Databricks to Excel

Exponam.Connect vs an ODBC-based solution – Which approach is best?

TL;DR

Exponam.Connect is the ideal choice for business and technical users to pull data from a Databricks Lakehouse into Excel. Exponam.Connect is the only option which accesses data via Delta Sharing. All other solutions on the market pull data via ODBC – with varying UI skins on top. As a result, Exponam.Connect is much faster (at least 10x) and much cheaper than all other options. And client testing routinely affirms that the Exponam.Connect experience is much easier and more intuitive than all others.

Summary Comparison

	Delta Sharing & Exponam.Connect	All Others
Databricks workspace credentials required		✓
Unity Catalog permissioning/governance	✓	
Technical users	✓	✓
Non-technical users	✓	
Data extract speed	10x	x
Data extract size limits (rows)	10,000,000 / unlimited*	1,000,000
Easy configuration / installation	✓	
Cost	User count based (free use option)	Usage based
Databricks ML Model Serving	✓ - run DBX models in Excel	x

Introduction

Accessing Databricks data in Excel is a powerful and efficient way to extend the value of the Databricks platform. Pairing the power of the Databricks 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-based connector and Databricks SQL cluster compute. Which approach is best? It depends on who will be using the data and how.

Considerations

Do the Excel users have Databricks workspace credentials?	
Delta Sharing & Exponam.Connect	**All Others**
Access is controlled via Unity Catalog recipients. 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.	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.	

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

If a data engineer or data scientist with a Databricks Workspace wishes to explore the Workspace's data in Excel, ODBC is a good option.

What is the user's level of technical competency?	
Delta Sharing & Exponam.Connect	**All Others**
○ No setup required. User is required to select the .Share file assigned to them via a typical Windows dialog screen. ○ Pulling data is as simple as clicking the table name and filtering in similar ways to filtering columns within Excel.	○ ODBC setup requires configuration of the Spark ODBC Driver with parameters from the Databricks cluster (again requiring Databricks workspace credentialed details) and a Databricks Personal Access Token ○ 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 other solutions – including SQL editors.	

Cost	
Delta Sharing & Exponam.Connect	**All Others**
○ **No DBU fees** associated with data discovery or data extract ○ Egress fees are minimized with compressed file downloads ○ Free and per-user licensing options are available for Exponam.Connect	○ Every data extract incurs DBU fees associated with query execution and workspace cluster activation ○ Standard egress fees for the JSON results downloaded
Exponam.Connect costs are driven by the number of users. ODBC-based system costs are driven by the number of queries executed and volume of data delivered. These ODBC processing costs dwarf Exponam.Connect User fees.	

Performance	
Delta Sharing & Exponam.Connect	**All Others**
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-based) due primarily to data compression during transfer. As data set size increases, the speed benefit is even more pronounced, 50x or more. E.g., a one million row narrow data set takes 9 seconds via Delta Sharing versus >100 seconds via ODBC.	

Data Queries	
Delta Sharing & Exponam.Connect	**All Others**
Row filtering and column inclusion/ordering via UI. Table joins via Views registered in the UC Catalog. Dynamic row filtering and column masking based on recipient properties.	Product specific capabilities to write direct SQL queries with any combination of table joins and selection criteria.
Exponam.Connect empowers non-technical users with most of the power of SQL select statements without writing SQL.	

Data Size	
Delta Sharing & Exponam.Connect	**All Others**
Data Extracts to Excel are limited to 10,000,000 rows (spread across 10 worksheets). Larger datasets can be	Data Extracts to Excel are bounded by the Excel limit of 1,048,576 rows.

Herman Weintraub
hweintraub@exponam.com
+1.646.360.0110

written directly to CSV or other data file from the UI within Excel.	
Exponam.Connect returns large sets efficiently and quickly. ODBC performance degrades as size increases – and often, Excel runs out of memory prior to returning the full 1 million row maximum data set.	

Herman Weintraub
hweintraub@exponam.com
+1.646.360.0110