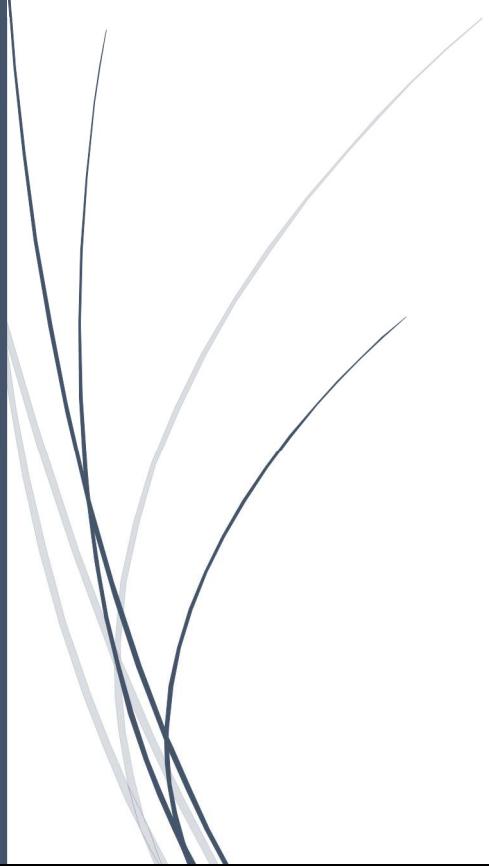




# Azure Data Factory Capstone - Covid Use Case



Name: Joshi Satya Vardan, Adina (Contractor)  
Employee ID:2320097  
Cohort Code: CSDAIA24AZ003

## Introduction:

Azure Data Factory, Synapse and Azure Data Lake Storage are some of the tools used to complete the project " Azure Data Factory Capstone - Covid Use Case " in accordance with business requirements. These Azure components help to extract data from its source (Data Lake), convert it in accordance with business needs, and load it into a destination data warehouse. The "Azure Data Factory Capstone - Covid Use Case " system was created to learn how to build a real-world data pipeline in Azure Data Factory (ADF) to analyze the covid trend across the regions using Azure cloud data services. By performing this case study, we will learn.

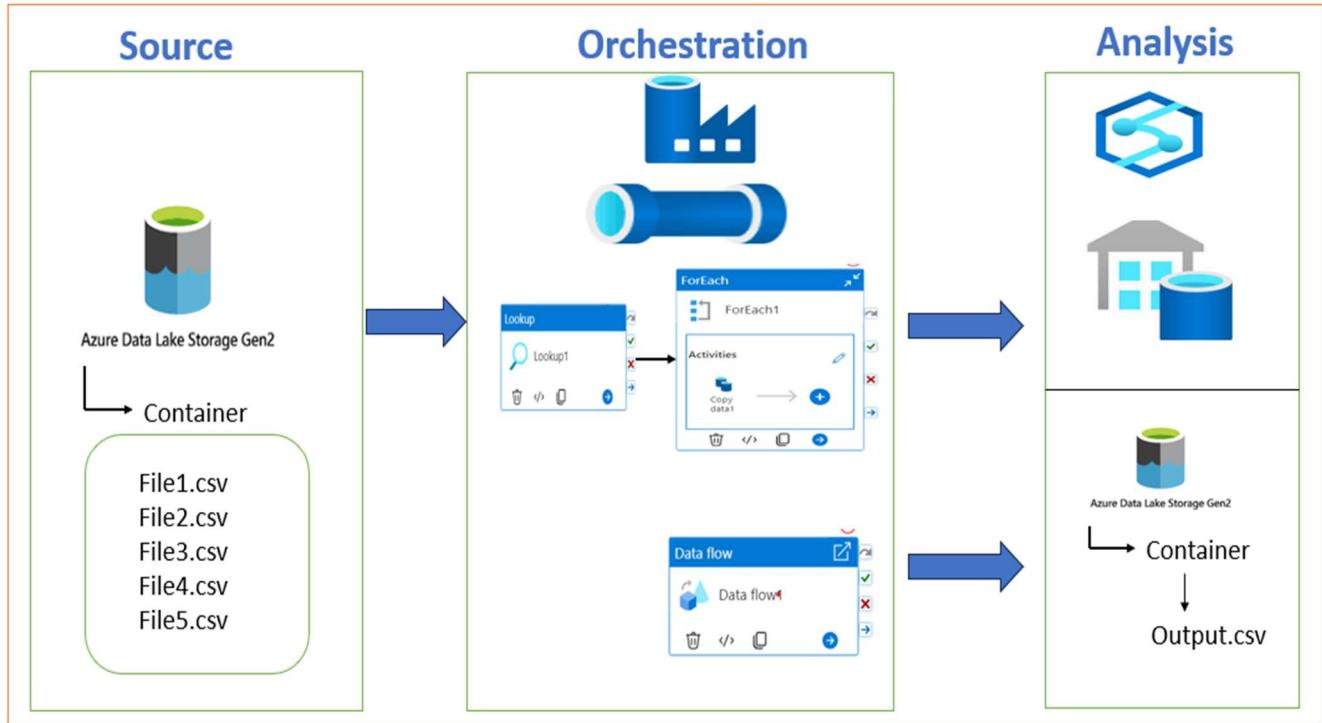
- How to ingest data from flat files into Azure Data Lake Gen2 and Azure Synapse using Azure Data Factory (ADF)
- How to transform data using Data Flows in Azure Data Factory (ADF) and load into Azure Synapse

## Objective:

The main aim of this project (Covid Use Case Exercise) is that we will be having a hands-on experience on Storage, ADF Pipeline, Mapping Dataflow, Azure Synapse along with getting to know how to ingest data from flat files into Azure Data Lake Gen2 and Azure Synapse using Azure Data Factory (ADF) and also knowing how to transform data using Data Flows in Azure Data Factory (ADF) and load into Azure data lake. This report gives a summary of the entire project making us realize and interpret the use case scenario of Azure and its applications.

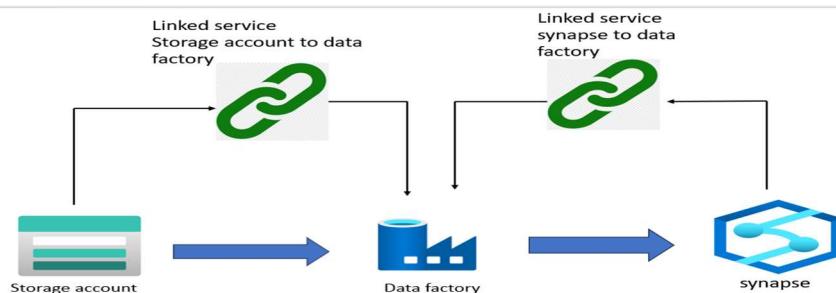
# Architecture Overview:

Azure Entra ID

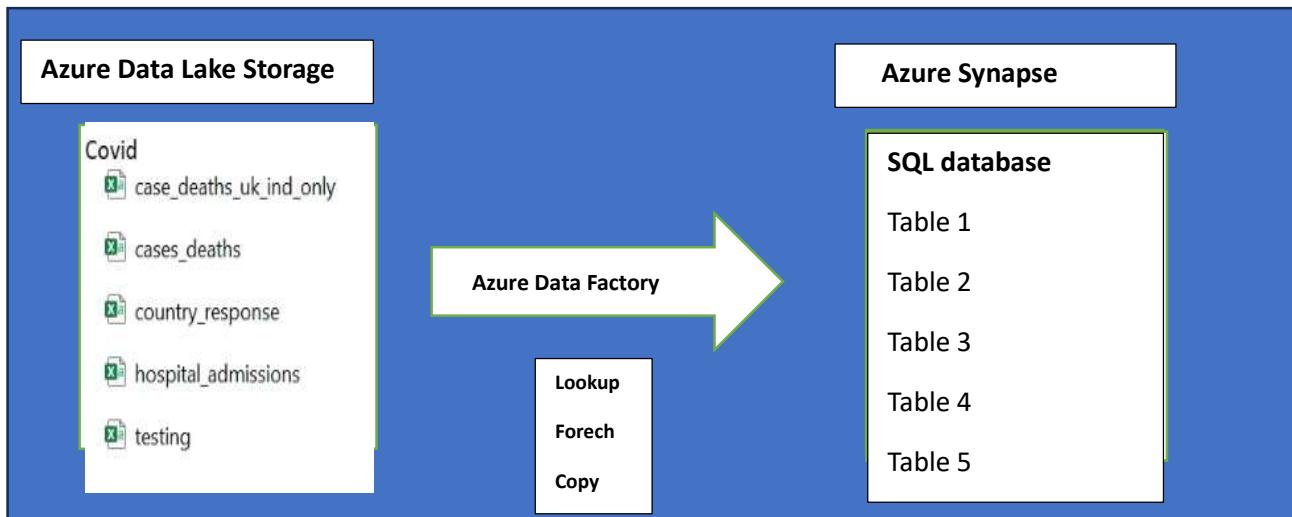


## Project Requirements:

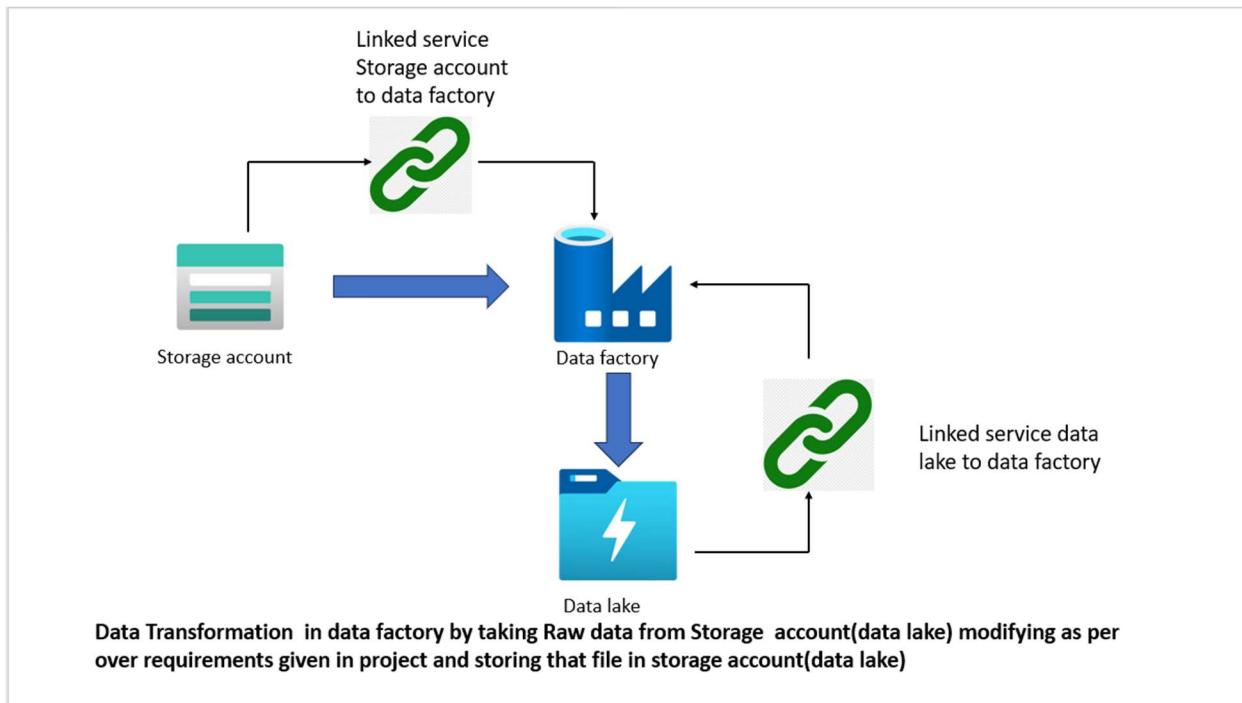
**Requirement 1:** Ingest raw flat files from data lake to synapse (data warehouse)

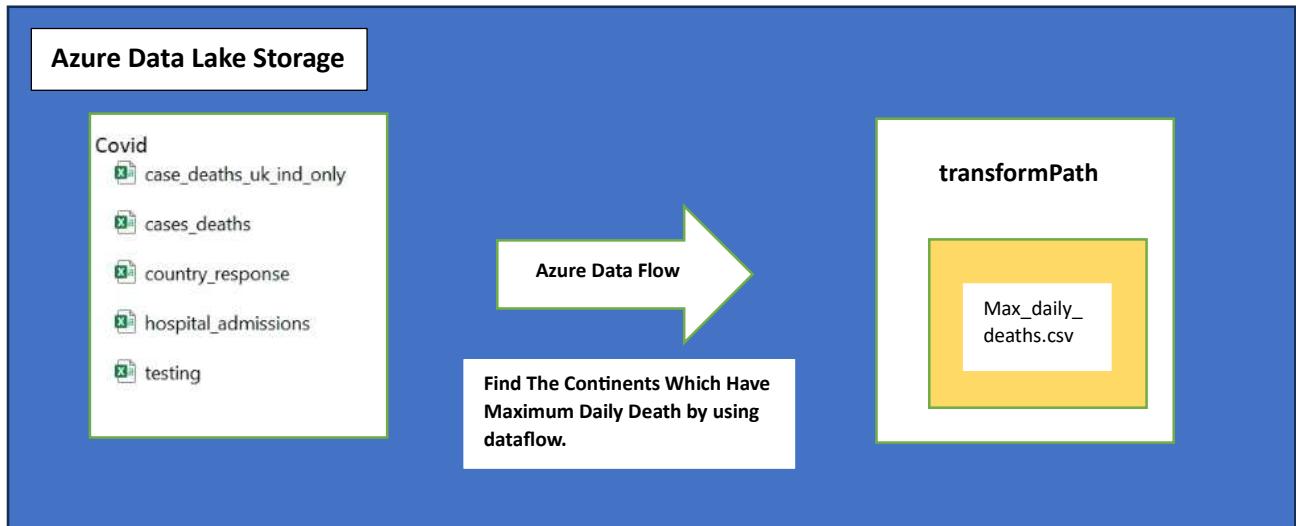


Ingesting process taken place in data factory by taking raw data from storage account(data lake) to relation database by creating tables in synapse(data warehouse)



## Requirement 2: Transform data using Data Flows in Azure Data Factory (ADF)





Microsoft Azure

Search resources, services, and docs (G+ /)

20B95A0201@srkrec.ac.in SRKR ENGINEERING COLLEGE (S...)

Azure services

- Create a resource
- Azure Synapse Analytics
- Storage accounts
- Quickstart Center
- Virtual machines
- App Services
- SQL databases
- Azure Cosmos DB
- Kubernetes services
- More services

Resources

Recent   Favorite

Name	Type	Last Viewed
covidadlsc3t3	Storage account	2 hours ago
covidsynapseworkspacec3t3	Synapse workspace	14 hours ago
covidadfc3t3	Data factory (V2)	14 hours ago
covid-rg-C3T3	Resource group	2 days ago

See all

**Step 1:** Created one Resource group(**covid-rg-C3T3**) and required resources for project like Storage account(**covidadlsc3t3**), Synapse workspace (data warehouse) (**covidsynapseworkspacec3t3**), Azure Data factory(**covidadfc3t3**).

The screenshot shows the Microsoft Azure Storage account interface for the container 'covidadlsc3t3'. The left sidebar lists various storage services like Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, and Storage browser. Under Data storage, 'Containers' is selected. The main area displays a table of containers with columns for Name, Last modified, Anonymous access level, and Lease state. Three containers are listed: '\$logs' (Last modified 04/03/2024, 11:28:57, Private, Available), 'covid' (Last modified 04/03/2024, 11:32:07, Private, Available), and 'transformation' (Last modified 06/03/2024, 09:20:51, Private, Available). A search bar at the top allows searching by prefix, and a toggle switch for 'Show deleted containers' is present.

**Step 2:** Created Container with name “**covid**” in Storage account (data lake) for holding folder which contains flat files.

The screenshot shows the Microsoft Azure Storage account interface for the container 'covid'. The left sidebar shows settings like Shared access tokens, Manage ACL, Access policy, Properties, and Metadata. The main area shows the 'Overview' tab selected. It displays authentication method as Access key (Switch to Microsoft Entra user account) and location as covid / Ingest. A search bar allows searching by blob prefix. The table below lists blobs with columns for Name, Modified, Access tier, Archive status, and Blob type. Several CSV files are listed under the 'Ingest' folder: 'case\_deaths\_uk\_ind\_only.csv', 'cases\_deaths.csv', 'country\_response.csv', 'hospital\_admissions.csv', and 'testing.csv'. All blobs are in the 'Hot (Inferred)' access tier and are Block blobs.

**Step 3:** Create Folder with name “**Ingest**” inside the Container “covid” and uploaded csv files (data sets) from personal computer which are there in zip file given in project document.

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. The left sidebar menu is open, showing options like Analytics pools, SQL pools, Data Explorer pools, External connections, Integration, Security, and others. The main content area is titled "SQL pools" and displays a table of existing pools:

Name	Type	Status	Size
Built-in	Serverless	Online	Auto
dwh_c3t3	Dedicated	Online	DW100c

**Step 4:** Create **Azure synapse resource** and one **dedicated pool** inside the azure synapse for data warehouse creation and it should be turn on.

The screenshot shows the Microsoft Azure Data Factory workspace interface. The left sidebar menu is open, showing options like General, Connections, Linked services, and others. The main content area is titled "Linked services" and displays a table of existing linked services:

Name	Type	Related	Annotations
adlslinkedservice	Azure Data Lake Storage Gen2	3	
dwLinkedService	Azure Synapse Analytics	2	

**Step 5:** Creating **two linked services** as per the project requirement in Azure data factory.

- **Storage account (data lake) to Azure data factory.**
- **Azure synapse workspace to Azure Data factory.**

The screenshot shows the Microsoft Azure Data Factory interface. In the top navigation bar, it says "Microsoft Azure | Data Factory > covidadfc3t3". The main area is titled "DS\_adlsV2Dataset". A banner at the top left announces the public preview of Microsoft Fabric. On the left, there's a sidebar with icons for Home, New, Pipelines, Datasets, and more. The main content area shows a "DelimitedText" dataset named "DS\_adlsV2Dataset". It has tabs for "Connection", "Schema", and "Parameters". Under "Connection", the "Linked service" is set to "adlslinkedservice", and the path is "covid / @dataset().FolderName / @dataset().FileName". Other settings include "Compression type" (Select...), "Column delimiter" (Comma (,), dropdown), "Row delimiter" (Default (\r\n, or \r\n)), "Encoding" (Default(UTF-8)), "Quote character" (Double quote (")), and "Escape character" (Backslash (\)).

**Step 6:** Created dataset (**DS\_adlsV2Dataset**) for fetching flat files from storage account (data lake) which are present in ingest folder inside covid container.

The screenshot shows the Microsoft Azure Synapse Analytics workspace. The top navigation bar says "Microsoft Azure | Synapse Analytics > covidsynapseworkspace". The main area is titled "Data". On the left, there's a sidebar with icons for Home, New, Pipelines, Datasets, and more. The "Data" section shows a "Workspace" tab selected. Below it, there's a "Linked" tab. A search bar says "Filter resources by name". Under "SQL database", there's one entry: "dwh\_c3t3 (SQL)". Inside "dwh\_c3t3", there are "Tables" (dbo.AllOverDeaths, dbo.config\_srcfilenames, dbo.CountryWiseResponse..., dbo.DeathsInUKAndIndia, dbo.HospitalAdmissions, dbo.Testing) and "Views". To the right, there's a placeholder icon with two cylinders and the text "Select an item" and "Use the resource explorer to select or create a new item".

**Step 7:** Created required **SQL Tables** in synapse SQL database by writing create table queries in Synapse workspace (SQL Script).

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists various components: Pipelines (1), Change Data Capture (preview) (0), Datasets (5), Data flows (1), and Power Query (0). The 'Datasets' section is expanded, showing five datasets: DS\_adlsV2Dataset, DS\_casesanddeath\_source, DS\_configDataSet, DS\_dwdataset (which is selected and highlighted in blue), and DS\_max\_daily\_deaths\_target. The main workspace displays the 'DS\_dwdataset' configuration. It shows it is an 'Azure Synapse Analytics' dataset. Under the 'Connection' tab, it is linked to 'dwLinkedService'. The 'Table' dropdown is set to 'dbo' with the sub-option '@dataset().sqlTableName' selected. A checkbox 'Enter manually' is checked. The top right corner of the interface has a 'Preview experience' toggle switch set to 'Off'.

**Step 8:** Created dataset (**DS\_dwdataset**) for inserting into SQL Tables created in synapse (data warehouse).

The screenshot shows the Microsoft Azure Synapse Analytics interface. On the left, the 'Develop' sidebar lists 'SQL scripts' (2), with 'Create\_table\_Script' and 'select\_queries\_SQL' selected. The main workspace shows a script titled 'Create\_table\_Script'. The code is as follows:

```
101
102
103 --create stored procedure
104 CREATE PROCEDURE sp_GetConfigSrcFileNames
105 As
106 Begin
107 Select * from dbo.config_srcfilenames;
108 End;
109 --Drop procedure dbo.sp_GetConfigSrcFileNames;
110
111
112
113
114
115
116
117
118
119
120
121 ---truncate table code
```

**Step 9:** Created **stored procedure(sp\_GetConfigSrcFileNames)** for fetching records from parameters table.

The screenshot shows the Microsoft Azure Data Factory interface. In the left sidebar, under 'Factory Resources', the 'Datasets' section is selected, showing five datasets: DS\_adlsV2Dataset, DS\_casesanddeath\_source, DS\_configDataSet, DS\_dwdataset, and DS\_max\_daily\_deaths\_target. The 'DS\_configDataSet' dataset is currently selected. The main pane displays the 'DS\_configDataSet' configuration, specifically the 'Connection' tab. It shows a linked service named 'dwLinkedService' and a table named 'Select...'. There are buttons for 'Test connection', 'Edit', 'New', and 'Learn more'. The top navigation bar includes a search bar, a message about the public preview of Microsoft Fabric, and a user profile for '20B95A0201@srkrec.ac.in SRKR ENGINEERING COLLEGE'.

**Step 10:** Created dataset(**DS\_configDataSet**) for fetching parameters table form synapse(data warehouse).

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar shows the same list of datasets: DS\_adlsV2Dataset, DS\_casesanddeath\_source, DS\_configDataSet, DS\_dwdataset, and DS\_max\_daily\_deaths\_target. The main pane is a resource explorer with a large icon of three cylinders and a grid, labeled 'Select an item' and 'Use the resource explorer to select or create a new item'. The top navigation bar is identical to the previous screenshot.

**Step 11:** Creating required datasets in Azure data factory.

- To fetch covid flat files from Storage account (data lake) →(DS\_adlsV2Dataset).
- To fetch parameters table from Synapse (data warehouse) →(DS\_configDataSet).
- To insert data into SQL tables in Synapse (data warehouse) →(DS\_dw\_dataset).
- To fetch cases\_deaths.csv file from storage account (data lake) →(DS\_casesanddeaths\_target).

- To insert data transformation done file by using dataflow into storage account (data lake) → **(DS\_max\_daily\_deaths\_target)**

The screenshot shows the Microsoft Azure Synapse Analytics interface. On the left, there's a navigation sidebar with options like 'Synapse live', 'Validate all', 'Publish all', 'Analytics pools', 'SQL pools' (which is selected), 'Apache Spark pools', 'Data Explorer pools (prev...)', 'External connections', 'Linked services', 'Microsoft Purview', 'Integration', 'Triggers', 'Integration runtimes', 'Security', 'Access control', 'Credentials', and 'Managed private endpoint'. The main area is titled 'SQL pools' and contains the following information:

Name	Type	Status	Size
Built-in	Serverless	Online	Auto
dwh_c3t3	Dedicated	Online	DW100c

**Step 12:** Before creating pipeline in data factory, we need **to turn on the dedicated pool**. we need to check these two to three times while moving on to creation of pipeline.

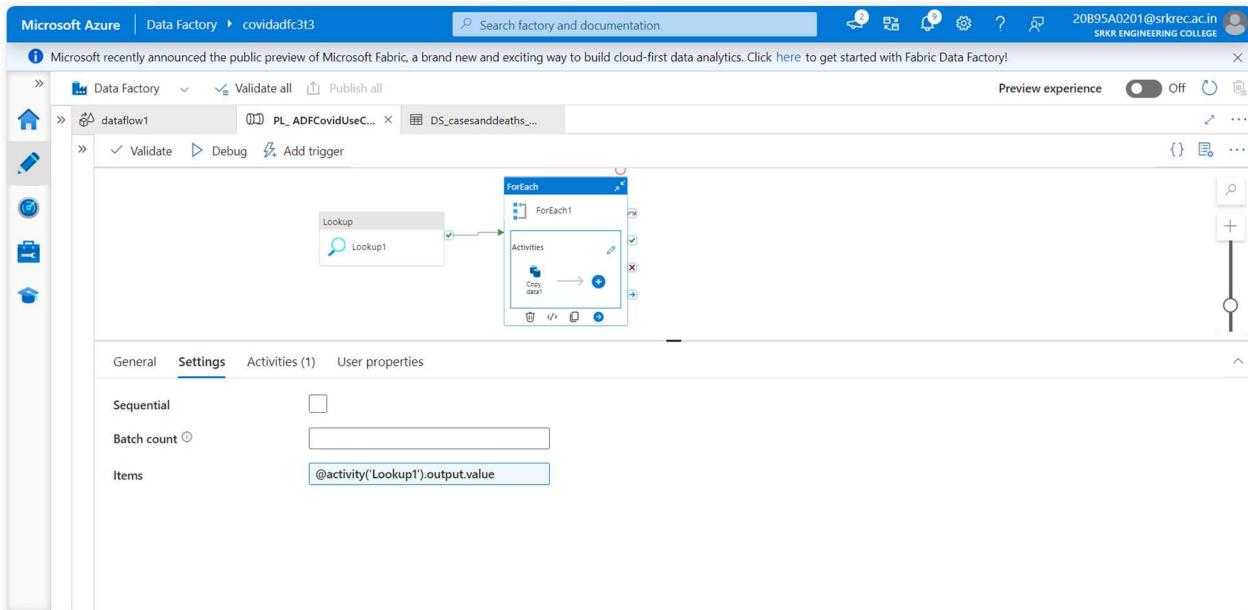
The screenshot shows the Microsoft Azure Data Factory workspace. The pipeline 'dataflow1' is displayed. The pipeline structure is as follows:

```

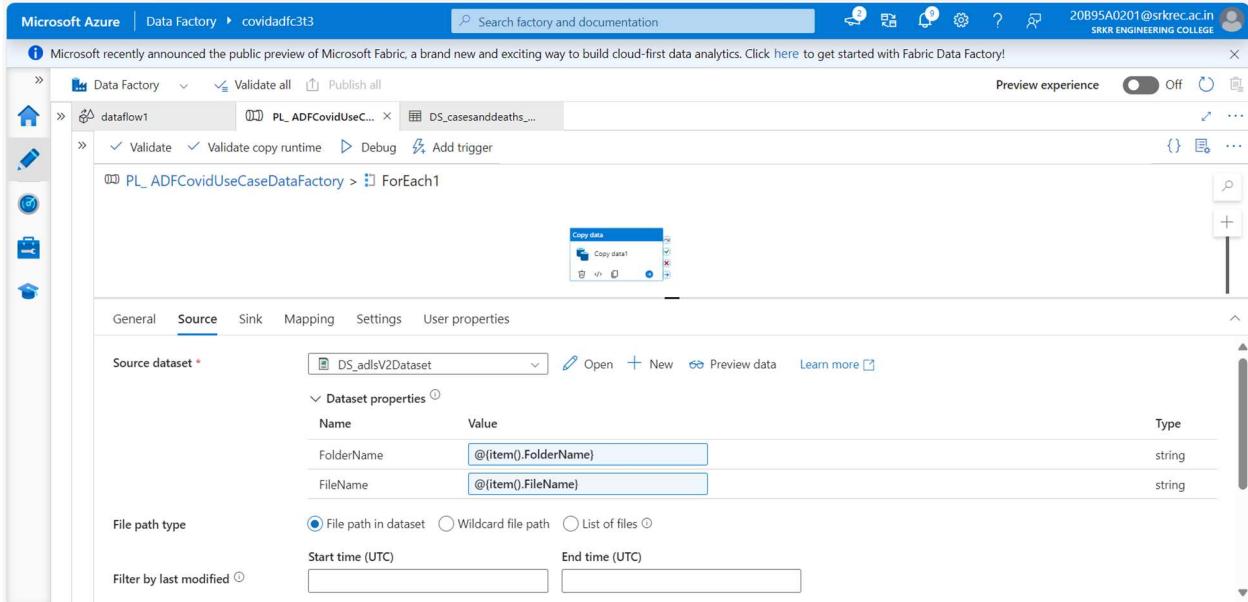
graph LR
    Lookup1[Lookup1] --> ForEach1[ForEach1]
    ForEach1 --> Activities[Activities]
    Activities --> CopyData[Copy data]
  
```

The 'Settings' tab is selected in the pipeline configuration pane. The 'Source dataset' is set to 'DS\_configDataSet'. The 'Stored procedure name' is set to '(dbo).[sp\_GetConfigSrcFileNames]'. The 'Use query' option is set to 'Stored procedure'.

**Step 13:** Create a **pipeline (PL\_ADFCovidUseCaseDataFactory)** and Drag and Drop the **Look up Activity** into pipeline workspace and set the **source dataset (DS\_configDataset)** for Lookup and choose the option **stored procedure** and given the stored procedure name created in synapse (data warehouse).



**Step 14:** Drag and Drop **For each activity** in pipeline workspace and configure the for each activity settings like **Items** with output of look up activity (**@activity(''Lookup1'').Output.Value**).



**Step 15:** Click on add activity symbol present on foreach activity and inside foreach activity add a **Copy activity** for copy data from CVS file into SQL table.

Configure settings at source side in copy activity by giving dataset (**DS\_adlsV2Dataset**) and giving folder name and file name dynamic by taking from foreach activity by **item**.

Folder Name (@ {item (). FolderName}), File Name (@ {item (). FileName})

The screenshot shows the Microsoft Azure Data Factory interface. A pipeline named 'PL\_ADFCovidUseCaseDataFactory' is selected. Within it, a 'ForEach1' activity is expanded, showing its child activities. One of these activities is a 'Copy data' activity. The 'Sink' tab is selected for this activity. In the 'Sink dataset' dropdown, 'DS\_dwdataset' is chosen. Under 'Dataset properties', there is a 'sqlTableName' field which contains the expression '@{item().sqltableName}'. Other settings like 'Copy method' (set to 'Bulk insert'), 'Bulk insert table lock' (set to 'No'), and 'Table option' (set to 'None') are also visible.

**Step 16:** Configure setting in copy activity at sink side by giving dataset (**DS\_dwdataset**) and giving sqltableName dynamically by taking from foreach activity by item→sqltableName (@ {item ()}.  
sqltableName})

The screenshot shows the Microsoft Azure Data Factory interface with the pipeline 'PL\_ADFCovidUseCaseDataFactory' selected. On the left, the 'Factory Resources' sidebar lists 'Pipelines', 'Datasets', 'Data flows', and 'Power Query'. The 'Output' tab is selected for the pipeline. It displays a table of pipeline runs. The first run listed is 'e67788ba-eb41-43d1-80d4-36c7fc432632' with a status of 'Succeeded'. The table includes columns for 'Activity name', 'Activity status', 'Activity type', 'Run start', 'Duration', and 'Integration runtime'. All activities listed in the run have a status of 'Succeeded'.

**Step 17:** After setting whole pipeline by using Lookup and Foreach activity recheck all parameters given in each configuration setting, check the dedicated pool is turn on and then turn on the debug option in pipeline. Finally, all the activities are successfully ruined.

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. In the center, there is a code editor window titled "SQL script 1" containing the following SQL query:

```

1 select * from AllOverDeaths
2 where country='India' and
3 [indicator]='confirmed cases' and
4 MONTH ([date])='3' and
5 YEAR ([date])='2020';
6
7

```

Below the code editor, the "Results" tab is selected, showing the output of the query. The output table has the following columns: country, country\_code, continent, population, indicator, daily\_count, date, rate\_14\_day, and source. The data for India is as follows:

country	country_code	continent	population	indicator	daily_count	date	rate_14_day	source
India	IND	Asia	1380004385	confirmed cases	0	2020-03-01T00:00:00	0.000000	Epidemic intelli...
India	IND	Asia	1380004385	confirmed cases	0	2020-03-02T00:00:00	0.000000	Epidemic intelli...
India	IND	Asia	1380004385	confirmed cases	2	2020-03-03T00:00:00	0.000144	Epidemic intelli...
India	IND	Asia	1380004385	confirmed cases	1	2020-03-04T00:00:00	0.000217	Epidemic intelli...
India	IND	Asia	1380004385	confirmed cases	22	2020-03-05T00:00:00	0.001811	Epidemic intelli...
India	IND	Asia	1380004385	confirmed cases	1	2020-03-06T00:00:00	0.001884	Epidemic intelli...
..								

At the bottom of the results pane, a message indicates "00:00:07 Query executed successfully."

**Step 18:** After successfully run of pipeline now we need check the data inserted into tables in data warehouse by preforming two SQL queries operation given in project documentation.

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. In the center, there is a code editor window titled "SQL script 1" containing the same SQL query as the previous screenshot. Below the code editor, the "Results" tab is selected, showing the output of the query. The output table has the same structure and data as the previous screenshot, indicating that the data has been successfully inserted into the data warehouse tables.

**Step 19:** After running first SQL Query written in SQL script and it is successfully ruined and given output as per the query.

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	country	country	contine	population	indicator	daily_cou	date	rate_14	source									
1	India	IND	Asia	1380004385	confirmed cases	0	3/1/2020	0	Epidemic intelligence, national daily data									
16670	India	IND	Asia	1380004385	confirmed cases	0	3/2/2020	0	Epidemic intelligence, national daily data									
17854	India	IND	Asia	1380004385	confirmed cases	2	3/3/2020	0.000145	Epidemic intelligence, national daily data									
19302	India	IND	Asia	1380004385	confirmed cases	1	3/4/2020	0.000217	Epidemic intelligence, national daily data									
20046	India	IND	Asia	1380004385	confirmed cases	22	3/5/2020	0.001812	Epidemic intelligence, national daily data									
21164	India	IND	Asia	1380004385	confirmed cases	1	3/6/2020	0.001884	Epidemic intelligence, national daily data									
21298	India	IND	Asia	1380004385	confirmed cases	2	3/7/2020	0.002029	Epidemic intelligence, national daily data									
22284	India	IND	Asia	1380004385	confirmed cases	3	3/8/2020	0.002246	Epidemic intelligence, national daily data									
22761	India	IND	Asia	1380004385	confirmed cases	0	3/9/2020	0.002246	Epidemic intelligence, national daily data									
24657	India	IND	Asia	1380004385	confirmed cases	10	3/10/2020	0.002971	Epidemic intelligence, national daily data									
26589	India	IND	Asia	1380004385	confirmed cases	6	3/11/2020	0.003406	Epidemic intelligence, national daily data									
28163	India	IND	Asia	1380004385	confirmed cases	23	3/12/2020	0.005072	Epidemic intelligence, national daily data									
29041	India	IND	Asia	1380004385	confirmed cases	2	3/13/2020	0.005217	Epidemic intelligence, national daily data									
29539	India	IND	Asia	1380004385	confirmed cases	8	3/14/2020	0.005797	Epidemic intelligence, national daily data									
30322	India	IND	Asia	1380004385	confirmed cases	7	3/15/2020	0.006304	Epidemic intelligence, national daily data									
30633	India	IND	Asia	1380004385	confirmed cases	3	3/16/2020	0.006522	Epidemic intelligence, national daily data									
35210	India	IND	Asia	1380004385	confirmed cases	32	3/17/2020	0.008696	Epidemic intelligence, national daily data									

**Step 20:** compare the output appear for previous SQL query written in SQL script with original CSV file in excel sheet.

```

4 MONTH ([date])='3' and
5 YEAR ([date])='2020';
6
7 select [country],[indicator],[daily_count],[date] from AlloverDeaths
8 where [indicator]='confirmed cases' AND
9 YEAR ([date])='2020'
10 group by [country],[indicator],[daily_count],[date]
11 order by [country] asc,[date] asc;
12
13
14

```

The results table shows the following data:

country	indicator	daily_count	date
India	confirmed cases	1380004385	2020-03-01
India	confirmed cases	1380004385	2020-03-02
India	confirmed cases	1380004385	2020-03-03
India	confirmed cases	1380004385	2020-03-04
India	confirmed cases	1380004385	2020-03-05
India	confirmed cases	1380004385	2020-03-06
India	confirmed cases	1380004385	2020-03-07
India	confirmed cases	1380004385	2020-03-08
India	confirmed cases	1380004385	2020-03-09
India	confirmed cases	1380004385	2020-03-10
India	confirmed cases	1380004385	2020-03-11
India	confirmed cases	1380004385	2020-03-12
India	confirmed cases	1380004385	2020-03-13
India	confirmed cases	1380004385	2020-03-14
India	confirmed cases	1380004385	2020-03-15
India	confirmed cases	1380004385	2020-03-16
India	confirmed cases	1380004385	2020-03-17

00:00:09 Query executed successfully.

**Step 21:** Wrote the second SQL Query as per the question given in the project documentation and click on run.

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. At the top, there's a navigation bar with 'Microsoft Azure' and 'Synapse Analytics'. Below it, a message says 'We use optional cookies to provide a better experience. Learn more' with 'Accept', 'Reject', and 'More options' buttons. The main area shows a list of items: 'Create\_table\_Script', 'select\_queries\_SQL', and 'SQL script 1'. 'SQL script 1' is selected and has a status message: 'Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.' It also shows 'dwh\_c3t3' as the connect to database and 'dwh\_c3t3' as the use database. Below this is a table titled 'Results' with columns: 'country', 'indicator', 'daily\_count', and 'date'. The data for Afghanistan shows daily confirmed cases from January 2nd to January 18th, 2020, all with a value of 0. A success message at the bottom says '00:00:09 Query executed successfully.'

**Step 22:** Successfully query is run and given output as per the query.

The screenshot shows a Microsoft Excel spreadsheet titled 'population'. The data is identical to the one shown in the previous screenshot, listing daily confirmed COVID-19 cases for Afghanistan from January 2nd to January 18th, 2020, with all values being 0. The Excel ribbon shows tabs like 'Home', 'Insert', 'Page Layout', etc. The status bar at the bottom indicates 'Ready 64635 of 129270 records found'.

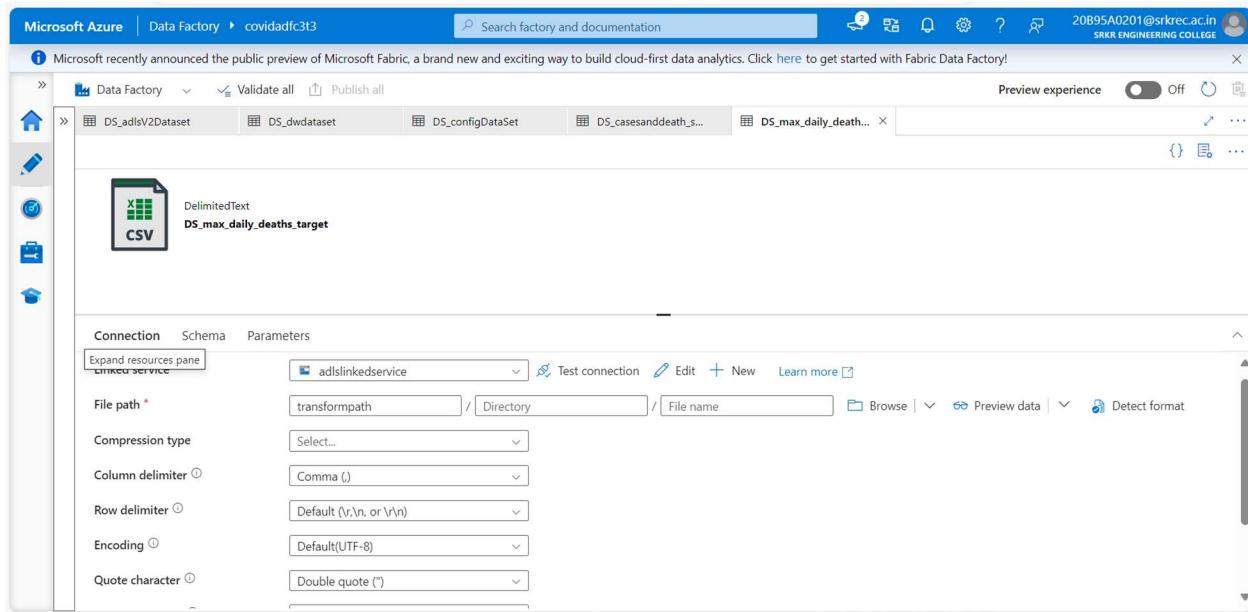
**Step 23:** Compare the above SQL query output with data which is appear from CSV file which is open in excel for reference check.

The screenshot shows the Microsoft Azure Storage account interface for 'covidadlsc3t3'. The left sidebar includes links for Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, and Storage browser. Under Data storage, 'Containers' is selected. The main area displays a table of containers with columns for Name, Last modified, Anonymous access level, and Lease state. The table contains three rows: '\$Logs' (Last modified 04/03/2024, 11:28:57, Private, Available), 'covid' (Last modified 04/03/2024, 11:32:07, Private, Available), and 'transformation' (Last modified 06/03/2024, 09:20:51, Private, Available). A search bar at the top allows searching by container prefix.

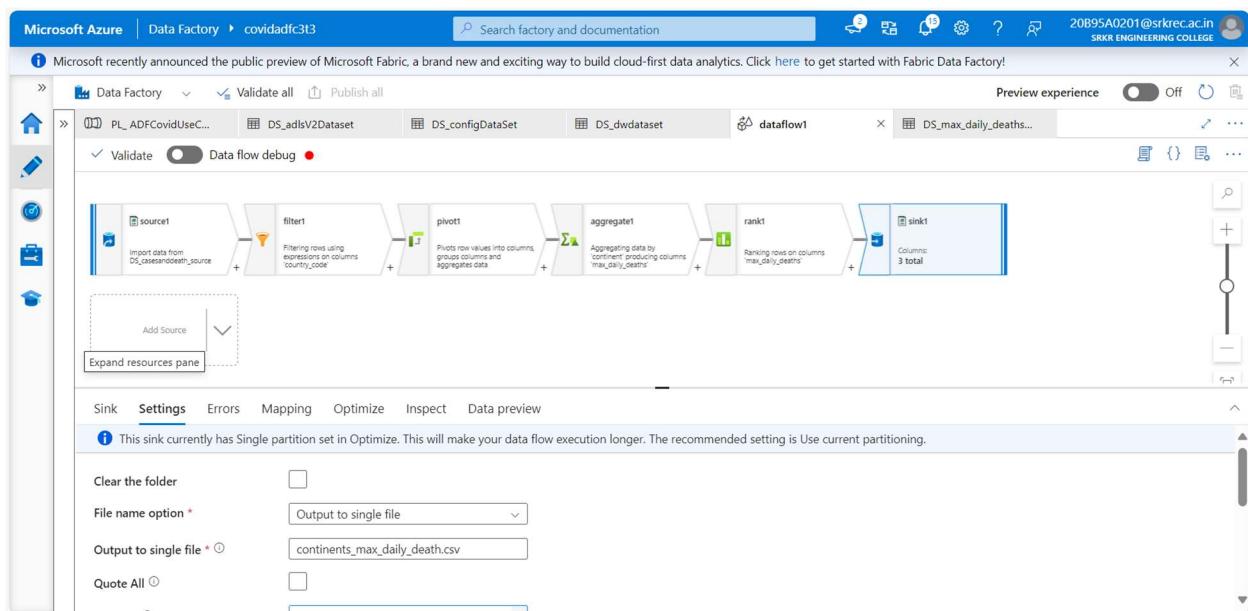
**Step 24:** Create another container with name “**transformation**” for second requirement given in project for storing transformed data file by using data flow.

The screenshot shows the Microsoft Azure Data Factory interface for 'covidadlfc3t3'. The left sidebar has icons for Data Factory, Validate all, Publish all, and a preview experience toggle (Off). The main area shows a dataset named 'DS\_casesanddeath\_source' with a CSV icon. Below it, the 'Connection' tab is selected, showing configuration for a linked service named 'adlslinkedservice' (covid) with an ingest path to 'cases\_deaths.csv'. Other tabs include Schema and Parameters. The right side of the screen shows a preview pane with data from the CSV file.

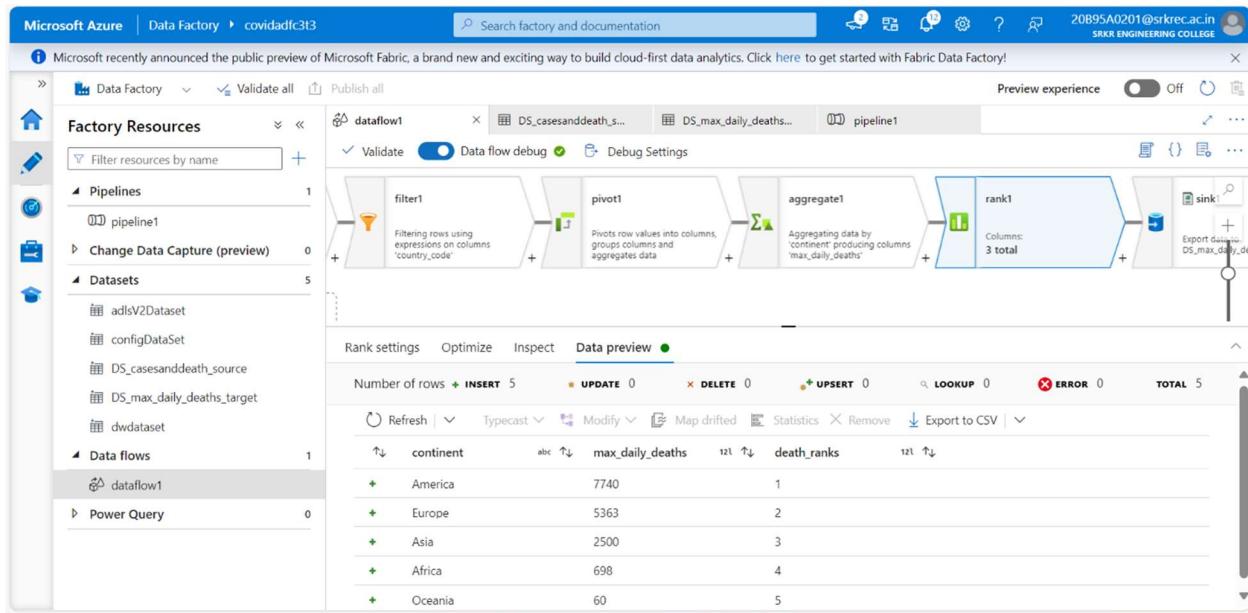
**Step 25:** Create source dataset (**DS\_casesanddeath\_source**) for dataflow by giving a file specific filename on which data transformation need to be taken place as per project requirement.



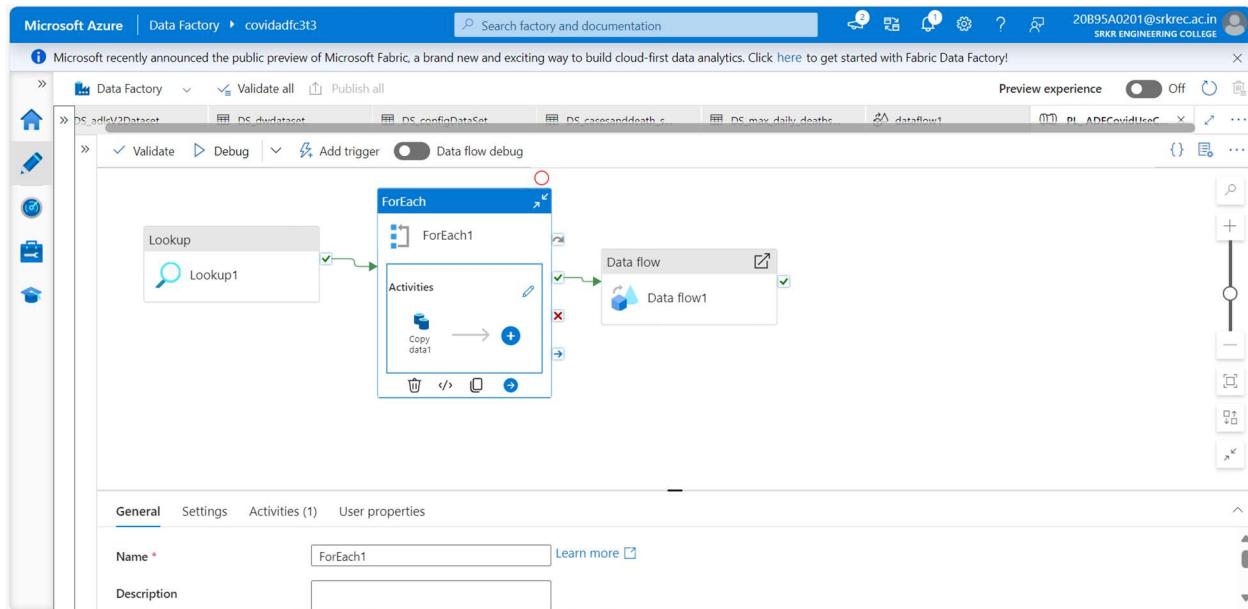
**Step 26:** Create target dataset (**DS\_max\_daily\_deaths\_target**) for dataflow to keep that data transformed file in specific place for further use.



**Step 27:** Develop dataflow by using some transformations like source, filter, pivot, aggregate, rank and sink as per the question given in project documentation.



**Step 28:** After successfully creating dataflow, I click on dataflow debug output appear as per the question.



**Step 29:** Drag and drop the data flow into the pipeline workspace and connect with before activity to run after successfully compilation of before activity.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (PL\_ADFCovidUseCaseDataFactory), 'Datasets' (DS\_adlsV2Dataset, DS\_configDataSet, DS\_casesanddeath\_source, DS\_max\_daily\_deaths\_target, DS\_dwdataset), and 'Data flows' (dataflow1). The main area displays the 'Pipeline run ID: e67788ba-eb41-43d1-80d4-36c7fc432632' status as 'Succeeded'. Below this, the 'Output' tab shows a table of activities: Copy data1 (5 rows), Data flow1 (1 row), ForEach1 (1 row), and Lookup1 (1 row). The 'Data flow1' row is highlighted with a red box.

**Step 30:** Dataflow is ruined successfully after click on debug as how in above picture.

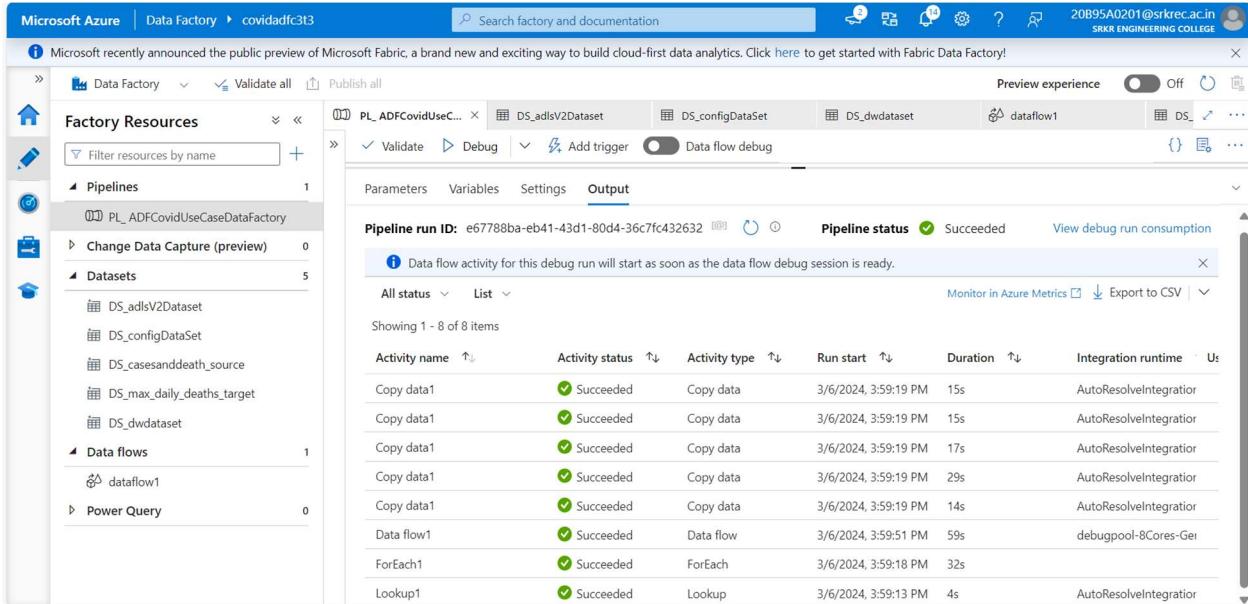
The screenshot shows the Microsoft Azure Storage Blob container for 'transformpath'. It displays the file 'continents\_max\_daily\_death.csv'. The file content is a CSV table with the following data:

continent	max_daily_deaths	death_ranks
America	7740	1
Europe	5363	2
Asia	2500	3
Africa	698	4
Oceania	60	5

**Step31:** After that we need to check file is appear in the transformation container in storage account (data lake). I successfully got that file in my container as per the question given in the project documentation.

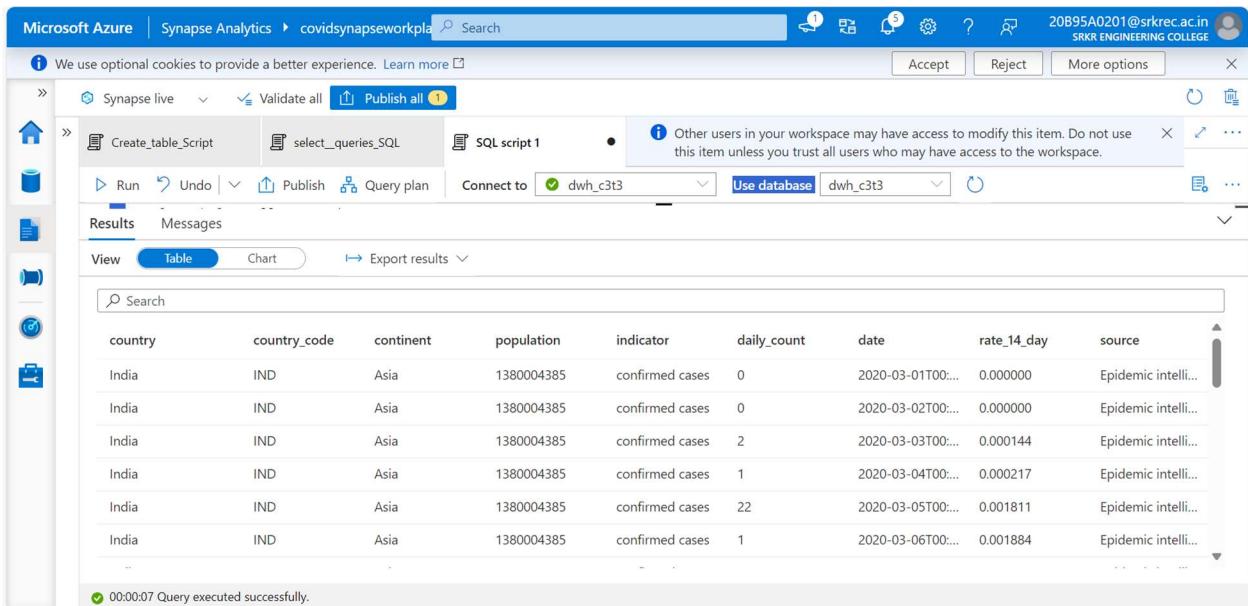
# CONCLUSION:

## Requirement 1 Output:



The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (PL\_ADFCovidUseCaseDatafactory), 'Datasets' (DS\_adlsV2Dataset, DS\_configDataSet, DS\_casesanddeath\_source, DS\_max\_daily\_deaths\_target, DS\_dwddataset), and 'Data flows' (dataflow1). The main pane displays the 'PL\_ADFCovidUseCaseDatafactory' pipeline run details. The 'Output' tab is selected, showing a table of data flow activities. The table has columns: Activity name, Activity status, Activity type, Run start, Duration, and Integration runtime. All activities show a green 'Succeeded' status. The table contains 8 items, including Copy data1, Data flow1, ForEach1, and Lookup1.

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime
Copy data1	Succeeded	Copy data	3/6/2024, 3:59:19 PM	15s	AutoResolveIntegration
Copy data1	Succeeded	Copy data	3/6/2024, 3:59:19 PM	15s	AutoResolveIntegration
Copy data1	Succeeded	Copy data	3/6/2024, 3:59:19 PM	17s	AutoResolveIntegration
Copy data1	Succeeded	Copy data	3/6/2024, 3:59:19 PM	29s	AutoResolveIntegration
Copy data1	Succeeded	Copy data	3/6/2024, 3:59:19 PM	14s	AutoResolveIntegration
Data flow1	Succeeded	Data flow	3/6/2024, 3:59:51 PM	59s	debugpool-8Cores-Gen1
ForEach1	Succeeded	ForEach	3/6/2024, 3:59:18 PM	32s	
Lookup1	Succeeded	Lookup	3/6/2024, 3:59:13 PM	4s	AutoResolveIntegration



The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. The left sidebar shows 'Synapse live' and 'Create\_table\_Script'. The main pane displays a table of data. The table has columns: country, country\_code, continent, population, indicator, daily\_count, date, rate\_14\_day, and source. The data shows multiple rows for India, with various values for daily\_count and rate\_14\_day. A message at the bottom indicates a successful query execution.

country	country_code	continent	population	indicator	daily_count	date	rate_14_day	source
India	IND	Asia	1380004385	confirmed cases	0	2020-03-01T00:00:00	0.000000	Epidemic intelli...
India	IND	Asia	1380004385	confirmed cases	0	2020-03-02T00:00:00	0.000000	Epidemic intelli...
India	IND	Asia	1380004385	confirmed cases	2	2020-03-03T00:00:00	0.000144	Epidemic intelli...
India	IND	Asia	1380004385	confirmed cases	1	2020-03-04T00:00:00	0.000217	Epidemic intelli...
India	IND	Asia	1380004385	confirmed cases	22	2020-03-05T00:00:00	0.001811	Epidemic intelli...
India	IND	Asia	1380004385	confirmed cases	1	2020-03-06T00:00:00	0.001884	Epidemic intelli...

Microsoft Azure | Synapse Analytics > covidsynapseworkpla Search

We use optional cookies to provide a better experience. Learn more

Synapse live Publish all

Create\_table\_Script select\_queries\_SQL SQL script 1

Run Undo Publish Query plan Connect to dwh\_c3t3 Use database dwh\_c3t3

Results Messages

View Table Chart Export results

Search

country	indicator	daily_count	date
Afghanistan	confirmed cases	0	2020-01-02T00:00:00.0000000
Afghanistan	confirmed cases	0	2020-01-03T00:00:00.0000000
Afghanistan	confirmed cases	0	2020-01-04T00:00:00.0000000
Afghanistan	confirmed cases	0	2020-01-05T00:00:00.0000000
Afghanistan	confirmed cases	0	2020-01-06T00:00:00.0000000
Afghanistan	confirmed cases	0	2020-01-07T00:00:00.0000000

00:00:09 Query executed successfully.

## Requirement 2 Output:

Microsoft Azure | Data Factory > covidadfc3t3 Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!

Preview experience

Data Factory Validate all Publish all

Factory Resources

- Pipelines 1
  - pipeline1
- Change Data Capture (preview) 0
- Datasets 5
  - adlsV2Dataset
  - configDataSet
  - DS\_casesanddeath\_source
  - DS\_max\_daily\_deaths\_target
  - dwdataset
- Data flows 1
  - dataflow1
- Power Query 0

dataflow1 DS\_casesanddeath\_s... DS\_max\_daily\_deaths... pipeline1

Validate

filter1 Filtering rows using expressions on columns 'country\_code' pivot1 Pivots row values into columns, groups columns and aggregates data aggregate1 Aggregating data by 'continent' producing columns max\_daily\_deaths rank1 Rank settings Columns: 3 total sink1 Export data to DS\_max\_daily\_deaths

Rank settings Optimize Inspect Data preview

Number of rows 5 0 0 0 0 0 TOTAL 5

Refresh Typecast Modify Map drifted Statistics Remove Export to CSV

continent	max_daily_deaths	death_ranks
America	7740	1
Europe	5363	2
Asia	2500	3
Africa	698	4
Oceania	60	5

Microsoft Azure Search resources, services, and docs (G+)

Home > covidadlsc3t3 | Containers > transformpath > continents\_max\_daily\_death.csv

Blob X

Save Discard Download Refresh Delete

Overview Versions Edit Generate SAS

continent	max_daily_deaths	death_ranks
America	7740	1
Europe	5363	2
Asia	2500	3
Africa	698	4
Oceania	60	5

Edit

My Account ?

Adina Joshi Satya...  
20B95A0201@srkrec.ac.in

Overview Security info

Devices Password

Organizations Settings & Privacy

My sign-ins Why can't I edit?

My Apps Give feedback

My Groups

Security info

Keep your verification methods and security info up to date.

UPDATE INFO >

Devices

## MY AZURE SUBSCRIPTION PROFILE

