***Northwind Data Warehouse Modeling***

***Table of Contents***

**1. Introduction**

**2. Technology Used**

**3. Requirements**

**4. Deliverable Requirements**

**5. Data Quality**

**6. Conceptual Model**

**7. Logical Model**

**8. Physical Model**

**9. Dashboard for KPIs**

***1. Introduction***

This document outlines the Northwind Data Warehouse Modeling project for a business that provides products to customers while managing suppliers, customers, employees, and shippers. The objective of this DWH is to support business intelligence and decision-making by consolidating, cleansing, and transforming data from various sources into a structured and optimized format.  
The data warehouse follows the Medallion Architecture with three stages:  
- Bronze Stage: Raw data is ingested from source systems.  
- Silver Stage: Data is cleansed, quality issues are resolved, and duplicates are removed with rejected records stored separately.  
- Gold Stage: Fact and dimension tables are structured for reporting and analysis.  
  
The data warehouse supports key performance indicators (KPIs) such as revenue analysis, customer insights, and operational efficiency tracking.

***2. Technology Used***

- Database: PostgreSQL.  
- ETL Pipeline: Extract, Transform, Load process.  
- BI Tools: Power BI.

***3. Requirements***

**Business Requirements:**  
- Centralized data repository for analysis and reporting.  
- Historical tracking using Type 2 Slowly Changing Dimensions (SCD2).  
- Ability to analyze revenue, growth trends, customer behavior, and delivery performance.  
- Integration of multiple data sources into a unified schema.  
**Data Sources:**  
- Operational database containing orders, customers, employees, suppliers, products, and shippers.  
- External data sources for enrichment if required.  
**Architectural Approach:**  
- ETL Pipeline: Extract data from source systems, transform it to ensure consistency and quality, and load it into the data warehouse.  
- Data Storage: Partitioned into Bronze, Silver, and Gold layers to manage data efficiently.

***4. Deliverable Requirements***

The final deliverables of the project include:  
- A fully designed data warehouse with fact and dimension tables.  
- ETL pipelines for data ingestion, transformation, and loading.  
- Data quality rules for cleansing and duplicate removal.  
- A business intelligence dashboard displaying KPIs.  
- Documentation of data models and processes.

***5. Data Quality***

To ensure high data quality, the following rules have been applied:  
- Future Date Handling: The date '2099-12-31' is used as a future date for the `Valid To` attribute in the data warehouse.  
- Missing Date Handling: The date '1937-01-01' is used as an old date for the `Valid From` attribute for any missing dates in the date warehouse.  
- Invalid Order Dates: If an `Order Date` is greater than the current date, it is considered invalid and added to the rejected tables.  
- Missing Region Values: If the `Region` attribute contains missing values, it is replaced with 'NA'.

- Missing Product Values: If the `Product Name` attribute contains missing values, it is replaced with 'NP'.

- Missing Employee Values: If the `Employee Name` attribute contains missing values, it is replaced with 'NE'.  
- Missing Discount Values: If the `Discount` attribute is missing, it is replaced with `0`.  
- Missing Shipped Date: If the `Shipped Date` is missing, it is considered an uncompleted order and added to the rejected tables. A future date '2099-12-31' is assigned.  
- Phone & Fax Formatting: Spaces or dots between numbers are replaced with hyphens ('-'), and missing values are replaced with 'NA'.  
 - Example:  
 - Input: `123.456 7890`  
 - Transformed Output: `123-456-7890`  
- Duplicate Data: Any duplicated data is identified and moved to the rejected tables.

***6. Conceptual Model***

The conceptual model represents the high-level business entities and their relationships. The core entity is Orders, which connects with Customers, Employees, Suppliers, Products, Shippers, and Countries. This model abstracts the real-world interactions and provides a foundation for logical and physical models.

A group of white rectangles with text

AI-generated content may be incorrect.

1.0 Conceptual Model.

***7. Logical Model***

The logical model defines the structure of the data warehouse without considering physical storage constraints. It consists of:  
- Fact Table: Orders Fact, which contains transactional data.  
- Dimension Tables: Customers, Employees, Suppliers, Shippers, Country, Date, Products, and Customer Segments.  
- The Customer Segment dimension is derived using RFM (Recency, Frequency, Monetary) analysis on orders data.  
This stage also defines the relationships between fact and dimension tables, ensuring efficient querying.

Several different types of names

AI-generated content may be incorrect.

2.0 Logical Model

***8. Physical Model***

The physical model details the actual implementation of the data warehouse, including:  
- Schema Design: Tables, primary keys and foreign keys.  
- Slowly Changing Dimensions (SCD2): Implemented to maintain historical changes in customer, product, shipper, employees, supplier and orders attributes.

A diagram of a computer network

AI-generated content may be incorrect.

3.0 Data Warehouse ER Diagram.

***9. Dashboard for KPIs***

The business users require a dashboard to track performance through the following KPIs:  
1. Revenue per Product, Country, and Customer Segment  
2. Growth Rate (Monthly & Yearly)  
3. Average Order Value  
4. Customer Lifetime Value  
5. Churn Rate  
6. On-Time Delivery Rate  
7. Average Delivery Delay  
  
Due to data limitations, Gross Profit Margin, Net Profit, and Lead Time are excluded.  
  
The dashboard is designed to provide interactive visualizations using Power BI tool, allowing users to filter, and analyze data efficiently.

A close-up of a data display

AI-generated content may be incorrect.

4.0 KPI’s Dashboard.