**DATA WAREHOUSE**

**(SS G515) PROJECT REPORT**



**Milk Marketing Company (Amul)**

Under the supervision of

**L Rajya Lakshmi**

**Assistant Professor**

**Department of Computer Science and Information Systems**

**TEAM 07**

**Sidharth Saxena (2022H1120282P, h20220282@pilani.bits-pilani.ac.in)**

**Shubham Shrivastav (2022H1120283P, h20220283@pilani.bits-pilani.ac.in)**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Title** | **Page No.** |
| 1 | Introduction | 3 |
| 2 | Information Package Diagrams | 4-5 |
| 3 | STAR Schema | 6-11 |
| 4 | Business Queries | 12-19 |

**INTRODUCTION**

**WHAT AND WHY ?**

A data warehouse is a large, central repository of data that is used to support business intelligence and decision-making processes. It is a system designed to store and manage data from various sources, such as transactional systems, operational databases, and other external sources.

The purpose of a data warehouse is to provide a single, unified view of data that is consistent and reliable. This allows organizations to analyze and extract insights from large amounts of data to support their decision-making processes. By storing historical data, a data warehouse enables organizations to identify trends and patterns over time, which can help in forecasting and planning.

Overall, a data warehouse plays a crucial role in helping organizations to leverage their data assets effectively, gain insights into their operations and customers, and make informed decisions.

**PROBLEM STATEMENT**

Design a data warehouse for a milk marketing company akin to Amul with a well-defined STAR Schema, information package diagrams and clearly show the business queries implemented on the warehouse.

Structure, store and analyze data associated with :

1. Procurement of milk
2. Manufacturing of milk products from the procured milk
3. Distribution of manufactured milk products to :
   1. Dedicated company outlets
   2. Milk products distributors

**BUSINESS REQUIREMENTS**

To extract strategic information such as :-

1. Calculating net profit by considering losses incurred on unsold items.
2. Calculating net profit that would have been yielded had all items that were manufactured were sold (What if? Analysis and Benchmarking).
3. Determining which items sell the least at dedicated company outlets with respect to quantity, amount, profit generated.
4. Calculating the profit earned at each store on a quarterly basis for a given year.
5. Calculating the difference in quantities of all products produced and sold.
6. Determining the top suppliers that provided the highest total quantity of milk in a given year.
7. Determining which plants have the highest and lowest production quantity in a given year.

**INFORMATION PACKAGE DIAGRAMS**

**Milk Procurement**

|  |  |  |  |
| --- | --- | --- | --- |
| **Time** | **Location** | **Supplier** | **Procurement Product** |
| Year | State | Name | Fat Content |
| Quarter | City | Address | Cost |
| Month | Pincode | Phone Number | Type |
| Date |  | Email |  |
| Day Number |  |  |  |
| **Facts :** Quantity, Procurement Cost | | | |

**Milk Products Manufacturing**

|  |  |  |  |
| --- | --- | --- | --- |
| **Time** | **Location** | **Plant** | **Product** |
| Year | State | Name | Name |
| Quarter | City | Address | Type |
| Month | Pincode |  | Sub-type |
| Date |  |  | Selling Price |
| Day Number |  |  | Cost |
|  |  |  | Shelf Life |
| **Facts :** Quantity, Manufacturing Cost | | | |

**Milk Products Distribution**

**Dedicated Store Outlets**

|  |  |  |  |
| --- | --- | --- | --- |
| **Time** | **Location** | **Store** | **Product** |
| Year | State | Name | Name |
| Quarter | City | Address | Type |
| Month | Pincode | Phone Number | Sub-type |
| Date |  | Email | Selling Price |
| Day Number |  |  | Cost |
|  |  |  | Shelf Life |
| **Facts :** Quantity, Sales, Profit | | | |

**Milk Products Distributors**

|  |  |  |  |
| --- | --- | --- | --- |
| **Time** | **Location** | **Distributor** | **Product** |
| Year | State | Name | Name |
| Quarter | City | Address | Type |
| Month | Pincode | Phone Number | Sub-type |
| Date |  | Email | Selling Price |
| Day Number |  |  | Cost |
|  |  |  | Shelf Life |
| **Facts :** Quantity, Sales, Profit | | | |

**STAR SCHEMA**

**Schema Diagram**



**Schema Definition**

|  |
| --- |
| **LEGEND** |
| **Dimension Table** |
| **Fact Table** |
| **Bridge Table** |

**Common Dimension Tables :-**

1. **Time\_Dim Table :** This table stores the dimension of time and is used in all fact tables. This table contains 2 years’ worth of data (2021-22). Its columns are :
   1. Time\_ID
   2. Date
   3. Day (Mon/Tue/Wed/Thu/Fri/Sat/Sun)
   4. Day\_Number (1-730)
   5. Month (1-12)
   6. Quarter (1-4)
   7. Year
   8. Holiday\_Flag (1-Holiday, 0-Working Day)
2. **Location\_Dim Table :** This table stores the dimension of location and is used in all fact tables. This table contains 4 cities having 4 pincodes each. Its columns are :
   1. Location\_ID
   2. City
   3. State
   4. Pincode
3. **Milk Procurement Tables :-**
   1. **Procurement\_Prod\_Dim Table :** This table stores the dimension of the raw material (milk in this case) and is used in the procurement fact table. This table contains 10 rows. Its columns are :
      1. Procurement\_Prod\_ID
      2. Fat\_Content (in %)
      3. Cost (per L)
      4. Type (Cow/Buffalo)
   2. **Supplier\_Dim Table :** This table stores the dimension of the suppliers providing the raw material (milk in this case) and is used in the procurement fact table. This table contains 40 rows. Its columns are :
      1. Supplier\_ID
      2. Location\_ID
      3. Supplier\_Name
      4. Supplier\_Address
      5. Supplier\_Phone\_Number
      6. Supplier\_Email
   3. **Procurement\_Fact\_Table :** This table stores the facts associated with raw material (milk in this case) procurement. This table contains 200 rows. Its columns are :
      1. Procurement\_ID
      2. Supplier\_ID
      3. Procurement\_Prod\_ID
      4. Time\_ID
      5. Quantity (in L)
      6. Procurement\_Cost

**Table common for 2 & 3 :-**

**Product\_Dim Table :** This table stores the dimension of the products manufactured / to be manufactured from raw material. It is used in all bridge tables. This table contains 25 rows. Its columns are :

* + 1. Product\_ID
    2. Product\_Name
    3. Product\_Type (Cow Milk/Buffalo Milk /Chocolate/Ice-Cream/Butter/Cheese/Cream/Ghee/Curd)
    4. Product\_Sub\_Type (Milk - Pasteurised, Skimmed, Toned ; Chocolate - Milk, Dark ; Ice-Cream : Cone, Family Pack, Cup ; Butter : Salted, Unsalted ; Cream : Fresh Cream , Ghee : Cow Ghee, Ghee , Curd : Curd, Sweet )
    5. Product\_Selling\_Price
    6. Product\_Cost (Individual Product Cost)
    7. Shelf\_Life (In days)

1. **Manufacturing Tables :-**
   1. **Plant\_Dim Table :** This table stores the dimension of the manufacturing plants processing the raw material into the final milk products to be sold. This table contains 4 rows. Its columns are :
      1. Plant\_ID
      2. Plant\_Name
      3. Plant\_Address
      4. Location\_ID
   2. **Plant\_Prod\_Bridge Table :** This bridge table stores the mapping between manufacturing plants and products indicating which products are manufactured at which manufacturing plants and is used in the manufacturing fact table. This table contains 100 rows. Its columns are :
      1. Plant\_Prod\_ID
      2. Plant\_ID
      3. Product\_ID
   3. **Manufacturing\_Fact\_Table :** This table stores the facts associated with the manufacturing of milk products from raw material. This table contains 200 rows. Its columns are :
      1. Production\_ID
      2. Plant\_Prod\_ID
      3. Time\_ID
      4. Quantity (Units)
      5. Manufacturing\_Cost (Cost of total product amount produced at a plant)
2. **Distribution Tables :-**
   1. **Store\_Dim Table :** This table stores the dimension of the dedicated company outlets where the manufactured milk products are sold and is used in the store sales fact table. This table contains 40 rows. Its columns are :
      1. Store\_ID
      2. Location\_ID
      3. Store\_Name
      4. Store\_Address
      5. Store\_Phone\_Number
      6. Store\_Email
   2. **Distributor\_Dim Table :** This table stores the dimension of the Milk products distributors where the manufactured milk products are distributed and is used in the distributor sales fact table. This table contains 20 rows. Its columns are :
      1. Distributor\_ID
      2. Location\_ID
      3. Distributor\_Name
      4. Distributor\_Address
      5. Distributor\_Phone\_Number
      6. Distributor\_Email
   3. **Store\_Prod\_Bridge Table :** This bridge table stores the mapping between company stores and products indicating which products are sold at which store and is used in the store sales fact table. This table contains 1000 rows. Its columns are :
      1. Store\_Prod-ID
      2. Store\_ID
      3. Product\_ID
   4. **Distributor\_Prod\_Bridge Table :** This bridge table stores the mapping between distributors and products indicating which products are sold at which distributor and is used in the distributor sales fact table. This table contains 200 rows. Its columns are :
      1. Distributor\_Prod\_ID
      2. Distributor\_ID
      3. Product\_ID
   5. **Store\_Sales\_Fact\_Table :** This table stores the facts associated with the sales of milk products from different company outlets. This table contains 200 rows. Its columns are :
      1. Sales\_ID
      2. Store\_Prod\_ID
      3. Time\_ID
      4. Quantity (Units)
      5. Sales
      6. Profit
   6. **Distributor\_Sales\_Fact\_Table :** This table stores the facts associated with the sales of milk products to different milk distributors. This table contains 400 rows. Its columns are :
      1. Sales\_ID
      2. Distributor\_Prod\_ID
      3. Time\_ID
      4. Quantity (Units)
      5. Sales
      6. Profit

**BUSINESS QUERIES**

1. **Calculate Net profit by subtracting losses on unsold items from total profit earned.**

**Query :-**

DECLARE @Total\_Sales DECIMAL(10,2);

SET @Total\_Sales = (SELECT SUM(Sales) from Store\_Sales\_Fact\_Table) + (SELECT SUM(Sales) from Distributor\_Sales\_Fact\_Table)

-- Manufacturing cost involves the costs associated with both sold and unsold items as all items were manufactured prior to distribution

DECLARE @Total\_Manufacturing\_Cost DECIMAL(10,2);

SET @Total\_Manufacturing\_Cost = (SELECT SUM(Manufacturing\_Cost) from Manufacturing\_Fact\_Table)

-- Subtracting sales and manufacturing cost will also take the manufacturing costs associated with unsold items into account

SELECT @Total\_Sales-@Total\_Manufacturing\_Cost AS 'Net Profit'

**Output :-**

Text

Description automatically generated with medium confidence

1. **Calculate Net profit that would have been yielded had all items that were manufactured were sold (What if? Analysis and Benchmarking).**

**Query :-**

/\* CREATING TEMPORARY TABLE :- \*/

SELECT MFT.Quantity \* PD.Product\_Selling\_Price AS Benchmark\_Sales INTO #TEMP\_TABLE FROM Manufacturing\_Fact\_Table MFT INNER JOIN Plant\_Prod\_Bridge PPB ON MFT.Plant\_Prod\_ID = PPB.Plant\_Prod\_ID

INNER JOIN Product\_Dim PD ON PD.Product\_ID = PPB.Product\_ID

DECLARE @Total\_Sales\_Benchmark DECIMAL(10,2);

SET @Total\_Sales\_Benchmark = (SELECT SUM(Benchmark\_Sales) from #TEMP\_TABLE)

-- Manufacturing cost involves the costs associated with both sold and unsold items as all items were manufactured prior to distribution

DECLARE @Total\_Manufacturing\_Cost DECIMAL(10,2);

SET @Total\_Manufacturing\_Cost = (SELECT SUM(Manufacturing\_Cost) from Manufacturing\_Fact\_Table)

-- Subtracting sales and manufacturing cost will also take the manufacturing costs associated with unsold items into account

SELECT @Total\_Sales\_Benchmark-@Total\_Manufacturing\_Cost AS 'Net Profit (Benchmark)'

**Output :-**

A picture containing text

Description automatically generated

1. **Calculate Net profit by subtracting losses on unsold items from total profit earned per year.**

**Query :-**

/\* CREATING TEMPORARY TABLE FOR STORES :- \*/

SELECT SFT.Sales,TD.Year INTO #TEMP\_TABLE\_STORE\_YEARS FROM Store\_Sales\_Fact\_Table SFT INNER JOIN Time\_Dim TD ON SFT.Time\_ID=TD.Time\_ID

SELECT SUM(Sales) AS 'Yearly Sales',Year INTO #TEMP\_TABLE\_YEARLY\_STORE\_SALES FROM #TEMP\_TABLE\_STORE\_YEARS GROUP BY Year

/\* CREATING TEMPORARY TABLE FOR DISTRIBUTORS :- \*/

SELECT SFT.Sales,TD.Year INTO #TEMP\_TABLE\_DISTRIBUTOR\_YEARS FROM Distributor\_Sales\_Fact\_Table SFT INNER JOIN Time\_Dim TD ON SFT.Time\_ID=TD.Time\_ID

SELECT SUM(Sales) AS 'Yearly Sales',Year INTO #TEMP\_TABLE\_YEARLY\_DISTRIBUTOR\_SALES FROM #TEMP\_TABLE\_DISTRIBUTOR\_YEARS GROUP BY Year

/\* CREATING TEMPORARY TABLE FOR SUM :- \*/

SELECT Year, SUM([Yearly Sales]) AS TotalSales INTO #TEMP\_TABLE\_YEARLY\_TOTAL\_SALES

FROM (

    SELECT Year, [Yearly Sales]

    FROM #TEMP\_TABLE\_YEARLY\_DISTRIBUTOR\_SALES

    UNION ALL

    SELECT Year, [Yearly Sales]

    FROM #TEMP\_TABLE\_YEARLY\_STORE\_SALES

) AS CombinedSales

GROUP BY Year;

/\* CREATING TEMPORARY TABLE FOR MANUFACTURING COST :- \*/

SELECT MFT.Manufacturing\_Cost,TD.Year INTO #TEMP\_TABLE\_COST\_YEARS FROM Manufacturing\_Fact\_Table MFT INNER JOIN Time\_Dim TD ON MFT.Time\_ID=TD.Time\_ID

SELECT SUM(Manufacturing\_Cost) AS 'Yearly Cost',Year INTO #TEMP\_TABLE\_YEARLY\_COST FROM #TEMP\_TABLE\_COST\_YEARS GROUP BY Year

SELECT t1.Year, (t1.TotalSales - t2.[Yearly Cost]) AS 'Difference'

FROM #TEMP\_TABLE\_YEARLY\_TOTAL\_SALES t1

INNER JOIN #TEMP\_TABLE\_YEARLY\_COST t2 ON t1.Year = t2.Year

**Output :-**

Table

Description automatically generated with low confidence

1. **See which items sell the least at stores with respect to :-**
   1. **Quantity**
   2. **Amount**
   3. **Profit Generated**

**Queries :-**

--a.) Quantity

SELECT PD.Product\_Name, SUM(SFT.Quantity) AS Total\_Sales\_in\_Units FROM Store\_Sales\_Fact\_Table SFT INNER JOIN Store\_Prod\_Bridge SB ON SB.Store\_Prod\_ID = SFT.Store\_Prod\_ID INNER JOIN Product\_Dim PD ON SB.Product\_ID = PD.Product\_ID GROUP BY Product\_Name ORDER BY Total\_Sales\_in\_Units ASC;

--b.) Amount

SELECT PD.Product\_Name, SUM(SFT.Sales) AS Total\_Sales\_in\_INR FROM Store\_Sales\_Fact\_Table SFT INNER JOIN Store\_Prod\_Bridge SB ON SB.Store\_Prod\_ID = SFT.Store\_Prod\_ID INNER JOIN Product\_Dim PD ON SB.Product\_ID = PD.Product\_ID GROUP BY Product\_Name ORDER BY Total\_Sales\_in\_INR ASC;

--c.) Profit Generated

SELECT PD.Product\_Name, SUM(SFT.Profit) AS Total\_Profit\_in\_INR FROM Store\_Sales\_Fact\_Table SFT INNER JOIN Store\_Prod\_Bridge SB ON SB.Store\_Prod\_ID = SFT.Store\_Prod\_ID INNER JOIN Product\_Dim PD ON SB.Product\_ID = PD.Product\_ID GROUP BY Product\_Name ORDER BY Total\_Profit\_in\_INR ASC;

**Outputs :-**

Graphical user interface, table

Description automatically generated**a.)**

Table

Description automatically generated**b.)**

**c.)**

1. **What is the profit earned at each store on a quarterly basis for the year 2022 ?**

**Query :-**

SELECT SPB.Store\_ID, TD.Quarter, SUM(SSFT.Profit) AS 'Total Profit' INTO #TEMP\_TABLE FROM Store\_Sales\_Fact\_Table SSFT

INNER JOIN Time\_Dim TD ON SSFT.Time\_ID=TD.Time\_ID INNER JOIN Store\_Prod\_Bridge SPB ON SSFT.Store\_Prod\_ID=SPB.Store\_Prod\_ID

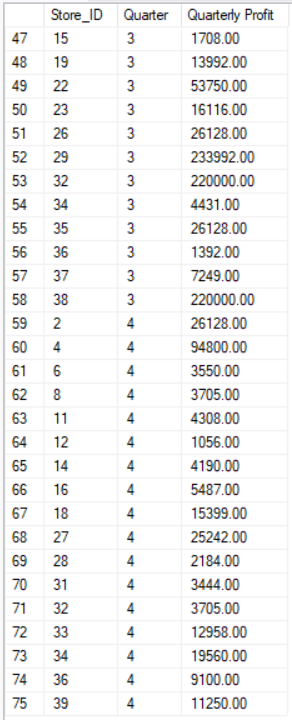
WHERE TD.Year = 2022 GROUP BY SPB.Store\_ID, TD.Quarter

SELECT Store\_ID, Quarter, SUM([Total Profit]) AS 'Quarterly Profit' FROM #TEMP\_TABLE GROUP BY Store\_ID,Quarter ORDER BY Quarter ASC

**Output :-**

Table

Description automatically generated with low confidenceA picture containing table

Description automatically generated

1. **Determine the quantities of all products produced vs the quantities of all products sold.**

**Query :-**

-- Products produced

SELECT PD.Product\_ID, PD.Product\_Name, SUM(MFT.Quantity) AS 'Quantity Produced (in units)' FROM Manufacturing\_Fact\_Table MFT INNER JOIN Plant\_Prod\_Bridge PPB ON MFT.Plant\_Prod\_ID = PPB.Plant\_Prod\_ID

INNER JOIN Product\_Dim PD ON PD.Product\_ID = PPB.Product\_ID GROUP BY PD.Product\_ID,PD.Product\_Name

-- Products Sold at stores

SELECT PD.Product\_ID, PD.Product\_Name, SUM(SFT.Quantity) AS 'Quantity Sold (in units)' FROM Store\_Sales\_Fact\_Table SFT INNER JOIN Store\_Prod\_Bridge SPB ON SFT.Store\_Prod\_ID = SPB.Store\_Prod\_ID

INNER JOIN Product\_Dim PD ON PD.Product\_ID = SPB.Product\_ID GROUP BY PD.Product\_ID,PD.Product\_Name

-- Products Sold to distributors

SELECT PD.Product\_ID, PD.Product\_Name, SUM(DFT.Quantity) AS 'Quantity Sold (in units)' FROM Distributor\_Sales\_Fact\_Table DFT INNER JOIN Distributor\_Prod\_Bridge DPB ON DFT.Distributor\_Prod\_ID = DPB.Distributor\_Prod\_ID

INNER JOIN Product\_Dim PD ON PD.Product\_ID = DPB.Product\_ID GROUP BY PD.Product\_ID,PD.Product\_Name

**Output :-**

Table

Description automatically generatedGraphical user interface, table

Description automatically generatedGraphical user interface, application, table

Description automatically generated

1. **Determine the top 5 suppliers that provided the highest total quantity of products in 2022.**

**Query :-**

SELECT TOP 5

    Supplier\_Name,

    SUM(Quantity) AS Total\_Quantity

FROM Procurement\_Fact\_Table

JOIN Supplier\_Dim ON Procurement\_Fact\_Table.Supplier\_ID = Supplier\_Dim.Supplier\_ID INNER JOIN Time\_Dim TD ON Procurement\_Fact\_Table.Time\_ID = TD.Time\_ID

WHERE TD.Year = 2022

GROUP BY Supplier\_Name

ORDER BY 2 DESC;

**Output :-**

Text, table

Description automatically generated with medium confidence

1. **Which plant has the highest and lowest production quantity in 2021?**

**Query :-**

SELECT TOP 1

    Plant\_Prod\_Bridge.Plant\_ID,

    SUM(Manufacturing\_Fact\_Table.Quantity) AS Total\_Quantity

FROM Manufacturing\_Fact\_Table

JOIN Plant\_Prod\_Bridge ON Manufacturing\_Fact\_Table.Plant\_Prod\_ID = Plant\_Prod\_Bridge.Plant\_Prod\_ID

JOIN Time\_Dim ON Manufacturing\_Fact\_Table.Time\_ID = Time\_Dim.Time\_ID

WHERE Time\_Dim.Year = 2021

GROUP BY Plant\_Prod\_Bridge.Plant\_ID

ORDER BY Total\_Quantity DESC; -- plant with highest production quantity

SELECT TOP 1

    Plant\_Prod\_Bridge.Plant\_ID,

    SUM(Manufacturing\_Fact\_Table.Quantity) AS Total\_Quantity

FROM Manufacturing\_Fact\_Table

JOIN Plant\_Prod\_Bridge ON Manufacturing\_Fact\_Table.Plant\_Prod\_ID = Plant\_Prod\_Bridge.Plant\_Prod\_ID

JOIN Time\_Dim ON Manufacturing\_Fact\_Table.Time\_ID = Time\_Dim.Time\_ID

WHERE Time\_Dim.Year = 2021

GROUP BY Plant\_Prod\_Bridge.Plant\_ID

ORDER BY Total\_Quantity ASC; -- plant with lowest production quantity

**Output :-**

Graphical user interface

Description automatically generated with medium confidenceA picture containing graphical user interface

Description automatically generated