

AZURE SYNAPSE ANALYTICS



Building analytics solutions using Serverless SQL Pools

Sidney Cirqueira

ABOUT ME

SIDNEY CIRQUEIRA - @SIDNEY.CIRQUEIRA

- +10 years of experience with Information Technology
- Azure Data Engineer at CI&T
- Big Data & Machine Learning Student and Enthusiast
- Microsoft Learn Student Ambassador
- Speaker in the MS Technical Community



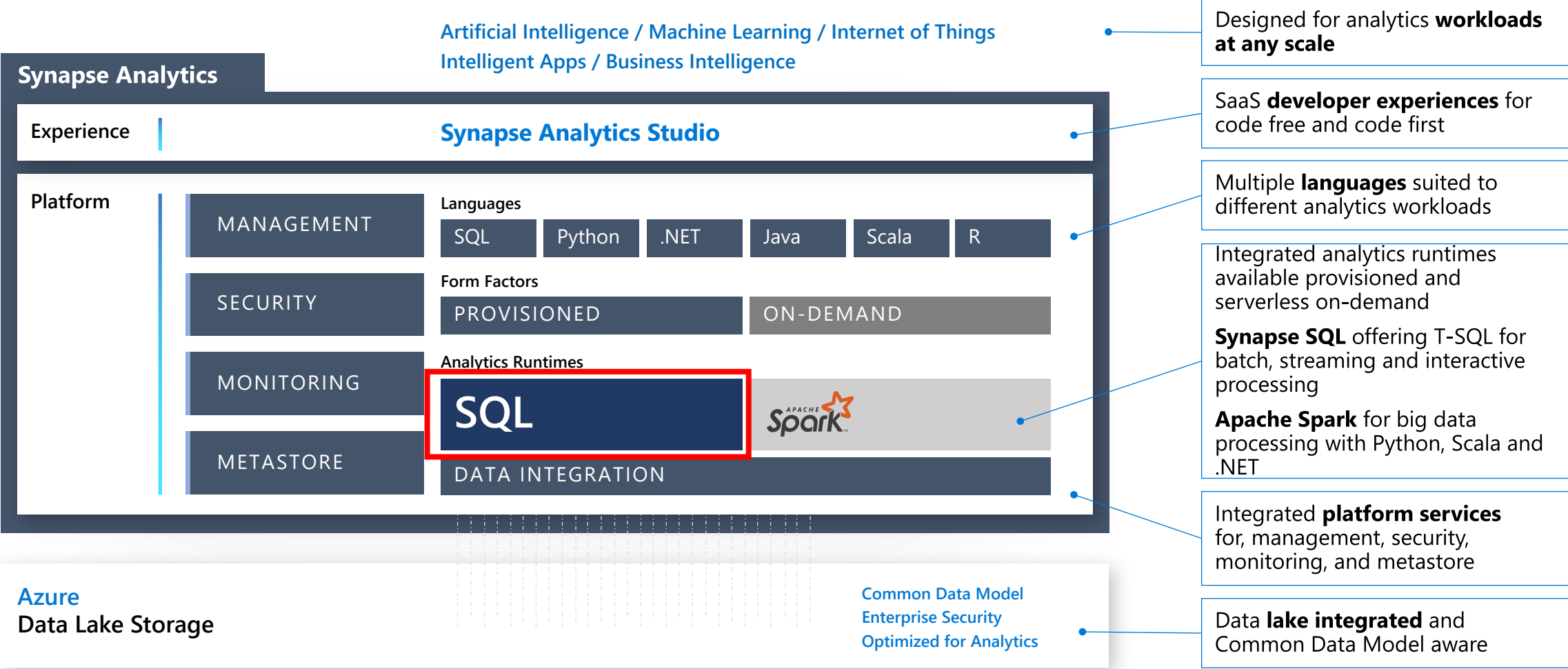
AGENDA

- Overview Azure Synapse Analytics
- SQL Pools – Key features
- Developer & Clients Tools
- serverless SQL Pool – Overview & Benefits
- Use Cases
- serverless SQL Pool - Architecture
- serverless SQL Pool - Key features
- Best Practices
- Demo
- Q&A



Azure Synapse Analytics

Limitless analytics service with unmatched time to insight



(gestãocomdados)

ELGIN



Key features – SQL Pools

Rich surface area

- T-SQL language for data analytics
- Supporting large number of languages and tools
- Enterprise-grade security

SQL Provisioned

- Modern Data Warehouse
- Indexing and caching
- Import and query external data
- Workload management

SQL Serverless

- Querying external data
- Model raw files as virtual tables and views
- Easy data transformation



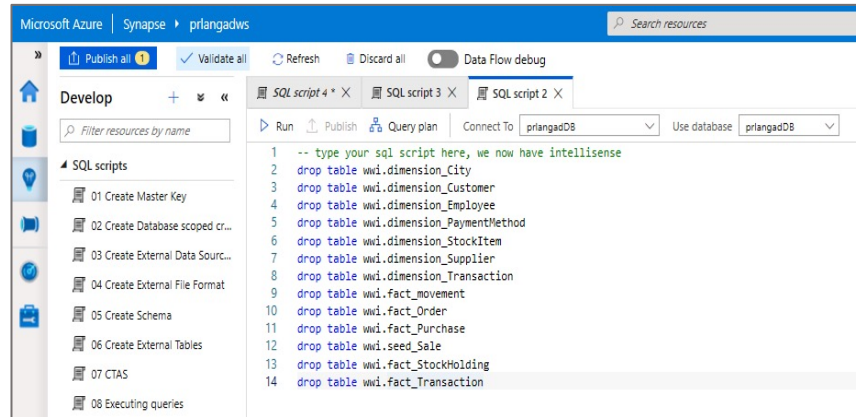
(gestãocomdados)

ELGIN

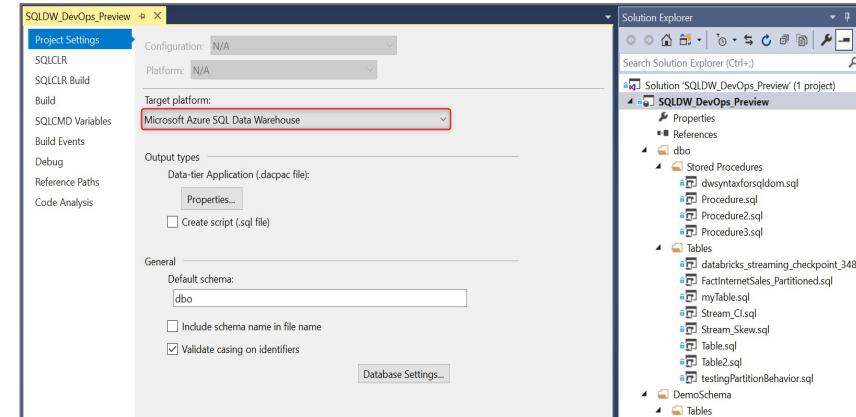


Developer Tools

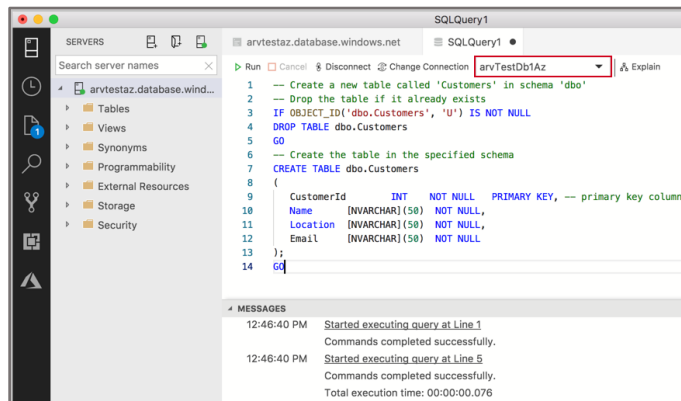
Azure Synapse Analytics



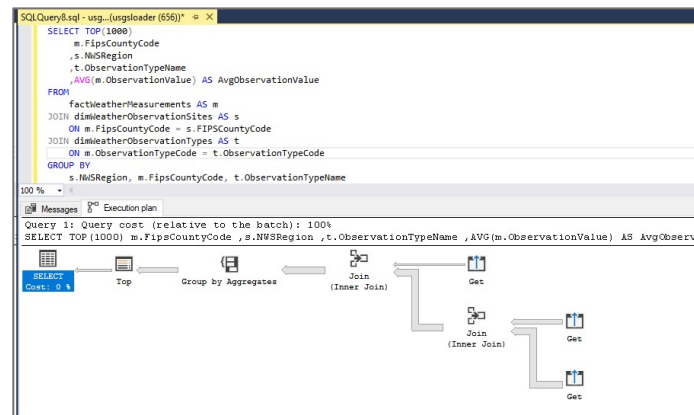
Visual Studio - SSDT database projects



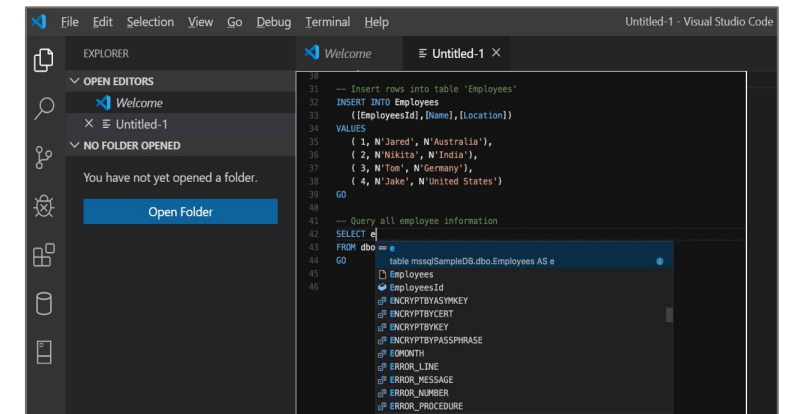
Azure Data Studio (queries, extensions etc.)



SQL Server Management Studio (queries, execution plans etc.)



Visual Studio Code



(gestãocomdados)

ELGIN

ELGIN developers community



Azure **Synapse** Analytics SQL Pool (serverless model)



(gestãocomdados)

ELGIN

ELGIN
developers
community

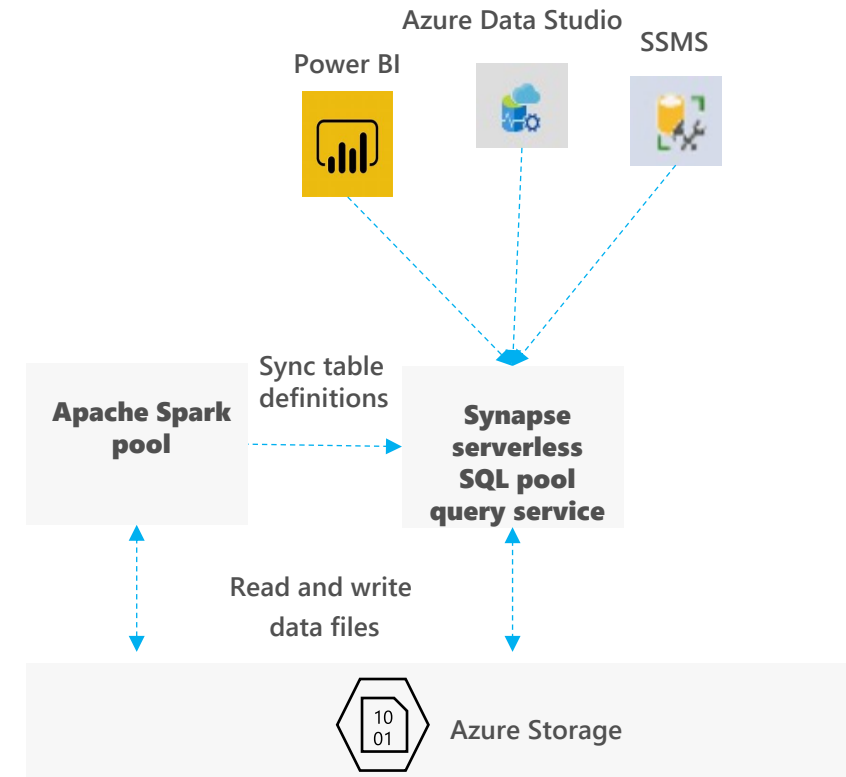
serverless SQL pool

Overview

An interactive query service that enables you to use standard T-SQL queries over files in Azure storage.

Benefits

- Use SQL to work with files on Azure storage
 - Directly query files on Azure storage using T-SQL
 - Logical Data Warehouse on top of Azure storage
 - Easy data transformation of Azure storage files
- Supports any tool or library that uses T-SQL to query data
- Automatically synchronize tables from Spark
- Serverless
 - No infrastructure, no upfront cost, no resource reservation
 - Pay only for query execution (per data processed)



Use Cases

Quick data exploration

- Easily explore schema and data in files on Azure storage
- Supports various file formats (Parquet, CSV, JSON)
- Direct connector to Azure storage for large BI ecosystem

Logical Data Warehouse

- Model raw files as virtual tables and views
- Use any tool that works with SQL to analyze files
- Use enterprise-grade security model

Easy data transformation

- Transform CSV to parquet format
- Move data between containers and accounts
- Save the results of queries on external storage



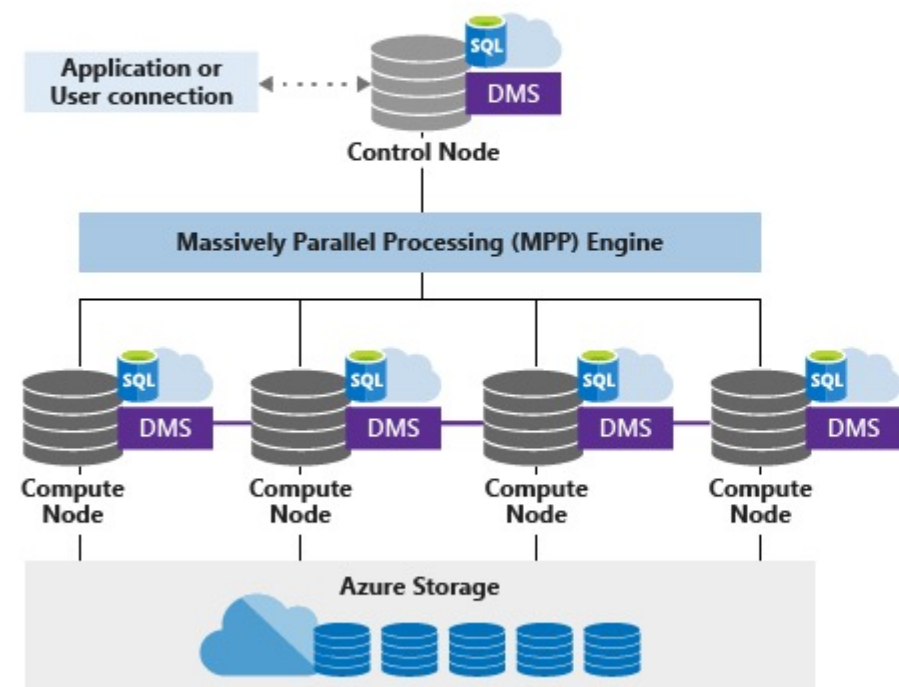
(gestãocomdados)

ELGIN

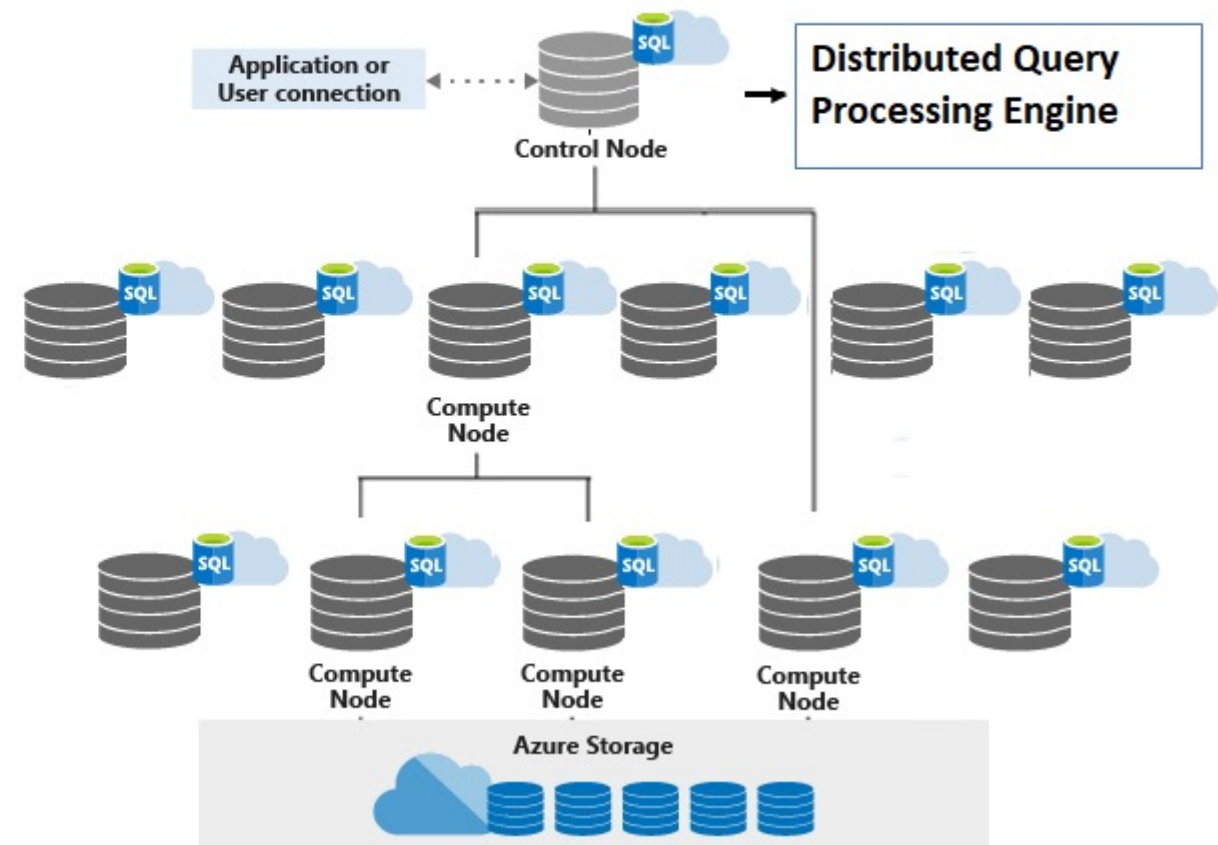
ELGIN
developers
community

SQL pools – Architecture

SQL pool - dedicated



SQL pool - serverless



Key Features - serverless

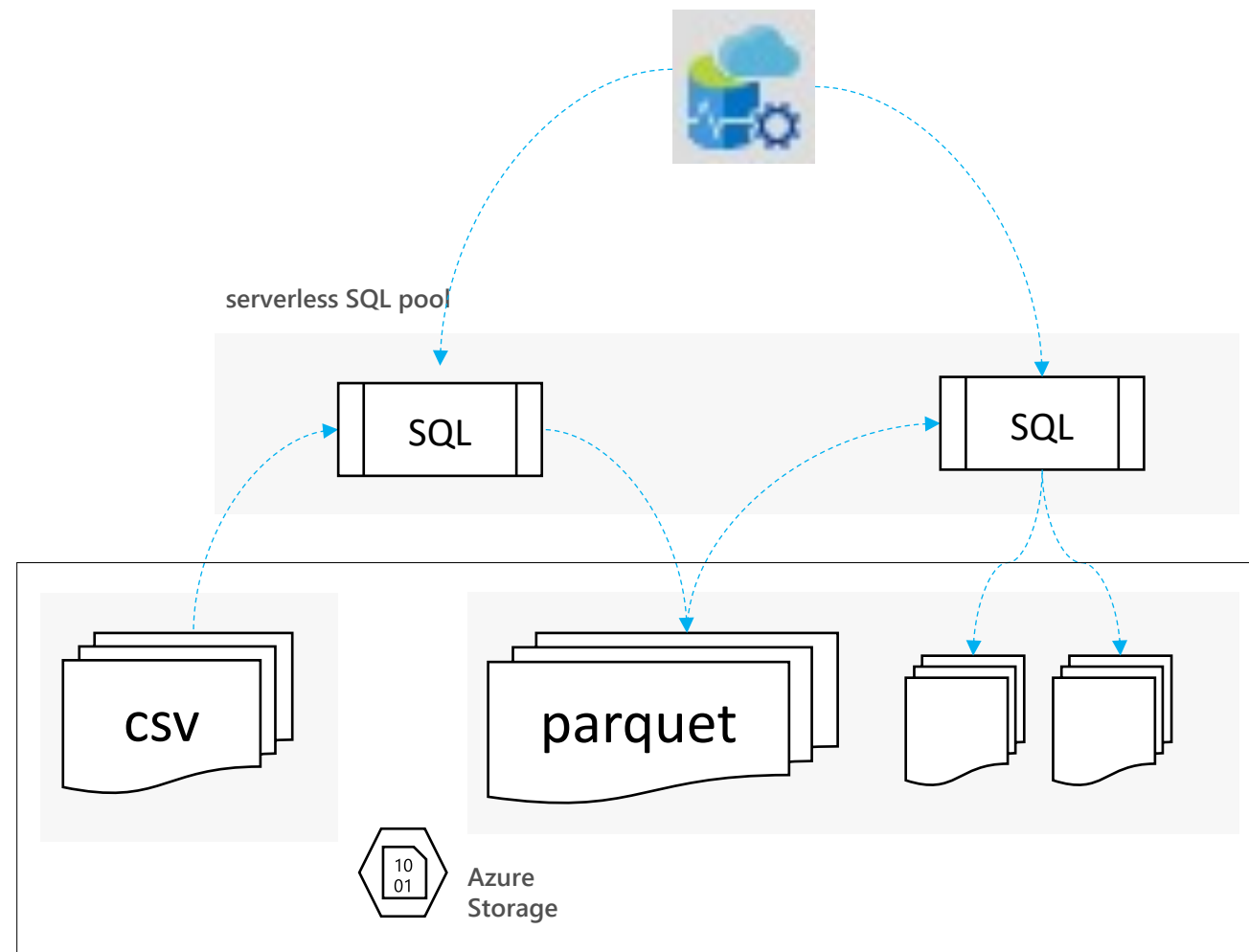
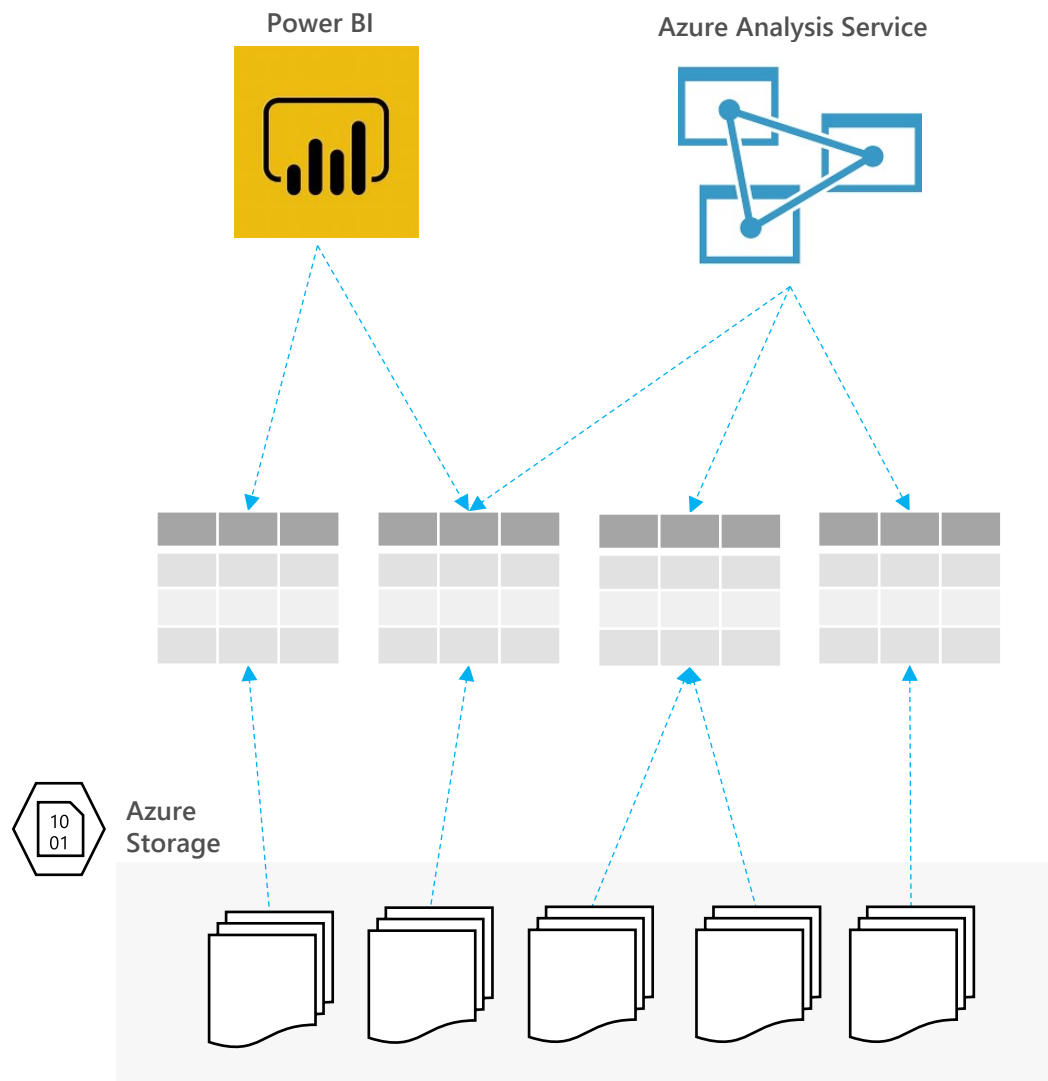
- Easily explore files on storage
- Easily query files in various formats
- Automatic schema inference
- Defined the query result schema inline
- Customize the content parsing to fit your case
- Easily query multiple files, with wildcards
- Query partitioned data, using the folder structure
- SQL serverless as a logical data warehouse
- Logical data warehouse views
- Logical data warehouse tables
- Easy data transformation with CETAS
- Automatic syncing of Spark tables
- Automatic syncing with Synapse Link



(gestão com dados)

ELGIN





Best practices

- Co-locate storage and serverless SQL pools
- Consider Azure Storage throttling
- Prepare files for querying (CSV, JSON -> Parquet)
- Push wildcards to lower levels in the path
- Use appropriate data types and check inferred data types
- Use filename and filepath functions to target specific partitions
- Use PARSER_VERSION 2.0 to query CSV files
- Use CETAS to enhance query performance and joins
- Choose SAS credentials over Azure AD pass-through (for now)



(gestãocomdados)

ELGIN

ELGIN
developers
community

Demo

Microsoft Azure | Synapse Analytics | internalsandboxwe5

Search resources

prlangad@microsoft.com

»

Publish all

Validate all

Refresh

Discard all

Home

Data

Filter resources by name

Storage accounts 1

internalsandboxwe (Prima...

Databases 3

Datasets 5

opendataset

New SQL script

New notebook

Upload

Download

New Folder

Select

opendataset > holidays

NAME

LAST MODIFIED

_SUCCESS

3/5/2020, 3:38:23 AM

part-00000-bd1aba93-a85a-4909-8bf4-f79afb6c946f-c000.snappy.parquet

3/5/2020, 3:38:11 AM

part-00001-bd1aba93-a85a-4909-8bf4-f79afb6c946f-c000.snappy.parquet

3/5/2020, 3:38:12 AM

part-00002-bd1aba93-a85a-4909-8bf4-f79afb6c946f-c000.snappy.parquet

3/5/2020, 3:38:23 AM

part-00003-bd1aba93-a85a-4909-8bf4-f79afb6c946f-c000.snappy.parquet

3/5/2020, 3:38:08 AM

New SQL script - Select TOP 100 rows

New SQL script - Create external table

New notebook

Copy ABFS path

Manage Access...

Rename...

Download

Delete

Properties...

Microsoft Azure | Synapse Analytics | internalsandboxwe5

Search resources

prlangad@microsoft.com

»

Publish all

Validate all

Refresh

Discard all

Home

Data

Filter resources by name

Storage accounts 1

internalsandboxwe (Prima...

Databases 3

Datasets 5

opendataset

SQL script 1

Run

Publish

Query plan

Connect to

SQL on-demand

```
1 SELECT
2   TOP 100 *
3 FROM
4   OPENROWSET(
5     BULK 'https://internalsandboxwe.dfs.core.windows.net/opendataset/holidays/part-00001-bd1aba93-a85a-4909-8bf4-f79afb6c946f-c000.snappy.parquet',
6     FORMAT='PARQUET'
7   ) AS [r];
8
```

Results

Messages

View

Table

Chart

Export results

Search

VENDORID

TPPEPICKUPDATETIME

TPPEDROPOFFDATET...

PASSENGERCOUNT

TRIPDISTANCE

PULOCATIONID

DOLOCATIONID

VT...

2009-05-07T23:1...

2009-05-07T23:2...

1

2.94

NULL

NULL

VT...

2009-05-07T16:3...

2009-05-07T16:3...

5

0.73

NULL

NULL

VT...

2009-05-08T14:5...

2009-05-08T15:0...

3

0.55

NULL

NULL

VT...

2009-05-07T15:5...

2009-05-07T16:1...

1

2.5

NULL

NULL

00:00:31 Query executed successfully.



Q&A



(gestão com dados)

ELGIN



References

[Serverless SQL pool - Azure Synapse Analytics | Microsoft Docs](#)

[Serverless Architecture and Concepts. What is it? - Microsoft Tech Community](#)

[POLARIS: the distributed SQL engine in azure synapse - Microsoft Research](#)

<https://www.vldb.org/pvldb/vol13/p3204-saborit.pdf>

[Create and use external tables in serverless SQL pool - Azure Synapse Analytics | Microsoft Docs](#)



(gestãocomdados)

ELGIN





(gestãocomdados)