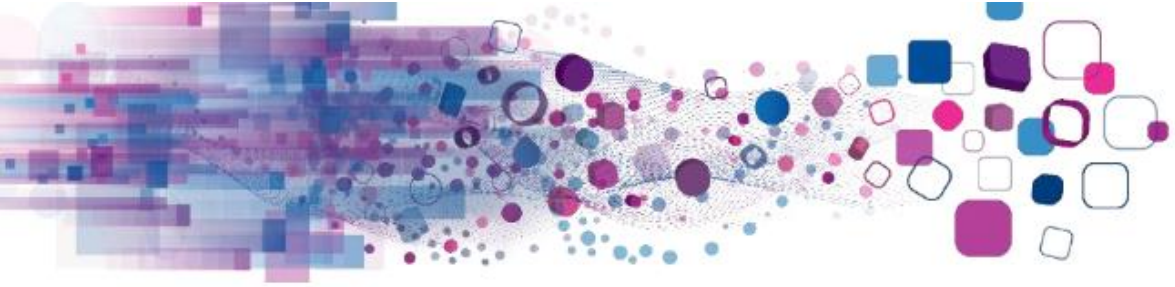
Experian Coffee Services OLAP Proposal

Prepared by: Marjay Pulumbarit

Data Engineer Applicant



The Challenge

Design an OLAP schema for Experian Coffee Services

Ingest and analyze the coffee data

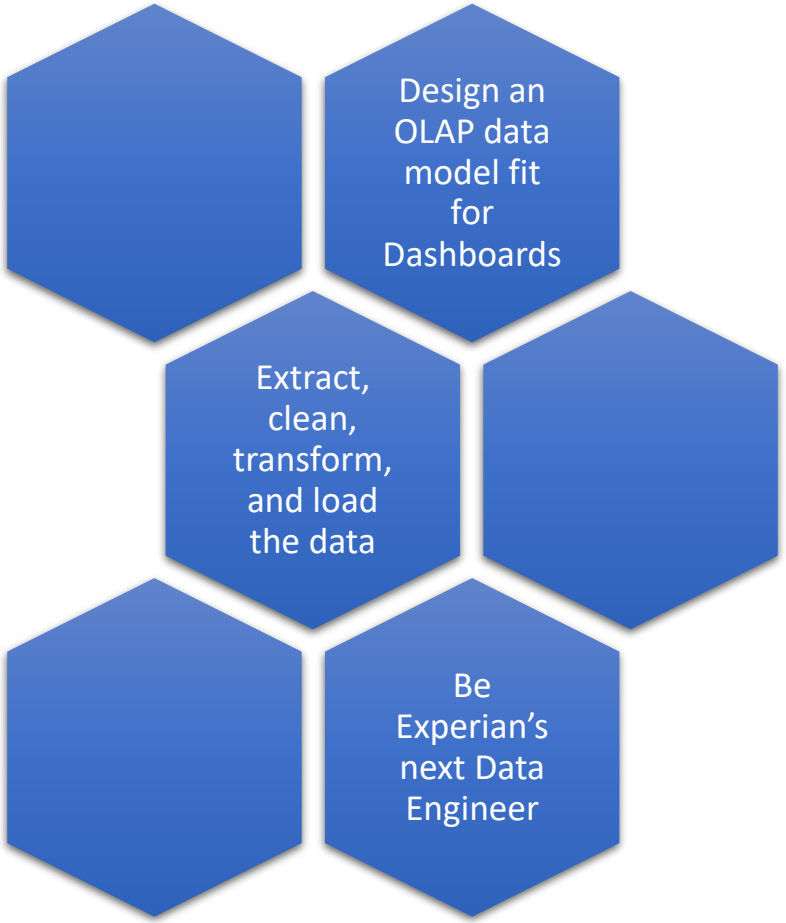
Work with Data Analysts to produce Dashboards

Help Subu find his perfect cup of coffee

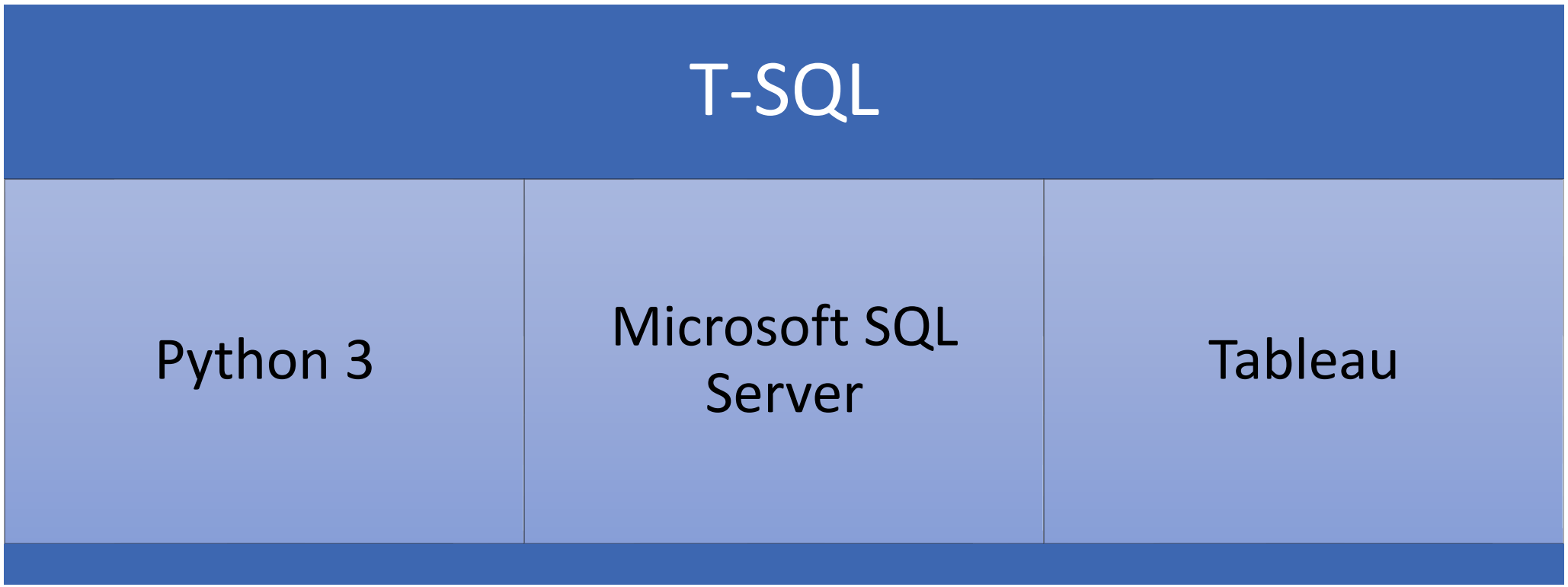
Propose the solution within 5 business days

experian™

The Goal

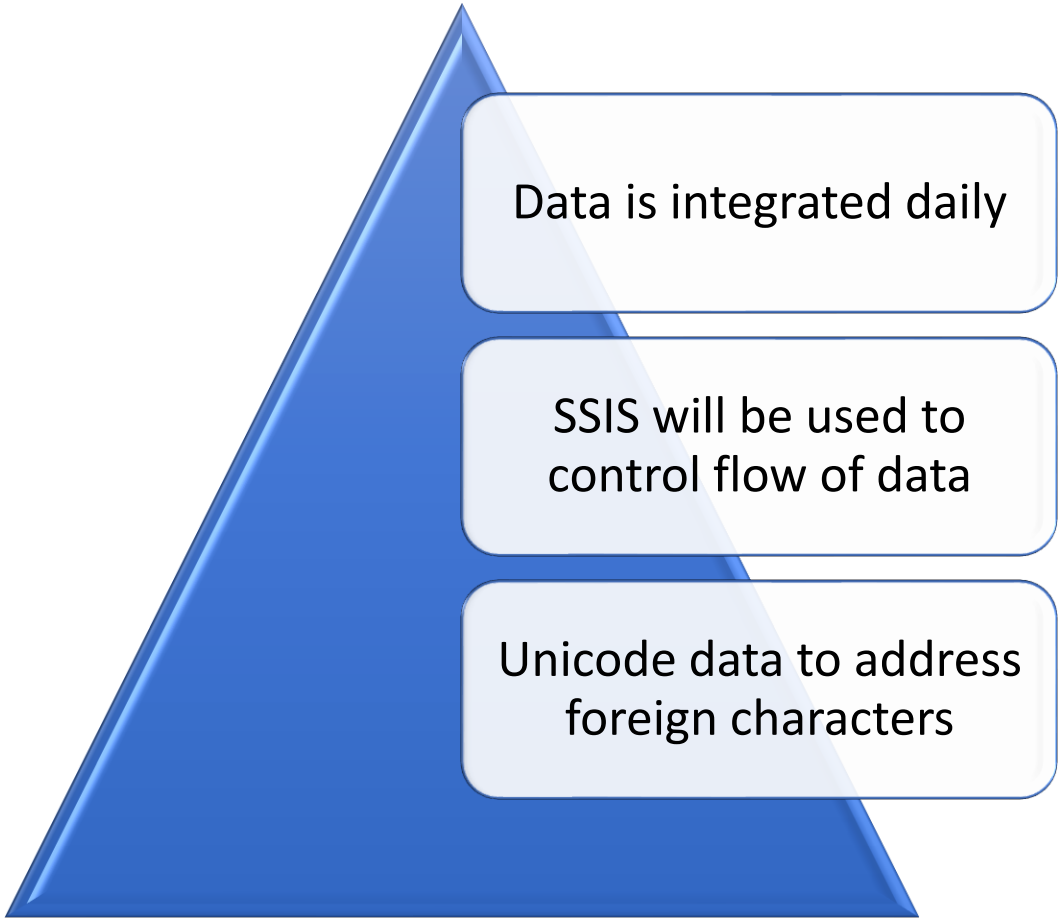

Design an OLAP data model fit for Dashboards
Extract, clean, transform, and load the data
Be Experian's next Data Engineer

Technology



experian™

Assumptions



experian™

Methodology

- Scrape data from CQI site using Python
- Design the table schema using https://dbdiagram.io/home
- Import csv file to SQL Server
- Clean and transform
- Stage then load

Scrape Latest Data

Python

beautifulsoup4

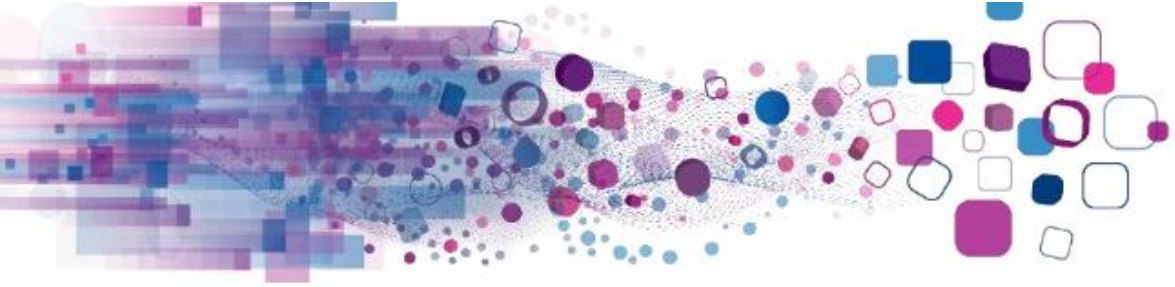
selenium

pandas

numpy

lxml

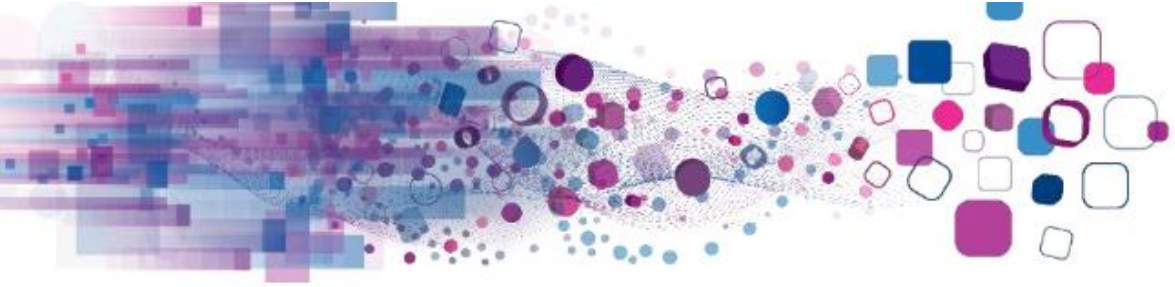
experian



Data Model

- Design Table Schema: Click here
- Store the data into 7 tables: 4 dimension tables + 3 fact tables
- Add index to keys
- Create views as source for dashboards
- Create logs table

experian



Extract

Truncate

- Ensure staging table is clean

Import flat file

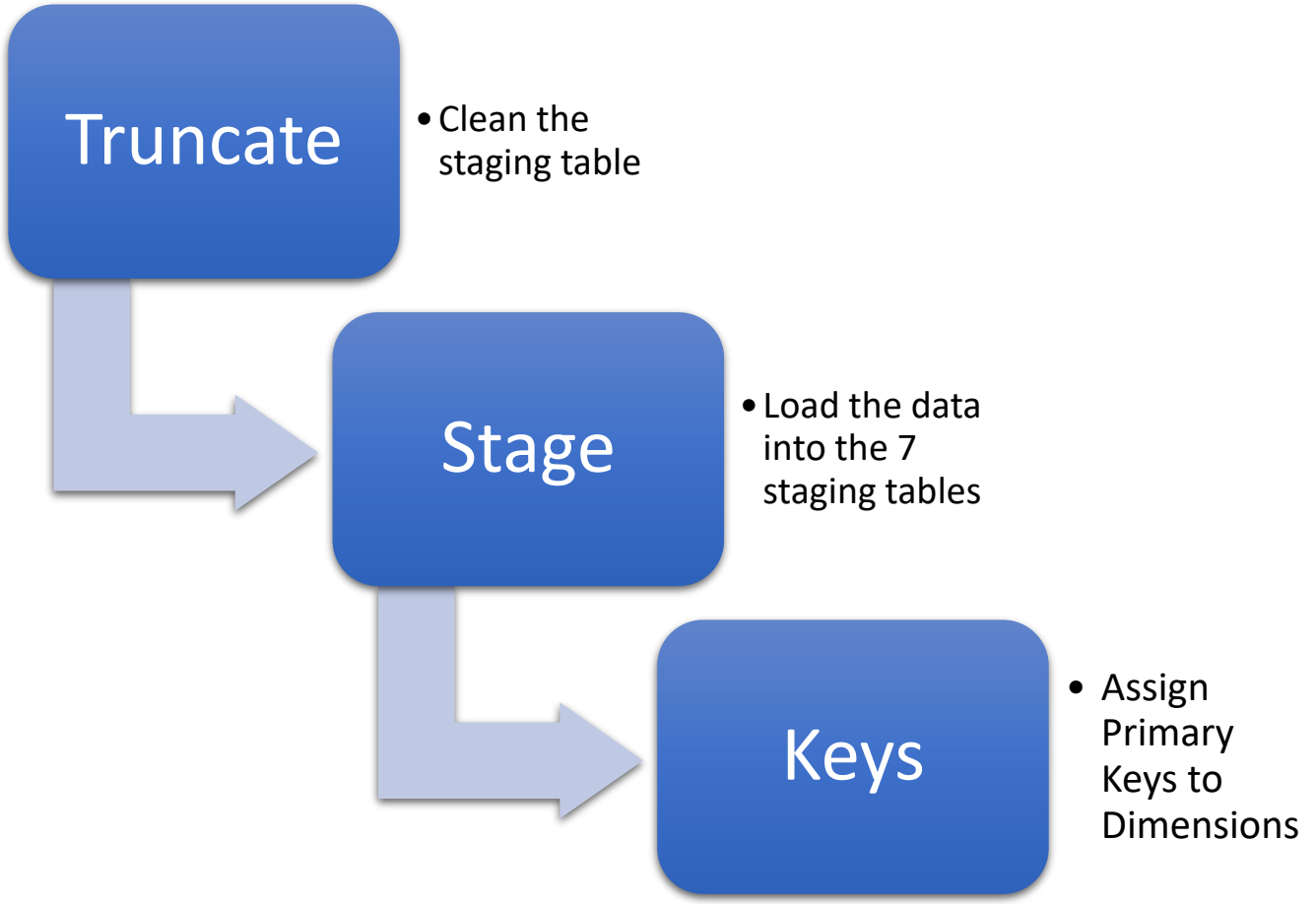
- Use SSIS, SSMS, or Bulk Insert

experian™

Clean and Transform

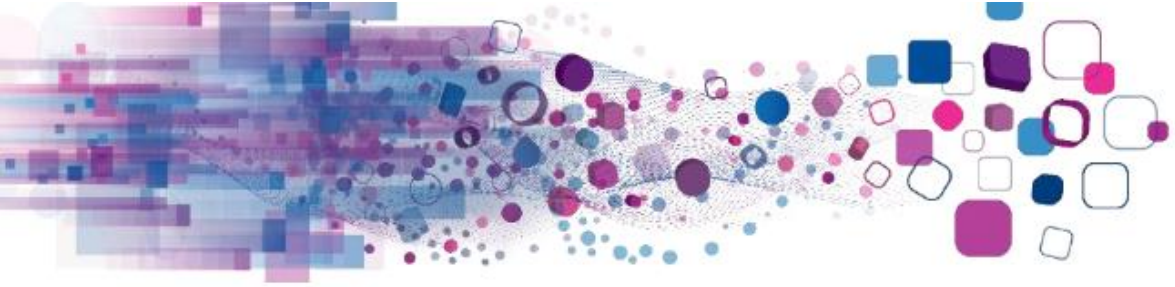
Delete Null Primary Dimensions → Cast Facts for presentation → Cast Altitude to Split values into Min and Max → Clean Defects to be used for data analysis

Stage

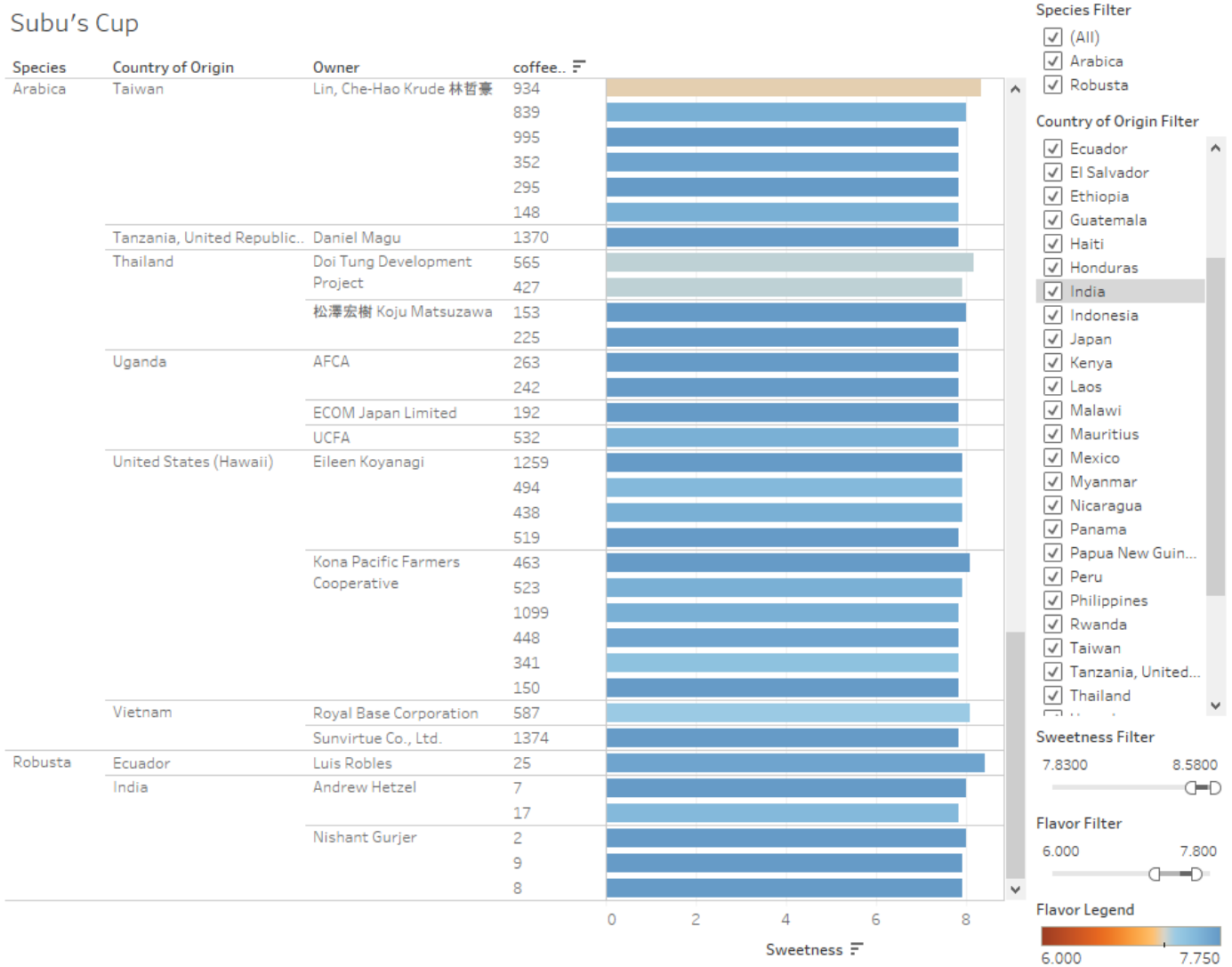

Truncate
• Clean the staging table
Stage
• Load the data into the 7 staging tables
Keys
• Assign Primary Keys to Dimensions

Load

- Update only changed data
- Insert new data
- Log the operations and add record_updated



Sample Output



About Me

Skills	Values	Master of Technology Management	Achievements
Data Pipelines	Learning	Information Security	Automated tasks saving 160 man-hours per month
Data Warehousing	Automation	Big Data and Deep Analytics	Reduce SSIS Package run time from 30 to 1 min.
SSIS Packages	Optimization	IT Project Management	Automated e-mails for quick and actionable insights
T-SQL	Attitude	Knowledge Management	Highest Customer Satisfaction Score for the Team
SQL Server	Continuous Improvement	Technology and Competitive Strategy	Recognized for automation and initiatives.

Questions?