

# What is Azure Synapse Analytics?

- **Azure Synapse** is an enterprise analytics service that accelerates time to insight across data warehouses and big data systems.
- Microsoft SaaS offering



# What is Azure Synapse Analytics?



SQL technologies used in enterprise data warehousing



Spark technologies used for big data



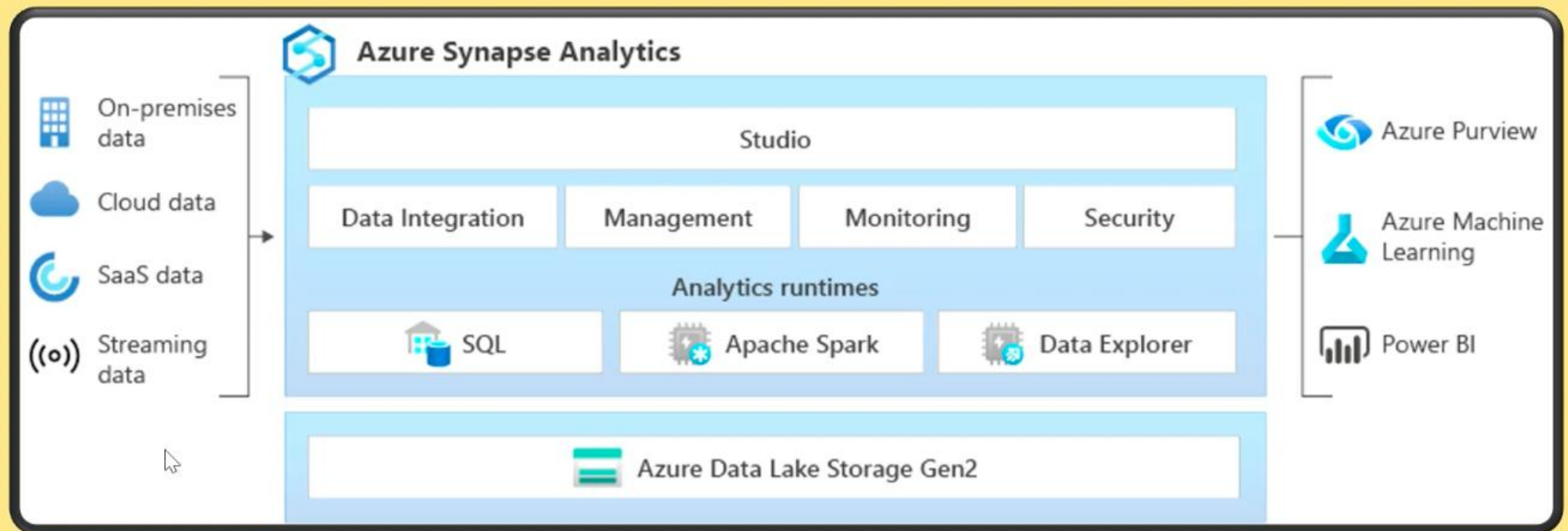
Data Explorer for log and time series analytics



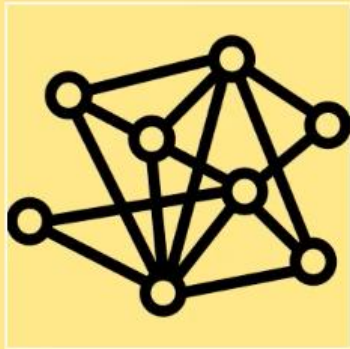
Pipelines for data integration and ETL/ELT



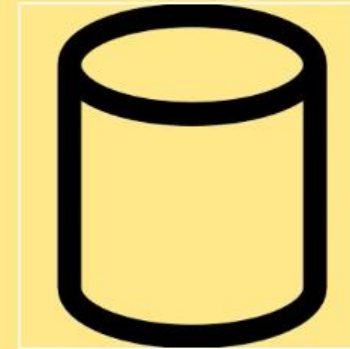
Deep integration with other Azure services such as **Power BI**, **CosmosDB**, and **AzureML**.



# Synapse workspace



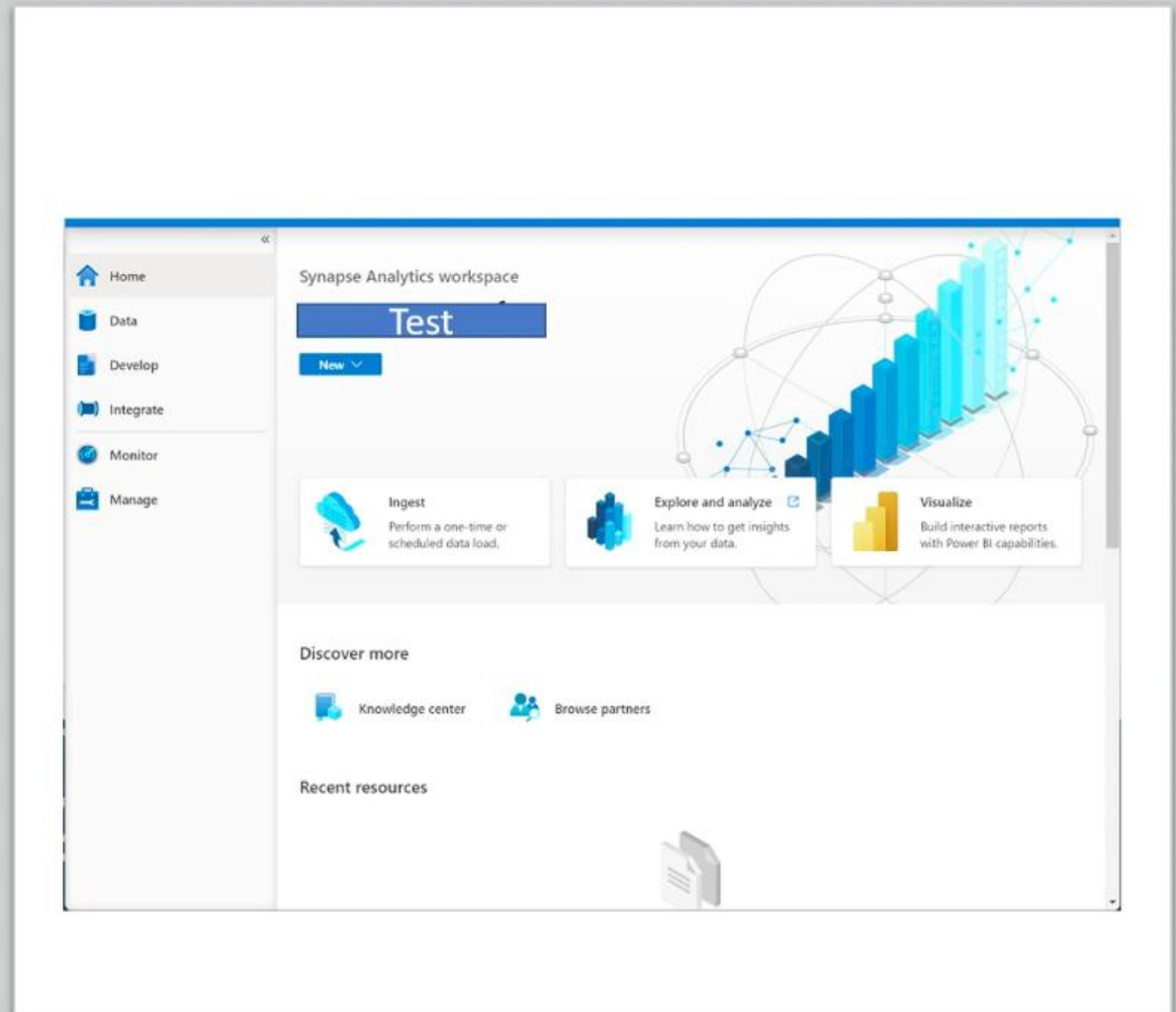
A Synapse Analytics workspace defines an instance of the Synapse Analytics



A workspace typically has a default data lake, which is implemented as a linked service to an Azure Data Lake Storage Gen2 container.

# Synapse Studio

A web-based portal for  
Azure Synapse Analytics.



# Pipeline

- Extracting data from multiple sources and collected in a central data storage
- Synapse includes built-in support for creating, running, and managing pipelines that orchestrate the activities necessary to retrieve data from a range of sources, transform the data as required, and load the resulting transformed data into an analytical store.

# SQL Querying

Azure Synapse Analytics supports SQL-based data querying and manipulation through two kinds of SQL *pool*

- Serverless
  - Dedicated SQL pool
- 
- The Azure Synapse SQL system uses a distributed query processing model to parallelize SQL operations



# Spark

Create one or more Spark pools and use interactive *notebooks* to combine code and notes as you build solutions for data analytics, machine learning, and data visualization.





# Data Explorer

- Data Explorer is a data processing engine in Azure Synapse Analytics.
- Data Explorer uses an intuitive query syntax named Kusto Query Language (KQL) to enable high performance, low-latency analysis of batch and streaming data.

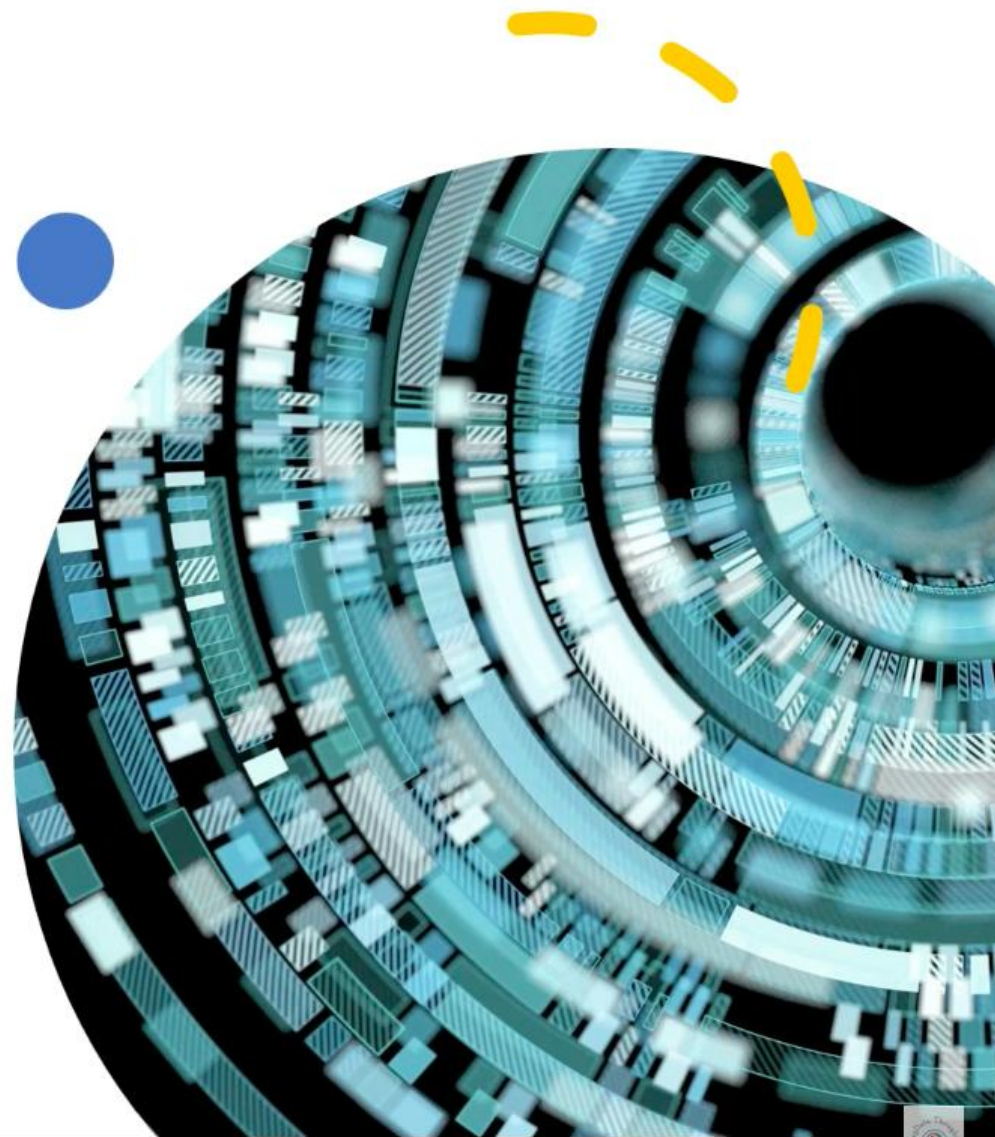
# Integration

- Azure Synapse Link
- Microsoft Power BI
- Microsoft Purview
- Azure Machine Learning integration

# What is Synapse SQL?

Azure Synapse Analytics supports SQL-based data querying and manipulation through two kinds of SQL *pool*

- Serverless
- Dedicated SQL pool
- The Azure Synapse SQL system uses a distributed query processing model to parallelize SQL operations



# Synapse SQL



SCALE OUT ARCHITECTURE



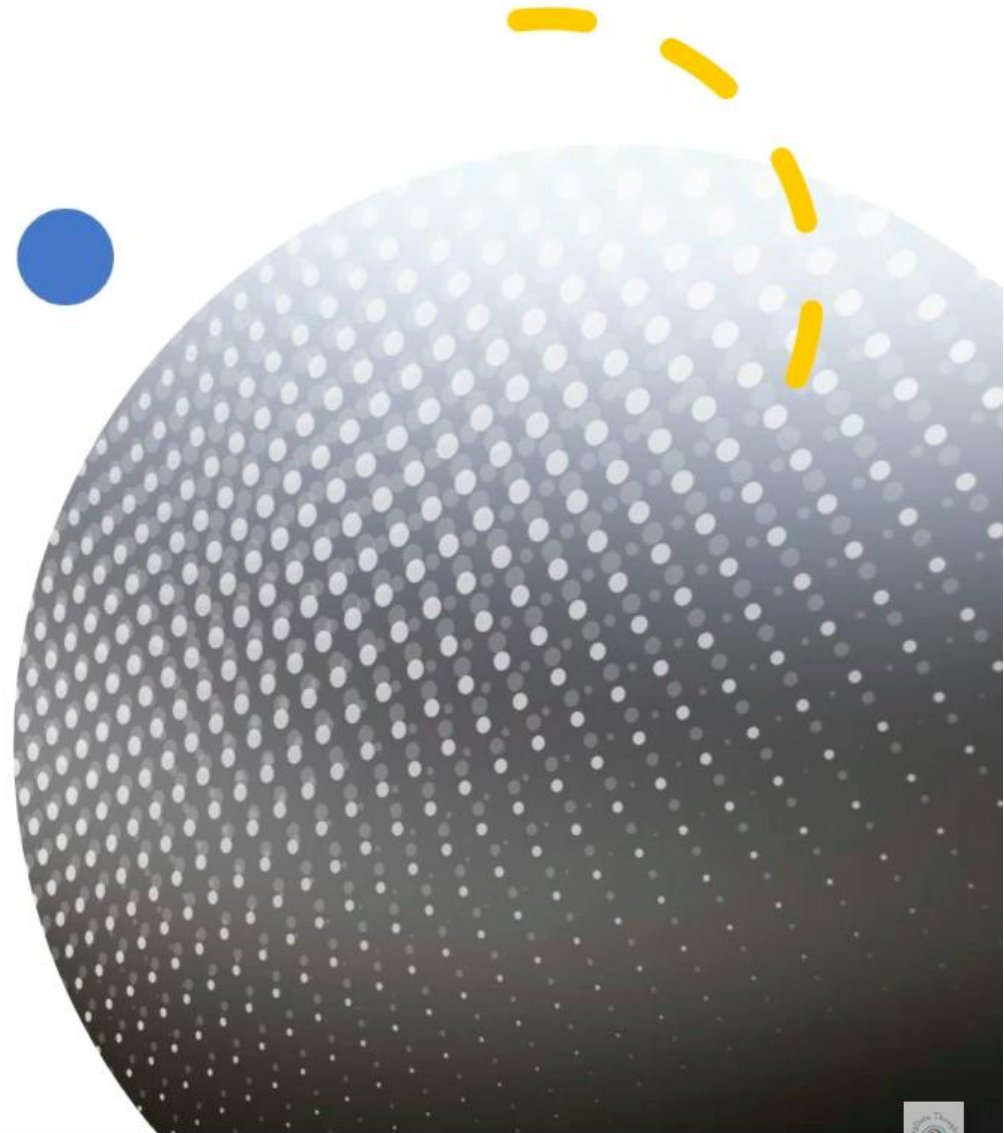
DISTRIBUTES  
COMPUTATIONAL  
PROCESSING OF DATA  
ACROSS MULTIPLE NODES.



COMPUTE AND STORAGE  
CAN SCALE INDEPENDENTLY

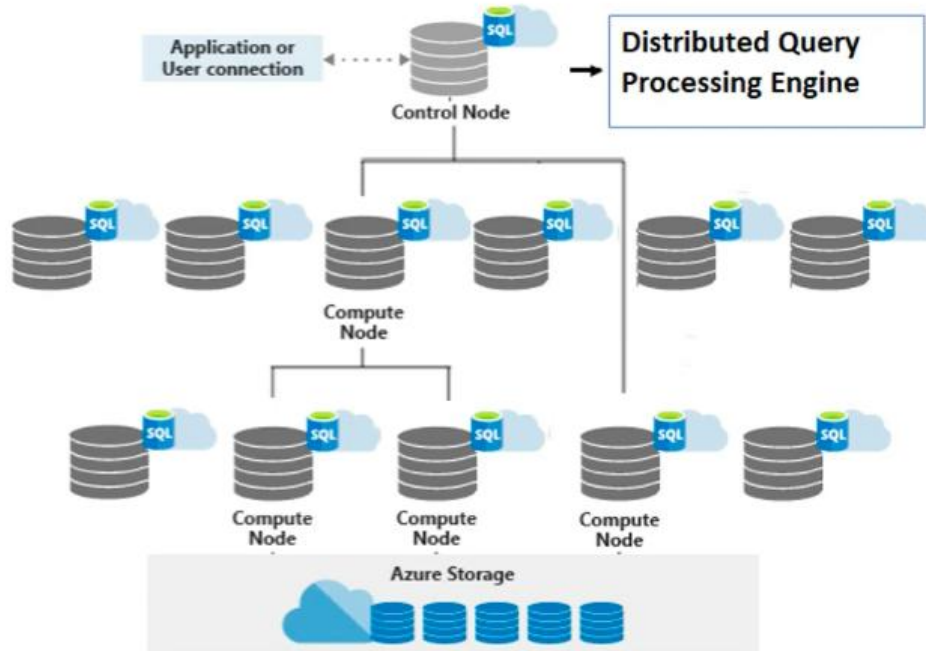


UNIT OF SCALE IS AN  
ABSTRACTION OF COMPUTE  
POWER - DWU



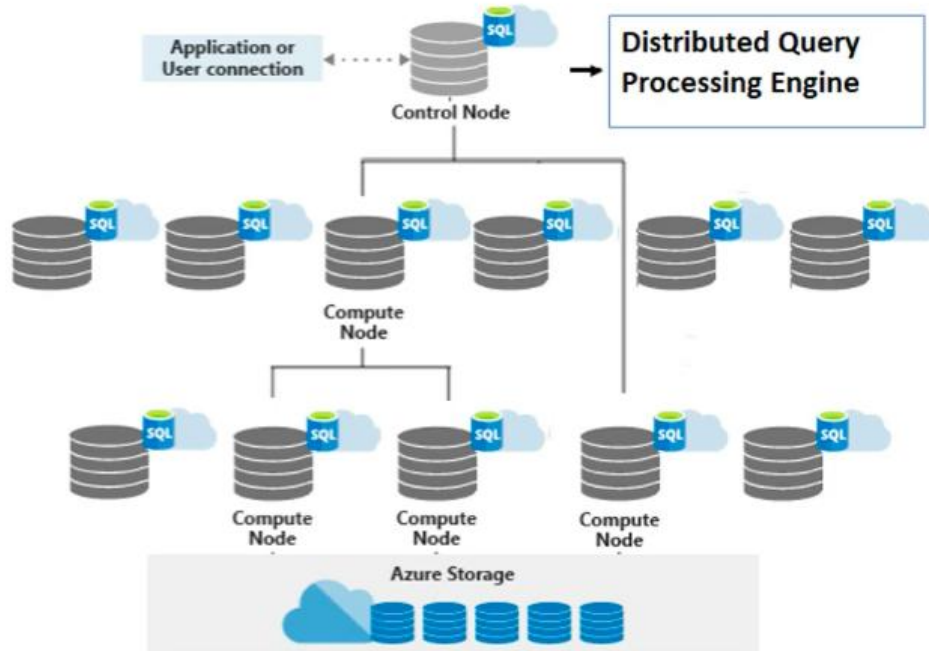


## How it works?

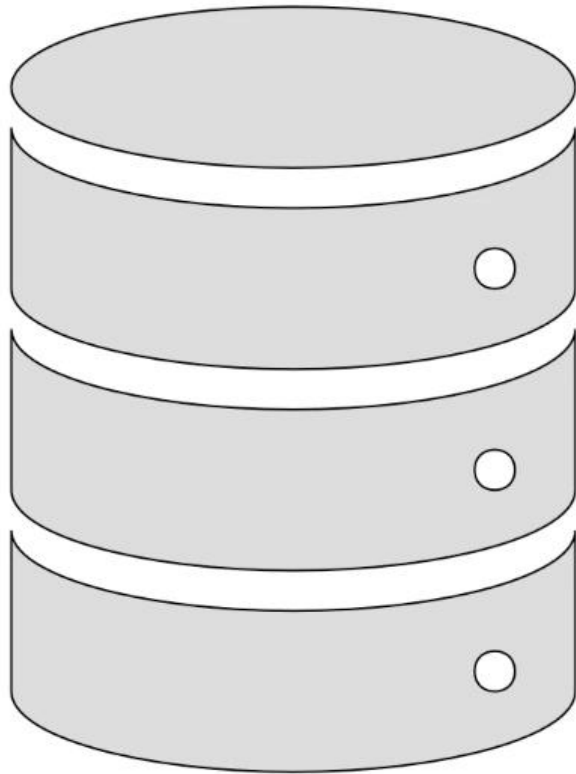


- Synapse SQL uses a node-based architecture.
- Applications connect and issue T-SQL commands to a Control node, which is the single point of entry for Synapse SQL.
- The Azure Synapse SQL Control node utilizes a distributed query engine to optimize queries for parallel processing,
- Control node passes operations to Compute nodes to do their work in parallel.

# Serverless SQL pool



- scaling is done automatically to accommodate query resource requirements.
- As topology changes over time by adding, removing nodes or failovers, it adapts to changes and makes sure your query has enough resources and finishes successfully.



# SQL Pools

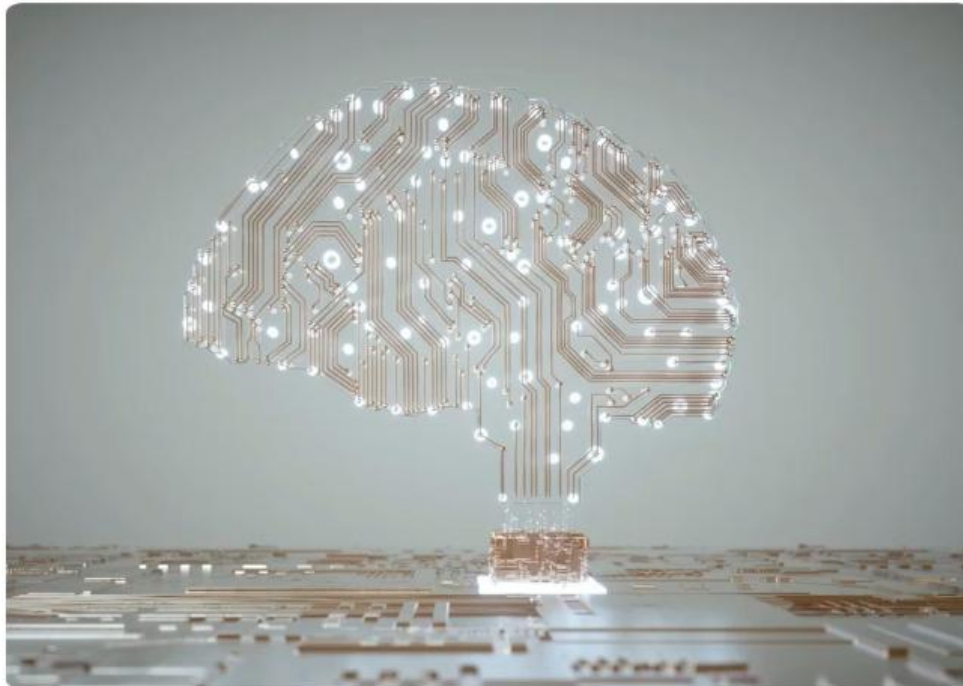
- **Dedicated Pool**
  - Grow or shrink compute power, within a dedicated SQL pool, without moving data.
  - Pause compute capacity while leaving data intact, so you only pay for storage.
  - Resume compute capacity during operational hours.
- **Serverless Pool**
  - Scaling is done automatically





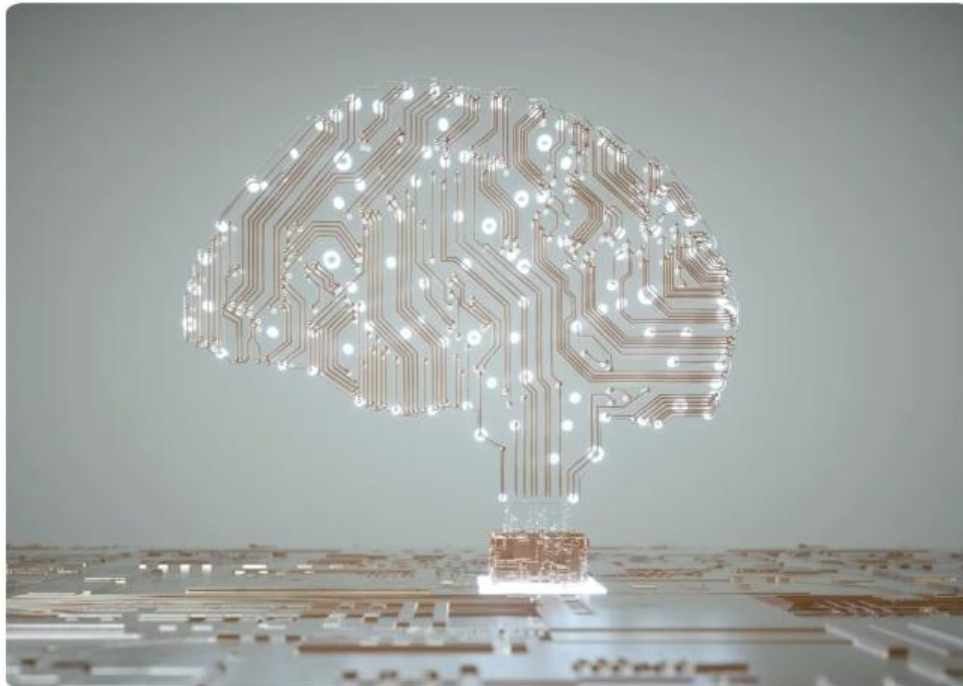
# Storage

- Synapse SQL leverages Azure Storage
- There is a separate charge for your storage consumption.
- When data is ingested into dedicated SQL pool, the data is sharded into **distributions** to optimize the performance of the system
- Sharding patterns supported:
  - Hash
  - Round Robin
  - Replicate



## Control Node

- The Control node is the brain of the architecture. It is the front end that interacts with all applications and connections.
- In serverless SQL pool, the DQP engine runs on Control node to optimize and coordinate distributed execution of user query by splitting it into smaller queries that will be executed on Compute nodes



## Compute Node

- The Compute nodes provide the computational power.
- In serverless SQL pool, each Compute node is assigned task and set of files to execute task on
- In dedicated SQL pool, distributions map to Compute nodes for processing.
- The number of compute nodes ranges from 1 to 60

---

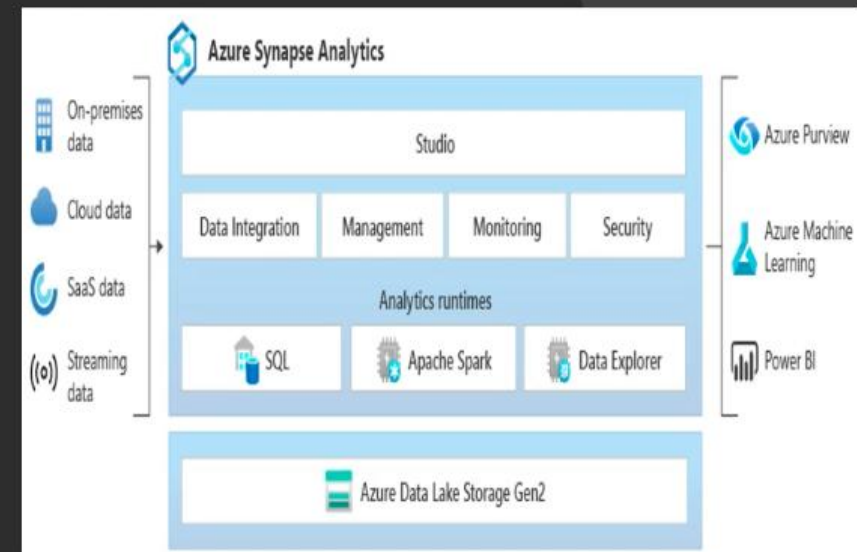
# What is Azure data explorer

---

- Azure Data Explorer is a big data analytics platform that makes it easy to analyze high volumes of data in near real time.
- The Azure Data Explorer toolbox gives you an end-to-end solution for data ingestion, query, visualization, and management

# What is Azure Synapse Data explorer?

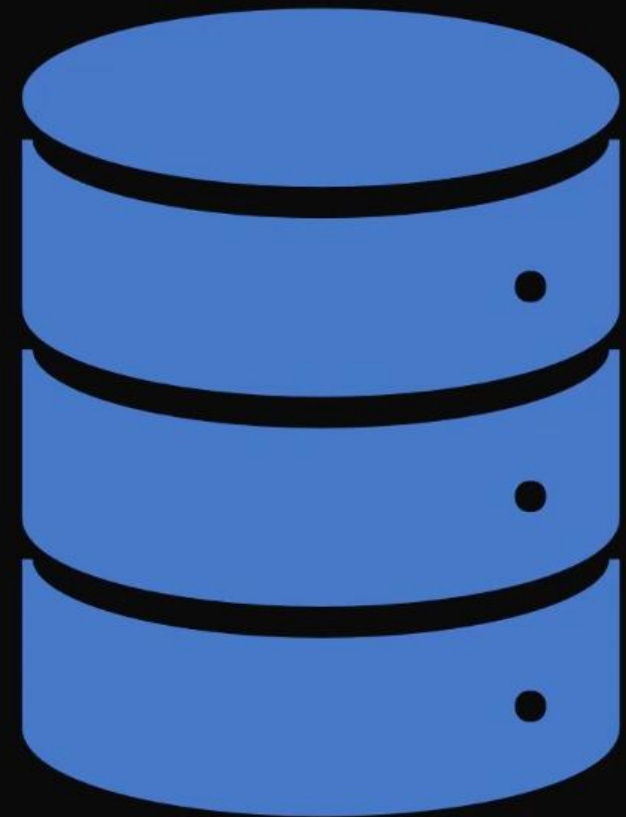
- Azure Synapse Data Explorer provides interactive query experience
- insights from log and telemetry data.
- It's a analytics runtime
- Has powerful indexing technology to automatically index free-text and semi-structured data commonly found in telemetry data.



# Features

---

- Easy ingestion
  - No complex data modeling
  - No index maintenance
  - Democratizing data analytics - Kusto Query Language
  - Independent Compute and storage scaling
  - Integration - Apache Spark, and SQL engines
- 





# Architecture

---

- Data Explorer pools implement a scale out architecture
- Storage and compute scale independently
- One set of compute - automatically indexing, compressing, caching, and serving distributed queries.
- Other set of computes - Data management service responsible for background system jobs, and managed and queued data ingestion
- Data is persisted on managed blob storage accounts using a compressed columnar format.





---

# Azure Data Explorer cluster

---

- Each Azure Data Explorer cluster can hold up to 10,000 databases and each database up to 10,000 tables.
- The data in each table is stored in data shards also called "extents". All data is automatically indexed and partitioned based on the ingestion time. Unlike a relational database, there are no primary foreign key constraints or any other constraints
- An Azure Data Explorer database can contain:
  - Tables: Made up of a set of columns. Each column has one of nine different data types.
  - External tables: Tables whose underlying storage is in other locations such as Azure Data Lake.



# Data Ingestion

- Data can be ingested in two modes: **Batching** or **Streaming**. Batching ingestion is optimized for high ingestion throughput and fast query results. Streaming ingestion allows near real-time latency for small sets of data per table.
- Data can be ingested using managed pipelines — Event Grid, IoT Hub, and Azure Data Factory.
- Connectors and plugins such as the Logstash plugin, Kafka connector, Power Automate, and Apache Spark connector can be used.
- Programmatic ingestion using SDKs, or LightIngest can be used.



# Data Analysis

---

- Azure Data Explorer uses the proprietary Kusto Query Language (KQL) to analyze data
- KQL is optimized for fast-flowing, diverse, big data exploration.
- Queries reference tables, views, functions, and any other tabular expressions. This can include tables in different databases or even clusters.



# Visualization

---

- Azure Data Explorer Web UI can be used to run queries and build dashboards. It supports a display of up to 500 K records and thousands of columns
- Native connectors can be used such as Power BI and Grafana.
- Azure Data Explorer also has ODBC and JDBC connector support to tools such as Tableau and Qlik.



# When to use ?

---

- IoT Analytics solutions
- Consolidate and correlate your logs and events data across on-premises, cloud, and third-party data sources.
- Analytics SaaS solutions
- Pattern recognition, Anomaly detection, forecasting etc.
- Log search solutions

