# Azure Data Factory Detailed Lab Guide

## Lab 1: Overview of Azure Data Factory (ETL vs ELT)

### Objective:

Understand the difference between ETL and ELT and how ADF supports both patterns.

### Steps:

1. Open [Azure Portal](https://portal.azure.com).
2. Navigate to your **Data Factory Studio**.
3. Review the **Copy Data tool** for ETL scenarios.
4. Explore **Data Flow** for ELT (data transformation post-load).

**Key Takeaway:** ETL = Transform before loading, ELT = Transform after loading into target system.

## Lab 2: ADF Components

### Objective:

Learn about and create basic ADF components.

### Steps:

1. Open **Author** hub in ADF Studio.
2. Explore the following:
   * **Linked Services** – define data connections.
   * **Datasets** – define schema/structure of data.
   * **Pipelines** – logical group of activities.
   * **Activities** – individual steps (Copy, Lookup, ForEach).
   * **Triggers** – schedule/event-based execution.

## Lab 3: Integration Runtimes

### Objective:

Configure integration runtimes for data movement.

### Steps:

1. Go to **Manage** hub → **Integration Runtimes**.
2. Add a new runtime:
   * Choose **Azure-hosted IR** (default, fully managed).
   * Choose **Self-hosted IR** if accessing on-premise or private network sources.
3. Test connection to validate configuration.

## Lab 4: Creating & Configuring Linked Services

### Objective:

Set up secure connections to common data sources.

### Steps:

1. Go to **Manage** hub → **Linked Services → + New**.
2. Create the following:
   * **ADLS Gen2** (provide storage account & authentication method).
   * **Azure SQL Database** (provide server, database, and credentials).
   * **Blob Storage** (storage account name and key).

## Lab 5: Dataset Creation

### Objective:

Define datasets for structured and semi-structured data.

### Sample Datasets:

* employee\_data.csv (CSV with columns: EmployeeID, Name, Department, Salary).
* department\_data.json (JSON file with DepartmentID and DepartmentName).

### Steps:

1. In **Author** hub → **+ New Dataset**.
2. For **CSV Dataset**:
   * Select **ADLS Gen2** linked service.
   * Point to employee\_data.csv.
   * Import schema.
   * Save as Employee\_CSV\_Dataset.
3. For **JSON Dataset**:
   * Select **Blob Storage** linked service.
   * Point to department\_data.json.
   * Import schema.
   * Save as Department\_JSON\_Dataset.

## Lab 6: Copy Activity Basics

### Objective:

Copy data between stores using ADF.

### Steps:

1. Create a new **Pipeline**.
2. Drag **Copy Data** activity to canvas.
3. Configure:
   * **Source**: Employee\_CSV\_Dataset.
   * **Sink**: Azure SQL Database (Employee\_SQL\_Dataset).
4. Run the pipeline manually using **Trigger Now**.

## Lab 7: Monitoring Pipeline Runs

### Objective:

Track and debug pipeline executions.

### Steps:

1. Go to **Monitor** hub in ADF Studio.
2. View pipeline run history.
3. Inspect success, failure, and duration.
4. Drill into failed runs to check error details.

## Lab 8: Error Handling & Retry Policies

### Objective:

Implement error handling and retry strategies.

### Steps:

1. Open your pipeline.
2. On **Copy Activity**, set **Retry** count = 3, Retry interval = 30 seconds.
3. Add a conditional activity to log errors or send notifications.
4. Test with incorrect credentials to simulate failure.

## Lab 9: ADF Pricing & Cost Estimation

### Objective:

Understand how ADF costs are calculated.

### Steps:

1. Open [Azure Pricing Calculator](https://azure.microsoft.com/pricing/calculator/).
2. Add **Data Factory**.
3. Estimate costs based on:
   * Number of pipeline activities.
   * Data movement volumes.
   * Integration Runtime usage.

## Lab 10: Source-to-Target Mapping

### Objective:

Map source fields to target fields in a pipeline.

### Steps:

1. In your **Copy Activity**, go to **Mapping tab**.
2. Enable **Auto-mapping**.
3. Adjust manual mappings:
   * EmployeeID → Emp\_ID
   * Name → Emp\_Name
   * Department → Dept
   * Salary → Emp\_Salary
4. Publish and re-run pipeline.

# 