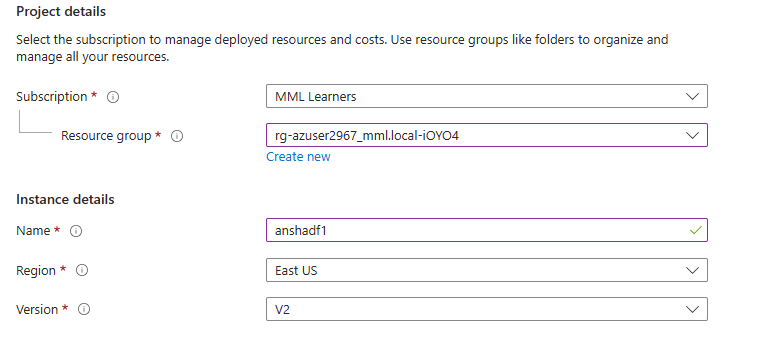
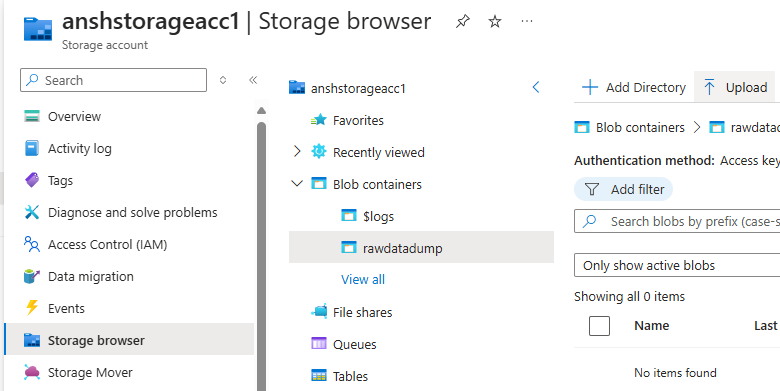
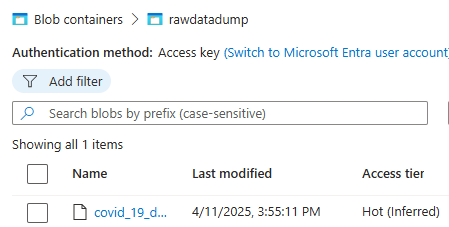
**Ansh Ranjan  
Azure Data Factory  
All Exercises**

**Exercise 1: Setting Up Azure Databricks & Spark Basics**

**TASK 1: Create a new Azure Databricks workspace.**

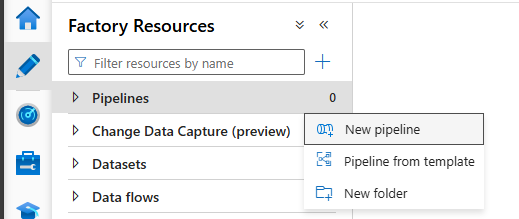
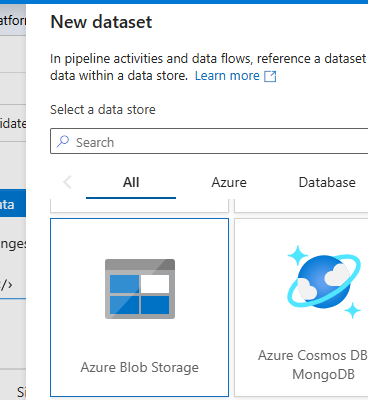
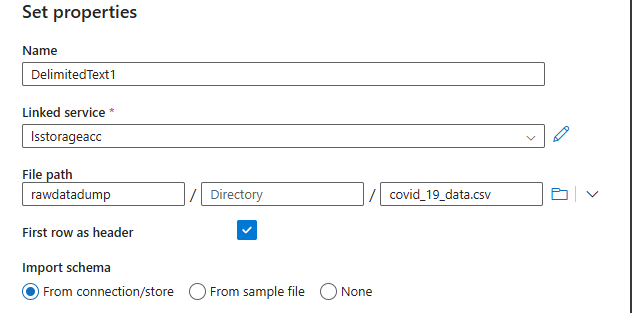
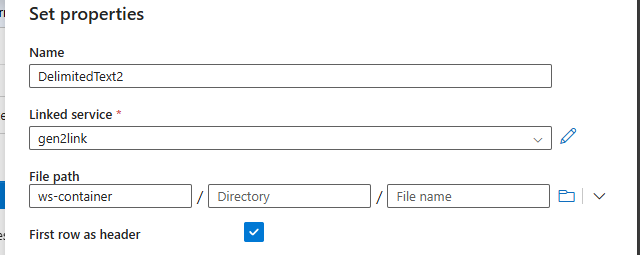
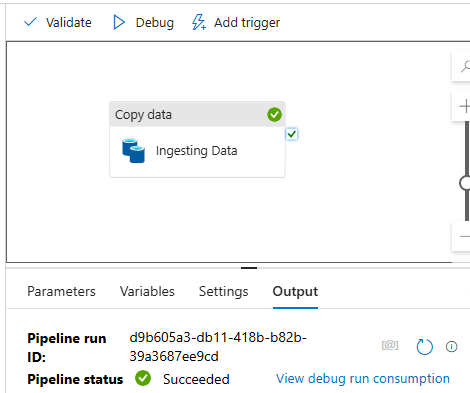
1. Go to Data Factories > Create > Enter details   
   

**TASK 2 & 3: Download the dataset and Upload to Blob Storage**

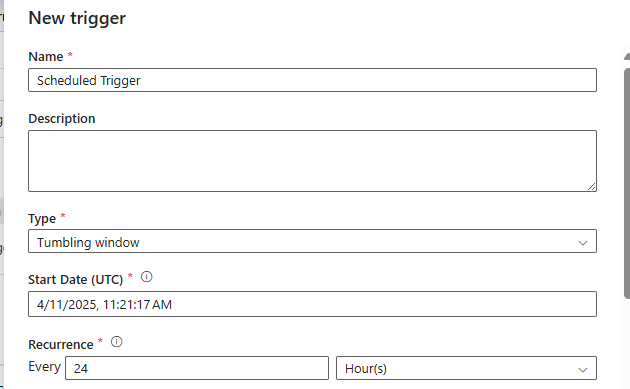
1. Go to your storage account > Storage Browser side tab > Blob Containers > select a blob container to upload your dataset > click on Upload   
   
2. Browse for your dataset csv file and hit Upload   
   

**Exercise 2: Building Pipelines for Covid-19 Data Ingestion**

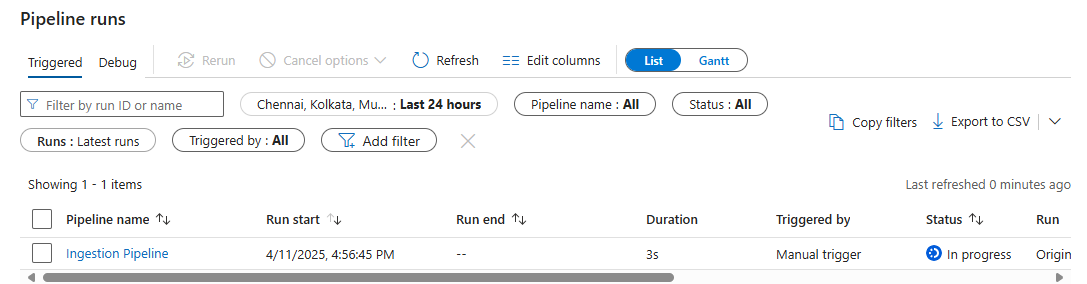
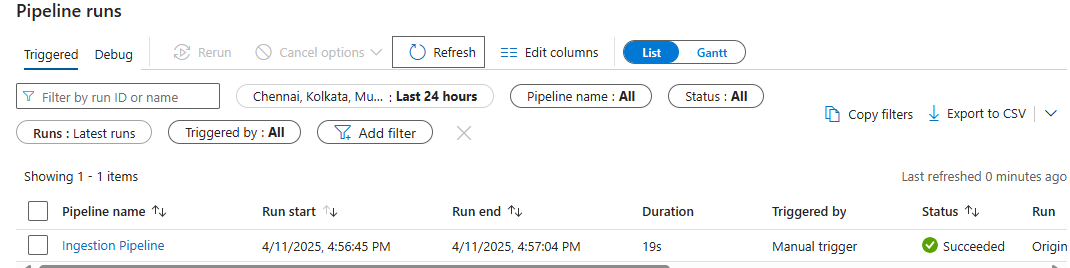
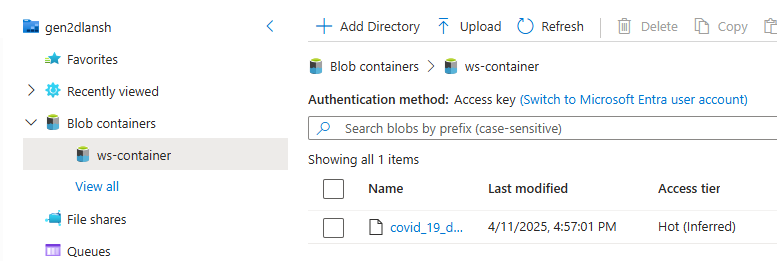
**TASK 1: Create an ADF pipeline to copy data**

1. Open your Data Factory and go to Author tab > Pipeline   
   
2. Drag a Copy Data activity onto stage
3. Go to source > New Dataset > Azure Blob Storage > Select CSV format   
   
4. Create a new linked service to your storage account. And select your path of csv file   
   
5. In Sink tab, select your destination. Click on New > Azure Gen 2 Data lake   
   
6. Click on Debug button to test your pipeline   
   

**TASK 2: Configure pipeline triggers for scheduled runs.**

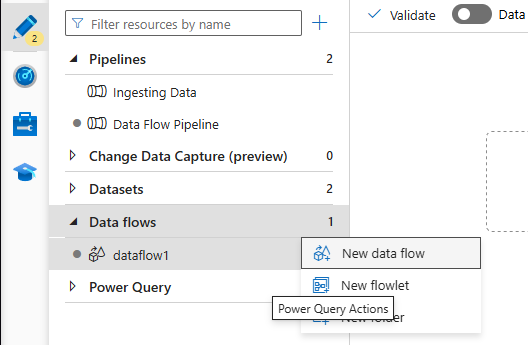
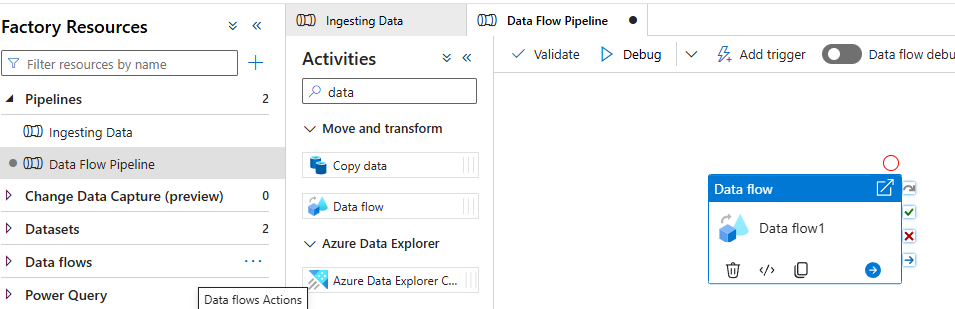
1. Click on Trigger > New/Edit > New > select type as Tumbling Runs > define start dates and reccurance time (going with 24 hrs)   
   
2. Publish your pipelines and Triggers.

**TASK 3: Monitor and troubleshoot the pipeline execution**

1. Trigger your pipeline once for execution. You can view the execution in Monitor tab   
   
2. The pipeline has successfully executed   
     
   

**Exercise 3: Data Transformation & Loading for Covid-19 Analysis**

**TASK 1: Use ADF's Mapping Data Flow**

1. Go to Author tab > Data Flow > New Data Flow  
   
2. Go to Author tab > Pipeline > New Pipeline > and grab the Data Flow activity  
   
3. Designing the Data Flow. To use adf’s data flow, for aggregate purpose first we must use derived columns, then aggregate and at last sink in storage which is accessible in databricks notebooks.  
   A screenshot of a computer

   AI-generated content may be incorrect.
4. In derived columns, make columns for date name for which you are going to do analytical process.  
   A screenshot of a computer

   AI-generated content may be incorrect.
5. After creating columns count the no of cases by date and grouping by each region.  
   A screenshot of a computer

   AI-generated content may be incorrect.

TASK 2: Load the transformed data into Azure Databricks for further analysis

1. Launch a Databricks workspace and Launch it
2. Create a new notebook  
   A screenshot of a computer program

   AI-generated content may be incorrect.  
   A screenshot of a computer

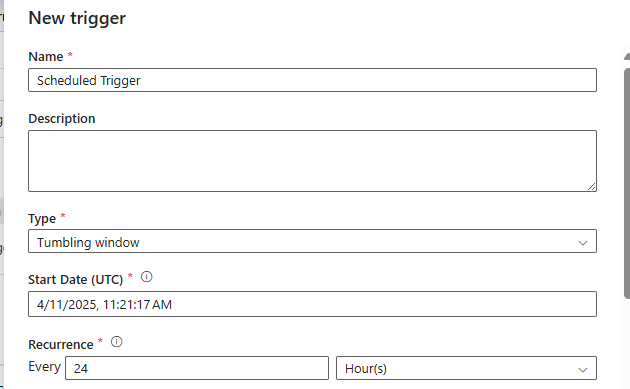
   AI-generated content may be incorrect.
3. Creating a visualization based on number of cases per day grouping A graph with a red line

   AI-generated content may be incorrect.

**Exercise 4: Orchestrating Covid-19 Data Updates with ADF**

**TASK 1: Schedule the pipeline from Exercise 2 to run daily**

We already accomplish this task while building the said pipeline earlier

1. Click on Trigger > New/Edit > New > select type as Tumbling Runs > define start dates and reccurance time (going with 24 hrs)   
   
2. Publish your pipelines and Triggers.

**TASK 2: Implement a notification system for pipeline success/failure**

1. For notification system > Go to alerts and metrics > Declare new alert rule > In target criteria configure, when a pipeline is succeeded or failed you will get a notification. > You can choose the method of notification by configuring settings  
   A screenshot of a computer

   AI-generated content may be incorrect.