**Documentation (IICS PROJECT)**

**By Sumeet Manoj Ghodke**

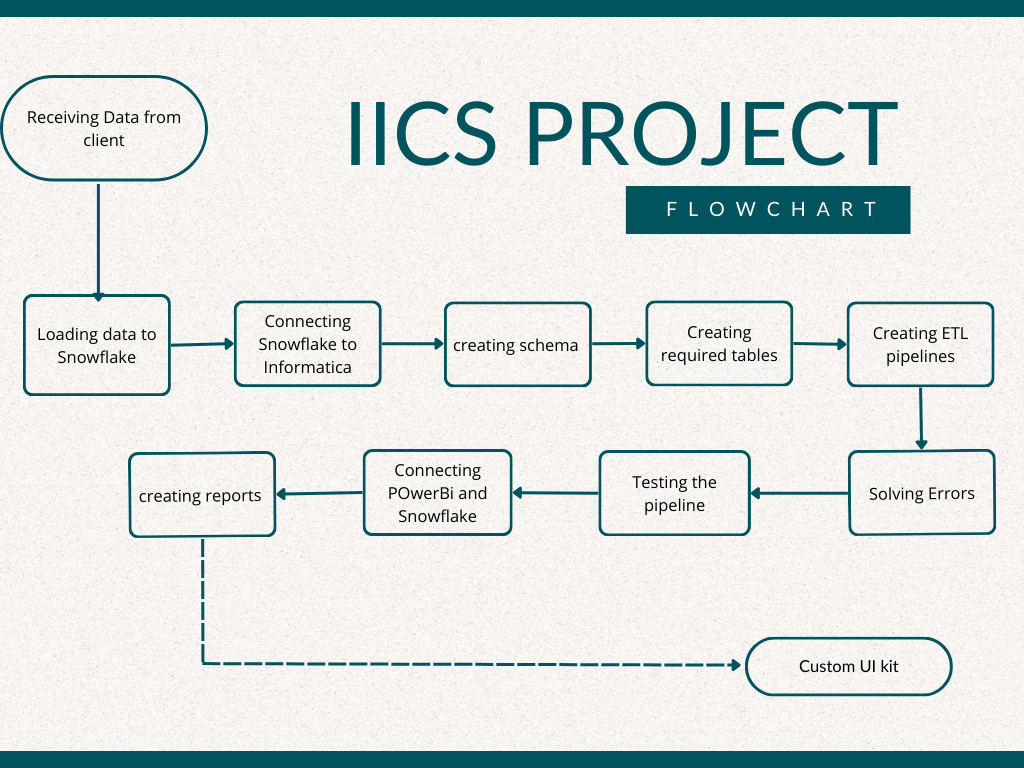
**Introduction:**  
  
In this comprehensive project, we embark on a mission to harness the **power of ETL (Extract, Transform, Load)** tools to meticulously process and reconfigure the Superstore dataset. This dataset is a rich repository of information encompassing vital aspects such as s**ales, profits, discounts, product details, geographic attributes, and more.**  
  
The primary, overarching objective of this endeavour is to orchestrate the metamorphosis of this raw and unstructured data into a meticulously structured and highly informative format. The resultant data structure will be purpose-built, catering to the specific needs of our client, and designed to serve as a robust foundation for effective decision-making.

**Key Objective**  
  
  
This objective encompasses the following key deliverables:  
  
1. A **well-structured and organized dataset** that aligns with the client's specific needs and objectives.

2. An **ETL process that is efficient, replicable, and adaptable to future data updates.**  
A data environment that is primed for analysis and further exploration.

3. A foundation for creating informative and visually engaging dashboards and reports using Power BI

**Flowchart**



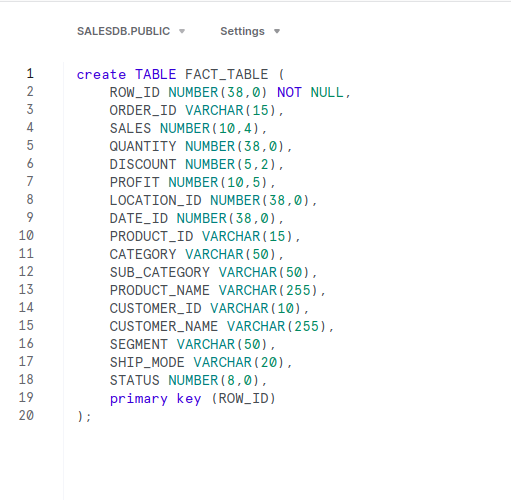
**Schema Overview**

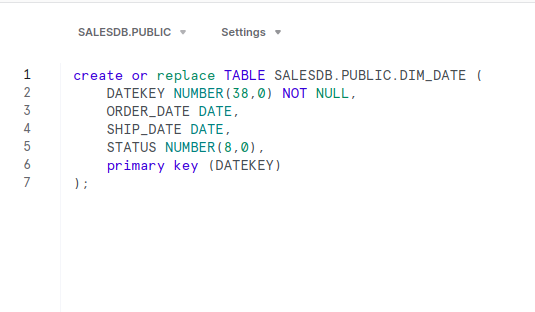
A screenshot of a computer

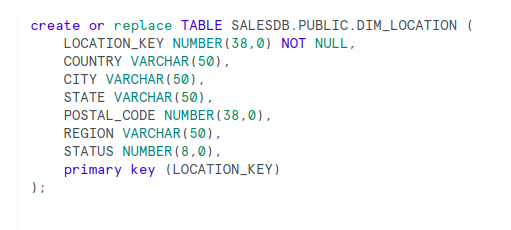
Description automatically generated

**Table Creation**

1. **Fact\_Table**
2. **Dim\_Date**
3. **Dim\_LOCATION**







**Planning**

**Day 1 - Project Kickoff and Environment Setup**  
  
Review project requirements.  
Set up Snowflake and IICS accounts.

**Days 2-3 - Data Upload and Initial Exploration**

Begin by uploading data from the local machine to Snowflake via a stage.  
Create the required schema in Snowflake based on project requirements.  
Verify data integrity and quality.

**Days 4-5 - IICS Exploration and Transformation**

Dive into Informatica Intelligent Cloud Services (IICS) to understand its capabilities.  
Experiment with various transformation options and explore data manipulation.  
Create a staging area using IICS for data processing.

**Days 6-8 - Fact and Dimension Table Setup in Snowflake**  
  
Create fact and dimension tables in Snowflake as per project requirements.  
Establish connections between Snowflake and IICS for data flow.  
Implement primary keys in dimension tables using sequence generators for uniqueness.

**Days 9-10 - Transformation and Data Load via IICS**

Apply update strategies using lookup and expression transformation in IICS.  
Implement data transformation logic as needed.  
Test the data load process from IICS to Snowflake.

**Days 11-12 - Lookup Transformation and Flags**

Apply lookup transformations to all three target tables in IICS.  
Introduce update and insert flags in the target tables.  
Ensure data consistency and accuracy.

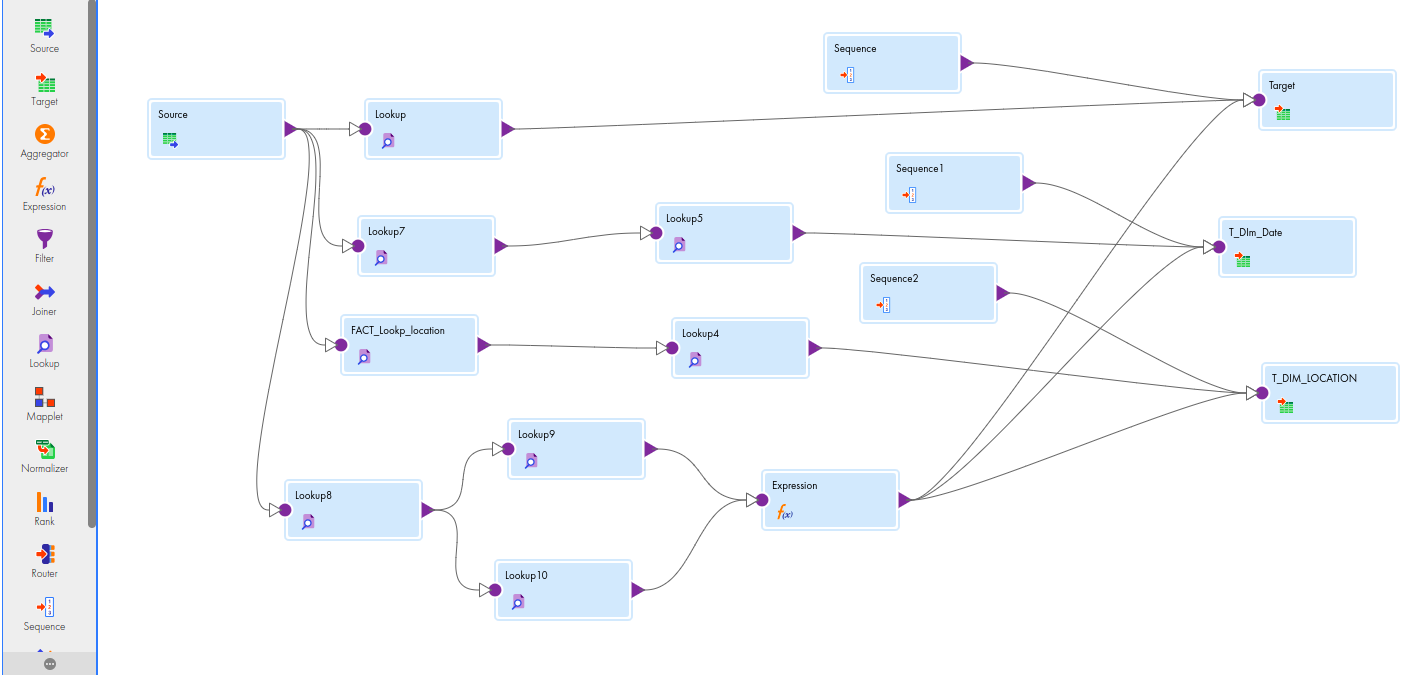
**Days 13-14 - Power BI Integration and Dashboard Creation**

Explore Power BI to understand its functionality.  
Connect Power BI to Snowflake for data retrieval.  
Solve any errors that occur during the integration process.  
Create a dashboard with relevant visualizations.

**Day 15 - Testing, Documentation, and Presentation**

Test the entire data pipeline and the created dashboards for accuracy and performance.  
Complete project documentation, ensuring it's well-structured and comprehensive.  
Prepare a PowerPoint presentation summarizing the project's key phases, challenges, and results.

**ETL Pipeline**

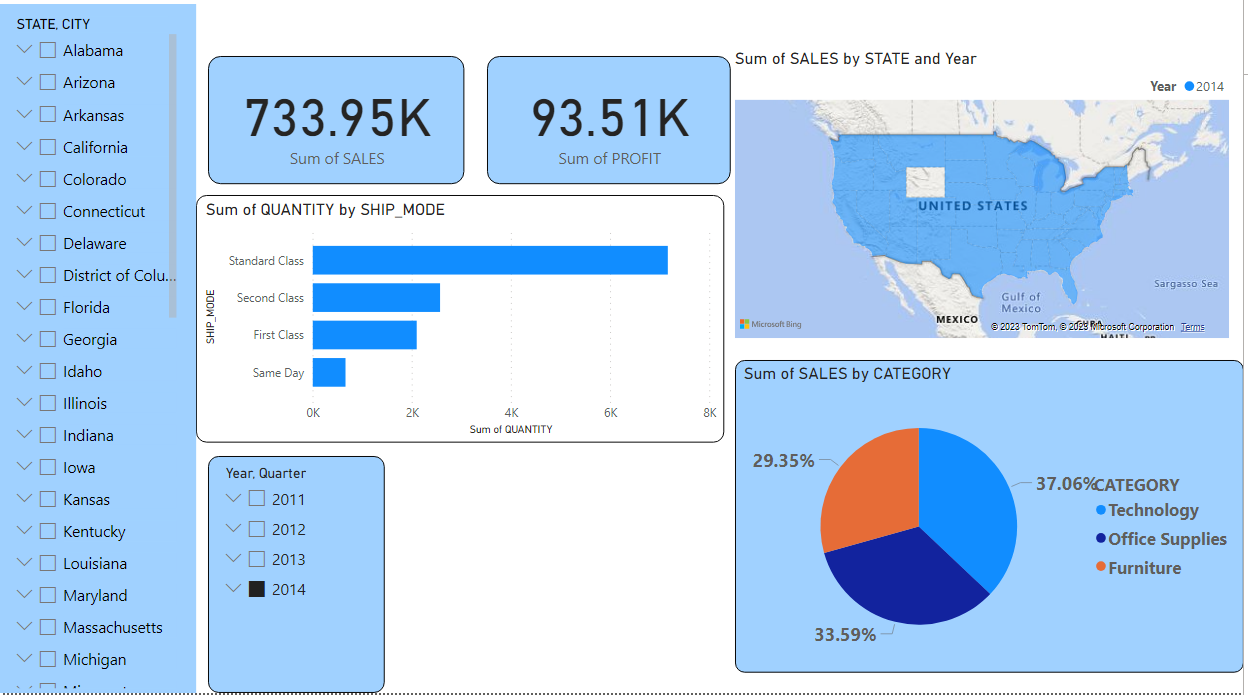


**Fig 1. ETL Pipeline**

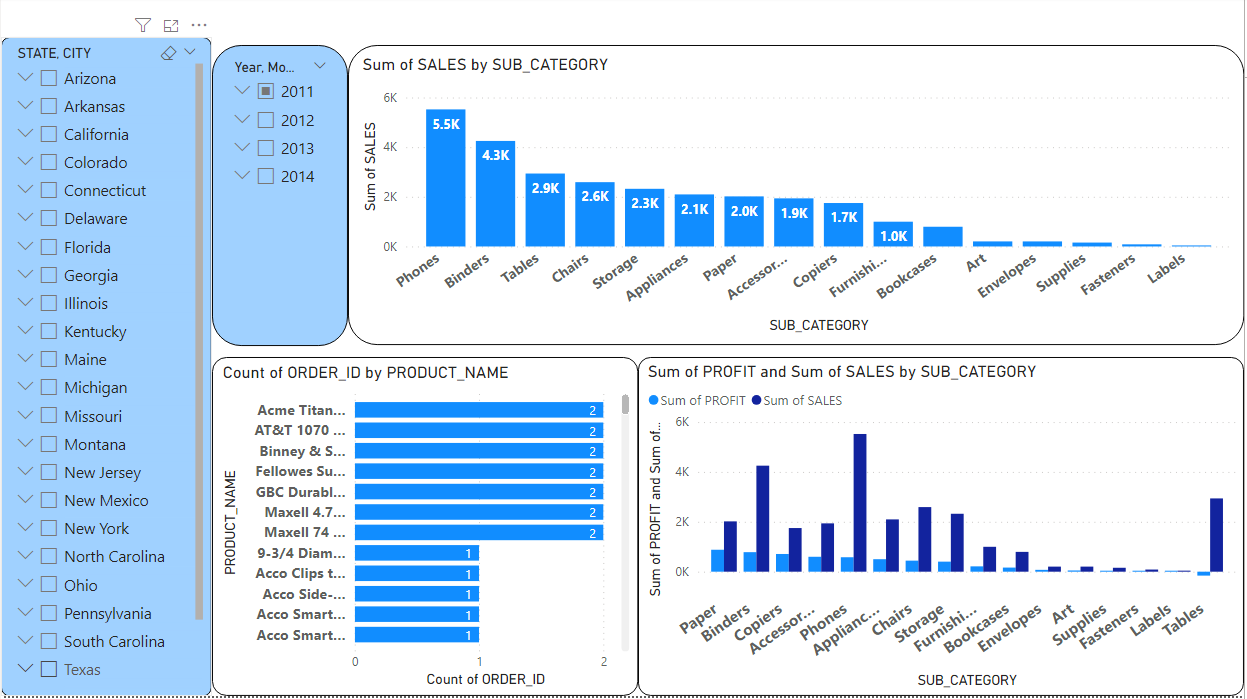
**Challenges**

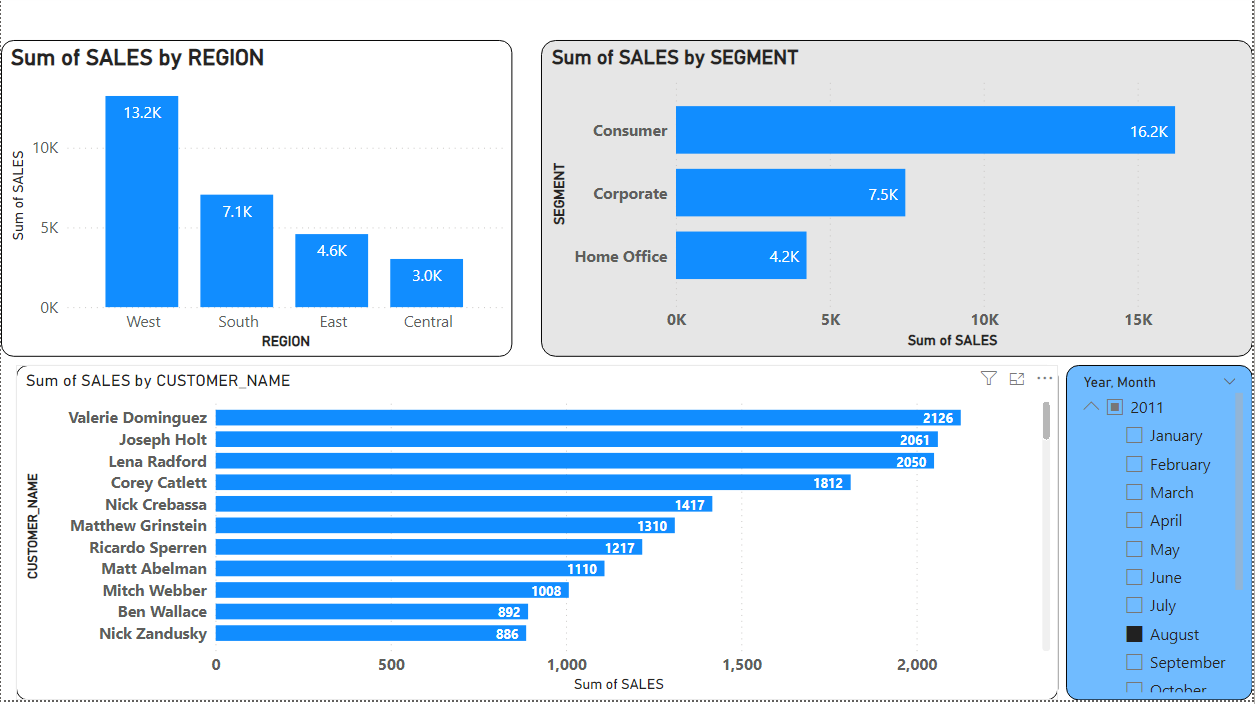
1. Begin by **uploading data from the local machine to Snowflake** via a stage.
2. The next challenge was that while **running the same pipeline two times** simultaneously an error would be occurred even though nothing was changed. (Restarting the pipeline)
3. Explore strategies to handle the **update flag in all three target tables** and how to avoid errors during updates.
4. While updating data of one row the **flag was updated to 1 for all the rows** was getting updated.

**Dashboards**



**Fig 2: Sales by Category**

  
 **Fig 3: Sum of Profit and Sales by SUB\_CATEGORY**



**Fig 4: Sales by Region, Segment**

**Summary**

This project involved the transformation of a complex dataset from a Superstore into a structured and accessible format. Key phases included dataset acquisition and setup, data transformation and integration, and finalization with Power BI integration.

**Initial Phase: Dataset Acquisition and Setup**

The project began with the receipt of the Superstore dataset. Accounts on Snowflake and IICS were set up. The dataset was uploaded into Snowflake after creating preliminary schemas.

**Data Transformation and Integration**

Data integration and transformation occurred in the second phase. A staging area was established in IICS, ensuring data quality by eliminating null values.

**Structured Data and Snowflake Schema**

Structured data was organized into fact and dimension tables within Snowflake. This involved creating mappings, adding primary keys, and implementing advanced data strategies.

**Finalization and Power BI Integration**

In the final phase, an indirect lookup table was created, and update flags were introduced to track data changes. Integration with Power BI resulted in the creation of an initial dashboard.

Overall, the project successfully transformed raw data into informative dashboards, providing valuable insights for data analysis and decision-making. It established a strong foundation for future data-driven initiatives.

**Conclusion**

This project is a testament to the capacity of data-driven insights to revolutionize decision-making. The Superstore dataset, initially a raw and disorganized resource, was meticulously transformed into a well-structured and invaluable asset. The data extraction and transformation phases ensured that the data's quality met the highest standards, making it ready for comprehensive analysis.

By embracing data-driven insights, the project provided our client with the tools and knowledge needed to make informed, data-driven decisions. This approach extends beyond data processing; it has the potential to reshape strategies in sales, product management, and resource allocation, ushering in a new era of efficiency and effectiveness.

Overcoming challenges was an integral part of the journey. Whether it was the need for a staging area or managing simultaneous pipeline runs, each challenge was met with innovation and problem-solving. These experiences strengthened the project's resilience, leaving us better equipped to tackle similar obstacles in the future.