Databricks to Excel

Exponam.Connect or ODBC? – Which approach is best?

TL;DR

Exponam.Connect is the ideal choice for business and technical users to pull data from Databricks into Excel. ODBC is a second option, but requires more technical expertise, has poorer performance, and lacks the UI and power features demanded in many business use cases.

Summary Comparison

	Delta Sharing & Exponam.Connect	ODBC
Databricks workspace credentials required		✓
Unity Catalog permissioning/governance	✓	
Technical users	✓	✓
Non-technical users	✓	
Data extract speed	10x	x
Data extract size limits (rows)	5,000,000 / unlimited*	1,000,000
Easy configuration / installation	✓	
Cost	User count based (free use option)	Usage based

Introduction

Accessing Databricks data in Excel is a powerful and efficient way to extend the value of Lakehouse data. Pairing the power of the Databricks platform with Excel – the world's most universal data platform – is required for many business solutions.

Users can connect to Excel via Delta Sharing and the Exponam.Connect Excel Add-in or an ODBC driver and Excel's Get Data query engine.

Which approach is best? It depends on who will be using the data and how.

Considerations

Are Excel users required to have Databricks workspace credentials?	
Delta Sharing & Exponam.Connect	**ODBC**
No. Access is controlled via Unity Catalog through a .share file. This type of access can be granted to thousands of users within and external to an organization. Users do not require Databricks workspaces or credentials.	Yes. Requires a Databricks Workspace credential with permissions to use and run the Databricks cluster.
If an organization wishes to share data internally (and externally) to be accessed by end-users who do not have permissions to run the Databricks environment, Delta Sharing & Exponam.Connect is appropriate. If a data engineer or data scientist with a Databricks Workspace wishes to explore their Workspace's data in Excel, ODBC is a good option.	

* Unlimited extract refers to writing the extract directly to a CSV file

What is the user's level of technical competency?

Delta Sharing & Exponam.Connect	ODBC
o No setup required. User selects the .Share file assigned to them via a typical Windows dialog screen. o Pulling data is as simple as clicking the table name and filtering - similar to filtering columns within Excel.	o ODBC setup requires configuration of the Simba Spark ODBC Driver with parameters from the Databricks cluster (requiring Databricks workspace credentialed details) o Pulling data via the Excel Microsoft Query interface involves writing SQL select statements
Business users familiar with Excel filters will be comfortable using the Exponam.Connect Add-in. Technical users fluent in writing SQL statements and familiar with their Databricks Workspace environments will find familiarity in the environment to pull via ODBC.	

Cost

Delta Sharing & Exponam.Connect	ODBC
o No DBU (Databricks) fees associated with data discovery or data extract o Egress fees are minimized with compressed file downloads o Free and per-user licensing options are available for Exponam.Connect	o ODBC queries incur DBU fees associated with query execution and workspace cluster activation o Standard egress fees for the JSON results downloaded
Exponam.Connect costs are driven by the number of users and, to a small degree, the volume of data delivered. ODBC costs are driven by the number of queries executed and volume of data delivered.	

Performance

Delta Sharing & Exponam.Connect	ODBC
Speed is a function of the size of compressed Parquet files to be downloaded based on the filter parameters. Initial filtering is based on partitioning and file metadata; with final processing done on the client.	Speed of export is a function of time to spin up the cluster, query execution time, and data transfer. As extract sizes increase, data transfer becomes the dominant variable – greatly impacted by the type and size of the data.
Performance benchmarks with both narrow (three columns) and wide (100 columns) data sets show a 10x speed benefit with Delta Sharing (over ODBC) due primarily to data compression during transfer. E.g., a one million row narrow data set takes 10 seconds via Delta Sharing versus 95 seconds via ODBC.	

Data Size

Delta Sharing & Exponam.Connect	ODBC
Data Extracts to Excel are limited to 5,000,000 rows (spread across 5 worksheets). Larger datasets can be written directly to CSV or other data file from the UI within Excel.	Data Extracts to Excel are bounded by the Excel limit of 1,048,576 rows.
Extracts under a million rows can be performed by either tool set. Extracting larger data sets (e.g. annual sales or trades; inventory details; sensor information) spanning multiple Excel sheets or even larger are easily accomplished with Delta Sharing and Exponam.Connect.	

Contact

Herman Weintraub

hweintraub@exponam.com

+1.646.360.0110