|  |  |  |
| --- | --- | --- |
|  |  | Azure Data Ecosystem  Automating ELT with Data Bricks, Data Lake & Data Factory to SQL DB/DW. |

# Description

Most of the data movement and data migration requires a lot of manual intervention and it requires a lot of engineering effort in developing and designing the solutions which can handle data from multiple sources and the data can be of any type . This tutorial considers automating 90% of the requirement of ELT / ETL from any sources and it can be implemented for any requirement.



### Part -2:

### Copying Data from Landing Data Lake to Output Data Lake Folder using Data Bricks.

After completing the copy activity from Blob to Landing Zone in Data Lake. The next most important thing is to utilize the power of data bricks to apply transformation and compression techniques to make it the best way to process and store the data into Output Zone in Data Lake.

1. Next step is to open the Data factory pipeline and drag and drop the Databricks notebook as below and give a meaningful name:

A screenshot of a social media post

Description automatically generated

1. If observed we have 2 important and mandatory tabs to be updated . Lets update those fields which require Data Bricks Linked Service and the Notebook Path .
2. In order to update them we need to have Linked Service for Data Bricks created for which the documentation is provided in the prerequisites folder. Create the Data Bricks LS and map it to the tab as below and also test connection to make sure the linked service is working fine.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

1. For the Notebook path, Navigate to Azure Data Bricks > Users> UserName > create a folder: bridge .
2. Upload the file which is in the scripts folder as import.

A screenshot of a social media post

Description automatically generated

1. Now update the same path into the TablestoProcess to automate the flow.
2. Open SSMS and execute the below script to update the notebook path in TablestoProcess table:

*update tablestoprocess*

*set NotebookPath = ‘<GIVE YOUR PATH>’*

*--'/Users/adffullload/full\_effex\_transformation'*

1. Now the next step is to mount the Data Bricks on top of Data Lake and we have provided the documentation in prerequisites for it .
2. Once the mounting is done , create a new notebook named ‘mount’ for defining the path of data lake as below :

A screenshot of a social media post

Description automatically generated

Code :

*configs = {"dfs.adls.oauth2.access.token.provider.type": "ClientCredential",*

*"dfs.adls.oauth2.client.id": "<APPLICATION ID>",*

*"dfs.adls.oauth2.credential": "APPLICATION AAD KEY",*

*"dfs.adls.oauth2.refresh.url": "https://login.microsoftonline.com/<DIRECTORY ID>/oauth2/token"}*

*dbutils.fs.mount(source = "adl://<NAME OF YOUR DATA LAKE>.azuredatalakestore.net/RAW", mount\_point = "/mnt/RAW", extra\_configs = configs)*

*Next Cell :*

*dbutils.fs.ls('/mnt/RAW/FILES')*

*Confirm once you are able to access the environment by above code .*

1. Now open the notebook which is shared to you and execute each code to confirm if all are working as expected:
2. Before executing passing static parameters into the widgets as below for temporary testing of code :

A screenshot of a cell phone

Description automatically generated

DatePath : TodaysDate

filename : Employee.csv

tableanem : Employee

1. Execute each cell to confirm if its working till the end . If any issues resolve them . Also make sure that the file is available in the path mentioned as below :

A screenshot of a cell phone

Description automatically generated

1. After successfully executing all the steps you should be able to see the file in the Output Zone in data lake.
2. Now we can define the Copy Activity to copy the data from Data Lake Output Zone to SQL DB using copy activity.

A screenshot of a cell phone

Description automatically generated

1. Source Connection in CopyActivity :

A screenshot of a cell phone

Description automatically generated

1. Sink Connection for Copy Activity :

A screenshot of a cell phone

Description automatically generated

1. Publish after creating the copy activity .
2. Trigger the pipeline to execute the whole pipeline .
3. Go to SQL DB and see if the data is present in the Employee\_Stage Table.
4. Once available (Assignment Extention to be done by everyone), Try to include an activity to move the data from stage table to main table.(Employee Table).