**MODULE -2**

Data pipelines -> sequence of activities that orchestrate overall process.

**E**xtracting data -> **L**oading to destination -> **T**ransforming it along the way

Run pipeline interactively UI/ schedule them (run automatically)

**Understand pipelines:**

Encapsulate sequence of activities that perform **data movement & processing tasks.**

Define **data transfer, transformation activities.**

Orchestrate these activities through **control flow** activities (manage **branching, looping, other processing logic)**.

**Graphical pipeline canvas** -> build complex pipelines with minimal or no coding required.

A screenshot of a computer

AI-generated content may be incorrect.

**Core pipeline concepts:**

1. Activities:

Executable tasks in pipeline.

Define activities flow by connecting them in sequence.

**Activity1 – output(success/fail/complete) 🡪 Activity2**

2 categories of activity in pipeline

* Data transformation: data transfer operations
  + Copy Data (extract from src 🡪 load to dest)
  + Data Flow 🡪 transformations to data as it is transferred
  + Notebook activities 🡪 run spark notebook
  + Stored procedure 🡪 run SQL code
  + Delete Data 🡪 delete existing data
* Control flow activities: used to implement
  + Loops
  + Conditional branching
  + Manage variable and parameter values

Enables you to implement complex pipeline logic to orchestrate data ingestion and transformation flow.

1. Parameters:

Pipelines🡪 can be parameterized 🡪 to provide specific values to be used each time a pipeline is run.

Increasing reusability of your pipelines🡪 enables you to create flexible data ingestion and transformation processes.

1. Pipeline runs:

Pipeline is executed => data pipeline run is initiated.

Runs:

* + On-demand
  + Scheduled (to start at a specified frequency)

**Run ID** -> to review run details to confirm they completed successfully and investigate the specific settings used for each execution.

**Use Copy Data activity:**

To ingest data from external source into lakehouse file or table.

Copy data activity + other activities 🡪 to create a repeatable data ingestion process

Example:

**Delete data** activity

Remove existing data

**Notebook** activity run Spark code that transforms data in file & loads it into table

**Copy data** activity to replace deleted data with a file containing the data from external src

**Copy Data activity settings**:

A screenshot of a computer

AI-generated content may be incorrect.

When to use Copy data activity:

If you want to **copy data** directly between supported src -> dest **without apply any transformations**/ when you want to import the raw data and apply transformations in later pipeline activities.

If you want to **apply transformations** to the ingested data/ merge data from multiple srcs, use **Data Flow activity** to run dataflow Gen2. Use **Power Query UI** to **define dataflow Gen2** that includes multiple transformation steps, and include it in a pipeline

**USE PIPELINE TEMPLATES:**

Create custom data ingestion and transformation processes.

A screenshot of a data pipeline

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**RUN AND MONITOR PIPELINES**

When you have completed a pipeline, you can use the **Validate** option to check that the configuration is valid, and then either run it interactively or specify a schedule.

A screenshot of a computer

AI-generated content may be incorrect.

**View run history**

You can view the run history for a pipeline to see details of each run, either from the pipeline canvas or from the pipeline item listed in the page for the workspace.

A screenshot of a computer

AI-generated content may be incorrect.

When you view a pipeline run history from the workspace page, you can select the **Run start** value to see the details of an individual run; including the option to view the individual execution time for each activity as a Gantt chart.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.