ASSIGNMENT 8

PROBLEM STATEMENT: -

Perform the Extraction Transformation and Loading (ETL) process to construct the database in the Sql server / Power BI.

OBJECTIVE:

1. To understand the concept of Extraction Transformation and Loading (ETL) process.
2. Construction of database in Power BI by performing ETL process.

THEORY:

Extraction Transformation and Loading (ETL)

- **Extraction**
  - During the first phase, data are extracted from the available internal and external sources.
  - The selection of data to be imported is based upon the data warehouse design, which in turn depends on the information needed by business intelligence analyses and decision support systems operating in a specific application domain.
  - Types of Data Extraction:
    - Full Extraction: All the data from source systems or operational systems gets extracted to staging area. (Initial Load)
    - Partial Extraction: Sometimes we get notification from the source system to update specific date. It is called as Delta load.
- **Transformation**
  - The goal of the cleaning and transformation phase is to improve the quality of the data extracted from the different sources, through the correction of inconsistencies, inaccuracies and missing values.

- During the transformation phase, additional data conversions occur in order to guarantee homogeneity and integration with respect to the different data sources.

- Examples

  - Standardizing data: Data is fetched from multiple sources so it needs to be standardized as per the target system.
  - Character set conversion: Need to transform the character sets as per the target systems.(First name and last name example)
  - Data Conversion in different formats: If in source system date in DDMMYY format and in target the date is in DDMONYYYY format then this transformation needs to be done at transformation phase.

- **Loading**

  Finally, after being extracted and transformed, data are loaded into the tables of the data warehouse to make them available to analysts and decision support applications.

There are many reasons for adopting ETL in the organization:

- It helps companies to analyze their business data for taking critical business decisions.
- ETL provides a method of moving the data from various sources into a data warehouse.
- As data sources change, the Data Warehouse will automatically update.
- Well-designed and documented ETL system is almost essential to the success of a Data Warehouse project.
- Allow verification of data transformation, aggregation and calculations rules.
- ETL helps to Migrate data into a Data Warehouse. Convert to the various formats and types to adhere to one consistent system.
- ETL in data warehouse offers deep historical context for the business.

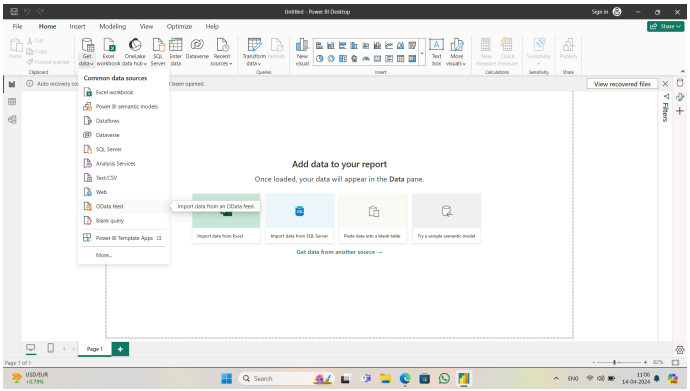
CONCLUSION:

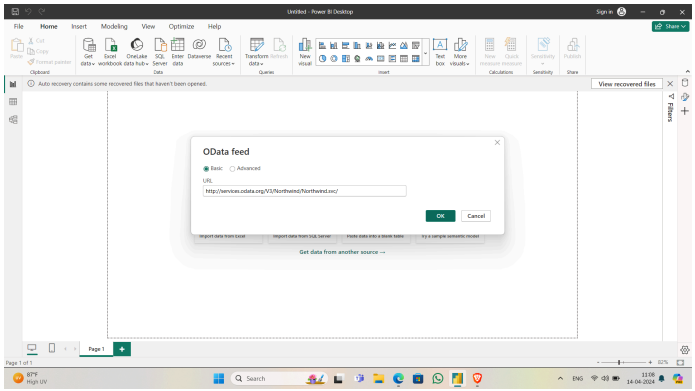
In this way we have performed Extraction Transformation and Loading (ETL) process to construct the database in Power BI.

ORAL QUESTION

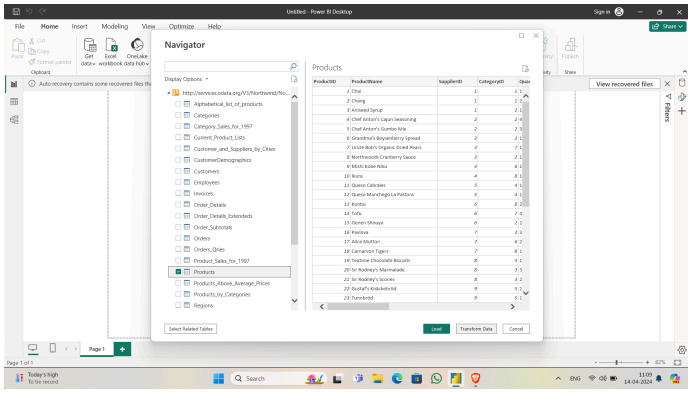
1. What is the purpose of the Extraction phase in the ETL process?
2. What factors should you consider when selecting data for extraction?
3. What is meant by the Transformation phase in the ETL process?
4. Explain the concept of loading data in the context of the ETL process.

Extracting data from source:

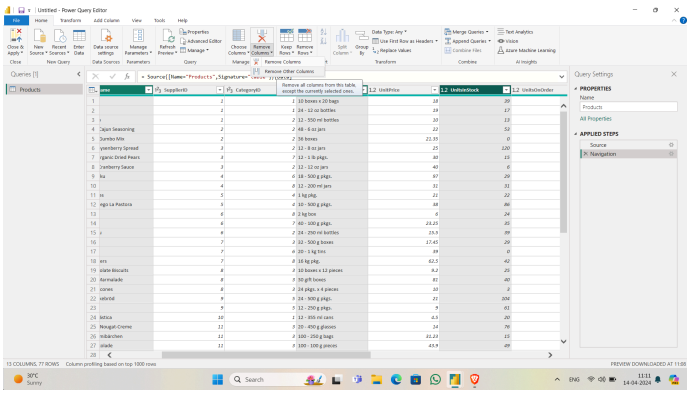


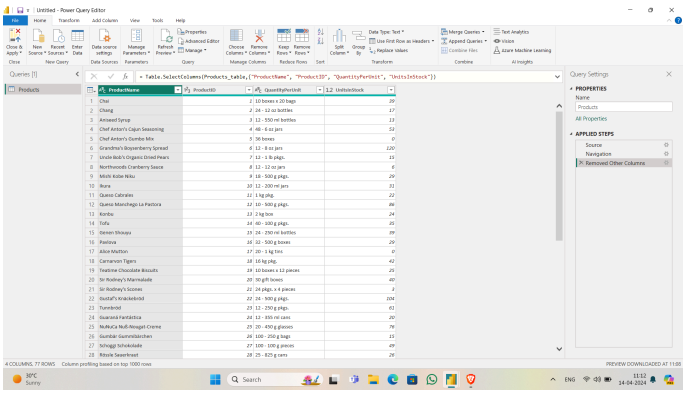


Transforming data:

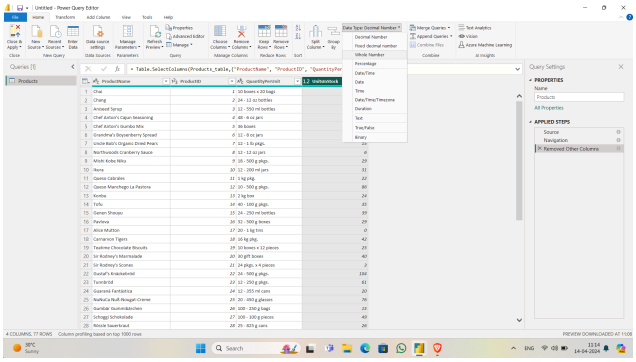


a) Removing unwanted columns

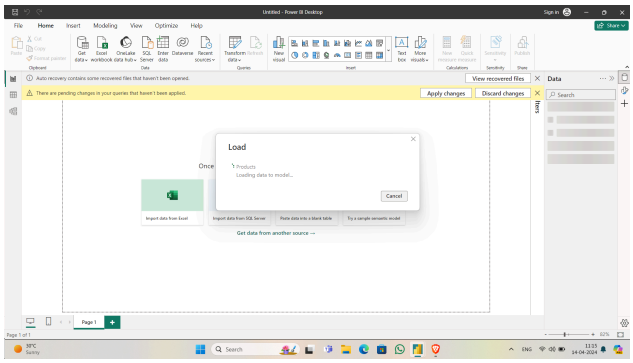




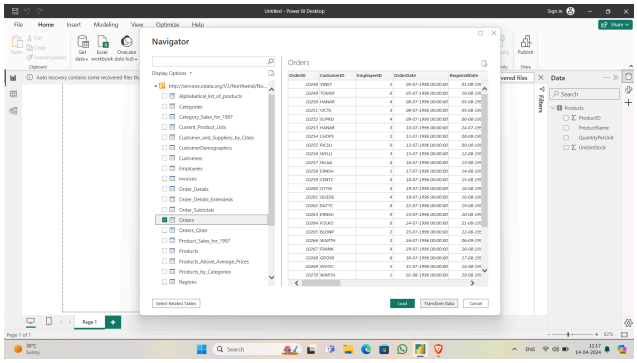
b) Changing Data Type

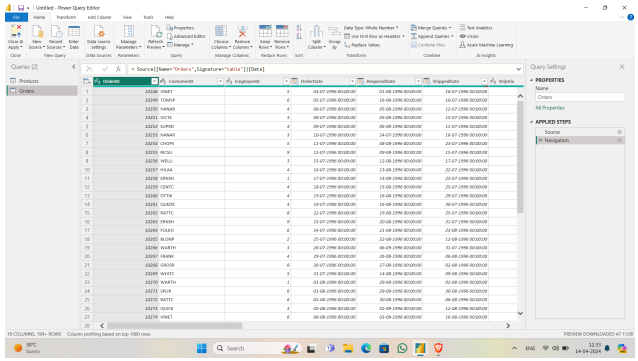


Loading the Data:



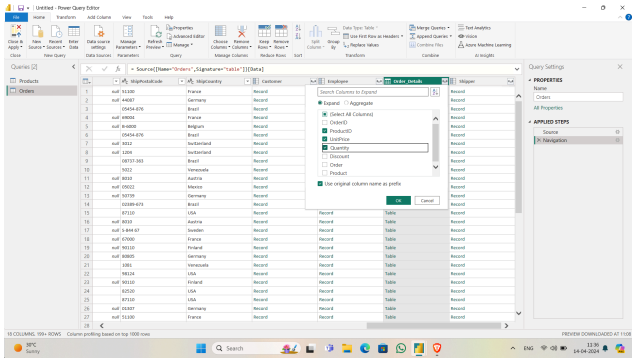
Extracting orders data:

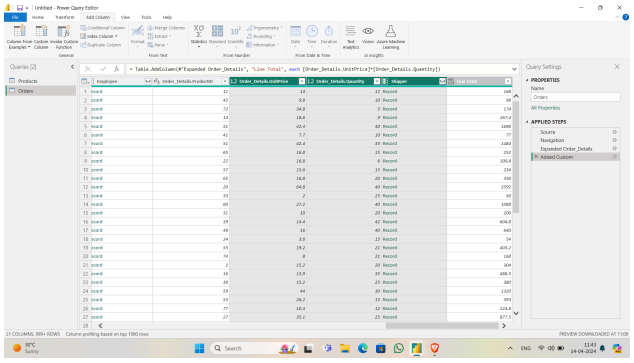




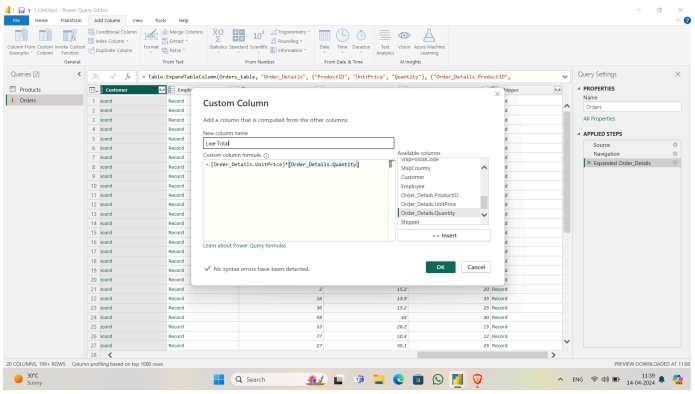
Transforming Orders Data:

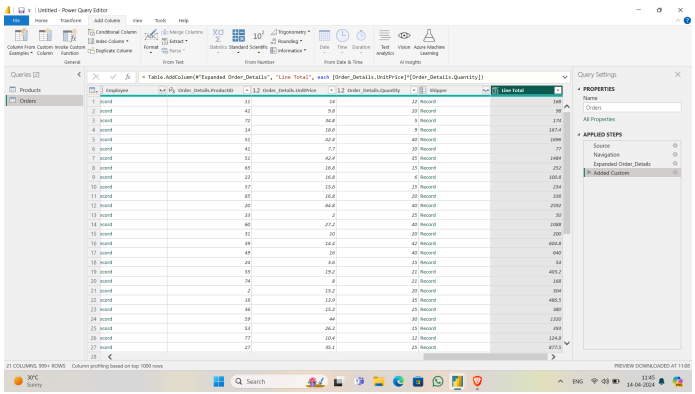
a) Expanding OrderDetails Column:





b) Creating a new Custom Columns:





Loading Data: