When Two Become One

Fabric SQL Data Warehouse

#DataWeekender 6.5

Andy Cutler

Andy Cutler

Independent BI/DW Consultant

Azure Data Platform & Power BI

datahai.co.uk

serverlesssql.com

twitter.com/MrAndyCutler

linkedin.com/in/andycutler/

datatoboggan.co.uk

SCAN ME

Agenda

Exploring the Timeline

Warehouses and Lakehouses in Fabric

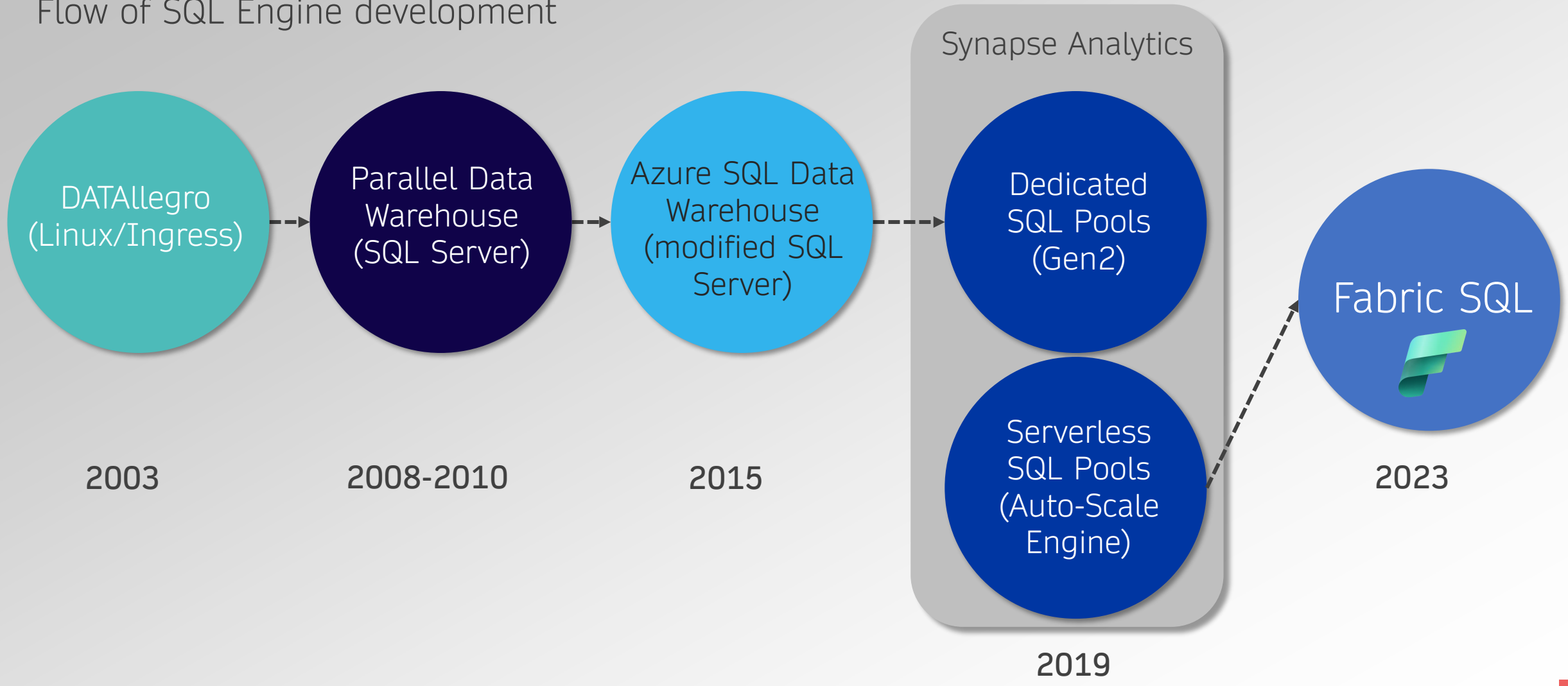
Demo

Timeline

...a 20 year journey

SQL Engine Timeline

Flow of SQL Engine development



Stu Frost

“It was clear that customers were frustrated by the cost and performance of incumbent systems – mainly Teradata”

“I stayed at Microsoft for 2 years and I was involved with PDW for the first year or so”

“Hmm, so many things! One might be to move it all to the cloud and pre-empt Snowflake!”

What is Fabric?

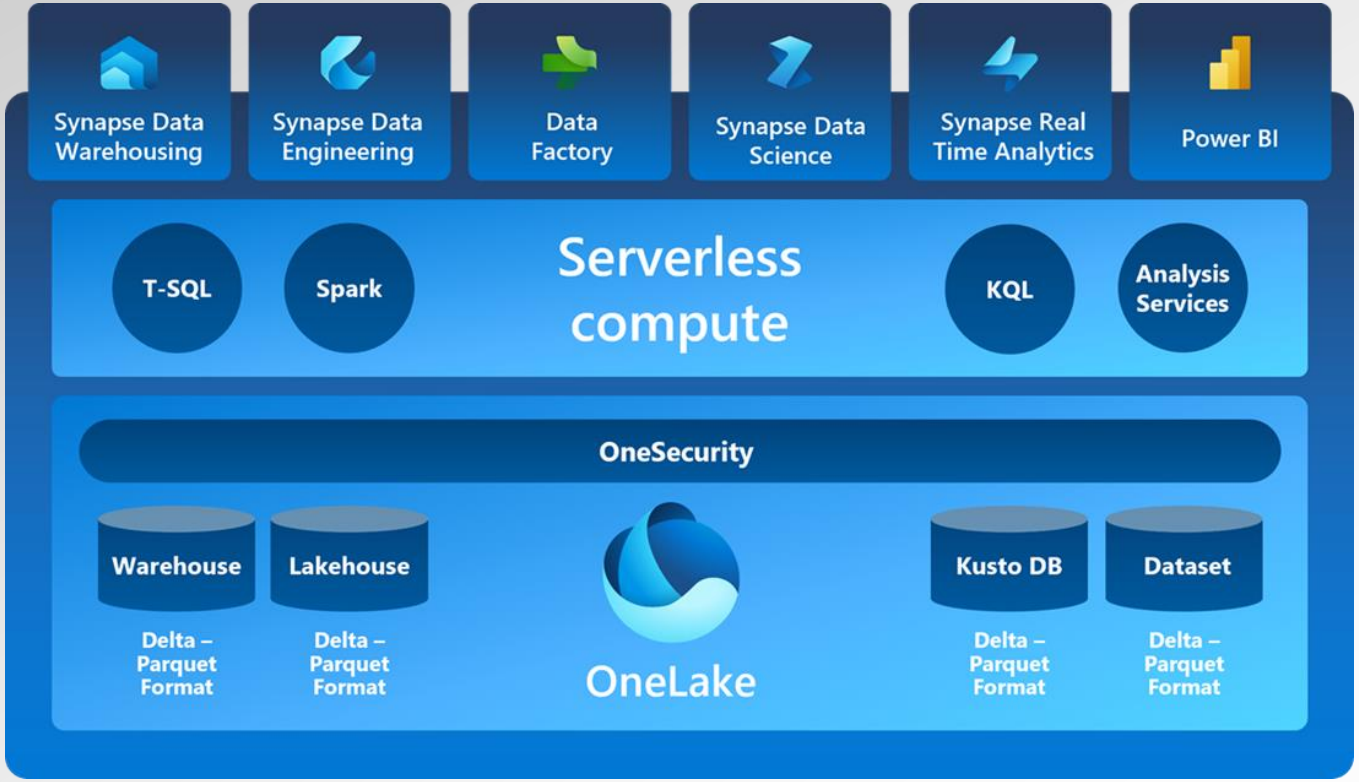
Adds “significant” functionality to a Power BI tenant

You are not purchasing a “server”, but rather compute in the form of CAPACITY UNITS (CU)

- E.G. Power BI Premium P1 = F64

Compute control abstracted away (Spark is configurable…)

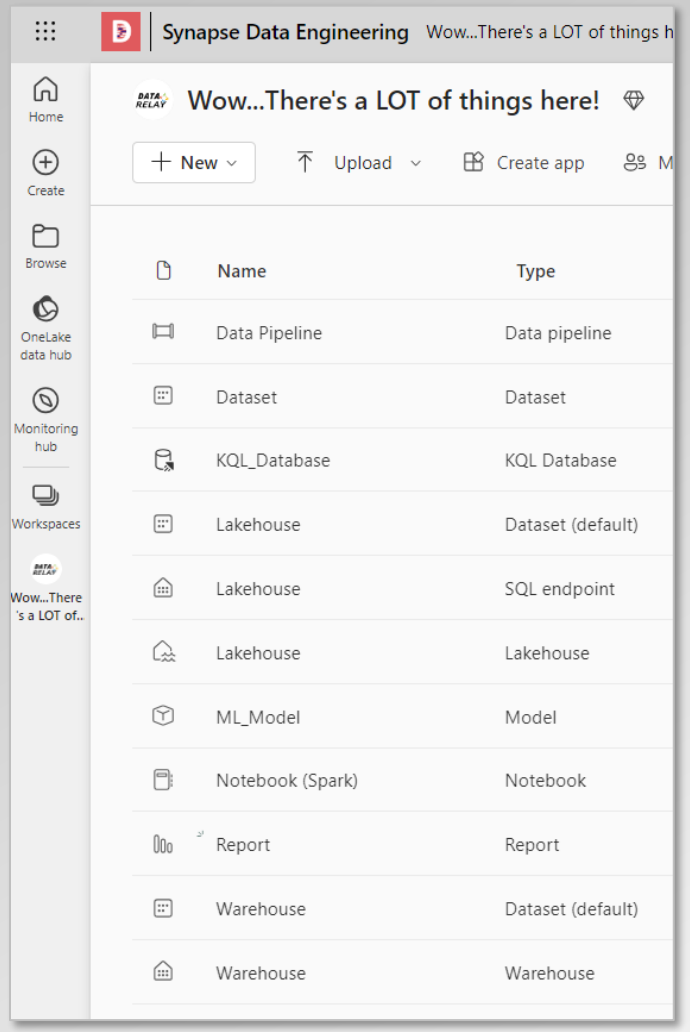
Storage underpinned by open-source Delta Lake format



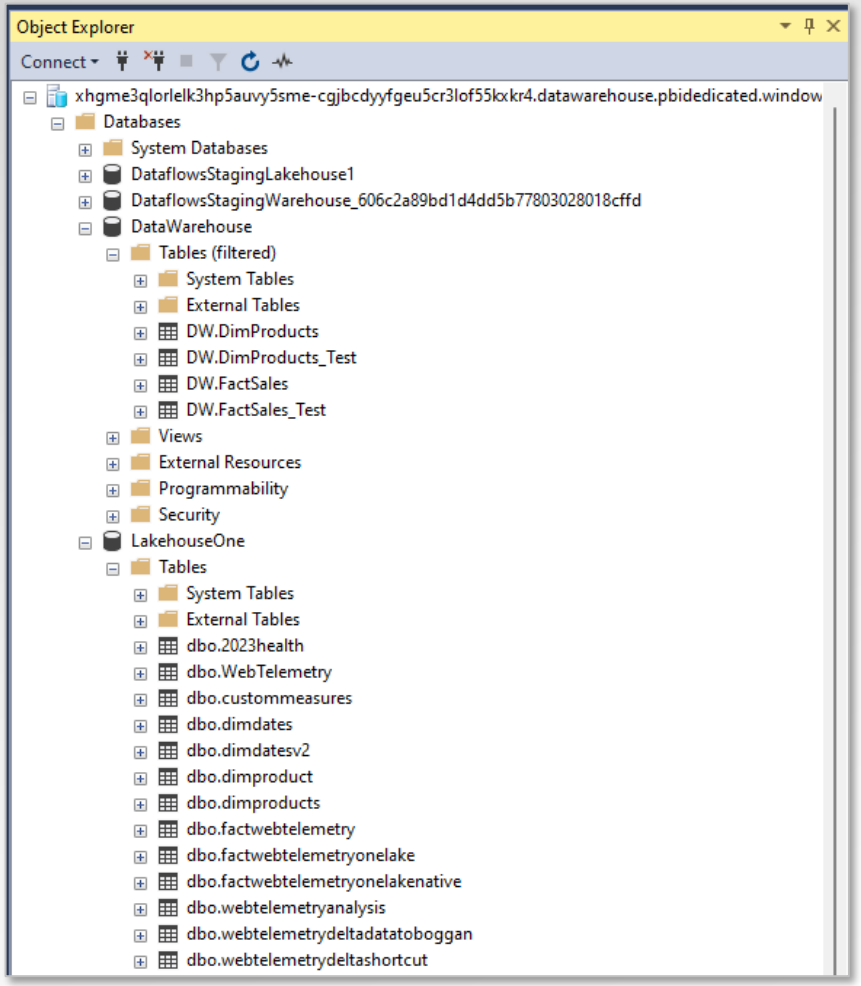
https://github.com/microsoft/Fabric-Readiness

How SQL tools see Workspaces

Workspace in the GUI



Workspace in SQL tools (E.G. SSMS)



Lakehouse

Persona

Workspace

Structured, Unstructured, Semi-Structured data across all ranges of file formats

Data Engineer (Spark, Scala), Data Scientist (Python)

Bronze, Silver, and Gold Layer

New DirectLake connection from Power BI Datasets

Warehouse

Persona

Workspace

Structured data in Databases and Tables

Data Warehouse Developer (SQL)

Can be used in the Gold layer

DirectQuery and Import modes only (currently)

What Powers the Warehouse?

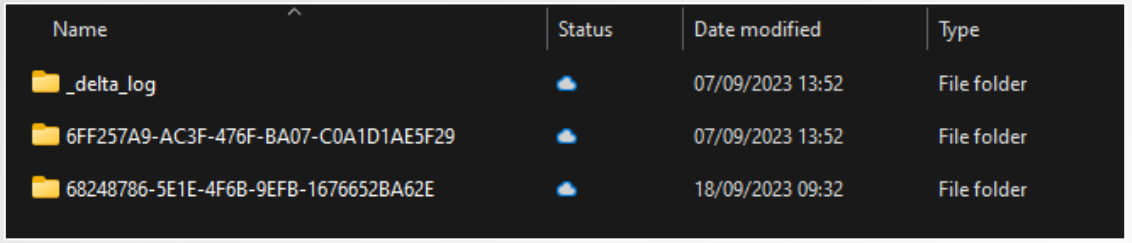
An auto-scale out, fault tolerant, distributed compute engine

Enhanced version of Serverless SQL Pools engine

Native storage uses Delta Lake

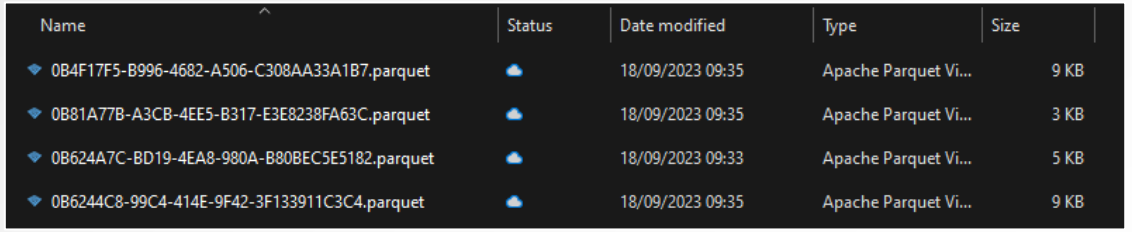
Transaction Log

Metadata



Parquet file format

Compressed “columnar”



Loading Options

Code

SQL COPY INTO... <from Azure storage>

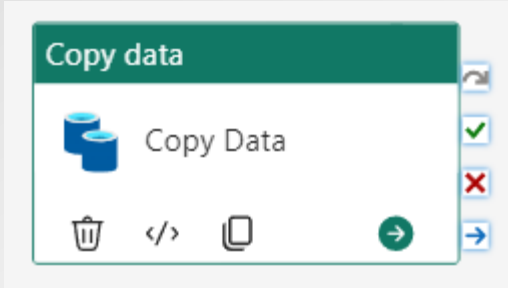
SQL CREATE TABLE AS...SELECT

SQL INSERT INTO...VALUES / SELECT

Pyodbc (from Notebooks)

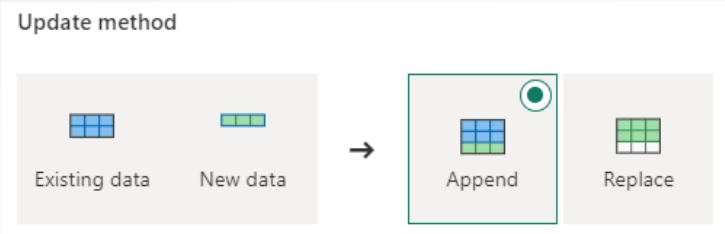
Low/No-Code

Pipelines



Dataflows Gen2

(Power Query)



Demo

Thank You

References

- Azure SQL Data Warehouse is now Azure Synapse Analytics | Azure Blog | Microsoft Azure
- Parallel Data Warehouse (PDW) benefits made simple | James Serra's Blog
- Parallel Data Warehousing (PDW) Explained | James Serra's Blog
- Microsoft Closes Acquisition of DATAllegro – Stories
- DATAllegro Placed in Visionaries Quadrant for Data (globenewswire.com)
- Data warehouse appliance – Wikipedia
- Microsoft Closes Acquisition of DATAllegro – Stories
- Ted Kummert: Microsoft Business Intelligence Conference – Stories
- Microsoft Business Intelligence Conference 2008 – Stories
- Fabric decision guide - choose a data store - Microsoft Fabric | Microsoft Learn
- Data Warehousing in Microsoft Fabric - Microsoft Fabric | Microsoft Learn
- T-SQL surface area - Microsoft Fabric | Microsoft Learn

History

2008-2010

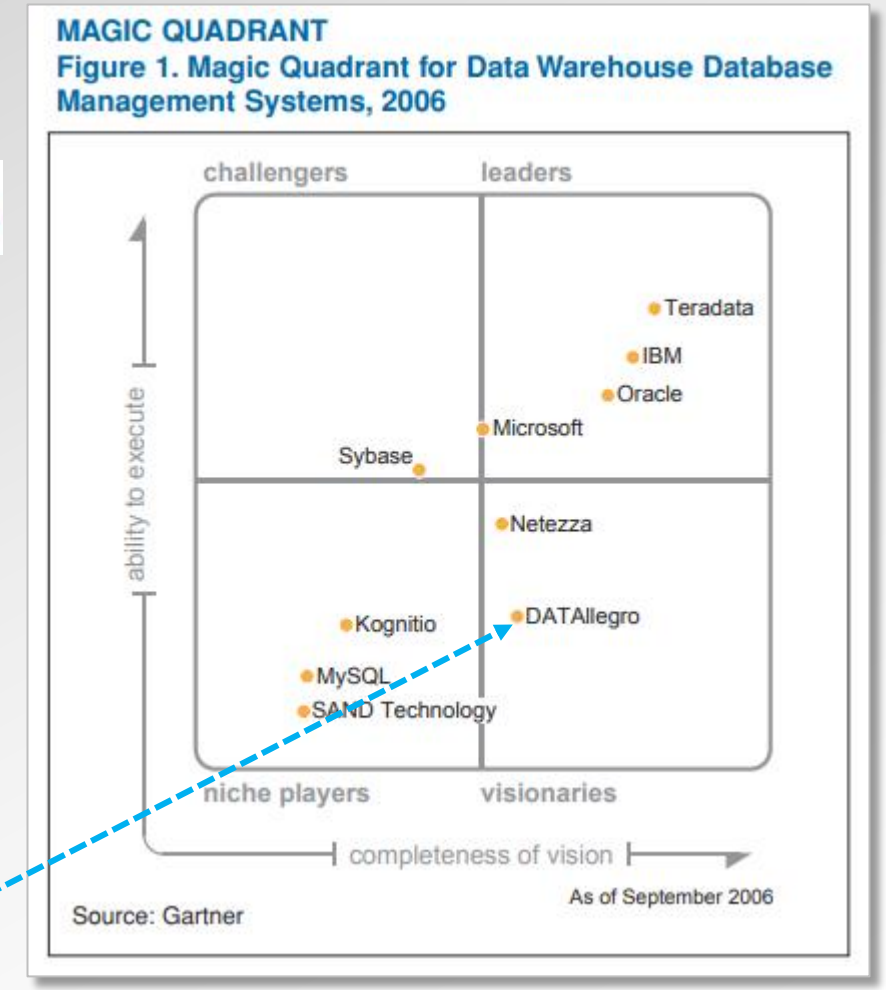
All started with **DATAllegro** (bought in 2008)

Evolved into **SQL Server 2008 Parallel Data Warehouse**

Installed on **commodity hardware** (Dell or HP hardware)

PDW was created to compete with other **multi-terabyte** data warehouse products like Oracle Exadata, Teradata.

DATAllegro were visionaries!



History

2015

Azure version of Parallel Data Warehouse with scalable Compute and Storage

Concept of Data Warehouse Units (DWU 100, 200...)

- Gen1 - 24GB per DWU100
- Gen2 – 60GB per DWU100

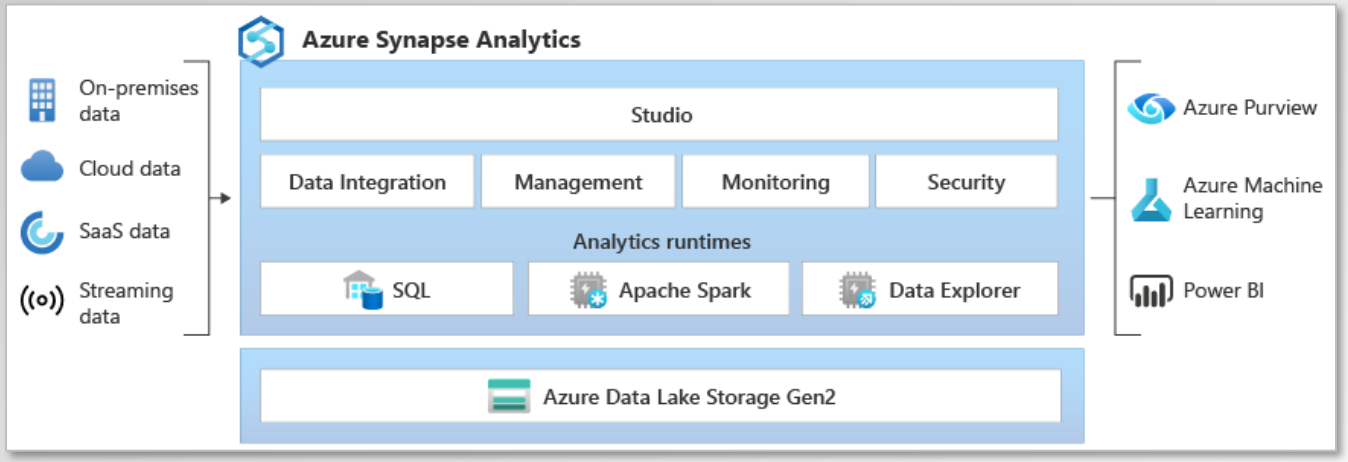
Easier to purchase and use!

Platform-as-a-Service (PaaS)

History

2019

Convergence of services



Platform-as-a-Service (PaaS)

Renamed Azure SQL Data Warehouse to ...**Dedicated SQL Pools**

Introduced a new SQL engine called .....**Serverless SQL Pools** (inspired by Polaris paper)

Added Spark, Data Explorer, Data Factory, Power BI integration (sort of...)

History

2023

Another convergence of services?!

A unified “Serverless” SQL Engine for both Lakehouse and Warehouse

Lakehouse SQL Endpoint
- SQL Read-Only over Tables in Lakehouse

Warehouse
- SQL Read/Write over Tables in Warehouse

Synapse name stays! (for now…)

Software-as-a-Service (SaaS)