

```
CREATE DATABASE IF NOT EXISTS task9_star_schema;
USE task9_star_schema;
-- Table name: globalsuperstore
-- DIMENSION TABLES
CREATE TABLE dim_customer (
    customer_key INT AUTO_INCREMENT PRIMARY KEY,
    customer_id VARCHAR(50),
    customer_name VARCHAR(100),
    segment VARCHAR(50)
);
CREATE TABLE dim_product (
    product_key INT AUTO_INCREMENT PRIMARY KEY,
    product_id VARCHAR(50),
    product_name VARCHAR(150),
    category VARCHAR(50),
    sub_category VARCHAR(50)
);
CREATE TABLE dim_region (
    region_key INT AUTO_INCREMENT PRIMARY KEY,
    country VARCHAR(50),
    region VARCHAR(50),
    state VARCHAR(50),
    city VARCHAR(50),
    postal_code VARCHAR(20)
);
CREATE TABLE dim_date (
    date_key INT AUTO_INCREMENT PRIMARY KEY,
    order_date DATE,
    year INT,
    quarter INT,
```

```

month INT,
month_name VARCHAR(15),
day INT,
weekday VARCHAR(15)

);

-- LOAD DIMENSIONS

INSERT INTO dim_customer (customer_id, customer_name, segment)
SELECT DISTINCT `Customer ID`, `Customer Name`, Segment
FROM globalsuperstore;

INSERT INTO dim_product (product_id, product_name, category, sub_category)
SELECT DISTINCT `Product ID`, `Product Name`, Category, `Sub-Category`
FROM globalsuperstore;

INSERT INTO dim_region (country, region, state, city, postal_code)
SELECT DISTINCT Country, Region, State, City, `Postal Code`
FROM globalsuperstore;

INSERT INTO dim_date (order_date, year, quarter, month, month_name, day, weekday)
SELECT DISTINCT
STR_TO_DATE(`Order Date`, '%m/%d/%Y'),
YEAR(STR_TO_DATE(`Order Date`, '%m/%d/%Y')),
QUARTER(STR_TO_DATE(`Order Date`, '%m/%d/%Y')),
MONTH(STR_TO_DATE(`Order Date`, '%m/%d/%Y')),
MONTHNAME(STR_TO_DATE(`Order Date`, '%m/%d/%Y')),
DAY(STR_TO_DATE(`Order Date`, '%m/%d/%Y')),
DAYNAME(STR_TO_DATE(`Order Date`, '%m/%d/%Y'))

FROM globalsuperstore;

FACT TABLE

CREATE TABLE fact_sales (
sales_key INT AUTO_INCREMENT PRIMARY KEY,
order_id VARCHAR(50),
customer_key INT,

```

```
product_key INT,  
region_key INT,  
date_key INT,  
sales DECIMAL(10,2),  
quantity INT,  
discount DECIMAL(5,2),  
profit DECIMAL(10,2),  
FOREIGN KEY (customer_key) REFERENCES dim_customer(customer_key),  
FOREIGN KEY (product_key) REFERENCES dim_product(product_key),  
FOREIGN KEY (region_key) REFERENCES dim_region(region_key),  
FOREIGN KEY (date_key) REFERENCES dim_date(date_key)  
);
```

-- LOAD FACT TABLE

```
INSERT INTO fact_sales (
```

```
    order_id, customer_key, product_key, region_key, date_key,  
    sales, quantity, discount, profit
```

```
)
```

SELECT

```
    g.`Order ID`,
```

```
    c.customer_key,
```

```
    p.product_key,
```

```
    r.region_key,
```

```
    d.date_key,
```

```
    g.Sales,
```

```
    g.Quantity,
```

```
    g.Discount,
```

```
    g.Profit
```

FROM globalsuperstore g

```
JOIN dim_customer c ON g.`Customer ID` = c.customer_id
```

```
JOIN dim_product p ON g.`Product ID` = p.product_id
```

```

JOIN dim_region r
ON g.Country = r.country
AND g.Region = r.region
AND g.State = r.state
AND g.City = r.city
JOIN dim_date d
ON STR_TO_DATE(g.`Order Date`, '%m/%d/%Y') = d.order_date;

```

INDEXES

```

CREATE INDEX idx_fact_customer ON fact_sales(customer_key);
CREATE INDEX idx_fact_product ON fact_sales(product_key);
CREATE INDEX idx_fact_region ON fact_sales(region_key);
CREATE INDEX idx_fact_date ON fact_sales(date_key);

```

2..star_schema_diagram

```

Table dim_customer {
    customer_key int [pk]
    customer_id varchar
    customer_name varchar
    segment varchar
}

```

```

Table dim_product {
    product_key int [pk]
    product_id varchar
    product_name varchar
    category varchar
    sub_category varchar
}

```

```

Table dim_region {

```

```
region_key int [pk]
country varchar
region varchar
state varchar
city varchar
postal_code varchar
}
```

```
Table dim_date {
    date_key int [pk]
    order_date date
    year int
    quarter int
    month int
    month_name varchar
    day int
    weekday varchar
}
```

```
Table fact_sales {
    sales_key int [pk]
    order_id varchar
    customer_key int
    product_key int
    region_key int
    date_key int
    sales decimal
    quantity int
    discount decimal
    profit decimal
}
```

}

Ref: fact_sales.customer_key > dim_customer.customer_key

Ref: fact_sales.product_key > dim_product.product_key

Ref: fact_sales.region_key > dim_region.region_key

Ref: fact_sales.date_key > dim_date.date_key

analysis_outputs.csv

-- 1. Total Sales by Year

```
SELECT d.year, SUM(f.sales) AS total_sales
```

```
FROM fact_sales f
```

```
JOIN dim_date d ON f.date_key = d.date_key
```

```
GROUP BY d.year;
```

-- 2. Profit by Category

```
SELECT p.category, SUM(f.profit) AS total_profit
```

```
FROM fact_sales f
```

```
JOIN dim_product p ON f.product_key = p.product_key
```

```
GROUP BY p.category;
```

-- 3. Sales by Region

```
SELECT r.region, SUM(f.sales) AS total_sales
```

```
FROM fact_sales f
```

```
JOIN dim_region r ON f.region_key = r.region_key
```

```
GROUP BY r.region;
```

-- 4. Top 10 Customers by Sales

```
SELECT c.customer_name, SUM(f.sales) AS total_sales
```

```
FROM fact_sales f
```

```
JOIN dim_customer c ON f.customer_key = c.customer_key
```

```
GROUP BY c.customer_name
```

ORDER BY total_sales DESC

LIMIT 10;