

**EY-GDS**

**D&A SQL DWH Python**

**Project Case Studies-1: 24<sup>th</sup> December 2023**

**Case Study No.01**

## **Data-Warehouse-Project**

### **Project Summary:**

The primary objective of this project is to create a concept data warehouse using dimensional modelling techniques. Identify the key stakeholders and business requirements for the same. Based on these requirements develop a design schema for the data warehouse. Create the warehouse using any ETL tools. Produce reports in support of the requirements using SSRS and R. Develop an XML schema based on the data warehouse. Implement a part of the data warehouse as graph database using Neo4j technologies.

### **Source and background:**

The data source that I have used for this project is Northwind.

The Northwind database is a fictional B2B ( Business to Business) trading company that imports and exports specialty food items. It includes data consisting of product information, orders, order details, customer, employee and a few more. This is one of the most dynamic database as it can be applied to a variety of applications which is why I was very keen on working with this data set. There is a wide variety of dimensions that can be created for a number of analysis. Using this source I have created a data warehouse called NW\_Trading\_DW. This warehouse is created to maintain optimal ratio of stock to WIP ( Work In Progress) based on their reorder level and ordered quantities. The idea is to ensure customer satisfaction with timely product delivery. It also helps keep a track of premium customers, most selling products, supply chain management (SCM) and employee performance. The key stakeholders include the Employees, Customers and Suppliers.

### **Business Requirements:**

1. Implement J.I.T.(Just in time): A Japanese model to manage deliveries.
2. Customer retention.

3. Employee incentive Program.
4. Optimize supply chain management.
5. Optimize delivery time.

## Design schema for the data warehouse:

### Entity Relationship Diagram

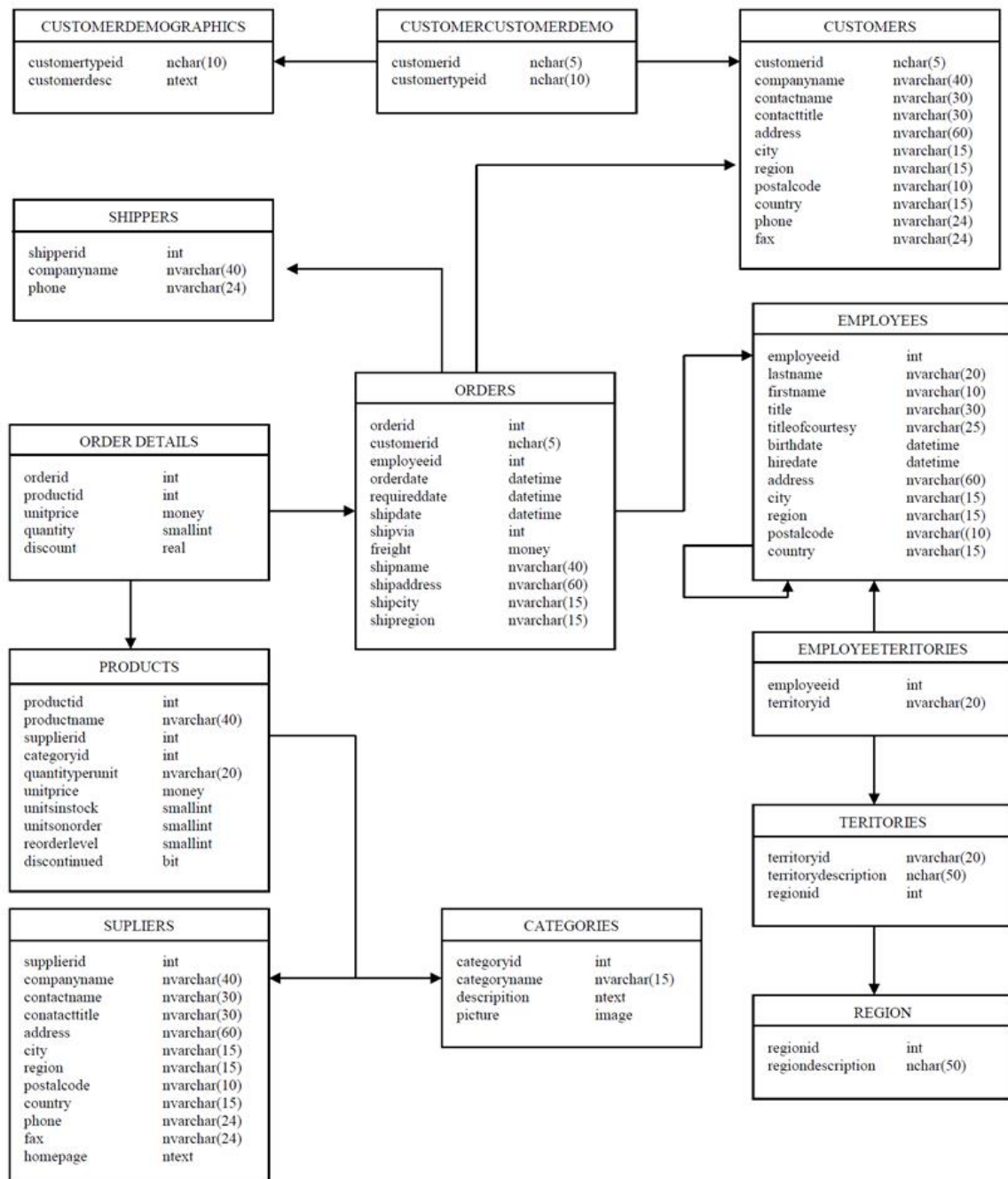
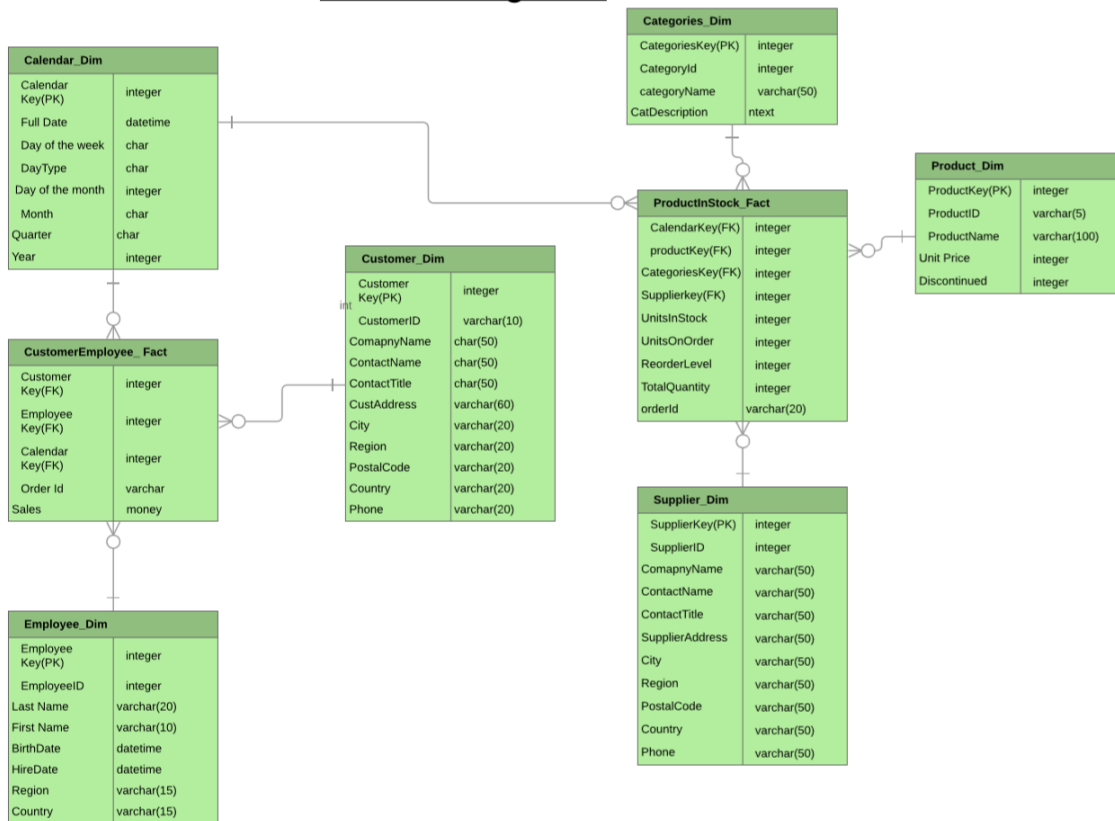


Fig 1

## NW Trading DW



**Fig 2**

A Constellation/ Galaxy of stars in data warehouse modelling is a dimensional model having multiple facts. I have used this model to create a star schema with two fact tables. The dimensions and facts and their relations are displayed in the above schema that I have created using Lucid chart.

### **Dimensions:**

The information obtained from the dimension table is used to do the analysis. The dimension tables contain descriptive and numeric values. Each dimension table has a unique Primary Key which is then used to create the linking relation with the fact tables. For the NW\_Trading\_DW the dimensions that I have created are Customer, Employee, Supplier which are also my key stakeholders. These contain various attributes that provide further information about the stake holders. These attributes are mentioned above and are self-explanatory. The other dimensions are : Product, which are being traded by Northwind, Category, categories information in which the

products are divided into, Calendar, which gives the date and time specific information of sales. It also gives a quarterly/ monthly or annual details.

## **Fact Tables:**

The fact tables usually contain numeric measures of the subject of analysis. I have created two aggregate fact tables:

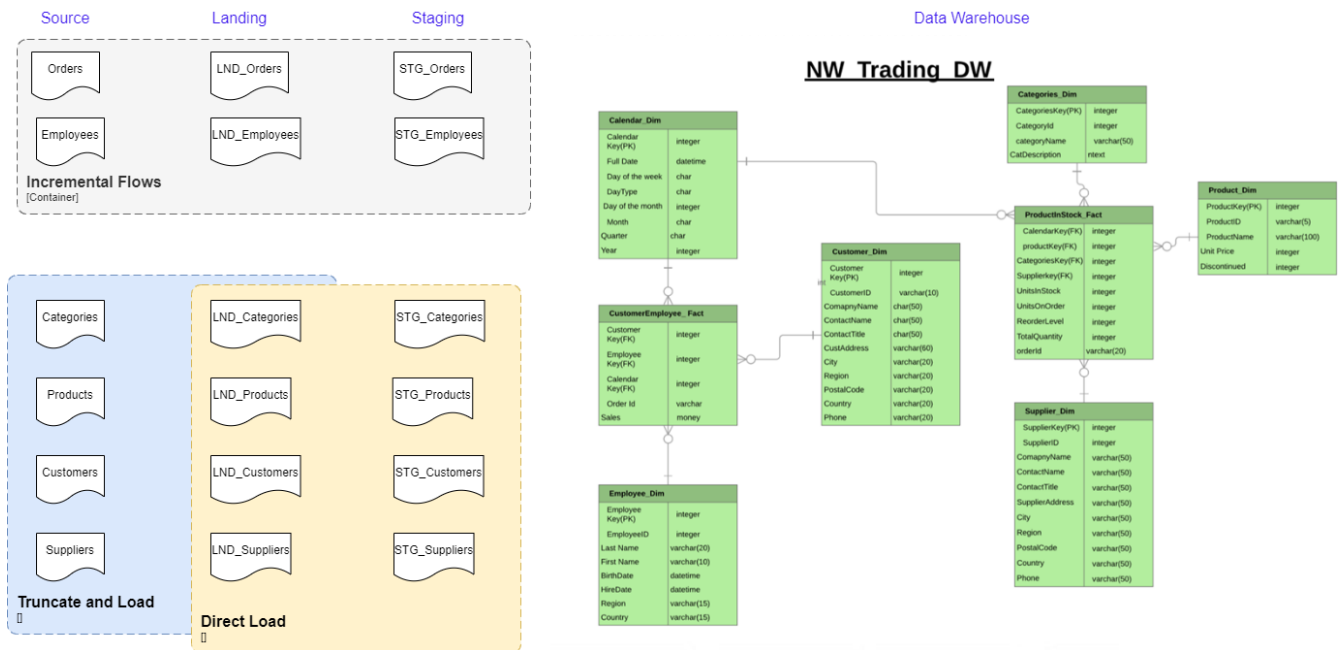
### **1. CustomerEmployee\_Fact:**

This fact table takes the Employee details from the Employee\_Dim using EmployeeKey as foreign key , Similarly, Customer details from the Customer\_Dim using CustomerKey as foreign key and date and time specific details from the Calendar\_Dim using CalendarKey as foreign key. This tables takes the dynamic value of Order Id from the source database Northwind. Together with the order id and the other three dimensions I have created an attribute of Sales that determines the total sales by each customer and the sales under each employee. This fact table can be used to determine the premium customers and best employees. Thus fulfilling the business requirements 2 and 3.This tables can be used to do specific time related analysis. Also, we can determine region specific sales and many other analysis few of which are displayed in the upcoming sections.

### **2. ProductInStock\_Fact:**

This fact is used to keep a track of the units in stock and the units on order of each product based on its reorder level. This is done to maintain the correct balance of the products stocked to their orders. It helps implement the business requirement 1 mention in section 2.1. This fact takes the order details from the source database Northwind using the attribute orderid. It also takes the Product details from the Product\_Dim using the productKey as foreign key. The category details from Categories\_Dim using CategoriesKey as foreign key and date and time specific information from the Calendar\_Dim similar to the CustomerEmployee\_Fact. This fact table also takes the Supplier information from the Supplier\_Dim the SupplierKey as Foreign key. This fact can be used to perform multiple analysis like the most selling product in each category and a few more mentioned in further sections. This fact helps fulfil requirement 4 and 5 from the business requirements section.

## Flow Diagram:



## Detailed Requirement:

Using the data from the Transactional System as the ER-Diagram in **Fig 1**. The requirement is to build a NW Trading Data warehouse System following the Dimensional Model in **Fig 2**.

1. Calender\_Dim –
2. Customer\_Dim
3. Employee\_Dim
4. Categories\_Dim
5. Product\_Dim
6. Supplier\_Dim
7. Customer\_Employee\_Fact -
8. ProductInStock\_Fact

On Creation of the DW Objects, perform the below analysis:

1. discontinued\_product\_report
2. Best and least selling product in each category
3. min max avg customer billing
4. Best Salesperson