

Enable Hybrid Analytics with Dell ECS and Snowflake

Hybrid Analytics with Dell ECS, Snowflake Data Cloud, Microsoft Power BI, and Faction Multicloud Enablement

May 2023

H19480

White Paper

Abstract

This paper describes how to maximize your on-prem Dell ECS investment for multicloud business intelligence with Snowflake Data Cloud and cloud adjacent data centers powered by Faction.

Dell Technologies Solutions

Copyright

The information in this publication is provided as is. Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © 2023 Dell Inc. or its subsidiaries. Published in the USA May 2023 [H19480].

Dell Inc. believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

Contents

- Executive summary.....4**
- Use case6**
- Solution overview.....6**
- Architecture.....7**
- Creating external stages and tables8**
- Configuring Power BI with Snowflake10**
- Conclusion.....14**
- References.....15**

Executive summary

Overview

Many organizations have made migrating analytics to the cloud a corporate imperative. Object storage is a key business driver for cloud migration, providing a cost-effective repository for data and rapid scaling when those data volumes grow. New forms of cloud architecture isolate workloads and support multitenancy across teams and projects. It frees analytics teams to run traditional Business Intelligence (BI) and reporting workloads independent of data science projects or data loads. The teams can all leverage the same source data without competing for CPU and memory resources.

However, migrating all your data to the cloud carries a lot of risks. There may be unintended costs and complexities, and businesses often decide that certain data is too mission-critical or too private to place on a public cloud.

About Dell Technologies

Dell Technologies helps organizations and individuals build their digital future and transform how they work, live and play. The company provides customers with the industry's broadest and most innovative technology and services portfolio for the data era.

ECS is the leading object storage platform from Dell Technologies, with unmatched scalability, performance, resilience, and economics¹. ECS delivers rich S3-compatibility on a globally distributed architecture, empowering organizations to support enterprise workloads such as cloud-native, archive, IoT, AI, and big data analytics applications at scale.

About Faction

Faction provides a fully managed cloud-data-services platform. Faction also offers patented, low-latency, high-throughput network connectivity that can deliver ultrahigh performance from ECS systems hosted adjacent to native cloud services.

About Snowflake

Snowflake enables every organization to mobilize their data with Snowflake's Data Cloud. Customers use the Data Cloud to unite siloed data, discover and securely share data, and execute diverse analytic workloads. Wherever data or users live, Snowflake delivers a single data experience that spans multiple clouds and geographies.

About Power BI

Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights.

Audience

Organizations that want to visualize data using Power BI with Snowflake and Dell ECS.

¹ Based on Dell analysis comparing cyber-security software capabilities offered for Dell ECS vs. competitive products, September 2022.

Revisions

Date	Part number/ revision	Description
May 2023	H19480	Initial release

We value your feedback

Dell Technologies and the authors of this document welcome your feedback on this document. Contact the Dell Technologies team by [email](#).

Authors: Chetan Somaiah and Rich Paulson

Note: For links to other documentation for this topic, see the [ECS Info Hub](#), and [guides that pertain to data analytics](#).

Use case

Customers often have data distributed across on-prem, public clouds, and co-located datacenters to meet specific business needs. Having distributed data repositories makes it challenging to discover, access, and process data on demand. The problem is further aggravated when data is stored in proprietary formats and is not interoperable with your tools of choice.

In this use case, the customer has data stored in Dell ECS that is located either on-premise or in a cloud adjacent datacenter such as Faction, and in an open format, that is, Parquet file format. They also have data stored in a Snowflake Database within the same Snowflake Data Cloud. This paper demonstrates how to access the data stored in Dell ECS using the Snowflake query engine and visualize the results using business intelligence software.

In this paper we demonstrate two use cases:

1. Visualize data using Microsoft Power BI by accessing the data stored in Dell ECS on-premise or located in a Faction cloud adjacent datacenter, using Snowflake Data Cloud in AWS.
2. Visualize data using Microsoft Power BI by accessing the data stored in Dell ECS and combine it with data stored in Snowflake Database using Snowflake Data Cloud in AWS.

Solution overview

Dell has partnered with Faction Inc. to deliver a fully managed, cloud-based service to address various cloud use cases. Faction is a multicloud platform-as-a-service provider and VMware partner that offers multicloud attached storage from various co-locations. The service is available in several locations in the United States and in Europe. In this hybrid-cloud solution, Dell ECS powered by Faction provides the storage layer or data lake for the Snowflake Data Cloud.

Cloud adjacent, multicloud connectivity

The Dell ECS object storage powered by Faction is integrated into Faction's patented low-latency, high-throughput FIX (Faction Internetwork eXchange) infrastructure, thereby providing instant cloud adjacent connectivity to any native cloud provider. In this solution the Snowflake instance is hosted in AWS, as is the Power BI app, allowing data hosted on-prem to be used by Snowflake for unique business intelligence dashboards in AWS.

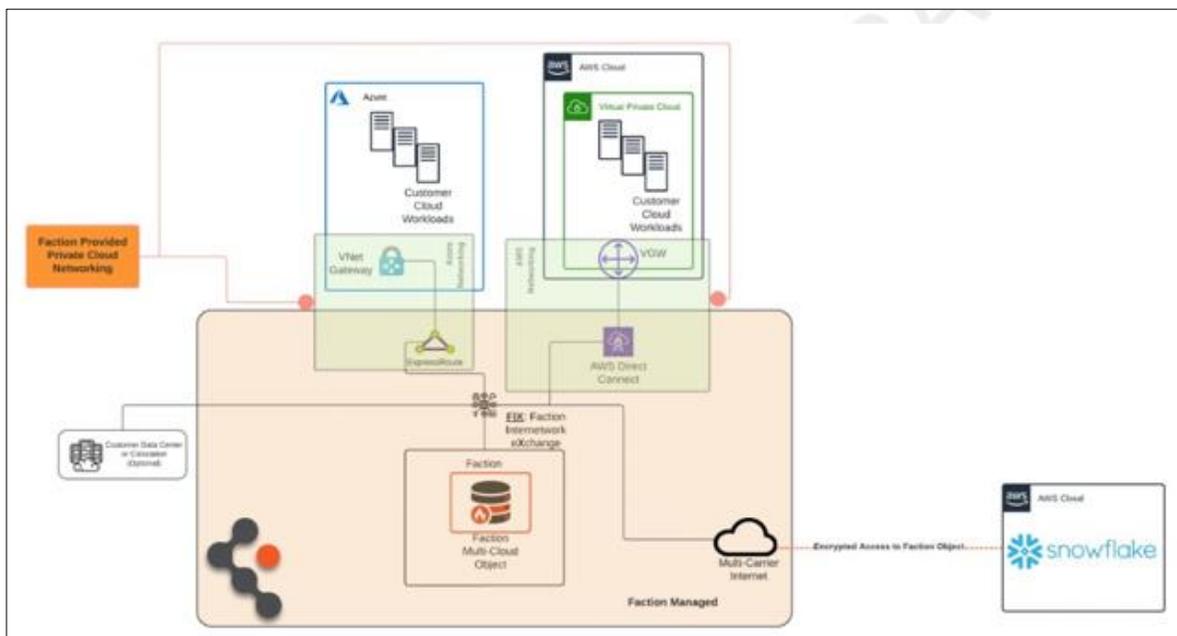


Figure 1. The FIX multicloud connectivity provided by Faction and the collocated infrastructure used to connect to the Dell ECS within the Faction datacenter

Architecture

Power BI can access Snowflake external tables through an ODBC Connector. This means that Power BI can query an external table for data that resides in Dell ECS to create insights against data processed in Snowflake using virtual warehouses.

The access pattern that Snowflake supports today traverses the AWS network backbone and exits to the public internet to reach the on-premises Dell ECS. This means that the ECS S3 endpoint must be publicly accessible.

Snowflake, Faction, and Dell ECS use Transport Security Protocol (TLS) 1.2 for data in-transit. All data traffic is encrypted using TLS, thus preventing vulnerable access points from cyber-attacks.

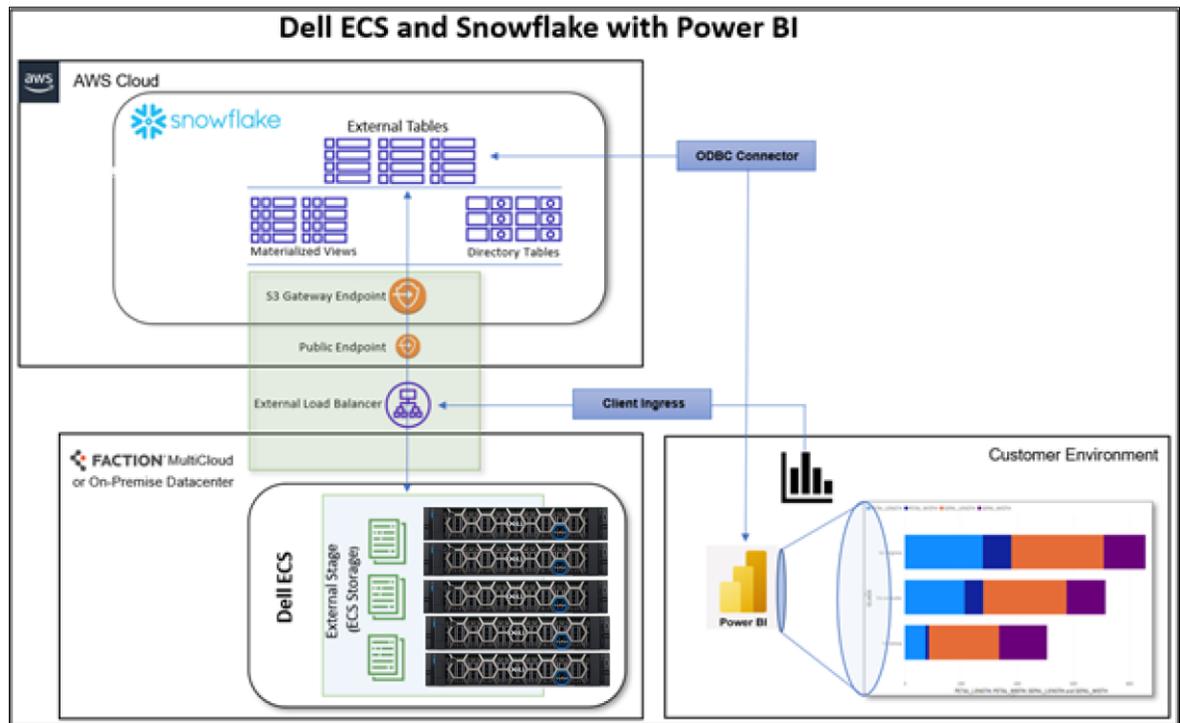


Figure 2. The architecture and data path from Power BI, to Snowflake, accessing the data from the ECS system that resides on-prem or in Faction’s multicloud datacenter

Terminology

The following table provides definitions for some of the terms used in this document.

Table 1. Terminology

Term	Definition
External Stage	An external (that is, S3) stage specifies where data files are stored so that the data in the files can be loaded into a table or queried in-place.
External Table	External tables store file-level metadata about the data files, such as the filename, a version identifier, and related properties. This enables querying data stored in files in an external stage as if it were inside a database.

Creating external stages and tables

External stage

Create named external stages using CREATE STAGE. A named stage object can be used to list files, load data, and unload files, in addition to other SQL actions.

```
create or replace stage ext_nyc_tripdata_stage
url='s3compat://datalake/warehouse/nyc-parquet/'
endpoint='<ECS Faction endpoint>'
'credentials=(
  AWS_KEY_ID='<ECS Access Key>'
  AWS_SECRET_KEY='<ECS Secret Key>');
```

External table

External tables enable querying data stored in ECS for analysis without first loading it into Snowflake. The CREATE EXTERNAL TABLE command references the EXTERNAL STAGE to allow connectivity to ECS located in the Faction Datacenter.

The following example creates a table using the Parquet format. Apache Parquet is an open source, column-oriented file format designed for efficient data storage and retrieval. It provides efficient data compression and encoding schemes with enhanced performance to handle complex data in bulk.

```
create or replace external table ext_nyc_tripdata (
  vendor_name TEXT as (value:vendor_name::TEXT),
  Trip_Pickup_DateTime TEXT as (value:Trip_Pickup_DateTime::TEXT),
  Trip_Dropoff_DateTime TEXT as (value:Trip_Dropoff_DateTime::TEXT),
  Passenger_Count NUMBER(38, 0) as (value:Passenger_Count::NUMBER(38,
0)),
  Trip_Distance REAL as (value:Trip_Distance::REAL),
  Start_Lon REAL as (value:Start_Lon::REAL),
  Start_Lat REAL as (value:Start_Lat::REAL),
  rate_code REAL as (value:rate_code::REAL),
  store_and_forward REAL as (value:store_and_forward::REAL),
  End_Lon REAL as (value:End_Lon::REAL),
  End_Lat REAL as (value:End_Lat::REAL),
  Payment_Type TEXT as (value:Payment_Type::TEXT),
  Fare_Amt REAL as (value:Fare_Amt::REAL),
  surcharge REAL as (value:surcharge::REAL),
  Tip_Amt REAL as (value:Tip_Amt::REAL),
  Tolls_Amt REAL as (value:Tolls_Amt::REAL),
  Total_Amt REAL as (value:Total_Amt::REAL)
)
location = @ext_nyc_tripdata_stage
auto_refresh = false
file_format = (type = parquet);
```

Use case 1: SQL query

The following SQL query was used to build the Power BI Dashboard. The script populates the `ext_nyc_tripdata` table with the data for the Power BI dashboard. This data is sourced from existing object data located in the `external_s3` bucket on ECS.

```
select t.range, count(*) as "Number of Occurrences",
ROUND(AVG(fare_amount),2) as "Avg",
  ROUND(MAX(fare_amount),2) as "Max" ,ROUND(MIN(fare_amount),2)
as "Min"
FROM (
  select
  case
    when trip_distance between 0 and 9 then '0-9 '
    when trip_distance between 10 and 19 then '10-19'
    when trip_distance between 20 and 29 then '20-29'
    when trip_distance between 30 and 39 then '30-39'
    else '> 39'
```

Configuring Power BI with Snowflake

```
end as range ,fare_amount
FROM ext_nyc_tripdata ) t
where fare_amount > 1 and fare_amount < 401092
group by t.range;
```

Use case 2: SQL query

The following SQL query is used to access data residing in Dell ECS and within a Snowflake Database to visualize the combined sources in Power BI dashboard. Federated queries enable business users, data scientists, and data analysts to run queries across distributed data stores. A user can submit a SQL query that is executed across multiple data sources in place for a holistic view of their data.

```
Select * from snowflake_sample_data.tpch_sf10.lineitem as t1,
ext_nyc_tripdata as t2 where t2.VENDORID=t1.L_ORDERKEY;
```

Configuring Power BI with Snowflake

This section describes how to configure Power BI to connect with Snowflake. When the connection is made to Snowflake, Power BI can work with table data stored in ECS referenced by the Snowflake database.

Note: This example is using Use Case 1, where the data is being accessed directly from ECS residing on-premise.

1. Launch Power BI then click Get Data from the Tool Bar.

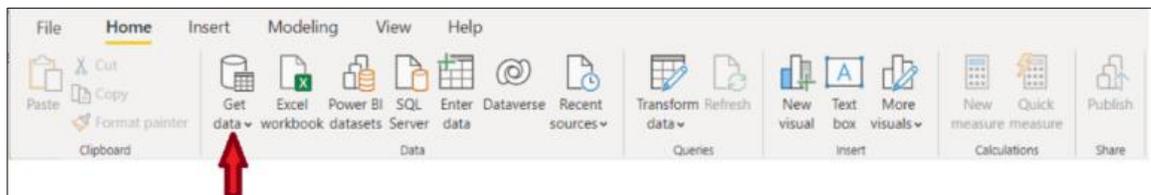


Figure 3. Click 'Get data' to select a connector

2. Select 'More...' and search for Snowflake, then select Snowflake and click 'Connect'.

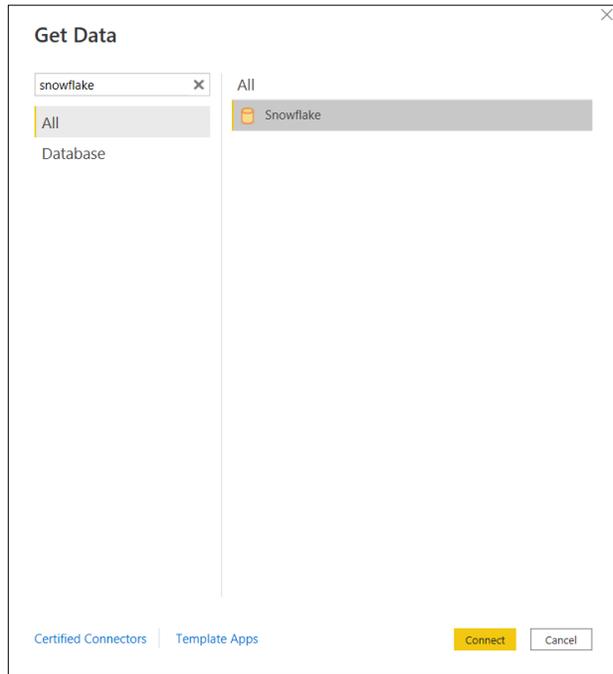


Figure 4. Search for the Snowflake connector

3. Enter the Snowflake Data Cloud account URL, Warehouse, Database, and SQL query to execute, then click 'OK'.

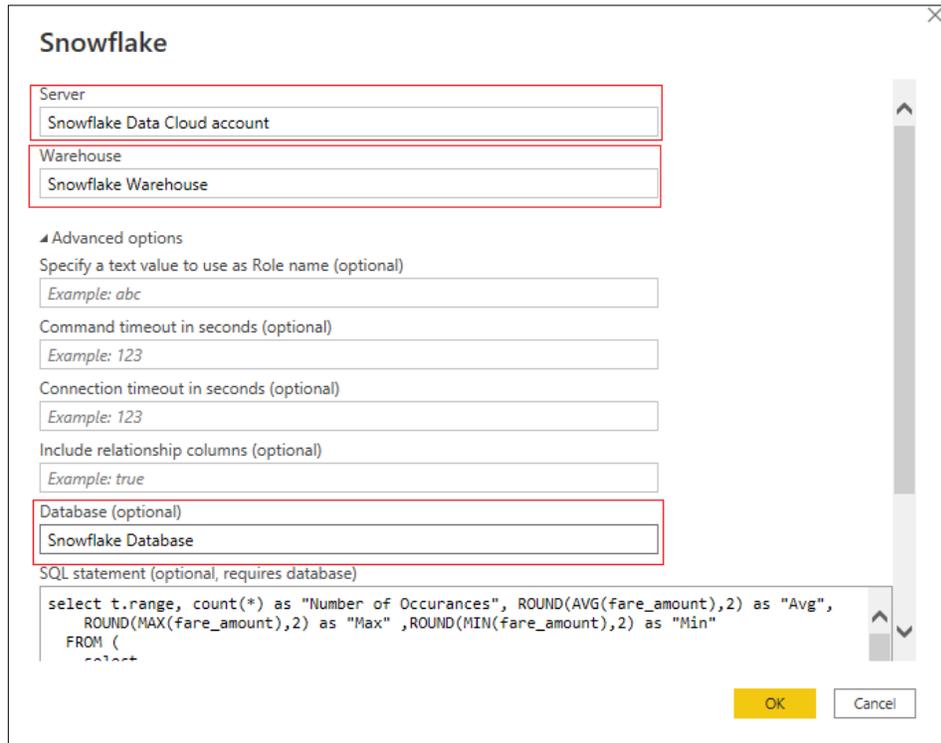


Figure 5. Enter the Snowflake account details

4. Enter the login credentials for your Snowflake account.

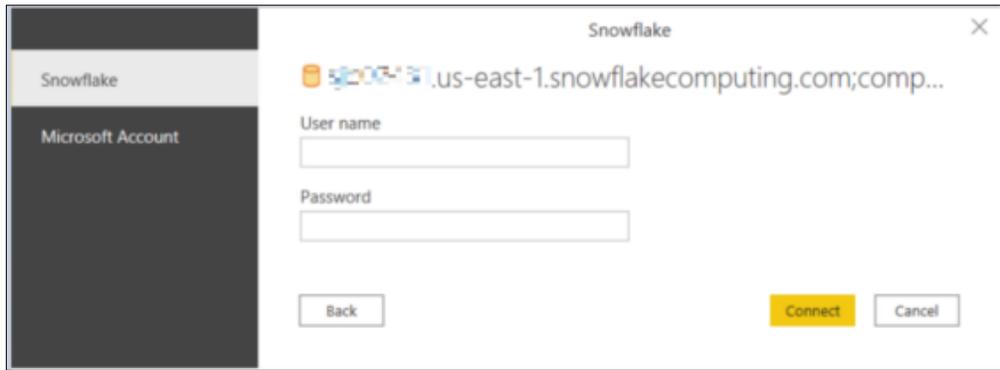


Figure 6. Enter the Snowflake account login credentials

5. When connected, Power BI executes the query that was entered when setting up the Snowflake connector in the previous step and shows the results. From here you can load or transform the data. In our case, we select Load.

RANGE	Number of Occurances	Avg	Max	Min
0-9	2889923	11.84	1274	1.1
10-19	229745	45.53	533	2
> 39	72167	39.71	1233.5	2
30-39	1447	117.13	342.5	2.5
20-29	33430	59.07	412	1.41

Figure 7. SQL Query results

6. Select whether to Import the data or query the data directly. Select either option based on the use case, then click OK.

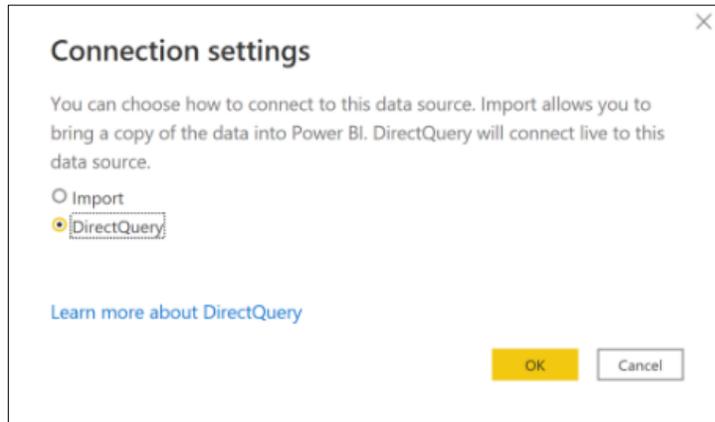


Figure 8. Choose whether to import the data locally or execute a direct query against Snowflake

7. When the connection to Snowflake is established, start building your dashboard by selecting the fields on the right-most panel of the Power BI desktop.

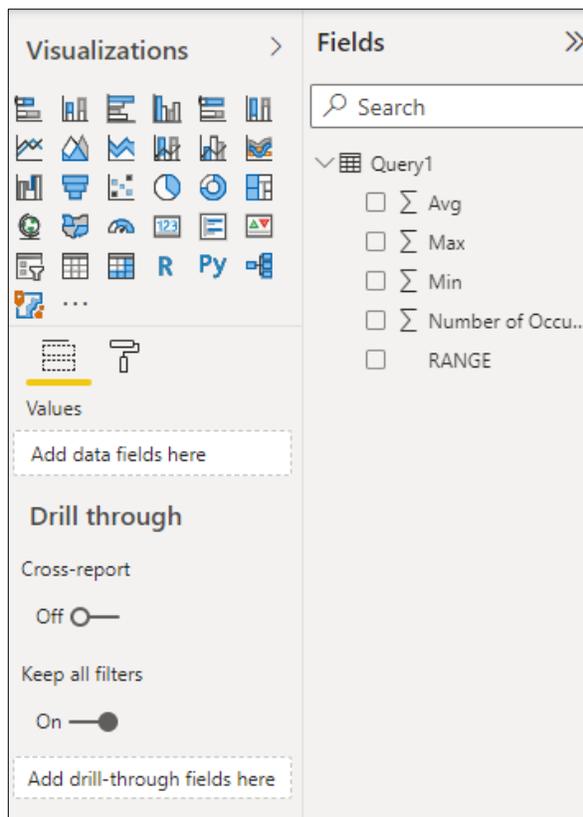


Figure 9. Start building your Dashboard

Example dashboard

After populating the data into Power BI, the dashboard begins to take shape and visually displays the results for analysis and interpretation.

Conclusion

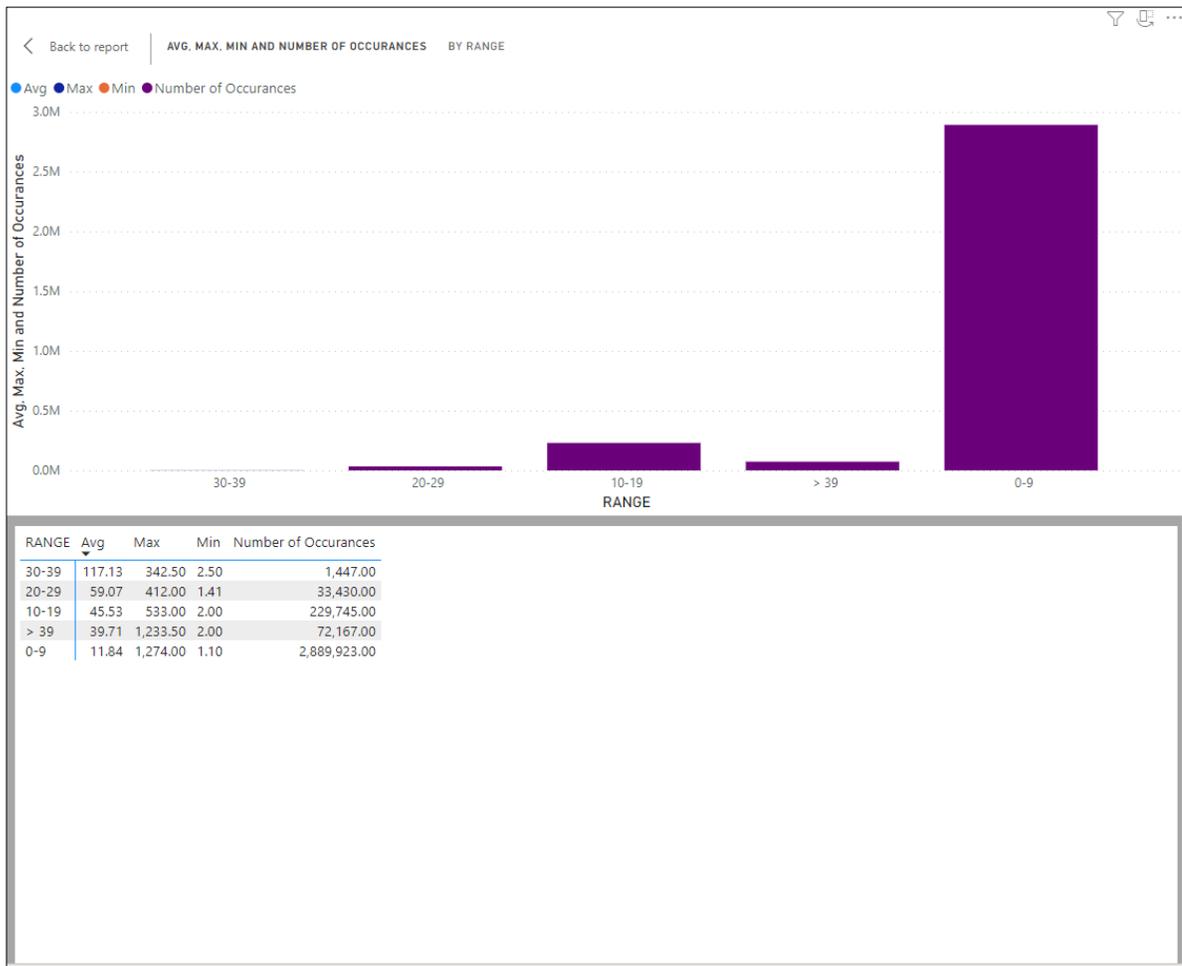


Figure 10. Example Dashboard

The Snowflake History tab will show the SQL executed in Snowflake by the ODBC connector. This can be a useful tool for troubleshooting issues when executing SQL queries.

Status	Query ID	SQL Text	User	Warehouse	Clust...	Size	Session ID	Bytes Scanned	Client Info
✓	01aa2198-...	select "RANGE", "C1", "C2", "C3", "C4" from (select "RANGE", sum("Number of Occurrences") as "...	DEMO	COMPUTE...			1652623958...		✓ ODBC 2.24.6

Figure 11. Snowflake history

Conclusion

It is common to maintain large on-prem storage repositories for a variety of business purposes. Historically the data in these repositories was only available to the applications and infrastructures running on premises. Now these repositories can be accessed by applications running in native cloud providers through high-speed cloud adjacent connectivity and the data used for temporary business insights or long-term research. This is a major change to how businesses have traditionally thought about using their data and a key value add to organizations thinking about implementing hybrid infrastructures to broaden business capabilities.

References

Dell Technologies documentation

The following Dell Technologies documentation provides other information related to this document. Access to these documents depends on your login credentials. If you do not have access to a document, contact your Dell Technologies representative.

- [Dell ECS Info Hub](#)
- [Data analytics guides](#)

Snowflake documentation

The following Snowflake documentation provides detailed information related to this solution.

- [Working with Amazon S3 Compatible Storage](#)
- [Introduction to External Tables](#)

Faction Inc.

Learn more about Faction's multicloud data services with Dell Technologies by visiting <https://www.factioninc.com/dell/>.