VISHWAKARMA
UNIVERSITY
Maximising Human Potential

T. Y. B. Tech Computer Engineering
2024-2025
Pursued in Department of Computer Engineering Faculty of Science & Technology
Vishwakarma University, Pune-411048

NAME	**Swanand Mangesh Garge**
YEAR	**THIRD YEAR**
DIV	**D**
BATCH	**D2**
ROLL NO	**39**
SRN NO	**202201589**
PRN NO	**2280030433**
COURSE NAME	**DATAWAREHOUSE & DATA MINING LAB**
COURSE CODE	**BTECCE22509**
COURSE TEACHER NAME	**PROF. RAHUL PAPALKAR**

ASSIGNMENT NO: 3

Problem Statement:

Consider a vaccination management operational database that tracks different vaccine types, doses administered, sessions, sites, and adverse events following immunization (AEFI). It also tracks demographic information such as age group, gender, and location by state. You have to design a data warehouse that will be updated from the operational database and should support decision-making by helping to answer analytical questions about the total doses administered, sessions conducted, individuals vaccinated, and AEFI reported per vaccine type, age group, gender, date, and state.

THEORY:

- **Star Schema:**

- A type of database schema that organizes data into fact and dimension tables.
- The fact table is at the center, and it connects to dimension tables, which "radiate" out like the points of a star.
- It is simple, easy to understand, and supports efficient queries.

- **Fact Table:**

- Stores quantitative data (measurable metrics) related to business processes, like sales, revenue, or performance metrics.
- Each record in a fact table is a combination of foreign keys to dimension tables and facts (numeric data).
- Typically, it has many rows and fewer columns.

- **Dimension Table:**

- Contains descriptive attributes (context) about the facts, like product details, time, location, or customer information.

- Dimension tables are used to filter, group, and label the data in the fact table.
- They usually have fewer rows but more columns compared to fact tables.

State_dim

State_Key	State_Name	Region
1	Maharashtra	Western
2	Tamil Nadu	Southern
3	West Bengal	Eastern
4	Gujarat	Western
5	Karnataka	Southern
NULL	NULL	NULL

Date_dim

Date_Key	Date	Day	Month	Year	Quarter
1	2021-01-01	1	1	2021	1
2	2021-02-01	1	2	2021	1
3	2021-03-01	1	3	2021	1
4	2021-04-01	1	4	2021	2
5	2021-05-01	1	5	2021	2

Age_group_dim

Age_Group_Key	Age_Group
1	0-18 years
2	19-30 years
3	31-45 years
4	46-60 years
5	60+ years
NULL	NULL

Age__dim

Gender_Key	Gender
1	Male
2	Female
3	Other
4	Non-binary
5	Prefer not to say

Vaccine_Type_Key	Vaccine_Type
1	Covaxin
2	Covishield
3	Sputnik V
4	Pfizer
5	Moderna
NULL	NULL

Vaccination_fact

Vaccination_ID	State_Key	Date_Key	Vaccine_Type_Key	Gender_Key	Age_Group_Key
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	4
5	5	5	5	5	5
NULL	NULL	NULL	NULL	NULL	NULL

Total_Doses_Administered	Sessions	Sites	First_Dose_Administered	Second_Dose_Administered
5000	50	10	2500	2500
10000	100	20	5000	5000
7500	75	15	3750	3750
12000	120	25	6000	6000
8000	80	18	4000	4000
NULL	NULL	NULL	NULL	NULL

AEFI	Total_Individuals_Vaccinated
5	5000
10	10000
7	7500
12	12000
8	8000
NULL	NULL

SOURCE CODE:

```
CREATE DATABASE vaccinedb;
USE vaccinedb;


CREATE TABLE state_dim (
    State_Key INT PRIMARY KEY,
    State_Name VARCHAR(100),
    Region VARCHAR(100)
);

CREATE TABLE date_dim (
    Date_Key INT PRIMARY KEY,
    Date DATE,
    Day INT,
    Month INT,
    Year INT,
    Quarter INT
);

-- Create age group dimension table
CREATE TABLE age_group_dim (
    Age_Group_Key INT PRIMARY KEY,
    Age_Group VARCHAR(50)
);

-- Create gender dimension table
CREATE TABLE gender_dim (
    Gender_Key INT PRIMARY KEY,
    Gender VARCHAR(50)
);

-- Create vaccine type dimension table
CREATE TABLE vaccine_type_dim (
    Vaccine_Type_Key INT PRIMARY KEY,
    Vaccine_Type VARCHAR(100)
);

-- Create vaccination fact table after all dimension tables
CREATE TABLE vaccination_fact (
    Vaccination_ID INT PRIMARY KEY,
    State_Key INT,
    Date_Key INT,
    Vaccine_Type_Key INT,
    Gender_Key INT,
```

```
    Age_Group_Key INT,
    Total_Doses_Administered INT,
    Sessions INT,
    Sites INT,
    First_Dose_Administered INT,
    Second_Dose_Administered INT,
    AEFI INT,
    Total_Individuals_Vaccinated INT,
    FOREIGN KEY (State_Key) REFERENCES state_dim(State_Key),
    FOREIGN KEY (Date_Key) REFERENCES date_dim(Date_Key),
    FOREIGN KEY (Vaccine_Type_Key) REFERENCES vaccine_type_dim(Vaccine_Type_Key),
    FOREIGN KEY (Gender_Key) REFERENCES gender_dim(Gender_Key),
    FOREIGN KEY (Age_Group_Key) REFERENCES age_group_dim(Age_Group_Key)
);
```

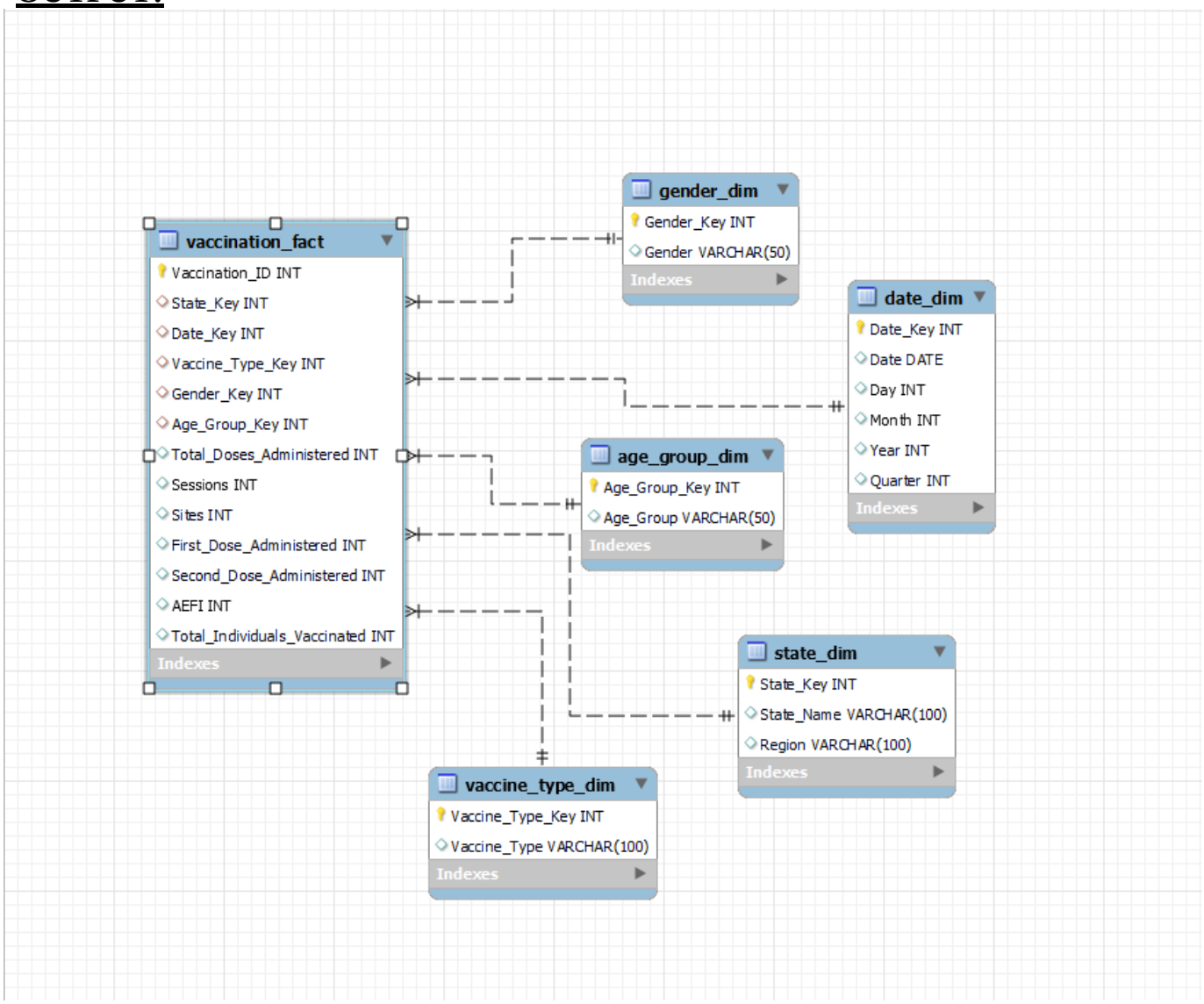
```
-- Insert values into state_dim
INSERT INTO state_dim (State_Key, State_Name, Region)
VALUES
(1, 'Maharashtra', 'Western'),
(2, 'Tamil Nadu', 'Southern'),
(3, 'West Bengal', 'Eastern'),
(4, 'Gujarat', 'Western'),
(5, 'Karnataka', 'Southern');

select * from state_dim;
```

```
-- Insert values into date_dim
INSERT INTO date_dim (Date_Key, Date, Day, Month, Year, Quarter)
VALUES
(1, '2021-01-01', 1, 1, 2021, 1),
(2, '2021-02-01', 1, 2, 2021, 1),
(3, '2021-03-01', 1, 3, 2021, 1),
(4, '2021-04-01', 1, 4, 2021, 2),
(5, '2021-05-01', 1, 5, 2021, 2);

select * from date_dim;
```

```
-- Insert values into age_group_dim
INSERT INTO age_group_dim (Age_Group_Key, Age_Group)
VALUES
(1, '0-18 years'),
```

```
(2, '19-30 years'),
(3, '31-45 years'),
(4, '46-60 years'),
(5, '60+ years');

select * from age_group_dim;

-- Insert values into gender_dim
INSERT INTO gender_dim (Gender_Key, Gender)
VALUES
(1, 'Male'),
(2, 'Female'),
(3, 'Other'),
(4, 'Non-binary'),
(5, 'Prefer not to say');

select * from gender_dim;

-- Insert values into vaccine_type_dim
INSERT INTO vaccine_type_dim (Vaccine_Type_Key, Vaccine_Type)
VALUES
(1, 'Covaxin'),
(2, 'Covishield'),
(3, 'Sputnik V'),
(4, 'Pfizer'),
(5, 'Moderna');
select * from vaccine_type_dim;

-- Insert values into vaccination_fact with matching Date_Key values
INSERT INTO vaccination_fact (Vaccination_ID, State_Key, Date_Key, Vaccine_Type_Key,
Gender_Key, Age_Group_Key, Total_Doses_Administered, Sessions, Sites, First_Dose_Administered,
Second_Dose_Administered, AEFI, Total_Individuals_Vaccinated)
VALUES
(1, 1, 1, 1, 1, 1, 1000, 100, 10, 4000, 1000, 10, 1000),
(2, 2, 2, 2, 2, 2, 10000, 100, 20, 5000, 5000, 10, 10000),
(3, 3, 3, 3, 3, 3, 7500, 75, 15, 3750, 3750, 7, 7500),
(4, 4, 4, 4, 4, 4, 12000, 120, 25, 6000, 6000, 12, 12000),
(5, 5, 5, 5, 5, 5, 8000, 80, 18, 4000, 4000, 8, 8000);


SELECT * FROM vaccination_fact;
```

OUTPUT:


vaccination_fact
Vaccination_ID INT
State_Key INT
Date_Key INT
Vaccine_Type_Key INT
Gender_Key INT
Age_Group_Key INT
Total_Doses_Administered INT
Sessions INT
Sites INT
First_Dose_Administered INT
Second_Dose_Administered INT
AEFI INT
Total_Individuals_Vaccinated INT
Indexes
gender_dim
Gender_Key INT
Gender VARCHAR(50)
Indexes
date_dim
Date_Key INT
Date DATE
Day INT
Month INT
Year INT
Quarter INT
Indexes
age_group_dim
Age_Group_Key INT
Age_Group VARCHAR(50)
Indexes
state_dim
State_Key INT
State_Name VARCHAR(100)
Region VARCHAR(100)
Indexes
vaccine_type_dim
Vaccine_Type_Key INT
Vaccine_Type VARCHAR(100)
Indexes

CONCLUSION:

The design of this data warehouse schema provides a well-structured and effective framework for managing and analyzing vaccination data. By implementing a star schema with a central fact table (vaccination_fact) surrounded by dimension tables (state_dim, date_dim, age_group_dim, gender_dim, and vaccine_type_dim), it ensures efficient querying and aggregation of key metrics such as the total doses administered, sessions held, and individuals vaccinated across various dimensions like state, vaccine type, and demographic categories.

The relationships between the fact and dimension tables enable comprehensive, multi-dimensional analysis. This structure enhances the ability to understand vaccination performance across different aspects, including time, geographic regions, age groups, gender, and vaccine types. Organizing data in this manner allows the data warehouse to support informed decision-making, offering stakeholders valuable insights into critical public health areas like vaccination coverage, demographic targeting, and regional distribution.

This schema not only ensures data consistency and integrity through enforced foreign key relationships but also empowers users to perform complex analytical queries with ease. It serves as a robust foundation for tracking and analyzing vaccination efforts, ultimately aiding in the effective management and optimization of vaccination campaigns.