Air Quality Check – End to End Snowflake Project

**Requirements**:

* Extract the basic json data for air pollutants from **https://www.data.gov.in/.**
* Load the json file into an internal stage (**AIRQUALITY\_STG)** under Database “**DevDB**”.
* The json data is pulled into a raw table(**RAW\_AIRQUALITY**) using “$” notation.
* Then de-duplication and flattening of the json data is done and converted into individual records in a table(**clean\_qi\_dt**).
* The pollutant data in the table(**clean\_aqi\_dt**) is transposed and average pollutant data is extracted for various states and timings.
* Seggregation of Fact Tables(**Air\_Quality\_Fact**) and dimensions table(Date and Location dimensions)
* Final table (**AQI\_FINAL\_WIDE\_DT**)is created to calculate the AIR QUALITY INDEX(AQI) with various calibrations for the pollutant data (SO2,CO,PM25,NO2,NH3,O3,PM10)
* This final data of AQI is pulled into a StreamLit based dashboard and published into Streamlit community cloud.
* The files upload into internal stages are automated using data ingestion process using a “ingest.py” file which is configured in GitHub.

**Architecture Diagram**:

|  |
| --- |
| DATA SOURCE - **https://www.data.gov.in/.** |

|  |
| --- |
| API – Integrate the json data daily from datasource into internal stage **AIRQUALITY\_STG** |

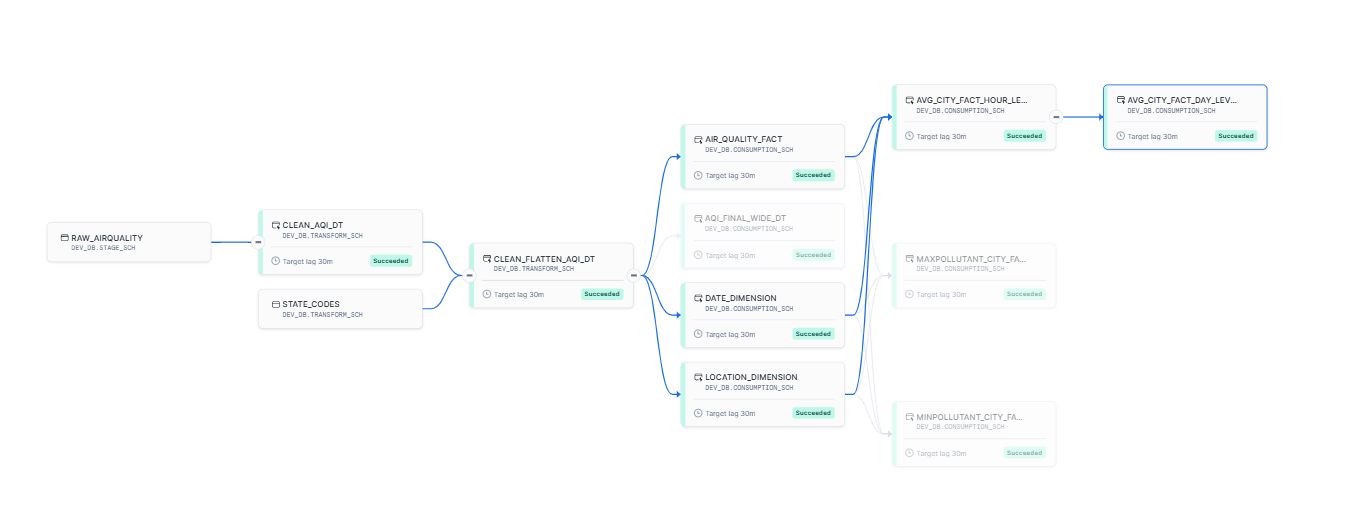
|  |  |  |  |
| --- | --- | --- | --- |
| **STAGE\_SCH SCHEMA** | **TRANSFORM\_SCH SCHEMA** | **CONSUMPTION\_SCH SCHEMA** | **PUBLISH\_SCH** |
| Load json data into internal stage(**AIRQUALITY\_STG)** | * Convert json rows to table (***raw\_airquality***) * Extract the json data into separate columns using lateral Flatten command. * Deduplication of table(Clean\_AQI\_DT) | * Transpose pollutant data (SO2,CO,PM25,NO2,NH3,O3,PM10) rows into columns. * Segregate table into FACT and DIMENSION TABLES * FACT TABLE – AIR QUALITY DATA. Calculate average pollutant data. * DIMENSION TABLE – DATE AND LOCATION TABLE .Correlating average pollutant data with respect to Date and location parameters. | AQI\_FINAL\_WIDE\_DT   * Final Air quality index is calculated with the help of average pollutants data. |
|  |  |  |  |

**Technical Entities**:

Database: **Dev\_DB**

Warehouses: **Load\_WH,ETL\_WH,Reporting\_WH**

**Airquality Index Table – Graph**

****

**Dashboard link:**

Streamlit Dashboard

<https://dashboardaqi-hon4gkxwfbi4dq35rsv5ba.streamlit.app/>

GitHub

<https://github.com/parudeploy/dashboard_AQI>

View My Portfolio

<https://www.datascienceportfol.io/paru172007>