Experiment No: 01

Rebecca Dias

TE CMPN A2 / 182027

Roll no:19

# Aim:

Build Data Warehouse/Data Mart for a given problem statement.

# Theory:

## 1. Why need a DW for this system

* Consolidation of information resources from different data source
* Improved query performance
* Separate research and decision support functions from the operational systems
* Foundation for data mining, data visualization, advanced reporting and OLAP tools

## 2. What criteria need to be measured?

* Scalability : with no on-premise software or hardware, it’s easy, cost-effective, simple and flexible to scale with cloud services
* Low entry cost: with no servers, hardware, IT work or operational costs, cloud services cost substantially less up-front.
* Connectivity: With cloud services, it is much easier to connect to other cloud services, which means that it’s easier to digest, store and file data.
* Security: Typical cloud providers stay hyper up-to-date with security patches and protocols to keep their host of customers safe and happy
* Availability/Reliability: If you have a strong provider, cloud solutions can have a very high up-time, which makes them reliable no matter what.
* Security: on-premise warehouses allow organizations to have full control over their security and access, which is important for businesses in which this is a big priority.

## 3. What goals need to be set?

* Present the organization's information consistently.
* Consistent information means high-quality information.
* The data warehouse must effectively control access to the organization's confidential information.

## 4. What problems need to be solved?

* The initial warehouse was created as a full refresh every day.
* The indexing is non-existent or ill-conceived.
* The data types are inconsistent or incorrect.
* Centralized Storage and Unified Reporting.
* Data and Reporting Automation.

## 5. Identify what is needed to build a DW

* Choosing the process-deciding subjects
* Choosing the grain- the level of data
* Identifying and confirming dimensions
* Choosing the facts
* Choosing the duration of the database

## Convert the system to DW:

* Determine the purpose of the mart (Business Scenario)
* Identify an association table as the central fact table
* Determine facts to be included
* Replace all keys with surrogate keys and put it in fact table
* Promote foreign keys in related tables to the fact table
* Add time dimension
* Refine the dimension tables

# Implementation:

● Choose a case study of a data mart/ data warehouse

Pantaloons, fashion retail company

● Write Detail Problem Statement specifying functionalities of the system.

Various brands like pantaloons is focusing on effective warehouse operations for satisfying customers by delivering goods on time and maintaining quality of the product.Customer data, product data, Store Data is being stored in dataware houses , helping the company to analyse various trends in the upcoming years.

● Identify facts/measures and dimension

➔ Dimension: Time, Customer, Product, Store

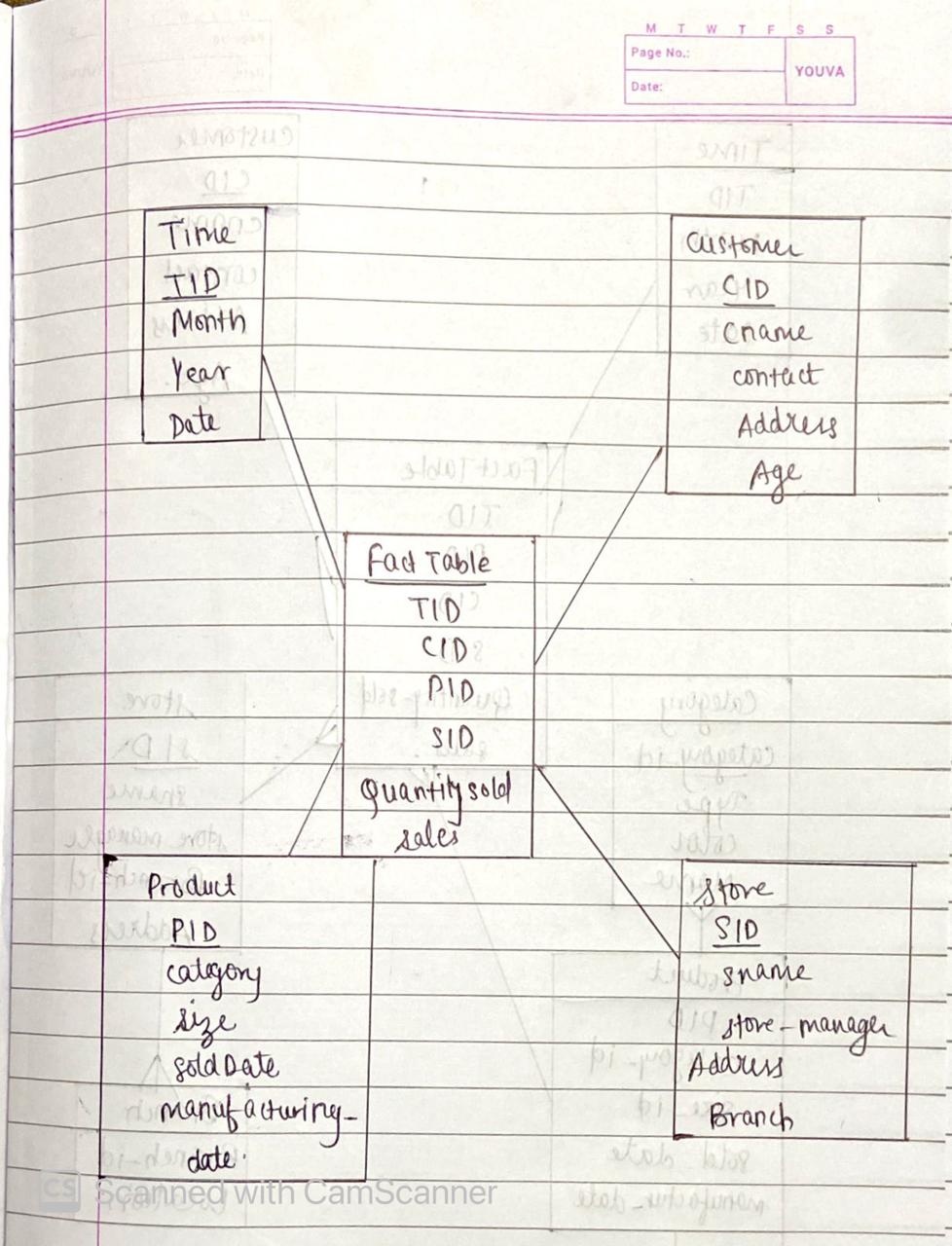
➔ Facts: Quantity\_sold ,sales

● Design Information package diagram

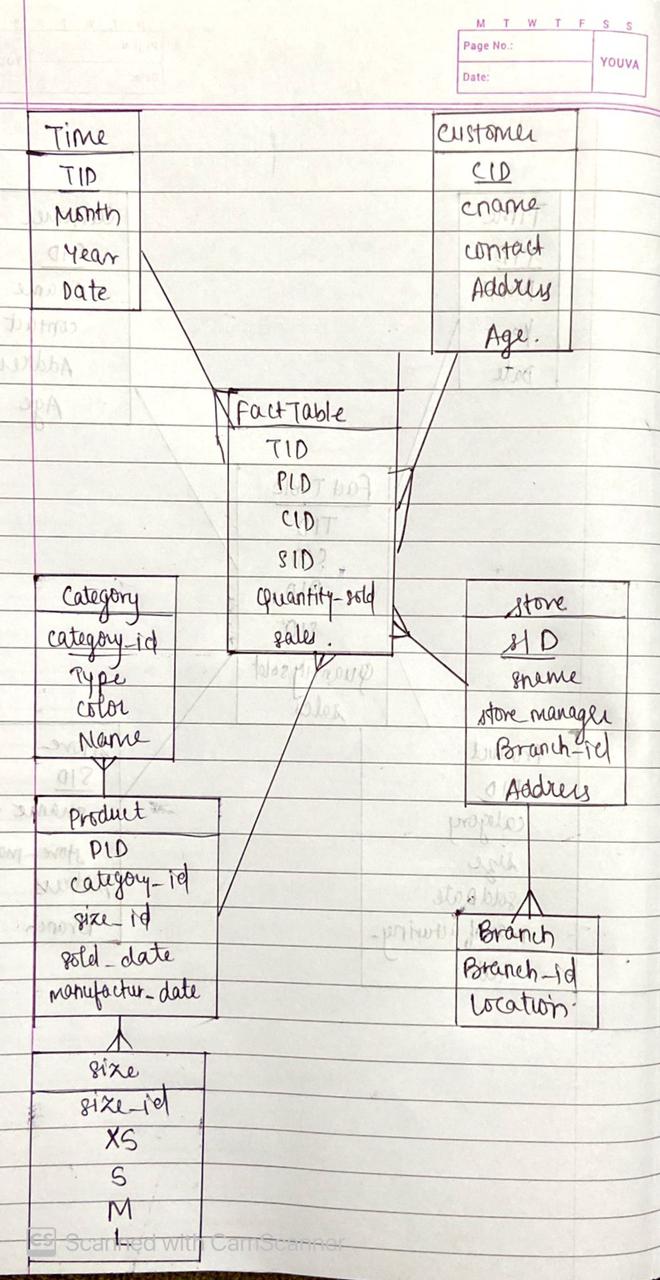
|  |  |  |  |
| --- | --- | --- | --- |
| Time | Product | Customer | Store |
| TID | PID | CID | SID |
| Date | Category | CustomerName | StoreName |
| Month | Size | Age | Address |
| Year | Manufacturing\_date | Address | Branch |
|  | Sold\_date | Contact |  |
| Facts: Quantity\_sold, Sales | | | |

● Design Dimensional model: Star Schema and Snowflake Schema

➔Star Schema



➔ Snowflake Schema



● Write 10 business rules or application

1. Which Store branch has the most sales?

2. Which Category was sold least in last year?

3. Which Store branch has the least sales?

4. Which Size was most sold?

5. Which Size was sold least?

6. What type of product was sold most?

7. Which colours were sold most?

8. On which day of the week was the sale the least?

9. Which Category was sold most in the last 6 months?

10.What was the average bill amount of each branch?

# Conclusion:

A Data warehouse is a large collection of data which helps a company make future decisions. It helps the company make future predictions much simpler by looking at the data in the data warehouse. In this experiment we successfully learned the process of making a data warehouse by using Information package diagram (IPD) by identifying the facts and dimensions. We learned how to create 2 schemas, Star Schema and Snowflake Schema