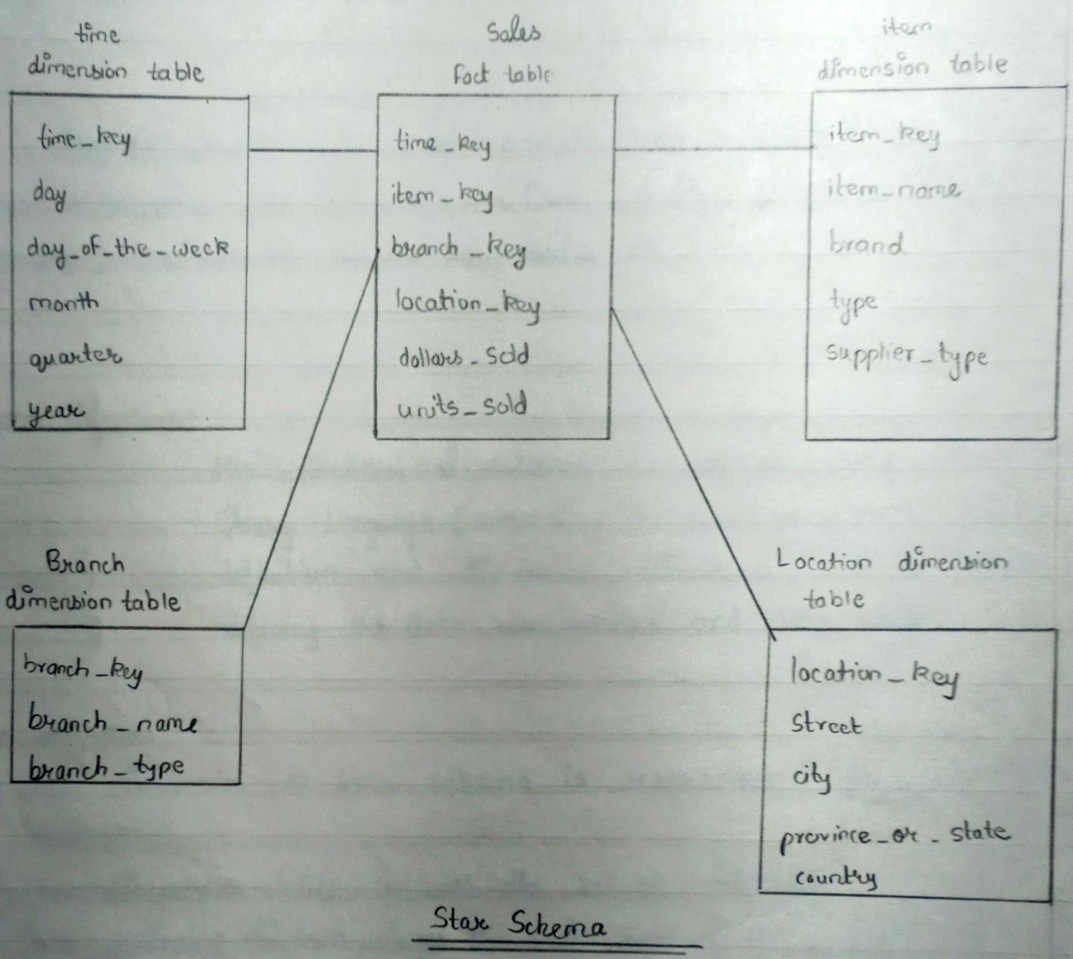
Experiment No. 01

Aim

Identify the fundamental concepts of data warehouse and data mining.

Diagrams:



Star Schema

Experiment No. 01

Aim : Identify the fundamental concepts of data warehouse and data mining.

Problem Statement :

Design multi-dimensional data models namely Star, SnowFlake and Fact Constellation schemas for ~~one~~ one enterprise (ex. Banking, insurance, finance, healthcare, manufacturing, Automobiles, sales, etc.)

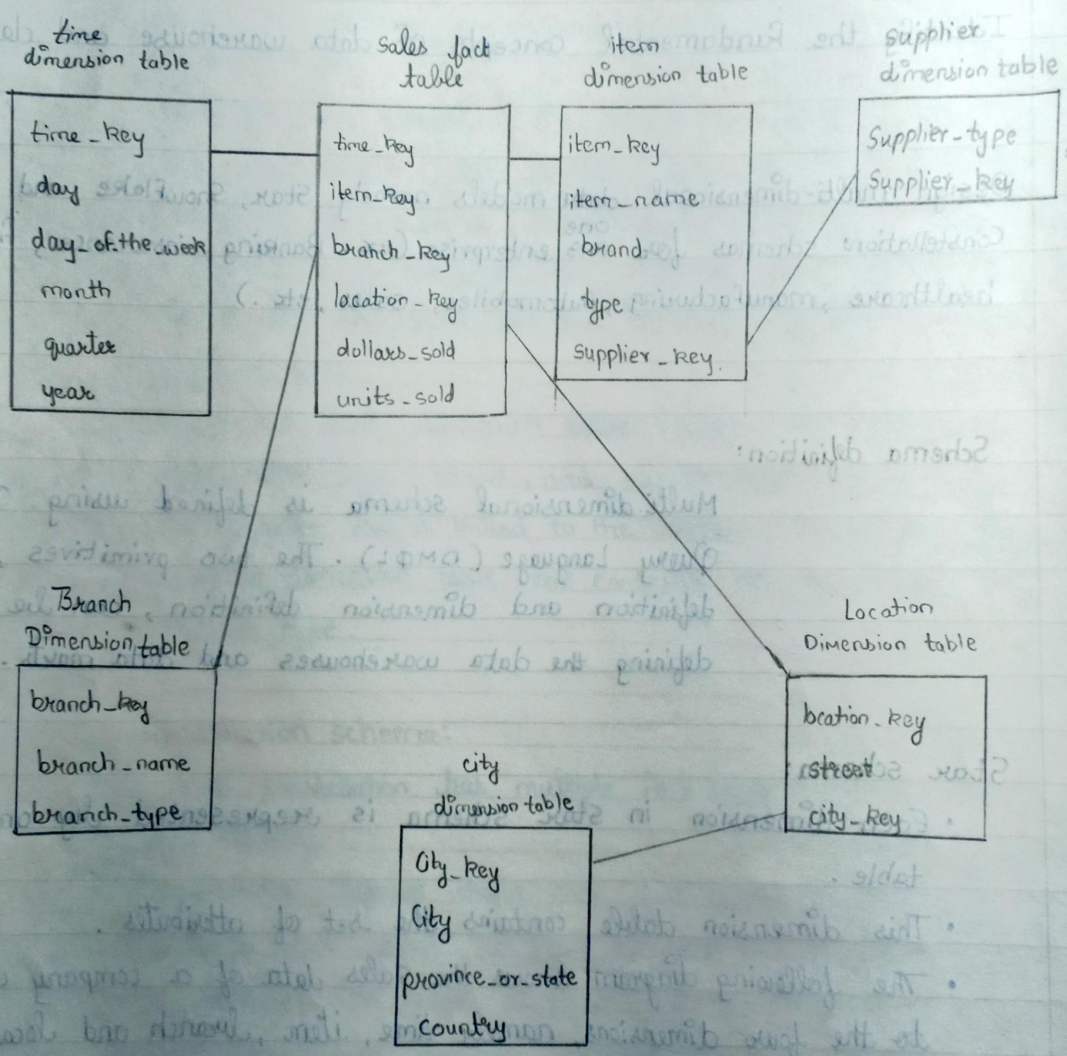
Theory :

Schema definition:

Multi dimensional schema is defined using Data Mining Query Language (DMQL). The two primitives, cube definition and dimension definition, can be used for defining the data warehouses and data marts.

Star schema:

- Each dimension in star schema is represented by one dimensional table.
- This dimension table contains the set of attributes.
- The following diagram shows the sales data of a company with respect to the four dimensions, namely time, item, branch and location.
- There is a fact table at the center. It contains the keys to each of four dimensions.
- The fact table also contains the attributes, namely dollars sold and units sold.



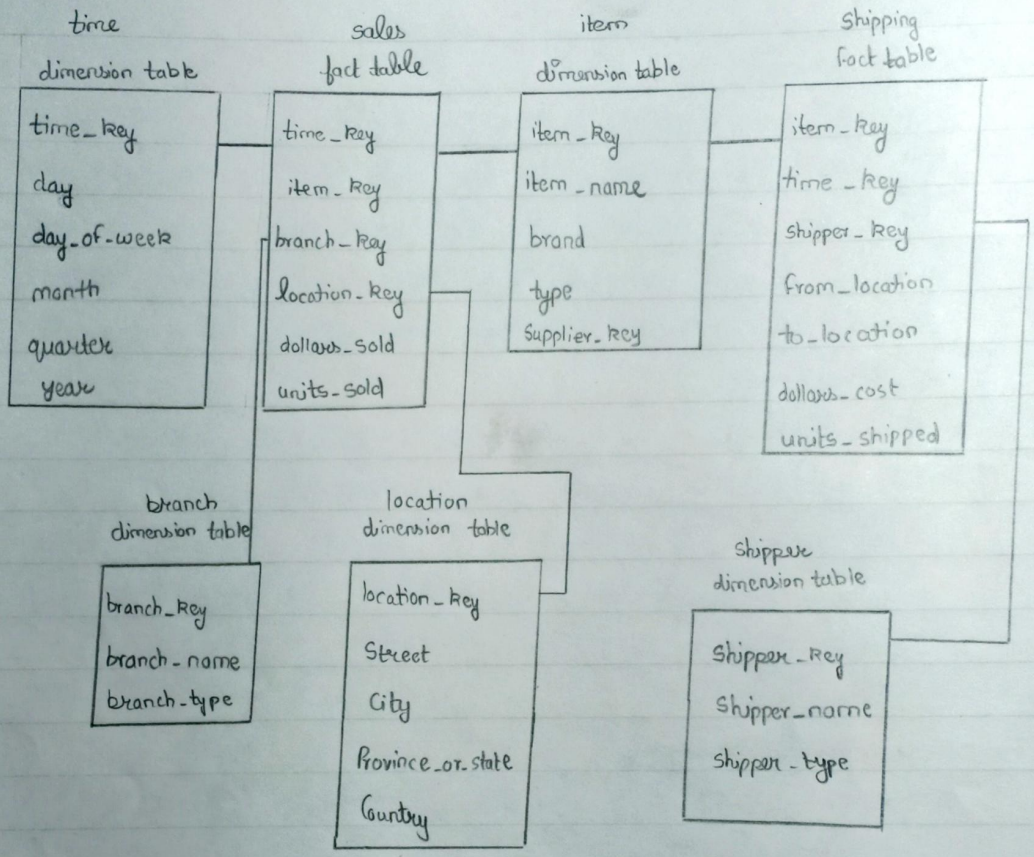
Snowflake Schema

Snowflake schema:

- Some dimension tables in the Snowflake schema are normalized.
- The normalization splits up the data into additional tables.
- Unlike star schema, the dimensions table in a snowflake schema is normalized. For example, the item dimension table in star schema is normalized and split into two dimensional tables, namely item and supplier table.
- Now the item dimension table contains the attributes item_key, item_name, type, brand, and supplier_key.
- The supplier key is linked to the supplier dimension table. The supplier dimension table contains the attributes supplier_key and supplier_type.

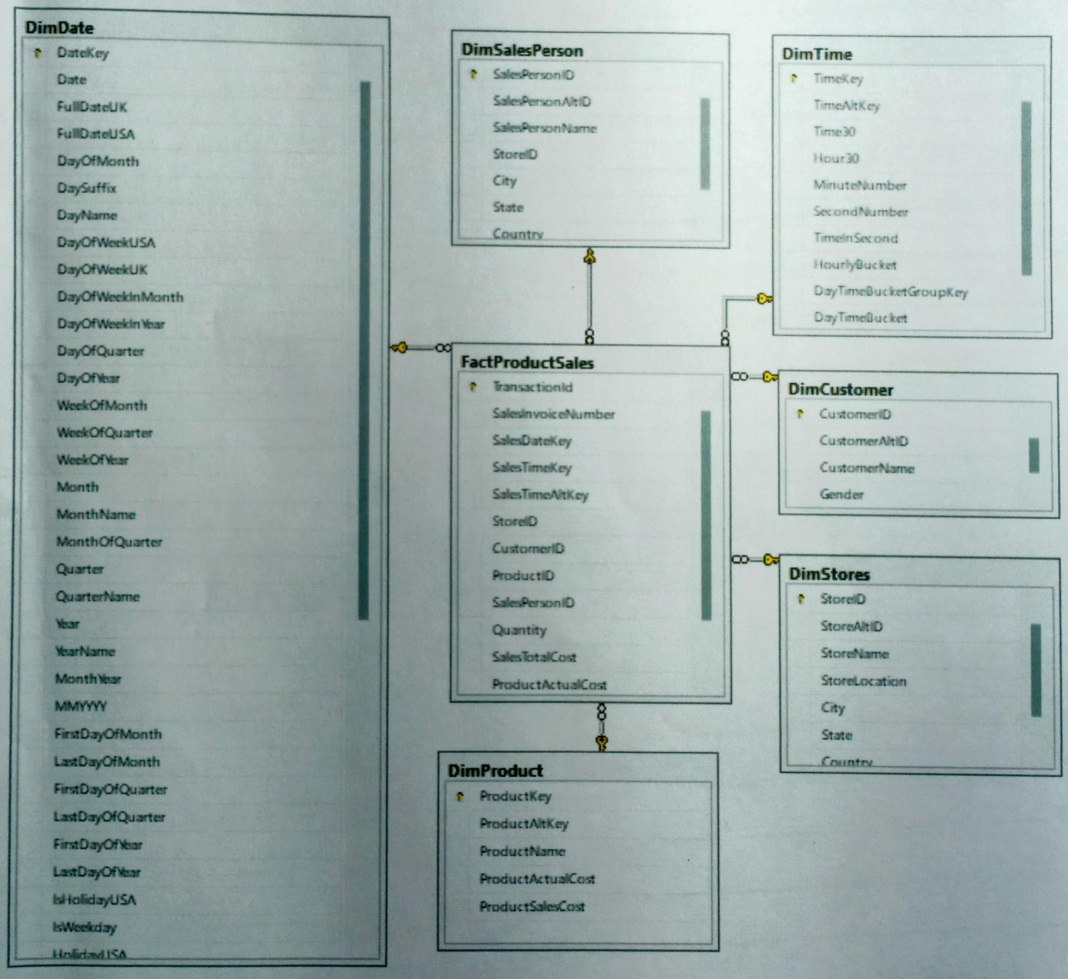
Fact constellation schema:

- A fact constellation has multiple fact table. It is also known as galaxy schema.
- The following diagram shows two fact tables, namely sales and shipping.
- The sales fact table is same as that in the star schema.
- The shipping fact table has the five dimensions, namely item_key, time_key, shipper_key, from_location, to_location.
- The shipping fact table also contains two measures, namely dollars sold and units sold.
- It is also possible to share dimension tables between fact tables. For example, time, item and location dimension tables are shared between the sales and shipping fact table.



Fact Constellation Schema

Conclusion: Hence, we successfully designed multi dimensional data models namely star, snowflake and fact constellation.


DimDate
DateKey
Date
FullDateUK
FullDateUSA
DayOfMonth
DaySuffix
DayName
DayOfWeekUSA
DayOfWeekUK
DayOfWeekInMonth
DayOfWeekInYear
DayOfQuarter
DayOfYear
WeekOfMonth
WeekOfQuarter
WeekOfYear
Month
MonthName
MonthOfQuarter
Quarter
QuarterName
Year
YearName
MonthYear
MMYYYY
FirstDayOfMonth
LastDayOfMonth
FirstDayOfQuarter
LastDayOfQuarter
FirstDayOfYear
LastDayOfYear
IsHolidayUSA
IsWeekday
HolidayUSA
DimSalesPerson
SalesPersonID
SalesPersonAltID
SalesPersonName
StoreID
City
State
Country
DimTime
TimeKey
TimeAltKey
Time30
Hour30
MinuteNumber
SecondNumber
TimeInSecond
HourlyBucket
DayTimeBucketGroupKey
DayTimeBucket
FactProductSales
TransactionId
SalesInvoiceNumber
SalesDateKey
SalesTimeKey
SalesTimeAltKey
StoreID
CustomerID
ProductID
SalesPersonID
Quantity
SalesTotalCost
ProductActualCost
DimCustomer
CustomerID
CustomerAltID
CustomerName
Gender
DimStores
StoreID
StoreAltID
StoreName
StoreLocation
City
State
Country
DimProduct
ProductKey
ProductAltKey
ProductName
ProductActualCost
ProductSalesCost

Conclusion: Hence, we successfully designed multi dimensional data models namely Star, Snowflake and fact for automobile enterprise.