**Data Analysis and Integration**

Project

**Group 24:**

João Novo – 190113

João Paquete - 189477

**Data Warehouse tables**

DROP DATABASE IF EXISTS northwind\_dw;

CREATE DATABASE northwind\_dw;

USE northwind\_dw;

CREATE TABLE dim\_customer (

    CustomerID VARCHAR(5),

    CompanyName VARCHAR(40),

    City VARCHAR(15),

    Country VARCHAR(15),

    PRIMARY KEY (CustomerID)

);

CREATE TABLE dim\_product (

    ProductID INTEGER,

    ProductIdentifier INTEGER,

    ProductName VARCHAR(40),

    CategoryName VARCHAR(15),

    VERSION INT,

    DATE\_FROM DATETIME,

    DATE\_TO DATETIME,

    PRIMARY KEY (ProductID)

);

CREATE TABLE dim\_supplier (

    SupplierID INTEGER,

    CompanyName VARCHAR(40),

    City VARCHAR(15),

    Country VARCHAR(15),

    PRIMARY KEY (SupplierID)

);

CREATE TABLE dim\_shipper (

    ShipperID INTEGER,

    CompanyName VARCHAR(40),

    PRIMARY KEY (ShipperID)

);

CREATE TABLE dim\_time (

    TimeID DATE,

    YearID INTEGER,

    MonthID INTEGER,

    MonthName VARCHAR(255),

    DayID INTEGER,

    PRIMARY KEY (TimeID)

);

CREATE TABLE fact\_order (

    OrderID INTEGER,

    CustomerID VARCHAR(5),

    ProductID INTEGER,

    SupplierID INTEGER,

    ShipperID INTEGER,

    TimeID DATE,

    UnitPrice DECIMAL(10,4),

    Quantity SMALLINT(2),

    Discount REAL,

    Sales DOUBLE,

    PRIMARY KEY (OrderID),

    FOREIGN KEY (CustomerID) REFERENCES dim\_customer (CustomerID),

    FOREIGN KEY (ProductID) REFERENCES dim\_product (ProductID),

    FOREIGN KEY (SupplierID) REFERENCES dim\_supplier (SupplierID),

    FOREIGN KEY (ShipperID) REFERENCES dim\_shipper (ShipperID),

    FOREIGN KEY (TimeID) REFERENCES dim\_time (TimeID)

);

**Transformations**

Dim\_customer

Graphical user interface, text, application, Word

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Dim\_product

Diagram, text

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Dim\_Supplier

Graphical user interface, text, application, Word

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Dim\_shipper

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, Word

Description automatically generated

Graphical user interface, text, application, Word

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Dim\_time

Graphical user interface, text, application, Word, email

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application, Excel

Description automatically generated

Graphical user interface, application, Excel

Description automatically generated

Graphical user interface, text, application, table, Excel

Description automatically generated

Graphical user interface, text, application, table, Excel

Description automatically generated

Graphical user interface, application, Excel

Description automatically generated

Graphical user interface, application, Excel

Description automatically generated

Fact\_order

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, table

Description automatically generated

Graphical user interface, table

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

Graphical user interface, application, table, Excel

Description automatically generated

Graphical user interface, application, table, Excel

Description automatically generated

Job

Graphical user interface, text, application, email

Description automatically generated

**XML code for the cube definition**

<Schema name="northwind\_dw">

  <Cube name="Orders" visible="true" cache="true" enabled="true">

    <Table name="fact\_order">

    </Table>

    <Dimension type="StandardDimension" visible="true" foreignKey="CustomerID" highCardinality="false" name="Customer">

      <Hierarchy name="Customer Hierarchy" visible="true" hasAll="true" allMemberName="All Customers" primaryKey="CustomerID">

        <Table name="dim\_customer">

        </Table>

        <Level name="Country" visible="true" column="Country" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

        <Level name="City" visible="true" column="City" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

        <Level name="Company Name" visible="true" column="CompanyName" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="ProductID" highCardinality="false" name="Product">

      <Hierarchy name="Product Hierarchy" visible="true" hasAll="true" allMemberName="All Products" primaryKey="ProductID">

        <Table name="dim\_product">

        </Table>

        <Level name="Category Name" visible="true" column="CategoryName" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

        <Level name="Product Name" visible="true" column="ProductName" ordinalColumn="ProductName" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

        <Level name="Product Identifier" visible="true" column="ProductIdentifier" type="Integer" uniqueMembers="false" levelType="Regular">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="TimeDimension" visible="true" foreignKey="TimeID" highCardinality="false" name="Time">

      <Hierarchy name="Time Hierarchy" visible="true" hasAll="true" allMemberName="All years" primaryKey="TimeID">

        <Table name="dim\_time">

        </Table>

        <Level name="Year" visible="true" column="YearID" type="Integer" uniqueMembers="false" levelType="TimeYears" hideMemberIf="Never">

        </Level>

        <Level name="Month" visible="true" column="MonthName" ordinalColumn="MonthID" type="String" uniqueMembers="false" levelType="TimeMonths" hideMemberIf="Never">

          <Annotations>

            <Annotation name="AnalyzerDateFormat">

              <CDATA[[yyyy].[MMM]]]>

            </Annotation>

          </Annotations>

        </Level>

        <Level name="Day" visible="true" column="DayID" type="Integer" uniqueMembers="false" levelType="TimeDays" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="ShipperID" highCardinality="false" name="Shipper">

      <Hierarchy name="Shipper Hierarchy" visible="true" hasAll="true" allMemberName="All Shippers" primaryKey="ShipperID">

        <Table name="dim\_shipper">

        </Table>

        <Level name="Company Name" visible="true" column="CompanyName" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="SupplierID" highCardinality="false" name="Supplier">

      <Hierarchy name="Supplier Hierarchy" visible="true" hasAll="true" allMemberName="All Suplliers" primaryKey="SupplierID">

        <Table name="dim\_supplier">

        </Table>

        <Level name="Country" visible="true" column="Country" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

        <Level name="City" visible="true" column="City" type="String" uniqueMembers="false" levelType="Regular">

        </Level>

        <Level name="Company Name" visible="true" column="CompanyName" type="String" uniqueMembers="false" levelType="Regular">

        </Level>

      </Hierarchy>

    </Dimension>

    <Measure name="Quantity" column="Quantity" datatype="Integer" formatString="#,###" aggregator="sum" visible="true">

    </Measure>

    <Measure name="Sales" column="Sales" datatype="Numeric" formatString="$ #,###.00" aggregator="sum" visible="true">

    </Measure>

  </Cube>

</Schema>

**Analysis queries**

Query a) Analyze sales by customer country and year to discover the country, the year, and the pair country-year with the most sales.

The country with the most sales is USA.

The year in which the most sales happened is 1997.

The pair country-year with the most sales is USA – 1998.

![Graphical user interface, table

Description automatically generated

Query b) Analyze sales by product category and year to discover the category, the year, and the pair category-year with the most sales.

The product category with the most sales is Seafood.

The year in which the most sales happened is 1997.

The pair category-year with the most sales is Seafood – 1997.

Graphical user interface, application

Description automatically generated

Query c) Analyze quantity by shipping company and year to discover the shipper, the year, and the pair shipper-year with the most quantity.

The shipper with the most quantity is United Package.

The year in which the most quantity was shipped is 1997.

The pair shipper-year with the most quantity is United Package - 1997.

Graphical user interface

Description automatically generated

Query d) Analyze sales by customer country and product category to identify the pairs of country-category with no sales at all.

The pairs country-category with no sales at all are:

Argentina – Beverages;

Austria – Beverages;

Belgium – Beverages; Belgium – Condiments; Belgium – Confections; Belgium – Produce;

Canada – Beverages;

Denmark – Beverages;

Ireland – Beverages; Ireland – Condiments;

Mexico – Produce; Norway – Beverages;

Norway – Confections; Norway – Meat/Poultry; Norway – Produce;

Poland – Beverages; Poland – Condiments; Poland – Confections; Poland – Dairy Products;

Portugal – Condiments; Portugal – Confections;

Spain – Condiments; Spain – Dairy Products;

Sweden – Condiments;

Switzerland – Beverages; Switzerland – Condiments; Switzerland-Confections; Switzerland – Dairy Products;

Table

Description automatically generated

Query e) Analyze quantity by supplier country and customer country to identify the pairs of countries with no quantities being shipped between them.

The pairs of countries with no quantities being shipped between them are:

Argentina - Australia; Argentina – Brazil; Argentina – Denmark; Argentina – Netherlands; Argentina – Norway; Argentina – Singapore; Argentina – Spain; Argentina – Sweden;

Austria – Brazil; Austria – Denmark; Austria – Spain; Austria -UK;

Belgium – Australia; Belgium – Brazil; Belgium – Denmark; Belgium – Netherlands; Belgium – Spain; Belgium – Sweden; Belgium – UK; Belgium – USA;

Brazil – Japan;

Canada – Brazil; Canada – Denmark; Canada – Finland; Canada – Spain; Canada – USA;

Denmark – Brazil; Denmark – Denmark; Denmark – Japan; Denmark – Netherlands; Denmark – Norway; Denmark – Singapore; Denmark – Spain; Denmark – Sweden; Denmark – UK;

Finland – Denmark; Finland – Netherlands; Finland – Sweden; Finland – Singapore;

France – Netherlands; France – Sweden;

Germany – Brazil;

Ireland – Brazil; Ireland – Denmark; Ireland – Japan; Ireland – Netherlands; Ireland – Spain; Ireland – Sweden; Ireland – UK;

Italy – Norway; Italy – Singapore; Italy – Spain;

Mexico – Denmark; Mexico – Sweden; Mexico – UK;

Norway – Brazil; Norway – Canada; Norway – Denmark; Norway – Finland; Norway – France; Norway – Italy; Norway – Japan; Norway – Netherlands; Norway – Spain; Norway – Sweden; Norway – UK;

Poland – Brazil; Poland – Denmark; Poland – Finland; Poland – Italy; Poland – Japan; Poland – Netherlands; Poland – Norway; Poland – Singapore; Poland – Spain; Poland – Sweden; Poland – USA;

Portugal – Brazil; Portugal – Denmark; Portugal – Japan; Portugal – Netherlands; Portugal – Norway; Portugal – Spain; Portugal – Sweden;

Spain – Brazil; Spain – Denmark; Spain – Netherlands; Spain – Norway; Spain – Singapore; Spain – Spain; Spain – Sweden;

Sweden – Brazil; Sweden – Denmark; Sweden – Japan; Sweden – Netherlands; Sweden -Singapore; Sweden – Spain; Sweden – Sweden;

Switzerland – Brazil; Switzerland – Denmark; Switzerland – Netherlands; Switzerland – Singapore; Switzerland – Spain; Switzerland – Sweden; Switzerland – UK;

UK – Brazil; UK – Netherlands;

USA – Japan; USA – Sweden;

Venezuela – Brazil; Venezuela – Denmark; Venezuela – Netherlands; Venezuela – Sweden;

A screenshot of a computer

Description automatically generated with medium confidence

Query f) Analyze quantity by product category and shipping company to identify the pairs of category-shipper with no quantity at all.

There are no pairs of product category – shipper with no quantity at all.

Table

Description automatically generated with medium confidence