Northwind Traders Data Warehouse Documentation

This documentation describes the data warehouse implementation for the Northwind Traders dataset. It provides an overview of the design and stages of building the data warehouse, including conceptual, logical, and physical models, ETL processes, administration, and analytics.

Contents

Project Overview

The **Digital Egypt Pioneers Initiative, NTI,** initiated the **Northwind Traders** data warehouse project to create a robust data warehouse for better business intelligence and reporting.

The Northwind dataset contains sales data for a fictitious company called **Northwind Traders**, which imports and exports specialty foods globally.

Key Data Warehouse Components:

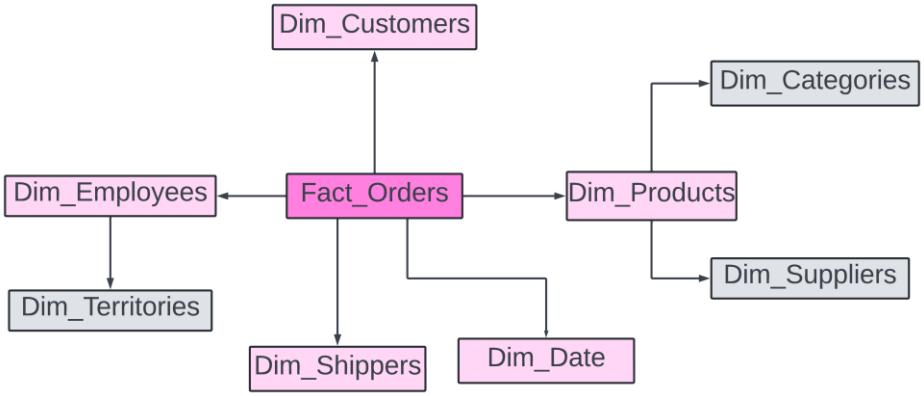
- Suppliers
- Customers
- Employees
- Products
- Shippers
- Orders and Order Details

Data Modeling

1. Conceptual Model

The conceptual model outlines the **high-level entities** and their **relationships** without going into the technical details like data types or specific keys. It provides a simplified view of the database's core entities.

Conceptual Diagram:



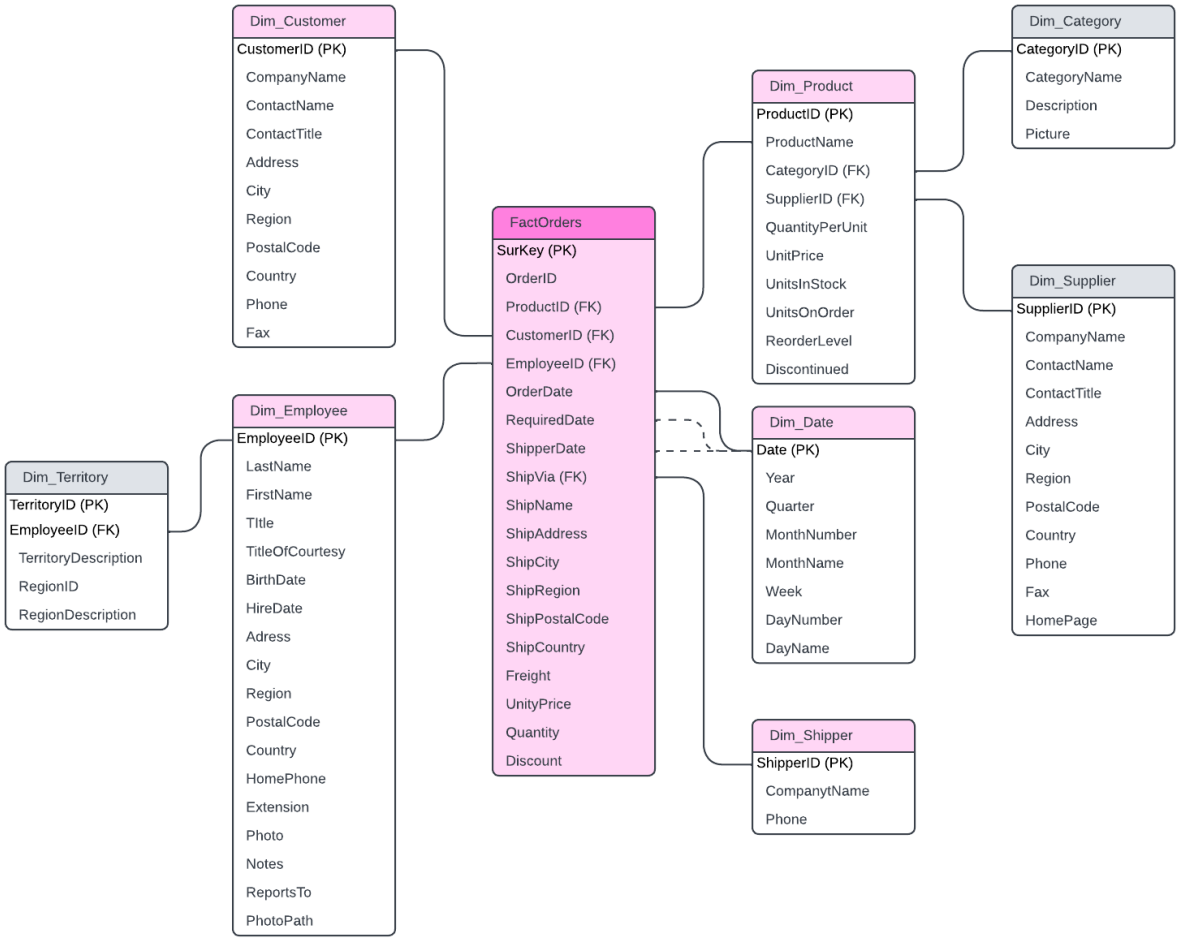
High-Level Relationships:

- **Products ↔ Categories:** Each product belongs to one category.
- **Products ↔ Suppliers:** Each product is supplied by one supplier.
- **Orders ↔ Customers:** Each order is placed by one customer.
- **Orders ↔ Employees:** Each order is handled by one employee.
- **Orders ↔ Shippers:** Each order is shipped by one shipper.
- **Employees ↔ Territories:** Each employee is assigned to one territory.

2. Logical Model

The logical model builds on the conceptual model by specifying the attributes of each entity, relationships, and keys, without focusing on the physical storage.

Logical Diagram:

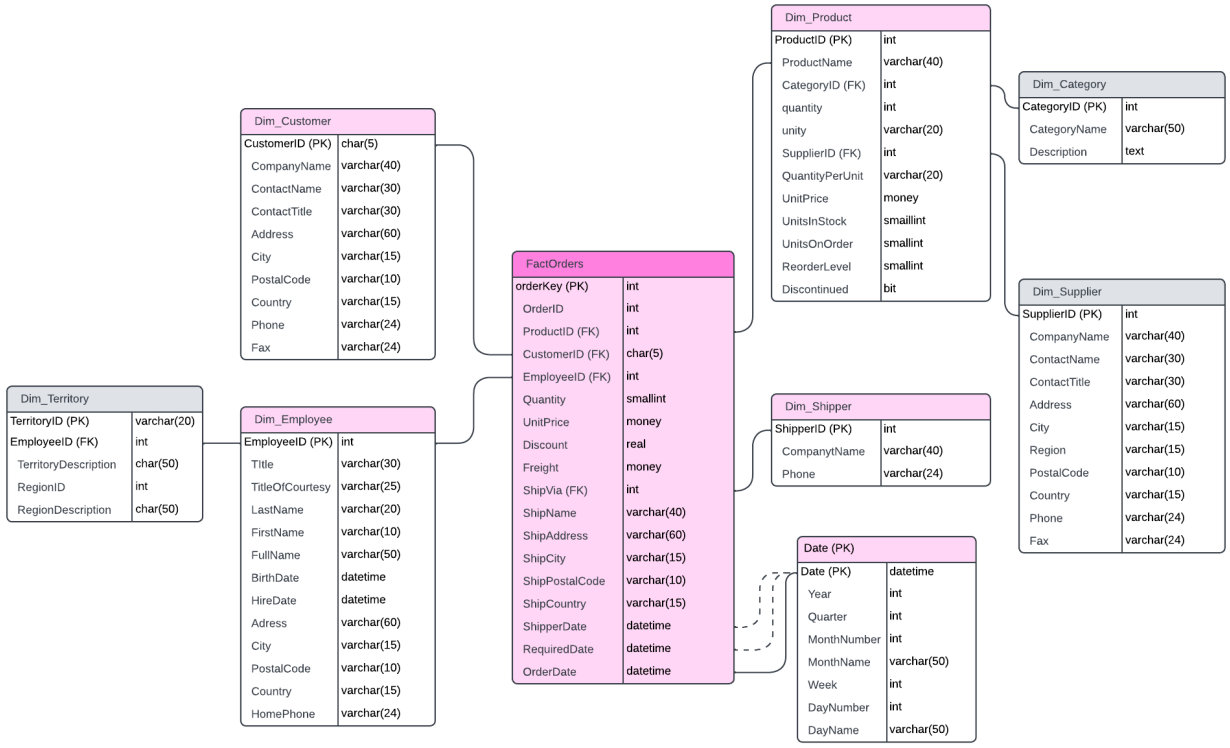


- **Fact_Orders:** Stores transaction-level details about orders.
- **Dim_Product:** Contains detailed information about each product sold.
- **Dim_Customer:** Contains customer-specific details, including geographic and contact information.
- **Dim_Employee:** Information about employees, including their roles and managers.
- **Dim_Shipper:** Details of shipping companies used in transactions.
- **Dim_Supplier:** Information about suppliers of products.
- **Dim_Territory:** Combines territory and region information into a single dimension.

3. Physical Model

The physical data model defines the actual structure of the tables and their columns, as well as the relationships between them.

Physical Diagram:



Key Adjustments:

1. Removal of Unessential Columns

During the design process, we carefully evaluated the source database and identified columns that did not add value to our analytical needs. **We removed these columns** to streamline the data warehouse, improve query performance, and reduce storage needs.

For example, columns like "Picture" in the Category table and "Home Page" in the supplier table were deemed unnecessary for the analysis and thus were excluded from the physical model.

2. Next Steps After the Physical Model

ETL Development:

- The Extract, Transform, Load (ETL) processes are developed to move data from the source operational database to the data warehouse.
- Data transformations are applied to align the source data with the new data warehouse schema. This includes tasks such as joining tables, cleaning data, and loading it into the fact and dimension tables

3. Design Decision: Separating Dimensions

Why Dim_Product and Dim_Category are Separate

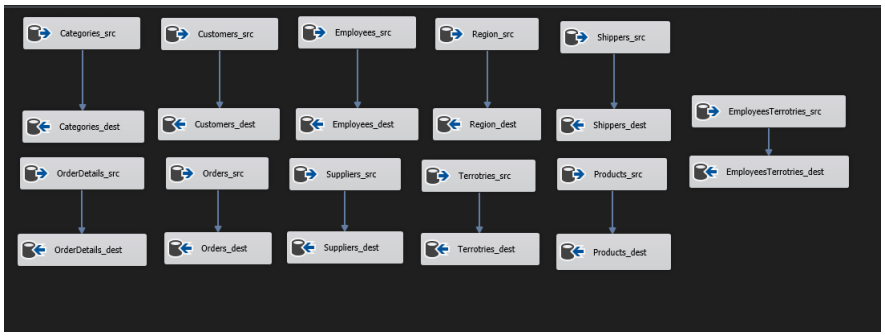
- **Normalization:** Dim_Product and Dim_Category are kept separate to avoid redundancy and make it easier to update category information without affecting the product data.
- **Flexibility:** Keeping them separate allows for more flexible queries, such as analyzing sales by product categories without repeating category data in every product row. It also enables easier grouping and filtering based on categories.
- **Hierarchical Relationship:** Each product belongs to a single category, but a category can have multiple products. This many-to-one relationship is better represented by separate tables to avoid unnecessary data duplication.

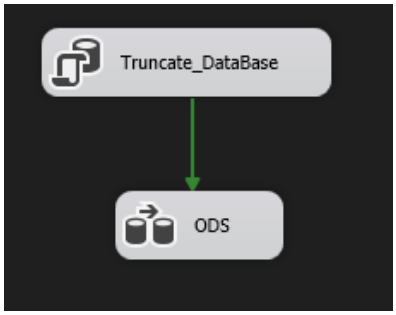
ETL Process

The ETL (Extract, Transform, Load) process extracts data from the source OLTP database, transforms it to fit the data warehouse schema, and loads it into the data warehouse.

ODS (Operational Data Store):

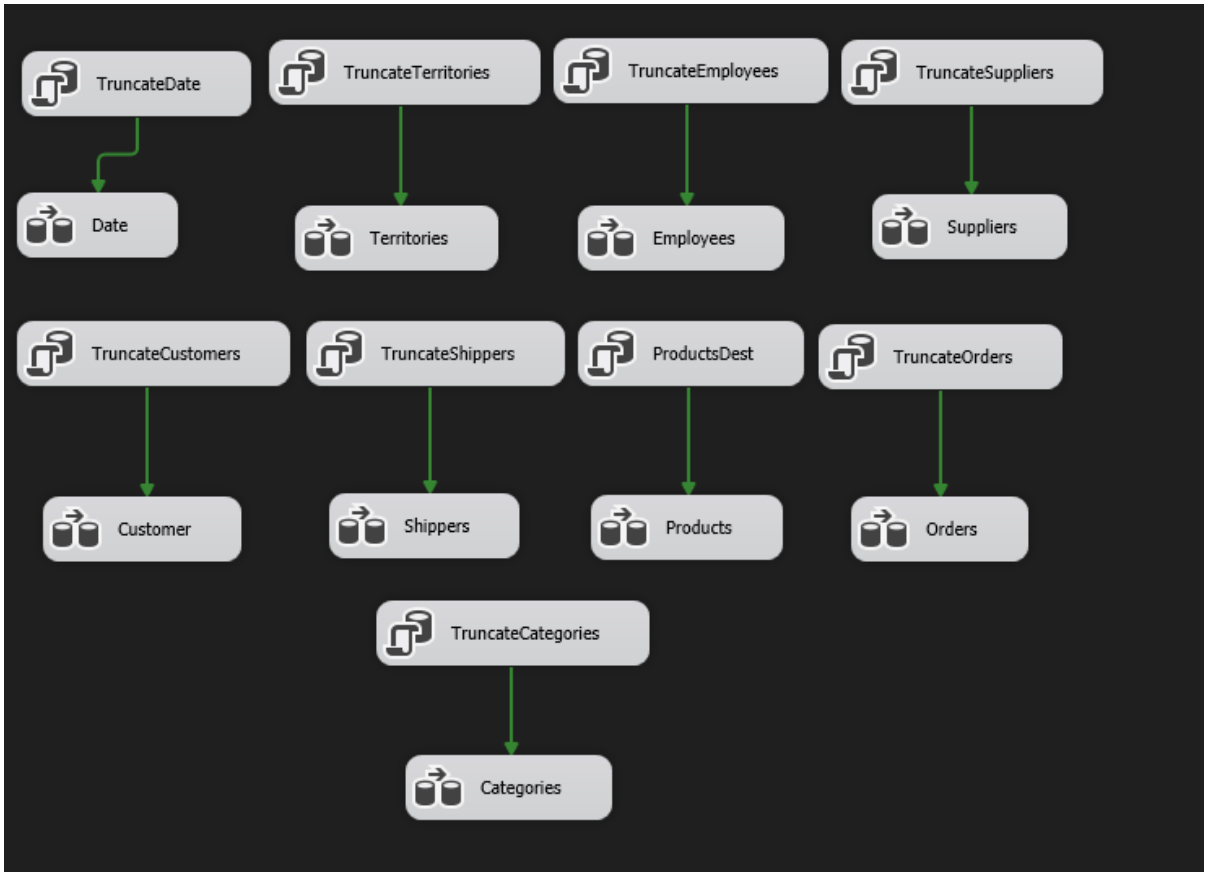
- Temporary storage for raw data from the source systems before transformation.





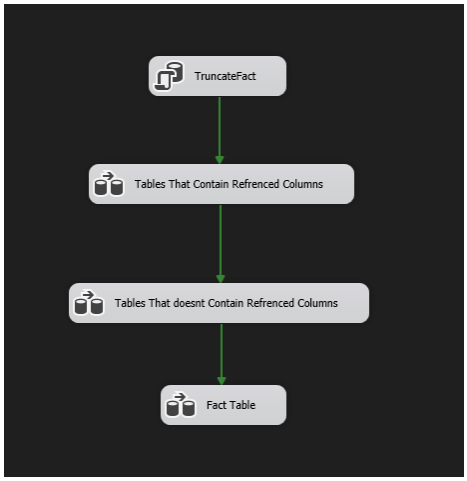
Stage Area:

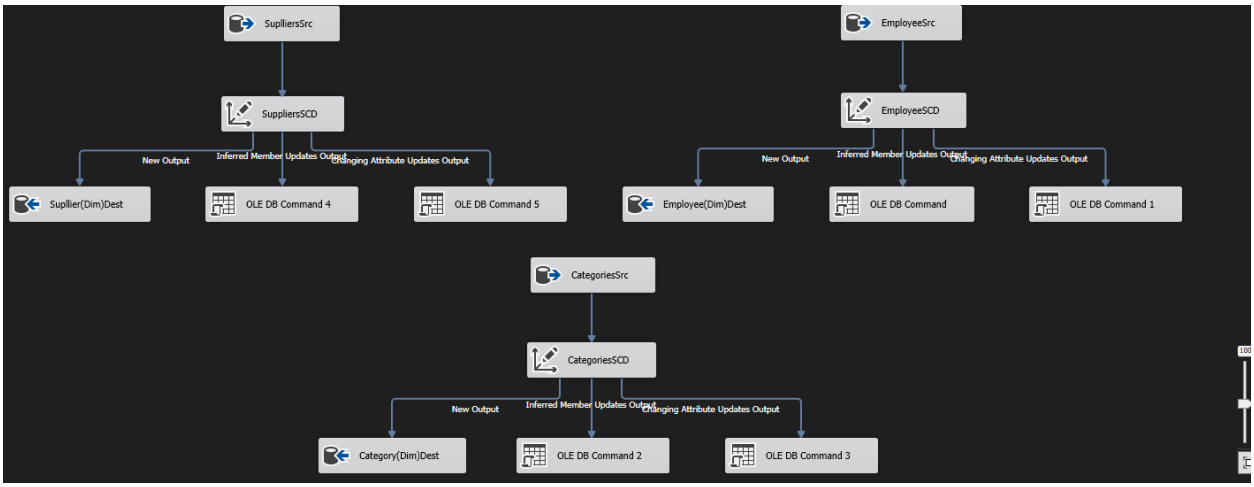
- The staging area stores data before it is loaded into the data warehouse. Transformation processes clean, filter, and normalize the data in this step.

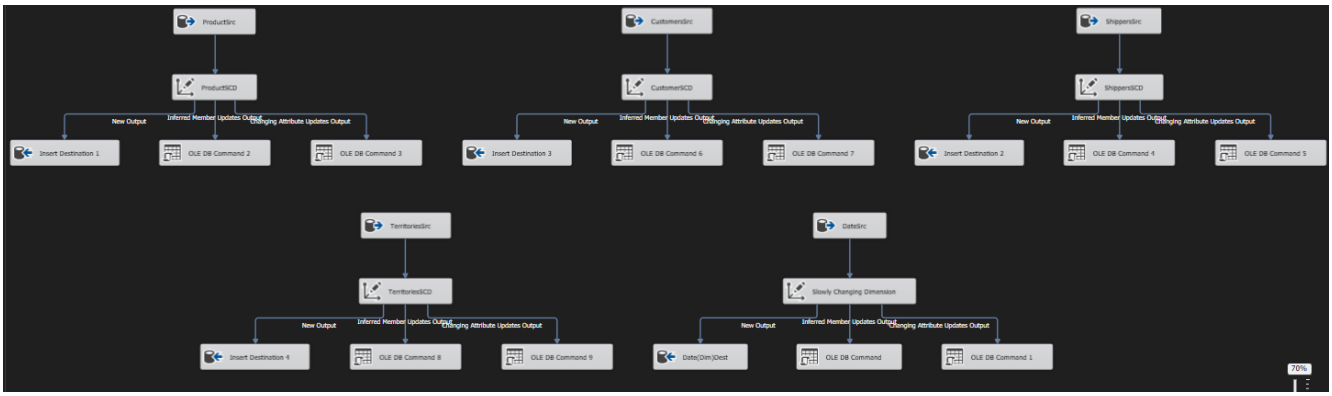


Data Warehouse:

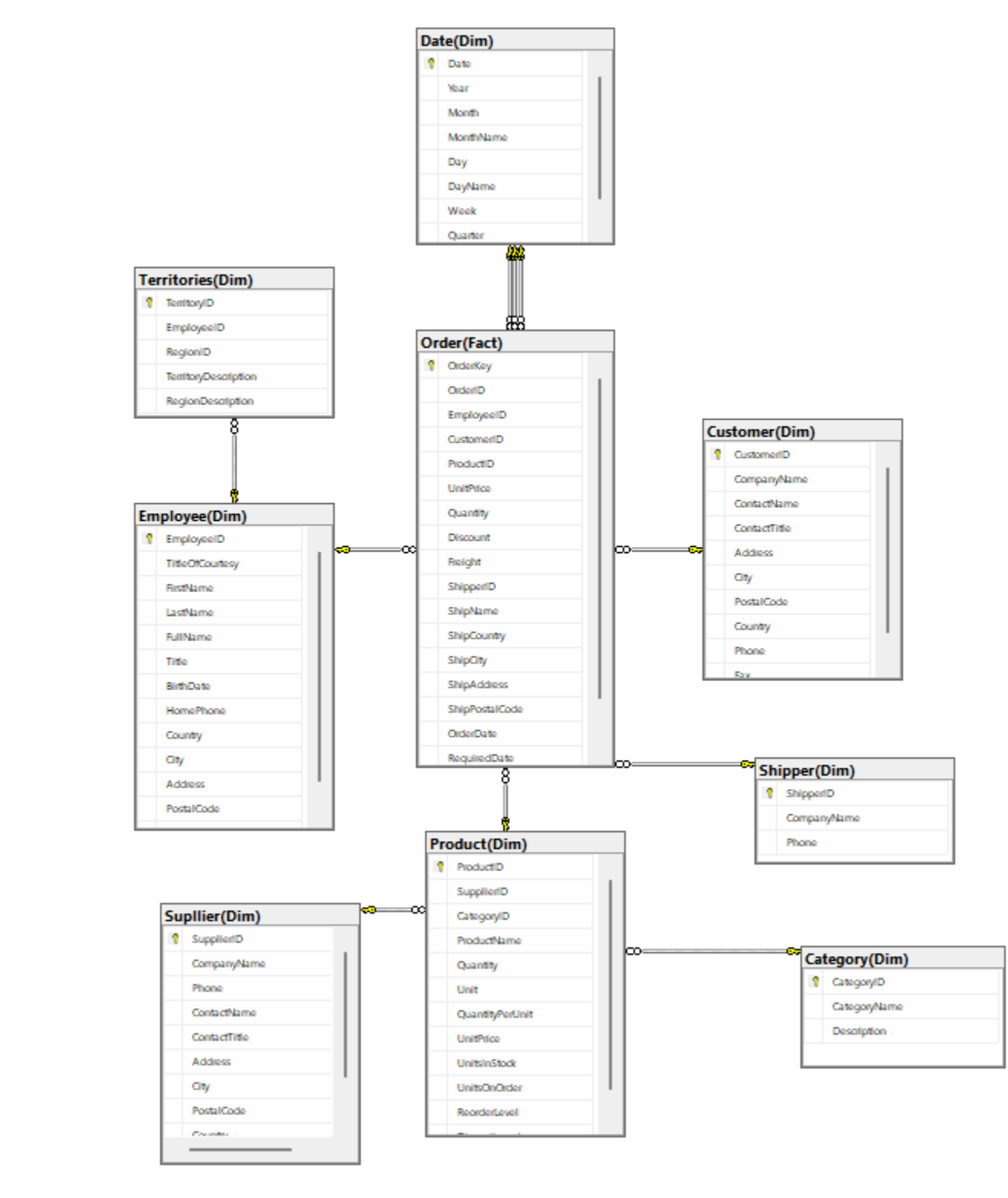
- The final, structured data warehouse where cleaned data is stored in the star schema







Final Model


Date(Dim)
Date
Year
Month
MonthName
Day
DayName
Week
Quarter
Territories(Dim)
TerritoryID
EmployeeID
RegionID
TerritoryDescription
RegionDescription
Order(Fact)
OrderKey
OrderID
EmployeeID
CustomerID
ProductID
UnitPrice
Quantity
Discount
Freight
ShipperID
ShipName
ShipCountry
ShipCity
ShipAddress
ShipPostalCode
OrderDate
RequiredDate
Customer(Dim)
CustomerID
CompanyName
ContactName
ContactTitle
Address
City
PostalCode
Country
Phone
Fax
Employee(Dim)
EmployeeID
TitleOfCourtesy
FirstName
LastName
FullName
Title
BirthDate
HomePhone
Country
City
Address
PostalCode
Shipper(Dim)
ShipperID
CompanyName
Phone
Product(Dim)
ProductID
SupplierID
CategoryID
ProductName
Quantity
Unit
QuantityPerUnit
UnitPrice
UnitsInStock
UnitsOnOrder
ReorderLevel
Supllier(Dim)
SupplierID
CompanyName
Phone
ContactName
ContactTitle
Address
City
PostalCode
Category(Dim)
CategoryID
CategoryName
Description

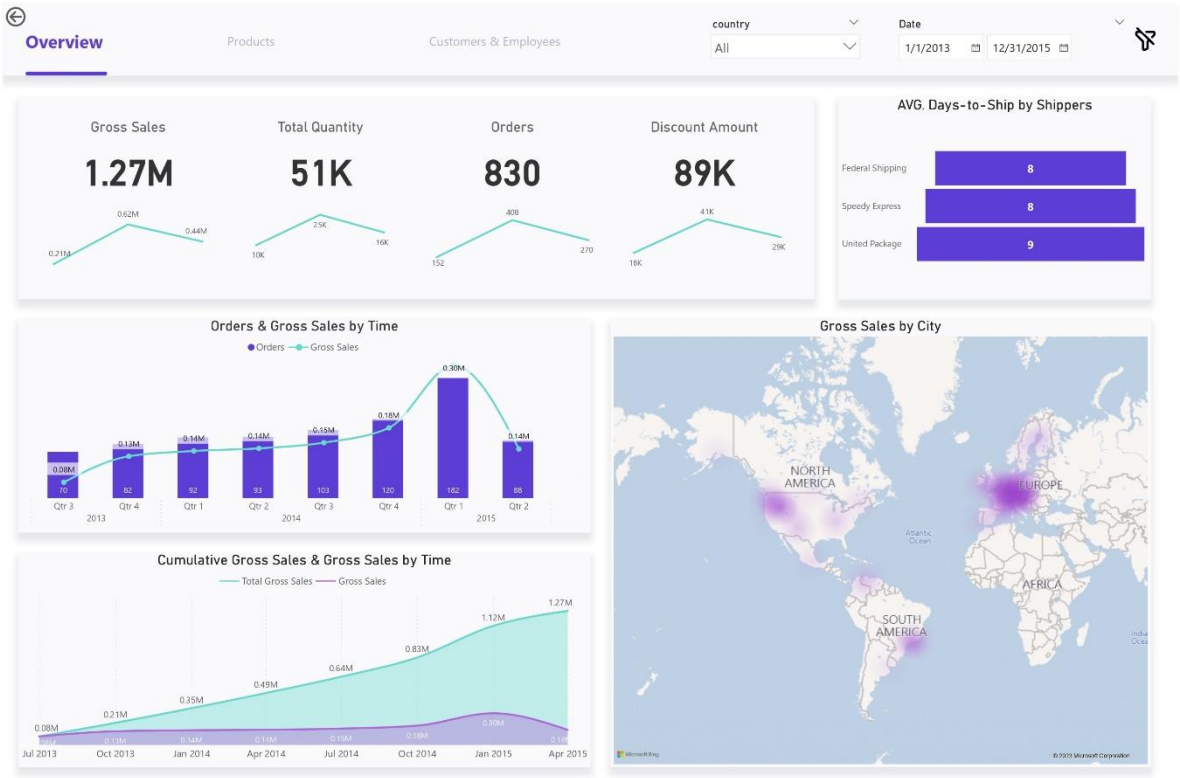
Administration

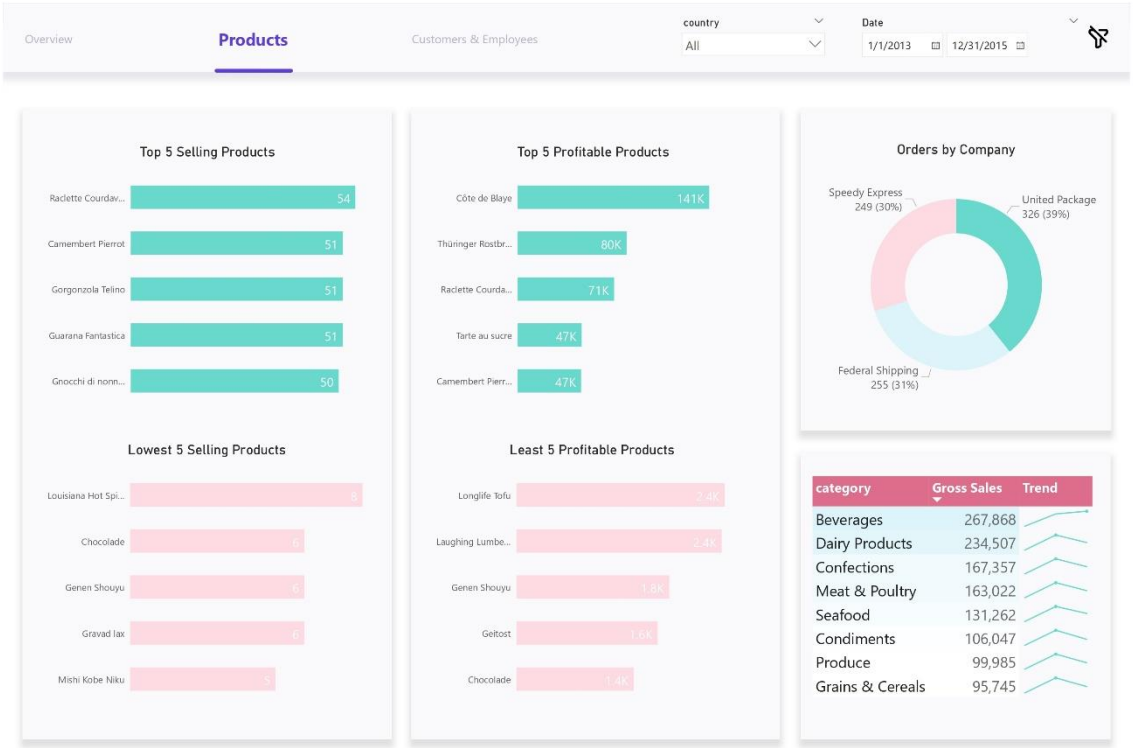
- **SSIS Package Automation**: The ETL process is automated with SSIS (SQL Server Integration Services) packages, scheduled to run daily at **9:00 PM**.
- **Job Tracking**: The status and progress of the scheduled jobs can be tracked using SQL Server Agent.
- **Alerts**: The system sends alerts to users upon job failure or success to ensure proper monitoring.

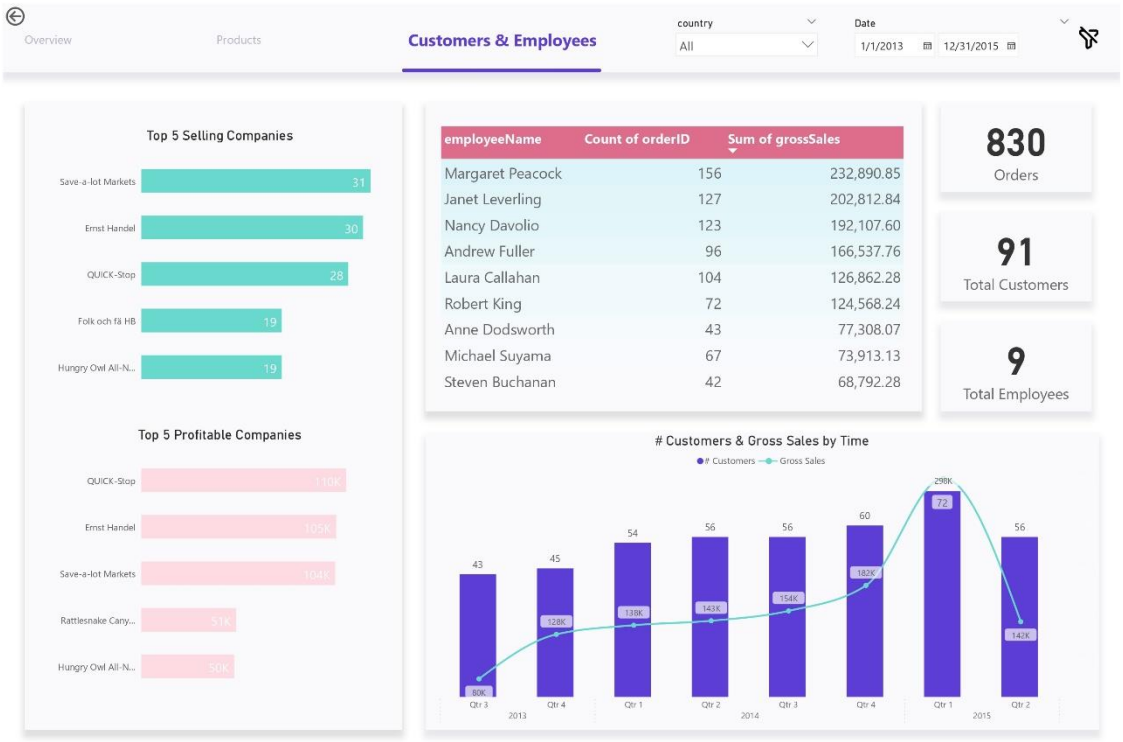
Analytics

Power BI Dashboard:

- A **Power BI** dashboard is developed to provide real-time visualizations of key business metrics, such as sales performance, order trends, inventory levels, and customer behavior.




Overview
Products
Customers & Employees
country
All
Date
1/1/2013
12/31/2015
Top 5 Selling Products
Raclette Courdav...
54
Camembert Pierrot
51
Gorgonzola Telino
51
Guarana Fantastica
51
Gnocchi di nonn...
50
Top 5 Profitable Products
Côte de Blaye
141K
Thüringer Rostbr...
80K
Raclette Courda...
71K
Tarte au sucre
47K
Camembert Pierr...
47K
Orders by Company
Speedy Express
249 (30%)
United Package
326 (39%)
Federal Shipping
255 (31%)
Lowest 5 Selling Products
Louisiana Hot Spi...
Chocolade
Genen Shouyu
Gravad lax
Mishi Kobe Niku
Least 5 Profitable Products
Longlife Tofu
Laughing Lumbe...
Genen Shouyu
Geitost
Chocolade
category
Gross Sales
Trend
Beverages
267,868
Dairy Products
234,507
Confections
167,357
Meat & Poultry
163,022
Seafood
131,262
Condiments
106,047
Produce
99,985
Grains & Cereals
95,745


Overview
Products
Customers & Employees
country
All
Date
1/1/2013
12/31/2015
Top 5 Selling Companies
Save-a-lot Markets
31
Ernst Handel
30
QUICK-Stop
28
Folk och fä HB
19
Hungry Owl All-N...
19
Top 5 Profitable Companies
QUICK-Stop
Ernst Handel
Save-a-lot Markets
Rattlesnake Cany...
Hungry Owl All-N...
employeeName
Count of orderID
Sum of grossSales
Margaret Peacock
156
232,890.85
Janet Leverling
127
202,812.84
Nancy Davolio
123
192,107.60
Andrew Fuller
96
166,537.76
Laura Callahan
104
126,862.28
Robert King
72
124,568.24
Anne Dodsworth
43
77,308.07
Michael Suyama
67
73,913.13
Steven Buchanan
42
68,792.28
830
Orders
91
Total Customers
9
Total Employees
Customers & Gross Sales by Time
Customers
Gross Sales
43
45
54
56
56
60
72
56
80K
128K
138K
143K
154K
182K
298K
142K
Qtr 3
Qtr 4
Qtr 1
Qtr 2
Qtr 3
Qtr 4
Qtr 1
Qtr 2
2013
2014
2015

Conclusion

The data warehouse for **Northwind Traders** is designed to provide a clear and efficient structure for analyzing business performance across multiple dimensions.

The **fact and dimension tables** store transactional and descriptive data, while the **ETL process** ensures that the data is clean, consistent, and ready for analytics.

Project Workflow

Team Members

Name	Role	Email
Ahmed Fawzy	Team Leader	ahmedfawzy2306@gmail.com
Abdallah Ashraf	Task Owner	body01021722406@gmail.com
Mamdouh Sabry	Member	mamdouh.eltehawey@gmail.com
Abdulrahman Mostafa	Member	abdelrahmanmustafa466@gmail.com
Mohammed Hani	Member	mohamedhani9000@gmail.com
Ebrahim Ayman	Member	ebrahimaymenzaki55@gmail.com
Ahmed Younis	Task Owner	ahmedyounisokal@gmail.com
Mohammed Ashraf	Member	ashrafmedo736@gmail.com
Fouad Khaled	Task Owner	foukha49@gmail.com

Project Tasks

Task	Assigned Members	Priority	Status	Start date	End date
Business Requirements	Ahmed Fawzy	P0	Completed	9/29/2024	9/29/2024
Data Modeling	Abdallah Ashraf, Mamdouh Sabry, Abdulrahman Mostafa	P0	Completed	9/29/2024	9/30/2024
ETL	Ahmed Younis, Mohammed Hani, Ebrahim Ayman	P0	Completed	10/1/2024	10/4/2024
Administration	Fouad Khaled, Mohammed Ashraf	P1	Completed	10/3/2024	10/4/2024
Analysis	Ahmed Fawzy, Fouad Khaled, Mohammed Ashraf	P2	Completed	10/3/2024	10/4/2024