

K.O



SQLBITS

2022



The Dream Team

Synapse Analytics Serverless SQL Pools and Pipelines

SQL Bits 2022

Thursday March 10th 1:40PM

Andy Cutler

K.O



SQLBITS

2022



Please
It h
Th

SQLBits
@SQLBits

GUESS WHAT!! 🌳 #SQLBits wants your help to plant some saplings with the #NationalTrust 🌱 For every 10 feedback forms we receive, we promise to plant one tree! So.. Get that feedback submitted!! 🌱

ession
ontent!

<https://sqlb.it/?6996>

Microsoft
Partner

Silver Data Analytics
Silver Data Platform



Microsoft®
Most Valuable
Professional

Andy Cutler

Independent BI/DW Consultant

Azure Data Platform & Power BI



datahai.co.uk/blog

serverlesssql.com

twitter.com/MrAndyCutler

linkedin.com/in/andycutler/



SCAN ME



- **Azure Synapse Analytics Overview**
- **Serverless SQL Pools**
- **The Data Lake and Serverless SQL Pools**
- **Pipelines for Data Movement & Orchestration**
- **Real-time and Batch solution for web store analysis**
- **Demo**

Synapse Analytics



Azure Cloud Analytics service

Data Warehousing

Big Data Analytics

Data Integration

Analytics

The Pools

Dedicated SQL

Import data into a large-scale database

Serverless SQL

Query data in Azure Storage, Cosmos DB, and Dataverse

Spark

Run Data Engineering and Machine Learning workloads

Data Explorer

Analyse log and telemetry data

Data Integration

Pipelines

Orchestrate data movement and transformation

Analytics

Power BI

Create Datasets and Reports within Synapse Studio

Development

Synapse Studio

Web Studio for developing SQL, Notebooks, Pipelines, and Power BI datasets/reports

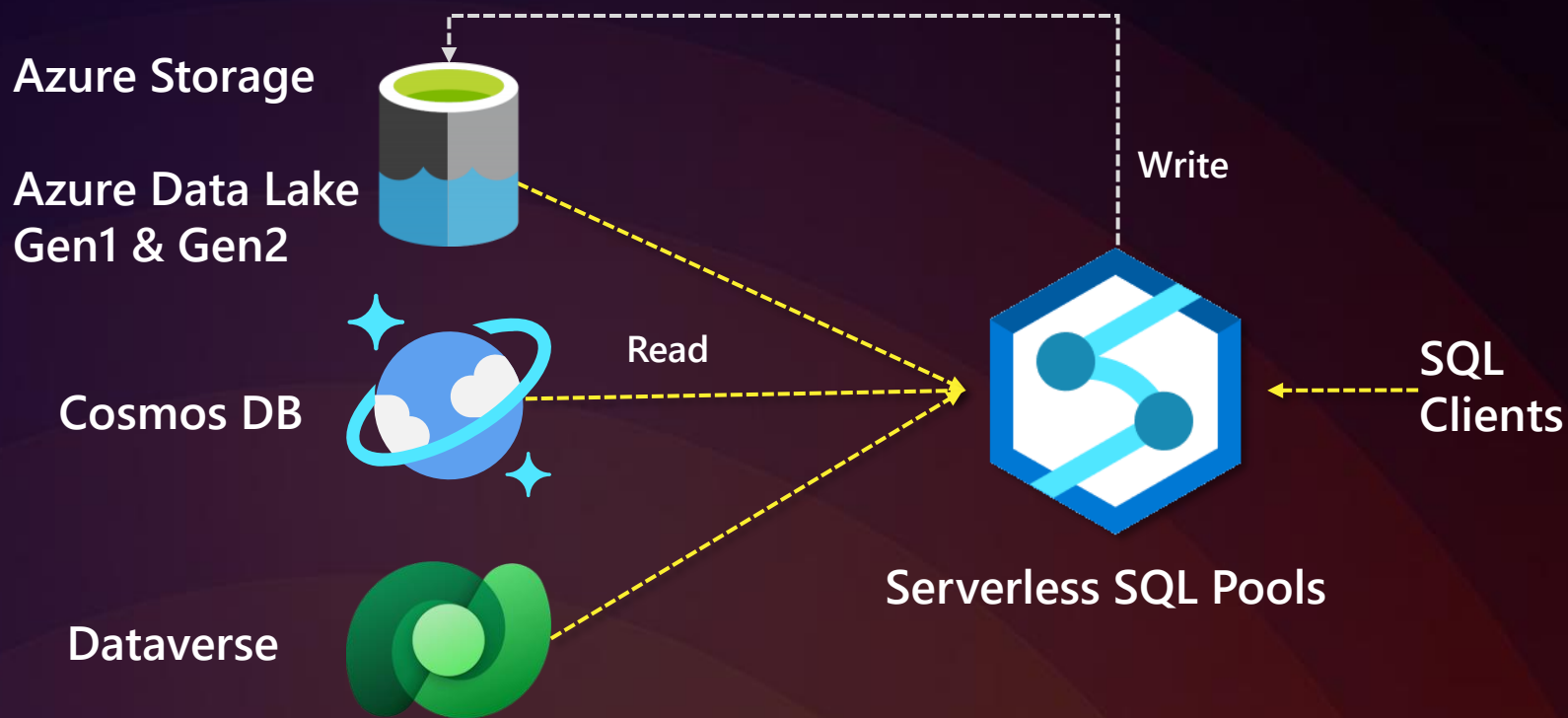
Serverless SQL Pools



Query external data in Azure Storage, Cosmos DB and Dataverse

Familiar SQL objects

- Databases
- Stored Procedures
- DMVs
- Views
- External Tables



Serverless SQL Pools cost is based on the amount of data processed and not compute/time to execute

~\$5 per 1TB Data Processed (Write/Read)

No data is stored within Serverless SQL Pools

Serverless SQL Pools Scenarios

...and the role of the Data Lake



Microsoft state 3 scenarios

Data Exploration

Adhoc analysis of external data

Logical Data Warehouse

Create a structure over external data without heavy-lifting moving data

Data Transformation

Use familiar T-SQL to transform data for downstream clients E.G Power BI

When working with Azure Storage and Serverless SQL Pools, we can connect using the full URL or by creating Data Sources in a Serverless SQL Pools database



Azure Storage



Serverless SQL Pools

←----- SQL Clients

Serverless SQL Pools can connect to Azure Storage using:

- Shared Access Signature
- Service Principal
- Managed Identity
- User Identity

Active Directory users can be assigned to groups, those groups can be assigned to database and object roles

SQL authentication is also supported

Pipelines



We can use Synapse Pipelines to copy and transform data into a format better suited for Serverless SQL Pools querying

We can orchestrate the data transformation process and also trigger the creation of Views within Serverless SQL Pools

Copy and Transform Data

A Microsoft best practice for Serverless SQL Pools is to convert CSV and JSON files to Parquet. Parquet is a columnar file format which compresses data and supports filtering

- ⚙️ Reduces query processing time
- ⚙️ Reduces data processed amount (which reduces cost!)

Orchestrate Serverless SQL View Creation

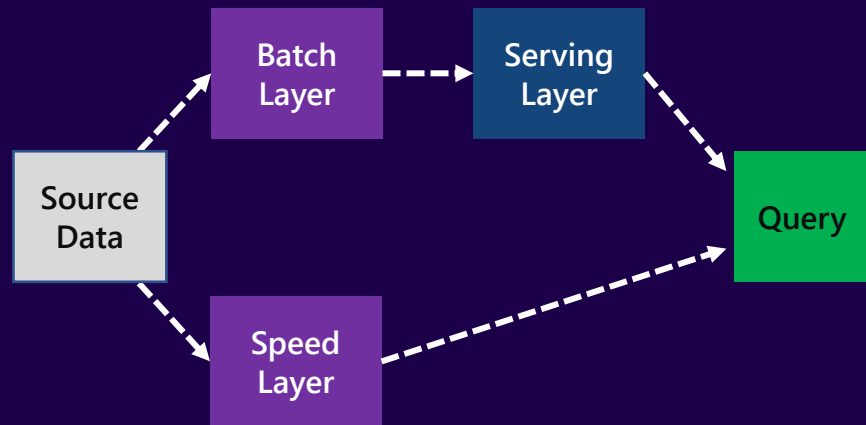
As we load data into Azure Storage, we can pass in relevant folder names to a stored procedure (or script) to dynamically create SQL Views in a Serverless SQL Pools database

- ⚙️ Reduces effort creating structure over data in Azure Storage

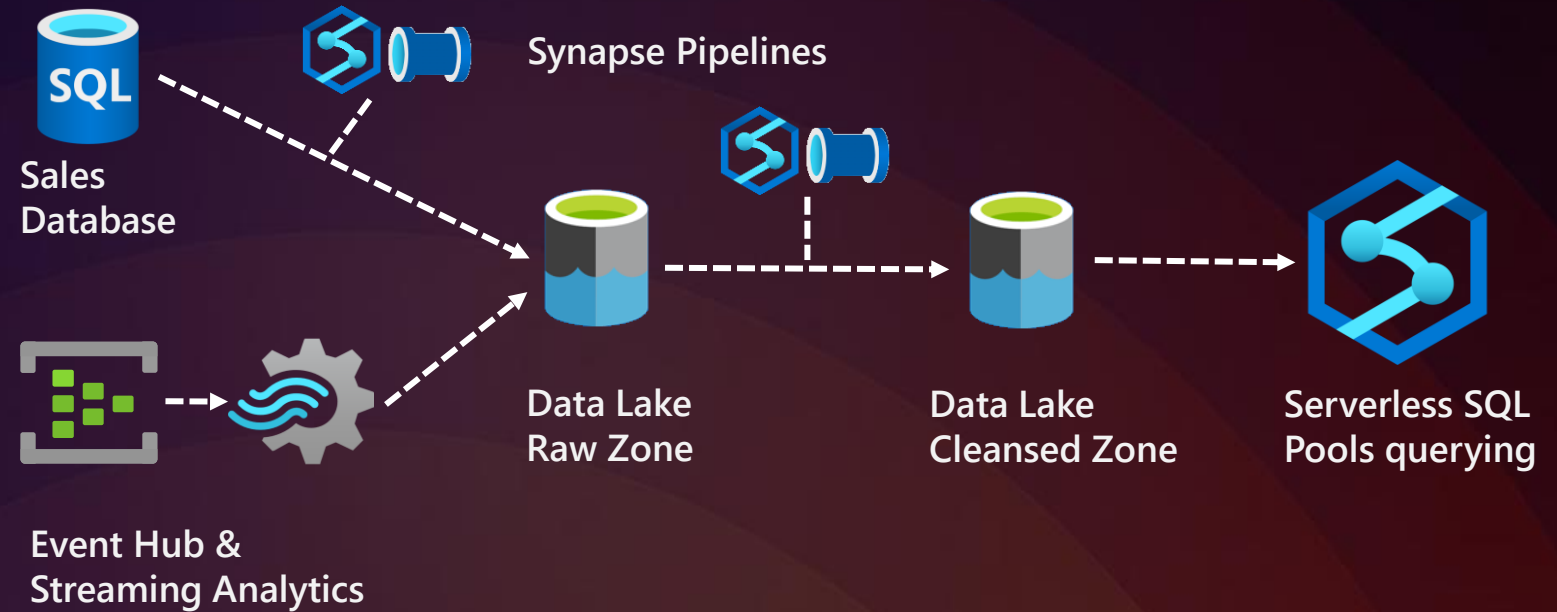
Solution



This solution allows for real-time querying and batch processing for richer and higher accuracy datasets



Batch & Speed Layers delivered by Pipelines and Serverless SQL Pools



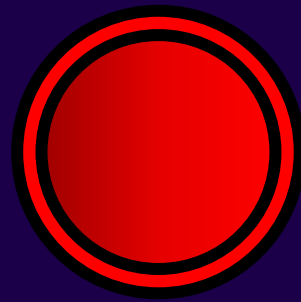
Speed Layer

- We can query CSV files as new data is being appended
- We can query more recent files by using Serverless SQL Pools **filepath** function to reduce data processed and increase query performance

Batch Layer

- We can join data together with different load frequencies
- We can join the Product & Sales data together with the Web Telemetry data

Press Start button for demo!



K.O



SQLBITS

2022



Please

It h

Th



ession

content!

<https://sqlb.it/?6996>

Game Over

Thank you for playing!



SCAN ME