ASSIGNMENT No: 08

Title: Data Visualization from Extraction Transformation and Loading (ETL) Process.

Problem Statement: Data Visualization from Extraction Transformation and Loading (ETL) Process.

Prerequisite:

Basics of Python

Software Requirements: Jupyter

Hardware Requirements:

PIV, 2GB RAM, 500 GB HDD

Learning Objectives:

Learn Data Visualization from Extraction Transformation and Loading (ETL) Process

Outcomes:

After completion of this assignment students are able to understand how Data Visualization is done through Extraction Transformation and Loading (ETL) Process

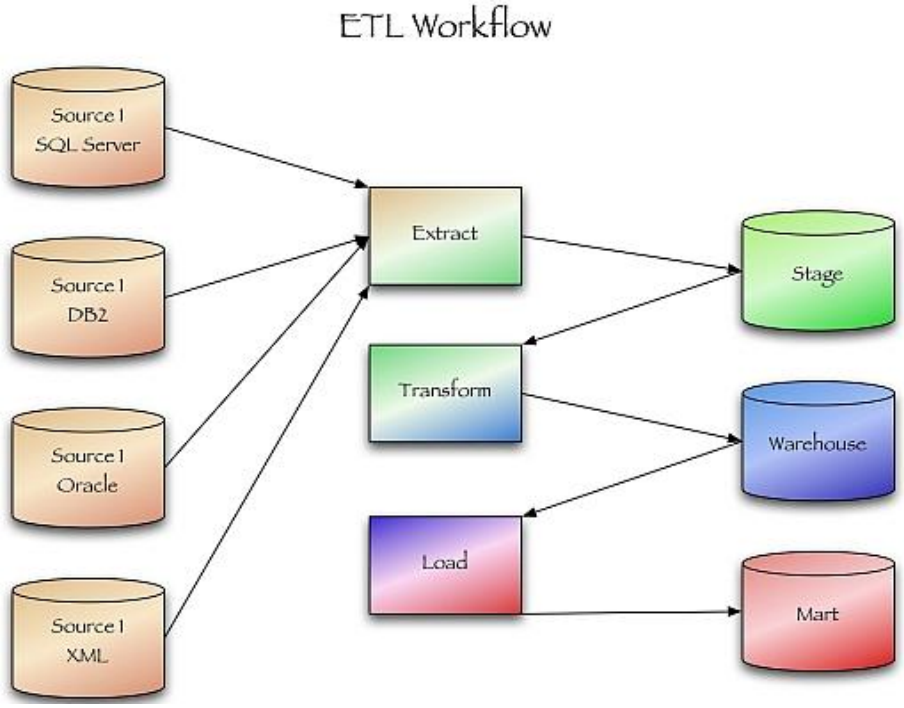
Theory:

Extract, transform, and load (ETL) are 3 data processes, followed after data collection. Extraction takes data, collected in data sources like flat files, databases (relational, hierarchical etc.), transactional datastores, semi-structured repositories (e.g. email systems or document libraries) with different structure and format, pre-validating extracted data and parsing valid data to destination (e.g. staging database)

Transformation takes extracted data and applies predefined rules and functions to it, including selection (e.g. ignore or remove NULLs), data cleansing, encoding (e.g. mapping "Male" to "M"), deriving (e.g. calculating designated value as a product of extracted value and predefined constant) , sorting, joining data from multiple sources (e.g. lookup or merge), aggregation (e.g. summary for each month), transposing (columns to rows or vice versa), splitting, disaggregation, lookups (e.g. validation through dictionaries), predefined validation etc. which may lead to rejection of some data. Transformed data can be stored into Data Warehouse (DW).

Load takes transformed data and places it into end target, in most cases called Data Mart (sometimes they called Data Warehouse too). Load can append, refresh or/and overwrite preexisting data, apply constraints

and execute appropriate triggers (to enforce data integrity, uniqueness, mandatory fields, provide log etc.) and may start additional processes, like data backup or replication.



Conclusion:- This way Data Visualization from Extraction Transformation and Loading (ETL) Process is done.