**ADF ASSIGNMENT**

1) Create a container in ADLS with the project name as sales\_view\_devtst

* Create a folder for customer, product, store, and sales - Upload the files in sequence to make it work in real-time.

2) Create an ADF Pipeline to get the latest modified files the folders of the ADLS

* Parameterize the Pipeline to work dynamically.
* Values need to be passed through parameters and it should work for all day's file

3) Create a Bronze/sales\_view/ (Bronze – container)

* Subfolders-> customer, product, store, sales, and store the raw data copied from the ADF pipeline

**Database name**: sales\_view

5) Customer file

* All the column header should be in snake case in lower case (By using UDF function dynamically works for all camel case to snake case)
* By using the "Name" column split by " " and create two columns first\_name and last\_name
* Create column domain and extract from email columns Ex: Email = "josephrice131@slingacademy.com" domain="slingacademy"
* Create a column gender where male = "M" and Female="F"
* From Joining date create two colums date and time by splitting based on " " delimiter.
* Date column should be on "yyyy-MM-dd" format.
* Create a column expenditure-status, based on spent column is spent below 200 column value is "MINIMUM" else "MAXIMUM"
* Write based on upsert [table\_name: customer] (in silver layer path is silver/sales\_view/tablename/{delta pearquet}

6) Product File

* All the column header shlould be in snake case in lower case (use same UDF Function)
* Create a column sub\_category (Use Category columns id category\_id=1, "phone"; 2, "laptop"; 3,"playstation"; 4,"e-device"
* Write based on upsert [table\_name: product](in silver layer path is silver/sales\_view/tablename/{delta pearquet}

7) Store

* Read the data make sure header shlould be in snake case in lower case (use same UDF Function)
* Create a store category columns and the value is exatracted from email Eg: "electromart" from [johndoe@electromart.com](mailto:johndoe@electromart.com)
* created\_at, updated\_at date as yyyy-MM-dd format
* Write based on upsert [table\_name: store] (in silver layer path is silver/sales\_view/tablename/{delta pearquet}

8) Sales

* Read the data make sure header shlould be in snake case in lower case (use same UDF Function)
* Write based on upsert [table\_name: customer\_sales] (in silver layer path is silver/sales\_view/tablename/{delta pearquet}

Note: all date needs to be manintained in yyyy-MM-dd format only

8) In gold layer

* using product and store table get the below data

store\_id,store\_name,location,manager\_name,product\_name,product\_code,description,category\_id,price,stock\_quantity,supplier\_id,product\_created\_at,product\_updated\_at,image\_url,weight,expiry\_date,is\_active,tax\_rate.

* Read the delta table (using UDF functions)
* Using the above data & customer\_sales and get the below data

OrderDate,Category,City,CustomerID,OrderID,Product ID,Profit,Region,Sales,Segment,ShipDate,ShipMode,latitude,longitude,store\_name,location,manager\_name,product\_name,price,stock\_quantity,image\_url

* Write based on overwrite (table\_name : StoreProductSalesAnalysis )(in gold layer path is gold/sales\_view/tablename/{delta pearquet}

So, at the end below are things you need to have.

**In ADLS**

Paths: Are mentioned above

**Bronze Layer (4 folder and raw files inside) Copied from ADF**

1)customer

2)product

3)store

4)sales

**Silver Layer (4 folders with table name inside silver)**

1)customer

2)product

3)store

4)customer\_sales

**Gold Layer (1 folder with table name inside gold)**

1)StoreProductSalesAnalysis

**In databricks**

bronze\_to\_silver - creating of silver tables

silver\_to\_gold - creating of gold tables

Question:

What is Data profiling and how it is done.

For the provided data what data profiling steps was done

Complete the above practical usecase given

Now write a usecase based on your understanding what is the architecture followed, completed understanding from end to end.