# Scenario-Based Hands-On Assessment: Data Transformation and Loading using Azure Data Factory

\*\*Total Marks:\*\* 50  
\*\*Duration:\*\* 90–120 minutes  
\*\*Tools:\*\* Azure Portal, Azure Data Factory, Azure Blob Storage, Azure SQL Database

## Business Scenario

Your company, Contoso Logistics Pvt. Ltd., collects daily shipment tracking data from its regional branches. These CSV files are stored in an Azure Blob Storage container called `raw-tracking-data`. Management wants this raw data to be cleaned, transformed, and loaded into an Azure SQL Database to analyze delivery performance metrics.

## Dataset Description

You are provided a CSV file named `shipment\_data.csv` located in your Blob Storage container `raw-tracking-data`.

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| Shipment\_ID | int | Unique identifier for each shipment |
| Origin\_City | string | City from where shipment originated |
| Destination\_City | string | Shipment destination city |
| Distance\_KM | float | Distance between origin and destination in kilometers |
| Shipment\_Date | datetime | Date when shipment was sent |
| Delivery\_Date | datetime | Date when shipment was delivered |
| Status | string | Delivery status (Delivered, In-Transit, Returned) |

### Sample Data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Shipment\_ID | Origin\_City | Destination\_City | Distance\_KM | Shipment\_Date | Delivery\_Date | Status |
| 501 | Bangalore | Chennai | 340 | 2024-09-10 | 2024-09-11 | Delivered |
| 502 | Mumbai | Delhi | 1420 | 2024-09-12 | 2024-09-14 | Delivered |
| 503 | Kolkata | Hyderabad | 1480 | 2024-09-13 |  | In-Transit |
| 504 | Pune | Bangalore | 845 | 2024-09-11 | 2024-09-12 | Delivered |

## Tasks

\*\*Task 1: Environment Setup (10 marks)\*\*  
1. Create an Azure Data Factory instance.  
2. Create linked services for Azure Blob Storage and Azure SQL Database.  
3. Create datasets for the input (Blob CSV) and output (SQL table).  
✅ Evaluation: Proper creation and validation of linked services and datasets.

\*\*Task 2: Data Transformation (20 marks)\*\*  
1. Use a Data Flow in ADF to read from Blob Storage.  
2. Derive a new column `Delivery\_Days` = DATEDIFF(Delivery\_Date, Shipment\_Date).  
3. Replace blank `Delivery\_Date` values with the current date for in-transit shipments.  
4. Filter out records where Distance\_KM is less than or equal to 0.  
5. Select relevant columns for output.  
✅ Evaluation: Correct transformation logic and expressions used.

\*\*Task 3: Load Transformed Data to Azure SQL Database (10 marks)\*\*  
1. Create a table `Shipment\_Transformed` in your Azure SQL Database with appropriate schema.  
2. Configure the sink in ADF to write the transformed data into the SQL table.  
3. Validate successful data load.  
✅ Evaluation: Successful data load and correct column mapping.

\*\*Task 4: Pipeline Execution and Monitoring (10 marks)\*\*  
1. Create a pipeline that runs your data flow.  
2. Trigger the pipeline manually.  
3. Monitor the pipeline and ensure all activities succeed.  
✅ Evaluation: Successful execution and monitoring screenshot.

\*\*Bonus Task (Optional, +5 Marks)\*\*  
Store the transformed data into another Blob container `processed-tracking-data` in Parquet format.

## Evaluation Rubric (50 Marks)

|  |  |  |
| --- | --- | --- |
| Criteria | Description | Marks |
| Task 1 | Linked services and dataset setup | 10 |
| Task 2 | Transformation logic implementation | 20 |
| Task 3 | Data load and SQL table configuration | 10 |
| Task 4 | Pipeline orchestration and monitoring | 10 |