

Anzo[®] 5.4 Getting Started Guide

Last Updated: 1/29/2024

Online documentation is available at docs.cambridgesemantics.com

Table of Contents

- About This Doc 3
- Anzo Concepts and Vocabulary 4
 - Glossary 5
 - Graphmart Concepts 11
 - Access Control Concepts 13
 - Graph Storage Concepts 14
- Anzo Application Overview 17
 - Accessing the Application 18
 - Application Overview 19
 - General Interface Elements 20
 - Onboard 24
 - Structured Data 25
 - Model 28
 - Blend 29
 - Access 31
- Tutorials 34
 - Building a Knowledge Graph from a CSV File 35
 - Prepare the Environment 36
 - Download the Sample Data 37
 - Add a CSV Data Source for the Sample Data 38
 - Create a Graphmart from the Data Source 43
 - Explore the Data in a Hi-Res Analytics Dashboard 45

About This Doc

The Getting Started Guide helps you get started with using Anzo. The topics in this guide introduce Anzo concepts, provide an overview of the user interface, give guidance on basic setup tasks, and include tutorials on building a sample knowledge graph from scratch.

Tip

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

The following list describes the sections in this guide.

- [Anzo Concepts and Vocabulary](#): Introduces key features, concepts, and vocabulary to know when working with Anzo.
- [Anzo Application Overview](#): Includes instructions for accessing the Anzo user interface and a walkthrough of the design, layout, and general interface elements.
- [Tutorials](#): Gives guidance on building a knowledge graph from scratch using the automated direct data load workflow.

Anzo Concepts and Vocabulary

The topics in this section introduce you to key features, concepts, and vocabulary to know when working with Anzo.

In this section:

- Glossary 5
- Graphmart Concepts11
- Access Control Concepts13
- Graph Storage Concepts14

Glossary

This topic defines commonly used Anzo terms and phrases.

- [Anzo Data Store](#)
- [Artifacts](#)
- [Blend](#)
- [Data Layers](#)
- [Datasets Catalog](#)
- [ELT](#)
- [File-Based Linked Data Set](#)
- [File Store](#)
- [Frame Graph](#)
- [Graph Data Interface](#)
- [Graphmarts](#)
- [Hi-Res Analytics](#)
- [IRI](#)
- [Journal or Volume](#)
- [Linked Data Set](#)
- [Model](#)
- [Managed Model](#)
- [Named Graph](#)
- [NLP](#)
- [OData](#)
- [OSGi](#)
- [Onboard](#)
- [Registry](#)
- [URI](#)

Term	Definition
Anzo Data Store	An Anzo data store defines an endpoint for writing data. It specifies the directory on the file store where Anzo can generate file-based linked data sets for unstructured pipelines (see File-Based Linked Data Set). It can also be used as a designated a directory to use for saving data sets created from graphmart exports. The data store also defines write properties such as the maximum file size and whether files should be compressed.
Artifacts	Artifacts are all of the objects that are created in Anzo during initial configuration

Term	Definition
	and the data onboarding process. For example, when you connect to a database or 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.
Blend	Using semantic models, separate data sets or new data sources can be blended into a knowledge graph. Anzo can combine and align any data set as well as apply data cleansing and/or transformation steps. Anzo delivers blending and access through Graphmarts , which give users the flexibility to combine and analyze any subset of data in Anzo.
Data Layers	Data layers enable users to enhance graphmarts dynamically. Users can create layers to load additional data sets, clean, conform, or transform data, infer new information, or export data to a file-based linked data set (FLDS).
Datasets Catalog	<p>Anzo's Dataset catalog combines traditional technical, operational, and business metadata with a semantic layer to describe all aspects of enterprise data elements. The catalog enables Anzo's unique use of semantics and graph models and is the system of record for data in Anzo.</p> <p>Anzo collects and generates metadata at every stage in the data discovery and integration process. Metadata in the catalog documents how data is converted during the onboarding process from its original format into a graph model. Subsequent data blending, transformation, and preparation steps are captured as additional metadata. Anzo also captures new metadata to describe all actions taken against data within Anzo. The metadata enables users to visualize their data, understand business contexts, identify connections, and blend and prepare data.</p>
ELT	Structured and semi-structured data sources are onboarded using an extract, load, and transform (ELT) workflow as opposed to a traditional ETL flow. Data Layers are Anzo's mechanism for flexibly transforming data in memory.

Term	Definition
File-Based Linked Data Set	<p>A File-Based Linked Data Set (FLDS) is a Named Graph that contains a collection of ontologies and the location of RDF data files that share common structure, purpose, meaning, or permissions.</p> <p>When the unstructured pipeline workflow is used to onboard data, Anzo creates a dataset in the Datasets catalog. The dataset in the catalog is registered in the Anzo system data source (see Journal or Volume) and includes metadata about the data, including a pointer to the data store location for the RDF files generated by the pipeline. The catalog dataset and the files on disk are an FLDS. An FLDS is also generated when an Export Step is included in a graphmart.</p>
File Store	<p>The Anzo platform components, AnzoGraph, Anzo Unstructured, and Elasticsearch share a file system for maintaining onboarded data and supporting files. A file store is the shared file storage system, such as NFS, HDFS, or cloud storage, that is shared between the servers.</p>
Frame Graph	<p>Each ontology in Anzo has a corresponding frame ontology or frame graph. A frame graph is generated when a new ontology is added, and it is regenerated each time the ontology is modified. By generating a frame graph, Anzo can pre-process the ontology rather than waiting until runtime to do the calculations. During frame graph generation Anzo performs activities like finding all of the properties available to a class, identifying the superclasses of each class, and determining whether a property is required or multi-valued.</p>
Graph Data Interface	<p>The Graph Data Interface (GDI) (sometimes called the Data Toolkit) is a flexible AnzoGraph extension that enables users to access a variety of data sources via SPARQL queries. The GDI has built-in, native support for various file format types, HTTP/REST endpoints, and JDBC connections to common database sources. For more information about the GDI, see Introduction to the GDI in the User Guide.</p>

Term	Definition
Graphmarts	Graphmarts are collections of knowledge graphs that users can blend and enhance. Graphmarts can combine any subset of data in Anzo for analysis. For more information about graphmarts, see Graphmart Concepts .
Hi-Res Analytics	Hi-Res Analytics enable users to explore and ask questions across all of their data. Using model-guided dashboards, users can perform computations across multi-dimensional data. Hi-Res Analytics dashboards generate complex graph queries dynamically based on user input.
IRI	An Internationalized Resource Identifier (IRI) is similar to URI but allows a greater range of characters. URI and IRI are often used interchangeably.
Journal or Volume	A journal, also known as a volume, refers to data that is stored in Anzo's embedded graph store. The graph store is transactional and is used to persist metadata, which is written to disk in a .jnl file. The system volume (or system data source) is the default, required volume where Anzo stores ontologies as well as system configuration, dataset, catalog, registry, and access control metadata. Users can create secondary local volumes that are used for more compartmentalized data and can be created and deleted without affecting the core system.
Linked Data Set	A linked data set (LDS) is a fundamental concept. Anzo organizes all data, including system data, into linked data sets. An LDS is associated with a data model and can be searched, discovered, shared, and protected with access control. For example, graphmarts are organized in a linked data set or registry of graphmarts, the Activity Log is a linked data set, data source configurations exist in a linked data set, and so on.
Model	Anzo establishes the semantic layer by enabling users to convert diverse enterprise data models into graph data models and then enhance the data by adding new business definitions, names, and tags. Further insight is added when data from separate sources are linked, connecting shared business

Term	Definition
	definitions across previously siloed sources. Anzo employs open World Wide Web Consortium (W3C) standards, including Web Ontology Language (OWL), RDF, and SPARQL to model, connect, and query interconnected graphs. For more information, see Model Concepts and Vocabulary in the User Guide.
Managed Model	A managed model is a model that is generated by the Graph Data Interface (GDI) and owned and managed by a data layer with a Direct Load Step. Any manual changes made to a managed model outside of a Direct Load step, such as from the Model editor, will be overwritten any time the graphmart or layer is refreshed or reloaded. For more information, see Managed Model Concepts in the User Guide.
Named Graph	Anzo implements the RDF Named Graph abstraction. These are the atomic units of storage in Anzo. Each named graph can be access controlled, and each graph has a corresponding “metadata graph” that includes the access control information, the last modified date, and which user created and modified the associated named graph. For more information about named graph storage, see Graph Storage Concepts .
NLP	Anzo performs named-entity recognition (NER) using knowledge bases and can interface with natural language processing (NLP) tools. It serves as a platform that enables text analytics through interplay between best-of-breed NLP tools.
OData	Open Data Protocol (OData) facilitates the creation of interoperable RESTful APIs. The Anzo Data on Demand service provides OData-based feeds that can be used to query graphmart data from third-party business intelligence tools.
OSGi	The Open Service Gateway Initiative (OSGi) is the open-standard architecture upon which Anzo is built. It is a Java framework for developing and deploying software programs and libraries. OSGi enables Cambridge Semantics to compartmentalize Anzo into "bundles" that can be deployed, activated, and removed independently without affecting other bundles in the system.

Term	Definition
Onboard	When data is ingested from its source platform to Anzo, it is converted from its original format to a new format that describes the data as a graph data model. This format, Resource Description Framework (RDF), captures each data value and relationship. RDF data is loaded to Anzo's in-memory graph engine, AnzoGraph for transformation and analysis.
Registry	Anzo manages configurations in system-level registries. Each registry is a collection of application and system component configurations of the same type. Like data, registries are stored and managed with RDF named graphs according to ontologies. Technically, a registry is a Linked Data Set .
URI	A Uniform Resource Identifier (URI) is a globally unique identifier for a piece of information. A URL (Uniform Resource Locator) is a URI that specifies a location, such as a web address.

Graphmart Concepts

Graphmarts are containers that enable users to create, enhance, and manage knowledge graphs. Graphmarts are made up of data layers. Each data layer in a graphmart is one named graph. Data layers are containers for steps. Users add steps to layers to enhance graphmarts dynamically. The steps in a layer perform all of the data operations. They are used to load additional datasets, infer new data automatically, or run SPARQL queries to create views or clean, conform, transform, or validate the data in the layer. The sections below provide details about layers and steps.

Data Layers

The following list introduces the key features of data layers:

- You can create any number of layers in a graphmart and control which users and groups have access to the layers and the data within the layers.
- Users can toggle data layers on and off in Hi-Res Analytics dashboards.
- You can configure execution conditions on layers so that the layer is processed only when certain conditions are met.
- You cannot share data layers between graphmarts, but you can clone layers and include a copy in multiple graphmarts.
- You control the source data for steps in a layer. Layers can build upon the data generated by steps in previous layers or can be self-contained, applying changes only to the data defined in the layer.

Steps

You can add any number of steps to a layer and can create the following types of steps:

- **Direct Load Step:** Runs a SPARQL query that invokes the Graph Data Interface (GDI) RDF and Ontology Generators to directly load a data source to a graphmart.
- **Elasticsearch Indexing Step:** Creates an Elasticsearch index to associate with a layer.

- **Elasticsearch Snapshot Step:** Creates an Elasticsearch snapshot of the index that is associated with a layer.
- **Export Step:** Exports the graphmart data in memory to a file-based linked data set (FLDS).
- **Load Dataset Step:** Loads a dataset from the Datasets catalog into a data layer graph. When you create a graphmart from the catalog, Anzo automatically creates a layer with a Load Dataset Step.
- **Pre-compile Query Step:** Runs the included query immediately after a graphmart is loaded so that the query is pre-compiled by AnzoGraph. Pre-compiling a query reduces execution time when a user runs that query for the first time.
- **Query Driven Templated Step:** Enables users to create reusable query-driven templates for quickly creating additional query steps. Unlike the Templated Step, where users define each key-value pair, this step runs a query to identify all of the key-value pairs. Then the template query is run for each key-value solution from the first query.
- **Query Step:** Provides a SPARQL query template that you can use for writing a query that creates, cleans, conforms, or transforms data in the data layer.
- **RDFS+ Inference Step:** Uses RDFS and OWL rules to generate new data in a layer based on the vocabularies in the existing data.
- **Templated Step:** Enables users to create reusable templates for quickly creating additional query steps. The query in a Templated Step uses parameters to represent key-value pairs. When reusing the step, users modify the values for the keys rather than rewriting the query.
- **Validation Step:** Enables users to write a query that validates the data in a data layer and is a required component when setting up an execution condition.
- **View:** Enables users to write a SPARQL CONSTRUCT query that defines a view of the data but does not alter the source data or create new data unless you choose to materialize the view.

For more information about graphmarts, layers, and steps, see [Working with Graphmarts](#) in the User Guide.

Access Control Concepts

Data access control in Anzo is an aggregation of three concepts:

1. [Default Access Policies](#)
2. [Permission Inheritance](#)
3. [Artifact Sharing](#)

Default Access Policies

These are the base permissions that are applied to artifacts by default when they are created. Artifacts are all of the objects that are created in Anzo during the initial system configuration and the data onboarding process. For example, when a user connects to a database or 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. 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, permission inheritance and sharing.

Permission Inheritance

To facilitate common workflows, Anzo 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 its permissions from the data source. This permission inheritance is applied in addition to the applicable Default Access Policy.

Artifact Sharing

When you onboard data to Anzo, you become the creator of several types of [Artifacts](#). Artifacts that you create can be shared with other groups (or users) from the **Sharing** tab (for that artifact) in the Anzo application. When an artifact is shared, those user-configured permissions are applied in addition to any permissions that were inherited.

For more detailed information about the access control concepts as well as user management, see in the Administration Guide.

Graph Storage Concepts

This topic describes the way onboarded graph data is shared between and stored in the Anzo and AnzoGraph graph stores.

The onboarding process generates different types of graph data artifacts. Storage of the artifacts differs based on the type of data that is being stored and the purpose of the data. The list below describes the artifacts and storage methods:

- The metadata, such as data models, data source configuration details, datasets catalog entries, registries, and access control definitions, are stored in Anzo's embedded graph store. The Anzo graph store is a transaction-oriented store that is built for processing many updates to small amounts of data. Data is persisted to disk in a journal, also known as a volume. The system volume (or system data source) is the default, required volume where Anzo stores models as well as system configuration, dataset, registry, and access control metadata. Users can create secondary local volumes that are used for more compartmentalized data and can be created and deleted without affecting the core system.
- When an unstructured pipeline is run or data is exported from a graphmart, the instance data and copies of the data models are written to a file-based linked data set (FLDS) on the shared file store. Each FLDS is represented as a dataset in Anzo's Datasets catalog. The catalog entry includes a pointer to the RDF files on disk.
- When a dataset from the catalog is added to a graphmart and the graphmart is activated, Anzo loads the data from the FLDS into the AnzoGraph graph store. Once the data is in memory, the rest of the graphmart's data layer steps are executed by AnzoGraph (known as the ELT process). Each data layer becomes a graph in AnzoGraph, and each layer graph includes the instance data created by that layer as well as the related models.
- Anzo system ontologies and metadata remain in Anzo's graph store, the system data source. They are not loaded to AnzoGraph unless the system data is added to a graphmart and the graphmart is activated.

- AnzoGraph does not have a "graphmart" or "step" construct, and graphmart and step URIs do not exist in the database. Though a graphmart acts as a container for data layers and its metadata can be queried in Anzo's embedded graph store, it does not include instance data that is needed by AnzoGraph.

Typically organizations manage all data with Anzo, i.e., data is onboarded to Anzo through unstructured pipelines or the Graph Data Interface. Anzo then loads the data to AnzoGraph for analytics. When data is loaded to AnzoGraph through Anzo, Anzo manages the reloading of graphmarts if AnzoGraph is restarted. Though users can load data and create named graphs directly in AnzoGraph, AnzoGraph is not configured by default to persist the data in memory to disk. Graphs that do not originate in Anzo must be reloaded manually any time AnzoGraph is restarted.

Example

An Anzo instance has two active graphmarts. Each graphmart has two data layers, one for loading datasets into memory and another for creating views and running ELT queries. When the following query is run against AnzoGraph to return a list of all distinct graphs, the results show that there are five graphs:

```
SELECT DISTINCT ?graph
WHERE {
  GRAPH ?graph {
    ?s ?p ?o
  }
}
```

graph

```
-----
http://cambridgesemantics.com/Layer/546fb89ac6d245f8bea2777a52077bc9
http://cambridgesemantics.com/Layer/1162fb0d0b724a18b4133c10d69f16b7
http://cambridgesemantics.com/Layer/12c7eedddff9449ab4b133373b56e65c
http://cambridgesemantics.com/Layer/b69bb3295ba3434e846b1ed372039416
http://cambridgesemantics.com/GqeDatasource/guid_10492203b5aa4a54f217ababb3dc6dee
5 rows
```

The first four graphs are the data layers for the two graphmarts. The graph URIs match the data layer URIs in Anzo. The last graph in the results above is the AnzoGraph data source graph. This graph contains one triple that records a timestamp for the last time the data source was updated. If

Anzo loses the connection to AnzoGraph, it checks this timestamp when it reconnects. The last updated time is used to determine whether the Anzo and AnzoGraph graph stores are in sync or if the graphmarts need to be reloaded to AnzoGraph.

Anzo Application Overview

This topic provides instructions for accessing the Anzo user interface and gives a general overview of the design and layout. For information about the Anzo concepts and procedures that influence the user interface's design and use, see [Anzo Concepts and Vocabulary](#).

Accessing the Application

To access the user interface:

1. Go to the following URL in your browser:

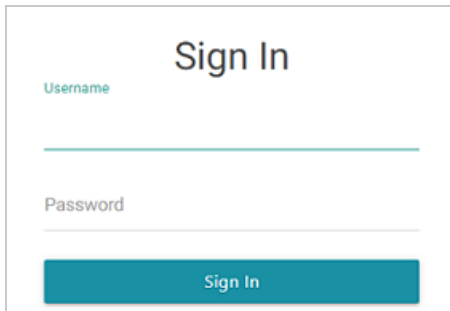
```
https://<hostname>
```

Where <hostname> is the Anzo server DNS name or IP address.

Note

Your browser may warn you that the server's SSL certificate has not been signed by a trusted authority. This is normal behavior. To continue, click the **Advanced** link at the bottom of the page and then click the **Proceed** link. If you require a trusted site certificate, you can obtain one from a Certificate Authority and add it to the Anzo server.

2. On the Sign In screen, type your username and password and then click **Sign In**.

A mockup of a 'Sign In' web form. The title 'Sign In' is centered at the top. Below it, the label 'Username' is positioned above a text input field. Further down, the label 'Password' is positioned above another text input field. At the bottom of the form is a teal-colored button with the text 'Sign In' in white.

Application Overview

The user interface is organized and compartmentalized by the processes involved with building solutions. The design accommodates use cases where users with different permissions and responsibilities build various parts of a solution. The sections below introduce you to each of the user interface elements.

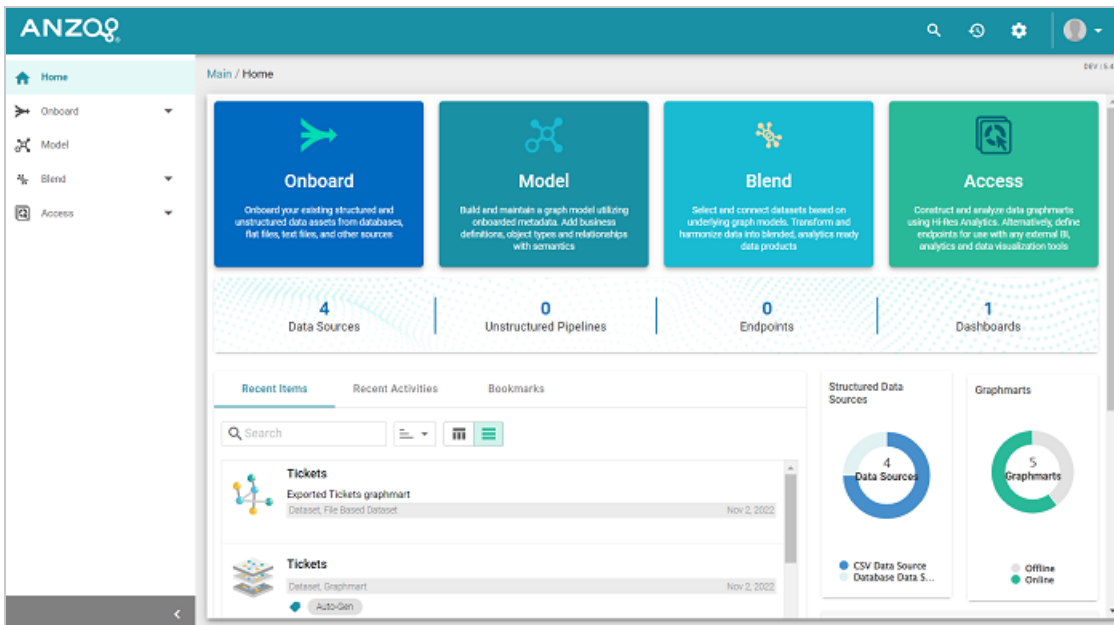
Note

The following sections show an administrator view of the user interface. Some users might not see all components. For information about default Anzo roles and permissions, see [Predefined Anzo Roles and Permissions](#) in the Administration Guide.


- [General Interface Elements](#)
- [Onboard](#)
- [Model](#)
- [Blend](#)
- [Access](#)

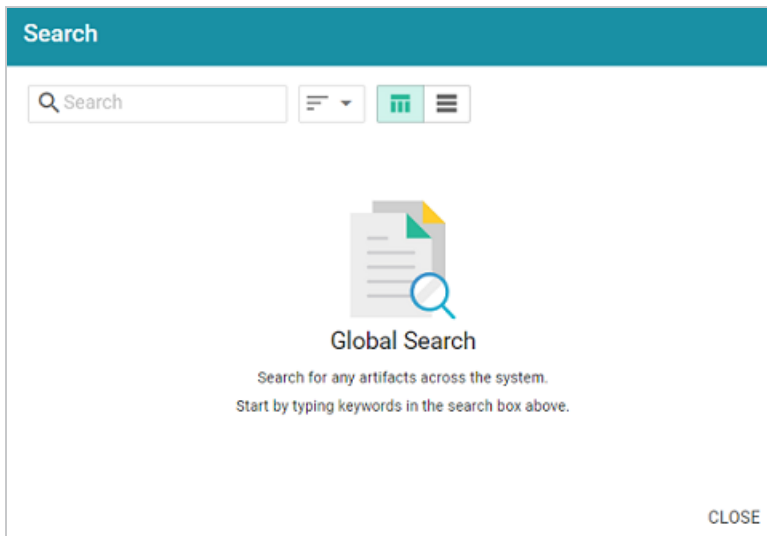
General Interface Elements

Each time you log into the application, the Home page is displayed and provides links to the features that you have access to. A dashboard displays an overview of the system artifacts and recently updated or bookmarked items. The dashboard becomes populated once you start onboarding data sources.

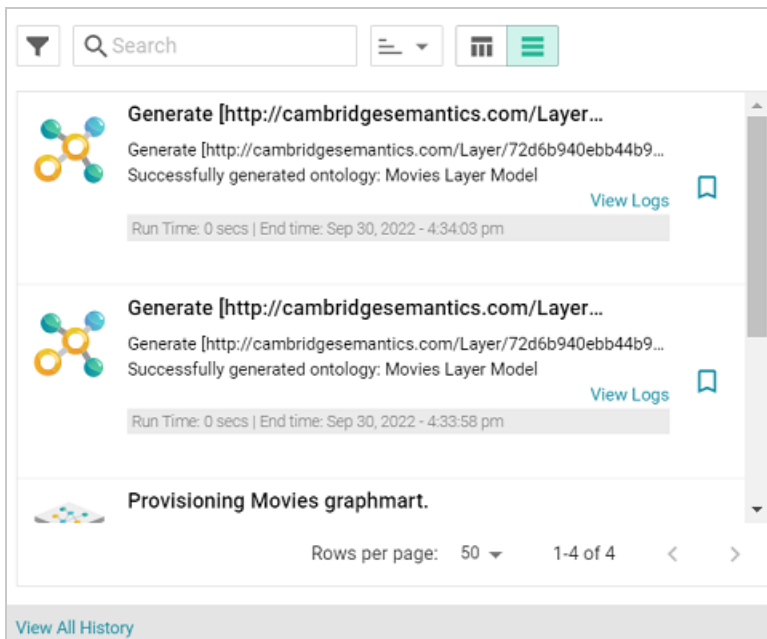


On the bottom left, you can click the collapse icon (<) to expand and collapse the left navigation menu. The right side of the top menu bar includes the following options:

- The Search icon () opens the Search dialog box where you can do a global search for a resource, such as a data source, dataset, or graphmart, by title. The Search field accepts wildcard characters.




- The Activity Log icon (🕒) opens the Activity Log, which offers a high-level view of individual activities, such as graphmart activity. The ability to filter activities based on their status and type enables users to drill down to activities of interest as well as profile user behavior. Additionally, the ability to include system data expands the scope of monitoring and aids in troubleshooting.



Once you start using Anzo and there are events that show up in the log, you can click **View All History** to open the full log for searching and filtering.


Title	↑	Description	Operation ID	Event Message	Progress ID	Start Time	End Time	Is Error	Is Cancelable	⋮
Generate [htt		Generate [htt...	1d586049-f2...	Successfully ...	http://cambri...	2022-09-30T...	2022-09-30T...	false	false	View Logs
Generate [htt		Generate [htt...	52e03602-79...	Successfully ...	http://cambri...	2022-09-30T...	2022-09-30T...	false	false	View Logs
Processing f		Movies	81o87k96tsf...	All files were ...	loadFilesInBa...	2022-09-30T...	2022-09-30T...	false	true	
Provisioning		Provisioning ...	81o87k96tsf...	Successfully ...	Remove.http...	2022-09-30T...	2022-09-30T...	false	true	View Logs

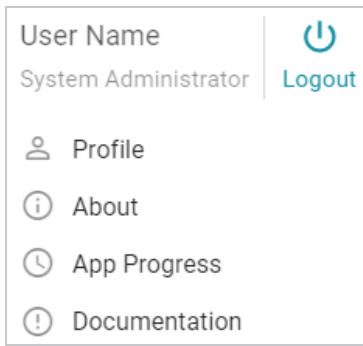
Rows per page: 25 1-4 of 4

- The Administration icon () opens the Administration menu. Clicking a menu option opens the Administration user interface for managing Anzo server settings, connections, users, and other components.

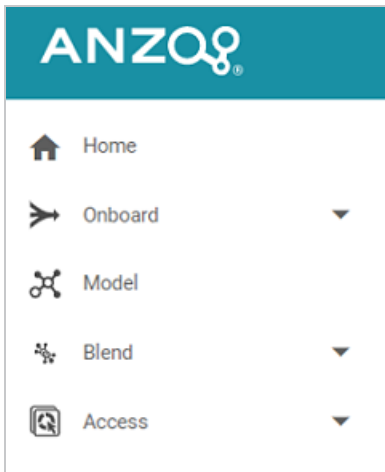
SERVICES	CONNECTIONS	TOOLS	USER MANAGEMENT	MONITORING & DIAGNOSTICS
Server Settings	File Store	Workflow Manager	Users	Logging
Server Certificates	Anzo Data Store	Migration Packages	Groups	System Query Audit
Licensing	AnzoGraph		Roles	Semantic Services
Volume Manager	Unstructured Clusters		Permissions	System Information
Plugin Configuration	Elasticsearch Config		Default Access Policies	Dynamic Resource Log
Advanced Configuration	Cloud Locations		Directory	
			SSO Config	

For information about the Administration application, see the Administration Guide.

- The User menu () provides access to your user profile, the About screen, and the Application Progress window, which lists recent application activity. It also includes the ability to log out of the application and a Documentation link that opens this guide.

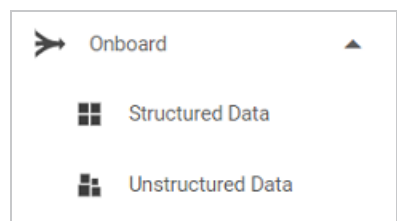


The navigation menu on the left provides access to all of the Anzo features.



Onboard

The Onboard menu provides access to the Structured and Unstructured Data components that users configure to ingest data from various sources.

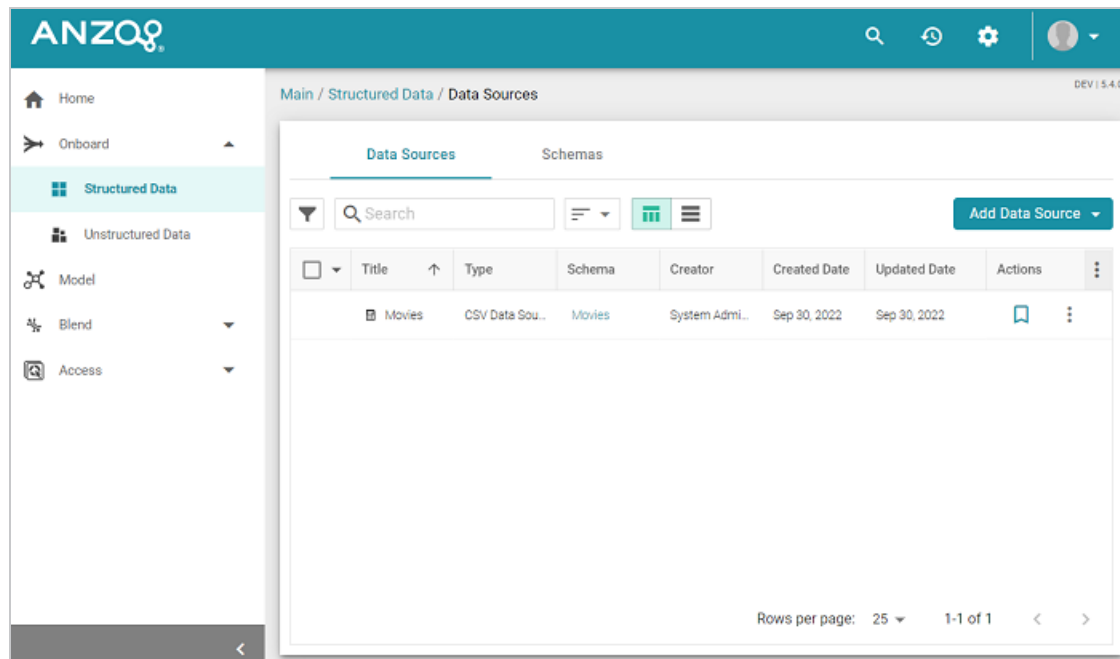


Structured Data

From **Structured Data**, users access data sources and schemas for structured data sources: database connections, CSV, JSON, XML, Parquet, and SAS files.

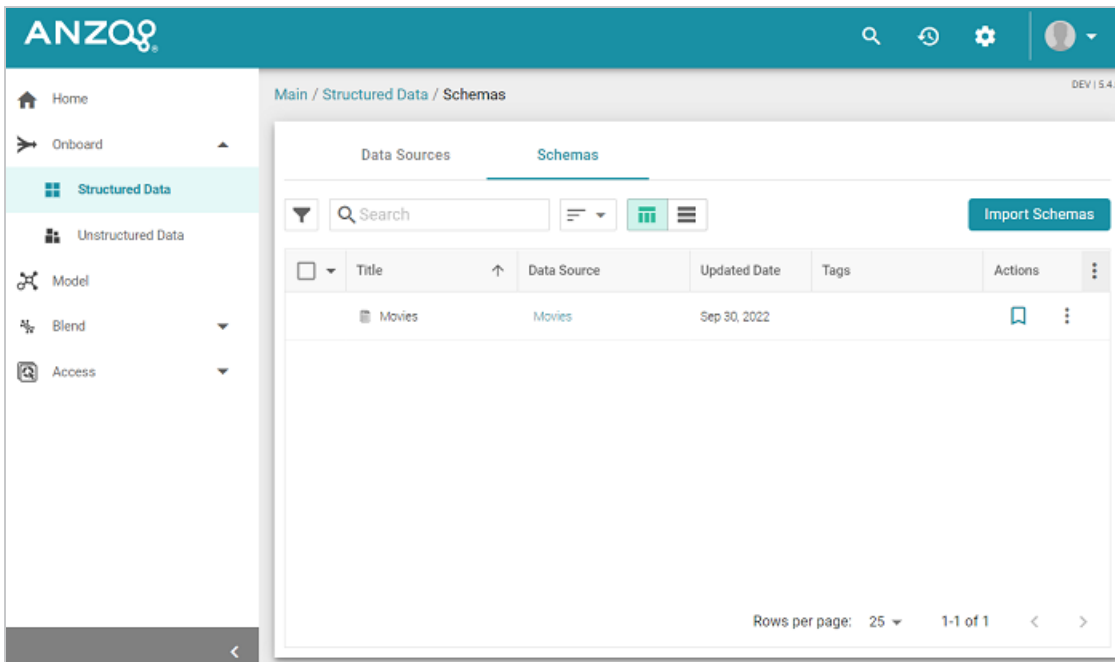
Data Sources

From the **Data Sources** tab, users connect to the files and databases that contain the data to onboard:



Schemas

From the **Schemas** tab, users view, create, and manage the schemas that define the data to onboard:

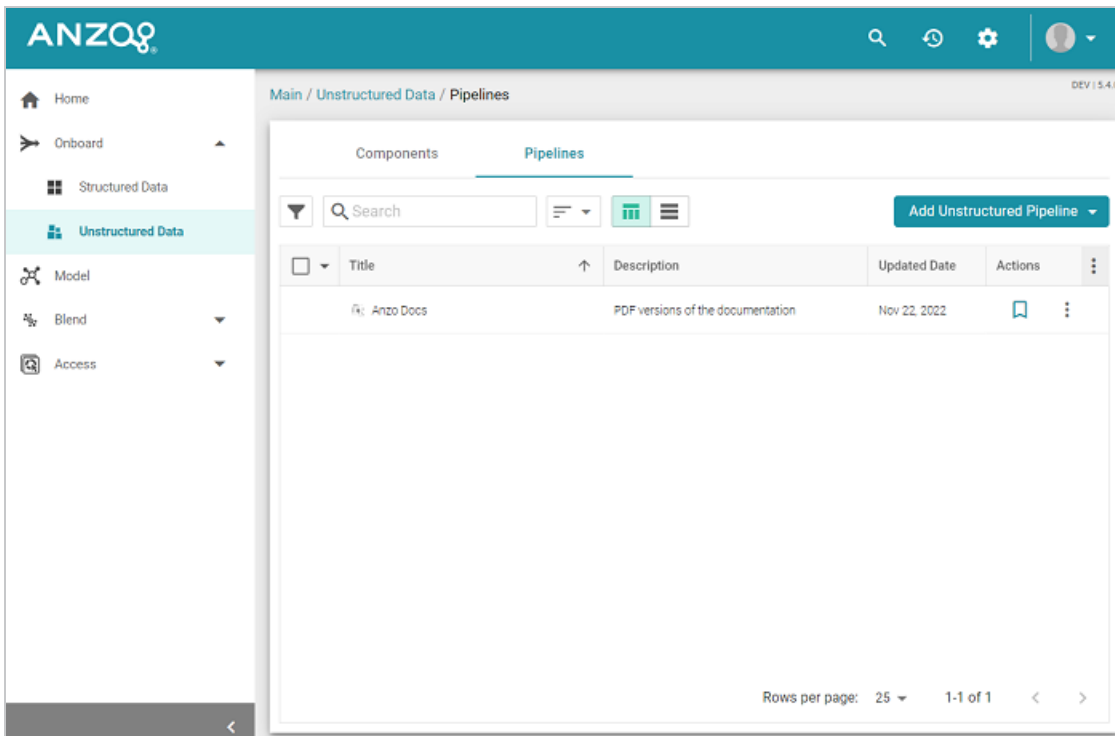


Unstructured Data

From **Unstructured Data**, users access and create pipelines for onboarding data from unstructured sources such as Office documents, PDFs, web pages, email messages, and knowledgebases.

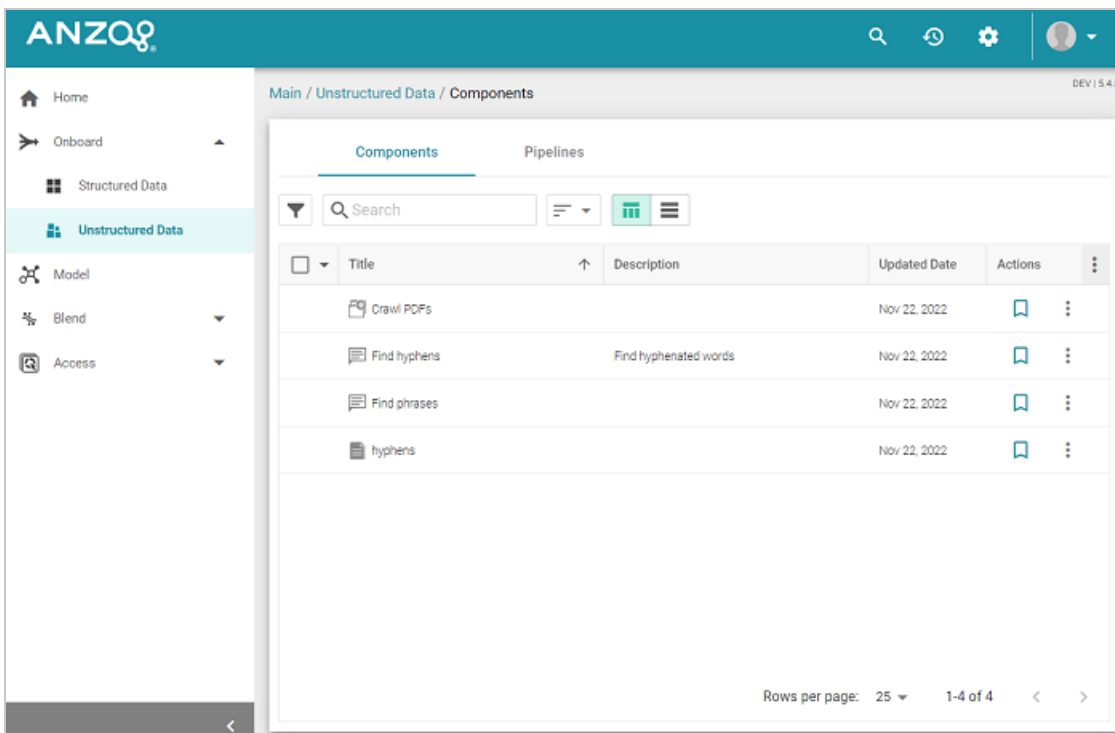
Pipelines

From the **Pipelines** tab, users can create and manage unstructured pipelines:



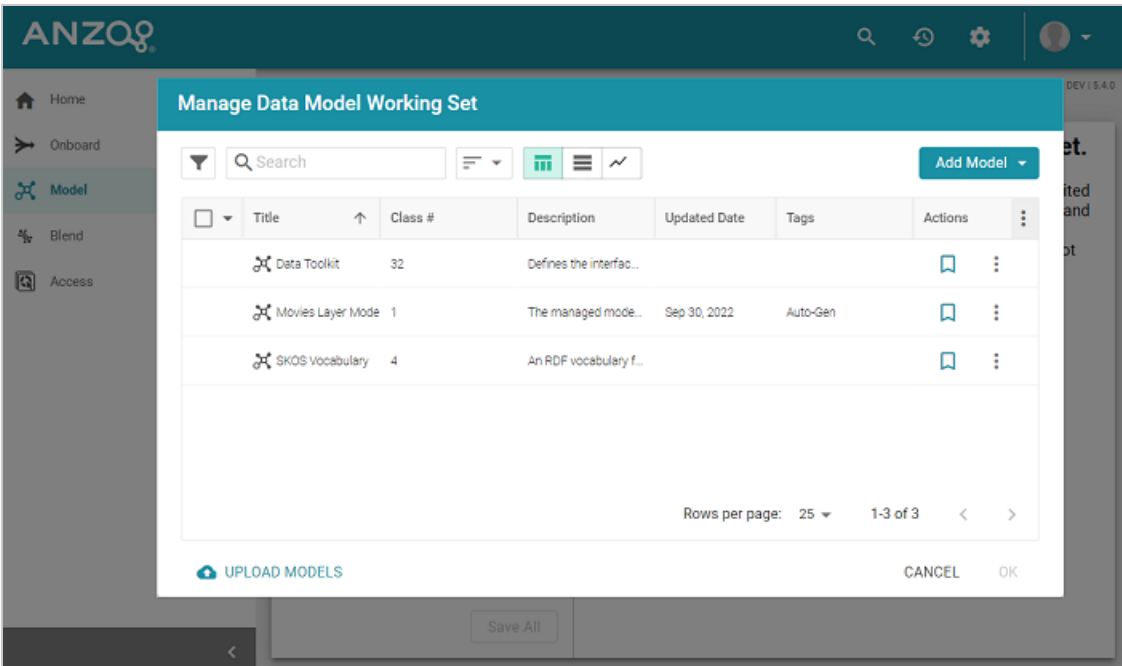
Components

From the **Components** tab, users can view and configure unstructured pipeline components:



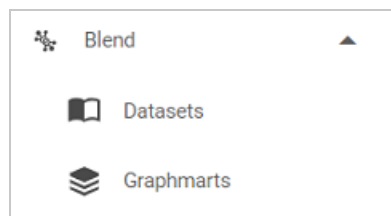
Model

The **Model** manager enables users to view, create, and manage data models, which describe the concepts, attributes, and relationships in or across datasets.



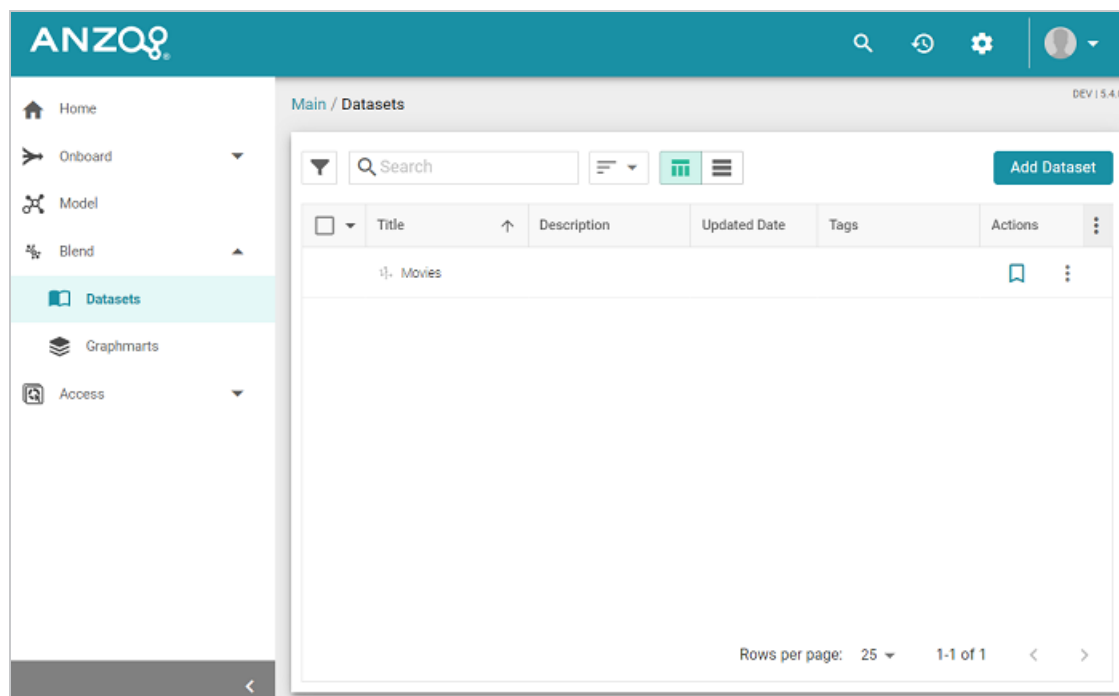
Blend

The **Blend** menu provides access to features that combine various datasets from different sources.



Datasets

The **Datasets** catalog is an inventory of the datasets in Anzo for which a file-based linked data set (FLDS) has been created. Users can sort the list by tags, classes, and creators. Users can also add datasets to graphmarts for loading into AnzoGraph.



Graphmarts

The **Graphmarts** page lists all of the existing graphmarts. Users can designate favorites and sort the list by tags and graphmart creators. Users can click a graphmart in the list to view details such as the datasets included in the graphmart, data layer details, and associated dashboards.

ANZOQ

Home

Onboard

Model

Blend

Datasets

Graphmarts

Access

Main / Graphmarts

DEV | 5.4.0

Search

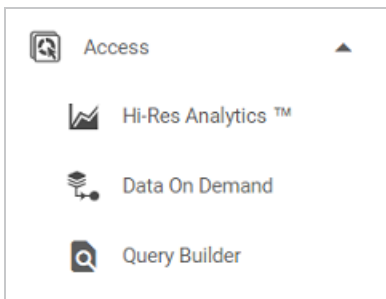
Add Graphmart

	Title	Description	Status	Statements	Creator	Tags	Actions
<input type="checkbox"/>	Movies		Ready to us	1,974	System Admi...	Auto-Gen	<div><div></div><div></div><div></div></div>

Rows per page: 25 1-1 of 1

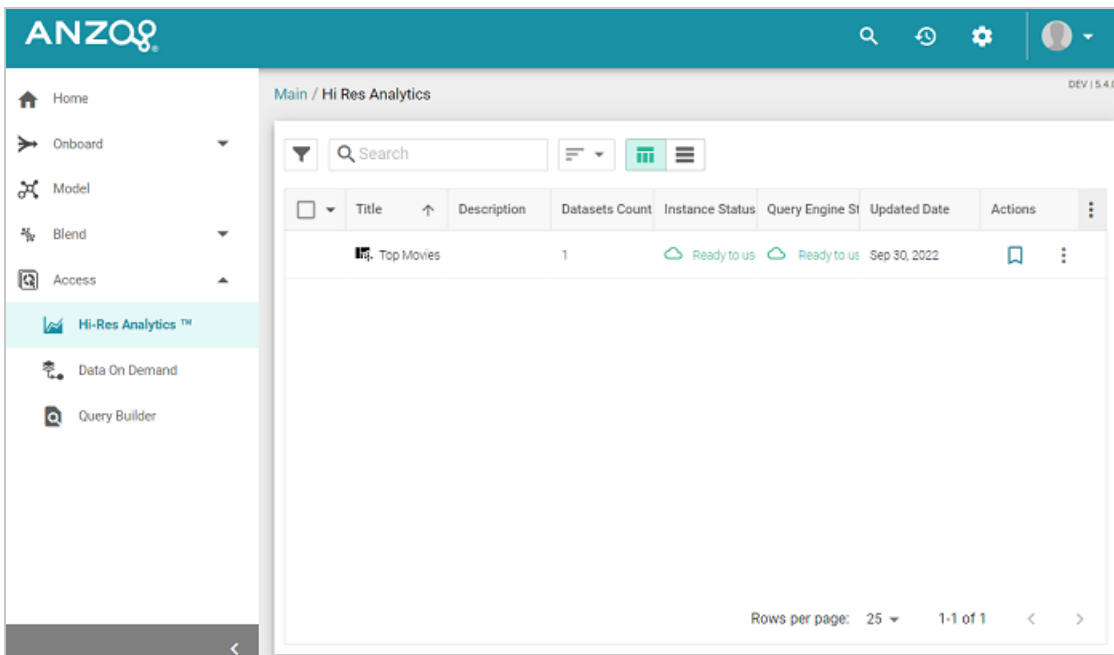
Access

The **Access** menu provides quick access to commonly used features.



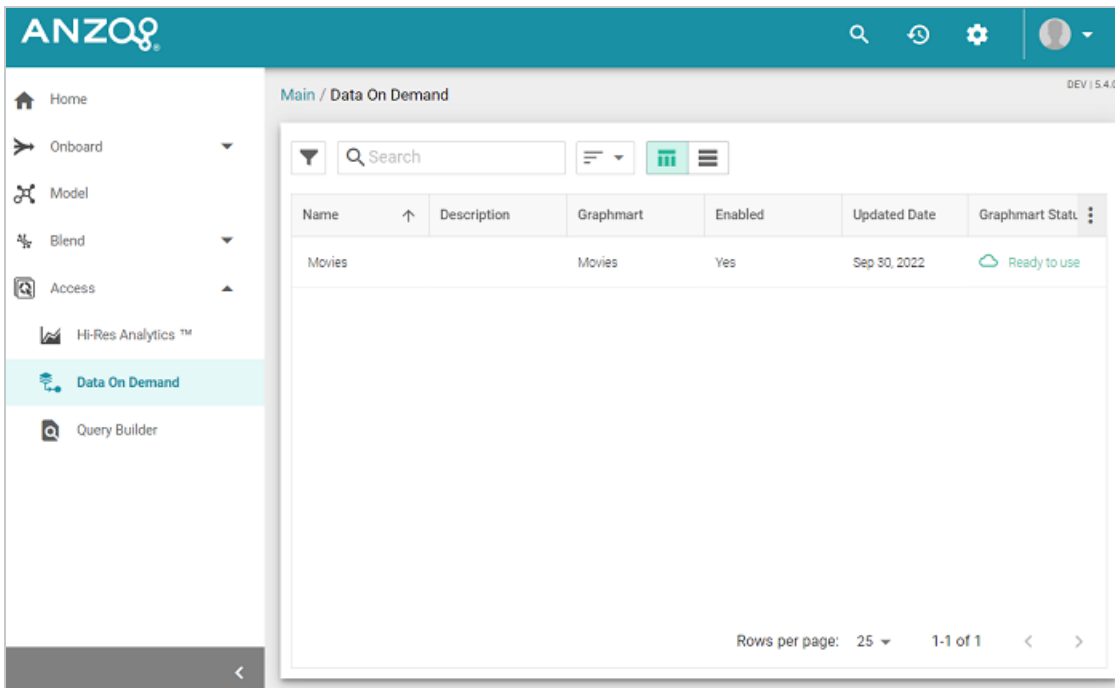
Hi-Res Analytics

The **Hi-Res Analytics** page lists all of the existing dashboards. Users can click a dashboard in the list to open it in the Hi-Res Analytics application. Users can designate favorites and sort the list by tags, graphmarts, and dashboard creators.



Data On Demand

The **Data On Demand** page provides a list of the Data on Demand endpoints that have been created.

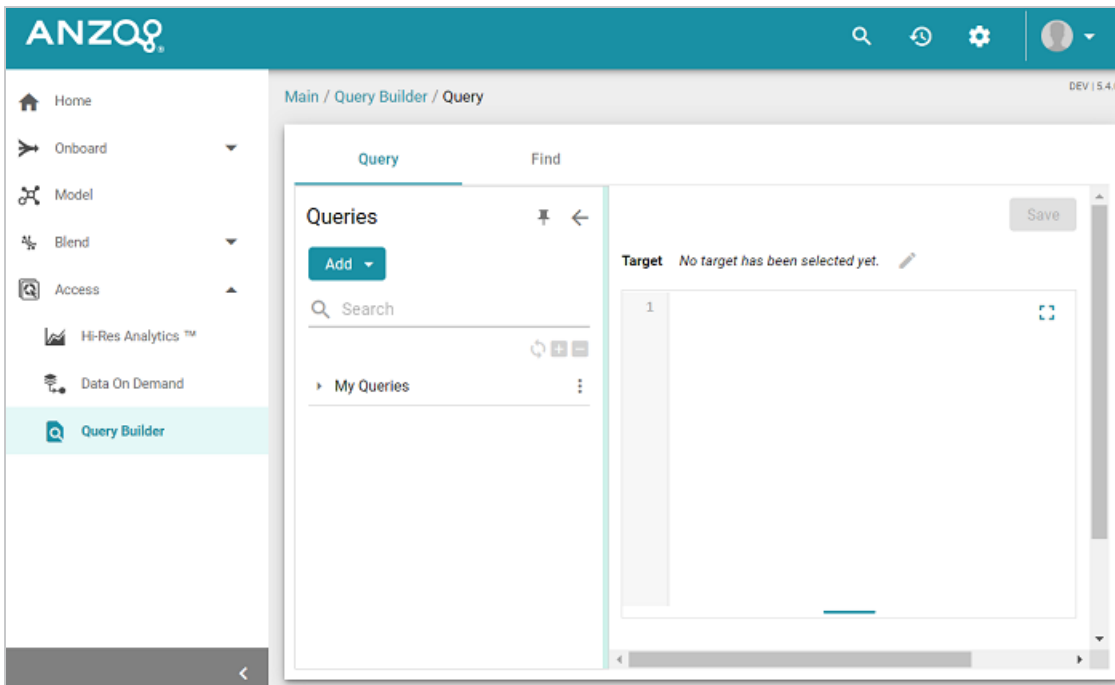


Query Builder

The **Query Builder** provides options to write and run SPARQL queries or find quads.

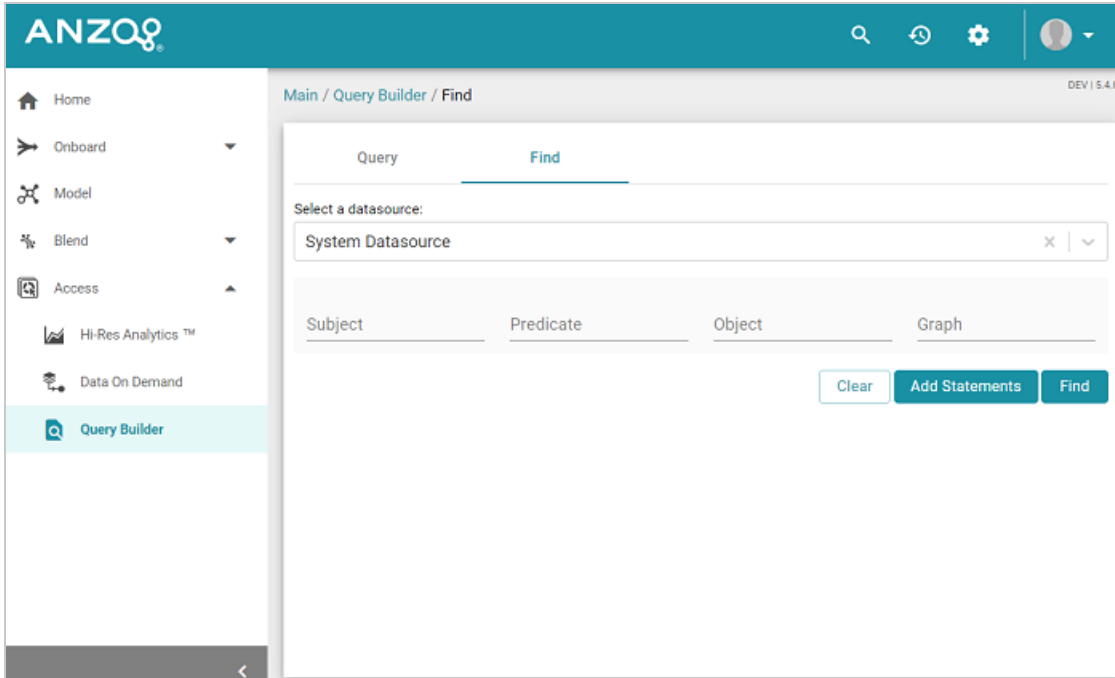
Query

The **Query** editor provides syntax assistance, type-ahead suggestions for model entity names, and automated prefix creation and query formatting for readability. It also includes the option to save queries for later use.



Find

The **Find** page enables users to search for quads by specifying a single subject, object, predicate, or graph name.



Tutorials

There are two ways to onboard structured and semi-structured data to Anzo:

Automated Direct Data Load Workflow

The Direct Data Load workflow follows an extract, load, and transform (ELT) process. In the ELT workflow, data sources are onboarded directly to graphmarts. Data layers with SPARQL queries are automatically generated to transform and blend the data to an analytics-ready knowledge graph. The AnzoGraph Graph Data Interface (GDI) Java plugin (sometimes called the Data Toolkit) is used to connect to the sources, create a model, and generate the data layer queries.

This section includes a tutorial for helping you get started with the direct data load ingestion method:

- [Building a Knowledge Graph from a CSV File](#)

Manual Ingestion with SPARQL Queries

For advanced users who are familiar with SPARQL, the GDI can also be invoked by writing queries from scratch. The GDI is extremely flexible, allowing you to connect directly to sources via queries and control all aspects of the extract, load, and transform process. In addition to the data sources that the automated workflow supports, you can also ingest raw data and data from HTTP/REST endpoints with manually written SPARQL queries.

For information about the advanced method of writing GDI SPARQL queries for ingestion, see [Onboarding or Virtualizing Data with SPARQL Queries](#) in the User Guide.

Building a Knowledge Graph from a CSV File

This tutorial helps get you started with Anzo by teaching you the basic steps to follow when onboarding data using the automated direct data load workflow. The instructions below provide a sample CSV file and guide you through the steps required from creating the data source to visualizing the data in a Hi-Res Analytics dashboard.

1. [Prepare the Environment](#)
2. [Download the Sample Data](#)
3. [Add a CSV Data Source for the Sample Data](#)
4. [Create a Graphmart from the Data Source](#)
5. [Explore the Data in a Hi-Res Analytics Dashboard](#)

Prepare the Environment

This tutorial requires a connection to AnzoGraph. Before starting the procedure, ensure that Anzo has a connection to a static AnzoGraph instance or is configured for dynamic Kubernetes-based AnzoGraph deployments. For instructions on configuring the connection to a static instance, see [Connecting to AnzoGraph](#) in the Administration Guide. For information about Kubernetes deployments, see [Using K8s for Dynamic Deployments of Anzo Components](#) in the Deployment Guide.

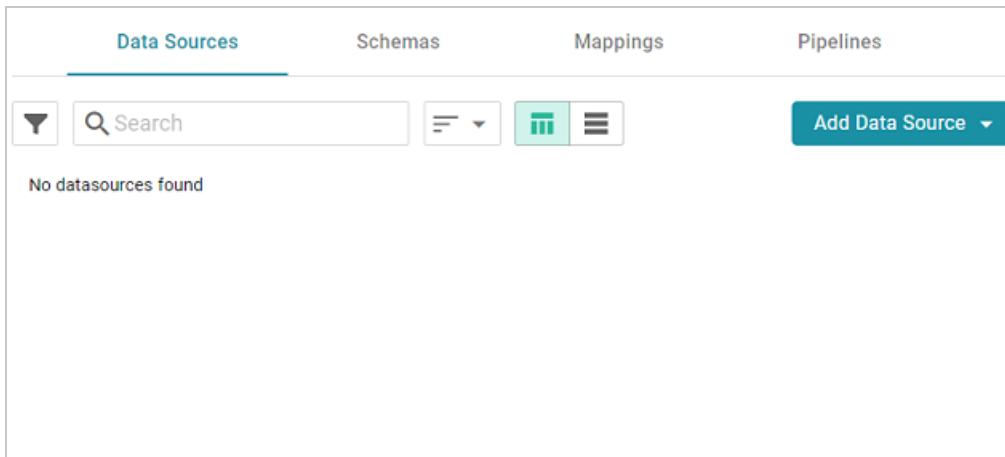
Download the Sample Data

This tutorial uses a small sample data set from [Kaggle](#). The data is in CSV format and includes information about the top 200 movies according to IMDB members. Click the link below to download Top_200_Movies.zip to your computer. Then extract the ZIP file to make **Top_200_Movies.csv** available for importing.

 [Top_200_Movies.zip](#)

Add a CSV Data Source for the Sample Data

1. In the Anzo application, expand the **Onboard** menu and click **Structured Data**. Anzo displays the Data Sources screen, which lists any existing data sources:



2. Click the **Add Data Source** button and select **File Data Source > CSV Data Source**. Anzo opens the Create CSV Data Source screen.

A screenshot of the 'Create CSV Data Source' form. The form has a teal header bar with the text 'Create CSV Data Source'. Below the header, there are two input fields: 'Title *' and 'Description'. The 'Title *' field is required, as indicated by an asterisk. At the bottom of the form, there are two buttons: 'CANCEL' and 'SAVE'.

3. Type a name for the source in the **Title** field. For example, "Top Movies." Then type an optional description in the **Description** field. For example:

Create CSV Data Source

Title *

Top Movies

Description

Top 200 IMDB movies

CANCEL

SAVE

- Click **Save**. Anzo saves the source and displays the Tables tab for the new Data Source.

Top Movies

Not Versioned

Overview

Tables

Versions

Discussion

Sharing

CSV

Please add new files to get started.

Add New File

- Click the **Add New File** button. Anzo displays the Add New File dialog box, and the Source is set to **From Your Computer**.

Add New File

Source
☒ From Your Computer ☐ From File Store

Upload to
 sysadmin User Folder [Change](#)

The location where your uploaded files will be put on the Anzo Server. Defaults to your user folder.

Drag and drop files or [browse](#) from computer.

Please select the .csv files to import

Selected: None [CLEAR ALL](#)

[CANCEL](#) [NEXT](#)

- As a best practice, check the upload location that is listed in the **Upload To** field by hovering your pointer over the value to view the tooltip. Make sure the upload location is a directory on the shared File Store and not a location in the server installation path. If the file is not uploaded to the shared File Store it is not accessible by applications like AnzoGraph. In addition, other users cannot publish pipelines for the Data Source because they typically do not have access to the file.

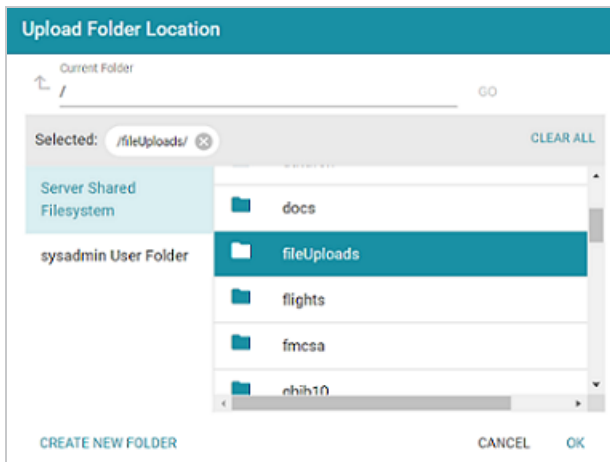
For example, viewing the Upload To location for the screen above shows that the file will be uploaded to the server installation path, `/opt/Anzo/Server/data...`

Source
☒ From Your Computer ☐ From File Store

Upload to
 sysadmin User Folder [Change](#)

/opt/Anzo/Server/data/userUploads/sysadmin_20235 is on the Anzo Server. Defaults to your user folder.

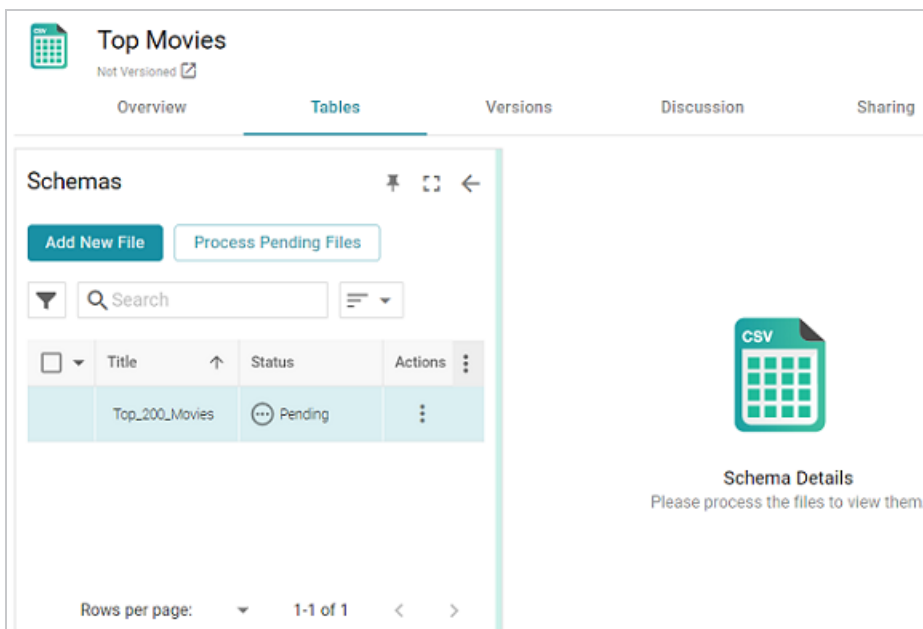
If your Upload To location is configured to upload the file to the server installation path, it is important to click **Change** and select another upload location that is on the shared File Store. For example, the image below shows the Upload Folder Location dialog box that is presented after clicking Change. A folder called **fileUploads** is selected on the File Store.



Tip

An administrator can configure the base upload path so that it points to a location on the File Store by default. For more information, see [Setting the Default Base File Store Path for File Uploads](#) in the Administration Guide.

- Next, drag and drop the `Top_200_Movies.csv` file onto the screen or click **browse** and navigate to the file and double-click it to select it. Anzo attaches the file and the **Next** button becomes active. Click **Next**. Anzo lists the file on the left side of the screen with a status of Pending. For example:



8. Click the **Process Pending Files** button to import the data to Anzo. Anzo imports the data and the status changes to **Processed**. Additional buttons are also displayed on the right side of the screen. You can click the table row for **Top_200_Movies** to display the schema details.

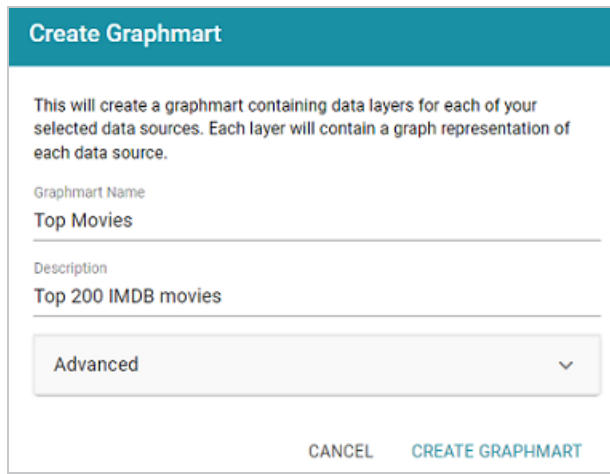
The screenshot shows the Anzo interface for a data source named 'Top Movies'. The 'Tables' tab is active, and the 'Top_200_Movies' table is selected. The table's status is 'Processed'. The schema details for 'Top_200_Movies' are displayed, showing columns: Title, Year of Release, Rating, Number of Reviews, Popularity Index, Movie Cast, Director, and Description. The table contains 8 rows of data.

Title	Year of Release	Rating	Number of Reviews	Popularity Index	Movie Cast	Director	Description
Pulp Fiction	1994	8.9	2M	128	[John Travolta, U]	Quentin Tarantino	The lives of two m
The Godfather: Pa	1974	9.0	1.2M	102	[Al Pacino, Rober	Francis Ford Copp	The early life and c
12 Angry Men	1957	9.0	771K	446	[Henry Fonda, Le	Sidney Lumet	The jury in a New
The Dark Knight	2008	9.0	2.6M	91	[Christian Bale, H	Christopher Nolan	When the menace
The Shawshank R	1994	9.3	2.6M	66	[Tim Robbins, M	Frank Darabont	Two imprisoned m
The Godfather	1972	9.2	1.8M	17	[Marlon Brando, F	Francis Ford Copp	The aging patriar
Schindler's List	1993	9.0	1.3M	245	[Liam Neeson, R	Steven Spielberg	In German-occup

Now that the data source has been added, proceed to the next step to create a Graphmart and load the data to AnzoGraph.

Create a Graphmart from the Data Source

1. From the Tables screen (from the last step of the procedure above), click the **Create Graphmart** button at the top of the screen. The Create Graphmart dialog box is displayed:



Create Graphmart

This will create a graphmart containing data layers for each of your selected data sources. Each layer will contain a graph representation of each data source.

Graphmart Name
Top Movies

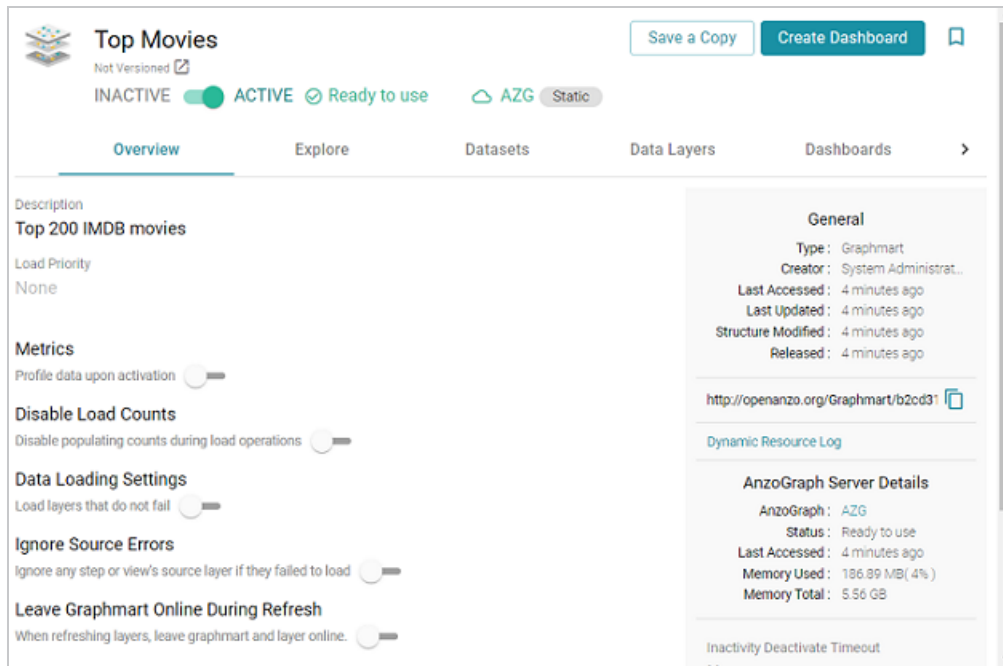
Description
Top 200 IMDB movies

Advanced ▼

CANCEL CREATE GRAPHMART

2. On the Create Graphmart dialog box, the **Graphmart Name** is populated with the name of the data source, and the **Description** is populated with the description from the data source. You can edit either field if necessary. No further configuration is required. For details about the Advanced options that are on this screen, refer to [Direct Load Advanced Settings Reference](#) in the User Guide.

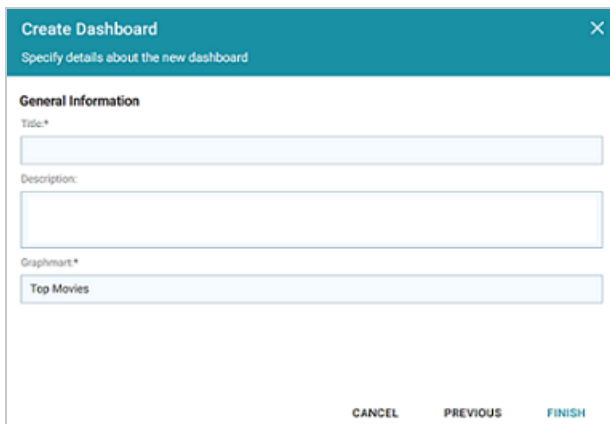
3. Click **Create Graphmart**. The new Graphmart is created and activated.



When the activation is complete, you can click the **Data Layers** tab to review the layer and step that were created. In addition to the layer and step, a **Managed Model** was also created. Now that the data is loaded into AnzoGraph, it is available to explore and analyze in dashboards. Proceed to the next step to explore the sample data in the Hi-Res Analytics application.

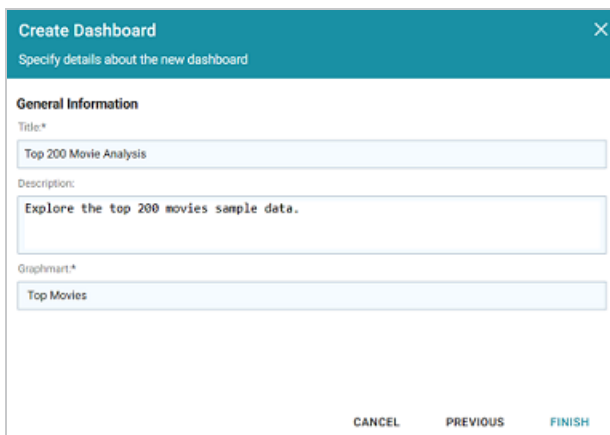
Explore the Data in a Hi-Res Analytics Dashboard

1. On the Graphmart screen (as shown in the last step above), click the **Create Dashboard** button. Anzo opens the Hi-Res Analytics application and displays the Create Dashboard dialog box.
2. Select **Graphmart Dashboard** and click **Next**. The Create Dashboard dialog box is displayed:



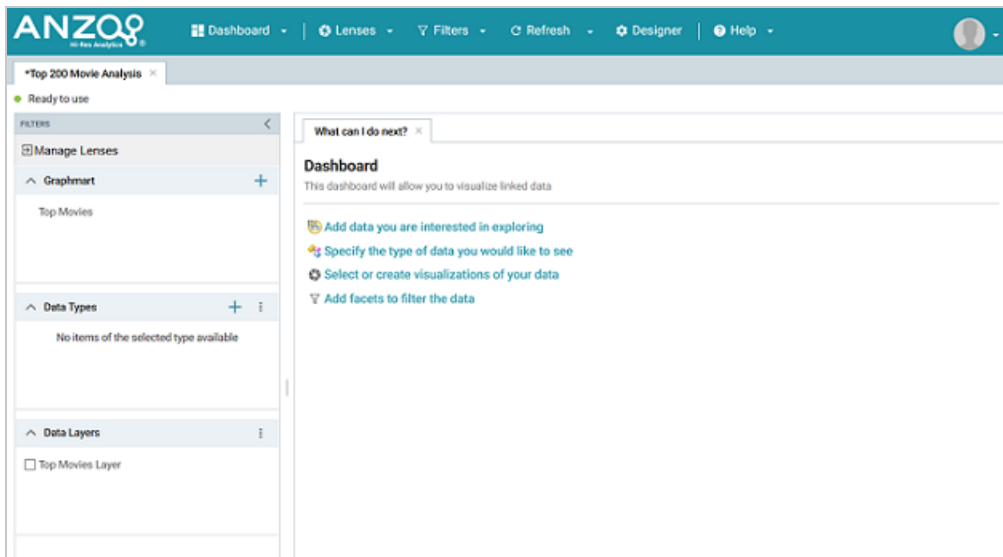
The image shows a 'Create Dashboard' dialog box with a teal header. Below the header, there's a section titled 'General Information'. It contains three input fields: 'Title*' (empty), 'Description:' (empty), and 'Graphmart*' (containing 'Top Movies'). At the bottom, there are three buttons: 'CANCEL', 'PREVIOUS', and 'FINISH'.

3. On the Create Dashboard screen, specify a **Title** for the dashboard and add an optional **Description**. Leave the default **Graphmart dashboard** value in the **Type** field. For example:

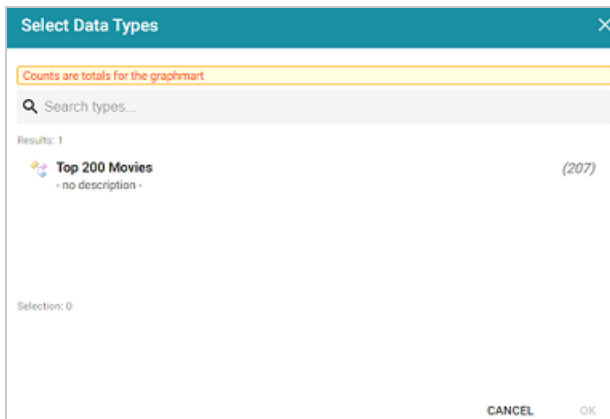


The image shows the same 'Create Dashboard' dialog box, but now the input fields are filled. The 'Title*' field contains 'Top 200 Movie Analysis', the 'Description:' field contains 'Explore the top 200 movies sample data.', and the 'Graphmart*' field still contains 'Top Movies'. The buttons at the bottom remain the same.

4. Click **Finish** to create the dashboard. The new Dashboard appears as a new tab on the screen and contains a sub-tab titled **What can I do next?**. This tab acts as a wizard to guide you through the initial Dashboard creation. Anzo populates the Graphmart and Data Layers panels with the sample movie data graphmart name and default layer for the graphmart.



5. To define the type of data to display on the dashboard, click the **Specify the type of data you would like to see** link on the What can I do next tab. The Select Data Types dialog box appears and displays the available data types. Since the sample data set has one type or class of data, only **Top 200 Movies** is listed. The value in parentheses shows the total number of instances of that type.



6. Click **Top 200 Movies** to select it, and then click **OK** to add the data type to the dashboard. The selected type is added to the Data Types panel.
7. In the main Hi-Res Analytics toolbar, click the **Dashboard** button and select **Save** to save the Dashboard.

Tip

To start to explore the data and see what values exist for the properties in the Top_200_Movies class, it can help to add a filter to the dashboard. Filters reveal the values associated with properties. Learning more about the values enables you to start making decisions about what properties to group on, for example, what properties have relationships, and what results you want to visualize.

8. To create a filter, click the **Add facets to filter the data** link on the What can I do next tab. Anzo displays the Create Filter dialog box:

The image shows a 'Create Filter' dialog box with a teal header bar containing the title 'Create Filter' and a close button (X). The main area is white and contains two sections: 'Fields:*' with a light blue button labeled 'Click to edit', and 'Filter:*' with a light gray rectangular area. At the bottom, there are two buttons: 'CANCEL' and 'OK'.

9. In the Create Filter dialog box, click the **Fields** field and browse the available properties. For this tutorial, select **Rating**, and then click **Close** to close the Fields drop-down list. Additional options are displayed:

Create Filter

×

Fields:*

Rating

Filter:*

List

Filter Properties

Title:

Label field:

Click to edit

Exclude: ☐ Show Bars: ☐ Show Blanks: ☐ Always Show Checkboxes: ☒

Show counts: ☒ Respond to other filters: ☒

Format

Type:

Automatic

Subfilters

+ Create Filter

CANCEL

OK

10. Click the **Filter** drop-down list and browse the available filter types. For this tutorial, select **Single Select List**.
11. Click **OK** to close the dialog box and add the new filter to the dashboard. Anzo adds the Rating filter to the left panel. The new filter will enable you to narrow the scope of the data when it is visualized.

^ Rating

:

|

×

8.1 (47)

8.2 (43)

8.3 (36)

8.4 (28)

8.5 (21)

8.6 (14)

8.7 (4)

8.8 (6)

8.9 (1)

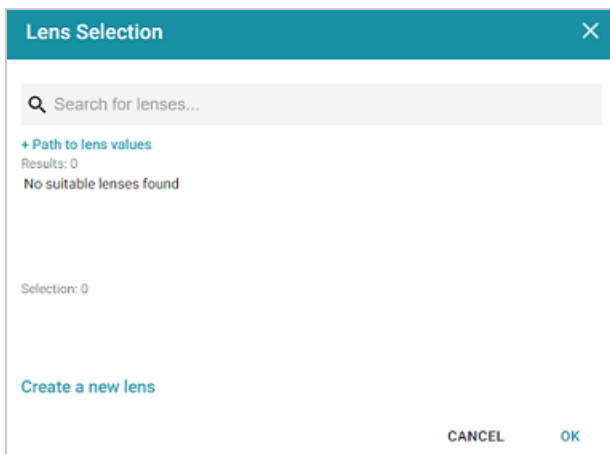
9 (5)

12 choices | [more](#)

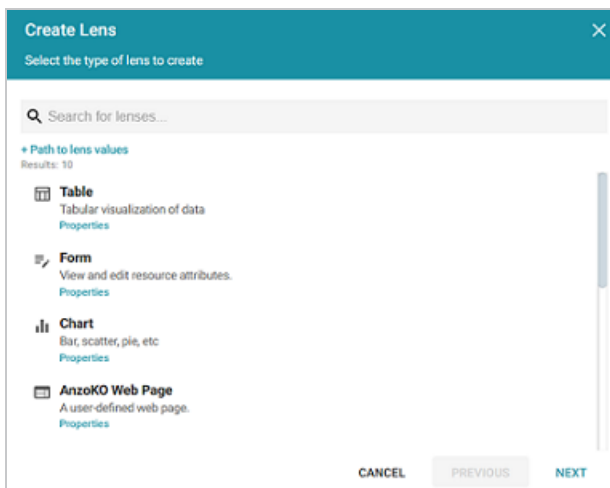
Tip

Once you have a good understanding of the values and relationships that exist in the data set, you can experiment with lenses and decide on the most appropriate way to display the data. Creating a Table lens is a quick way to get started with learning about the values and relationships in the data and to see how the new filter works.

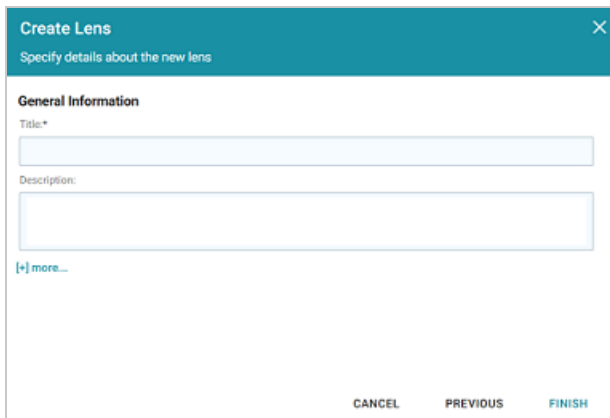
12. To create a Table lens, click the **Select or create visualizations of your data** link on the What can I do next tab. Anzo displays the Lens Selection dialog box. Since lenses have not been created previously, the list of lenses is empty:



13. Click the **Create a new lens** link at the bottom of the screen. Anzo displays the Create Lens dialog box:

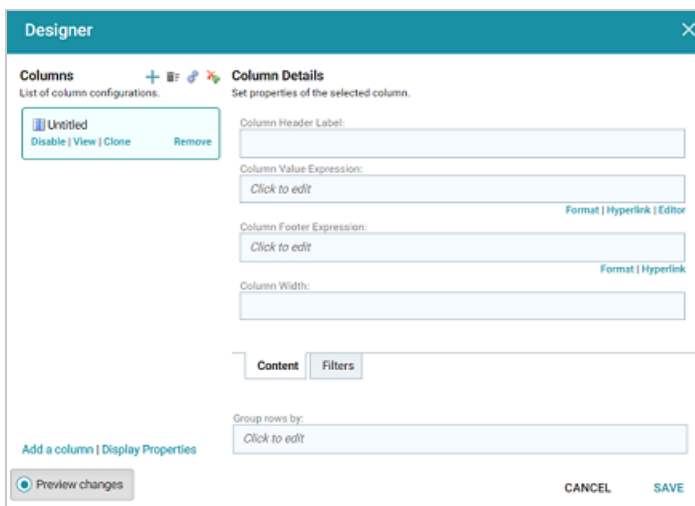


14. Select the **Table** lens and then click **Next**. Anzo displays the Create Lens dialog box:



The 'Create Lens' dialog box has a teal header with the title 'Create Lens' and a close button. Below the header is a subtitle 'Specify details about the new lens'. The main section is titled 'General Information' and contains a 'Title*' text field, a 'Description:' text area, and a '+ more...' link. At the bottom are three buttons: 'CANCEL', 'PREVIOUS', and 'FINISH'.

15. Type a **Title** for the lens, for example "Movie Table," and then click **Finish**. Anzo adds the lens as a new sub-tab on the dashboard and opens the Table Designer:



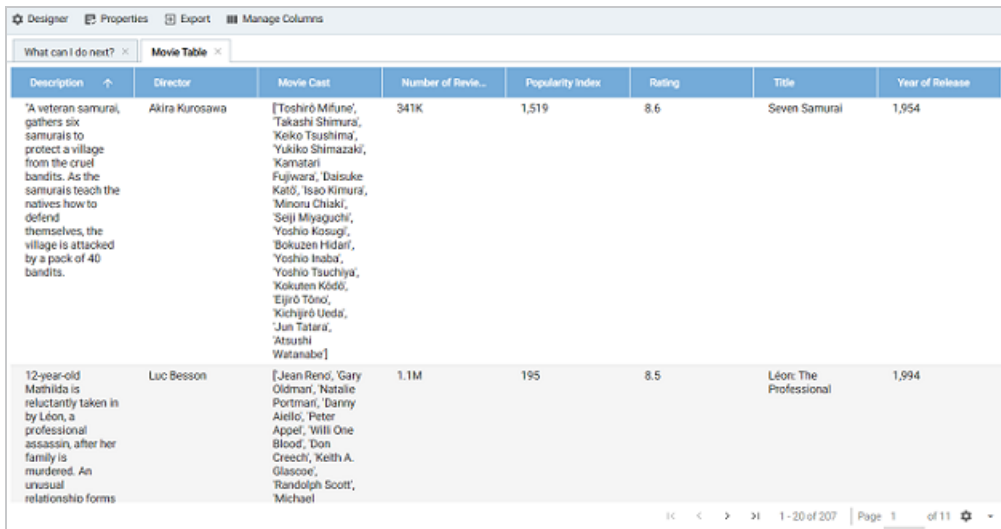
The 'Designer' dialog box has a teal header with the title 'Designer' and a close button. It is divided into two main sections. The left section, titled 'Columns', shows a list of column configurations with one entry 'Untitled' and buttons for 'Disable', 'View', 'Clone', and 'Remove'. The right section, titled 'Column Details', allows setting properties for the selected column, including 'Column Header Label', 'Column Value Expression', 'Column Footer Expression', and 'Column Width'. Each of these has a 'Click to edit' button. There are also links for 'Format', 'Hyperlink', and 'Editor'. At the bottom, there are tabs for 'Content' and 'Filters', a 'Group rows by:' section with a 'Click to edit' button, and buttons for 'Add a column', 'Display Properties', 'Preview changes', 'CANCEL', and 'SAVE'.

16. In the Designer, click the **Auto-generate columns** icon (⚙️) to add all available columns to the table.

Tip

Since the sample data set includes only 8 columns, it is not overwhelming to view all columns at once. For larger data sets, you might want to be more selective when adding columns to table lenses.

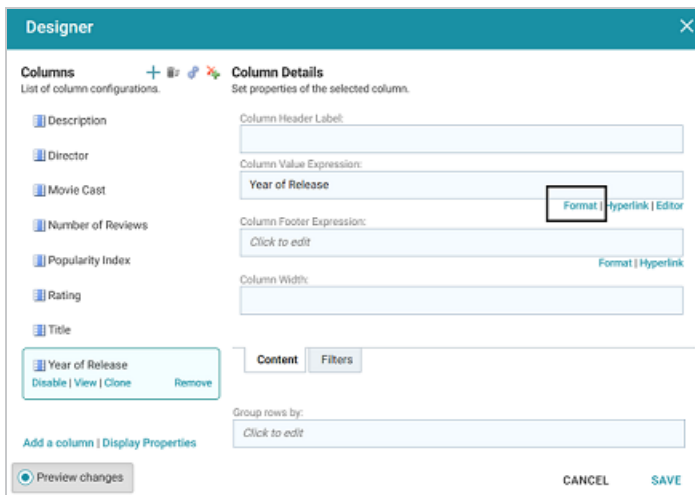
17. Click **Save**. The table lens is added to the dashboard and is populated with data. For example:



The screenshot shows a table lens titled "Movie Table" with the following columns: Description, Director, Movie Cast, Number of Reviews, Popularity Index, Rating, Title, and Year of Release. The table contains two rows of data.

Description	Director	Movie Cast	Number of Reviews	Popularity Index	Rating	Title	Year of Release
'A veteran samurai, gathers six samurais to protect a village from the cruel bandits. As the samurais teach the natives how to defend themselves, the village is attacked by a pack of 40 bandits.	Akira Kurosawa	[Toshiro Mifune, Takashi Shimura, Keiko Tsushima, Yukiko Shimazaki, Kamatari Fujiwara, Daisuke Katô, Isao Kimura, Minoru Chiaki, Seiji Miyaguchi, Yoshio Kosugi, Bokuzen Hidari, Yoshio Inaba, Yoshio Tsuchiya, Kokuten Kôdô, Eijirô Tônô, Kichijirô Ueda, Jun Tataru, Atsushi Watanabe]	341K	1,519	8.6	Seven Samurai	1,954
12-year-old Mathilda is reluctantly taken in by Léon, a professional assassin, after her family is murdered. An unusual relationship forms	Luc Besson	[Jean Reno, Gary Oldman, Natalie Portman, Danny Aiello, Peter Appel, Willi One Blood, Don Creech, Keith A. Glasore, Randolph Scott, Michael	1.1M	195	8.5	Léon: The Professional	1,994

18. If you want to make adjustments to the table, such as to reformat the **Year of Release** column to remove the commas from the values, click the **Designer** button for the lens. In the Designer, select the **Year of Release** column, and then click **Format** under Column Value Expression in the Column Details:



The Designer dialog box shows the "Year of Release" column selected. In the "Column Details" section, the "Column Value Expression" is set to "Year of Release". The "Format" button is highlighted, indicating the next step in the process.

In the Format Column dialog box, click the **Type** drop-down list and select **Year** or **No format**, and then click **Close**. Click **Save** to change the lens change and close the Designer.

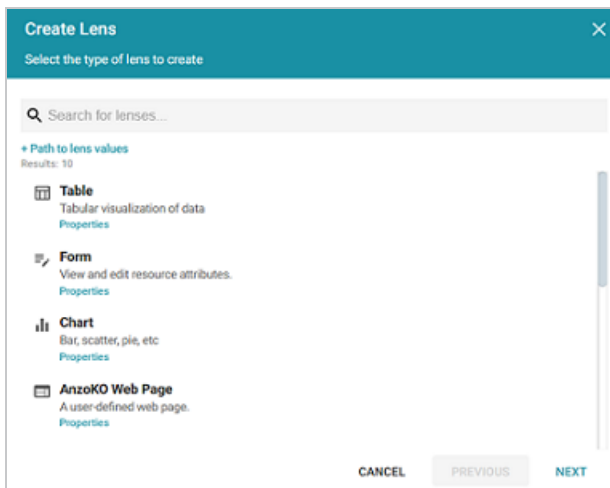
19. Now, you can filter the data and reduce the number of results that are displayed by clicking a value in the **Rating** filter. In the filter, click any one of the ratings in the list to display only the

data about movies with that rating. You can click another rating to filter the table again, and so on.

20. In the Hi-Res Analytics main toolbar, click the **Dashboard** button and select **Save** to save the changes. Remember to save dashboards periodically.

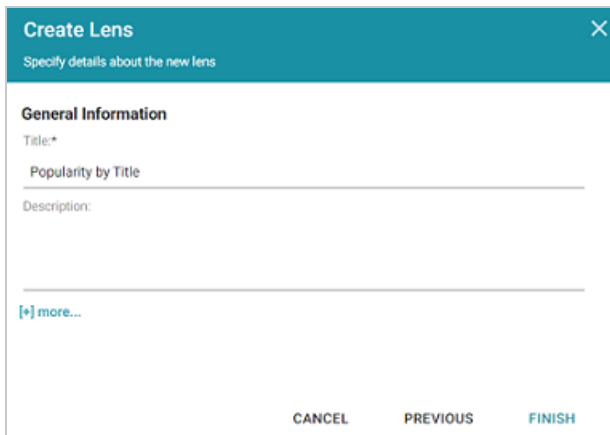
Now that you can view a summary of the data in a table, it can help you determine how to further narrow or expand the results by adding, changing, or removing filters. In addition, you can experiment by adding other lenses to the dashboard to find the ideal way to display the data to answer the questions that you have. For example, the table includes a **Popularity Index** column. It might be interesting to see which movies were the most popular and how the popularity compares to the chosen rating in the filter. The next steps guide you through creating a chart lens to display movie popularity.

1. In the Hi-Res Analytics main toolbar, click **Lenses** and select **New**. The Create Lens dialog box opens.



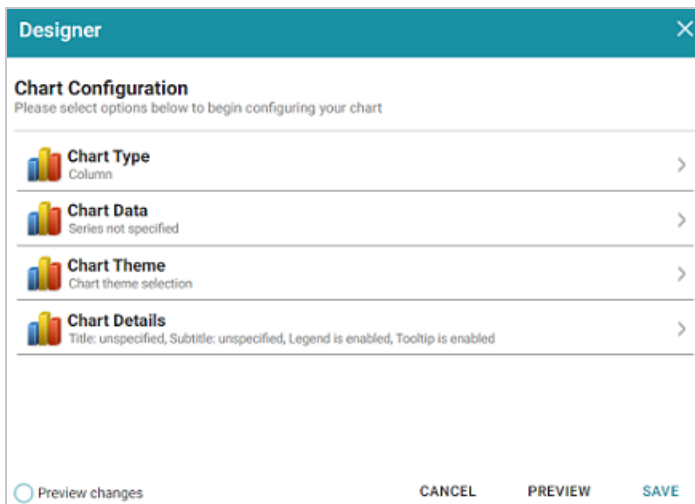
2. Select **Chart** and click **Next**. Anzo displays the Create Lens dialog box.

3. Specify a **Title** and add an optional **Description** for the new lens. For example:



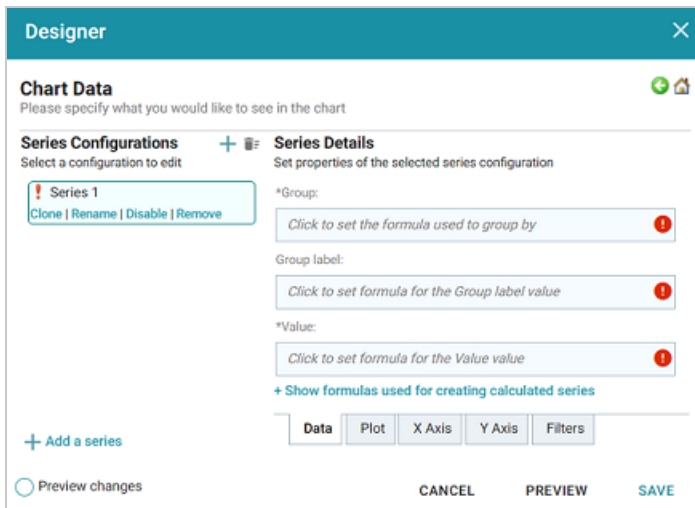
The 'Create Lens' dialog box has a teal header with the title 'Create Lens' and a close button. Below the header is a subtitle 'Specify details about the new lens'. The main area is titled 'General Information' and contains a 'Title:' label with a required asterisk, a 'Popularity by Title' label, and a 'Description:' label. There are text input fields for each. A '[+] more...' link is at the bottom left. At the bottom right are three buttons: 'CANCEL', 'PREVIOUS', and 'FINISH'.

4. Click **Finish**. Anzo displays the Chart Designer, which enables you to configure the type of chart to create, such as column, pie, or line, specify the data that will populate the chart, and customize details such as the chart's theme and fonts, legend, title, and tooltip formats.

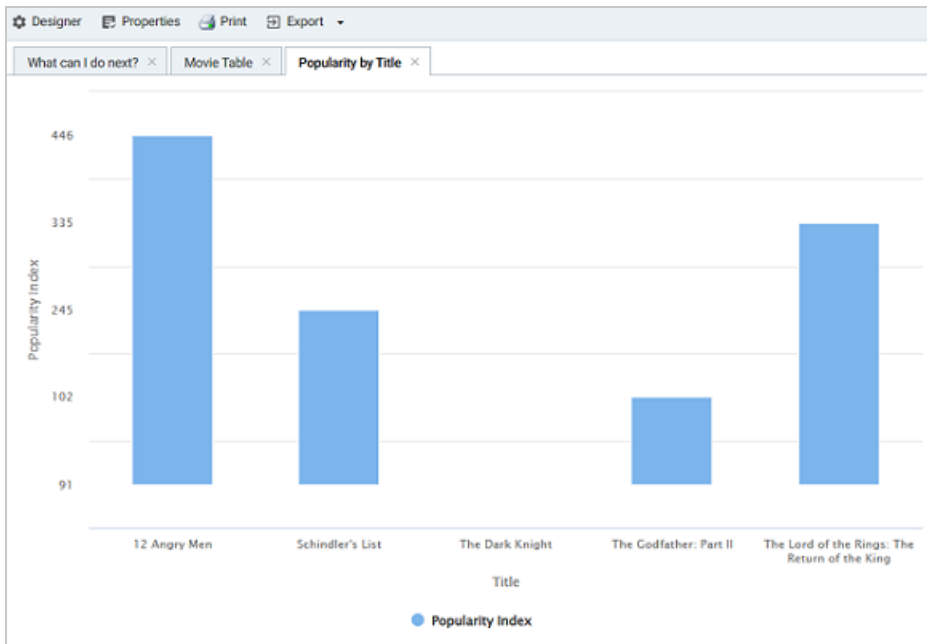


The 'Designer' dialog box has a teal header with the title 'Designer' and a close button. Below the header is the section 'Chart Configuration' with the subtitle 'Please select options below to begin configuring your chart'. There are four expandable sections: 'Chart Type' (Column), 'Chart Data' (Series not specified), 'Chart Theme' (Chart theme selection), and 'Chart Details' (Title: unspecified, Subtitle: unspecified, Legend is enabled, Tooltip is enabled). Each section has a right-pointing arrow. At the bottom left is a radio button labeled 'Preview changes'. At the bottom right are three buttons: 'CANCEL', 'PREVIEW', and 'SAVE'.

5. For this tutorial, accept the default Chart Type of **Column**, and click **Chart Data** to view the Chart Data screen and configure the data to display.



6. On the left side of the screen, click **Rename** under **Series 1**. Specify a new name in the Rename Series field, for example, "Popularity Index," and then click **OK**.
7. On the right side of the screen, click the **Group** field and select the **Title** property from the drop-down list. The Group field sets the values for the X Axis on the column chart. Then click **Close** to close the drop-down list. Anzo also populates the Group label with the Title property.
8. Click the **Value** field and select the **Popularity Index** property from the drop-down list. The Value field sets the Y Axis values. Click **Close** to close the drop-down list.
9. Click **Save**. Anzo adds the new Chart lens to the dashboard. To narrow the results that are displayed and filter on one rating at a time, click a rating in the filter. For example, by clicking **9** in the filter, the chart lists the popularity index for the titles that earned the highest rating. Of the highest rated movies, "12 Angry Men" had the highest popularity index:



Tip

You can hover the pointer over a bar in the chart to view details.

10. If you want to add more data to the Chart, such as to compare Popularity versus Reviews for each Title, follow these steps:
 - a. Open the Chart Designer by clicking the **Designer** button above the lens tabs.
 - b. In the Designer, click **Chart Data**.
 - c. On the Chart Data screen, click the **Add a series** link at the bottom of the screen. A new series, named Series 2, is added to the list.

Designer

Chart Data

Please specify what you would like to see in the chart

Series Configurations

Select a configuration to edit

Popularity by Title

Series 2

Clone | Rename | Disable | Remove

+ Add a series

Preview changes

Series Details

Set properties of the selected series configuration

*Group:

Click to set the formula used to group by

Group label:

Click to set formula for the Group label value

*Value:

Click to set formula for the Value value

+ Show formulas used for creating calculated series

Data

Plot

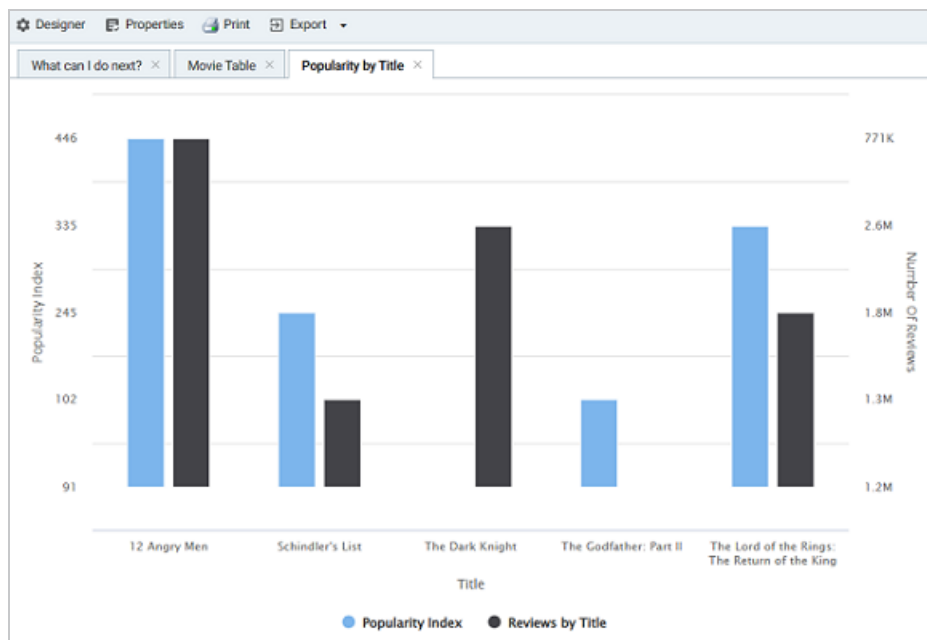
X Axis

Y Axis

Filters

CANCEL PREVIEW SAVE

- Under **Series 2**, click **Rename** and specify a name for the series, such as "Reviews by Title." Then click the **Group** field on the right side of the screen and select the **Title** property from the drop-down list. Click **Close** to close the list.
- Click the **Value** field and select the **Number of Reviews** property from the drop-down list. Close the list.
- Click **Save** to save the changes. The Chart is updated to display the number of reviews in addition to the popularity. For example:



Note

If you filtered the data in a previous step, the lens remains filtered on the chosen rating. You can de-select the rating in the filter or choose an alternate rating to explore the reviews versus popularity for additional titles.

For more information about working with Dashboards, see [Access Data with Hi-Res Analytics Dashboards](#) in the User Guide.