Data Mining and Business Intelligence

Assignment 1

Q1. Define the data warehouse.

Q2. Differentiate Between Database and Data Warehouse.

Q3. Explain Data Warehouse Components.

Q4. What do you understand by Data Pre-Processing and Data Cleaning?

Q5. Explain ETL process.

Ans 1.

Data warehouse : A Data warehouse, also known as an enterprise data warehouse, is a system used for reporting and data analysis, and is considered a core component of business intelligence. DWs are central repositories of integrated data from one or more disparate sources.

Ans 2.

- Database is a collection of related data that represents some elements of the real world whereas Data warehouse is an information system that stores historical and commutative data from single or multiple sources.

- Database is designed to record data whereas the Data warehouse is designed to analyze data.

- Database is application-oriented-collection of data whereas Data Warehouse is the subject-oriented collection of data.

- Database uses Online Transactional Processing (OLTP) whereas Data warehouse uses Online Analytical Processing (OLAP).

- Database tables and joins are complicated because they are normalized whereas Data Warehouse tables and joins are easy because they are denormalized.

- ER modeling techniques are used for designing Database whereas data modeling techniques are used for designing Data Warehouse.

Ans 3.

A data warehouse design mainly consists of five key components.

Components of Data warehouse :

- Data Warehouse Database
- Extraction, Transformation, and Loading Tools (ETL)
- Metadata
- Data Warehouse Access Tools
- Data Warehouse Bus

Ans 4.

Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format.

Data wrangling is the process of transforming and mapping data from one “data form” into another format with the intent of making it more appropriate and valuable for a variety of downstream purposes such as analytics.

The main difference is that when we speak about data wrangling, then we speak about the “cleaning” of the data. It includes data cleansing and feature engineering. It puts data into the right shape and quality for analysis.

Ans 5.

ETL tools are central to a data warehouse architecture. These tools help with extracting data from different sources, transforming it into a suitable arrangement, and loading it into a data warehouse.

The ETL tool you choose will determine:

- The time expended in data extraction
- Approaches to extracting data
- Kind of transformations applied and the simplicity to do so
- Business rule definition for data validation and cleansing to improve end-product analytics
- Filling mislaid data
- Outlining information distribution from the fundamental depository to your BI applications