DELTA Pg No.
Date / /

Experiment - 1

Aim :- Study of ETL Process and its tools.

Software used :- Weka

Theory :-

ETL (Extract, Transform and Load) Process :-
ETL is a process in Data Warehousing. It is a process in which an ETL tool extracts the data from various data source systems, transforms it in the staging area and then finally, loads it into the Data Warehouse system.

1. Extraction :-
   - Data from various source systems is extracted which can be in various formats like relational databases, NoSQL, XML and flat files into the staging area.
   - Staging area is required when you want to get the data from multiple data sources together or if you want to join two or more systems together.

2. Transformation :-
   - A set of rules or functions are applied on the extracted data to convert it into a single standard format.
   - It may involve Filtering, Cleaning, Joining, Splitting and Sorting processes.

DELTA Pg No.
Date / /

3. Loading :-
- The transformed data is finally loaded into the data warehouse.
- Sometimes the data is updated by loading into the data warehouse very frequently and sometimes it is done after longer but regular intervals.

ETL Tools :- Many data warehousing projects use ETL Tools for ETL process. Some of the tools are :-

1. Oracle Warehouse Builder (OWB) :- It provides ETL capabilities and takes advantage of inherent database abilities.

2. MarkLogic :- It is a data warehousing solution which makes data integration easier and faster using an array of enterprise features.

3. Amazon Redshift :- It is an easy to manage, simple and cost effective data warehouse tool. It can analyze almost every type of data using standard SQL.

4. DBMiner :- A data mining system for interactive mining of multiple level knowledge in large relational databases.

Conclusion :- Successfully studied ETL process & its tools.