

Assignment

This week's assignment will be interesting, where you can work on any domain of your choice. There are no restrictions with the Datasets that are used for the assignment.

You can feel free to explore and choose datasets of your choice (It is important to have multiple datasets belonging to the domain that you would like to work on for this assignment).

Question 1.

Create a Resource Group to organise all the resources you will be using for this particular assignment. (Always use proper naming conventions with meaningful names while creating the resources.)

Question 2.

Create a normal Blob storage account and upload all the datasets that you have chosen to work on.

Question 3.

Ingest the data present in the Blob storage (Source) by using the HTTP base URL of the Datasets and loading it to ADLS GEN2 (Target)

(reference steps to perform the data ingestion)

3.1 Create a Data Factory

3.2 Create a Linked Service pointing to HTTP with base URL (Source)

3.3 Create a Linked Service pointing to ADLS Gen2 (Sink)

3.4 Create a Source Dataset to fetch data from HTTP and mention the relative URL and select the source http linked service

3.5 Create a Target Dataset - Choose the format and the file path(target) where the data needs to be copied.

3.6 Test the connections and publish.

3.7 Launch Azure Data Factory Studio.

3.8 Create a pipeline and a copy activity in the Data Factory and Debug the pipeline.

Note : Depict the complete process with relevant screenshots.

Question 4.

Optimize the above activities by reducing the number of components to be created with the use of **Pipeline Parameterization** feature. Perform the complete data ingestion process using the parameterization approach. Also highlight the advantages of pipeline parameterization.

Question 5.

Create a scalable solution for the above activity by completely automating the process without any manual intervention using the **Lookup & Execute Pipeline** Activities. Make sure to organize the pipelines and datasets following the industry standards.

Process to Submit the Assignment -

You need to create a Google Document consisting of answers to all the above questions. Name the Google Document as

<your-name>_week22_assignment

Please upload your solution by filling the following form -

<https://forms.gle/AYGek3D6Jtsf1bjc7>

Top 5 answers will be selected and they will be compiled into a solution document and added to the Learning portal. Exceptional solutions will be chosen and you will receive a token of appreciation.