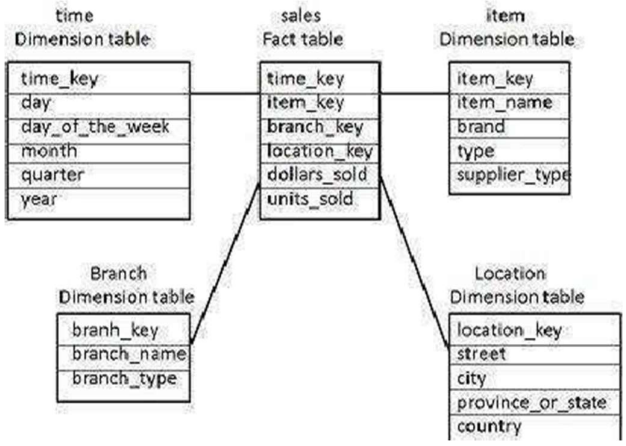
DEPARTMENT OF COMPUTER ENGINEERING
[NBA Accredited]

EXPERIMENT 1

Title:	One case study on building Data warehouse/Data Mart ✓ Write Detailed Problem statement and design dimensional modelling (creation of starand snowflake schema)
Objective:	✓ To learn fundamental of data warehousing ✓ To learn concepts of dimensional modeling ✓ To learn star, snowflake & Galaxy schema
Theory:	Dimensional modeling (DM) is the name of a logical design technique often used for data warehouses. Dimensional modeling always uses the concepts of facts, measures, and dimensions. Facts are typically (but not always) numeric values that can be aggregated, Dimensions are groups of hierarchies and descriptors that define the facts. For example, sales amount is a fact; timestamp, product, register#, store#, etc. are elements of dimensions. Dimensional models are built by business process area, e.g. store sales, inventory, claims, etc. **Fact table :** The fact table is not a typical relational database table as it is de-normalized on purpose to enhance query response times. The fact table typically contains records that are ready to explore, usually with ad hoc queries. Records in the fact table are often referred to as events, due to the time-variant nature of a data warehouse environment. The primary key for the fact table is a composite of all the columns except numeric values/scores (like QUANTITY, TURN OVER, exact invoice date and time). Typical fact tables in a global enterprise at a ware house are (usually there may be additional company or business specific fact tables) **Dimension table:** Nearly all of the information in a typical fact table is also present in one or more dimension tables. The main purpose of maintaining Dimension Tables is to allow browsing the categories quickly and easily. The primary keys of each of the dimension tables are linked together to form the composite primary key of the fact table. In a star schema design, there is only one de- normalized table for a given dimension. Typical dimension tables in a data warehouse are: Time dimension table Customers dimension table Products dimension table Key account managers (KAM) dimension table Sales office dimension table

Star schema architecture:

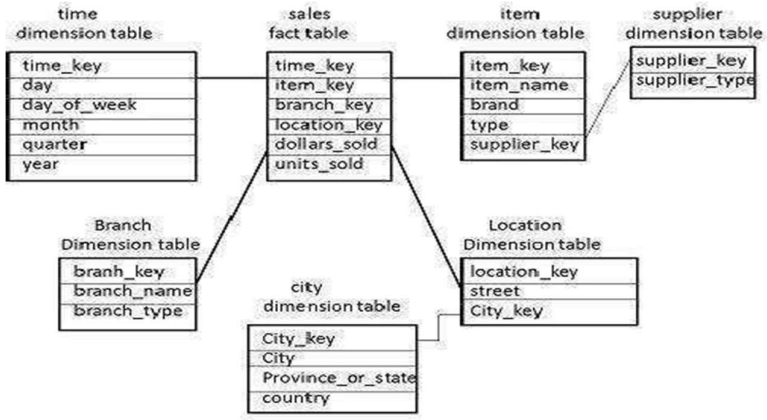
Star schema architecture is the simplest data warehouse design. The main feature of a star schema is a table at the centre, called the fact table and the dimension tables which allow browsing of specific categories, summarizing, drill-downs and specifying criteria. Typically, most of the fact tables in a star schema are in database third normal form, while dimensional
Tables are de-normalized (second normal form). Despite the fact that the star schema is the simplest data warehouse architecture; it is most commonly used in the data warehouse implementations across the world today (about 90-95% cases).



Snowflake Schema architecture:

Snow flake schema architecture is a more complex variation of a star schema design. The main difference is that dimensional tables in a snow flake schema are normalized, so they have a typical relational database design. Snow flake schemas are generally used when a dimensional table becomes very big and when a star schema can't represent the complexity of a data structure.

For example if a PRODUCT dimension table contains millions of rows, the use of snow flake schemas should significantly improve performance by moving out some data to other table. The problem is that the more normalized the dimension table is, the more complicated SQL joins must be issued to query them. This is because in order for a query to be answered, many tables need to be joined and aggregates generated.



	Fact constellation/Galaxy schema Architecture For each star schema or snow flake schema it is possible to construct a fact constellation schema. This schema is more complex than star or snowflake architecture, which is because it contains multiple fact tables. This allows dimension tables to be share amongst many fact tables. In a fact constellation schema, different fact tables are explicitly assigned to the dimensions, which are, for given facts relevant. This may be useful in cases when some facts are associated with a given dimension level and other facts with a deeper dimension level.
Deliverables:	One case study given to a group of 3 /4 students of a data mart/ data warehouse. 1. Write detail Problem Statement highlighting required information related to creation of Data Warehouse. 2. Information Package Diagram 3. Schema Design with identified fact table/s, dimensions and description
Conclusion:	A schema is a logical description of database where fact and dimension tables are joined in a logical manner. Data Warehouse is Constellation schema.