements with a simple quad pattern match against the entire database, use the `find` subcommand. You can find statements by specifying one value in the quad (subject, predicate, object, or named graph) or any combination of values. Any URIs and/or literal values that you specify must match the value in the data. Partial values, wildcard characters, and regular expressions are not supported. This section lists example usage of `anzo find`. To view the help for `find`, run `anzo help find`.

Find all statements about a particular resource (subject):

```
anzo find -sub http://film.com/films/PulpFiction
```

Find all `rdf:type` statements for a particular type:

```
anzo find -pred rdf:type -uri film:Movie
```

Find all `rdf:type` statements for a particular type in a given named graph:

```
anzo find -pred rdf:type -uri film:Movie films:PulpFiction
```

Count all `rdf:type` statements for a particular type:

```
anzo find -n -pred rdf:type -uri film:Movie
```

## Running Queries

To run a SPARQL query against the database, use the `query` subcommand. This section lists example usage of `anzo query`. To view the help for `query`, run `anzo help query`.

Execute a query from a file. The named graphs to run against are defined in the query:

```
anzo query -f filmQuery.rq
```

Execute a query from a file against all named graphs:

```
anzo query -a -f filmQuery.rq
```

Execute a query specified on the command line (not recommended):

```
anzo query "SELECT ?name ..."
```

Execute a query from a file and pipe the results to another file:

```
anzo query -f film.rq > film.txt
```

## Changing the Output Format

The anzo CLI enables you to request results in the following formats: TriG (default), RDF, RDFS, XML, NT, N3, TTL, TriX, CSV, and JSON. To change the format for results, you use the `-o` option with subcommands such as `find`, `get`, and `query`.

For example, the following `get` subcommand returns data set details in XML format:

```
anzo get -o xml http://csi.com/FileBasedLinkedDataSet/059060234accd1d2d44b6bbb4207ee54
```

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ld="http://cambridgesemantics.com/ontologies/2009/05/LinkedData#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description
  rdf:about="http://csi.com/DataLocation/059060234accd1d2d44b6bbb4207ee54">
<fileConnection xmlns="http://cambridgesemantics.com/ontologies/DataSources#"
  rdf:resource="http://cambridgesemantics.com/File_Connection/local"/>
<filePath xmlns="http://cambridgesemantics.com/ontologies/DataSources#"
  /nfs/data/store/LoadMovies_223d3/</filePath>
<isPrimary xmlns="http://cambridgesemantics.com/ontologies/DataSources#"
  rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean">true</isPrimary>
<rdf:type
  rdf:resource="http://cambridgesemantics.com/ontologies/DataSources#DataLocation"/>
<rdf:type
  rdf:resource="http://cambridgesemantics.com/ontologies/DataSources#PathConnection"/>
</rdf:Description>
```

# Ongoing Operations

The topics in this section focus on administrative tasks that are suggested to be performed on a regular basis to monitor component connections, the overall usage and performance of the platform, and the quality of the data in the system.

**In this section:**

Platform Usage and Performance Monitoring .....	340
Data Quality Monitoring .....	347

# Platform Usage and Performance Monitoring

This topic provides queries that administrators can run to monitor platform usage and performance. You can run the queries from the Query Builder (See [Running SPARQL Queries in the Query Builder](#) in the User Guide) or against the SPARQL endpoint (See [Access the SPARQL Endpoint](#)). Each of the queries below target the System Datasource or System Tables data source, which are accessible to the sysadmin user or users with the Anzo Administrator role.

## Tip

You can generate a cURL request against the SPARQL endpoint from a query in the Query Builder. Once you have a valid query, click the **More** button under the query and select **Copy CURL Command**.

- Return the number of queries that were run in the last N hours
- Return the average and maximum runtime of the queries run in the last N hours
- Return the number of active graphmarts
- Return the number of triples that are loaded in each active graphmart
- Return the total number of dashboards
- Return file system usage details
- Return information about the last N events
- Return a list of queries that took longer than N seconds to run
- Return the number of queries run per day
- Return the number of active users per day
- Return the number of times each user logged in during the given time period

## Return the number of queries that were run in the last N hours

You can run the following aggregation query against the **System Tables** data source (<http://cambridgesemantics.com/datasource/SystemTables>) to return the number of queries that were executed in the specified number of hours:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX System: <http://openanzo.org/ontologies/2008/07/System#>
SELECT (COUNT(?query) as ?Nr_Of_Queries)
WHERE {
# VALUES ?hour { "N" }
# Example: in the last 5 hours
VALUES ?hour { "05" }
BIND(STRDT(CONCAT("P0Y0M0DT",?hour,"H00M00S"), xsd:duration) as ?duration)
?query a System:QueryExecution;
System:dateCreated ?time .
FILTER (?time + ?duration > NOW())
}
```

## Return the average and maximum runtime of the queries run in the last N hours

You can run the following aggregation query against the **System Tables** data source (<http://cambridgesemantics.com/datasource/SystemTables>) to return the average and maximum runtime of all of the queries that were executed in the specified number of hours:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX System: <http://openanzo.org/ontologies/2008/07/System#>
SELECT (AVG(?queryTime) as ?avgQueryTime) (MAX(?queryTime) as ?maxQueryTime)
WHERE {
# VALUES ?hour { "N" }
# Example: in the last 5 hours
VALUES ?hour { "05" }
BIND(STRDT(CONCAT("P0Y0M0DT",?hour,"H00M00S"), xsd:duration) as ?duration)
?query a System:QueryExecution;
System:queryTime ?queryTime;
System:dateCreated ?time .
FILTER (?time + ?duration > NOW())
}
```

## Return the number of active graphmarts

You can run the following query against the System Datasource

(<http://openanzo.org/datasource/systemDatasource>) to return the number of active graphmarts, i.e., the number of graphmarts with an active query engine:

```
PREFIX gmart: <http://cambridgesemantics.com/ontologies/Graphmarts#>
SELECT (COUNT(?graphmart) as ?runningGraphmartCount)
WHERE {
  ?graphmart a gmart:Graphmart ;
  # The running graphmarts are those with a query engine
  gmart:graphQueryEngineUri ?o .
}
```

## Return the number of triples that are loaded in each active graphmart

You can run the following query against the System Datasource

(<http://openanzo.org/datasource/systemDatasource>) to return the number of triples that are loaded in each graphmart that is active:

```
PREFIX System: <http://openanzo.org/ontologies/2008/07/System#>
PREFIX Graphmarts: <http://cambridgesemantics.com/ontologies/Graphmarts#>
PREFIX graphmartStatus: <http://cambridgesemantics.com/ontologies/GraphmartStatus#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?graphmart ?graphmartTitle ?statementCount
WHERE {
  ?graphmart dc:title ?graphmartTitle .
  SERVICE <http://cambridgesemantics.com/datasource/SystemTables> {
    SELECT ?graphmart ?statementCount
    FROM <http://openanzo.org/namedGraphs/reserved/graphs/ALL>
    WHERE {
      # Get the graphmarts that are online
      ?graphmart a graphmartStatus:GraphmartStatus ;
      graphmartStatus:status System:Online ;
      # Get their triple count
      graphmartStatus:totalStatements ?statementCount .
    }
  }
}
```

## Return the total number of dashboards

You can run the following aggregation query against the System Datasource

(<http://openanzo.org/datasource/systemDatasource>) to return the total number of dashboards in the system:

```
SELECT (COUNT(?dashboard) as ?dashboardCount)
WHERE {
  ?dashboard a
  <urn:com.cambridgesemantics.application.anzoweb.lens.linkeddataset.view.GraphmartViewLe
ns> .
  FILTER(?dashboard NOT IN
  (<urn:com.cambridgesemantics.application.anzoweb.lens.linkeddataset.view.GraphmartViewL
ens>))
}
```

## Return file system usage details

You can run the following query against the System Tables data source

(<http://cambridgesemantics.com/datasource/SystemTables>) to return information about used and free space on the specified file system locations:

```
SELECT ?dir ?free_gb ?available_gb ?total_gb
WHERE {
  ?s a <http://openanzo.org/ontologies/2008/07/System#FilesystemInfo> ;
  <http://openanzo.org/ontologies/2008/07/System#dirName> ?dir ;
  <http://openanzo.org/ontologies/2008/07/System#fsFree> ?free ;
  <http://openanzo.org/ontologies/2008/07/System#fsTotal> ?total ;
  <http://openanzo.org/ontologies/2008/07/System#fsAvailable> ?available .
  BIND(?free/1000000000 as ?free_gb)
  BIND(?total/1000000000 as ?total_gb)
  BIND(?available/1000000000 as ?available_gb)
  VALUES ?dir {
#    "location1"
#    [ "location2" ]
#    [ "... " ]
# Example with / and /mnt as locations:
#    "/"
#    "/mnt"
  }
}
```

## Return information about the last N events

You can run the following query against the System Tables data source

(<http://cambridgesemantics.com/datasource/SystemTables>) to return details about the N most recent activities:

```
SELECT DISTINCT ?desc ?event ?user ?startTime ?completedTime ?source
WHERE {
  ?s a <http://openanzo.org/ontologies/2008/07/System#ActivityAuditEvent> ;
    <http://purl.org/dc/elements/1.1/description> ?desc ;
    <http://openanzo.org/ontologies/2008/07/System#activitySource> ?source ;
    <http://openanzo.org/ontologies/2008/07/System#eventMessage> ?event ;
    <http://openanzo.org/ontologies/2008/07/System#userUri> ?user ;
    <http://openanzo.org/ontologies/2008/07/System#activityStarted> ?startTime ;
    <http://openanzo.org/ontologies/2008/07/System#activityCompleted> ?completedTime .
}
# LIMIT N
LIMIT 100
```

## Return a list of queries that took longer than N seconds to run

You can run the following query against the System Tables data source

(<http://cambridgesemantics.com/datasource/SystemTables>) to return a list of queries that took longer than N seconds to complete:

```
SELECT ?dateCreated (?queryTime/1000 as ?seconds) ?datasource
WHERE {
  {
    SELECT *
    WHERE {
      ?queryEvent a <http://openanzo.org/ontologies/2008/07/System#QueryEvent> ;
        <http://openanzo.org/ontologies/2008/07/System#queryTime> ?queryTime ;
        <http://openanzo.org/ontologies/2008/07/System#dateCreated> ?dateCreated ;
        <http://openanzo.org/ontologies/2008/07/System#datasourceUri> ?datasource .
      FILTER(CONTAINS(STR(?queryEvent), "queryStack"))
    }
  }
  ORDER BY DESC(?dateCreated)
  LIMIT 1000
}
# FILTER(?queryTime > number_of_milliseconds)
# For example, 10000 milliseconds=10 seconds:
```



```
FILTER(?queryTime > 10000)
}
```

## Return the number of queries run per day

You can run the following query against the System Tables data source

(<http://cambridgesemantics.com/datasource/SystemTables>) to return a count of the total number of queries that were run per day since the last time Anzo was restarted:

```
PREFIX System: <http://openanzo.org/ontologies/2008/07/System#>
SELECT ?date (COUNT(?query) as ?queryPerDayCount)
WHERE {
  ?query a System:QueryExecution;
  System:dateCreated ?time .
  BIND(DATEPART(?time) as ?date)
}
GROUP BY ?date
ORDER BY ?date
```

## Return the number of active users per day

You can run the following query against the System Tables data source

(<http://cambridgesemantics.com/datasource/SystemTables>) to return the number of active users per day since Anzo was last restarted:

```
PREFIX System: <http://openanzo.org/ontologies/2008/07/System#>
SELECT (COUNT(DISTINCT(?user)) as ?user_count)
FROM <http://openanzo.org/namedGraphs/reserved/graphs/ALL>
WHERE {
  {
    SELECT DISTINCT ?user ?date
    WHERE {
      ?event a system:QueryEvent ;
      System:userUri ?user ;
      System:dateCreated ?date .
    }
    ORDER BY DESC(?date)
    LIMIT 100
  }
}
```

## Return the number of times each user logged in during the given time period

You can run the following query against the System Tables data source

(<http://cambridgesemantics.com/datasource/SystemTables>) to return the number of times each user logged in between certain dates:

```
PREFIX System: <http://openanzo.org/ontologies/2008/07/System#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT DISTINCT ?user (COUNT(?loginDate) as ?loginCount)
WHERE {
    ?s a System:UserAuditEvent ;
        System:eventMessage ?eventMessage ;
        System:userUri ?user ;
        System:dateCreated > ?loginDate .
    FILTER(contains(?eventMessage, "Connect"))
# BIND(xsd:date("start_date") as ?startDate)
# BIND(xsd:date("end_date") as ?endDate)
# Specify the start and end dates to find the total number of logins
# between a range of time. For example, July 2023:
    BIND(xsd:date("2023-07-01") as ?startDate)
    BIND(xsd:date("2023-07-31") as ?endDate)
    FILTER(?loginDate > ?startDate && ?loginDate < ?endDate)
}
GROUP BY ?user
```

# Data Quality Monitoring

This topic provides queries that administrators can run to monitor the quality of the data based on the validity and completeness of the ontologies in the system. You can run the queries from the Query Builder (See [Running SPARQL Queries in the Query Builder](#) in the User Guide) or against the SPARQL endpoint (See [Access the SPARQL Endpoint](#)).

Each of the queries below target the System Datasource, which is accessible to the sysadmin user or users with the Anzo Administrator role.

## Tip

You can generate a cURL request against the SPARQL endpoint from a query in the Query Builder. Once you have a valid query, click the **More** button under the query and select **Copy CURL Command**.

- [Return properties that are missing a domain or range definition](#)
- [Return classes without class-specific properties](#)
- [Return similar classes](#)

## Return properties that are missing a domain or range definition

You can run the following query against the System Datasource (<http://openanzo.org/datasource/systemDatasource>) to check for properties that do not have a range or domain.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>

SELECT DISTINCT ?predicate
WHERE {
  GRAPH ?ont {
    ?ont a owl:Ontology .
    ?predicate a owl:ObjectProperty .
    OPTIONAL { ?ont <http://openanzo.org/2008/07/Anzo#isSystem> ?isSystem . }
    FILTER(!BOUND(?isSystem) || ?isSystem != "true").
  }
}
```

```

{ ?predicate a owl:ObjectProperty .
  FILTER NOT EXISTS {?predicate rdfs:domain ?domain} } UNION
{ ?predicate a owl:ObjectProperty .
  FILTER NOT EXISTS {?predicate rdfs:range ?range} }
}
}

```

## Return classes without class-specific properties

You can run the following query against the System Datasource

(<http://openanzo.org/datasource/systemDatasource>) to check for classes that do not have any properties that belong only to that class. The query finds classes that do not have properties or have only inherited properties.

```

PREFIX owl: <http://www.w3.org/2002/07/owl#>

SELECT DISTINCT ?class
WHERE {
  GRAPH ?ont {
    ?ont a owl:Ontology .
    ?predicate a owl:ObjectProperty .
    OPTIONAL { ?ont <http://openanzo.org/2008/07/Anzo#isSystem> ?isSystem . }
    FILTER(!BOUND(?isSystem) || ?isSystem != "true").
    ?class a owl:Class .
    FILTER NOT EXISTS { ?predicate rdfs:domain ?class } .
  }
}

```

## Return similar classes

You can run the following query against the System Datasource

(<http://openanzo.org/datasource/systemDatasource>) to check for classes that are similar based on the instances that are assigned to them.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>

SELECT DISTINCT ?class1 ?class2
WHERE {
  GRAPH ?ont { ?ont a owl:Ontology .

```

```
?predicate a owl:ObjectProperty .
OPTIONAL { ?ont <http://openanzo.org/2008/07/Anzo#isSystem> ?isSystem . }
FILTER(!BOUND(?isSystem) || ?isSystem != "true") .
?class1 a owl:Class .
?class2 a owl:Class .
?instance a ?class1, ?class2.
FILTER ( ?class1 != ?class2 ) .
FILTER NOT EXISTS { ?anything a ?class1 .
    FILTER NOT EXISTS { ?anything a ?class2 . } }
FILTER NOT EXISTS { ?anything a ?class2 .
    FILTER NOT EXISTS { ?anything a ?class1 . } }
}
}
```

# Troubleshoot

The topics in this section include an error message reference as well as instructions on retrieving diagnostic files and resolving certain issues.

## Tip

For information about monitoring Anzo usage on a regular basis, see [Platform Usage and Performance Monitoring](#).

### In this section:

<a href="#">Error Message Reference</a> .....	351
<a href="#">Investigating when Anzo is Unresponsive</a> .....	358
<a href="#">Updating an Expired License</a> .....	363
<a href="#">Restoring the Server ID</a> .....	364
<a href="#">Viewing the Current Stack in a Browser</a> .....	366
<a href="#">Taking AnzoGraph X-Rays from the Command Line</a> .....	368

# Error Message Reference

This topic describes the possible causes and solutions for error messages. Click a message in the list below to view details about that error:

- [File uses multiline records and cannot be segmented - disable segmenting for this file](#)
- [Elasticsearch exception: Data too large for \[http\\_request\]](#)
- [Error creating/mounting volume: User does not have read/write permission to location specified for new volume](#)
- [Error generating frame graph for ontology: <ontology\\_uri>  
java.util.NoSuchElementException](#)
- [LDAPException: Client request timed out or LdapException: searching for user](#)
- [QueryServlet - Could not flush org.eclipse.jetty.io.EofException](#)
- [CloudProviderException Failed to get node pools: Unable to execute HTTP request](#)
- [Unstructured Pipeline Errored: Cluster has no workers for pipelineWorkerService](#)
- [GqeAnzoException: Exception thrown from within an extension jvm -  
java.lang.OutOfMemoryError: Java heap space](#)
- [AnzoGraph: Cannot execute as user 'root'. Invalid user id](#)
- [AnzoGraph: Invalid Certificate](#)
- [AnzoGraph: Exception thrown from within the JVM - error -78 starting the JVM](#)
- [AnzoGraph: stg: Cannot allocate memory - heap is exhausted](#)
- [AnzoGraph: localhost:8256: Connection refused](#)
- [AnzoGraph: "Compilation Failed" at Startup](#)
- [AnzoGraph: Fatal Error. Caught Signal 15](#)

## File uses multiline records and cannot be segmented - disable segmenting for this file

This message indicates that you onboarded a CSV file that has embedded new lines and the Graph Data Interface (GDI) is unable to segment (partition) the file. Segmenting is enabled by default when using the automated direct load workflow and when running manually written GDI queries that do not explicitly disable segments. To resolve the issue, you have two options:

1. The most straightforward option is to edit the Direct Load Step query, even if it was automatically generated. To do so, open the step for editing and add the following content to the query below the `FileSource` object. After editing the query and saving the step, you can refresh or reload the layer or graphmart to onboard the file.

```
s:format [  
  s:segment false ;  
] ;
```

For example, the `s:segment` property is added to the `WHERE` clause below:

```
WHERE {  
  SERVICE <http://cambridgesemantics.com/services/DataToolkit>  
  {  
    ?data a s:FileSource ;  
    s:url "/opt/shared-files/data/csv/post_6_0/" ;  
    s:pattern "post_[0-9]*_[0-9]*.csv" ;  
    s:format [  
      s:delimiter "|" ;  
      s:segment false ;  
    ] ;  
    ...  
  }  
}
```

2. The second way to resolve the issue is to invoke the direct load workflow again. This time, expand the **Advanced** settings and disable the **Enable Partitioning** option. When you rerun the workflow with **Enable Partitioning** disabled, the GDI does not try to segment the file. This method can be more complicated than the first option, however, because the previous graphmart (or layer that was added to an existing graphmart) from the first run will not be overwritten by the second run. Another graphmart (or layer) is created, and the existing one with the error must be deleted or disabled.



## Elasticsearch exception: Data too large for [http\_request]

This message indicates that the Elasticsearch heap size is not large enough to process the request. By default, Elasticsearch is configured to use a maximum heap size of 1 GB. Cambridge Semantics recommends that you increase the amount to 50% of the memory that is available on the server. To change the configuration, open the `<elasticsearch_install_dir>/config/jvm.options` file in an editor. At the top of the file, modify the **Xms** and **Xmx** values to replace the **1** with the new value. For example:

```
# Xms represents the initial size of total heap space
# Xmx represents the maximum size of total heap space

-Xms15g
-Xmx15g
```

## Error creating/mounting volume: User does not have read/write permission to location specified for new volume

If a new volume is being created via the Administration application and the user who is creating the volume has read/write permission to the file system location, it is likely that this message was returned because a relative path to the volume location was used. To resolve the issue, type the absolute path to the location for the new volume.

## Error generating frame graph for ontology: <ontology\_uri> java.util.NoSuchElementException

If this message is returned when importing an ontology, it likely means that a multi-domain of a property is not constructed correctly as a list. For example, a multi-domain property may be defined incorrectly like the following example:

```
:isPublic a owl:DatatypeProperty, owl:FunctionalProperty ;
  rdfs:label "Is Public" ;
  restapi:apiIgnore true ;
  rdfs:domain   :ComponentBasedLinkedDataSet , :DatasetEdition ;
  rdfs:range   xsd:boolean .
```

The corrected syntax is below:

```
:isPublic a owl:DatatypeProperty, owl:FunctionalProperty ;
  rdfs:label "Is Public" ;
  restapi:apiIgnore true ;
  rdfs:domain [ a owl:Class ; owl:unionOf ( :ComponentBasedLinkedDataSet
:DatasetEdition ) ] ;
  rdfs:range xsd:boolean .
```

The ontology must be corrected and re-imported.

## **LDAPException: Client request timed out or LdapException: searching for user**

If users have intermittent problems logging in to Anzo and an LDAP timeout or "operations error" is logged, it may be that the connection to the directory server times out before the server can return the user authorization data to Anzo. To help resolve the issue, you can configure an LDAP heartbeat so that Anzo maintains the connection to the server. See [Setting a Heartbeat for LDAP Connections](#) for instructions.

## **QueryServlet - Could not flush org.eclipse.jetty.io.EofException**

Typically this error means that the client that called the Anzo SPARQL endpoint disconnected. This may mean that you ran a query that returns more results than the client can handle. Adding a LIMIT to the query may resolve the issue.

## **CloudProviderException Failed to get node pools: Unable to execute HTTP request**

This exception typically means that you are connecting to a Cloud Location and the firewall is blocking access to the pricing information from the cloud service provider. To resolve the issue, you can disable the retrieval of pricing information. See [Disabling Cloud Location Pricing Information](#) for instructions.

## **Unstructured Pipeline Errored: Cluster has no workers for pipelineWorkerService**

If the unstructured cluster is running but this exception is returned when an unstructured pipeline is run, it is likely that the connection to the cluster was lost and the worker nodes are reported as "Removed." To correct the issue, reconnect the cluster to Anzo by clicking the **Connect** button on

the Unstructured Clusters > Configuration tab in the Administration application.

### **GqeAnzoException: Exception thrown from within an extension jvm - java.lang.OutOfMemoryError: Java heap space**

This message indicates that you ran a query that calls an AnzoGraph Java extension and the AnzoGraph JVM is not allocated enough memory to execute the query. To resolve the issue, modify the AnzoGraph server configuration file, `<install_path>/config/settings.conf`, to increase the value for the `jvm_max_memory` setting. For instructions on changing `settings.conf`, see [Changing AnzoGraph Settings](#).

### **AnzoGraph: Cannot execute as user 'root'. Invalid user id**

This message indicates that you tried to start AnzoGraph as the root user and root access is disabled. Log in as the correct user, and then run the command again.

### **AnzoGraph: Invalid Certificate**

This message indicates that you replaced the default AnzoGraph certificates with your own trusted certificates and the certificates are invalid. Certificates can be invalid because they expired or they were generated or signed incorrectly. For information about replacing certificates, see [Replace the Default Self-Signed Certificates with Trusted Certificates](#) in the Deployment Guide.

### **AnzoGraph: Exception thrown from within the JVM - error -78 starting the JVM**

This error indicates that AnzoGraph cannot find the installed JVM and likely means that the `$JAVA_HOME` variable is not set. Typically AnzoGraph is started via systemd services and `$JAVA_HOME` is set in the `azgmgrd.service` unit. If AnzoGraph is started from the executables in `<install_path>/anzograph/bin`, however, the `$JAVA_HOME` variable must be set for the instance. To resolve the issue, start AnzoGraph with `systemctl` or set `$JAVA_HOME` on each instance in the cluster.

## AnzoGraph: stg: Cannot allocate memory - heap is exhausted

This message indicates that all of the memory that is available to AnzoGraph is in use and there is not enough left to run queries. The solution is to free memory by restarting AnzoGraph. Then either adjust your workload to reduce the amount of data that is loaded or increase the amount of RAM on the host server(s).

## AnzoGraph: localhost:8256: Connection refused

This message indicates that you sent a request to the SPARQL HTTPS port, such as when using the AnzoGraph CLI (`azgi`), and SSL protocol is disabled. SSL access is controlled by the `enable_ssl_protocol` setting. If HTTPS access is disabled and you want to enable it so that you can use the CLI, see [Changing AnzoGraph Settings](#) for instructions.

### Important

Enabling the SPARQL HTTPS protocol opens the standard SPARQL-compliant HTTPS endpoint. Unlike the Anzo gRPC protocol endpoint, the SPARQL HTTPS endpoint is encrypted but not authenticated.

## AnzoGraph: "Compilation Failed" at Startup

If AnzoGraph fails to start and you receive a "Compilation failed" message, it may indicate that some of the required GNU Compiler Collection (GCC) libraries are missing. Specifically, AnzoGraph requires the `glibc`, `glibc-devel`, and `gcc-c++` libraries. Typically when you install GCC by running `yum install gcc` those libraries are included as part of the package. In some cases, depending on the host server configuration, installing GCC excludes certain libraries. To install the missing libraries, run the following command and then start AnzoGraph again:

```
sudo yum install glibc glibc-devel gcc-c++
```

## **AnzoGraph: Fatal Error. Caught Signal 15**

This error indicates that a process external to AnzoGraph stopped the AnzoGraph processes, such as if the host machine was shut down while AnzoGraph was running. Restart AnzoGraph to proceed with normal usage.

# Investigating when Anzo is Unresponsive

If the Anzo server seems unresponsive or crashes, it is important to gather diagnostic information while the server is in the unresponsive state to aid in the investigation of the issue. This topic describes some of the symptoms you may encounter and includes recommendations on retrieving information to help diagnose the issue.

## Symptom

The Hi-Res Analytics and Anzo applications are inaccessible. When trying to reach the applications, you experience one of the following symptoms:

- A "404 not found" error.
- A "server connection was lost" message, and clicking refresh does not reconnect to the server.
- Pages remain blank and never load.

## Steps for Investigating the Cause

The suggestions below provide options for investigating the issue and collecting diagnostic information to review and, if necessary, send to Cambridge Semantics Support for further investigation.

- [Make sure that Anzo is running](#)
- [Check basic resource utilization](#)
- [View the Anzo log files](#)
- [View the JVM stack](#)
- [Investigate with the OSGi console](#)
- [Generate a thread dump](#)

## Make sure that Anzo is running

First, it is helpful to make sure that Anzo is still running. From the command line on the Anzo server, you can run the following command to check whether the `anzo` process is running: `ps aux | grep anzo`. If the process is not found, Anzo is not running and can be restarted with `sudo systemctl start anzo`.

## Check basic resource utilization

Next, check basic resource utilization to determine whether the server is close to the limit of memory or disk resources. From the command line, you can run the following commands:

- **top**: To display CPU usage and threads.
- **free g**: To display memory usage.
- **df -h**: To display disk usage.

If disks are close to full or memory usage is close to the limit, you may need to increase the resources on the server or delete files to reclaim disk space.

### Tip

There are ways that you can limit the disk space that is used to store logs and unstructured status journals if you run unstructured pipelines. See topics such as [Limiting the Age/Size of Audit Logs](#), [Limiting the Size/Number of anzo\\_full Logs](#), and [Limiting the Number of Unstructured Status Journals](#) in [Advanced Semantic Service Configuration](#).

### Note

Adding more memory to your machine does not automatically allocate more memory to Anzo. To increase memory for Anzo if you have added memory to the server, modify the `-XmxNNNm` value in the `AnzoServer.vmoptions` file in the `<install_path>/Server` directory. **It is important that you do not over-allocate memory.** Make sure that the value you specify is no more than 70 to 80 percent of the total memory on the server. After editing the file, restart Anzo to configure the change.

## View the Anzo log files

Review the Anzo log files for errors. The `anzo_full.log` contains the most comprehensive information, but the `_error` logs are more focused. For instructions on viewing the logs, see [Viewing Log Files](#).

## View the JVM stack

The `sysadmin` user can view the stack to see the current state of the JVM and look for deadlocks or blocked threads. For instructions, see [Viewing the Current Stack in a Browser](#).

## Investigate with the OSGi console

The `sysadmin` user can also SSH into OSGi console. To do so, follow the steps below:

### Note

If the OSGi console is inaccessible, retrieve a thread dump instead. See [Generate a thread dump](#) below.

1. On the Anzo server, make sure that you are logged in as the user that is running Anzo.
2. Run the following command to log in to the console:

```
ssh -p 8022 sysadmin@127.0.0.1
```

3. When prompted, enter the `sysadmin` password.
4. At the `osgi>` prompt, you can type `help` and press **Enter** to view the available commands. The list below describes the commands that are helpful for troubleshooting:

- **queries**: Lists any active local volume queries being run. This command only reports on active requests against a journal, not AnzoGraph. If there are long-running local volume queries, you can run `cancelQuery` to cancel them.
- **sysinfo**: Outputs information about total memory, free memory, max memory, and CPU usage.
- **gc**: Means "garbage collect." When run, Anzo frees memory that is not being used.



- **ss**: Displays a list of all installed bundles and their status. This command can be useful for locating and stopping a bundle that is stuck in the initialization phase and causing problems. To find a particular bundle, you can run `ss | grep <search_string_in_the_bundle_name>`. For example:

```
ss | grep asdl
```

Returns the bundle ID, the bundle status, and the bundle name. For example:

```
24    ACTIVE    com.cambridgesemantics.anzo.asdl.services_
5.4.5.r202311151600
```

To stop a bundle, run `ss <ID>`. For example, `ss 24`.

- **deadlock /local\_file\_path/file\_name.txt**: Outputs to the specified file information about deadlocked threads.
- **blocked /local\_file\_path/file\_name.txt**: Outputs to the specified file information about threads that are blocking other threads.
- **topStack /local\_file\_path/file\_name.txt**: Outputs to the specified file information about the top thread stacks (according to CPU usage).
- **stack /local\_file\_path/file\_name.txt**: Outputs to the specified file information about all thread stacks.
- **heap /local\_file\_path/file\_name.hprof**: Outputs a heap dump to the specified file. Though the difference between heap dumps and stack dumps is nuanced, in general heap dumps are useful for diagnosing memory issues if you suspect the issue is memory-related, and stack or thread dumps are useful for diagnosing CPU issues if you suspect the issue is CPU-related.

### Important

Heap dumps are very large. At a minimum, the generated file is several GB in size. Make sure that you save the file to a disk with plenty of available space. In addition, heap dumps may contain sensitive data that you might not want to share

outside of your organization. Stack and thread dumps do not contain any sensitive data.

## Generate a thread dump

If the OSGi console is unresponsive, you can retrieve a thread dump with a script that is included in the installation. To generate the thread dump files, run the `Anzo-detailed-thread-dump.sh` script in the `<install_path>/Server/scripts` directory. For example:

```
cd /opt/Anzo/Server/scripts
```

```
./Anzo-detailed-thread-dump.sh
```

The script output lists the names of the files that are generated in the `scripts` directory. For example:

```
Detailed thread dump information written to the following files:  
./anzo-thread-detail-jcmd-dump-2023-12-13-20-39-17.txt  
./anzo-thread-detail-native-dump-2023-12-13-20-39-17.txt
```

# Updating an Expired License

If your license is expired and Anzo will not start or Anzo is running but the Server Licensing screen in the Administration application shows an `Access Denied/Forbidden License is invalid` error message, follow the steps below to update the license key.

1. First, log in to the Anzo host server and stop Anzo with the following command:

```
sudo systemctl stop anzo-server
```

Or, if you do not use the systemd service, you can stop Anzo with the command below:

```
<install_path>/Anzo/Server/AnzoServer stop
```

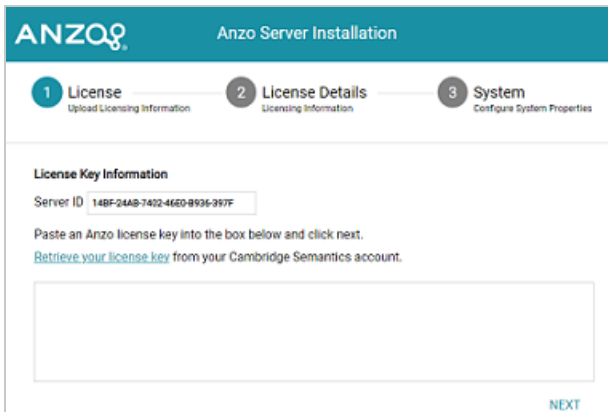
2. Next, start Anzo with the appropriate command below:

```
sudo systemctl start anzo-server
```

Or

```
<install_path>/Anzo/Server/AnzoServer start
```

3. Once Anzo is restarted, you will be presented with the same license key entry screen that is displayed during installation (shown below).



4. Paste the license key in the text field and click **Next**. Anzo completes the startup process and you can resume usage. It may take Anzo noticeably longer to start for the first time after the license is updated. Subsequent starts will return to the usual startup time.

# Restoring the Server ID

If Anzo is inadvertently updated by a user account that is different than the one used for the initial installation, the best way to resolve the issue is to revert the server ID to its original value by rolling back the update.

- If it was a new installation that used the wrong user account, uninstall Anzo. Then change to the correct user and run the installation script again.
- If your backup is a snapshot of the previous application disk, restore the disk. Then change to the correct user and update the installation.
- If it was an upgrade that used the wrong user account, follow the appropriate instructions below to restore Anzo from the backup that was saved before the upgrade:
  - [Restore from a Copy of the System Journal](#)
  - [Restore from a Copy of the Installation Directory](#)

## Restore from a Copy of the System Journal

1. Uninstall Anzo.
2. Change to the correct user account.
3. Reinstall the previous version of Anzo using the original installation script.
4. After the installation, replace the **anzo.jnl** file in the `install_path/Server/data/journal` directory with the backup version of the file.

At this point, Anzo is restored to the previous version and has the server ID that is associated with the license.

5. Now Anzo can be re-upgraded to the later release.

## Restore from a Copy of the Installation Directory

1. Uninstall Anzo.
2. Change to the correct user account.

3. Move the copy of the previous Anzo installation directory to the original location on the file system.

At this point, Anzo is restored to the previous version and has the server ID that is associated with the license.

4. Now Anzo can be re-upgraded to the later release.

## Viewing the Current Stack in a Browser

When the System Monitor service is configured to save heap and/or stack dumps (as described in [Enabling the System Monitor Service](#)), those dumps are saved to disk and cannot be viewed from the Administration application. However, the `sysadmin` user can quickly review the stack for the current state of the JVM in a browser. Follow the instructions below to view the stack.

### Note

Only a user with `sysadmin` access can view the stack in a browser. The `sysadmin` credentials are required to log in to the stack page.

To review the stack for the current state, go to the following URL in a browser:

```
https://<Anzo_server>:<HTTPS_admin_port>/status?stack
```

Where `<Anzo_server>` is the IP address or host name for the Anzo server and `<HTTPS_admin_port>` is the HTTPS port for the Administration application. For example:

```
https://10.11.0.12:8946/status?stack
```

The browser prompts you to log in as the `sysadmin` user. Supply the credentials and click **Sign in**.

The current state is displayed. For example:

```
2:Reference Handler
  Cpu: 0.00%
  Priority: 10 WAITING
  BlockedCount:487 BlockedTime:-1
  WaitedCount:482 WaitedTime:-1
  LockName:java.lang.ref.ReferenceLock@b3a29cf
  LockOwnerId:-1
  LockOwnerName:null
  LockClassName:java.lang.ref.ReferenceLock
  LockMonitors:
  LockSynchronizers:
  Stack:
    java.lang.Object.wait(Native Method)
    java.lang.Object.wait(Object,java:502)
    java.lang.ref.Reference.tryHandlePending(Reference,java:191)
    java.lang.ref.Reference$ReferenceHandler.run(Reference,java:153)

3:Finalizer
  Cpu: 0.00%
  Priority: 0 WAITING
  BlockedCount:1599 BlockedTime:-1
  WaitedCount:416 WaitedTime:-1
  LockName:java.lang.ref.ReferenceQueue$Lock@7941cc81
  LockOwnerId:-1
  LockOwnerName:null
  LockClassName:java.lang.ref.ReferenceQueue$Lock
  LockMonitors:
  LockSynchronizers:
  Stack:
    java.lang.Object.wait(Native Method)
    java.lang.ref.ReferenceQueue.remove(ReferenceQueue,java:144)
    java.lang.ref.ReferenceQueue.remove(ReferenceQueue,java:165)
    java.lang.ref.Finalizer$FinalizerThread.run(Finalizer,java:216)
```

You can also check specifically for blocked or deadlocked threads by replacing **stack** in the URL with **block** or **deadlock**. To check for blocked threads, go to the following URL:

```
https://<Anzo_server>:<HTTPS_admin_port>/status?block
```

For example:

```
https://10.11.0.12:8946/status?block
```

To check for deadlocks, go to the URL below:

```
https://<Anzo_server>:<HTTPS_admin_port>/status?deadlock
```

For example:

```
https://10.11.0.12:8946/status?deadlock
```

# Taking AnzoGraph X-Rays from the Command Line

If the Anzo Administration application is inaccessible and you need to generate AnzoGraph diagnostic files for troubleshooting, you can use the system manager on the AnzoGraph leader node to generate the files. The topics in this section provide instructions for generating diagnostic files using the system manager in AnzoGraph 2.5 and 3.1 releases. For instructions on retrieving diagnostic files from the Anzo Administration application, see [Retrieving AnzoGraph Diagnostic Files](#).

**In this section:**

<a href="#">Generating Diagnostic Files in AnzoGraph 2.5</a> .....	369
<a href="#">Generating Diagnostic Files in AnzoGraph 3.1</a> .....	371



## Generating Diagnostic Files in AnzoGraph 2.5

This topic provides instructions for using the system management CLI, **azgctl**, to retrieve diagnostic files for troubleshooting issues with an AnzoGraph 2.5 release. The CLI is in the `<install_path>/bin` directory. If you encounter an error and the database remains running, you take an x-ray to produce a tarball of the appropriate system diagnostic files. If you encounter an error that crashes the database, you generate a "crash dump" that retrieves the crash-related diagnostic files.

- [Taking an X-Ray](#)
- [Generating a Crash Dump](#)

### Taking an X-Ray

Run the following command on the leader server to take an x-ray on a running database. The result is a tarball that includes the historical system records.

```
azgctl -xray /path/name.xray
```

Where `/path/name` is the path on the file system where you want to save the file and the name of the tarball. All x-ray tarballs must include the `.xray` extension.

For example, the following command generates an x-ray named `query_error.xray` that is written to the `/tmp` directory.

```
/opt/bridgesemantics/anzograph/bin/azgctl -xray /tmp/query_error.xray
```

### Generating a Crash Dump

If you encounter an issue that stops the database, AnzoGraph automatically generates diagnostic files. Follow the instructions below to retrieve the files after a crash.

#### Note

The database does not need to be running to collect the crash dump.

1. Run the following command on the leader server to view a list of the available crash dumps.

```
azgctl -crashlist
```

For example:

```
/opt/bridgesemantics/anzograph/bin/azgctl -crashlist
```

The results show a list of available crash dumps by timestamp. For example:

Crash ID	Time
520460982	2023-06-28 20:30:35
520457655	2023-06-28 20:28:25

2. Run the following command to retrieve the appropriate crash files. This command creates a tarball that includes the diagnostic files:

```
azgctl -crashfetch [ crash_id ] /path/name.xray
```

Include the `crash_id` when you want to retrieve a specific crash dump that is listed in the crash list. Omit the crash ID to retrieve the latest files. All crash dump tarballs must include the `.xray` extension.

For example, the following command captures the most recent crash files. The tarball is named `latest_crash.xray` and it is saved to the `/tmp` directory.

```
/opt/bridgesemantics/anzograph/bin/azgctl -crashfetch /tmp/latest_crash.xray
```

The example below captures the crash dump for ID 520457655:

```
/opt/bridgesemantics/anzograph/bin/azgctl -crashfetch 520457655 /tmp/crash_520457655.xray
```

### Tip

You can run `azgctl -crashtoss` to remove all crash dumps from the server.

## Generating Diagnostic Files in AnzoGraph 3.1

This topic provides instructions for using the system management CLI, **azgctl**, to retrieve diagnostic files for troubleshooting issues with an AnzoGraph 3.1 release. The CLI is in the `<install_path>/bin` directory. If you encounter an error and the database remains running, you take an x-ray to produce a tarball of the appropriate system diagnostic files. If you encounter an error that crashes the database, you generate a "crash dump" that retrieves the crash-related diagnostic files.

- [Taking an X-Ray](#)
- [Generating a Crash Dump](#)

### Taking an X-Ray

Run the following command on the leader server to take an x-ray on a running database. The result is a tarball that includes historical system records from the specified time period. All flags for specifying a time period are optional. If you omit the options, the resulting x-ray will include the last 24 hours of historical system data.

#### Note

The system manager interprets time specifications using the system's local time and converts the timestamps to UTC when starting the x-ray.

```
azgctl -xray /path/name.xray [ -f <time> ] [ -t <time> ] [ -d <num_days> ]  
[ -h <num_hours> ] [ -m <num_minutes> ]
```

Option	Description
<code>/path/name</code>	The path on the file system where you want to save the tarball and the name of the tarball. All x-rays must be named with the <code>.xray</code> extension.
<code>-f &lt;time&gt;</code>	The <code>-f &lt;time&gt;</code> (or <code>--from &lt;time&gt;</code> ) option can be used to specify the time to start the system data capture, i.e., omit all of the records from before the specified time. Time must be specified in the following format: <code>YYYY-MM-DD[:HH[:MM]]</code> . For example, <code>-f 2024-01-10:15:00</code> sets the start time to 3:00 p.m.

Option	Description
	(local system time) on January 10, 2024.
<b>-t &lt;time&gt;</b>	<p>The <code>-t &lt;time&gt;</code> (or <code>--to &lt;time&gt;</code>) option can be used to specify the time to end the system data capture, i.e., omit all of the records after the specified time. Time must be specified in the following format: <code>YYYY-MM-DD[:HH[:MM]]</code>. For example, <code>-t 2024-01-09:19:30</code> sets the end time to 7:30 p.m. (local system time) on January 9, 2024.</p>
<b>-d &lt;num_days&gt;</b>	<p>The <code>-d &lt;num_days&gt;</code> (or <code>--days &lt;num_days&gt;</code>) option can be used to specify the number of days to include in the x-ray. The value must be a positive integer.</p> <ul style="list-style-type: none"> <li>• When combined with <code>-t &lt;time&gt;</code> or <code>-f &lt;time&gt;</code>, the number of days is relative to the <code>from</code> or <code>to</code> value. For example, <code>-f 2024-01-10 -d 2</code> means two days starting from 1/10/24 (i.e., 1/10/24 – 1/12/24). And <code>-t 2024-01-10 -d 2</code> is two days before 1/10/24 (i.e., 1/8/24 – 1/10/24).</li> <li>• When included without <code>-f</code> or <code>-t</code>, the number of days is relative to the current local system time. For example, <code>-d 2</code> captures the last 2 days of data starting from <code>now()</code>.</li> </ul>
<b>-h &lt;num_hours&gt;</b>	<p>The <code>-h &lt;num_hours&gt;</code> (or <code>--hours &lt;num_hours&gt;</code>) option can be used to specify the number of hours to include in the x-ray. The value must be a positive integer.</p> <ul style="list-style-type: none"> <li>• When combined with <code>-t &lt;time&gt;</code> or <code>-f &lt;time&gt;</code>, the number of hours is relative to the <code>from</code> or <code>to</code> value. For example, <code>-f 2024-01-10:12:00 -h 5</code> means the 5 hours after 12:00 p.m. on 1/10/24. And <code>-t 2024-01-10:12:00 -h 5</code> means the 5 hours before 12:00 p.m. on 1/10/24.</li> <li>• When included without <code>-f</code> or <code>-t</code>, the number of hours is relative to the current local system time. For example, <code>-h 3</code> captures the last 3 hours of data starting from <code>now()</code>.</li> </ul>

Option	Description
<b>-m &lt;num_minutes&gt;</b>	<p>The <code>-m &lt;num_minutes&gt;</code> (or <code>--minutes &lt;num_minutes&gt;</code>) option can be used to specify the number of minutes to include in the x-ray. The value must be a positive integer.</p> <ul style="list-style-type: none"> <li>When combined with <code>-t &lt;time&gt;</code> or <code>-f &lt;time&gt;</code>, the number of minutes is relative to the <code>from</code> or <code>to</code> value. For example, <code>-f 2024-01-10:12:00 -m 30</code> means the 30 minutes after 12:00 p.m. on 1/10/24. And <code>-t 2024-01-10:12:00 -m 30</code> means the 30 minutes before 12:00 p.m. on 1/10/24.</li> <li>When included without <code>-f</code> or <code>-t</code>, the number of minutes is relative to the current local system time. For example, <code>-m 30</code> captures the last 30 minutes of data starting from <code>now()</code>.</li> </ul>

## Examples

The following example generates an x-ray that includes the last 24 hours of system data. A tarball named `24hr_errors.xray` is written to the `/tmp` directory.

```
/opt/cambridgesemantics/anzograph/bin/azgctl -xray /tmp/24hr_errors.xray
```

The example below captures the last 12 hours worth of data. A tarball named `last12hours.xray` is written to the `/tmp` directory.

```
/opt/cambridgesemantics/anzograph/bin/azgctl -xray /tmp/last12hours.xray -h 12
```

The example below captures the last two days of data from before 5:00 p.m. on 1/18/24. A tarball named `1-16_to_1-18.xray` is written to the `/opt/shared/xrays` directory.

```
/opt/cambridgesemantics/anzograph/bin/azgctl -xray /opt/shared/xrays/1-16_to_1-18.xray
-t 2024-01-18:17:00 -d 2
```

## Generating a Crash Dump

If you encounter an issue that stops the database, AnzoGraph automatically generates diagnostic files. Follow the instructions below to retrieve the files after a crash.

## Note

The database does not need to be running to collect the crash dump.

1. Run the following command on the leader server to view a list of the available crash dumps.

```
azgctl -crashlist
```

For example:

```
/opt/cambridgesemantics/anzograph/bin/azgctl -crashlist
```

The results show a list of available crash dumps by timestamp. For example:

Crash ID	Time
520460982	2023-12-28 20:30:35
520457655	2023-12-28 19:01:25

2. Run the following command to retrieve the appropriate crash files. This command creates a tarball that includes the diagnostic files:

```
azgctl -crashfetch [ crash_id ] /path/name.xray
```

Include the `crash_id` when you want to retrieve a specific crash dump that is listed in the crash list. Omit the crash ID to retrieve the latest files. All crash dump tarballs must include the `.xray` extension.

## Examples

The following command captures the most recent crash files. The tarball is named `latest_crash.xray` and it is saved to the `/tmp` directory.

```
/opt/cambridgesemantics/anzograph/bin/azgctl -crashfetch /tmp/latest_crash.xray
```

The example below captures the crash dump for ID 520457655:

```
/opt/cambridgesemantics/anzograph/bin/azgctl -crashfetch 520457655 /tmp/crash_520457655.xray
```

## Tip

You can run `azgctl -crashtoss` to remove all crash dumps from the server.