

Adding Query-Based Datasets

A query-based dataset allows you to select data from one or more tables in your data warehouse and save the query in AtScale as a new dataset.

About Reusing Datasets

Consider that reusing Datasets between the Dimension canvas and the Cube Canvas - or between multiple Dimensions - can cause the Engine to take unexpected join-paths at deployment time.

If a data architect wishes to reference the same table from multiple points in the same model, it is recommended to define separate Query Datasets (QDS) that select from the desired table. The separate datasets should be used to back each dimension or Fact model component. Doing so ensures the generation of run-time join-paths that are readily identified in the Design Center.

Note that each QDS must be created from scratch; converting an existing Dataset to a QDS does not satisfy the uniqueness requirement.

About This Task

In some cases you may want to create a dataset from a query. This allows you to manipulate the tables you have in your data warehouse and select just the rows you are interested in for your cube model. For example, you can select only the columns you need, join rows from multiple tables, and do other data manipulation functions. Query-based datasets do not alter your data warehouse.



Note: You cannot add calculated columns to a query-based dataset after it is created. Do all of the required data manipulation in the SELECT list of the query.

Procedure

1. Go to the main cube canvas.
2. Open the Library panel.
3. In the Library panel, click the menu icon and select **Add New Dataset**.
4. Give the dataset a name.
5. Select a data source:

- ▲ **Select a Table:** Specify the database, schema, and table.
- ▲ **Specify a Query:** Enter a valid SELECT statement. Ensure that the syntax of the statement is valid for the data warehouse you are using.

Some things to note:

- ▲ Only one `SELECT` statement is allowed.
- ▲ Make sure to include the schema name when referring to a table or view in the FROM clause.
- ▲ If your data warehouse is in a Hadoop cluster and you are using more than one SQL-on-Hadoop engine (for example SparkSQL and Hive), you must use SQL functions, data types, and SQL syntax that is supported by the SQL-on-Hadoop engines that you are using.
- ▲ You do not need to include a semicolon to denote the end of a statement.

6. If you specified a query, click **Preview SQL**.



Note: This button is disabled if your data warehouse is an instance of Google BigQuery. The reason is that Google BigQuery does not respect the LIMIT keyword in SQL statements that would limit the size of the result set in the preview. Rather, BigQuery performs a full table scan, which is costly.

If results appear, the SQL statement is valid. If you see an error or do not see data, double-check your SQL syntax and try running the statement directly against your data warehouse.

7. To create a limited dataset, turn on the **Limited Dataset** toggle. Limited datasets allow cube designers to create a smaller version of their primary dataset in order to quickly preview changes in AtScale or BI tools. If you turn on this option, you can either:
- ▲ **Select a Table:** Specify the database, schema, and table.
 - ▲ **Specify a Query:** If you have existing queries, you can select one from the drop down menu; or enter a custom query directly in the Selection SQL form field. Select the **Preview SQL** option to see the records returned from the query. If results appear, the SQL statement is valid.
8. Turn on or off **Enable Aggregate production from this table**. Turn on this option if you want the AtScale engine to be able to define aggregate tables that are based on queries against the query data set. If you do not select this option, the engine must run the query that the dataset is based on every time the dataset is queried.
9. Turn on or off **Allow incremental builds**. If this option is turned on, aggregate instances can append new rows and update existing rows that fall within the time period you specify. For more information, see [About Incremental Rebuilds](#).
10. Click **Save** to save this query dataset.

The new QDS would be displayed on the canvas using a color that differentiates it from the other datasets included in the cube. It would also be added to the Library; you can be edit it later as needed.

What To Do Next

If you need to edit the query dataset, locate it in the Library, click the ellipsis menu icon to the right of its name, and select **Edit**.