

Starburst Enterprise Power BI Guide

In this Document

- Overview1**
- Prerequisites.....1**
- Architecture.....2**
- Connectivity**
 - DirectQuery.....3
 - Import5
 - Direct ODBC.....5
- Summary.....8**



Overview

Microsoft's Power BI is a powerful business intelligence and reporting tool. It consistently leads the Gartner Magic Quadrant for Analytics and Business Intelligence Platforms and is used by many enterprise businesses today.

Starburst Enterprise is a fully scalable enterprise query engine that provides Power BI users the ability to query petabytes to terabytes of data on their data lake as well as federating data between many different source systems using standard ANSI SQL. Starburst Enterprise allows you to connect with all of your data no matter where it lives removing the need to copy or move your data.

Starburst has partnered with Microsoft to provide a DirectQuery connector. DirectQuery is a tool used to connect Power BI to a data source without the need to copy that data into Power BI. DirectQuery queries the data source every time there is a change in the filters.

Starburst's DirectQuery connector enables users to overcome some of the limitations of the tool. Starburst's DirectQuery connector allows you to query very large data sets as well as join data from multiple source systems using Starburst's powerful MPP query engine in real-time, eliminating the need to import data into the client's system first.

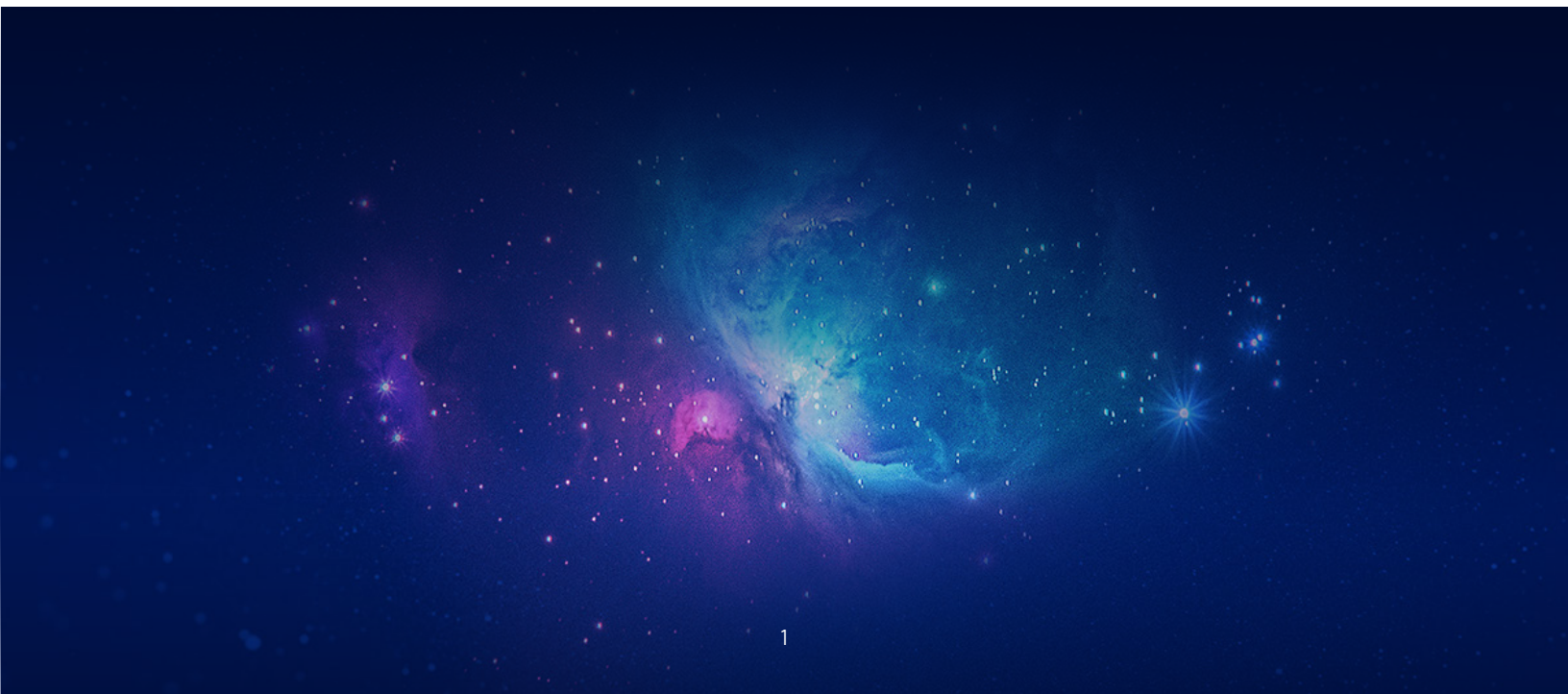
This guide will walk through the capabilities of the DirectQuery connector as well general usage of Power BI with Starburst.

Prerequisites

Power BI uses ODBC to communicate with a Starburst Enterprise cluster and issue SQL queries. The driver can be obtained from your Starburst representative and once installed, Power BI can connect and execute queries to any Starburst cluster.

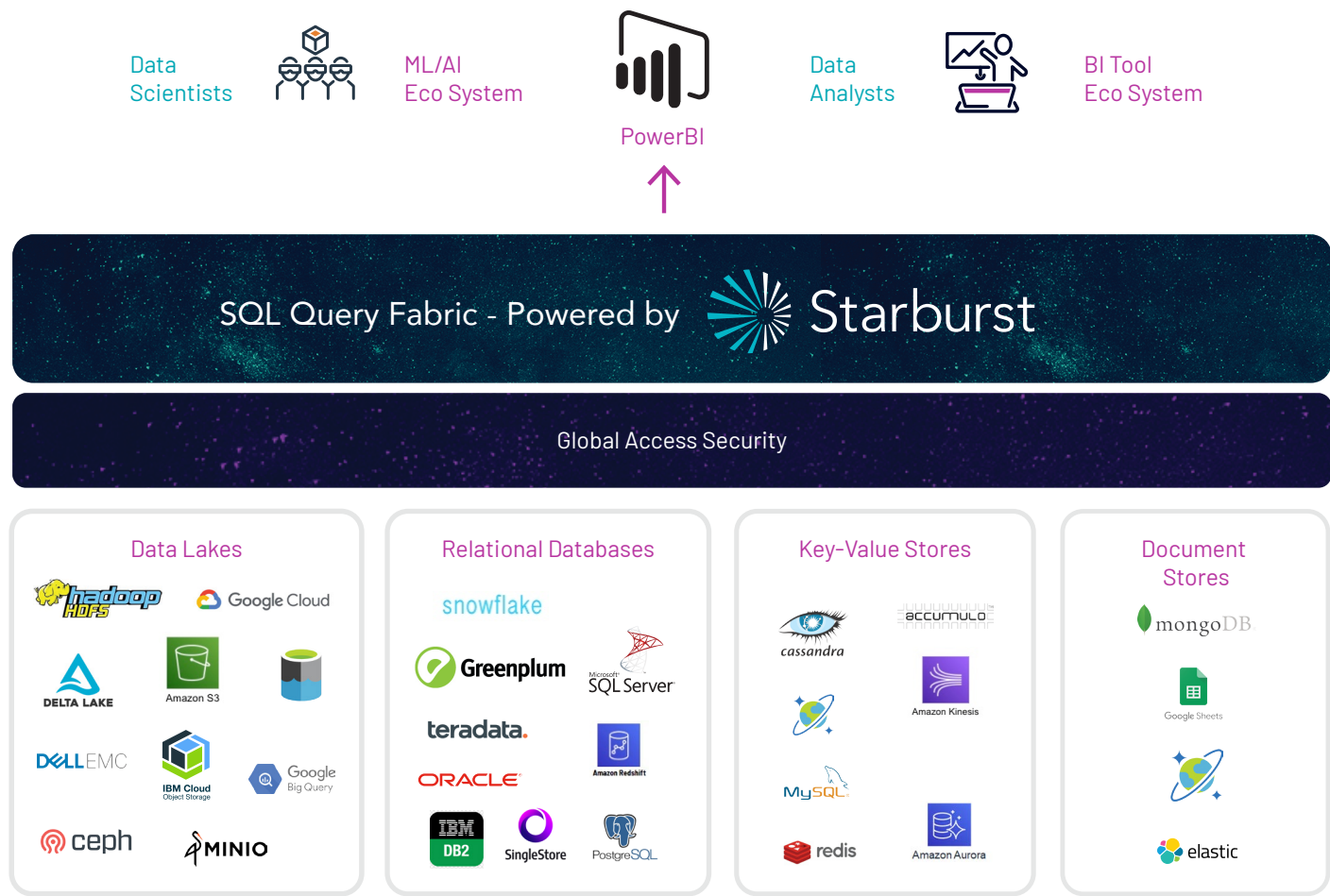
Note - The Starburst DirectQuery connector is built into Power BI and only requires the Starburst ODBC driver thus a ODBC profile is only needed when using the ODBC mode which will be covered later in this document.

If your Starburst Enterprise cluster has TLS enabled (recommended for production installs) then you must make sure the server's certificate has been added to the system trust store (Certificates -> Trusted Root Certification Authorities). The certificate must be added to the Gateway machine, or for each user running the Power BI connector.



Architecture

When using Power BI to access data in your data lake or many other supported data sources, it is important to provide seamless access while ensuring standard data security best practices.



Starburst’s Global Access Security enforces security policies down to the column and row level regardless of the access method. Using Power BI, users automatically adhere to these policies which also includes data masking for sensitive data. The diagram below illustrates the architecture of having a truly single point of access to all of your data using Power BI.

clerk	totalprice	orderdate
xxxxxxxxxx0000	97,027.36	Sunday, February 09, 1992
xxxxxxxxxx0000	22,957.22	Saturday, February 15, 1992
xxxxxxxxxx0000	96,085.16	Sunday, March 08, 1992
xxxxxxxxxx0000	131,802.19	Wednesday, April 29, 1992
xxxxxxxxxx0000	61,092.67	Monday, May 04, 1992
xxxxxxxxxx0000	324,708.22	Saturday, June 06, 1992
xxxxxxxxxx0000	250,407.06	Monday, June 15, 1992
xxxxxxxxxx0000	87,911.51	Monday, June 22, 1992
xxxxxxxxxx0000	137,822.11	Friday, June 26, 1992

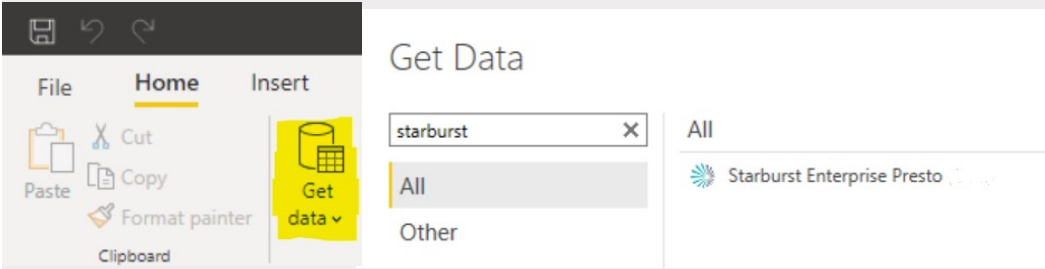
An example of Starburst’s Global Security can be seen in the screenshot of a report where the clerk column is masked. This is transparent to the user so data can be secured based on the user or group that is executing the report.

Connectivity

Connecting to a Starburst cluster can be accomplished either via Starburst DirectQuery or via the ODBC Standard Connector. Using the Starburst DirectQuery connector is the preferred way to access your data as it queries the data using the Starburst cluster without bringing all of the data back to the client machine running Power BI.


DirectQuery

This allows queries to span multiple terabytes along from either a single source or multiple sources all from a user's desktop. As mentioned above, the Starburst DirectQuery connector uses the ODBC driver already installed but does require any initial setup or profile.



The screenshot shows the Power BI ribbon with the 'Home' tab selected. The 'Get Data' button is highlighted in yellow under the 'Insert' group. The 'Get Data' dialog box is open, showing 'starburst' in the search bar and 'Starburst Enterprise Presto' as a result.

To connect using this method using the desktop Power BI application, you navigate to the Home/Get data/more option in the ribbon and choose Starburst Enterprise.



The screenshot shows the 'Starburst Enterprise Presto' dialog box. It has fields for 'Host' (presto-demo-ssl.az.starburstdata.net) and 'Port' (8443). Under 'Data Connectivity mode', 'DirectQuery' is selected. There are 'OK' and 'Cancel' buttons at the bottom.

A dialog box will appear asking for the host name of the Starburst cluster as well as the port. From here, you can choose Import or DirectQuery. Both options will issue queries to the Starburst cluster.



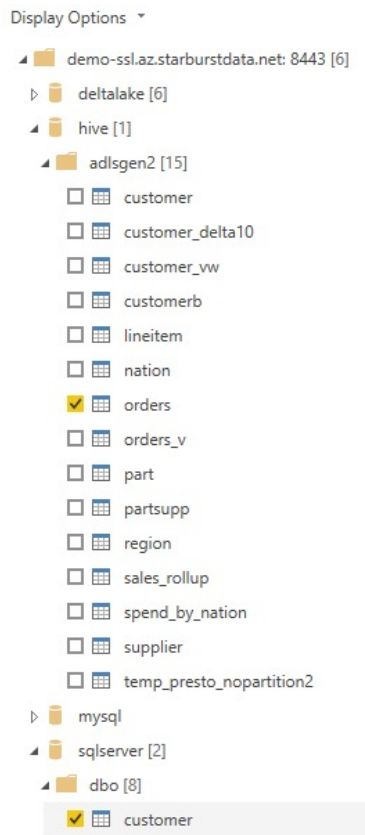
The screenshot shows the authentication dialog box for 'Starburst Enterprise Presto'. It has a sidebar with 'Anonymous', 'Kerberos', and 'LDAP' options. The 'User name' and 'Password' fields are visible. There are 'Back', 'Connect', and 'Cancel' buttons at the bottom.

The next dialog box will ask for an authentication method which depends upon how your Starburst cluster is set up. The LDAP option is the most common because it supports LDAP and Active Directory.

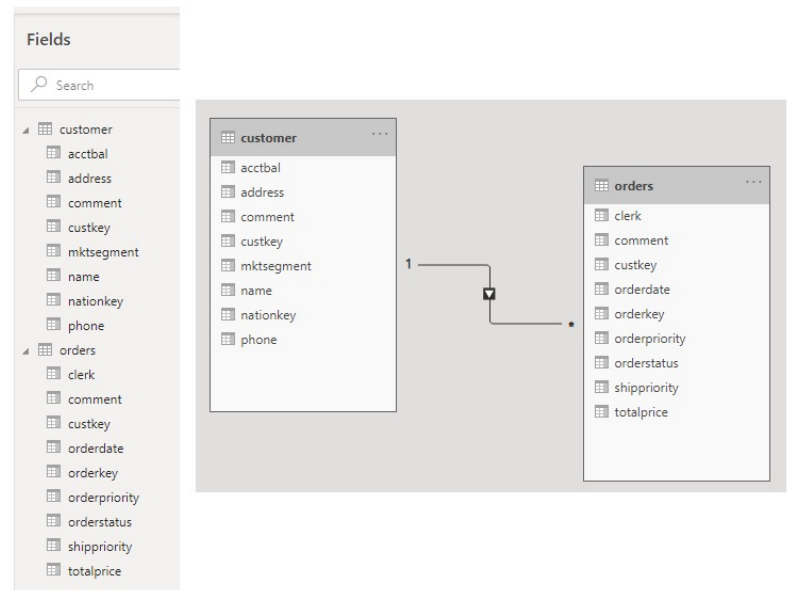
Once the user enters their credentials, the navigator screen will appear with all of the catalogs available for this user.

In the diagram below, this user has access to 4 different catalogs and some built-in ones which include system and tpch which is a data generator connector used for testing and benchmarking.

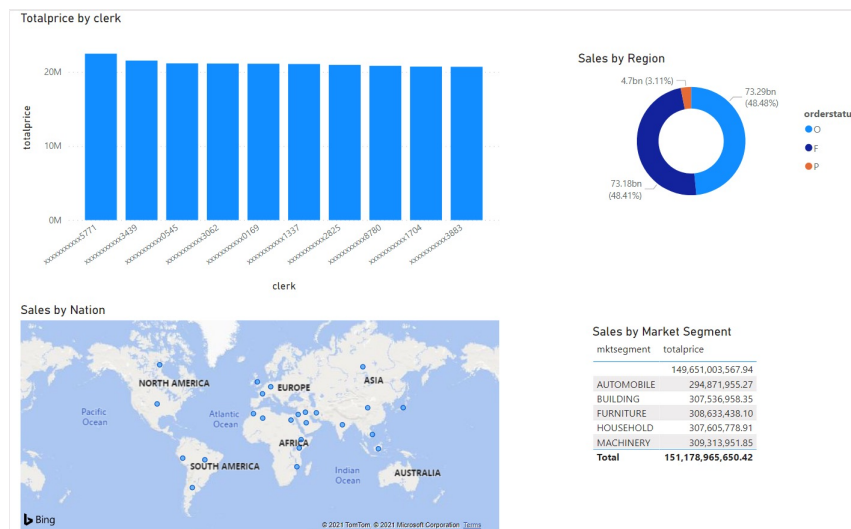
From here, each source can be expanded to list all of the tables available accessible for querying. In the example below, the ADLS GEN 2 datasets and SQL Server tables are chosen:



After the tables and columns are chosen, they will appear in the main Power BI report section as well as the modeling section as shown below:



Now, reports from different data sources spanning many terabytes of data can easily be generated:



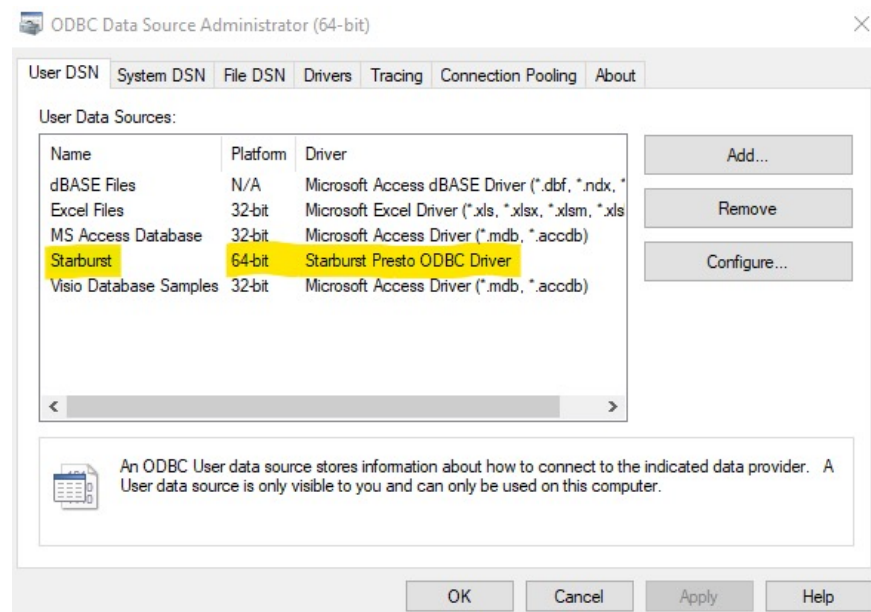
Import

Another option for querying data is using the Import mode instead of DirectQuery. This would be used if the user needs to bring a data set back to their desktop for further analysis such as using DAX functions, quick insights and calculated tables. Although this still uses the power of the Starburst cluster to query, it's limited to 1GB of data.

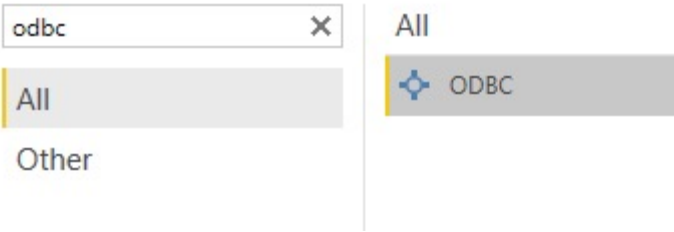
Direct ODBC

There are times where a connection to a Starburst cluster would be used using standard ODBC connectivity rather than going through the DirectQuery connector such as using custom SQL which isn't supported by Microsoft's DirectQuery framework. Using this method reduces the amount of data that can be queried because the data must be brought back to the Power BI client which is limited to 1GB data.

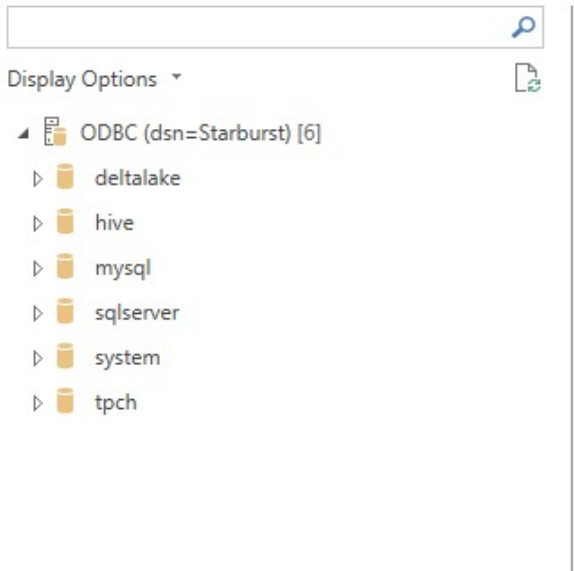
In order to query using ODBC, a DSN (data source name) needs to be created first in the end user's Windows machine. This is accomplished using the ODBC Data Source Administrator utility as shown below:



Get Data



Navigator



From here, a list of Starburst catalogs appear similar to the DirectQuery section above.

Summary

Power BI is one of the most popular BI tools in the industry. When paired with the power of a Starburst MPP query engine and the DirectQuery connector, users can finally overcome some of the tool’s limitations and unlock insights into their data lake and many other data sources using fast interactive queries over terabytes of data. Either through ad-hoc reporting or scheduled reports using the Power BI Gateway service, Starburst can provide a single point of access into all of your data sources in a highly performant and secure way.



Starburst
ANALYTICS ANYWHERE