**Cohort 5.2:**

Requirements:

Source:

Python script to send Realtime records.

Target: ADLS/Refined/<file>. parquet

Problem statement:

Fetch data from python script load stream data via EventHub to data bricks

Step 1 - Data preparation & Transformation (use adb): -

* Fetch data from given python script via event hub & process streaming data via data bricks.
* Read data in data brick from Json file & load in delta table with given schema.
* Check if the file is available in the path. If it’s not available, there should be timeout after 1 minute.
* In delta table maintain below schema:

Id int,

Timestamp timestamp,

Temperature int,

Humidity int,

Insertion time timestamp

Setup email configuration to get notify.

**Deliverables**:

1. Architecture diagram of flow
2. Best Practice in Pipeline
3. Best Practice for Database
4. Best Practice for ETL
5. Target Tables