DMW - Experiment 2

Name: Urmi Dedhia *SAPID: 60004200108*
Branch: Computer Engineering *Division: B1*

Aim:

Build Data Warehouse/Data Mart for a given problem statement

1. Identifying the source tables and populate sample data
2. Making information package diagram
3. Design dimensional data model i.e. Star schema, Snowflake schema and Fact Constellation schema (if applicable)

Problem Statement:

Apple is a multinational Technology Company. It specializes in consumer electronics, software and online services. Apple is the largest technology company by revenue (totaling US$365.8 billion in 2021) and, as of June 2022, is the world's biggest company by market capitalization, the fourth-largest personal computer vendor by unit sales and second-largest mobile phone manufacturer. It is one of the Big Five American information technology companies, alongside Alphabet, Amazon, Meta, and Microsoft.

Apple not only sells a number of Electronic devices, but also provides multiple services like Apple Music, etc. The Data Warehouse built should be constructed in such a way that strategic information for all these areas must be easily available. Important conclusions like which region in the world has a greater demand, which customer age group uses a particular service (so the services can be customized as per their interests), which product has an excess of supply, etc. should be easily drawn from the Data Warehouse.

Theory:

In a **data warehouse**, an information package is a logical grouping of related data that is organized and presented in a specific format. The data in an information package is typically organized into one or more data tables, and the package may also include supporting information such as a description of the data, a list of the data sources, and a glossary of terms.
An **information package** is often used to provide data to decision makers who need to analyze the data to make business decisions. For example, a data warehouse may contain information packages that contain data about customer purchasing habits, product sales, and inventory levels. Decision makers can use this data to make decisions about what products to stock, how to price products, and how to promote products.
An information package is defined as a conceptual representation of the data warehouse's information requirements. We gain a better understanding of the various insights, nebulous thoughts, and opinions expressed during the application process for collecting requirements by doing so.

Information Package Content: It contains the Content Data Object (physical object or digital object, i.e., bits), as well as the Representation Information required to explain the Content Data Object to the designated community.

Submission Information Package: Data producers create submission information packages (SIPs) from the Submission Information Package (SIP). A SIP is an archive deposit of digital data, as well as any documentation or metadata required to support the archive's long-term preservation and to provide consumers (e.g., reuse) with access to the archive.

Star Schema -
A star schema is a database organizational structure optimized for data warehousing or business intelligence that uses a single large fact table to store transactional or measure data and one or more smaller dimension tables to store attributes about the data.

Interview Questions:

Q. Where does your data come from?
A. We get our data from CRM, Facebook, Google Analytics and BI software.

Q. How do you measure profit margin?
A. Our main formula for measuring the profit accumulated is Net income / revenue. Our profit margin deteriorated from the financial year 2019-2020 but increased from the financial year 2020-2021.

Q. Which product is your biggest source of revenue?
A. Our iPhones produce the biggest source of revenue.

Q. What are the business segments of Apple?
A. Countries like Japan, America, China etc. are our business segment nations. Currently the US is leading but Asia is catching up. America is contributing 46B dollars in the first quarter of 2021. 31% of Macbook users in the US are between ages 25-34. 53% of all MacBook users are from small towns. 1/10 people own an Apple Watch.

Q. Customer demographics of your company?
A. More than 1.6B active apple devices are currently in use across the world out of which 1B are iPhones. iOS users account for 26.99% of all mobile users. As of 2019, 51% of all iOS users were female and 43% were male. It's the only major mobile vendor with max female users.

Q. Which kind of analysis helps you the most in making strategic decisions?
A. Correlation analysis. It measures the strength of the linear relationship between two variables and computes their association. Simply put - correlation analysis calculates the level of change in one variable due to the change in the other. Q. Relation b/n customer satisfaction and product sales?

A. According to our research, we have seen that the more satisfied the customers are, better are the sales.

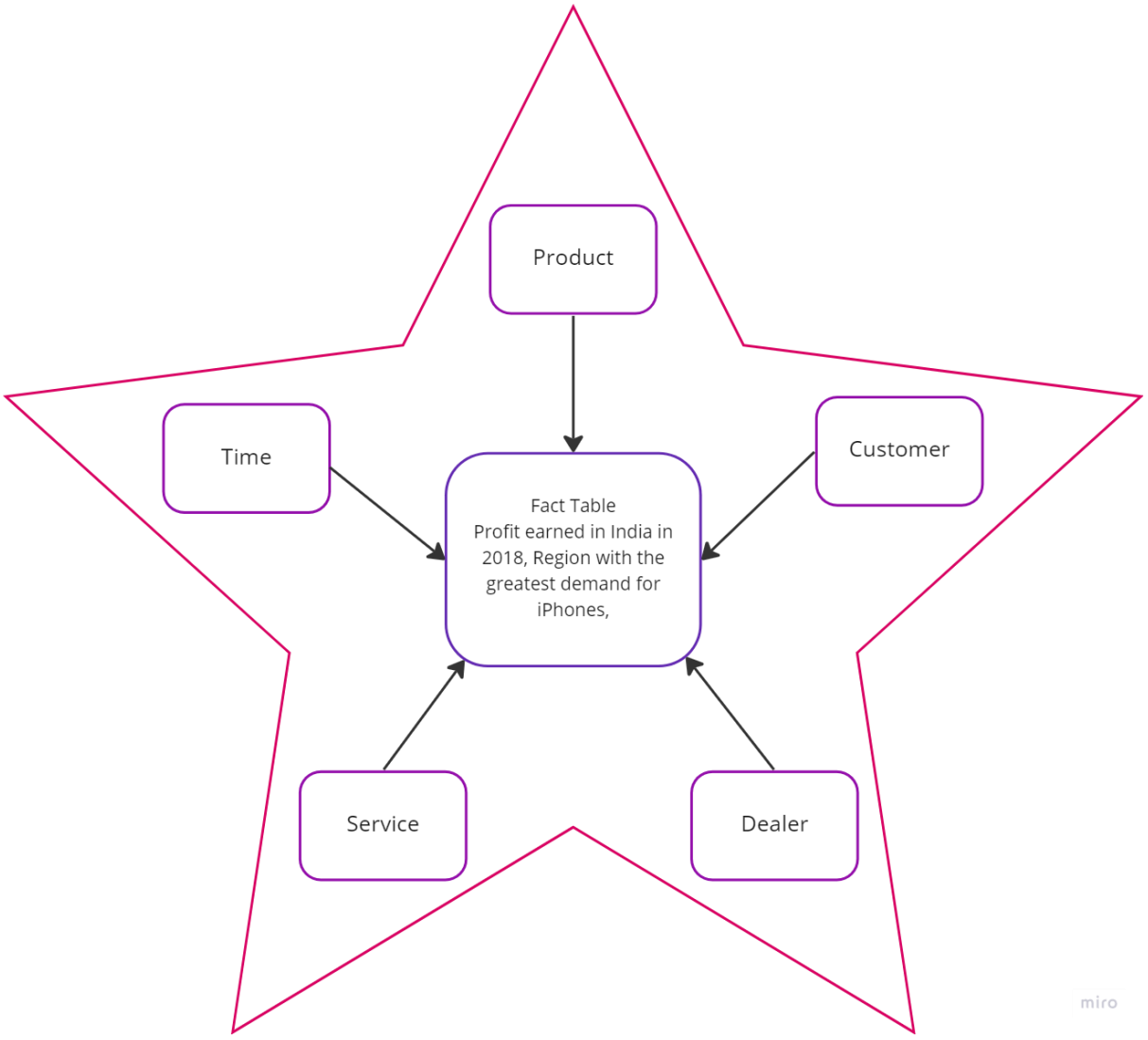
Q. How are apple services growing?
A. Apple Music had 98M users by 2021.

IP:

Time	Customer	Product	Service	Dealer
Yearly	Name	Model Name	Name	Name
Quarterly	DOB	Model Year	Launch Year	State
Monthly	Gender	Price	Subscribers	Country
Weekly	State	Category	Revenue	Region
Daily	Country	Colour	Subscription Type	Contact No.
Season	Region	Specifications	SubscriptionPeriod	Email
	Contact No.	Warranty	Price Plans	Sales
	Email	Storage	Region	Revenue
		Place of Assembly	Country	Date First Operation
		Revenue		Payment Methods
		Sales		
		Cost of Manufacture		

Facts: Profit earned in India in 2018, Region with the greatest demand for iPhones, Most popular product among the ages 18-25 years, Service that generates the most revenue in Australia, the most common nationality among employees

Data model:



Conclusion:

Hence, we have successfully understood the process of designing a data warehouse & the resources required to do so. We modelled an information package along with a star schema & fact table.