

Experiment No.1

Aim:

Build Data Warehouse/Data Mart for a given problem state

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

Theory:

Data Warehouse: A data warehouse is a large collection of business data used to help an organization make decisions. The concept of the data warehouse has existed since the 1980s, when it was developed to help transition data from merely powering operations to fueling decision support systems that reveal business intelligence.

A Data Warehousing (DW) is a process for collecting and managing data from varied sources to provide meaningful business insights. A data warehouse is typically used to connect and analyze business data from heterogeneous sources. The data warehouse is the core of the BI system which is built for data analysis and reporting.

It is a blend of technologies and components which aids the strategic use of data. It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users in a timely manner to make a difference.

Star Schema: Each dimension is represented with only one-dimensional table. This table contains a set of attributes. There is a fact table at the center, It contains keys to other dimensions.

Snowflake Schema: Some dimension tables normalized. The normalization splits up the data into additional tables. Due to normalization, the snowflake schema, the redundancy is reduced and therefore, it becomes easy to maintain and save storage space.

Information Package: An Information Package is a conceptual container of 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.

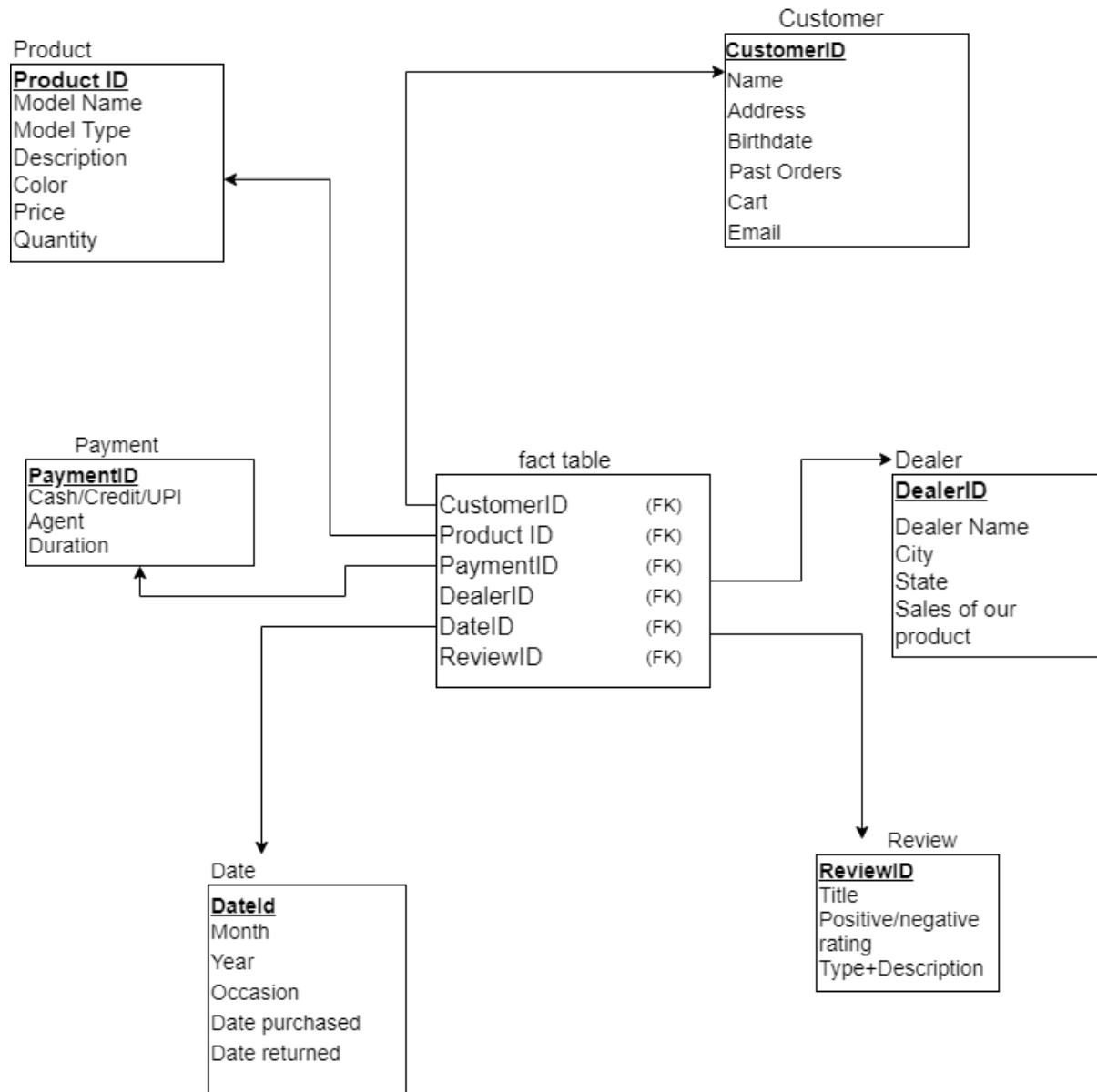
Questions:

- Q.1)** What is the percent increase in sales during IPO Announcement/Dividend Declaration
- Q.2)** Compare the Percent Increase in Customers during advertising Campaign and subsequently calculate the ROI (Return on Investment) with respect to advertising Investment
- Q.3)** Which schemes influenced maximum sales?
- Q.4)** What is the quarter-wise sales distribution of every product?
- Q.5)** Which payment method is common among the customers and for which product?
- Q.6)** Which product and sold by which agent has the most negative reviews/complaints?
- Q.7)** What product has the most customers opting for EMI?
- Q.8)** How many existing India Bulls customers have consistently renewed their membership year over year
- Q.9)** In which locations do we have customers greater than a specific number where we don't have dealers?
- Q.10)** What age groups prefer which type of membership
- Q.11)** How many Dhani App Members have enrolled for finance lending scheme?

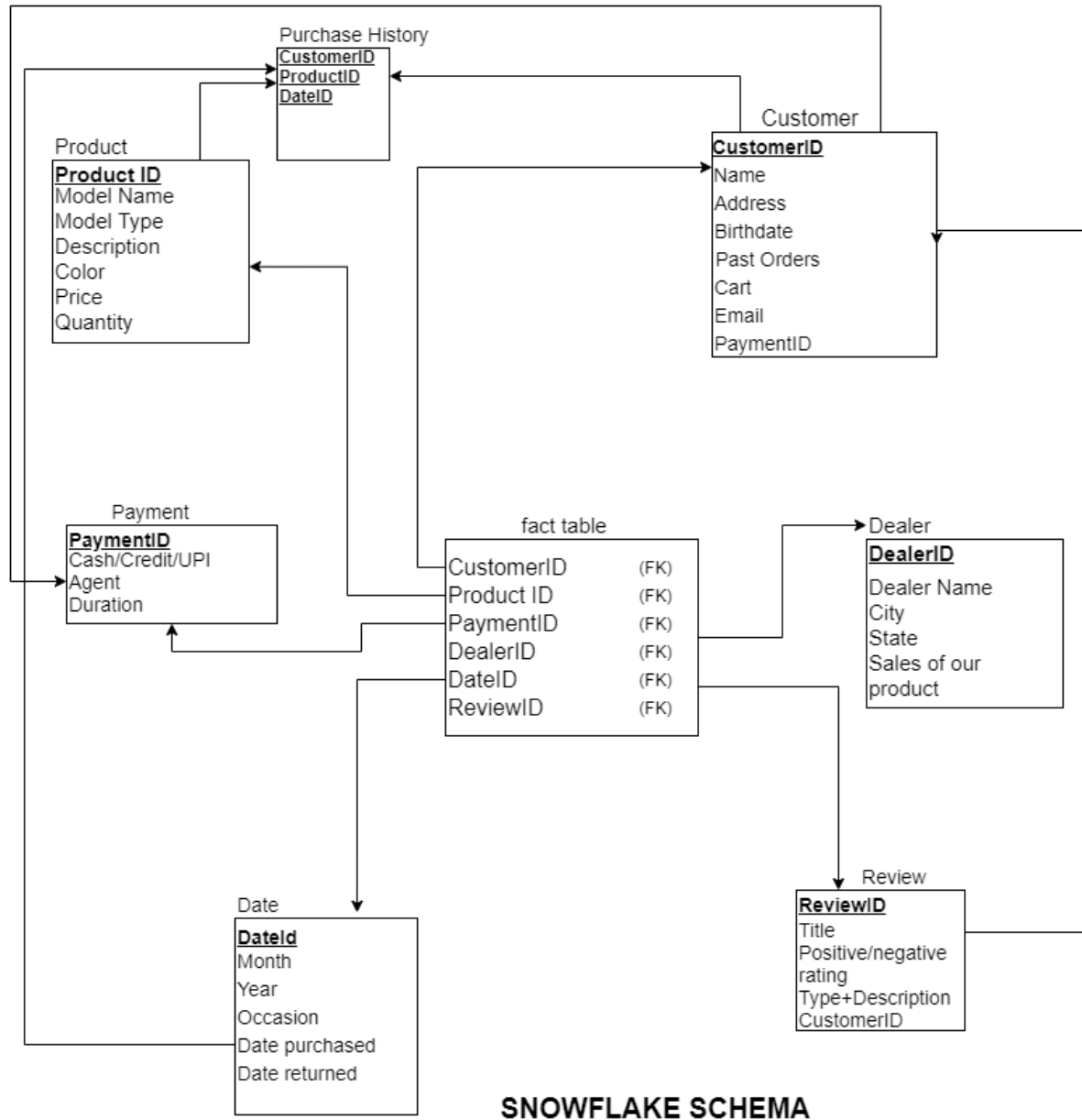
Dimensions

Categories	Time	Product	Customer Info	Dealer	Payment Method	Review
	Product launch date	Model name	Name	Dealer name	Cash, Credit, EMI or UPI	Title
	Year	Model type	Address	City	Agent	Positive negative rating
	Month	Description	Birthdate	State	Duration	Complaints type+ Description
	Quarter	Color	Past complications if any	Sales of our product		
	Day of week	Price	Cart			
	Day of month	Quantity	Email			

	Occasion flag					
	Product return flag					



STAR SCHEMA



Conclusion:

Hence, we successfully created star schema and snowflake schema based on information package diagram.