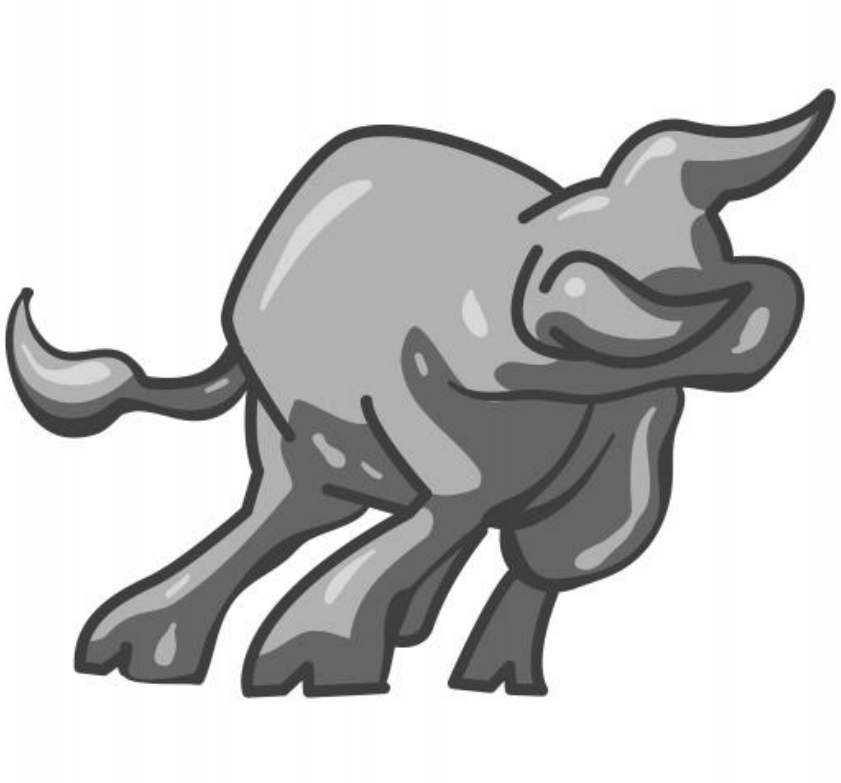
Synapse Analytics Serverless SQL

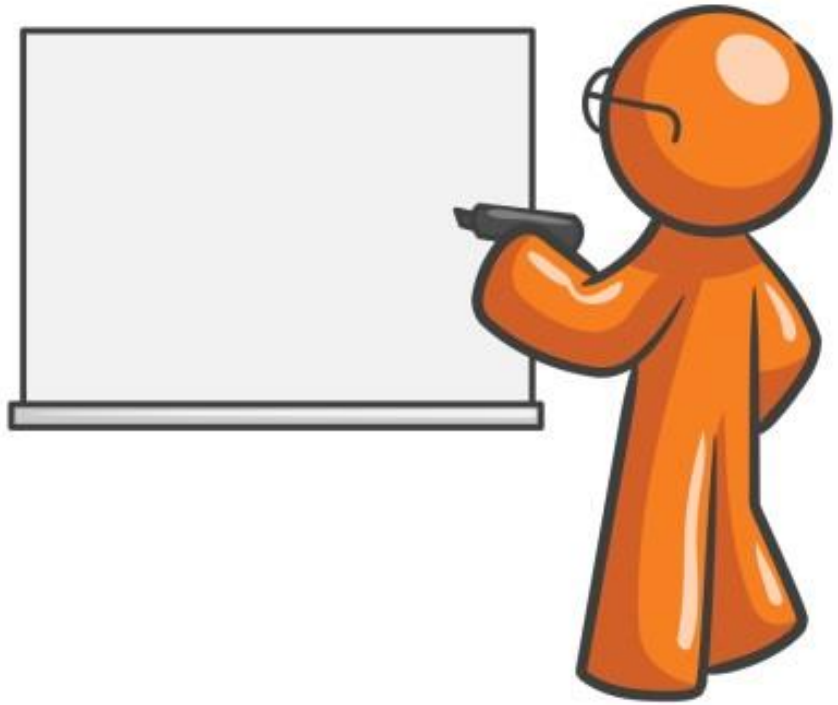
MSBID- 10. december 2020



ORANGEMAN

Agenda

- Quick Azure Synapse Analytics overview
- Serverless SQL
  - What is it
  - What can it be used for
- Cost management



ORANGEMAN

Azure Synapse Analytics

Limitless analytics service with unmatched time to insight

Azure Synapse is Azure SQL Data Warehouse evolved*—blending big data, data warehousing, and data integration into a* ***single service*** *for end-to-end analytics at cloud scale.*

ORANGEMAN

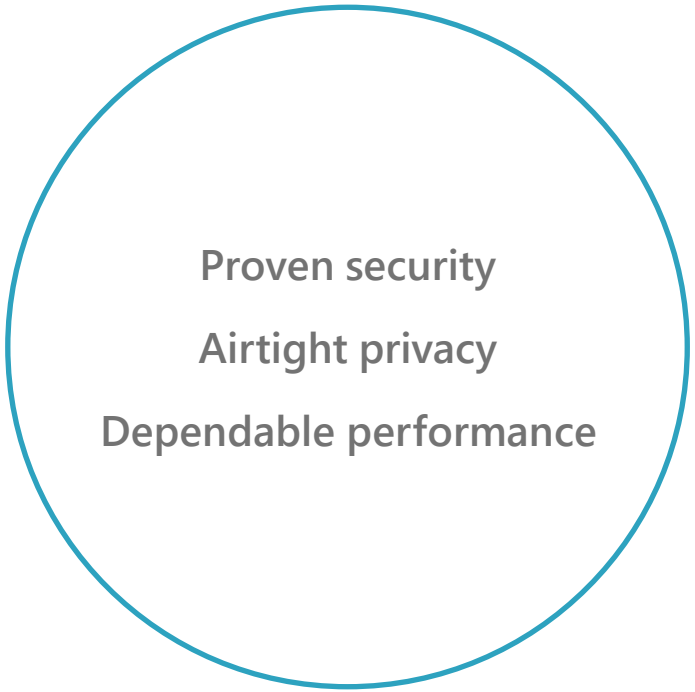
This is a result of businesses being forced to maintain two critical, yet independent analytics systems

Big Data

Data Lake

OR

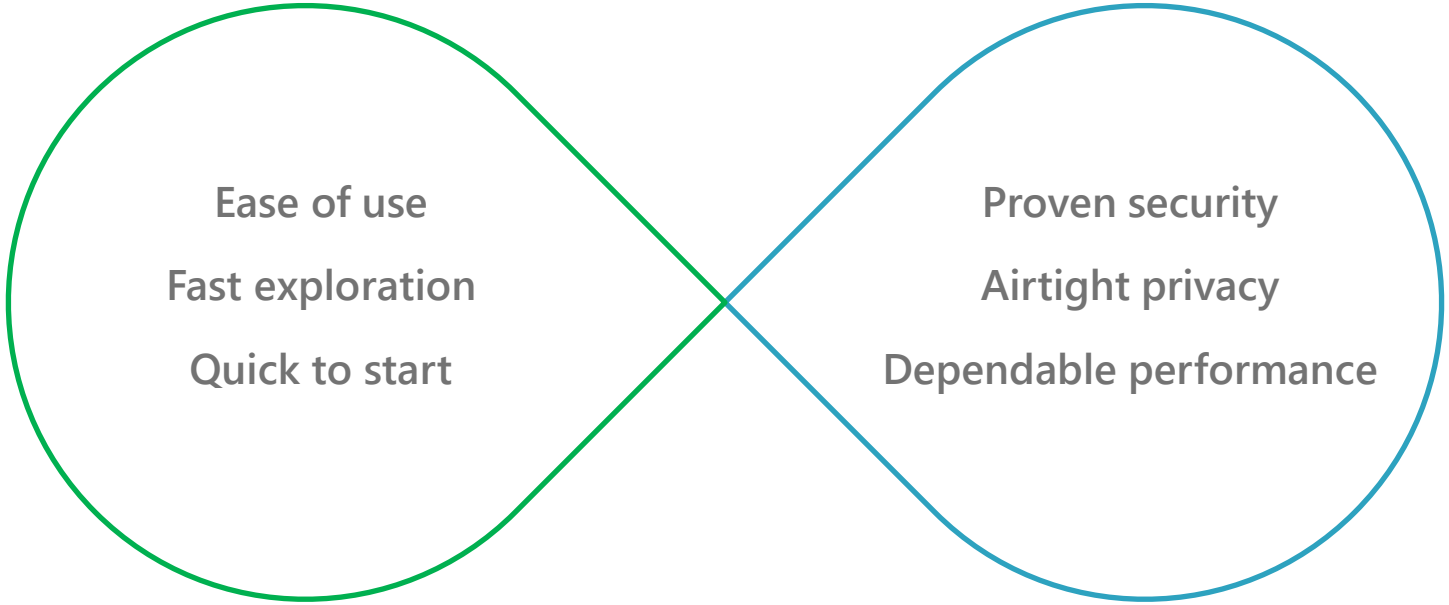
Relational Data



Data Warehouse

ORANGEMAN

Azure brings these two worlds together, in a single service

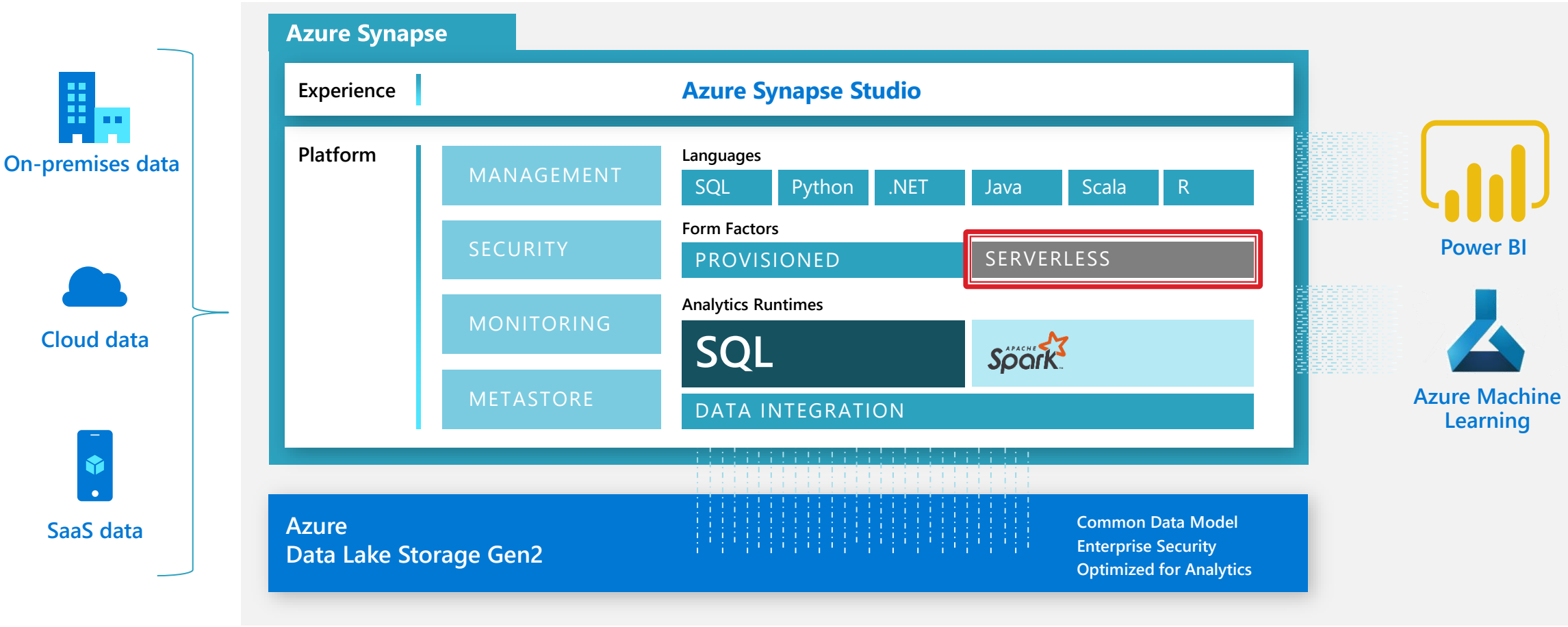


Welcome to Azure Synapse Analytics

ORANGEMAN

Analytics using Azure Synapse

Ingest and store data from multiple sources → Prepare your data → Analyze the prepared data → Serve your data with high concurrency support → Build BI dashboards and reports



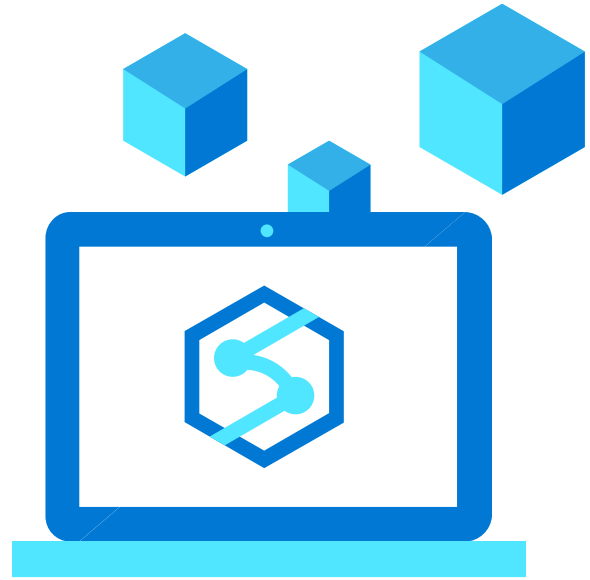
ORANGEMAN

Azure Synapse Analytics Terminology

- **Synapse workspace:** A securable collaboration boundary. Has an associated ADLS Gen2 account and file system (for temporary data).
- **Data Integration (Data Factory):** Ingest data between various sources and orchestrate activities running within or outside a workspace
- **SQL serverless:** Distributed data processing system that lets you run T-SQL queries over data in data lake. It is serverless.
- **SQL pool (SQL DW):** 0-to-N SQL provisioned resources with their corresponding databases
- **Apache Spark pool:** 0-to-N Spark provisioned resources with their corresponding databases.

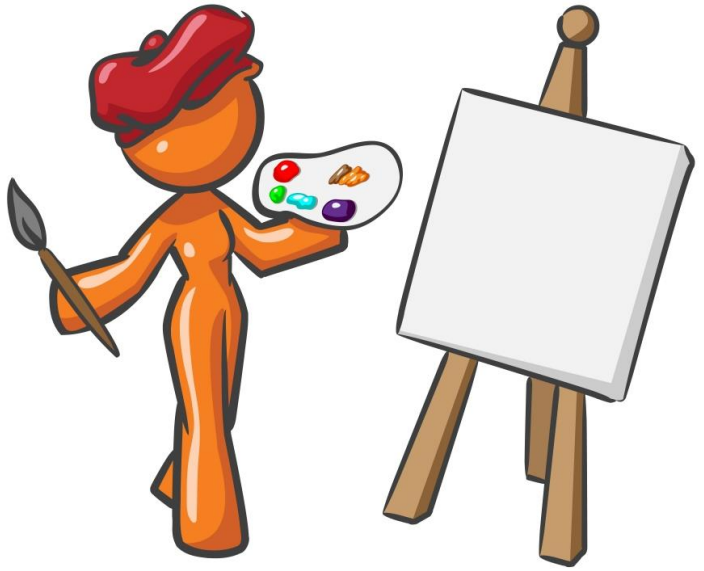
Synapse Studio

- Makes modern data warehouse solution-building easier than ever!
- Synapse Studio combines ingestion, preparation, analysis, and serving.
- Develop notebooks, SQL scripts, pipelines, Power BI reports, and more… all within one experience.
- Solution-level authoring and monitoring in a single pane of glass.
- Spark and SQL can operate on the same data. No need to duplicate.



ORANGEMAN

Demo

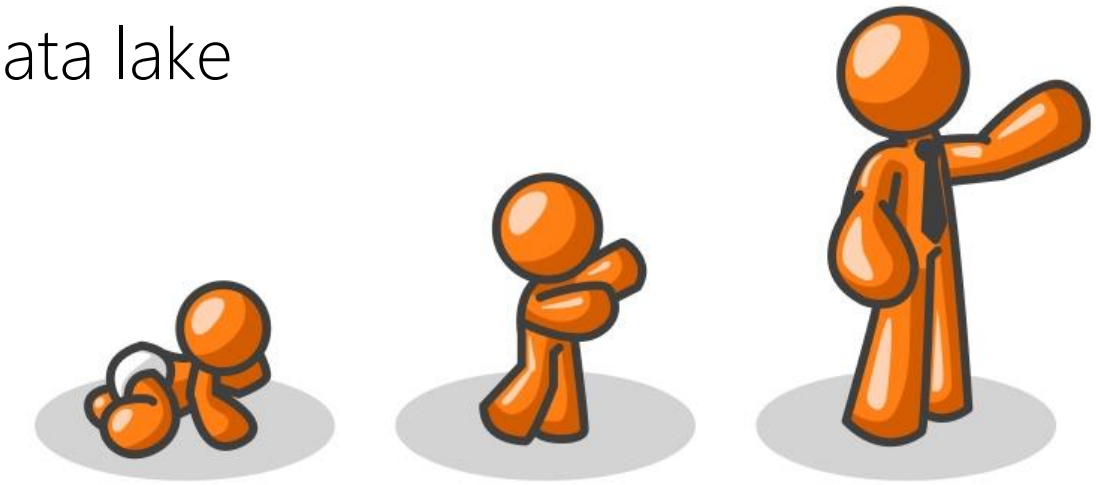


1. Create a Workspace and use existing Data Lake
2. Setup firewall rules
3. Set security on Data Lake
   - Storage Blob Reader or Contributor role
4. Launch Synapse Studio
   - Browse storage account
   - Query PARQUET file with the use of SQL serverless

Quickstart: https://docs.microsoft.com/en-us/azure/synapse-analytics/quickstart-create-workspace and https://docs.microsoft.com/en-us/azure/synapse-analytics/quickstart-synapse-studio

SQL serverless in Synapse

- Query service over the data in your data lake



ORANGEMAN

SQL serverless in Synapse

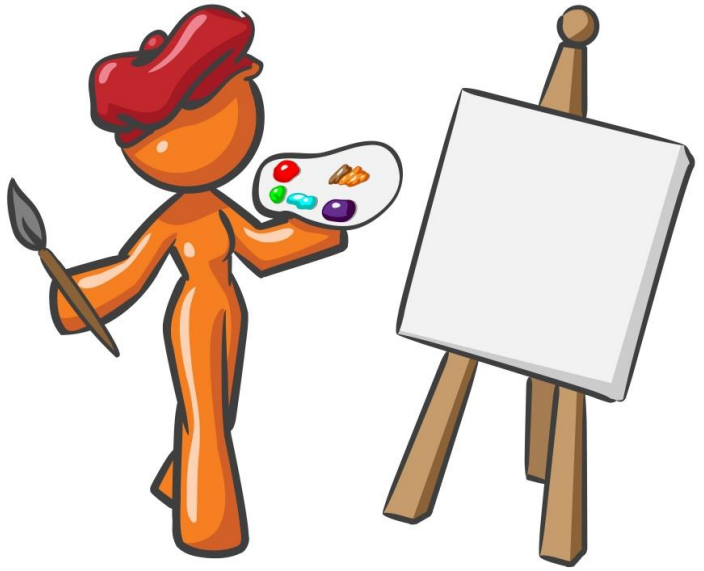
- Allows you to query files in your Azure storage accounts
- No local storage, only metadata objects are stored in databases
- Analyze your Big Data in seconds to minutes
- Supported T-SQL
  - Full SELECT surface area is supported, including a majority of SQL functions
  - CETAS - CREATE EXTERNAL TABLE AS SELECT
  - DDL statements related to views and security only
- There is no charge for resources reserved, you are only being charged for the data scanned by queries you run, hence this model is a true pay-per-use model. ($5 per TB of data processed)

SQL serverless scenarios

- **Basic discovery and exploration** - Quickly reason about the data in various formats (Parquet, CSV, JSON) in your data lake, so you can plan how to extract insights from it.
- **Logical data warehouse** – Provide a relational abstraction on top of raw or disparate data without relocating and transforming data, allowing always up-to-date view of your data.
- **Data transformation** - Simple, scalable, and performant way to transform data in the lake using T-SQL, so it can be fed to BI and other tools, or loaded into a relational data store (Synapse SQL databases, Azure SQL Database, etc.).

Demo

1. Create a Database
2. Create a data source
3. Query CSV files
4. Create an external table (CETAS)



Quickstart: https://docs.microsoft.com/en-us/azure/synapse-analytics/quickstart-sql-on-demand
and https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/create-external-table-as-select

Cost management

- $5 per TB of data processed
  - Data read while reading data and metadata.
  - Amount of data in intermediate results
  - Amount of data written to storage
- Each query has a minimum of 10 MB of data processed
- Cost control
  - Setup limits – Daily, weekly and monthly
  - Query sys.dm_external_data_processed

```
Started executing query at Line 60

(Statement ID: {584E51B0-4653-40DA-9220-B8EF535CF2B1} |
Query hash: 0xF436DC5C144EC80 | Distributed request ID:
{134C847F-4B23-43C2-B0EA-BE836C135B10}. Total size of data
scanned is 10823 megabytes, total size of data moved is 1
megabytes, total size of data written is 0 megabytes. )
Total execution time: 00:00:28.969
```

Request ID	Request content	Submit time	Duration	Data processed
42362	SELECT TOP 100 * FROM OPENF	12/10/20, 8:17:06 AM	18s	315 MiB
50189	*** Global stats query ***	12/10/20, 8:19:22 AM	1m 9s	3.95 GiB
53268	SELECT title = JSON_VALUE(doc	12/10/20, 8:20:15 AM	0s	1 MiB
54097	SELECT DateStruct, TimeStruct, T	12/10/20, 8:20:39 AM	0s	1 MiB
54180	SELECT CAST([tpepPickupDateT	12/10/20, 8:19:22 AM	1m 38s	1.39 GiB

ORANGEMAN