



Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



A.Y. 2021-2022

Meith Navlakha

60004190068 – TE B1

Name	SAP ID	Batch
Meith Navlakha	60004190068	TE - B1
Neil Mankodi	60004190073	TE - B1
Nishant Aridaman Kumar	60004190077	TE - B1
Rahul Raheja	60004190086	TE - B2

EXPERIMENT 6

AIM

1. Making information package diagram
2. Design dimensional data model i.e., Star schema, Snowflake schema and Fact Constellation schema (if applicable)

THEORY

Data Warehouses

Data warehousing is the process of constructing and using a data warehouse. A data warehouse is constructed by integrating data from multiple heterogeneous sources that support analytical reporting, structured and/or ad hoc queries, and decision making. Data warehousing involves data cleaning, data integration, and data consolidations.

Uses of Data Warehouses

There are decision support technologies that help utilize the data available in a data warehouse. These technologies help executives to use the warehouse quickly and effectively. They can gather data, analyse it, and make decisions based on the information present in the warehouse. The information gathered in a warehouse can be used in any of the following domains –



Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



A.Y. 2021-2022

Meith Navlakha

60004190068 – TE B1

- **Tuning Production Strategies:** The product strategies can be well tuned by repositioning the products and managing the product portfolios by comparing the sales quarterly or yearly.
- **Customer Analysis:** Customer analysis is done by analysing the customer's buying preferences, buying time, budget cycles, etc.
- **Operations Analysis:** Data warehousing also helps in customer relationship management, and making environmental corrections. The information also allows us to analyse business operations.

Functions of data warehouse tools and utilities

- **Data Extraction:** Involves gathering data from multiple heterogeneous sources.
- **Data Cleaning:** Involves finding and correcting the errors in data.
- **Data Transformation:** Involves converting the data from legacy format to warehouse format.
- **Data Loading:** Involves sorting, summarizing, consolidating, checking integrity, and building indices and partitions.
- **Refreshing:** Involves updating from data sources to the warehouse.

Information Package

An information package diagram defines the relationships between subject matter and key performance measures. The information package diagram has a highly targeted purpose, providing a focused scope for user requirements. Because information package diagrams target what the users want, they are effective in facilitating communication between the technical staff and the users, indicating any inconsistencies between the requirements and what the data warehouse will deliver.

An Information Package is a conceptual container having two types of information called Content Information and Preservation Description Information (PDI). The Content Information and PDI are viewed as being encapsulated and identifiable by the Packaging Information. The resulting package is viewed as being discoverable by virtue of the Descriptive Information.



A.Y. 2021-2022

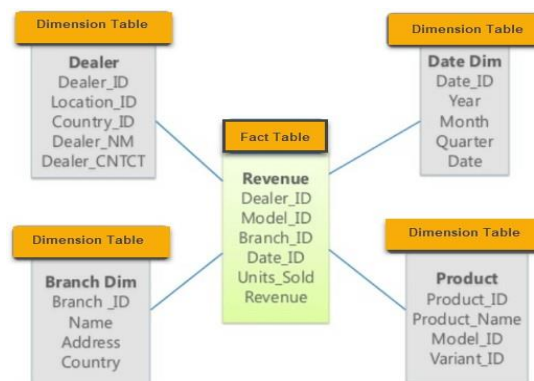
Meith Navlakha

60004190068 – TE B1

Types of Schemas

Star Schema

The star schema separates business process data into facts, which hold the measurable, quantitative data about a business, and dimensions which are descriptive attributes related to fact data. Examples of fact data include sales price, sale quantity, and time, distance, speed and weight measurements. Related dimension attribute examples include product models, product colours, product sizes, geographic locations, and salesperson names. A star schema that has many dimensions is sometimes called a centipede schema. Having dimensions of only a few attributes, while simpler to maintain, results in queries with many table joins and makes the star schema less easy to use.



Snowflake Schema

In computing, a snowflake schema is a logical arrangement of tables in a multidimensional database such that the entity relationship diagram resembles a snowflake shape. The snowflake schema is represented by centralized fact tables which are connected to multiple dimensions. "Snowflaking" is a method of normalizing the dimension tables in a star schema. When it is completely normalized along all the dimension tables, the resultant structure resembles a snowflake with the fact table in the middle. The principle behind snowflaking is normalization of the dimension tables by removing low cardinality attributes and forming separate tables.

The snowflake schema is similar to the star schema. However, in the snowflake schema, dimensions are normalized into multiple related tables, whereas the star schema's dimensions are denormalized with each dimension represented by a single table. A

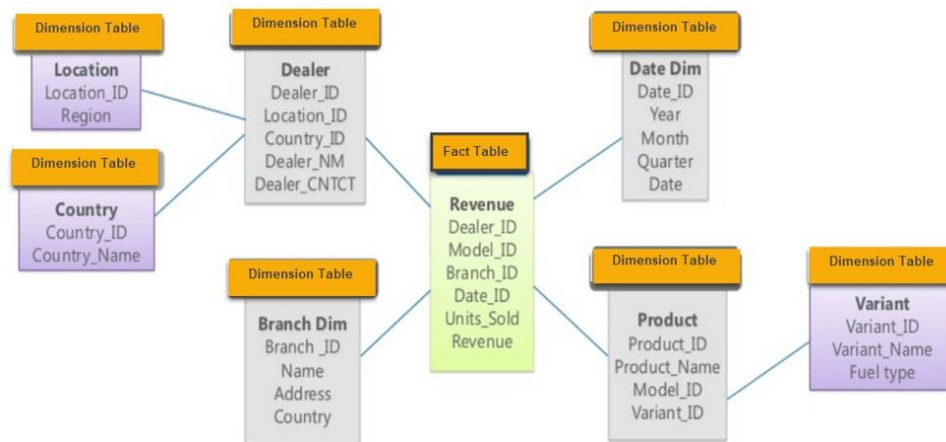


A.Y. 2021-2022

Meith Navlakha

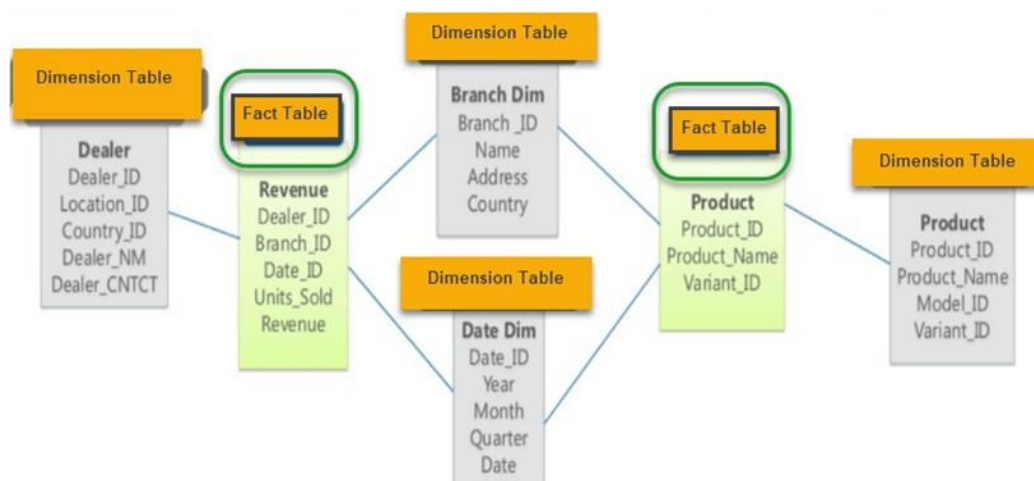
60004190068 – TE B1

complex snowflake shape emerges when the dimensions of a snowflake schema are elaborate, having multiple levels of relationships, and the child tables have multiple parent tables ("forks in the road").



Fact Constellation Schema

Fact Constellation is a schema for representing a multidimensional model. It is a collection of multiple fact tables having some common dimension tables. It can be viewed as a collection of several star schemas and hence, also known as Galaxy schema. It is one of the widely used schema for Data warehouse designing and it is much more complex than star and snowflake schema. For complex systems, we require fact constellations.





A.Y. 2021-2022
Meith Navlakha
60004190068 – TE B1

EXPERIMENTATION:

INFORMATION PACKAGE - Superstore

Time	Customer	Product	Branch	Suppliers
Year	Customer_ID	Product_ID	Branch_ID	Supplier_ID
Quarter	Mode of Payment	Brand	City	Delivery Time
Month	Address	Product Category	Rating	Product Category Procured
Day	Membership	Price	State	City
Season	Age	Tax	Number of Employees	State
Holiday Flag	Gender	Discount	Monthly expenses	
		Expiration Time		

Facts:

1. Average Delivery Time of all the Suppliers in business.
2. Profit Margin of the Branches.
3. Product wise sales.
4. Age group wise sales.
5. Seasonal sales.
6. Membership wise sales.

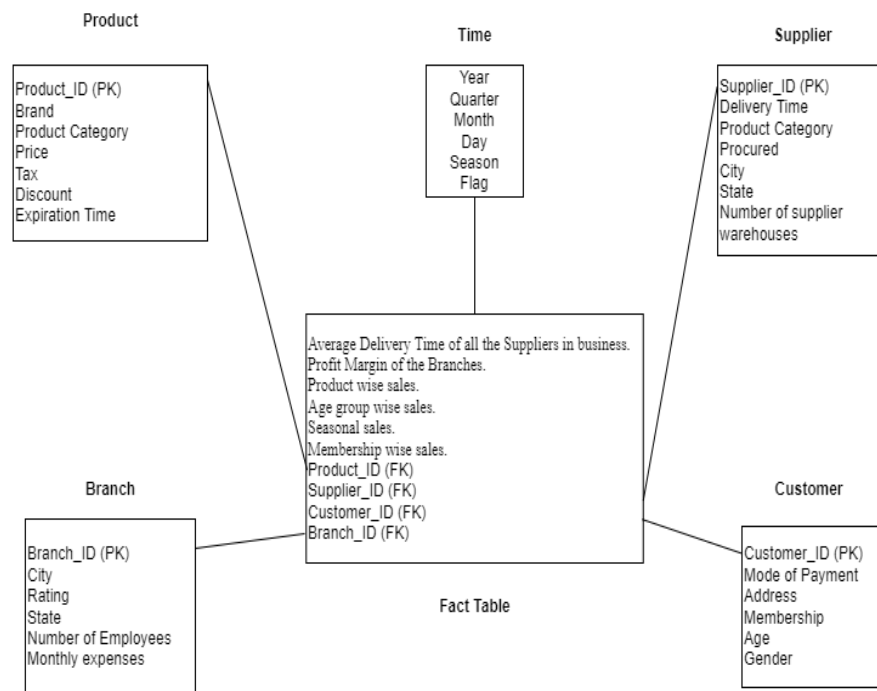


A.Y. 2021-2022

Meith Navlakha

60004190068 – TE B1

STAR SCHEMA



CONCLUSION

The main objective of this experiment was to design an Information Package and a Dimensional Model for our selected application of data warehousing. The selected application for which we performed this experiment was a Superstore. The Information Package is a novel idea for determining and recording information requirements for a data warehouse. The designing process of an Information Package usually involves several iterations of taking feedback from the stakeholders that will eventually make use of the data warehouse and making changes until a final design is created which satisfies all possible requirements. While the Snowflake schema reduces redundancies and is more compact, the Star Schema allows for faster and simpler execution of queries which coupled with the lack of complexity in our current warehouse model, made the latter a better option for our case. After evaluating the trade-offs that are present we decided to go with the Star Schema as the dimensional data model for our experiment. Thus, we have completed the experiment on Building a Data Warehouse which involved designing the Information Package and the Dimensional Data Model.