**CAPSTONE PROJECT – WORLD EXPORTS**

**INDEX**

INFORMATICA

- Understanding the Dataset

- Constructing Staging Table for all 4 Exports

- Performing SCD1 on Dimension Tables

- Constructing a Fact Table

ANALYSIS

-DATA INCONSISTENCIES & OBSERVATIONS

- Used SQL Queries to Understand and Perform Analyses

FLASK INTEGRATION

- Performed Visualizations

**INFORMATICA –**

Constructing Staging Tables for all 4 Exports

* Combined all four export data files into a single staging table.
* Used the source qualifier with the 'currently file read' option and added ports for currentlyprocessedfilename and export type to identify file origins and export categories.
* Set the source file type to INDIRECT in the workflow manager, using a .lst file to list all input files.
* By using this method , we can combine all 4 EXPORT Flat Files into One Table.
* QUERY –

A screenshot of a computer code

Description automatically generated

ALSO We have 4 INDICATOR FILES, Each file with a single row giving us –

* + - INDICATOR\_CODE
    - INDICATOR\_NAME
    - SOURCE\_NOTE
    - SOURCE\_ORGANIZATION

In SSMS, I have created those tables using import flat files method –

A close-up of a text

Description automatically generated

Apart from these tables, There was a given table (common for all)

METADATA TABLE –

* select \* from Metadata\_Country;

Column Names :

Country\_Code

Income\_Group

Region

Special\_Notes

Table Name(Country Name)

A diagram of a company

Description automatically generated

SO,THESE ARE THE TABLES IN HAND FOR US TO PERFORM ETL .

IN INFORMATICA WE CREATED THE DIMENSION TABLES AND A FACT TABLE –

SCHEMA :

A diagram of data processing

Description automatically generated

Kindly find the attached CAPSTONE\_sql file from the folder for table schema’s .

MAPPINGS – > WORKFLOWS

STGing\_Table :

A screenshot of a computer

Description automatically generated

DIM\_COUNTRY :

A screenshot of a computer

Description automatically generated

DIM\_INDICATOR :

A screenshot of a computer

Description automatically generated

DIM\_TIME :

A screenshot of a computer

Description automatically generated

FACT\_TABLE :

A screenshot of a computer

Description automatically generated

NOTE - The WORKFLOW MANAGER AND WORKFLOW MONITOR Screenshots are attached neatly in the INFORMATICA FOLDER,For each of the above mappings.

**PMCMD COMMANDS USED –**

* pmcmd startworkflow -service INF\_DEV\_INT -domain Domain -user Administrator -password Admin@12345 -folder Dev\_Practice wf\_fact\_exports\_fin
* Similarly for other workflows too.

**SAMPLE EVENT\_WAIT :**

I performed “Event Wait” in wf\_stg\_export –

A screenshot of a computer

Description automatically generated

**DATA INCONSISTENCIES & OBSERVATIONS**

1. **Missing Country Data:**
   * CountryCode = 'INX' appears in all four exports but has "NOT CLASSIFIED" as the CountryName and no values.
   * This CountryCode is not present in the Country Metadata file.
   * As a result, a total of **252 rows** are missing after normalization.
     + Calculation: 1 Country x 63 Years x 4 Exports = 252 rows.
2. **Negative and Positive Values in Energy Imports:**
   * The Energy imports, net (% of energy use) indicator contains both negative and positive values over the years.
     + **3350 negative values**
     + **4900 positive values**
3. **Country Names with Small States:**
   * The CountryName column includes entries such as "Small States" and "Other Small States."

**Note:** Missing or empty values are treated as NULLs across all exports.

**VISUALIZATIONS AND ANALYSIS:**

The analysis queries used for visualizations are attached in the folder with name – CAPSTONE\_Analysis\_queries.

The python codes for all the visualizations have been attached with a folder name – FLASK\_INT .

Here is the HOME SCREEN –

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

THE VISUALIZATIONS –

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A close-up of a text

Description automatically generated

2.

A screenshot of a computer

Description automatically generated

A close-up of a white background

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

3.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

5.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A close-up of a text

Description automatically generated