

# North South University



## Assignment On Data Warehouse

**Course Title** : Distributed Database Systems

**Course Code** : CSE 512

**Assignment Title:** Assignment On Data Warehouse

**Department** : Electrical & Computer Engineering

### Submitted By

**Name** : Md. Mintu Hossain  
**Student ID** : 2315034650  
**Program** : M.Sc. in CSE  
**Department** : Electrical and Computer Engineering (ECE)

### Submitted To

**Name** : Dr. Abu Sayed Md. Latiful Hoque  
**Designation** : Professor,  
Department of ECE, NSU.  
**Signature** : .....

**Date of Submission:** 31 / 05 / 2023

# **Analysis**

## **1. Analyze the operational database:**

The steps in the design and implementation of the nationwide chain of the superstore's e-commerce data warehouse include selecting pertinent entities and characteristics, integrating data, extracting data, preprocessing, and uploading to the data warehouse. Based on the aforementioned examples, we may identify the following characteristics and entities.

**Supplier:** sup\_id, name, product\_type, address

**Customer:** customer\_id, name, contact, nid, address

**Item:** item\_id, item\_name, item\_type

**Superstore:** store\_id, location, trans\_id, trans\_type, time\_hour, day, week, month, quarter, year, quantity, unit\_price, total\_price

We will pick the aforementioned entities and characteristics to upload to the data warehouse.

## **2. Analyze the activities required:**

Analyze the activities required in the source sites to design the wrappers for data integration, data extraction, pre-processing and upload to DW.

**Data Integration:** For analysis, we will combine data from many sources, including supplier, item, customer, and superstore data.

**Data Extraction:** Using Extract, Transform, and Load (ETL), we will extract the pertinent information from the operational database.

**Preprocessing:** We will perform data cleaning and data transformation in order to remove any inconsistencies and errors from the data. Additionally, we will format the data so that it can be transferred to the data warehouse and used for analysis.

**Uploading:** Finally, the preprocessed data will be uploaded into the data warehouse via ETL.

## **Design**

**Task - 1:** Design the architecture of the warehouse:

**Source Systems:** These systems, which are currently in use, contain the data that must be entered into the DW. The source systems in this scenario would be the systems at the suppliers that track shipments of orders and items along with the systems at the several superstores that manage transactions and inventories.

**Data Integration Layer:** The ETL process, which extracts data from the source systems, transforms it into the right format for the warehouse and loads it into the DW, is included in this tier.

**Business Intelligence Layer:** This layer consists of the technologies and tools that let users access and evaluate the data kept in the DW.

**Data Warehouse:** All the data is kept in this single repository. Now that we have this warehouse, we can create a star schema.

Now explain the sources, preprocessing, noise reduction, transformation and uploading.

**Preprocessing:** Before the data is imported into the warehouse, any inconsistencies must be found and fixed. For example, we may need to clean up the customer's address and update the product name in different systems.

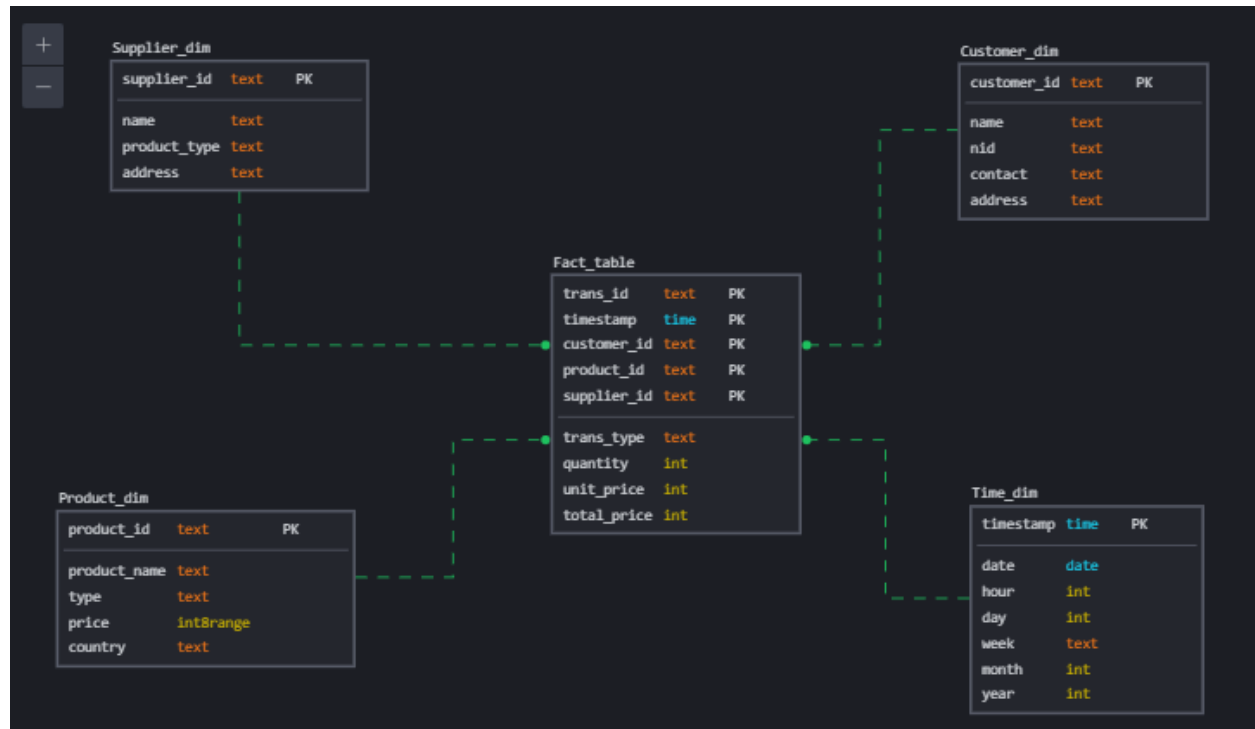
**Noise Reduction:** This entails filtering out any unnecessary or redundant data from the warehouse. For instance, we might be able to aggregate by week instead of storing every single transaction record from the superstores' systems.

**Transformation:** For the transformation, the information is in a form that the data warehouse can utilise. For instance, standardizing units of measurement or computing derived metrics like sales or profit margins may be necessary.

**Uploading:** Data warehouse loading and transformation are required during this procedure, which can be completed on a regular schedule.

**Task - 2:** Design the star schema for the warehouse using the scenario and the data set and explain how the data of the superstore database will be collected to the DW

## Star Schema



**Explanation:** Using a source-driven methodology, the data integration layer can extract data from the source systems and load it into the data warehouse in order to acquire the data from the superstore database. ETL processes, which are created to extract data from the many superstore systems, transform it into the proper format, and post it into the warehouse, can be used to do this.

**Task - 3:** Prepare a mapping of your design with the DW given in VIS and propose any change.

Both designs feature comparable columns in the fact\_table, such as customer\_key, item\_key, quantity, unit\_price, and total\_price. But unlike the e-commerce data warehouse architecture, the VIS DW design has an extra column for payment\_key, time\_key, and store\_key.

Both designs for the dimension table, such as the object and customer, have equal dimensions. A dimension for stores is added to the VIS DW design that is absent from the e-commerce data warehouse design. A transaction dimension for payment type and bank name is also present in the VIS design but not in the e-commerce data warehouse.

This comparison shows that if we were to convert the e-commerce data warehouse design to the VIS DW, we would need to add dimension tables for stores and transactions as well as columns in the fact tables for payment\_key, time\_key, and store\_key.

It should be noted that a data warehouse's design is heavily influenced by the objectives and requirements of the particular companies. Any adjustments should therefore be properly thought out in light of those factors.