



Case Study

on

Azure

Case study by,

Mouli S,
Batch – 02,
moulisankar2002@outlook.com.



Contents

1	Understandings of Case Study	3
2	Case Study Questions	4-11
2.1	Question 1	4
2.2	Question 2	6
2.3	Question 3	9



1. Understandings of Case Study

Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-COV-2 virus.

Most people infected with the virus will experience mild to moderate respiratory illness and recover without requiring special treatment. However, some will become seriously ill and require medical attention. Older people and those with underlying medical conditions like cardiovascular disease, diabetes, chronic respiratory disease, or cancer are more likely to develop serious illness. Anyone can get sick with COVID-19 and become seriously ill or die at any age.

European Centre for Disease Prevention and Control (also known as ECDC) - an agency of the European Union tracks COVID-19 cases and vaccination status worldwide.

Reference Website - <https://www.ecdc.europa.eu/en/>

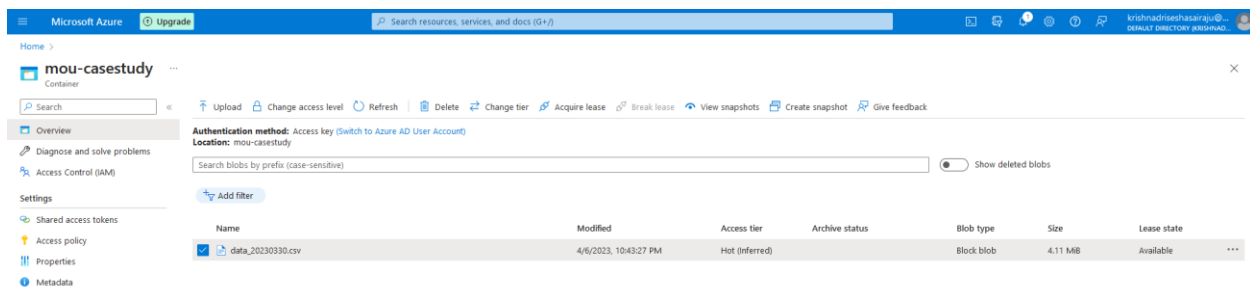


3. Case Study Questions

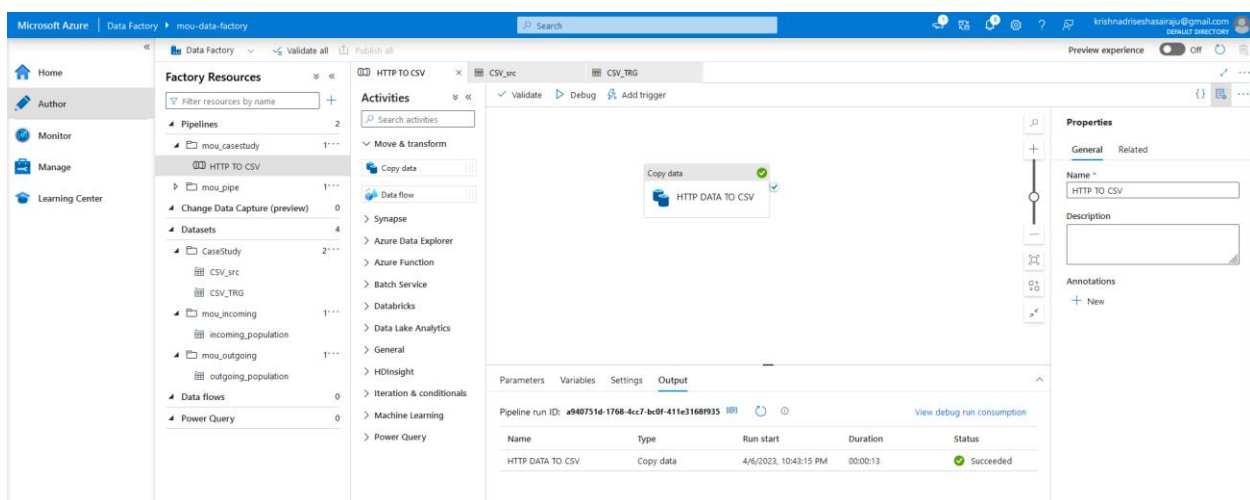
3.1 Question 1

Design Job (using ADF) to read (data.csv) ECDC website (<https://opendata.ecdc.europa.eu/covid19/casedistribution/>) and consume data and store in Blob storage in Azure and rename with by appending timestamp like data_20230330.csv.

File loaded in Blob storage:



Pipeline created for the ECDC Website to blob .csv file:





Linked services required for the pipeline and job run:

Microsoft Azure | Data Factory | mou-data-factory

Validate all | Publish all

Home | Author | Monitor | Manage | Learning Center

General | Factory settings | Connections | Linked services | Integration runtimes | Microsoft Purview | Source control | Git configuration | ARM template | Author | Triggers | Global parameters | Data flow libraries | Security | Credentials | Customer managed key | Managed private endpoints | Airflow (Preview) | Airflow

Linked services

Linked service defines the connection information to a data store or compute. [Learn more](#)

+ New

Filter by name | Annotations: Any

Showing 1 - 3 of 3 items

Name	Type	Related	Annotations
Http_ECDC	HTTP	1	
mou_blobstorage	Azure Blob Storage	2	
mou_datalake_storage	Azure Data Lake Storage Gen2	1	

Output file in the format of CSV:

Autosave | data_20230330

File | Home | Insert | Page Layout | Formulas | Data | Review | View | Developer | Help

Undo | Paste | Copy | Format Painter | Font | Alignment | Number | Conditional Formatting | Styles | Cells | Editing

dateRep	day	month	year	cases	deaths	countriesAndTerritories	geoid	countryterritoryCode	popData2019	continentExp	Cumulative_number_for_14_days_of_COVID-19_cases_per_100000
14-12-2020	14	12	2020	746	6	Afghanistan	AF	AFG	38041757	Asia	9.01377925
13-12-2020	13	12	2020	298	9	Afghanistan	AF	AFG	38041757	Asia	7.05277624
12-12-2020	12	12	2020	113	11	Afghanistan	AF	AFG	38041757	Asia	6.88876792
11-12-2020	11	12	2020	63	10	Afghanistan	AF	AFG	38041757	Asia	7.13426504
10-12-2020	10	12	2020	202	16	Afghanistan	AF	AFG	38041757	Asia	6.96865815
09-12-2020	9	12	2020	135	13	Afghanistan	AF	AFG	38041757	Asia	6.96340077
08-12-2020	8	12	2020	200	6	Afghanistan	AF	AFG	38041757	Asia	7.09483529
07-12-2020	7	12	2020	210	26	Afghanistan	AF	AFG	38041757	Asia	7.21575505
06-12-2020	6	12	2020	234	10	Afghanistan	AF	AFG	38041757	Asia	7.32616004
05-12-2020	5	12	2020	235	18	Afghanistan	AF	AFG	38041757	Asia	7.11586481
04-12-2020	4	12	2020	119	5	Afghanistan	AF	AFG	38041757	Asia	7.10797874
03-12-2020	3	12	2020	202	19	Afghanistan	AF	AFG	38041757	Asia	7.53645527
02-12-2020	2	12	2020	400	48	Afghanistan	AF	AFG	38041757	Asia	7.00545982
01-12-2020	1	12	2020	272	11	Afghanistan	AF	AFG	38041757	Asia	6.96077208
30-11-2020	30	11	2020	0	0	Afghanistan	AF	AFG	38041757	Asia	6.43603317
29-11-2020	29	11	2020	228	11	Afghanistan	AF	AFG	38041757	Asia	6.84510987
28-11-2020	28	11	2020	214	15	Afghanistan	AF	AFG	38041757	Asia	6.78464983
27-11-2020	27	11	2020	0	0	Afghanistan	AF	AFG	38041757	Asia	6.39560365
26-11-2020	26	11	2020	200	12	Afghanistan	AF	AFG	38041757	Asia	7.34193218
25-11-2020	25	11	2020	185	13	Afghanistan	AF	AFG	38041757	Asia	7.1999629
24-11-2020	24	11	2020	246	17	Afghanistan	AF	AFG	38041757	Asia	6.71367319
23-11-2020	23	11	2020	252	8	Afghanistan	AF	AFG	38041757	Asia	6.655844
22-11-2020	22	11	2020	154	12	Afghanistan	AF	AFG	38041757	Asia	6.20370926
21-11-2020	21	11	2020	232	25	Afghanistan	AF	AFG	38041757	Asia	6.13010593
20-11-2020	20	11	2020	282	5	Afghanistan	AF	AFG	38041757	Asia	5.6727138
19-11-2020	19	11	2020	0	0	Afghanistan	AF	AFG	38041757	Asia	5.03657073
18-11-2020	18	11	2020	383	13	Afghanistan	AF	AFG	38041757	Asia	5.35464227
17-11-2020	17	11	2020	65	6	Afghanistan	AF	AFG	38041757	Asia	4.57392123
16-11-2020	16	11	2020	163	9	Afghanistan	AF	AFG	38041757	Asia	4.65278194
15-11-2020	15	11	2020	205	12	Afghanistan	AF	AFG	38041757	Asia	4.57129254
14-11-2020	14	11	2020	66	10	Afghanistan	AF	AFG	38041757	Asia	4.23219148
13-11-2020	13	11	2020	360	14	Afghanistan	AF	AFG	38041757	Asia	4.47140211
12-11-2020	12	11	2020	146	4	Afghanistan	AF	AFG	38041757	Asia	3.84840269
11-11-2020	11	11	2020	0	0	Afghanistan	AF	AFG	38041757	Asia	3.4646139



3.2 Question 2

Design Job (using ADF) to read the file (data_20230330.csv) from blob storage and load into SQL Server.

List of Datasets:

1.

Microsoft Azure | Data Factory | mou-data-factory

Factory Resources

- Pipelines
- mov_casestudy
- BLOB_SQL_SERVER
- HTTP TO CSV
- pipe_3_execute_filter_dataflow
- TABLE TO CSV
- mov_pipe
- Change Data Capture (preview)
- Datasets
- AzureSqlTable1
- AzureSqlTable2
- output_dataset
- READ_BLOB**
- READCSV
- sqlwrite
- WRITE_BLOB
- CaseStudy
- CSV_src
- CSV_TRIG
- mov_incoming
- mov_outgoing
- Data flows
- year_2020
- Power Query

Connection Schema Parameters

Linked service * mou_blobstorage Test connection Edit + New Learn more

File path * mou-casestudy / data_20230330.csv Browse Preview data

Compression type None

Column delimiter Comma (,)

Row delimiter Default (\r\n, or \r\n)

Encoding Default(UTF-8)

Escape character Backslash (\)

Quote character Double quote (")

First row as header ☒

Null value

2.

Microsoft Azure | Data Factory | mou-data-factory

Factory Resources

- Pipelines
- mov_casestudy
- BLOB_SQL_SERVER
- HTTP TO CSV
- pipe_3_execute_filter_dataflow
- TABLE TO CSV
- mov_pipe
- Change Data Capture (preview)
- Datasets
- AzureSqlTable1**
- AzureSqlTable2
- output_dataset
- READ_BLOB
- READCSV
- sqlwrite
- WRITE_BLOB
- CaseStudy
- CSV_src
- CSV_TRIG
- mov_incoming
- mov_outgoing
- Data flows
- year_2020
- Power Query

Connection Schema Parameters

Linked service * AzureSqlDatabase1 Test connection Edit + New

Table dbro.csvh19_data_load Refresh Preview data

Preview data

Linked service: AzureSqlDatabase1

Object: dbro.csvh19_data_load

	dateRep	day	month	year	cases	deaths	countriesAndTerritories	geoid	countryterritoryCode
1	14/12/2020	14	12	2020	746	6	Afghanistan	AF	AFG
2	13/12/2020	13	12	2020	298	9	Afghanistan	AF	AFG
3	12/12/2020	12	12	2020	113	11	Afghanistan	AF	AFG
4	11/12/2020	11	12	2020	63	10	Afghanistan	AF	AFG
5	10/12/2020	10	12	2020	202	16	Afghanistan	AF	AFG
6	09/12/2020	9	12	2020	135	13	Afghanistan	AF	AFG
7	08/12/2020	8	12	2020	200	6	Afghanistan	AF	AFG
8	07/12/2020	7	12	2020	210	26	Afghanistan	AF	AFG
9	06/12/2020	6	12	2020	234	10	Afghanistan	AF	AFG
10	05/12/2020	5	12	2020	235	18	Afghanistan	AF	AFG



Pipe line to read csv file from blob and load to SQL Database:

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' pane lists various components including Pipelines, Datasets, and Data Flows. The 'BLOB_TO_SQL_SERVER' pipeline is selected. The 'Activities' pane shows a 'Copy data' activity. The 'Properties' pane on the right shows the activity's configuration, including the source 'BLOB_TO_SQL_SERVER' and the destination 'BLOB_TO_SQL_DATABASE'. The 'Output' tab at the bottom shows the pipeline run ID and a table of run details.

Name	Type	Run start	Duration	Status	Integration runtime
BLOB TO SQL DATABASE	Copy data	4/7/2023, 11:12:21 AM	00:00:17	Succeeded	AutoResolveIntegrationR

Loaded csv file queried in SQL Database:

The screenshot shows the Microsoft Azure Data Studio interface. The 'Query editor (preview)' pane displays a SQL query: `SELECT TOP (1000) * FROM [dbo].[covid19_data_load]`. The 'Results' pane shows the query output as a table with columns: dateRep, day, month, year, cases, deaths, and countriesAn. The 'Messages' pane shows a success message: 'Query succeeded | 8s'.

dateRep	day	month	year	cases	deaths	countriesAn
14/12/2020	14	12	2020	746	6	Afghanistan
13/12/2020	13	12	2020	298	9	Afghanistan
12/12/2020	12	12	2020	113	11	Afghanistan
11/12/2020	11	12	2020	63	10	Afghanistan
10/12/2020	10	12	2020	202	16	Afghanistan
09/12/2020	9	12	2020	135	13	Afghanistan
08/12/2020	8	12	2020	200	6	Afghanistan
07/12/2020	7	12	2020	210	26	Afghanistan



List of linked services:

Microsoft Azure | Data Factory | mou-data-factory

Search

krishnadreshashanku@gmail.com
Default directory

Validate all Publish all

Preview experience Off

General
Factory settings
Connections
Linked services
Integration runtimes
Microsoft Purview
Source control
Git configuration
ARM template
Author
Triggers
Global parameters
Data flow libraries
Security
Credentials
Customer managed key
Managed private endpoints
Airflow (Preview)
Airflow

Linked services

Linked service defines the connection information to a data store or compute. [Learn more](#)

+ New

Filter by name Annotations: Any

Showing 1 - 4 of 4 items

Name	Type	Related	Annotations	Using private endpoint
AzureSqlDatabase1	Azure SQL Database	1		---
Http_ECDC	HTTP	1		---
mou_blobstorage	Azure Blob Storage	3		---
mou_datalake_storage	Azure Data Lake Storage Gen2	1		---



3.3 Question 3

Design a Job (using ADF) to summarize data for all the countries monthly for year 2020 in a table and store a copy in blob storage too having timestamp field like (out_data_20230330.csv). Output table and file will consist of following columns:

- a. Continent (e.g... Asia)
- b. Countries (e.g... Afghanistan)
- c. Year Month (e.g., 2020-01)
- d. Cases (e.g., 0)
- e. Deaths (e.g., 0)

Pipeline for the Dataflow:

The screenshot displays the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' pane shows a list of pipelines, with 'pipe_3_execute_filter_dataflow' selected. The main canvas shows the pipeline diagram with a single activity named 'Data flow1'. The 'Properties' pane on the right shows the details for 'Data flow1', including its name and description. The 'Output' tab at the bottom shows a table with the following data:

Name	Type	Run start	Duration	Status	Integration runtime
Data flow1	Data flow	4/7/2023, 12:33:34 PM	00:00:59	Succeeded	debugpool-8Cores-Gen



Data flow creation for filter year 2020:

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' pane lists various components like Pipelines, Datasets, and Data flows. The main workspace displays a data flow named 'year_2020'. The flow consists of three activities: 'Import data from READCSV', 'filter1' (with the expression 'Filtering rows using expressions on columns year'), and 'outdatasets'. The 'Data preview' tab is active, showing a table with columns: dateRep, day, month, year, cases, deaths, country, geoid, and countryt. The table contains 10 rows of data for the year 2020.

dateRep	day	month	year	cases	deaths	country	geoid	countryt
14/12/20...	14	12	2020	746	6	Afghanistan...	AF	AFG
13/12/20...	13	12	2020	298	9	Afghanistan...	AF	AFG
12/12/20...	12	12	2020	113	11	Afghanistan...	AF	AFG
11/12/20...	11	12	2020	63	10	Afghanistan...	AF	AFG
10/12/20...	10	12	2020	202	16	Afghanistan...	AF	AFG
09/12/20...	9	12	2020	135	13	Afghanistan...	AF	AFG
08/12/20...	8	12	2020	200	6	Afghanistan...	AF	AFG
07/12/20...	7	12	2020	210	26	Afghanistan...	AF	AFG
06/12/20...	6	12	2020	234	10	Afghanistan...	AF	AFG
05/12/20...	5	12	2020	235	18	Afghanistan...	AF	AFG

Loaded data in SQL Database 'covid_data_3'

The screenshot shows the Microsoft Azure portal interface. On the left, the 'Query editor (preview)' pane is active, showing a query: `SELECT TOP (1000) * FROM [dbo].[covid_data_3]`. The 'Results' tab is active, displaying a table with columns: continents, countries, Year_Month, cases, and deaths. The table contains 10 rows of data for the year 2020.

continents	countries	Year_Month	cases	deaths
Africa	Algeria	2020-1	0	0
Africa	Algeria	2020-10	6283	230
Africa	Algeria	2020-11	24370	454
Africa	Algeria	2020-12	9881	186
Africa	Algeria	2020-2	1	0
Africa	Algeria	2020-3	583	35
Africa	Algeria	2020-4	3264	409
Africa	Algeria	2020-5	5419	202



Generated file in Blob Storage:

Microsoft Azure | Upgrade | Search resources, services, and docs (0+)

Home > moustorage | Containers >

mou-casestudy Container

Search | Upload | Change access level | Refresh | Delete | Change tier | Acquire lease | Break lease | View snapshots | Create snapshot | Give feedback

Overview

Authentication method: Access key (Switch to Azure AD User Account)

Location: mou-casestudy

Search blobs by prefix (case-sensitive) | Show deleted blobs

Settings

- Shared access tokens
- Access policy
- Properties
- Metadata

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> data_20230330.csv	4/6/2023, 10:43:27 PM	Hot (inferred)		Block blob	4.11 MiB	Available
<input type="checkbox"/> out_data_20230330.csv	4/7/2023, 12:59:59 PM	Hot (inferred)		Block blob	83.55 KiB	Available
<input type="checkbox"/> part-00000-85f382f-0123-43e8-b181-c3563fe3579-c000.csv	4/7/2023, 12:33:46 PM	Hot (inferred)		Block blob	3.94 MiB	Available

Output file in the CSV Format:

AutoSave | out_data_20230330 | Mouli Sankar

File Home Insert Page Layout Formulas Data Review View Developer Help

Undo | Paste | Copy | Format Painter | Font | Alignment | Number | Styles | Cells | Editing

contints

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	continent	countries	Year	Moncases	deaths																								
2	Africa	Algeria	2020-1	0	0																								
3	Africa	Algeria	2020-10	6283	230																								
4	Africa	Algeria	2020-11	24570	454																								
5	Africa	Algeria	2020-12	9881	186																								
6	Africa	Algeria	2020-2	1	0																								
7	Africa	Algeria	2020-3	583	35																								
8	Africa	Algeria	2020-4	3264	409																								
9	Africa	Algeria	2020-5	5419	202																								
10	Africa	Algeria	2020-6	4304	259																								
11	Africa	Algeria	2020-7	16260	295																								
12	Africa	Algeria	2020-8	14315	301																								
13	Africa	Algeria	2020-9	7222	225																								
14	Africa	Angola	2020-10	5164	96																								
15	Africa	Angola	2020-11	4834	71																								
16	Africa	Angola	2020-12	1085	25																								
17	Africa	Angola	2020-3	7	2																								
18	Africa	Angola	2020-4	20	0																								
19	Africa	Angola	2020-5	57	2																								
20	Africa	Angola	2020-6	192	7																								
21	Africa	Angola	2020-7	802	37																								
22	Africa	Angola	2020-8	1546	59																								
23	Africa	Angola	2020-9	2281	72																								
24	Africa	Benin	2020-10	303	1																								
25	Africa	Benin	2020-11	331	2																								
26	Africa	Benin	2020-12	116	1																								
27	Africa	Benin	2020-3	6	0																								
28	Africa	Benin	2020-4	63	2																								
29	Africa	Benin	2020-5	163	1																								
30	Africa	Benin	2020-6	955	16																								
31	Africa	Benin	2020-7	583	16																								
32	Africa	Benin	2020-8	375	5																								
33	Africa	Benin	2020-9	195	0																								
34	Africa	Botswana	2020-10	3470	8																								
35	Africa	Botswana	2020-11	9350	7																								

out_data_20230330 | Accessibility: Unavailable | Ready | 100%