



## *Azure Data Fundamental DP-900*

*HASSAN RAFIQUE MSc Data Science*

- **Describe core data concepts**
- 
- **Identify considerations for relational data on Azure**
- 
- **Describe considerations for working with non-relational data on Azure**
- 
- **Describe an analytics workload on Azure**
- 
- **Azure Synapse Analytics**
- 
- **Azure cosmos DB**
- 
- **Azure Data Factory**
- 
- **Stream and Batch processing**
- 
- **Data warehouse**

# Azure DP - 900

## 1 - Explore Core Data Concept.

Key-value, Document DB, Graph DB  
Relational DB is some types of DB.

### Transaction DB

Ex: Banking data.

It must deal to ACID.

✓ **Atomicity** :

- Execute in single unit
- Fail or Pass.

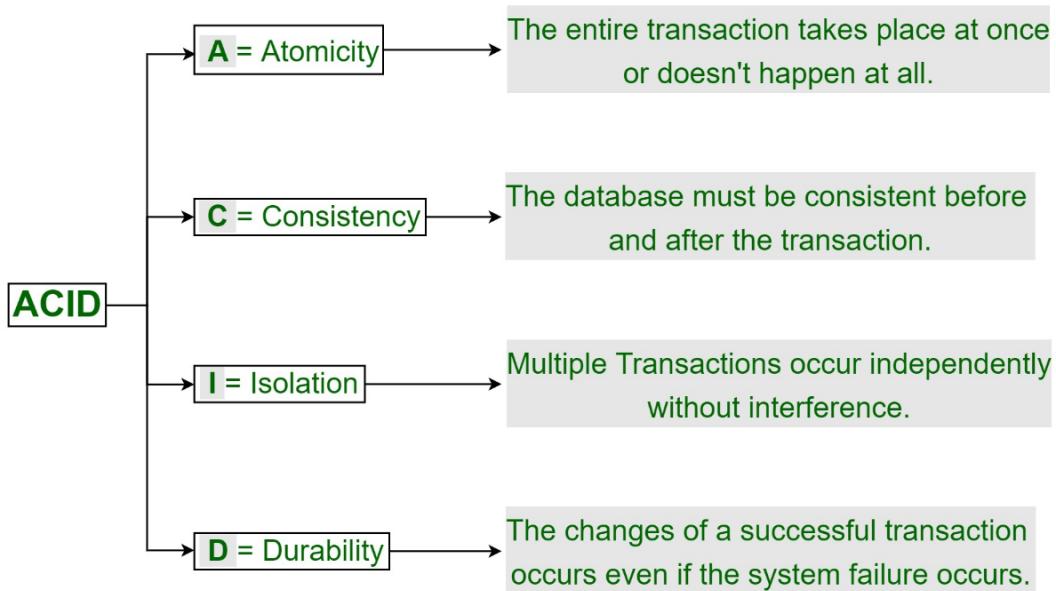
✓ **Consistency** .

Take data from valid state to another state.

✓ **Isolation** : multiple transaction occur independent

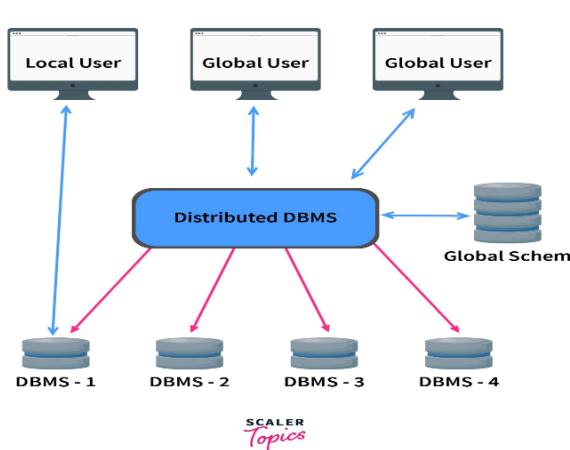
✓ Durability:- Committed transactions remain committed even if system failure occurs.

## ACID Properties in DBMS



DG

Distributed DB.  
Data store in different location on same or different network.



## Batch and Stream Process:-

- data in the form of group is Batch.
- apply OLAP.
- Data warehouse solution, Monthly analysis
- Stream OLTP. transactional data. New incoming data.

Ex: Stock Price

Online gaming company  
Real-estate website.

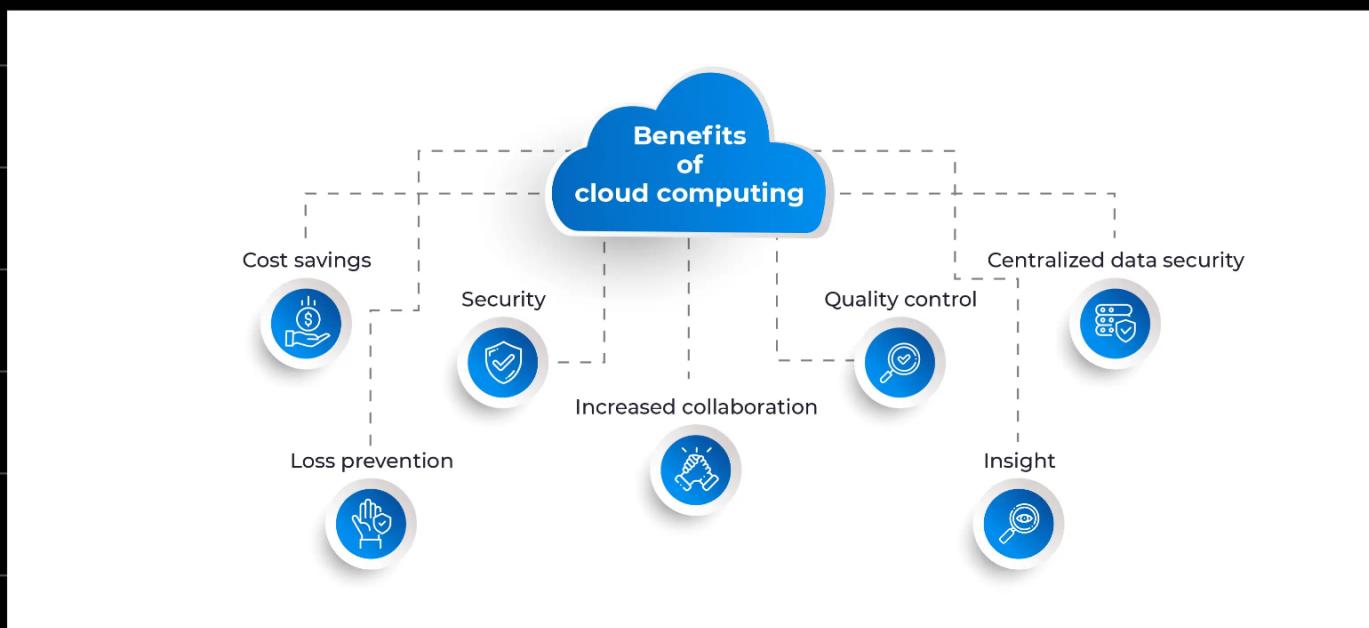


## 1 Role and Responsibilities in Data:

- Data Administrator .
- Handle data
- modify DB structure

- Control and monitoring
- Backup database
- Data replication.
- Install and upgrade DB server.

• EX:- Azure Data studio  
SQL Server Management Studio.



## Relational and Non relational data.

Entity, field , Table, Name

Relational DB structure is fixed format.

\* Primary Key : Identify



## \* Foreign Key:-

Primary key in another table called foreign key.

## \* Cluster Index / Non Cluster Index .

- Insert data in table with order Format
- Fast search
- Primary key create to create cluster automatically.

ID	Name
1	A
2	B
3	C

## \* View :-

Immediate access to relative data.

A virtual Table.

"Create VIEW as  
Select Name From mytable;

## Non-Relational

Azure Cosmos DB.

- semi-structure
- unstructured data } data structure.

JSON - semi-structure , XML,

### • AVRO - Apache -

Row based format. Header store JSON information store in Binary.

### • Apache ORC

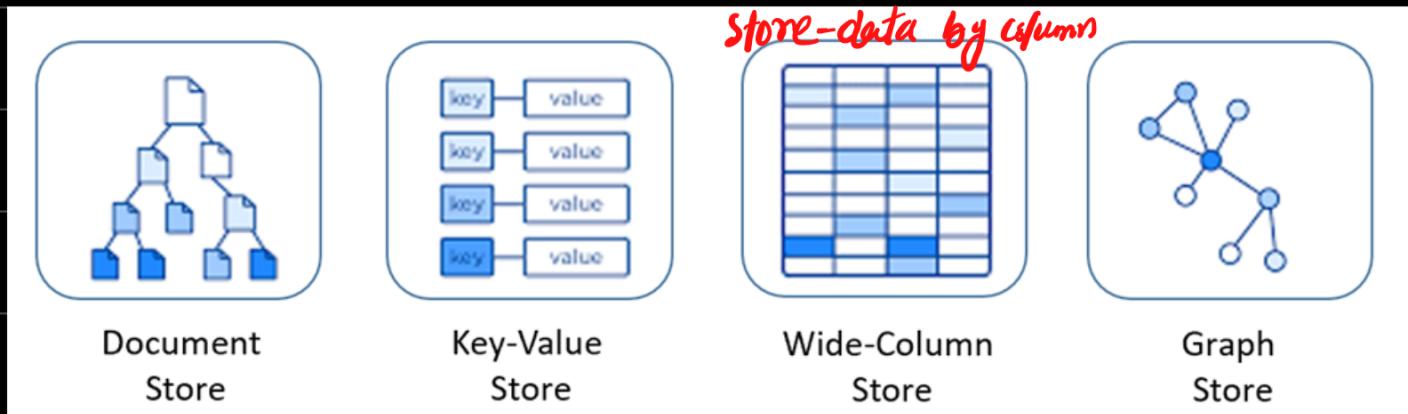
### • Parquet :-

Column data base.

## TYPES

Azure Cosmos DB.

- key-value — Azure Table Store.
- column — column family - Apache Cassandra
- graph — neo4j
- Document — Cosmos DB.



Ingestion → Process → Explore

ETL, ELT.

ELT use for complex data model.



Azure data factory.

Cloud base - data transform tools.



2- **SQL:**

DBMS..

on premiss      — cloud.

## SQL Server

Stored Procedure  
linked server ( SQL Server we join  
to one server to another server )

PaaS - Azure data Services are available  
to create.

SQL, Maria dB, Postgress > Relational-DB  
my SQL Server are popular tools .

On-premises migrate → Azure cloud.  
Data Data

- Azure resources managed by sever.
  - { Availability
  - Scalability
  - Security, [AD, key vault]
  - threat detection
  - Automatic update

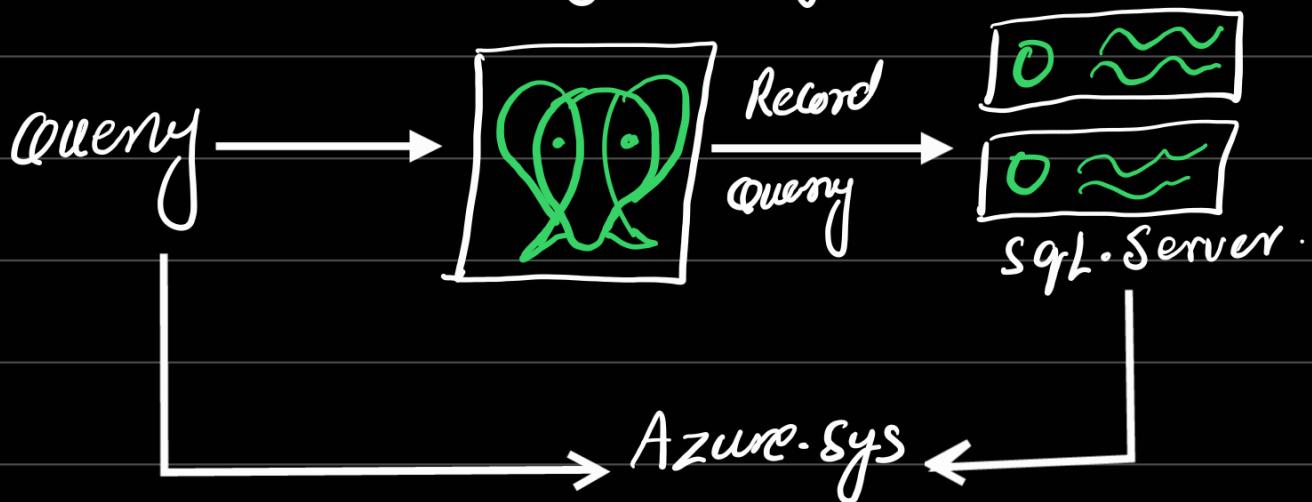
SQL Server login → Azure AD

SQL server managed is good for connecting more server.



## Postgres (Pg-admin tool)

- Hybrid relational-object database
- Relational and non-relational store
- Store geometric data [2d-data]
- Pgsql own Query Language.



- Support ultra high performance.
- Create <sup>read</sup> Replicas in Azure to 5 number virtual Network.

Azure VM  $\leftrightarrow$  AVN  $\leftrightarrow$  Azure service

## RBAC:-

allow resources who are use Resource

owner, Reader, Contributor.

Azure DB communicate over Port 1433.



3

## Azure Cosmos DB

Azure database services for non relational.

- (1) Azure Cosmos DB.
- (2) Table Storage
- (3) Blob Storage
- (4) File Storage [Continuous Availability]  
[Transfer legacy system]

✓ - Table Storage.

- un-structure data store
- Easy and Scalable.

- Partition Key, Rowkey.

✓ Blob Storage.

image, video, storage

① - Block blob

② - Page - 8Tb - VM

③ - append -

Blob → Container → Blobs

✓ File Share :-

- Data replicated
- Scalability
- move data from legacy system to cloud.
- automatically handle delta.
- Back-up file

✓

Cosmos DB -

- Multimodel NOSQL system. | store document data.
- API *Cosmos DB support*.
  - SQL API
  - MongoDB API
  - Table API
  - Cassandra API
  - Gremlin API (graph API)

Document in Cosmos db

Partition-key — Container

- automatically allocating space
- Height Availability
- Replicas

## Non-Relational.

Provision:-

action of providing and supply something like Azure-portal

Cosmos DB → Container → data

Data Lake Storage:

- use to store large amount of data.

Hot tier → High storage cost

Security Component:  
Firewall

# Active directory (AD)

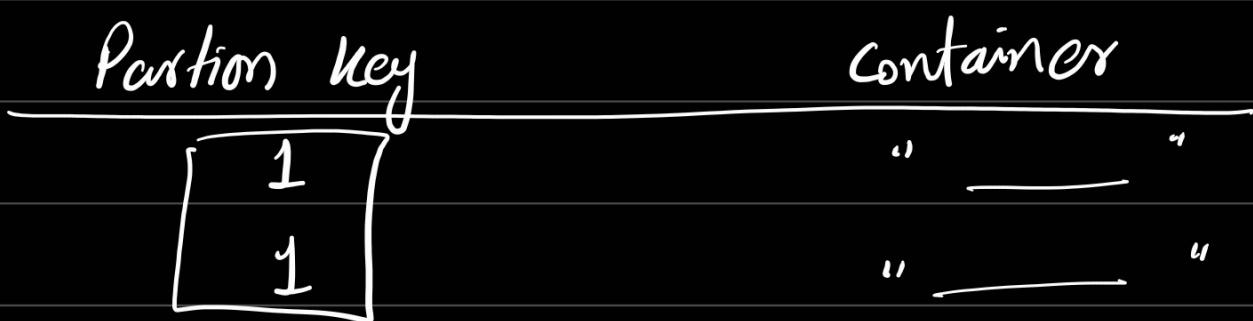
- user authentication
  - multi-user authentication
- etc = mobile
- Add user
  - Access control

# COSMOS-DB

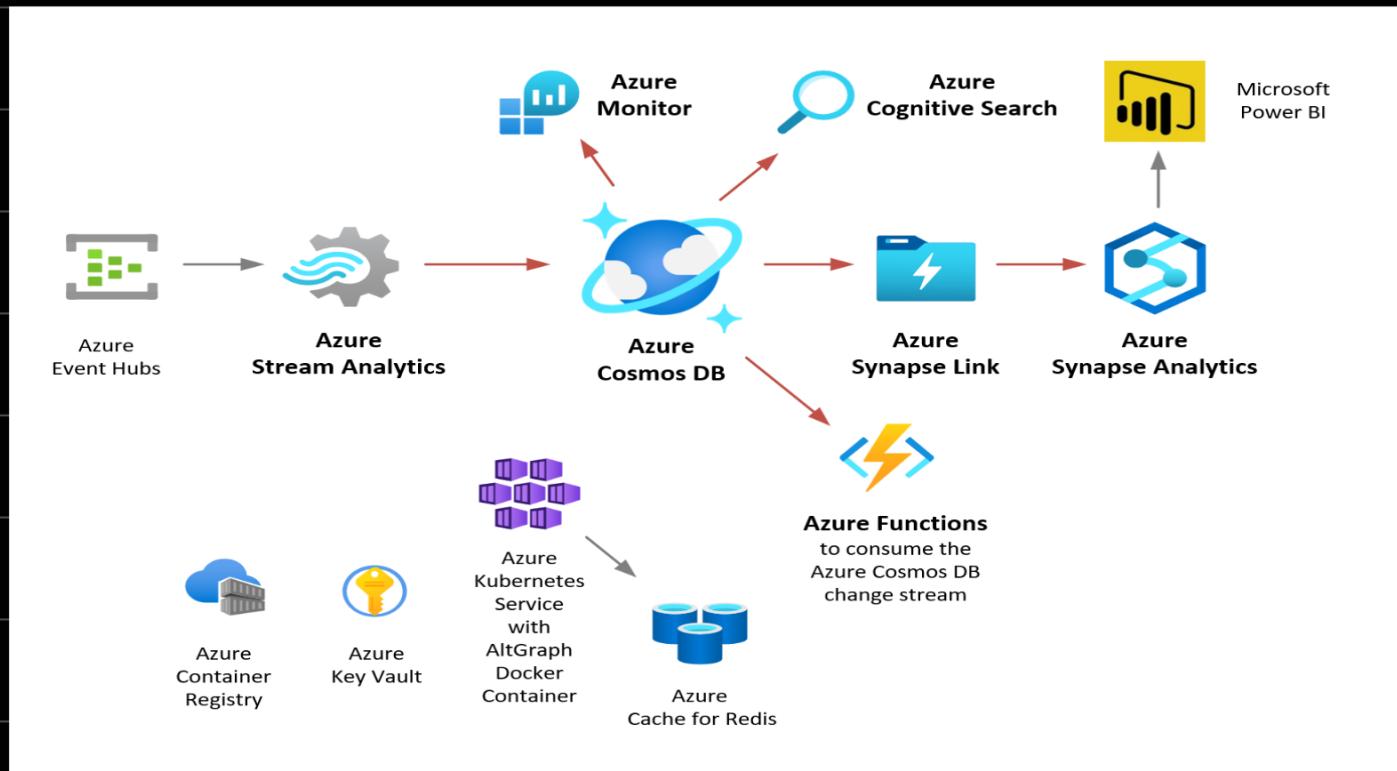
- Replication
  - in region - 4 time
  - outside region

(\*)

- It manage data as set of document.
- select partition to get all related containers.



(\*) Native API of Cosmos DB is SQL API.



## Data Warehouse:-

- Data ingestion
- store
- PowerBI
- Data factory. (handle ETL service)

Synapse:-

Dat warehouse + Bigdata Analytics.

4

Azure service to build datawarhous

Solution.

Data factory , Data bricks , Synapse  
Power BI .

Stream and batch data  
**Data warehouse** . Store Structured data.



## ① Data factory:-

- ETL / ECT Process.
- Ingest data
- Extract Interesting Data.
- Data Integration Service.

## ② Data Lake

- hold raw data
-  directory
- hierarchical Name Space

Store data in  
Folder in Folder.

Note - Blob only mimic data

- HDFS

Data lake → Factory → Synapse → Report

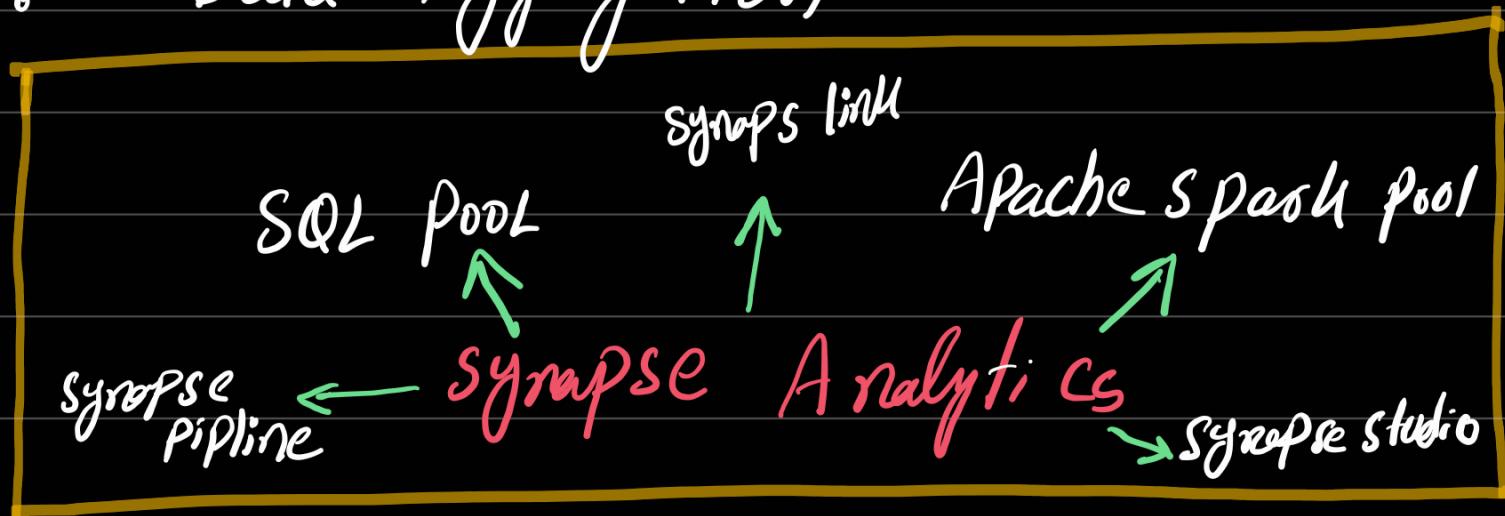
### ③ Data bricks

is an apache spark environment running on Azure, use for Machine learning.

Data Processing.

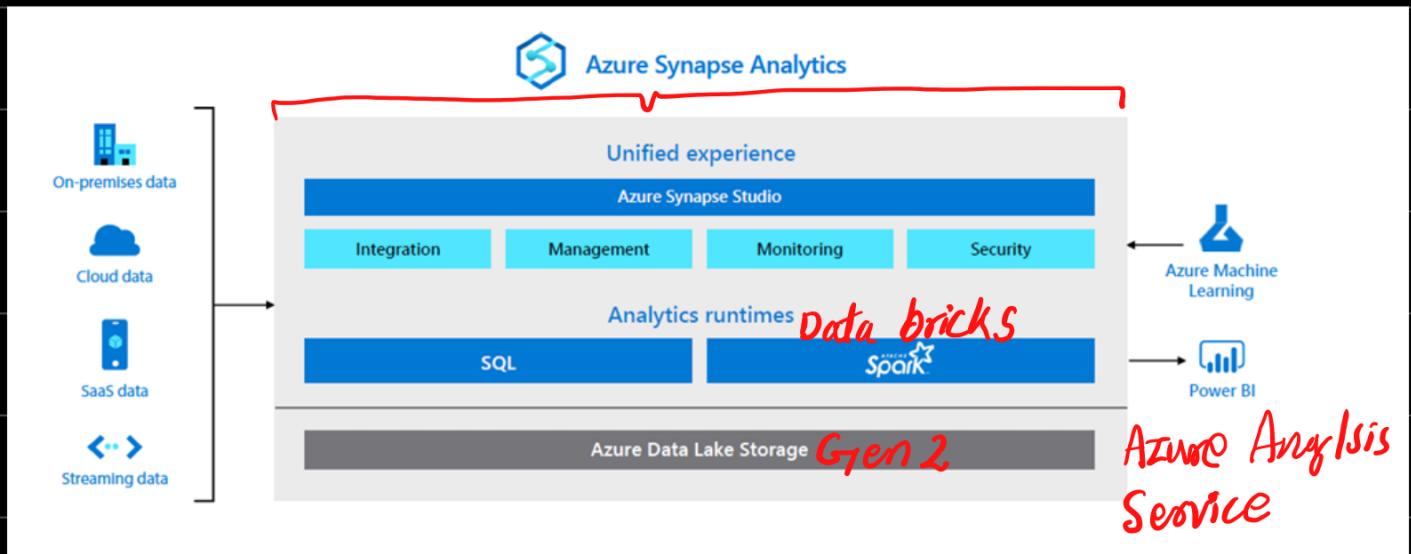
### ④ Synapse

- ingest data from different source
- Data Aggregation



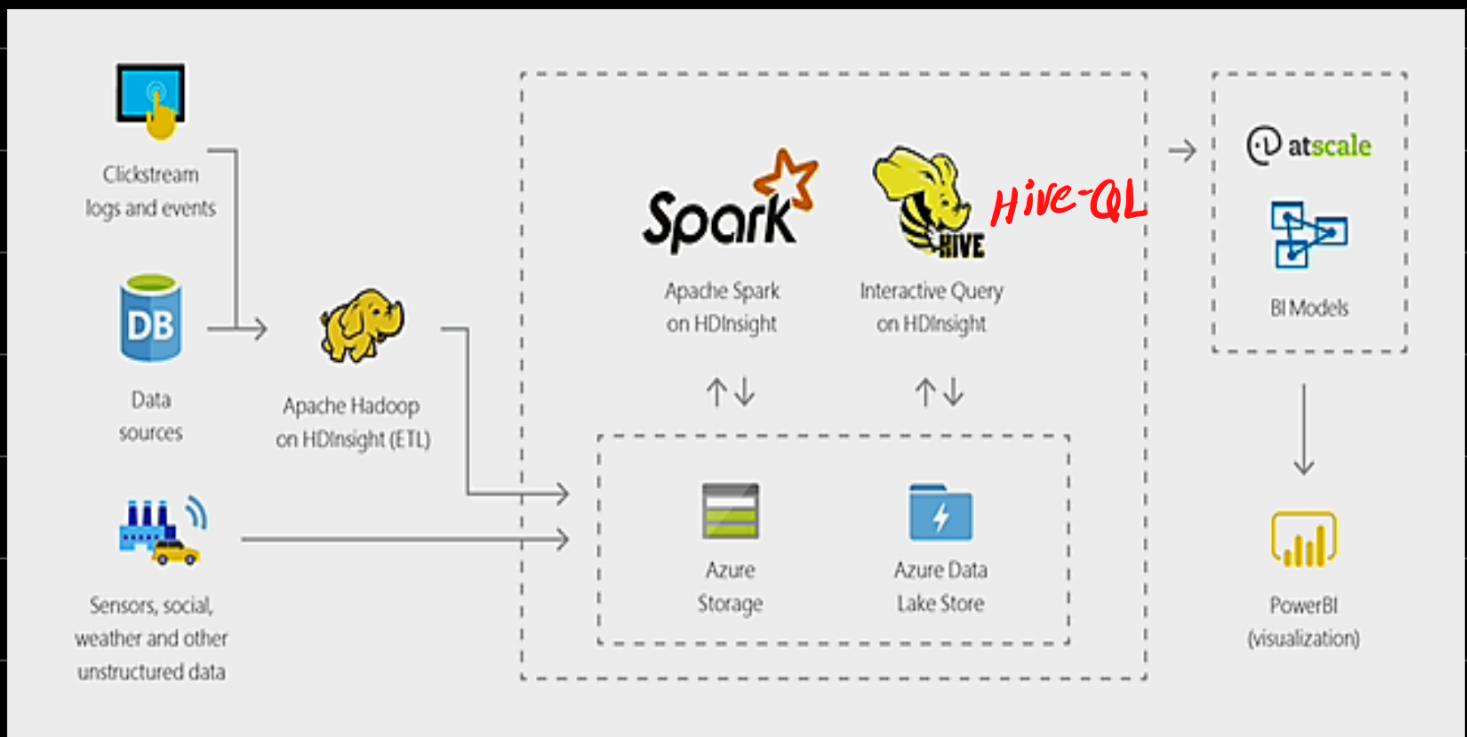
(\*) Spark is in memory computation tool.

- ETL operation



## Azure HD insight

- Big data Processing
- implement Cluster model.
- Support Streaming technology such as (Kafka).



## Azure Data bricks :-

- parallel process.
- streaming data
- use Apache Spark to distribute data.

## Data storage and processing :

★ Azure Synapse Analytics . ]

★ Data factory

★ Data bricks

★ Data Lake

} Data Process

## ★ PolyBase :-

- Polybase is technique to make external data → look like sql table.
- Run Queries against these table directly.

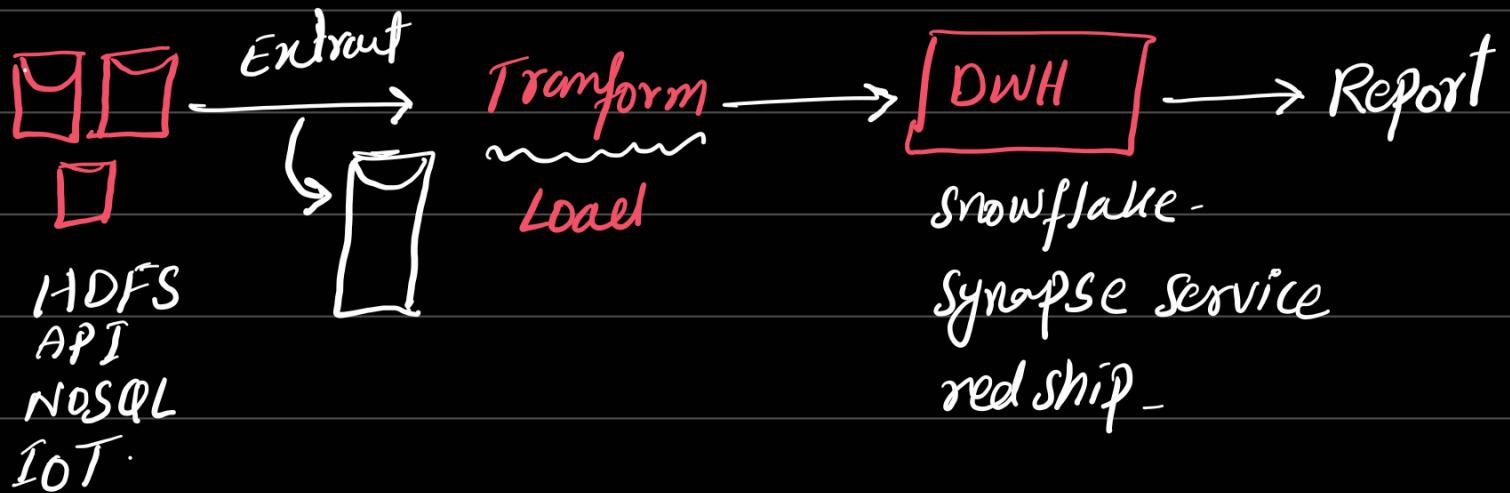
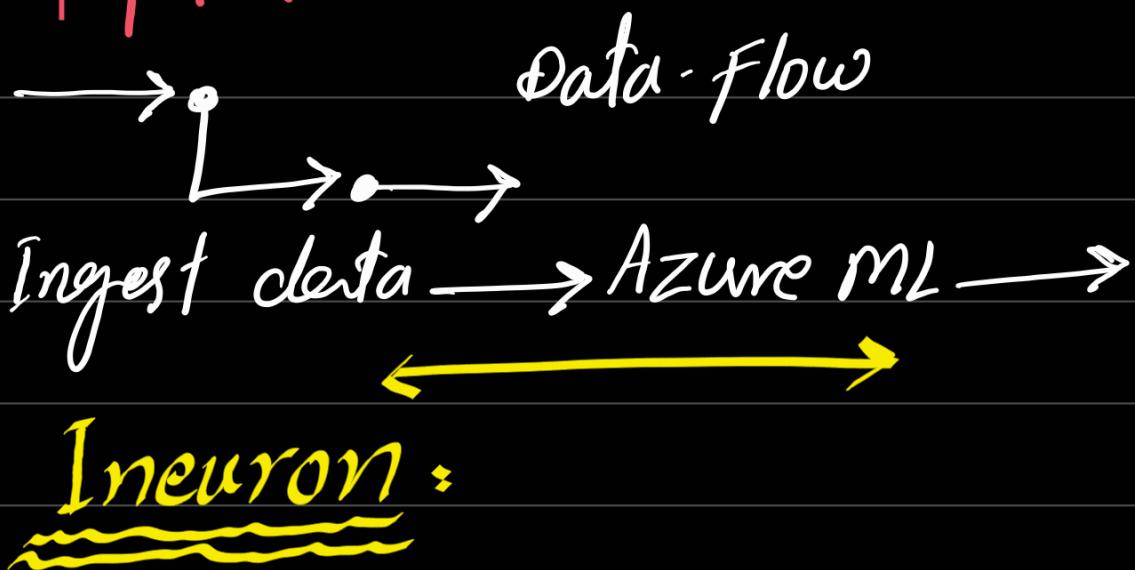
## [Pool used]

- complex data
- data ingest from many source

- Synapse Spark to write code base on Apache Spark.
- To train ML model.
- In memory Computing more faster.

## Component of Synapse

### Pipeline



# [ELT]



{ HDFS , Amazon S3 }  
{ Azure Blob }

## Data Lake :-



Categorical data in DWH for specific purpose called data-mart.

## OLAP / OLTP

### OLTP:-

If support ACID Property one

main goal to handle realtime data.

- RDMS
- MySQL, Oracle

## OLAP:-

- to handle large data
- Analytics on history data
- Data warehouse

Hive, Azure Synapse.

# OLAP Vs OLTP

