**Task Description**

**You need to use the same scripts from the NORTHWIND database that you worked with previously.**

**The task is divided into several steps, which must be completed in order.**

**Design Staging Tables and a Star Schema**

**The staging tables will be used for loading the initial data, and the star schema will be organized around a fact table and related dimension tables.**

***To start, we’ll create the staging tables and drop them if they already exist. Here’s the SQL script for creating each of the required staging tables.***

***Step 1: Create Staging Tables***

*-- Drop staging tables if they exist*

*DROP TABLE IF EXISTS northwind.staging\_orders;*

*DROP TABLE IF EXISTS northwind.staging\_order\_details;*

*DROP TABLE IF EXISTS northwind.staging\_products;*

*DROP TABLE IF EXISTS northwind.staging\_customers;*

*DROP TABLE IF EXISTS northwind.staging\_employees;*

*DROP TABLE IF EXISTS northwind.staging\_categories;*

*DROP TABLE IF EXISTS northwind.staging\_shippers;*

*DROP TABLE IF EXISTS northwind.staging\_suppliers;*

*-- Create staging tables*

*CREATE TABLE northwind.staging\_orders AS TABLE northwind.orders WITH NO DATA;*

*CREATE TABLE northwind.staging\_order\_details AS TABLE northwind.order\_details WITH NO DATA;*

*CREATE TABLE northwind.staging\_products AS TABLE northwind.products WITH NO DATA;*

*CREATE TABLE northwind.staging\_customers AS TABLE northwind.customers WITH NO DATA;*

*CREATE TABLE northwind.staging\_employees AS TABLE northwind.employees WITH NO DATA;*

*CREATE TABLE northwind.staging\_categories AS TABLE northwind.categories WITH NO DATA;*

*CREATE TABLE northwind.staging\_shippers AS TABLE northwind.shippers WITH NO DATA;*

*CREATE TABLE northwind.staging\_suppliers AS TABLE northwind.suppliers WITH NO DATA;*

1. Create the following staging tables:

*staging\_orders*

*staging\_order\_details*

*staging\_products*

*staging\_customers*

*staging\_employees*

*staging\_categories*

*staging\_shippers*

*staging\_suppliers*

Designing a star schema design involves creating dimension tables and a fact table.

2. Use the proposed set of dimension tables and their respective columns.

***Table - DimDate:***

DateID (Primary Key)

Date

Day

Month

Year

Quarter

WeekOfYear

***Table - DimCustomer:***

CustomerID (Primary Key)

CompanyName

ContactName

ContactTitle

Address

City

Region

PostalCode

Country

Phone

***Table - DimProduct:***

ProductID (Primary Key)

ProductName

SupplierID (FK)

CategoryID (FK)

QuantityPerUnit

UnitPrice

UnitsInStock

***Table - DimEmployee:***

EmployeeID (Primary Key)

LastName

FirstName

Title

BirthDate

HireDate

Address

City

Region

PostalCode

Country

HomePhone

Extension

***Table - DimCategory:***

CategoryID (Primary Key)

CategoryName

Description

***Table - DimShipper:***

ShipperID (Primary Key)

CompanyName

Phone

***Table - DimSupplier:***

SupplierID (Primary Key)

CompanyName

ContactName

ContactTitle

Address

City

Region

PostalCode

Country

Phone

And the table FactSales with the columns below:

*SalesID (Primary Key)*

*DateID (FK to Date Dimension)*

*CustomerID (FK to Customer Dimension)*

*ProductID (FK to Product Dimension)*

*EmployeeID (FK to Employee Dimension)*

*CategoryID (FK to Category Dimension)*

*ShipperID (FK to Shipper Dimension)*

*SupplierID (FK to Supplier Dimension)*

*QuantitySold*

*UnitPrice*

*Discount*

*TotalAmount (Calculated as QuantitySold \* UnitPrice - Discount)*

*TaxAmount*

**step-by-step through this data loading, transformation, and validation process.**

**Step 1: Load Data into Staging Tables**

**Assuming you have staging tables already created, load data from the source tables into each staging table. Here’s the SQL script:**

**-- Load data into staging tables**

**-- Customers**

**INSERT INTO northwind.staging\_customers**

**SELECT \* FROM northwind.customers;**

**-- Orders**

**INSERT INTO northwind.staging\_orders**

**SELECT \* FROM northwind.orders;**

**-- Order Details**

**INSERT INTO northwind.staging\_order\_details**

**SELECT \* FROM northwind.order\_details;**

**-- Products**

**INSERT INTO northwind.staging\_products**

**SELECT \* FROM northwind.products;**

**-- Employees**

**INSERT INTO northwind.staging\_employees**

**SELECT \* FROM northwind.employees;**

**-- Categories**

**INSERT INTO northwind.staging\_categories**

**SELECT \* FROM northwind.categories;**

**-- Shippers**

**INSERT INTO northwind.staging\_shippers**

**SELECT \* FROM northwind.shippers;**

**-- Suppliers**

**INSERT INTO northwind.staging\_suppliers**

**SELECT \* FROM northwind.suppliers;**

Load Data Into Staging, Transformation, and Star Schema

1. For each source table in the northwind\_pg database, you need to create a corresponding staging table and load data into it. Below is an example for the Customers table.

Assuming staging tables with the same structure as the source tables have already been created, load data into staging\_customer from the source Customers table.

INSERT INTO staging\_customers

SELECT \* FROM Customers;

2. Repeat this process for each table listed in step 1:

staging\_orders

staging\_order\_details

staging\_products

staging\_customers

staging\_employees

staging\_categories

staging\_shippers

staging\_suppliers

3. Transform the data from the staging tables and load it into the respective dimension tables. Here's an example for DimCustomer:

INSERT INTO DimCustomer (CustomerID, CompanyName, ContactName, ContactTitle, Address, City, Region, PostalCode, Country, Phone)

SELECT CustomerID, CompanyName, ContactName, ContactTitle, Address, City, Region, PostalCode, Country, Phone

FROM staging\_customers;

4. Repeat this process for other dimensions (DimProduct, DimEmployee, etc.), transforming the data as necessary.

5. Load it into the fact table, as shown below:

INSERT INTO FactSales (DateID, CustomerID, ProductID, EmployeeID, CategoryID, ShipperID, SupplierID, QuantitySold, UnitPrice, Discount, TotalAmount, TaxAmount)

SELECT

d.DateID,

c.CustomerID,

p.ProductID,

e.EmployeeID,

cat.CategoryID,

s.ShipperID,

sup.SupplierID,

od.Quantity,

od.UnitPrice,

od.Discount,

(od.Quantity \* od.UnitPrice - od.Discount) AS TotalAmount,

(od.Quantity \* od.UnitPrice - od.Discount) \* 0.1 AS TaxAmount

FROM staging\_order\_details od

JOIN staging\_orders o ON od.OrderID = o.OrderID

JOIN staging\_customers c ON o.CustomerID = c.CustomerID

JOIN staging\_products p ON od.ProductID = p.ProductID

LEFT JOIN staging\_employees e ON o.EmployeeID = e.EmployeeID

LEFT JOIN staging\_categories cat ON p.CategoryID = cat.CategoryID

LEFT JOIN staging\_shippers s ON o.ShipVia = s.ShipperID

LEFT JOIN staging\_suppliers sup ON p.SupplierID = sup.SupplierID

LEFT JOIN DimDate d ON o.OrderDate = d.Date;

6. After loading data into the fact and dimension tables, you should validate the data to ensure it is accurate and complete. This process typically involves:

\*Checking for data and referential integrity

\*Verifying that the data in the fact and dimension tables aligns with the source data

\*Ensuring that all records have been transferred correctly

### Step 2: Load Data into Dimension Tables

For each dimension table, we’ll select and transform data from the corresponding staging table and insert it into the dimension table. Below are the queries for each dimension table.

#### DimCustomer

INSERT INTO northwind.DimCustomer (CustomerID, CompanyName, ContactName, ContactTitle, Address, City, Region, PostalCode, Country, Phone)

SELECT CustomerID, CompanyName, ContactName, ContactTitle, Address, City, Region, PostalCode, Country, Phone

FROM northwind.staging\_customers;

INSERT INTO northwind.DimProduct (ProductID, ProductName, SupplierID, CategoryID, QuantityPerUnit, UnitPrice, UnitsInStock)

SELECT ProductID, ProductName, SupplierID, CategoryID, QuantityPerUnit, UnitPrice, UnitsInStock

FROM northwind.staging\_products;

…. Etc on sql files

### Step 3: Load Data into the Fact Table (FactSales)

**The FactSales table will pull data from various staging tables using joins to populate foreign keys and calculated fields. Here’s the script**

Prepare Business Report Queries

Now that you have built the DWH, create scripts (SQLs) to cover the following business requirements:

1.Display average sales (total amount, net amount, tax; number of transactions), the rolling average for three months (January–February; January–February–March; February–March–April) per day (specifying the month and date range) across all product categories (selected category, list of categories) in geographical sections (regions, countries, states), in gender sections (men, women), by age group (0–18, 19–28, 28–45, 45–60, 60+), by income (0–20000, 20001–40000, 40001–60000, 60001–80000, 80001-100000). This involves querying the FactSales and DimDate tables.

2.Display the top (worst) five products by number of transactions, total sales, and tax (add category section). This involves querying the FactSales table.

3.Display the top (worst) five customers by number of transactions and purchase amount (add gender section, region, country, product categories, age group). This involves querying the FactSales table.

4.Display a sales chart (with the total amount of sales and the quantity of items sold) for the first week of each month. This involves querying the FactSales and DimDate tables.

5.Display a weekly sales report (with monthly totals) by product category (period: one year). This involves querying the FactSales, DimDate, and DimProduct tables.

6.Display the median monthly sales value by product category and country. This involves querying the FactSales, DimProduct, and DimCustomer tables and requires a more complex query or a custom function to calculate the median.

7.Display sales rankings by product category (with the best-selling categories at the top). This involves querying the FactSales and DimProduct tables.

**Each query is structured to meet the specified business logic. Ensure you have the necessary tables and data in your DWH before running these queries.**

### 1. Average Sales with Rolling Average

**This query calculates average sales, including total amount, net amount, tax, and number of transactions over specified rolling periods.**

### 2. Top/Worst Five Products

**This query finds the top and worst five products by the number of transactions, total sales, and tax.**

### 3. Top/Worst Five Customers

**This query identifies the top and worst five customers based on the number of transactions and total purchase amounts.**

### 4. Sales Chart for the First Week of Each Month

This query generates a sales chart for the first week of each month.

### 5. Weekly Sales Report by Product Category

This query produces a weekly sales report by product category for the past year.

### 6. Median Monthly Sales Value by Product Category and Country

This query calculates the median sales value monthly by category and country.

### 7. Sales Rankings by Product Category

This query ranks product categories based on sales volume.