

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..... 2

Data Modeling 3

1. Conceptual Model..... 3

Conceptual Diagram:..... 3

High-Level Relationships: 3

2. Logical Model..... 4

Logical Diagram: 4

3. Physical Model..... 5

Physical Diagram: 5

Key Adjustments:..... 5

ETL Process 7

ODS (Operational Data Store): 7

Stage Area: 7

Data Warehouse: 8

Administration 10

Analytics 10

Power BI Dashboard: 10

Conclusion..... 12

Project Workflow 12

Team Members 12

Project Tasks..... 12

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:

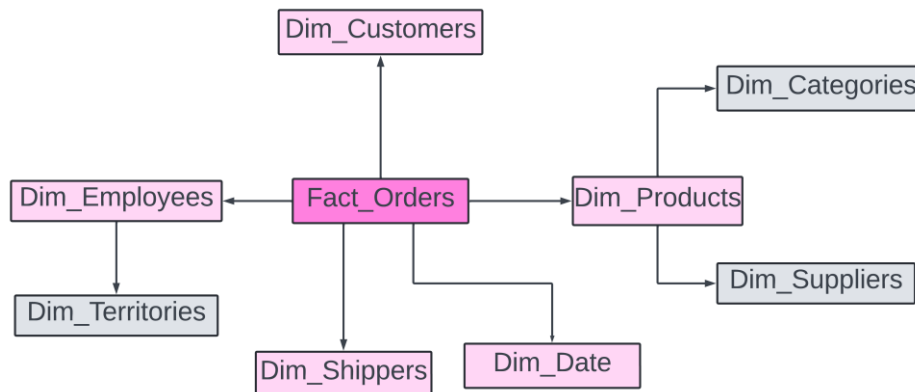
- Suppliers
- Customers
- Employees
- Products
- Shippers
- Orders and Order Details

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:



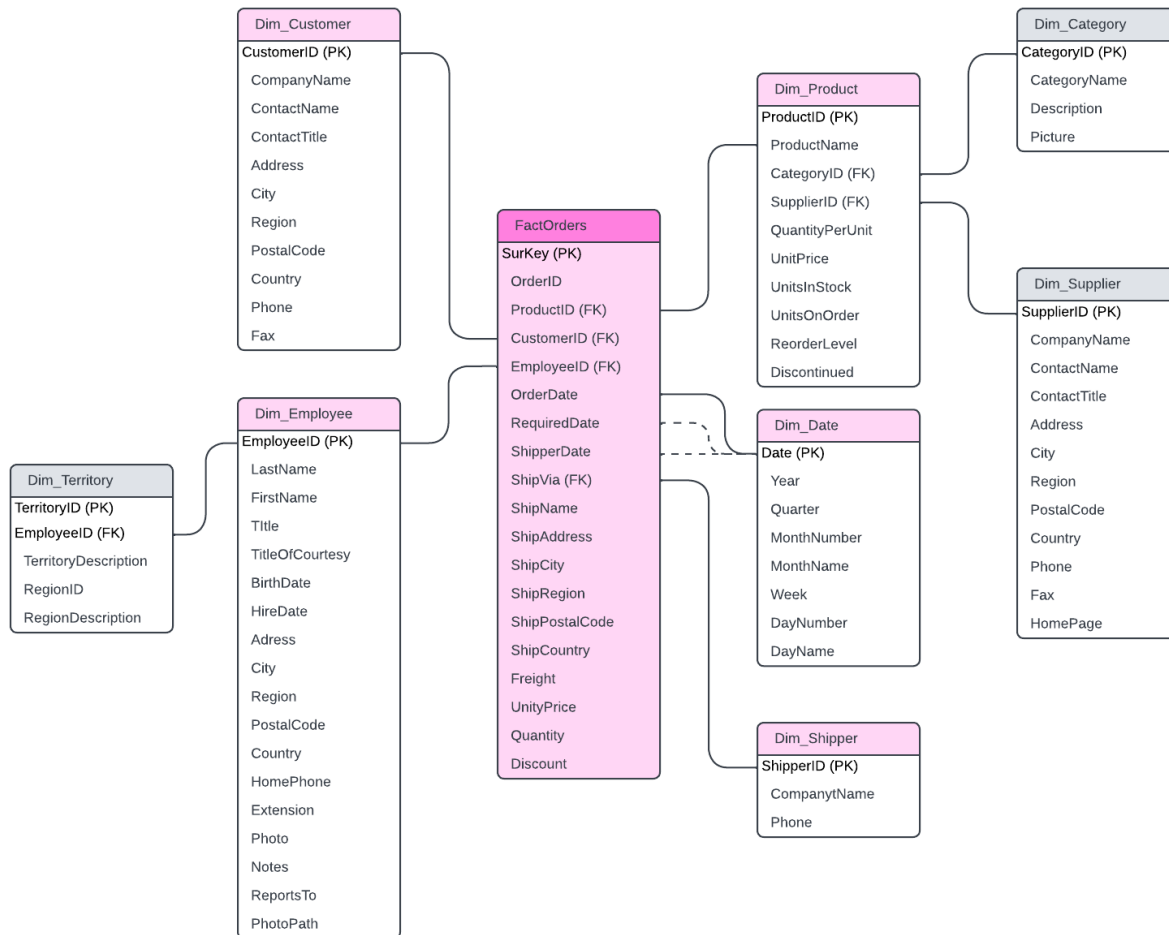
High-Level Relationships:

- **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:

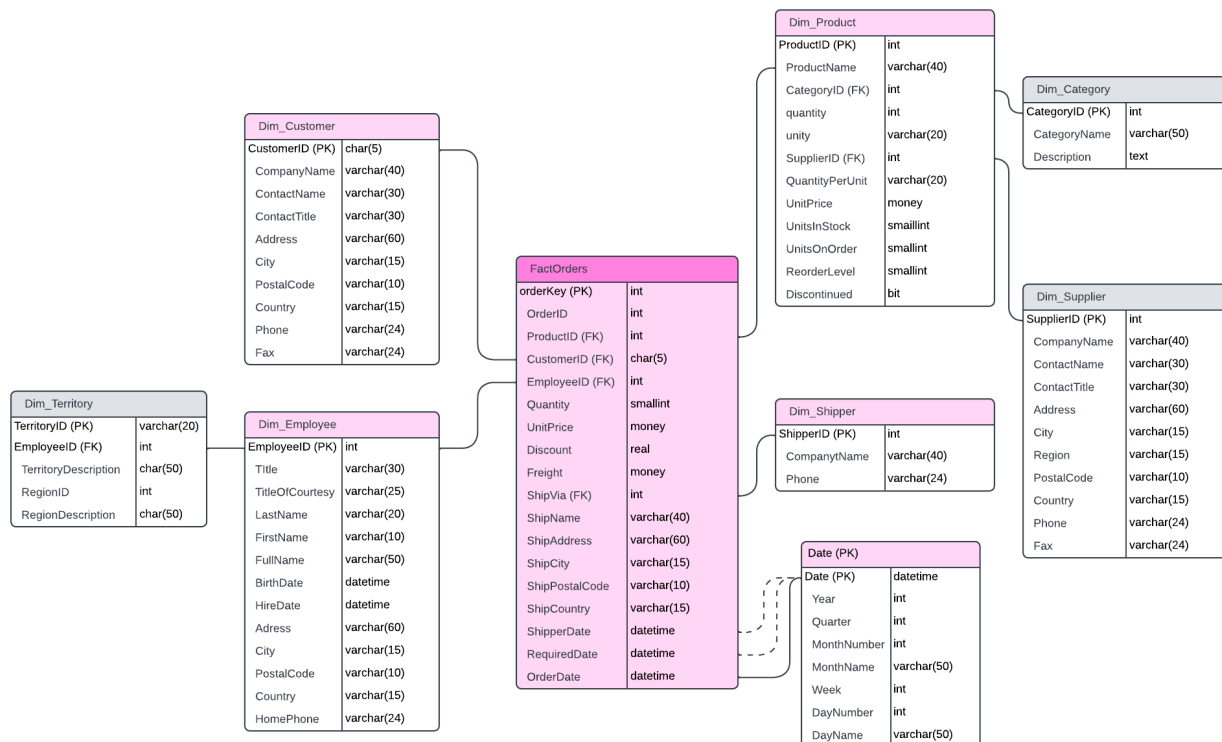


- **Fact_Orders:** Stores transaction-level details about orders.
- **Dim_Product:** Contains detailed information about each product sold.
- **Dim_Customer:** Contains customer-specific details, including geographic and contact information.
- **Dim_Employee:** Information about employees, including 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

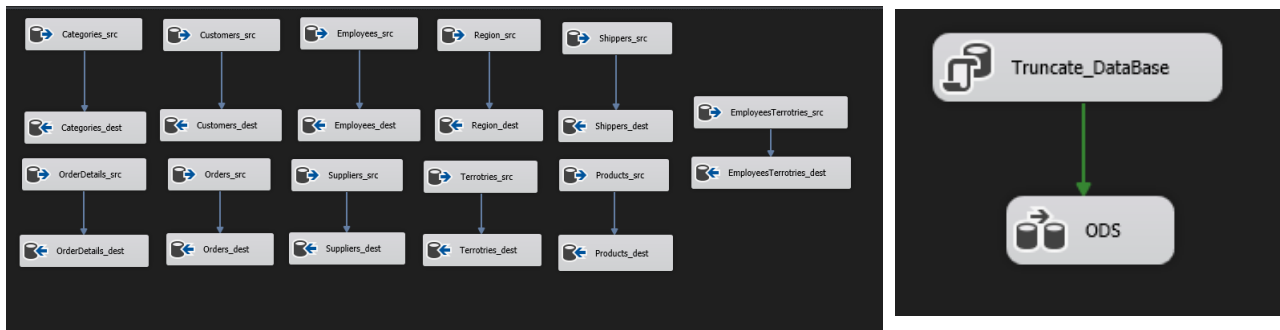
- **Normalization:** Dim_Product and Dim_Category are kept separate to avoid redundancy and make it easier to update category information without affecting the product data.
 - **Flexibility:** Keeping them separate allows for more flexible queries, such as analyzing sales by product categories without 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.

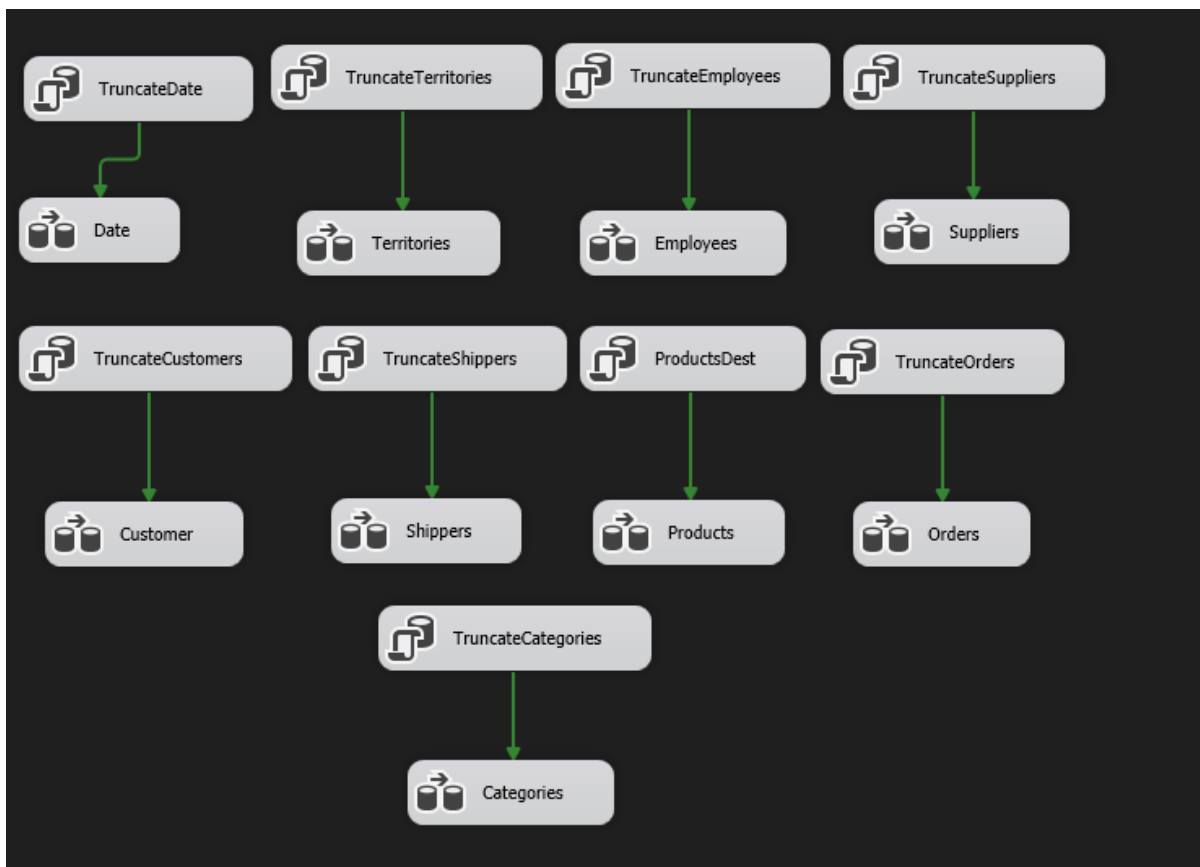
ODS (Operational Data Store):

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



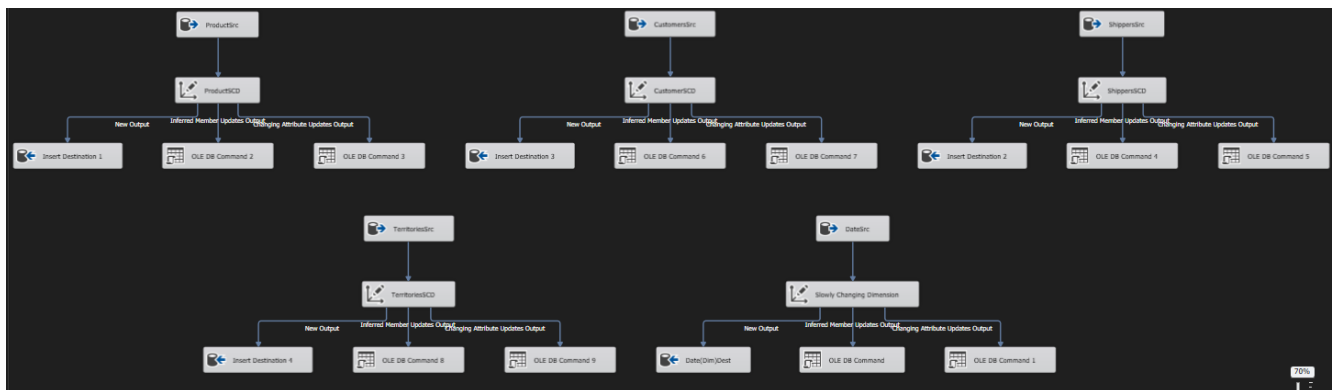
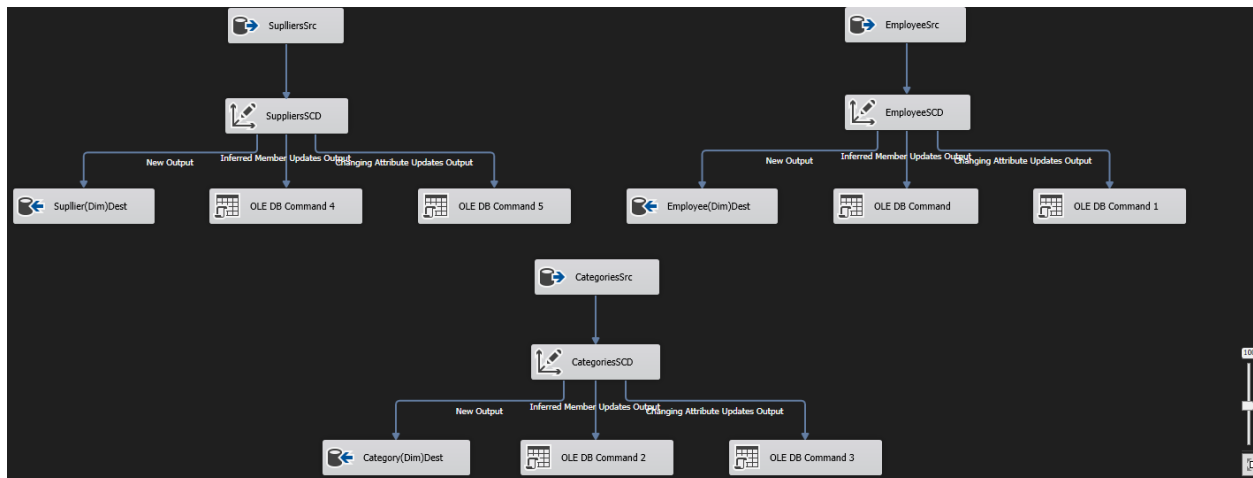
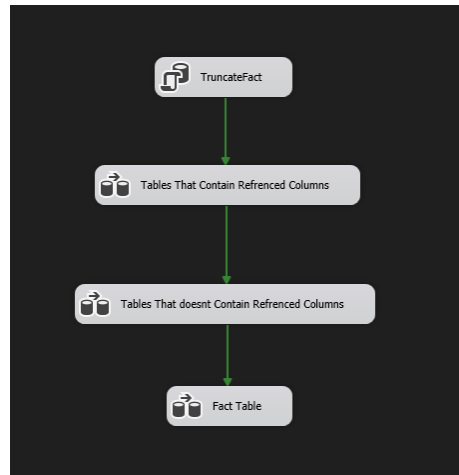
Stage Area:

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

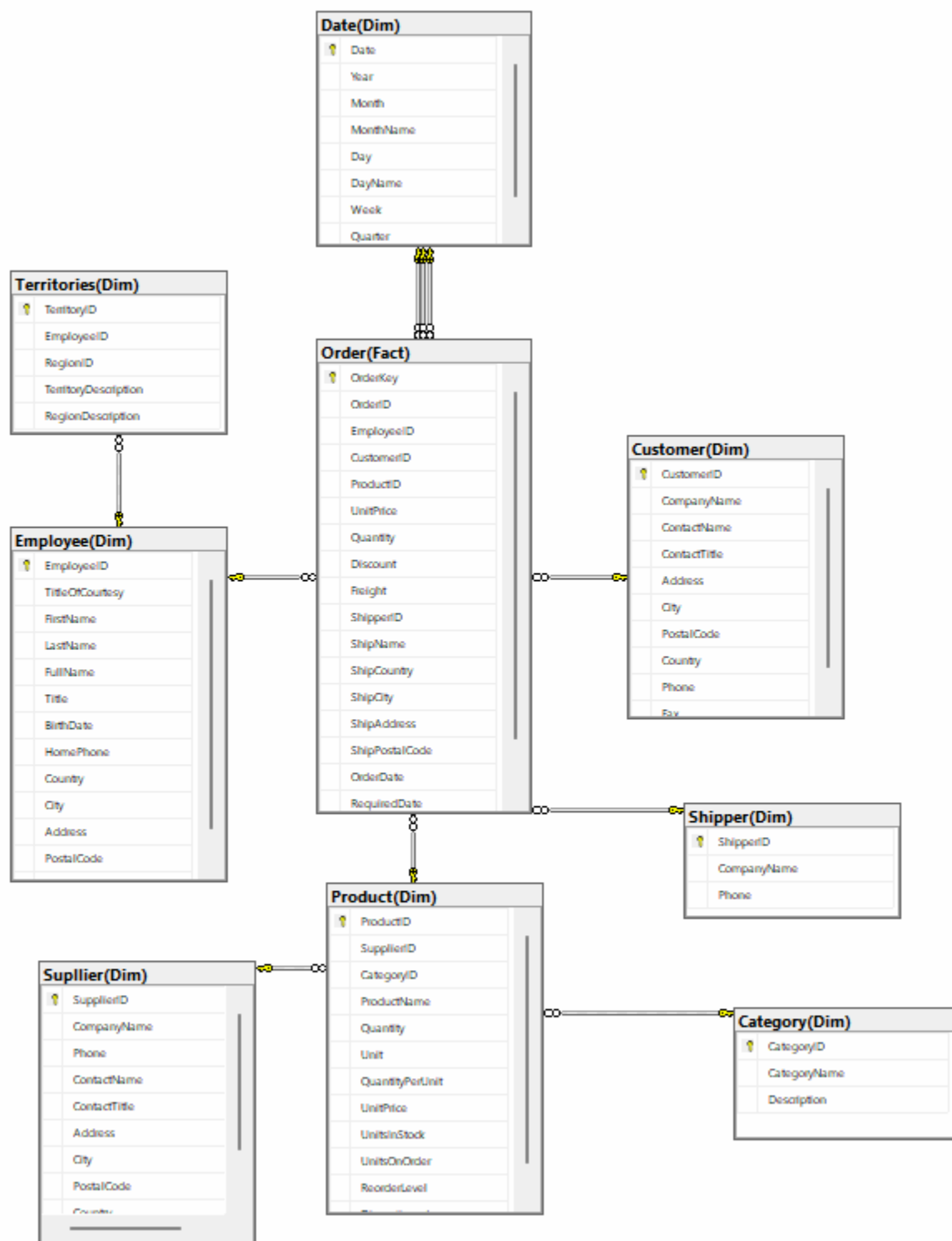


Data Warehouse:

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



Final Model



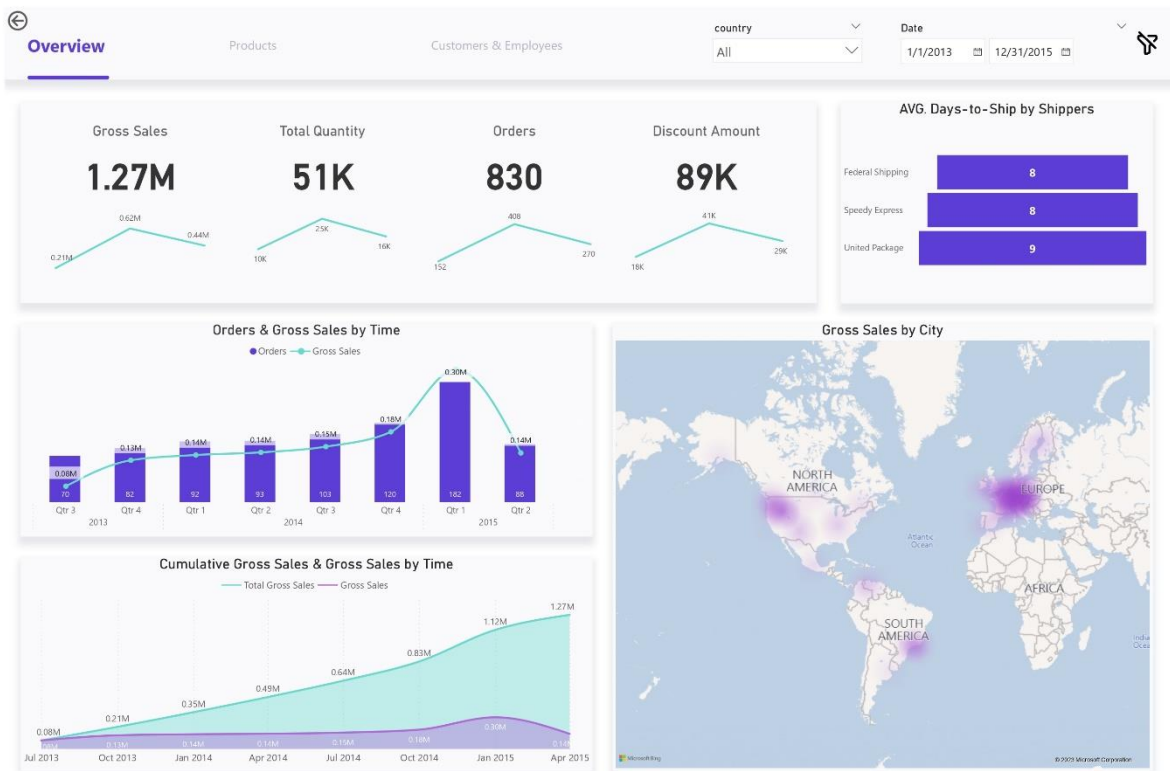
Administration

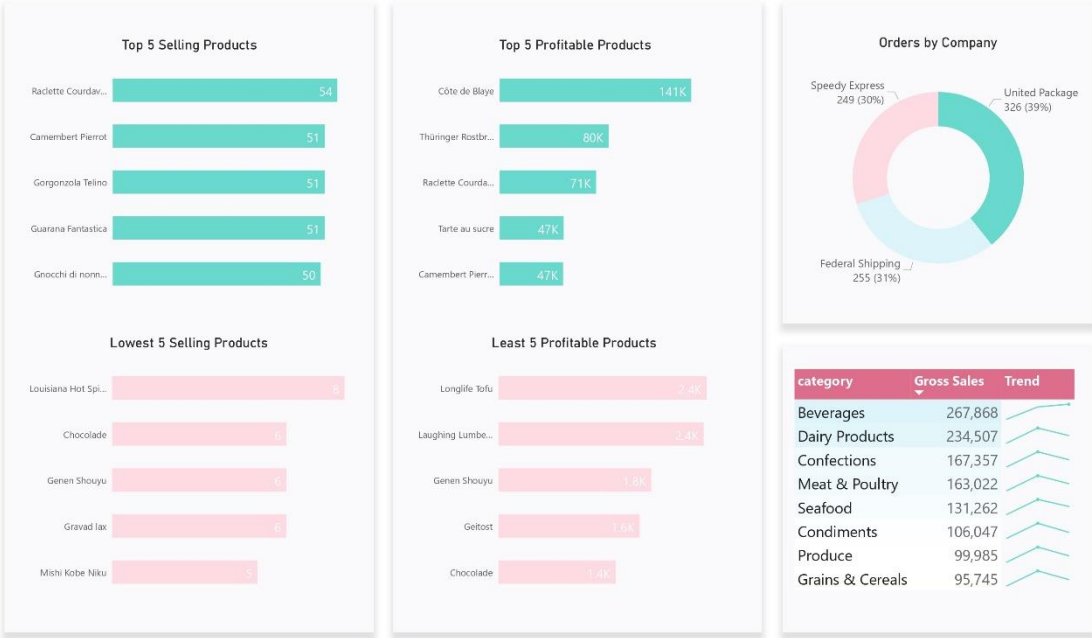
- **SSIS Package Automation:** The ETL process is automated with SSIS (SQL Server Integration Services) packages, scheduled to run daily at **9:00 PM**.
- **Job Tracking:** The status and progress of the scheduled jobs can be tracked using SQL Server Agent.
- **Alerts:** The system sends alerts to users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





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	Member	body01021722406@gmail.com
Ebrahim Ayman	Member	ebrahimaymenzaki55@gmail.com
Ahmed Younis	Task Owner	ahmedyounisokal@gmail.com
Fouad Khaled	Member	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, Ahmed Younis, Ahmed Fawzy	P0	Completed	9/29/2024	9/30/2024
ETL	Ahmed Younis, Ebrahim Ayman	P0	Completed	10/1/2024	10/4/2024
Analysis	Ahmed Fawzy, Fouad Khaled	P1	Completed	10/3/2024	10/4/2024