

Date ____ / ____ / ____

Assignment-1

* Title:- For an organization of your choice, choose a set of business processes. Design a star/snowflake schemas for analyzing these processes. Create a fact constellation schema by combining them. Extract data from different data source, apply suitable transformations and load into destination tables using an ETL TOOL.

Example:- Business organization: sales order, marketing process.

* Objective:- Implement the problem statement using ETL Tool. Study star/snowflake schema.

* Outcome:- Understood star/snowflake schemas. used on ETL tool to implement a given problem statement.

* S/W & H/W Requirements :- Pentaho data integration tool 9.0.0, 64-bit OS, open JDK-8, Machine with 64-bit processors.

* Theory:-

- 1.) Star schema:-
- (i) Star schema looks like a star, with the dimensions tables surrounding the central fact table.
 - (ii) Every dimension in a star schema is represented with the only one dimension table.
 - (iii) Dimension table contains set of attributes.
 - (iv) Dimension table is joined to fact table using a foreign key.
 - (v) Dimension tables are not joined to each other.
 - (vi) Star schema is easy to understand and provide

optimal disk usage.

- (vii) The schema is widely used by business integration tools.
- (viii) The dimension tables are not normalized.

- 2) Snowflake Schema :-
- i) A snowflake schema is an extension of star schema, it adds additional dimensions.
 - ii) The dimension tables are normalized which splits data into additional tables.
 - iii) Snowflake schema uses smaller disk space.
 - iv) It is easier to add a new dimension to schema.
 - v) Due to multiple tables query performance is reduced.

* ETL: ETL is an abbreviation of Extract, Transform and Load. In this process, an ETL tool extracts the data from diff RDBMS source systems then transform the data like applying calculations, concatenation etc and then load the data into the Data Warehouse system.

In ETL, data is taken from source to target. In ETL process, transformation engine takes care of any data changes.

Instead of transforming data before it's written, ETL lets the target system to do the transformation. The first data copied to the target and then transform in place.

* Pentaho :- It is an intuitive, graphical, drag and drop design environment and a proven, scalable, standard based architecture.

Data integration delivers powerful extraction, transformation and loading capabilities, using a grand

Date ___ / ___ / ___

breaking, meta-data driven approach.

* Conclusion:- Thus, we have learnt to extract data from diff. data source, apply suitable transformations and load into destination tables using ETL tool.