Ingest Data with Microsoft Fabric

"Ingesting Data" means bringing data from one place into another.

Ways to Ingest Data:

- Upload files (manual file upload)
- Dataflows Gen2 (Power Query transformations)
- Notebooks (Spark code)
- Data Factory pipelines (Copy Data activity- Automated data movement)

Example
Moving into a new apartment:

- Upload = carrying boxes yourself
- Dataflows = hiring movers who also help organize
- Notebooks = using power tools to assemble furniture
- Pipelines = setting up a conveyor belt that moves boxes automatically

Schema-on-Read
Fabric lakehouses use schema-on-read, meaning data doesn't need to be perfectly structured before loading. This makes ingestion more flexible.

Example
It's like dumping all your photos into a folder and organizing them later — instead of sorting them before copying.

Orchestrate Data with Pipelines
Pipelines are automated workflows that move and transform data.

Example
Making a cup of Coffee:

- Grinding beans
- Brewing
- Pouring
- Serving

A pipeline is like automating that entire process.
Key Activities

- Copy Data = move data without changes
- Dataflows = transform data

- Control flow = loops, conditions

Use Apache Spark in Microsoft Fabric

Spark is a system that processes large amounts of data by splitting the work across many computers (worker nodes) controlled by a driver node. This is possible with Spark pools, RDDs, DataFrames, and the Spark catalog.

Example

Cleaning a huge stadium after a concert:

- One supervisor (driver)
- Many cleaners (workers)
- Everyone cleans a section (partition)
- Work finishes faster than one person doing it alone

Key Spark Concepts

- RDD = raw distributed data
- DataFrame = table-like data you can filter, group, sort
- Spark SQL = SQL-like queries
- Temporary views = short-lived tables
- Spark catalog = storage for tables and views

Work with Real-Time Intelligence (KQL + Eventhouse)

Real-Time Data means Data that updates every second — like live website clicks or sensor readings.

Tools used:

- KQL queries
- Real-time dashboards
- Materialized views
- Eventhouse + KQL databases

Example

Think of a live scoreboard at a football match — it updates instantly as events happen.

KQL Basics

- Uses pipes |
- Case-sensitive
- Can visualize results
- Supports T-SQL too