Data Warehousing

Data Warehouse for the DTH Service Provider Organization

Instructor Incharge:
Dr. L Rajya Lakshmi
Assistant Professor

1] Aniket Mourya (2020H1120298P)
2] Vishal Jha (2020H1120269P)
3] Parth Pandya (2020H1120287P)

Problem Overview :

- DTH Service Provider Organization often need to analyze the historical data to make strategic decisions for the organization.
- From the Operational Data, Organization Can not make well informed and thoroughly analyzed strategic decision.
- Strategic information obtained from analysis helps to improve Sales.
- It also helps generate more revenue from existing recharge plans.

Solution :

- Data warehouse will be a good solution for this.
- From data warehouse company can analyze the trends of the business from different perspectives at runtime only.
- Data can be easily available to person who is responsible to make strategic decisions.
- Analysing the scope of business expansion, identification of regions where sales can be improved, combinations of plans and offers to increase revenue.

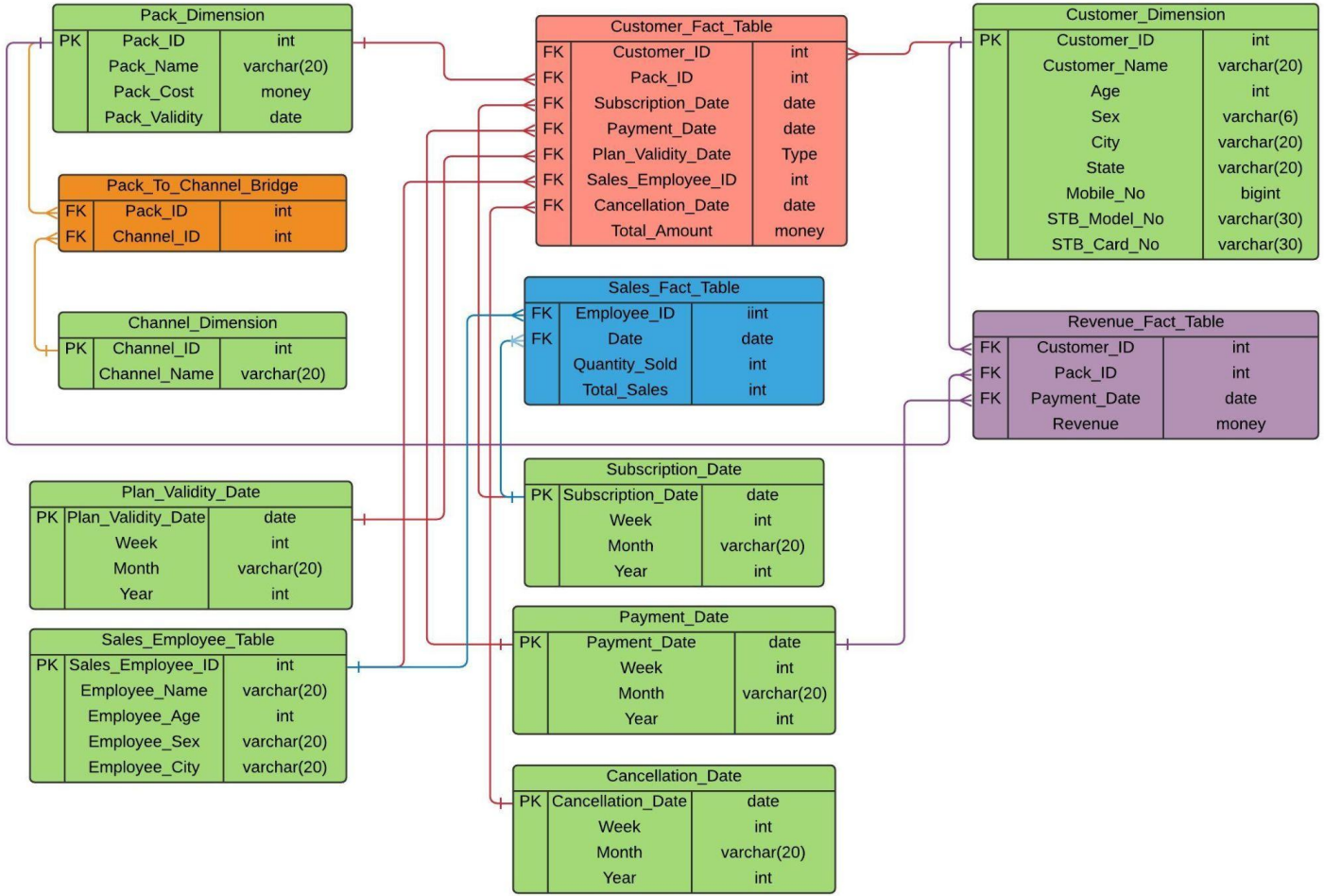
Business Analysis :

- The given data warehouse for “DTH Service Provider” analyses various business goals here to answer some of the strategic details:
- Which are the most preferred plans and STB Model?
- Is the number of subscribers increasing or decreasing. What can be the corrective action ?
- Analysis of transactions.(Are cancellations more prominent?)
- Can offering a certain plan combination help increase our revenue?
- Monitor Sales employee performance.

Identifying Trends (Queries)

- Names of top 3 employee who did maximum sales on a particular day
- Revenue of company year wise.
- The 3 top-selling plans in the customer base
- List of customers whose plan expires on a particular day
- Top 5 cities with the highest customer base
- List of top 5 cities which have the highest cancellation rate
- List of 5 cities with lowest subscription
- The 3 top-selling plans in particular city
- Show payment history of particular customer
- Find the STB model that has the highest sales
- Which month has the most new subscription

STAR SCHEMA


Pack_Dimension
PK Pack_ID int
Pack_Name varchar(20)
Pack_Cost money
Pack_Validity date
Pack_To_Channel_Bridge
FK Pack_ID int
FK Channel_ID int
Channel_Dimension
PK Channel_ID int
Channel_Name varchar(20)
Customer_Fact_Table
FK Customer_ID int
FK Pack_ID int
FK Subscription_Date date
FK Payment_Date date
FK Plan_Validity_Date Type
FK Sales_Employee_ID int
FK Cancellation_Date date
Total_Amount money
Sales_Fact_Table
FK Employee_ID iint
FK Date date
Quantity_Sold int
Total_Sales int
Customer_Dimension
PK Customer_ID int
Customer_Name varchar(20)
Age int
Sex varchar(6)
City varchar(20)
State varchar(20)
Mobile_No bigint
STB_Model_No varchar(30)
STB_Card_No varchar(30)
Revenue_Fact_Table
FK Customer_ID int
FK Pack_ID int
FK Payment_Date date
Revenue money
Plan_Validity_Date
PK Plan_Validity_Date date
Week int
Month varchar(20)
Year int
Sales_Employee_Table
PK Sales_Employee_ID int
Employee_Name varchar(20)
Employee_Age int
Employee_Sex varchar(20)
Employee_City varchar(20)
Subscription_Date
PK Subscription_Date date
Week int
Month varchar(20)
Year int
Payment_Date
PK Payment_Date date
Week int
Month varchar(20)
Year int
Cancellation_Date
PK Cancellation_Date date
Week int
Month varchar(20)
Year int

Schema Overview:

- This schema includes 8 Dimensions, 1 bridge Table and 3 Fact Tables.

Dimensions	Fact_Tables	Bridge_Table
Customer_Dimension	Customer_Fact_Table	Pack_to_Channel_Bridge
Pack_Dimension	Sales_Fact_Table	
Channel_Dimension	Revenue_Fact_Table	
Plan_Validity_Date		
Sales_Employee_Table		
Subscription_Date		
Payment_Date		
Cancellation_Date		

Schema for Dimension tables:

- Dimension Table:
  - **Customer_Dimension** : Customer_ID, Customer_Name, Age, Sex, City, State, Mobile_No, STB_Model_No, STB_Card_No
  - **Pack_Dimension** : Pack_ID, Pack_Name, Pack_Cost, Pack_Validity
  - **Sales_Employee_Dimension** : Sales_Employee_ID, Employee_Name, Employee_Age, Employee_Sex, Employee_City
  - **Channel_Dimension** : Channel_ID, Channel_Name
  - **Plan_Validity_Date** : Plan_Validity_Date, Week, Month, Year
  - **Subscription_Date** : Subscription_Date, Week, Month, Year
  - **Payment_Date** : Payment_Date, Week, Month, Year
  - **Cancellation_Date** : Cancellation_Date, Week, Month, Year

Schema for Fact Tables:

- Fact Tables
  - **Customer_Fact_Table** : Customer_ID, Pack_ID, Subscription_Date, Payment_Date, Plan_Validity_Date, Sales_Employee_ID, Cancellation_Date, Total_Amount
  - **Sales_Fact_Table** : Employee_ID, Date, Quantity_Sold, Total_Sales
  - **Revenue_Fact_Table** : Customer_ID, Pack_ID, Payment_Date, Revenue

- Bridge Tables
  - **Pack_to_Channel_Bridge :** Pack_ID, Channel_ID

SQL Queries

```
/* Name of top 3 employee who did maximum sales on a particular date */

SELECT TOP 3  se.Employee_Name, sum(sf.Quantity_Sold) AS "Units Sold"

FROM Sales_Fact_Table sf, Sales_Employee_Table se

WHERE se.Sales_Employee_ID=sf.Empolyee_ID AND Date = '11/07/2020'

GROUP BY se.Employee_Name

ORDER BY "Units Sold" DESC
```

	Employee_Name	Units Sold
1	Quinlan Q. Bass	3
2	Colby E. Macdonald	2
3	Yael Rivas	2

```
/* Revenue of company year wise */

SELECT year(Payment_Date) AS Revenue_Year ,

SUM(revenue) AS Total_revenue FROM Revenue_Fact_Table

GROUP BY year(Payment_Date) ORDER BY Total_revenue

DESC;
```

Results Messages

	Revenue_Year	Total_revenue
1	2020	318372.00
2	2021	134775.00

```
/* SELECT the 3 top-selling packs in the Customer base */

SELECT top 3 Pack_ID, COUNT(Customer_ID) AS No_of_Sell

FROM Customer_Fact_Table

GROUP BY Pack_ID

ORDER BY No_of_Sell DESC
```

Results | Messages

	Pack_ID	No_of_Sell
1	3	60
2	12	58
3	10	55

```
/* List of customers whose plan expires on a particular day */

SELECT cd.Customer_ID,cd.Customer_Name,pv.Plan_Validity_Date

FROM Customer_Dimension cd, Customer_Fact_Table cf, Plan_Validity_Date pv

WHERE cd.Customer_ID = cf.Customer_ID

 AND pv.Plan_Validity_Date = cf.Plan_Validity_Date

 AND pv.Plan_Validity_Date = '2020/05/02'
```

	Customer_ID	Customer_Name	Plan_Validity_Date
1	97	Ima D. Mcmillan	2020-05-02
2	206	Shea Mayo	2020-05-02

```
/* Top 5 cities with the highest customer base */

 SELECT Top 5 City, COUNT(DISTINCT Cu.Customer_ID) AS Subscribers

 FROM Customer_Fact_Table C,Customer_Dimension Cu, Subscription_Date S

WHERE C.Customer_ID = Cu.Customer_ID and C.Subscription_Date = S.Subscription_Date

GROUP BY City

ORDER BY Subscribers DESC;
```

Results | Messages

	City	Subscribers
1	Jaipur	69
2	Surat	61
3	Kolkata	57
4	Hydrabad	56
5	Delhi	54

```
/* List of top 5 cities which have the highest cancellation rate */

SELECT Top 5 City, COUNT(*)AS Cancellations

FROM Customer_Fact_Table C,Customer_Dimension Cu, Cancellation_Date S

WHERE C.Customer_ID = Cu.Customer_ID and C.Cancellation_Date = S.Cancellation_Date

GROUP BY City order by Cancellations DESC;
```

Results | Messages

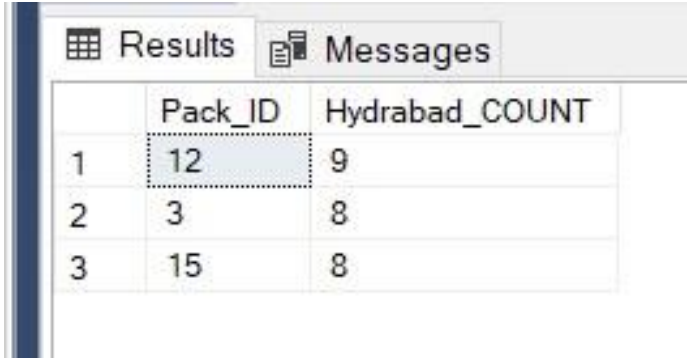
	City	Cancellations
1	Chennai	7
2	Hydrabad	7
3	Kolkata	6
4	Delhi	5
5	kanpur	4

```
/*  List of  5 cities with lowest subscription */

SELECT  top 5 Customer_Dimension.City,COUNT(Customer_Dimension.Customer_ID)

AS subscription_count

FROM Customer_Dimension

INNER JOIN Customer_Fact_Table ON Customer_Fact_Table.Customer_ID = Customer_Dimension.Customer_ID

GROUP BY Customer_Dimension.City

ORDER BY subscription_count ASC;
```

Results | Messages

	City	subscription_count
1	Hydrabad	149
2	Kolkata	149
3	Chennai	149
4	kanpur	150
5	Delhi	156

```
/*  SELECT the 3 top-selling pack in particular city */


SELECT top 3 Customer_Fact_Table.Pack_ID,COUNT(Customer_Fact_Table.Customer_ID) AS Hydrabad_COUNT

FROM Customer_Fact_Table

INNER JOIN Customer_Dimension ON Customer_Fact_Table.Customer_ID = Customer_Dimension.Customer_ID

 WHERE Customer_Dimension.City = 'Hydrabad'

GROUP BY Customer_Fact_Table.Pack_ID

ORDER BY Hydrabad_COUNT DESC;
```



Results | Messages

	Pack_ID	Hydrabad_COUNT
1	12	9
2	3	8
3	15	8

```
/* Show payment history of a particular customer */
SELECT cf.Customer_ID, cd.Customer_Name, cf.Payment_Date, cf.Total_Amount from Customer_Fact_Table

 AS cf,Customer_Dimension AS cd

WHERE  cf.Customer_ID = cd.Customer_ID and cf.Customer_ID = 285
```

Results | Messages

	Customer_ID	Customer_Name	Payment_Date	Total_Amount
1	285	Jamalia Fitzpatrick	2020-08-09	360.00
2	285	Jamalia Fitzpatrick	2020-03-01	401.00
3	285	Jamalia Fitzpatrick	2021-01-03	725.00
4	285	Jamalia Fitzpatrick	2020-03-25	194.00
5	285	Jamalia Fitzpatrick	2020-01-20	390.00
6	285	Jamalia Fitzpatrick	2020-10-22	407.00

```
/*  Find the model that has the highest quantity sold */

SELECT top 3 STB_Model_No,COUNT(Customer_ID) AS No_of_sell

FROM Customer_Dimension

GROUP BY STB_Model_No
```

Results | Messages

	STB_Model_No	No_of_sell
1	B4O1T7	96
2	E3P5U7	85
3	F8W5L5	71

```
/*  Which Month has the most new subscription */

 SELECT top 3 Month(Subscription_Date)

as Month_Subscription, COUNT(Customer_ID)

AS No_of_New_Sub FROM Customer_Fact_Table

GROUP BY Month(Subscription_Date) ORDER BY No_of_New_Sub DESC
```

Results | Messages

	Month_Subscription	No_of_New_Sub
1	2	163
2	5	157
3	3	156

CONTRIBUTIONS

S.NO	TASK	ANIKET MOURYA	PARTH PANDYA	VISHAL KUMAR JHA
1	SCHEMA DESIGN	25	25	50
2	TABLE & SQL QUERY	40	40	20
3	DATASET GENERATION	30	60	10
4	PPT & DOCUMENTATION	40	10	50

THANK YOU