Contents

PART A

PART B

1. Introduction

ABC Consumer Electronics Outlet Ltd is a multi-channel retailer of consumer electronics that has six stores operated in different locations in and around London. The Consumer Electronics retail industry is very competitive, hence the need to gain access to innovative information and communication technology to gain more advantage in this competitive business. ABC Consumer Electronics Outlet Ltd has already invested in information and communication technologies to support its business processes – retail activities, online business automation, and finance department support using Vend, Linnworks, and Xerox respectively which generate huge transactional data every day. However, the managers are still having a difficult time producing reports and performing analysis to gain insights for decision-making.

Several business requirements have been identified by the business analyst. This report is centered only on proffering a data warehousing solution for the inventory management business process. Simple Star schema dimensional data model design was adopted in this coursework with three (3) fact tables to cover the business activities of the company and four (4) dimensions to answer to the requested report and analysis needed for inventory management.

Daily, weekly and monthly reports demands are expected to be effectively generated from the warehouse system. This in turn will help the management to achieve the aim of increased revenue and customer satisfaction.

2. Business Requirements

Data warehousing business requirements help describe the solution needed in detail and business terms. They are usually deduced from the outcome of function user's requirement gathering conducted by business analysts. This will form the basis for developing a good data warehousing solution to meet all users' reporting and analysis needs. The aspect of business requirements in the development and design process of a data warehousing project cannot be overemphasized as it determines the success of the data warehousing solution. The scope of the business requirement for ABC Consumer Electronics Outlet Ltd is on Inventory Management which focuses on how managers can control and manage product stock levels. Table 1 and 2 below gives a more detailed description of the business requirement and the processes involved in developing the data warehousing solution.

Business Requirement	
Business Requirement Code	01
Business Requirement Name	Inventory Management
Priority	High
Description	

This requirement is focused on controlling and managing inventory or stocks. The main goal is to produce a good purchase order plan to ensure that products do not run out-of-stock and to also reduce overstocking of products to maximize profit and increase customer satisfaction.

Table 1: Business Requirement for Inventory Management

BUSINESS PROCESSES

Business Process Code	01
Business Process Name	Sending Purchase Orders
Priority	High
Description	
This process will keep a log of purchase orders sent, ordered quantity, sent date, product details, and supplier details for products with low/minimum levels of inventory.	
Process Events	Active, Discontinued
Measured By:	Ordered Quantity
Related To:	Product, Supplier, Date,Outlet

Business Process Code	02
Business Process Name	Receiving Purchase Orders
Priority	High
Description	
This process will manage all received orders while noting the destination outlet for storage. It will keep track of the received purchase order, quantity ordered, quantity received, received date, product details and supplier's details.	
Process Events	Active, Discontinued
Measured By:	Ordered Quantity, Received Quantity
Related To:	Product, Supplier, Date, Outlet

Business Process Code	03
Business Process Name	Controlling and Maintaining Stock
Priority	High
Description	
This process will manage stock levels for all products ensuring that new products are added while the stock levels is being adjusted for all existing products.	
Process Events	Active, Discontinued
Measured By:	StockLevel
Related To:	Product, Supplier, Date

Table 2: Business Processes for Inventory Management Requirements

3. Reporting Requirement

Inventory management is one of the most important business processes of a company. The main goal is to produce a good purchase order plan to ensure that product items are available when they are needed. The data warehousing solution is expected to meet the following reporting and analysis requirements:

• **A daily stock levels of all products for the last month**: This involves all available stocks of all products in the last month (a monthly report)

• **A weekly report of all products with minimum stock levels:** This is a report of all products that fall within the limit of the stock level (a weekly report). This helps the company to keep tabs on their stocks to know when they are over stock or understocked.

• **Analyzing stock levels by brand or product type or supplier**: This report involves segregation and analysis of stock levels by brand or product type or supplier. This report helps the company to have information on their stocks based on the brand or product type or supplier.

• **Daily and weekly sent and received stock orders for the last four weeks**: This report helps to monitor stock orders daily and weekly so the company can have information on their level of demand and supplies.

• **Analyzing received stock orders by the supplier and by the month**: This report helps to monitor who the suppliers are and how many orders they received from them every month.

Based on the problems the company is facing in controlling over-stocks and under-stocks of product items. The above report requirement will give an overview of the company's well-being and in turn reduce over-stocks and under-stocks.

The report when properly utilized will increase revenue and customer satisfaction.

4. Understanding the Business Data

This coursework utilizes the data sourced from Vend which is a cloud-based software used by the company to support its retail activities. The data contains details of the products, suppliers, and stock movements from suppliers to warehouse locations. There are five (5) data flat files provided, which are as follows:

- Generated Date and Time
- Products Details.
- Suppliers Details.
- Sent Purchase Orders.
- Received Purchase Orders.

The generated date and time are in a .txt flat-file format and it is within the year range 01/01/2006 - 01/01/2027. This data will be used to generate date and time dimensions. It contains the following columns:

Column Names	Data Types	Description
FullDate	DateTime	Stores date in (dd/mm/yyyy) format
DateName	DateTime	Stores date in (yyyy/mm/dd) format
DayOfWeek	Int	Stores day of the week- 7 days (values from 1-7)
DayNameOfWeek	Str	Stores day name of the week (Sunday to Monday)
DayOfYear	Int	Stores day of the year – 365 days a year
WeekOfYear	Int	Stores week of the year – 52 weeks of the year
MonthName	Str	Stores month Name – January to December
MonthOfYear	Int	Stores month number – 12 months a year
CalendarQuarter	Int	Stores quarter of the year.
CalendarYear	Str	Stores year in (yyyy) format.

Table 3: Generated Date Attributes and Description

The supplier data file include SupplierName, Description, Phone, Email, Fax, FirstLineAddress, PostCode, City, State, and CountryID. The solution will use the attributes to identify and differentiate each supplier.

The product data file is SKU, ProductName, Description, Condition, ProductType, Brand, SupplierName, Tags, CostPrice, RetailPrice, CurrentStockLevel, DateCreatedAt, DateDiscontinuedAt, IsActive. The CurrentStockLevel is used for observing and maintaining stock trends.

For stock movements, the tables SentPurchaseOrders and ReceivedPurchaseOrders are used.

The ReceivedPurchaseOrder has PurchaseOrderCode, ProductSKU, SupplierName, DestinationOutletID, SentDate, ReceivedDate, ReceivedQty, and OrderedQty attributes.

The SentPurchaseOrders include PurchaseOrderCode, ProductSKU, SupplierName, DestinationOutlet, SentDate, and OrderedQty.

5. Schema and Granularity

5.1. Identify Grain Level

The grain of a fact table **represents the atomic level by which the facts may be defined**. The grain of a sales fact table might be stated as "sales volume by day by product by store". Each record in this fact table is therefore uniquely defined by a day, product, and store.

5.2. Grain Declaration

The records in the fact table are on the day level, outlet level, and on product level. This granularity defines each record as uniquely as possible and all requirements from the managers can be fulfilled. The following identifies the grain of the three different central fact tables:

1. Sending Purchase Orders: A line item on a purchase order sent to suppliers daily
2. Receiving Purchase Orders: A line item on a purchase order received from suppliers and stored in a location daily.
3. Controlling and Maintaining Stocks: An individual line item is added to products while updating stock levels daily.

5.3. Identify Dimensions and their Attributes

This data warehousing solution identified five (5) Dimensions. These dimensions are briefly described and justified below:

- **DATE DIMENSION:** This dimension is very important as it will be useful in the daily, weekly, and monthly report generation and analysis to meet the reporting analysis. The attributes identified based on the available data source include:
  FullDate(PK)
  DateName
  DayOfWeek
  DayNameOfWeek
  DayOfYear
  WeekOfYear
  MonthName
  MonthOfYear
  CalendarQuarter
  CalendarYear

- **PRODUCT DIMENSION:** This is a slowly changing dimension that contains the product details. The attributes identified based on the available data source include:
  SKU (PK)
  ProductName
  Description
  Condition
  ProductType
  Brand
  Tags
  CostPrice
  RetailPrice
  CurrentStockLevel
  DateCreatedAt
  DateDiscountinuedAt

IsActive

- **SUPPLIER DIMENSION:** This is a slowly changing dimension that contains the supplier's details. The attributes identified based on the available data source include:
  SupplierID (PK)
  SupplierName
  Description
  Phone
  Email
  Fax
  FirstLineAddress
  PostCode
  City
  State
  CountryID

- **OUTLET DIMENSION:** This is a slowly changing dimension that contains the Outlet or storage location details. The available data source contains only the OutletName and does not include detailed information on the Outlet.
  OutletID (PK)
  OutletCode
  OutletName
  OutletAddress
  OutletPhone
  OutletEmail
  City
  State
  CountryID

- **PURCHASEORDER DIMENSION:** This is a degenerate dimension that stores the details of the purchase order.
  PurchaseOrderCode (PK)

5.4. Identify Dimensions and Measures of the Fact Tables

The dimensions and measures identified for each of the fact tables are as follows:

5.4.1. Sending Purchase Orders

The dimensions identified for the SendingPurchaseOrder Fact table are as follows:

- Date Dimension
- Product Dimension
- Supplier Dimension

The measure identified is

- Ordered Quantity

5.4.2. Receiving Purchase Orders

The dimensions identified for the ReceivingPurchaseOrder Fact table are as follows:

- Date Dimension
- Product Dimension
- Supplier Dimension
- Outlet Dimension

The measures identified are:

- Ordered Quantity
- Received Quantity

5.4.3. Controlling and Maintaining Stock

The dimensions identified for the ControllingandMantainingStocks Fact table are as follows:

- Date Dimension
- Product Dimension
- Supplier Dimension

The measure identified is

- CurrentStockLevel

6. Data Warehouse Schema and Design

A simple star schema dimensional data model design was used to design the models. Fig 1, 2 and 3 show a simple star schema for Controlling and Maintaining Stocks, Receiving Purchase Orders, and Sending Purchase Orders fact tables respectively. Fig 4 shows a combined view of all the schemas on a page.

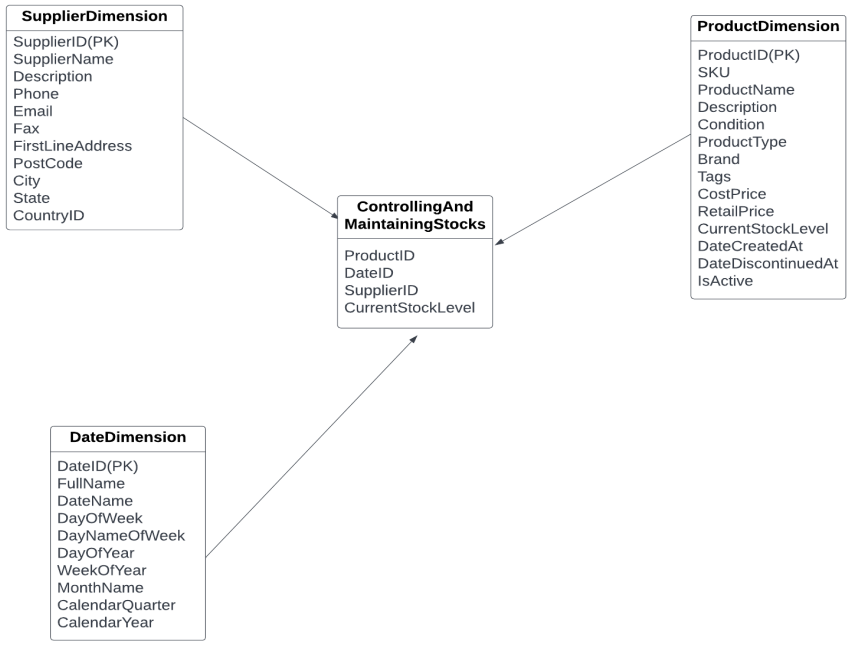


Figure 1: A simple Star Schema for Controlling and Maintaining Stocks

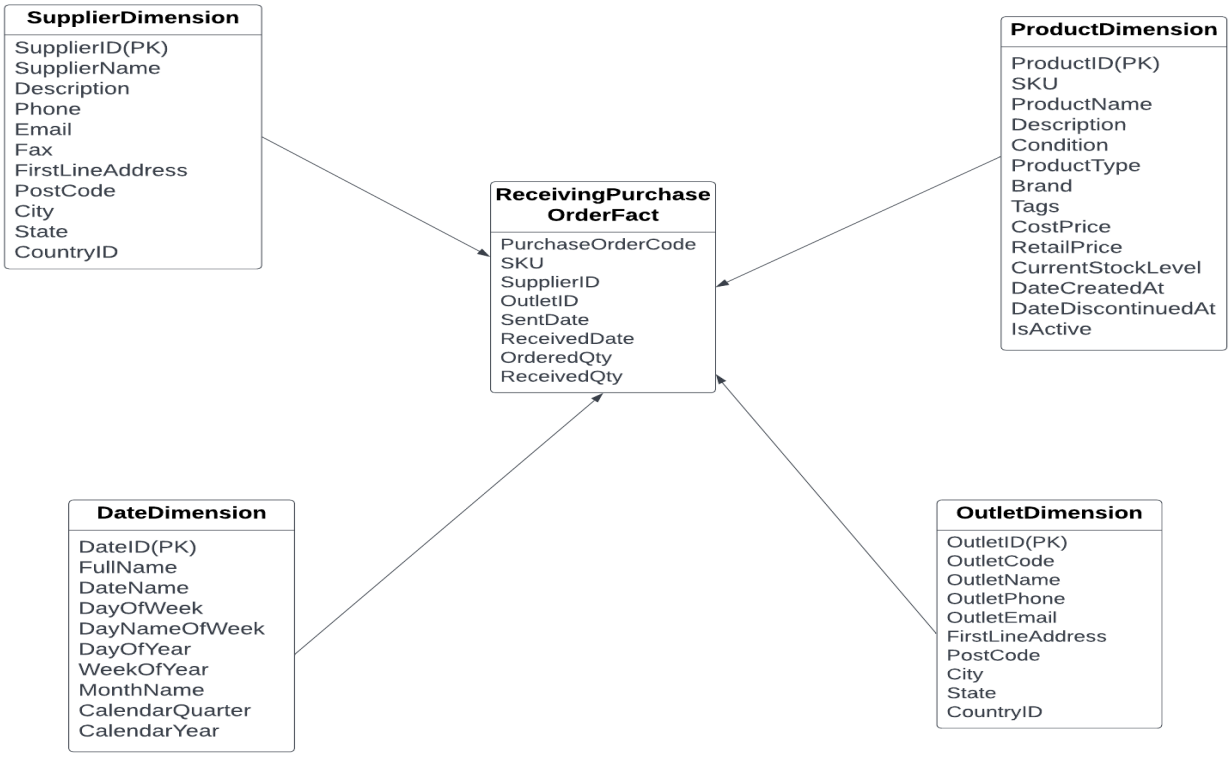


Figure 2: A simple Star Schema for Receiving Purchase Orders fact table

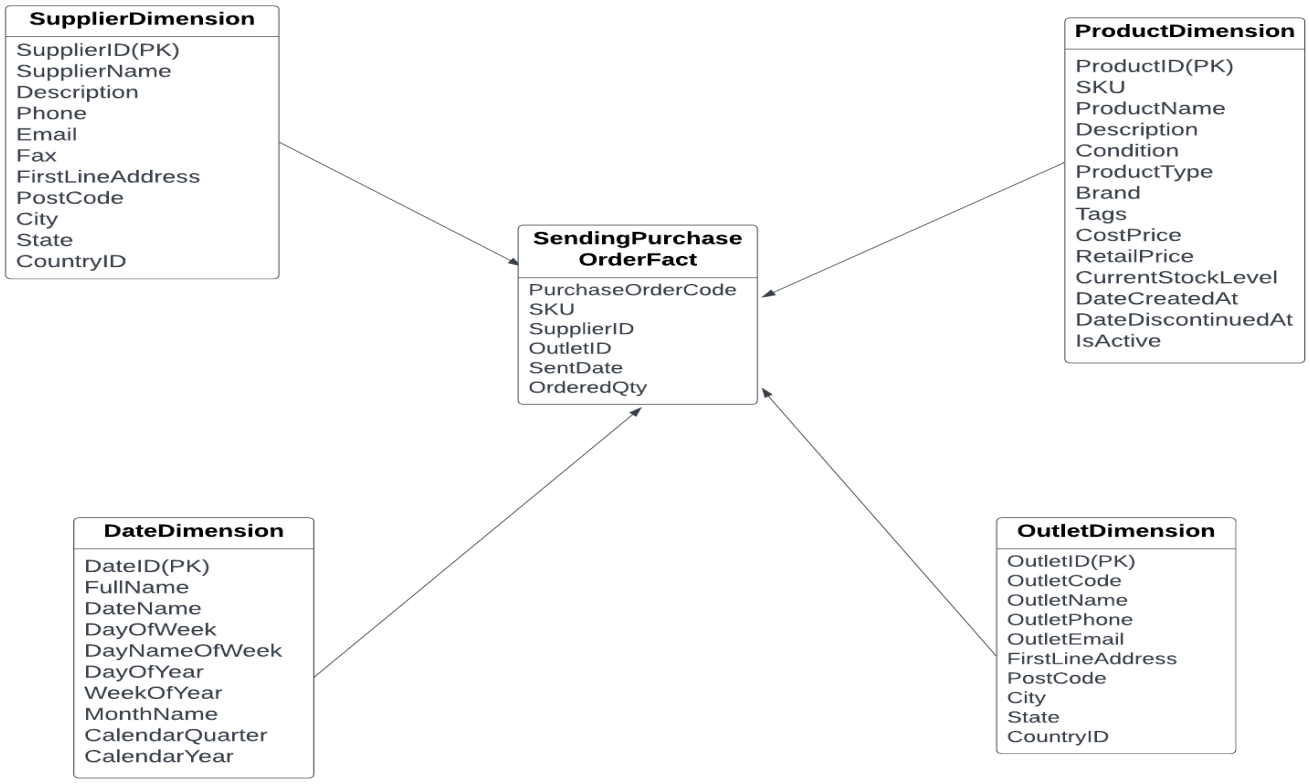


Figure 3: A simple Star Schema for Receiving Purchase Orders fact table

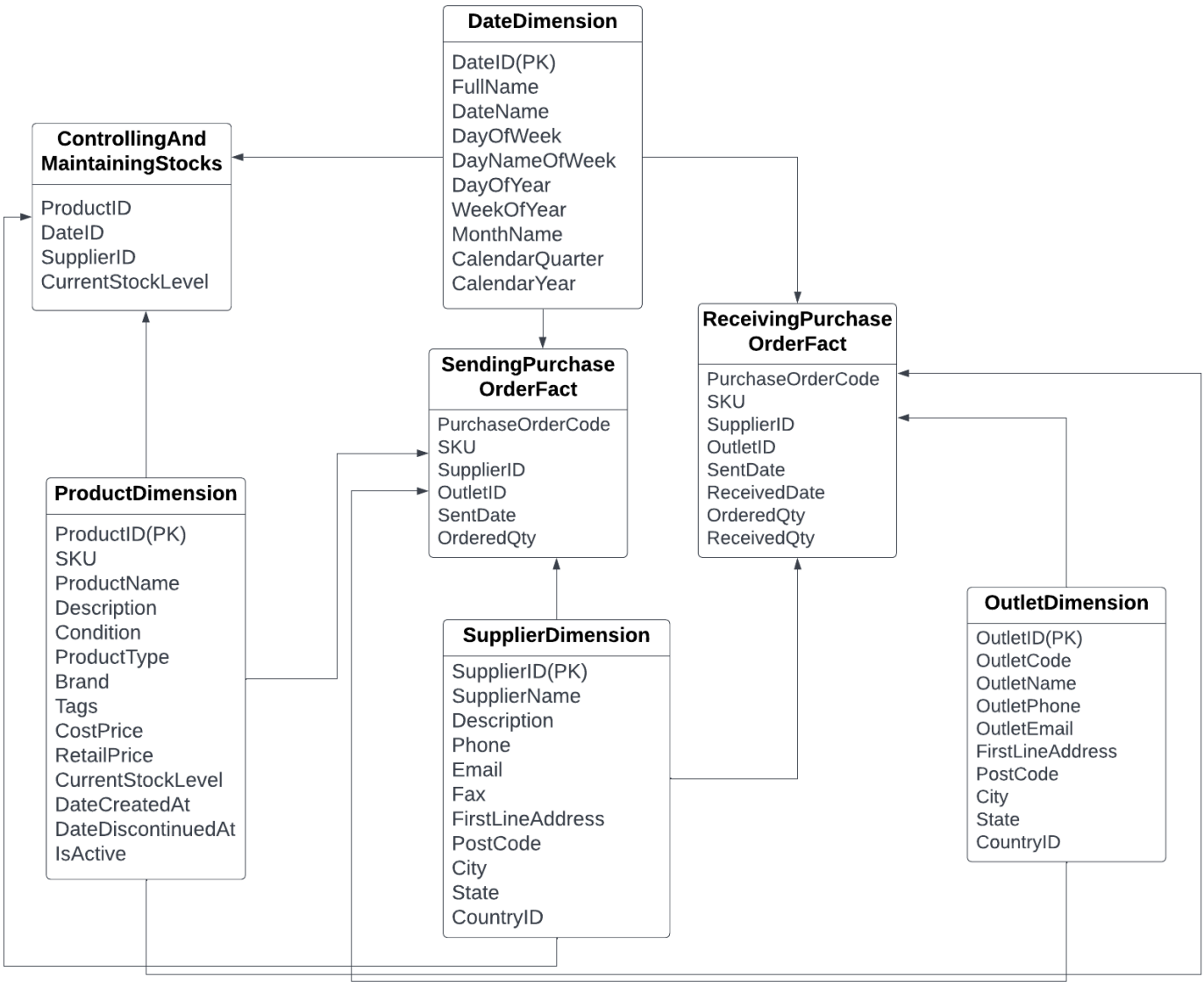


Figure 4: Overall Graphical Representation of all the schemas

7. Further Works

The design of the star schema will be implemented on the SQL server. The business and reporting requirements will be checked at all times during the implementation phase to ensure that the data warehouse conforms to the requirements for reporting and analysis. If the requirements are not being met, the design phase will be updated to meet the requirements.

PART B

1. Introduction

The business case scenario of this coursework is based on ABC Consumer Electronics Outlet Ltd- which is in the retail industry for consumer electronics in London and Europe. The company has invested in Vend, Linnworks, and Xerox to support its business processes which generate huge transactional data daily. Despite their investments in information and communication technologies, the managers still have difficulty in producing and performing analysis to gain insights for decision-making.

This coursework aims to design, implement, and test a data warehouse solution for the inventory management process of the ABC Consumer Electronics Outlet Ltd business case scenario, to export data from it and ingest it for further processing on a Big Data platform. The Objectives are as follows:

- Analyse and design of the data dimensional model.
- Create a Relational Database to store dimensions and fact tables using Microsoft SQL Server Management Studio.
- Migrate test data from the data warehouse to an Apache Hadoop platform (Hortonworks Data Platform) for further analysis of Big Data.
- Implement and test the Big Data Storage on Apache HDFS
- Implement a suitable data structure to load data into HIVE using Data Analysis Studio (DAS)
- Implement and manipulate the loaded data using Pig

2. Analysis and Design of a data dimension data model

The first step in the data dimensional modeling is business and reporting requirement gathering and grain declaration. The scope of the business requirement for ABC Consumer Electronics Outlet Ltd is on Inventory Management which focuses on how managers can control and manage product stock levels. Table 1 and 2 below gives a more detailed description on the business requirement and the processes involved in developing the data warehousing solution.

Business Requirement	
Business Requirement Code	01
Business Requirement Name	Inventory Management
Priority	High
Description	
This requirement is focused on controlling and managing inventory or stocks. The main goal is to produce a good purchase order plan to ensure that products do not run out-of-stock and to also reduce overstocking of products to maximize profit and increase customer satisfaction.	

Table 4: Business Requirement for Inventory Management

BUSINESS PROCESSES			
Business Process Code	01	02	03
Business Process Name	Sending Purchase Orders	Receiving Purchase Orders	Controlling
Priority	High	High	High
Description	This process will keep a log of purchase orders sent, ordered quantity, sent date, product details, and supplier details for products with low/minimum levels of inventory.	This process will manage all received orders while noting the destination outlet for storage. It will keep track of the received purchase order, quantity ordered, quantity received, received date, product details and supplier's details.	This process for all prod products are levels is bein products.
Process Events	Active, Discontinued	Active, Discontinued	
Measured By:	Ordered Quantity	Ordered Quantity, Received Quantity	Stock Level
Related To:	Product, Supplier, Date	Product, Supplier, Date, Outlet	Product, Sup

Table 5: Business Processes for Inventory Management Requirements

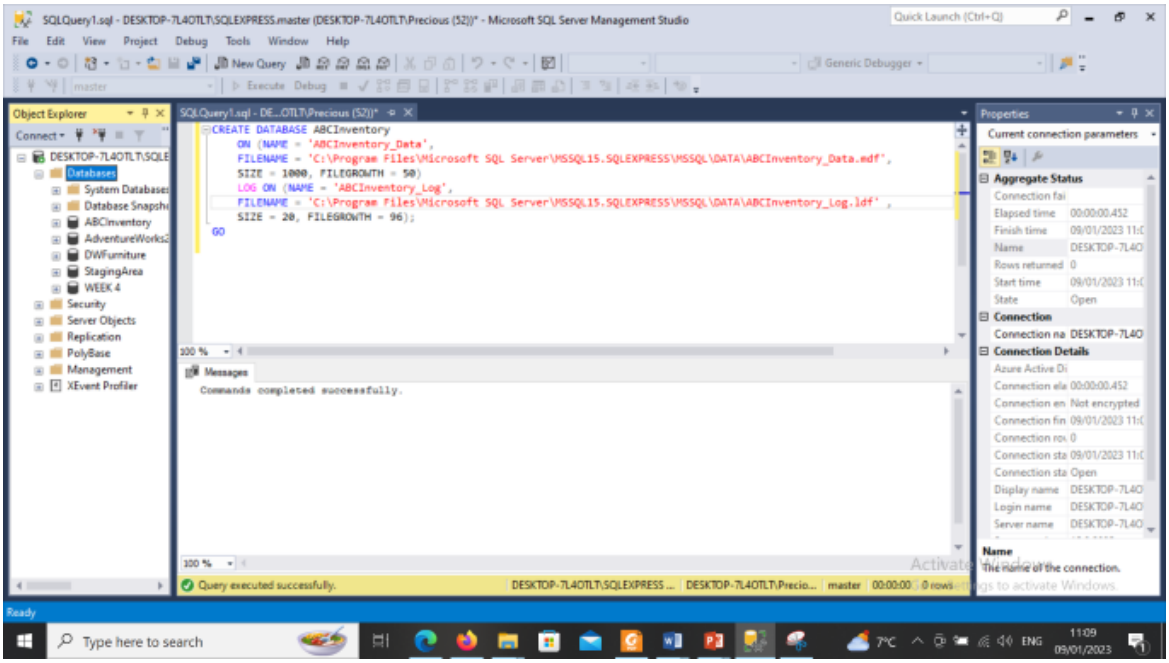
The reporting requirement required are as follows:

- A daily stock levels of all products for the last month
- A weekly report of all products with minimum stock levels
- Analyzing stock levels by brand or product type or supplier
- Daily and weekly sent and received stock orders for the last four weeks
- Analyzing received stock orders by supplier and by month

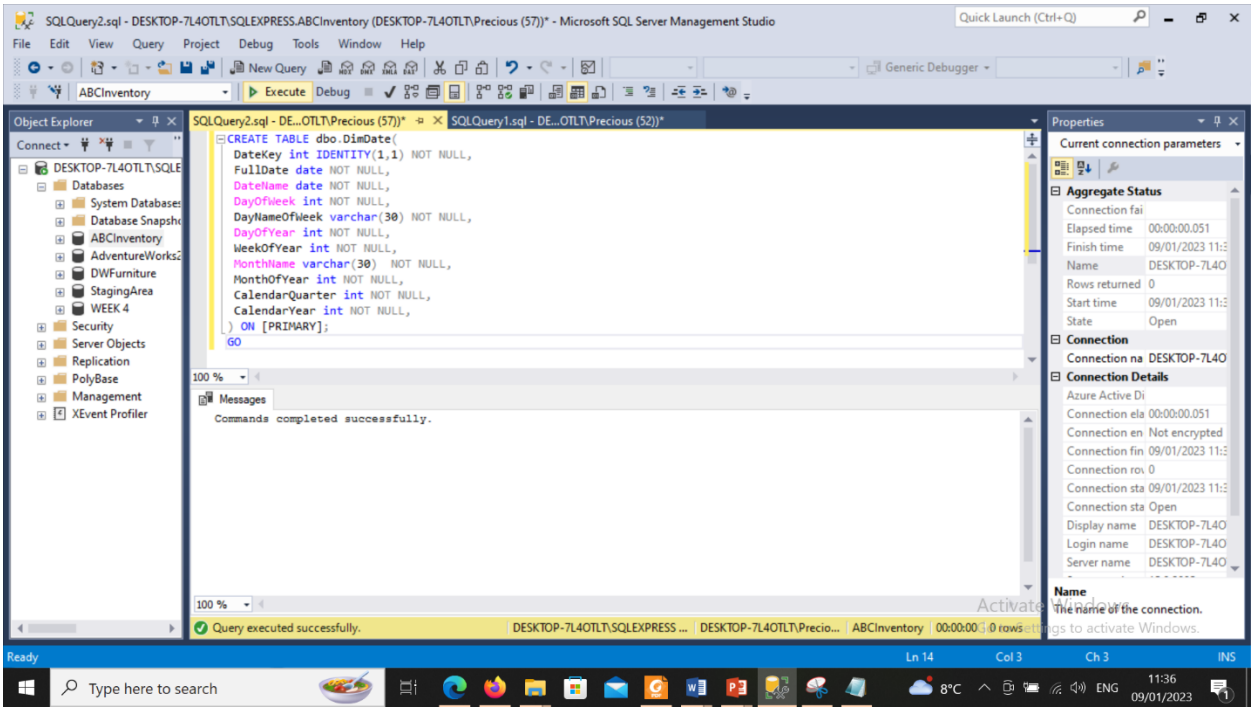
3. Implementation and Testing of Data Warehouse

The data warehouse was implemented using the SQL Server Management Studio (SSMS).

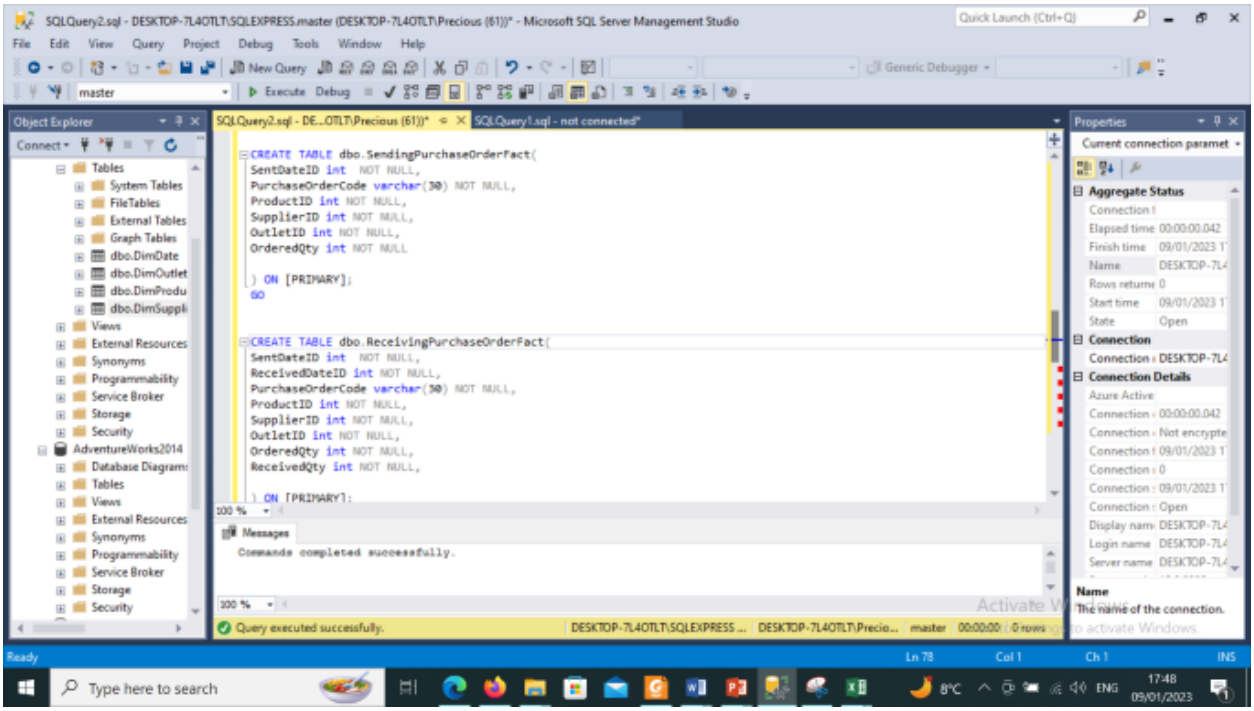
3.1. Creating Database



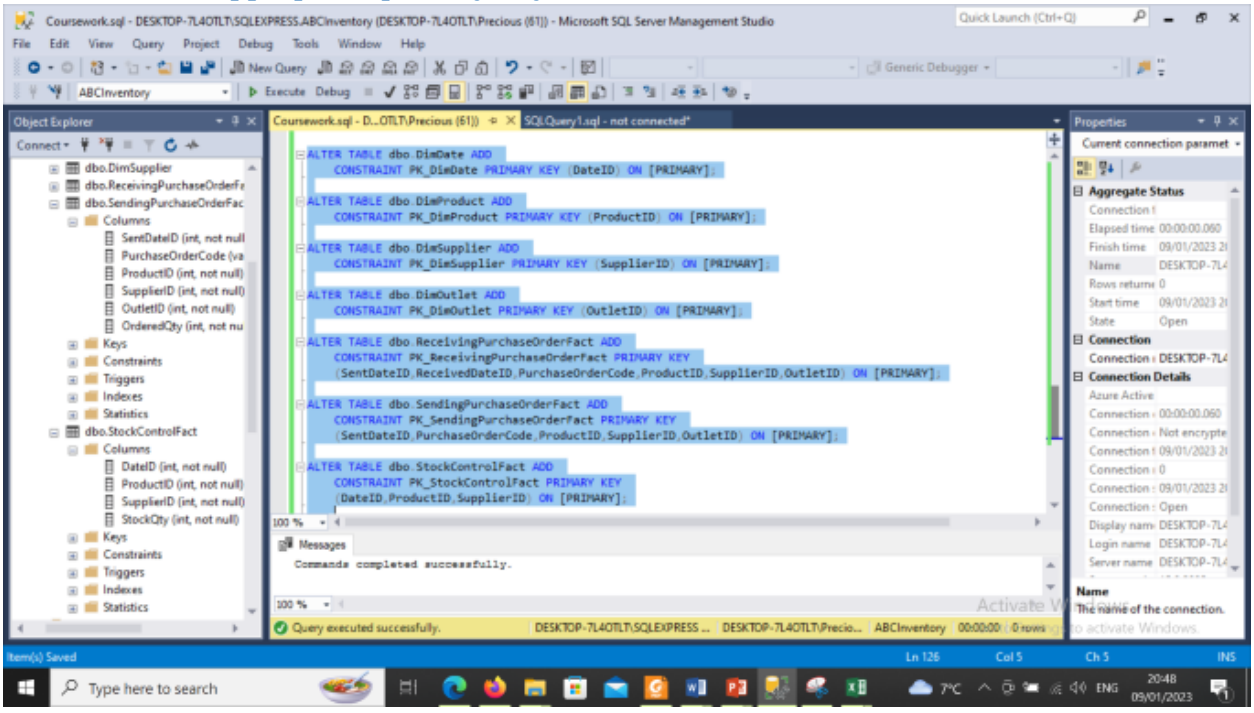
3.2. Create Dimension tables



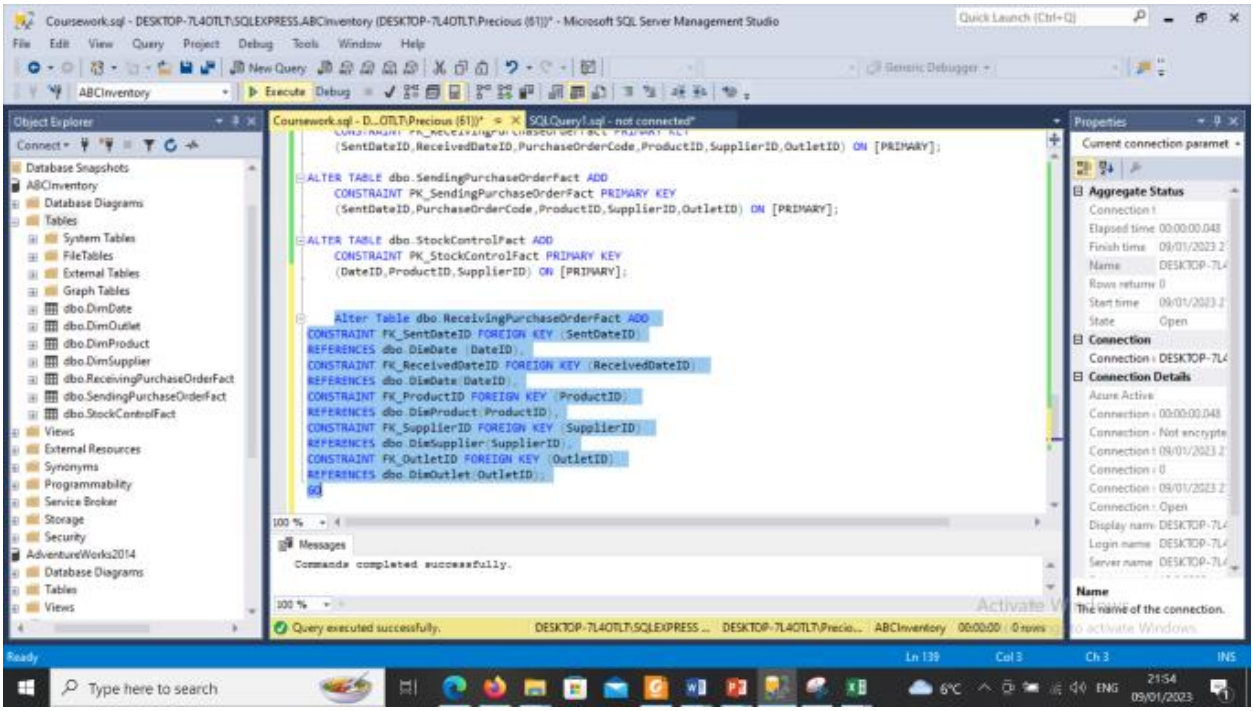
3.3. Create Fact tables



3.4. Add appropriate primary key constraints



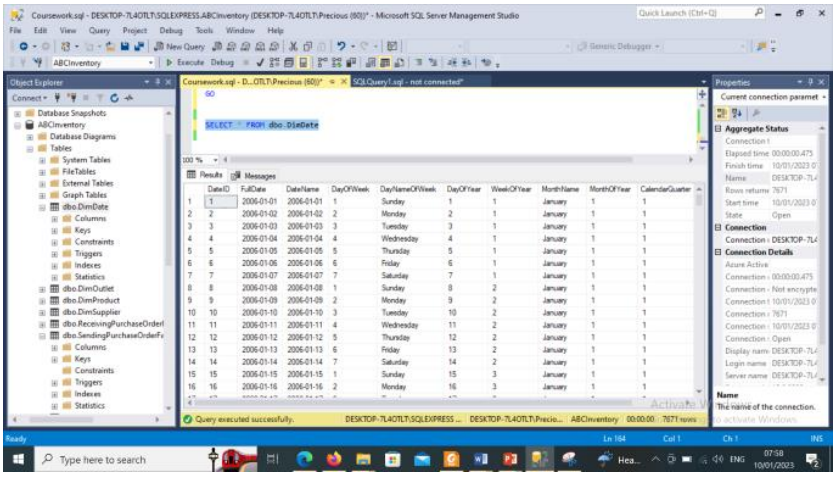
3.5. Add appropriate foreign key constraints



4. Populating the Data Warehouse

The SQL Server Import and Export Wizard was used to import all the test data into the datawarehouse

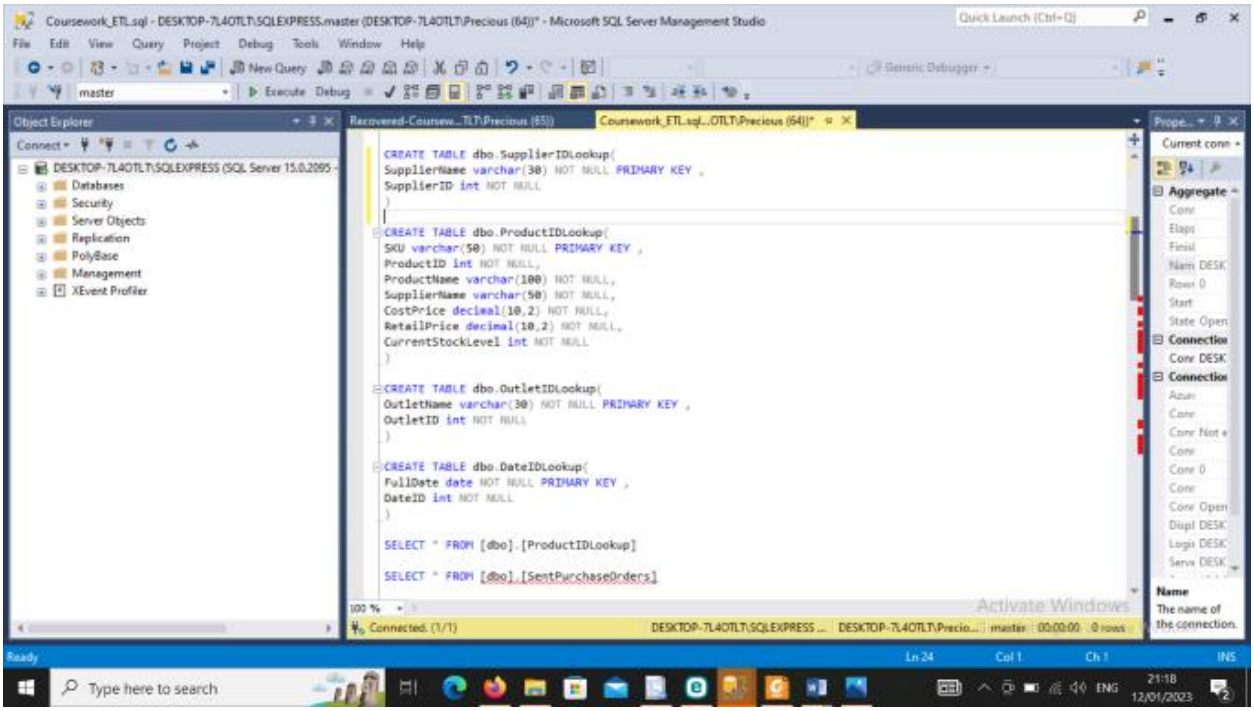
4.1. Populate Dimensions



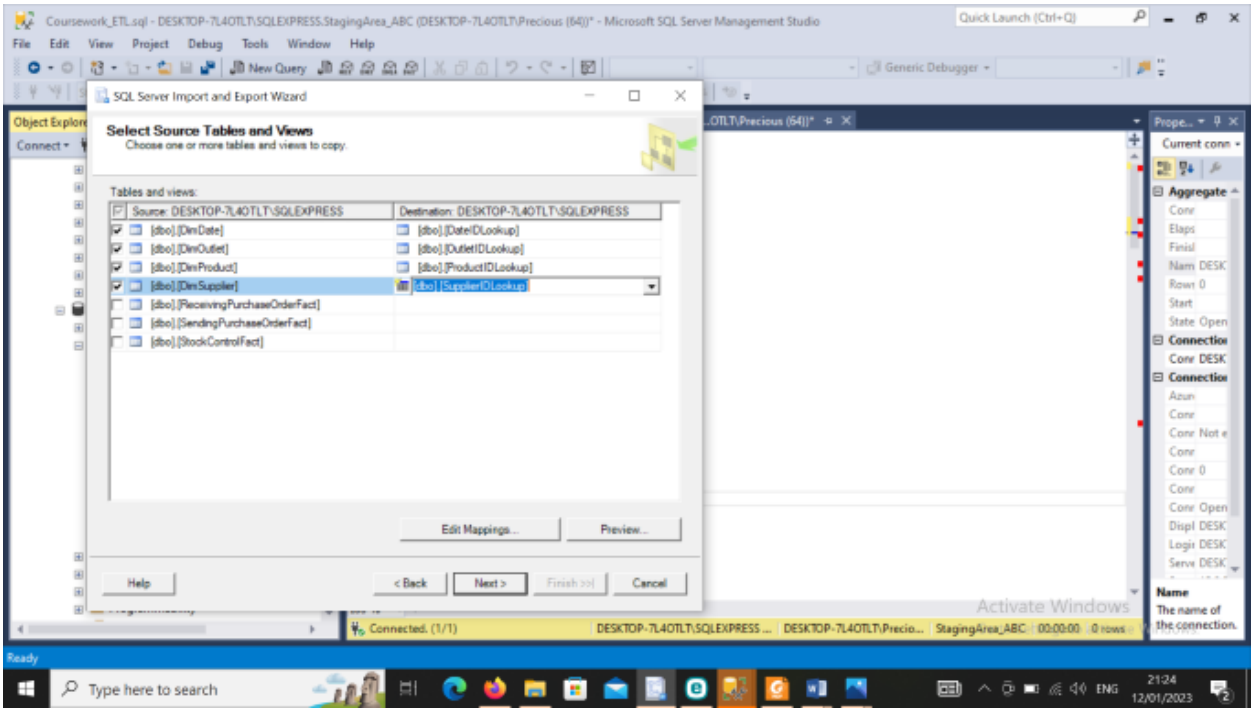
4.2. Populate Fact Tables

Inorder to populate the fact tables with test data, a staging area named “StagingArea_ABC” was created to house the lookup tables.

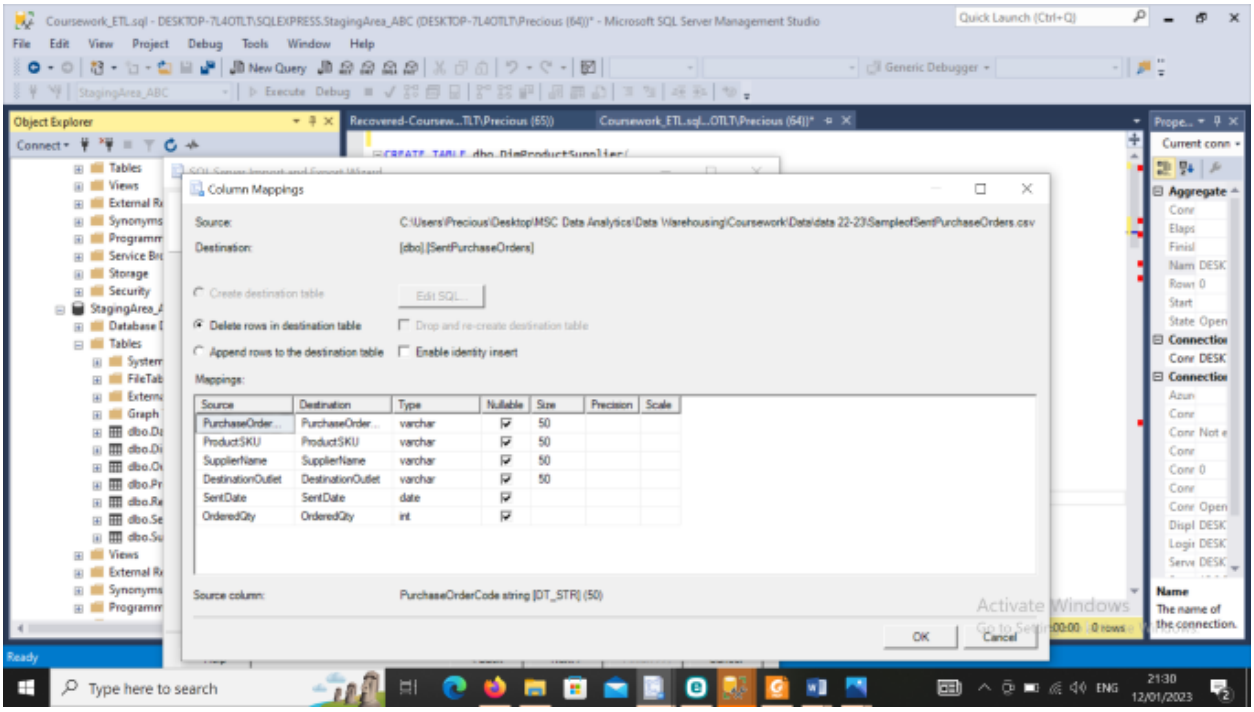
4.2.1. Creating Lookup tables in the Staging Area



4.2.2. Populating Lookup tables

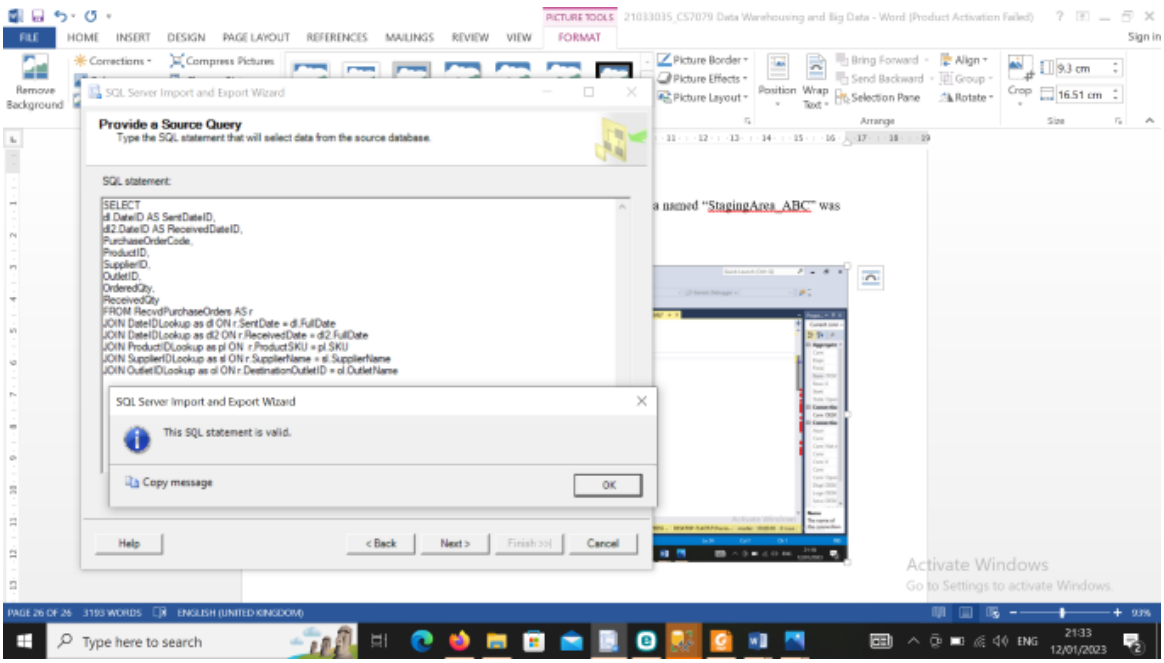


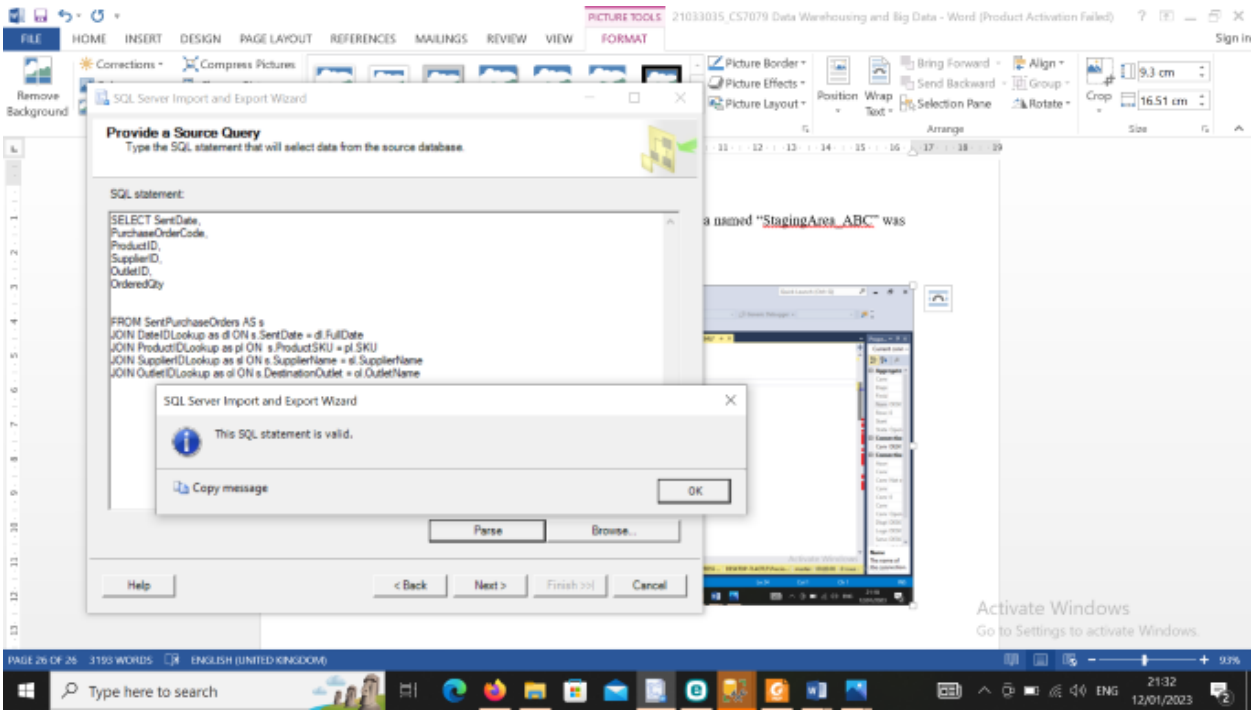
4.2.3. Populating the transaction lookup table in the Staging Area

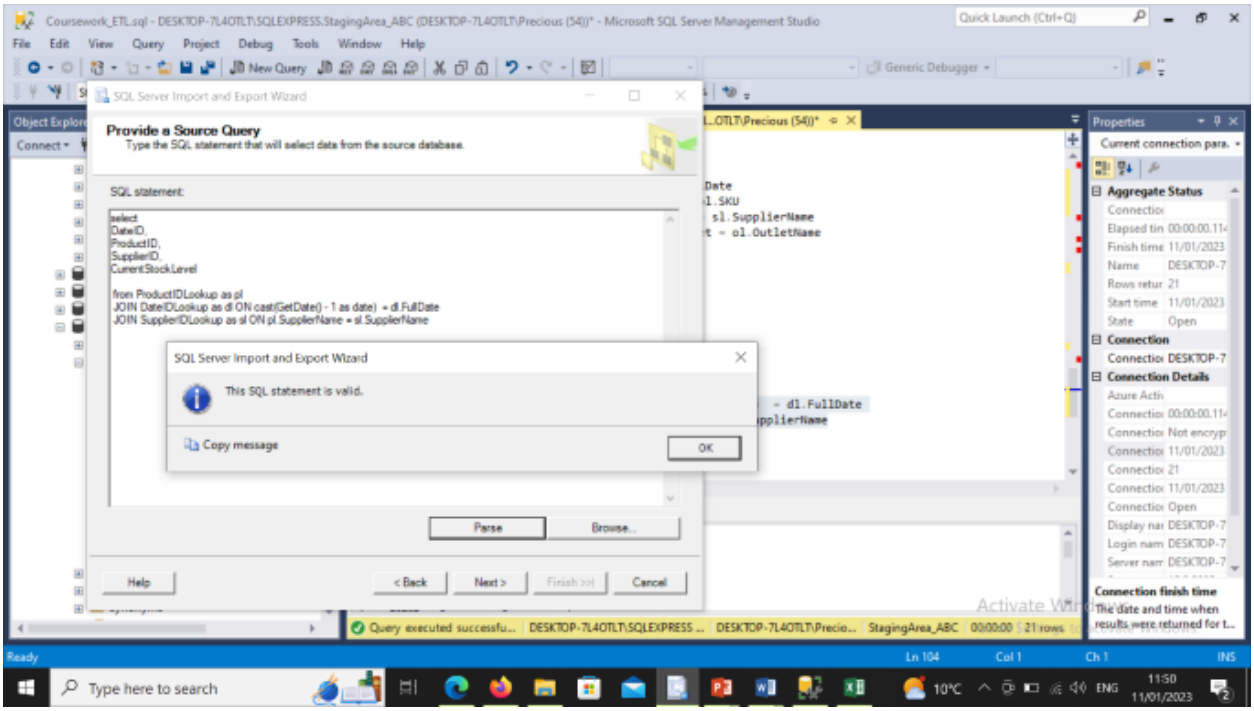


4.2.4. Populating the Fact tables

Once all the lookup tables in the staging area have been populated. The test data now needs to be transferred from the staging area to the main ABCInventory database.

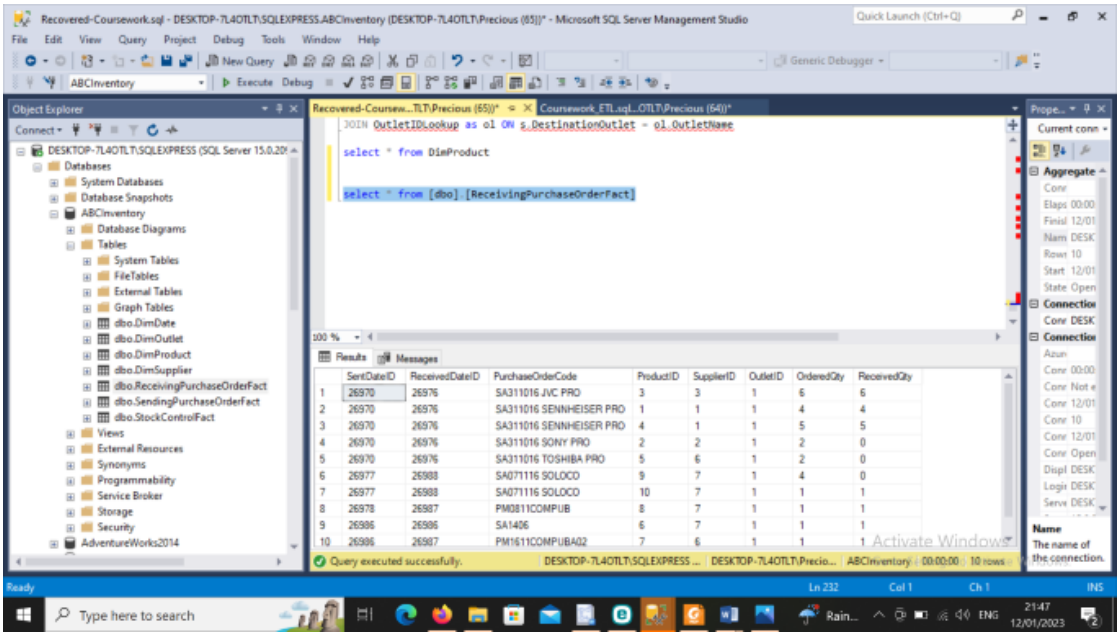


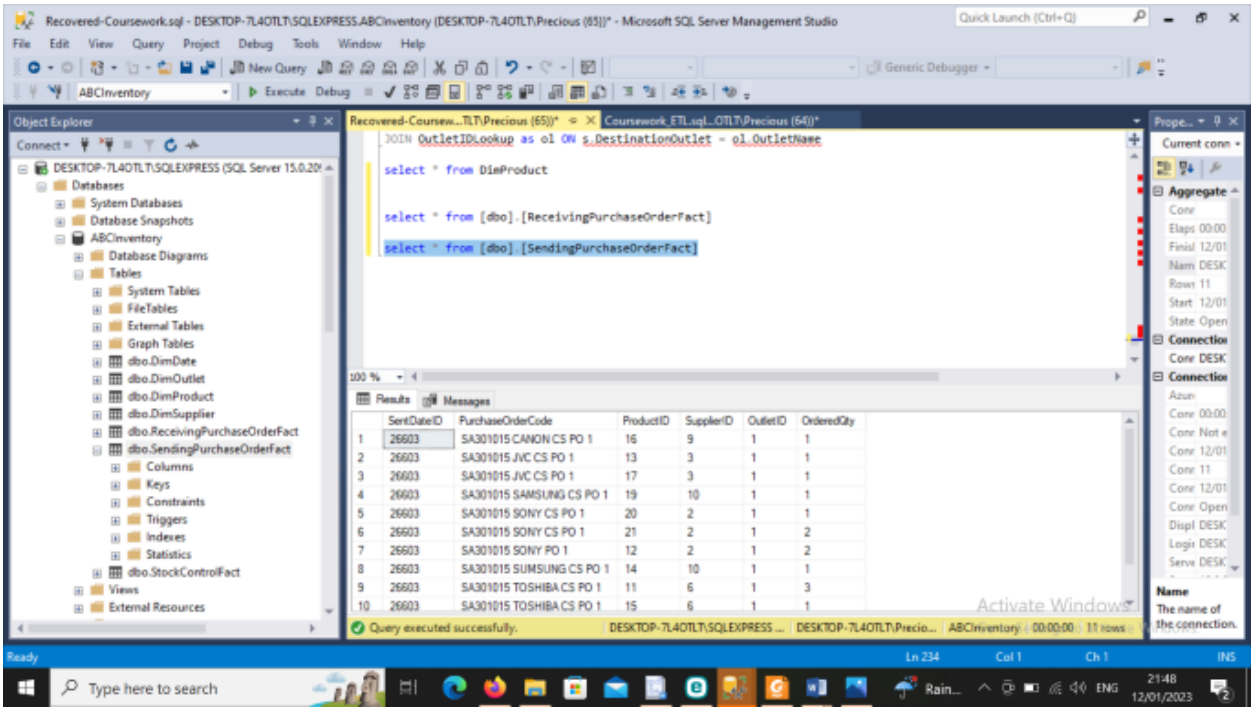


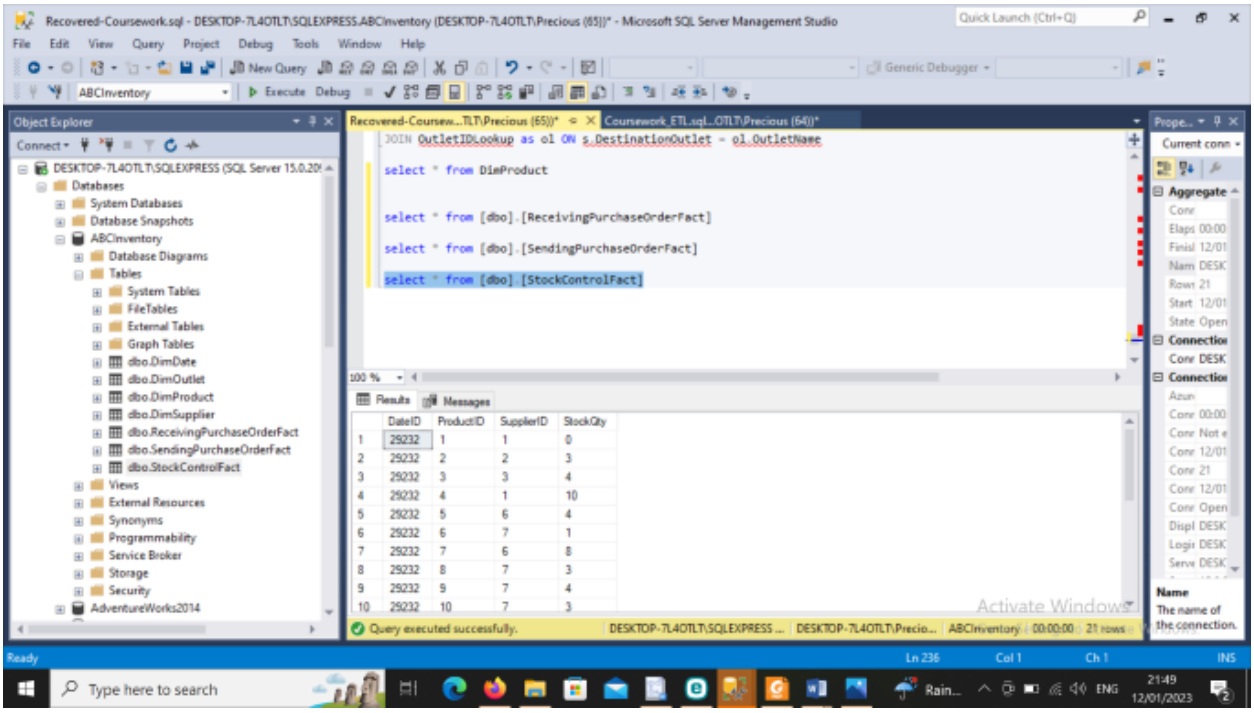


5. Test implementation

To test that the tables were fully populated, the figure below shows the results of the queries from all the tables.


Recovered-Coursework.sql - DESKTOP-7L4OTLT\SQLEXPRESS.ABCInventory (DESKTOP-7L4OTLT\Precious (65))* - Microsoft SQL Server Management Studio
JOIN OutletIDLookup as ol ON s.DestinationOutlet = ol.OutletName
select * from DimProduct
select * from [dbo].[ReceivingPurchaseOrderFact]
Query executed successfully.

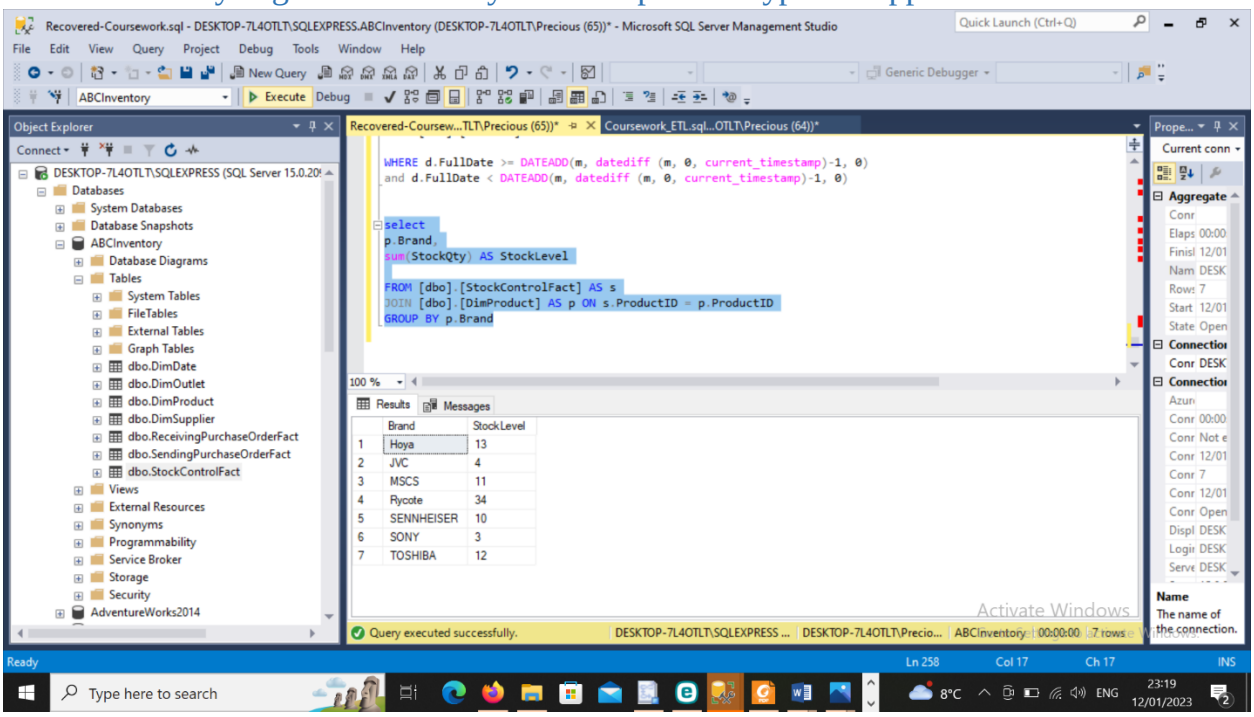

Recovered-Coursework.sql - DESKTOP-7L4OTLT\SQLEXPRESS.ABCInventory (DESKTOP-7L4OTLT\Precious (65))* - Microsoft SQL Server Management Studio
JOIN OutletIDLookup as ol ON s.DestinationOutlet = ol.OutletName
select * from DimProduct
select * from [dbo].[ReceivingPurchaseOrderFact]
select * from [dbo].[SendingPurchaseOrderFact]
Query executed successfully.

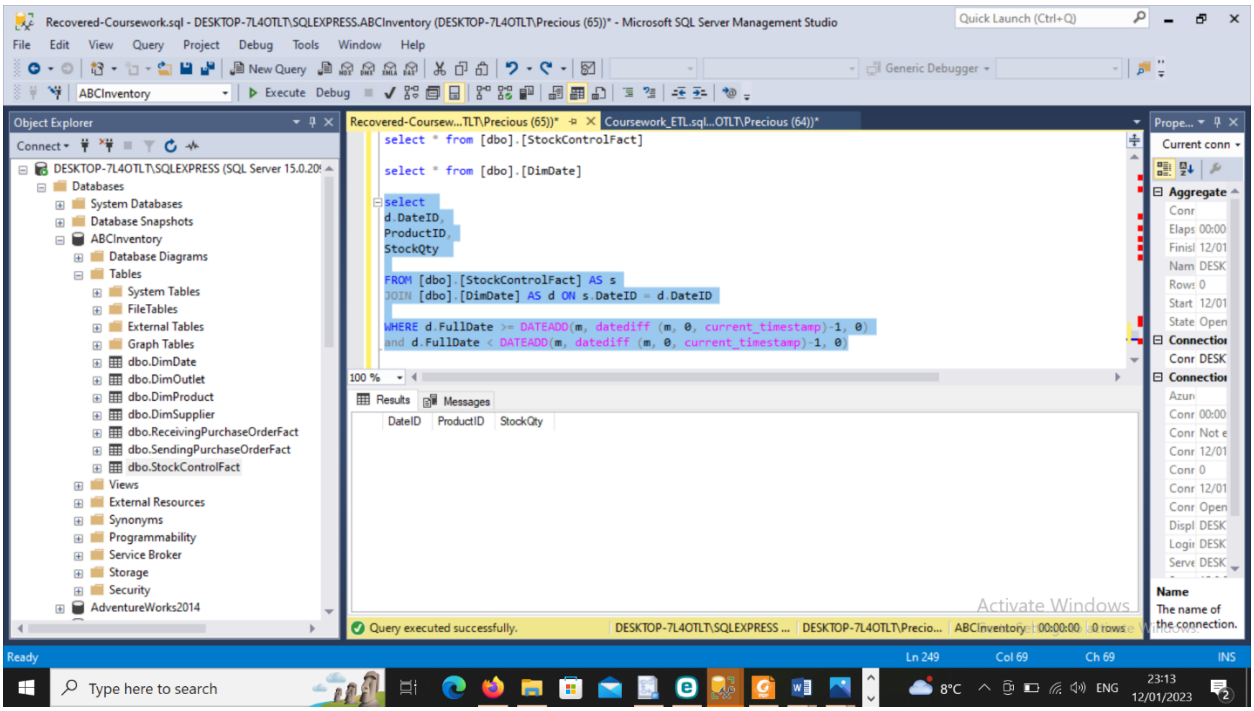


6. Testing some of the Business Requirements

To ensure the functionality of the data warehouse and its alignment with business and reporting needs, SQL queries were formulated and executed to verify compliance with these requirements.

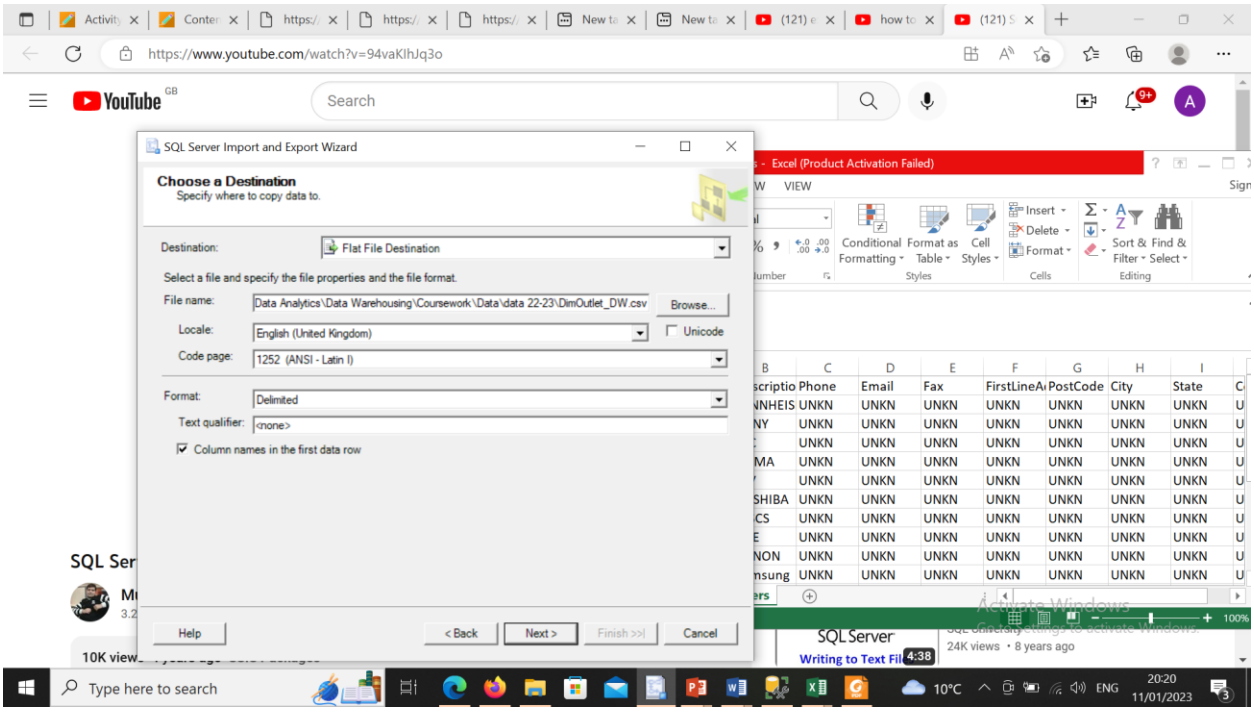
6.1. Analysing stock levels by brand or product type or supplier





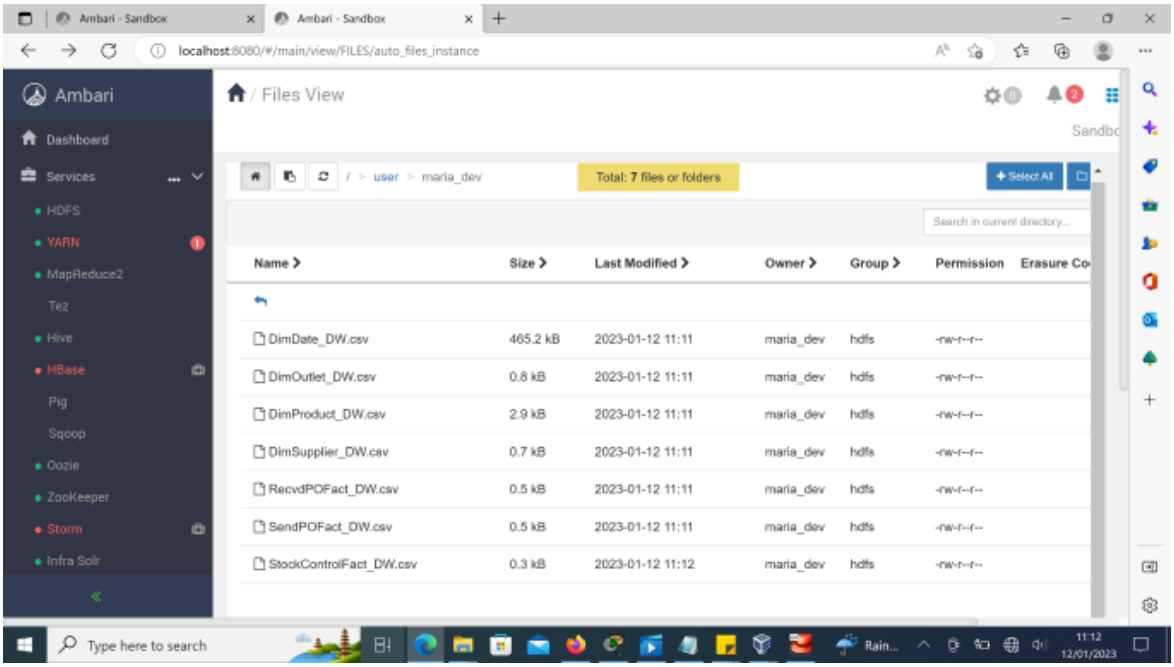
7. Exporting the data warehouse to a flat file destination.

The SQL Server import and export wizard was used to export the datawarehouse tables to external files. The Datasource is the ABCInventory Database while the destination is Flat File Destination

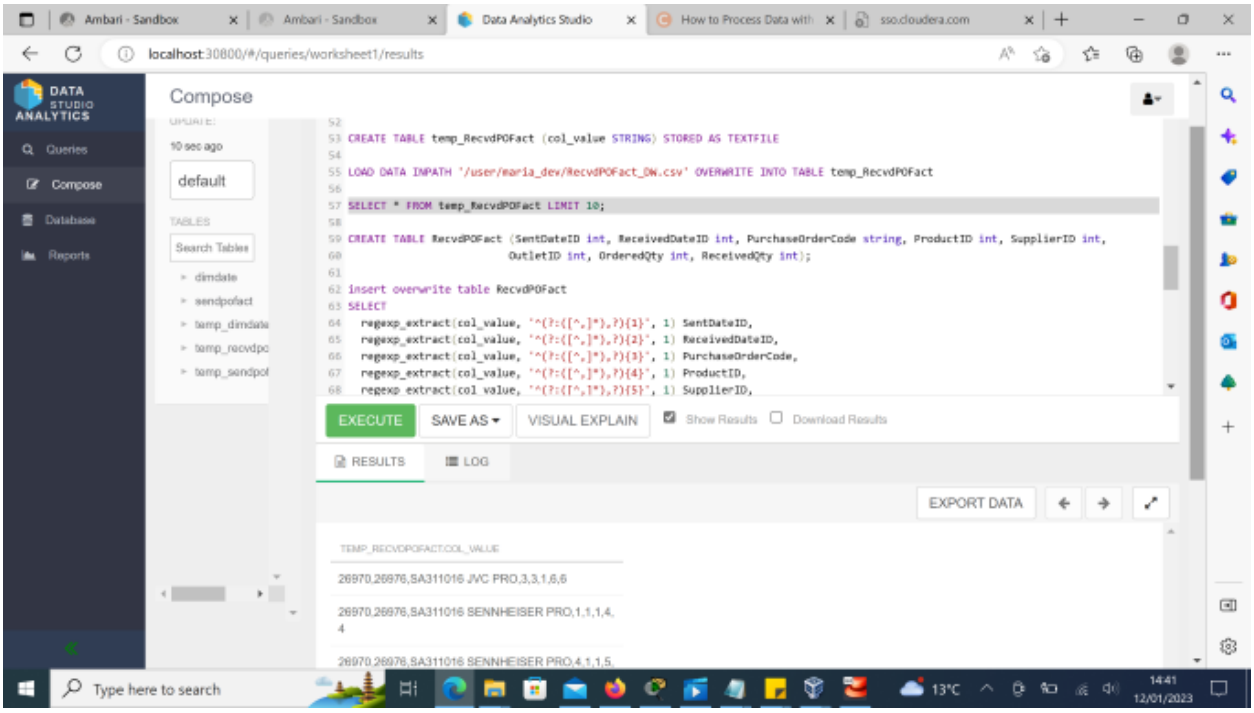


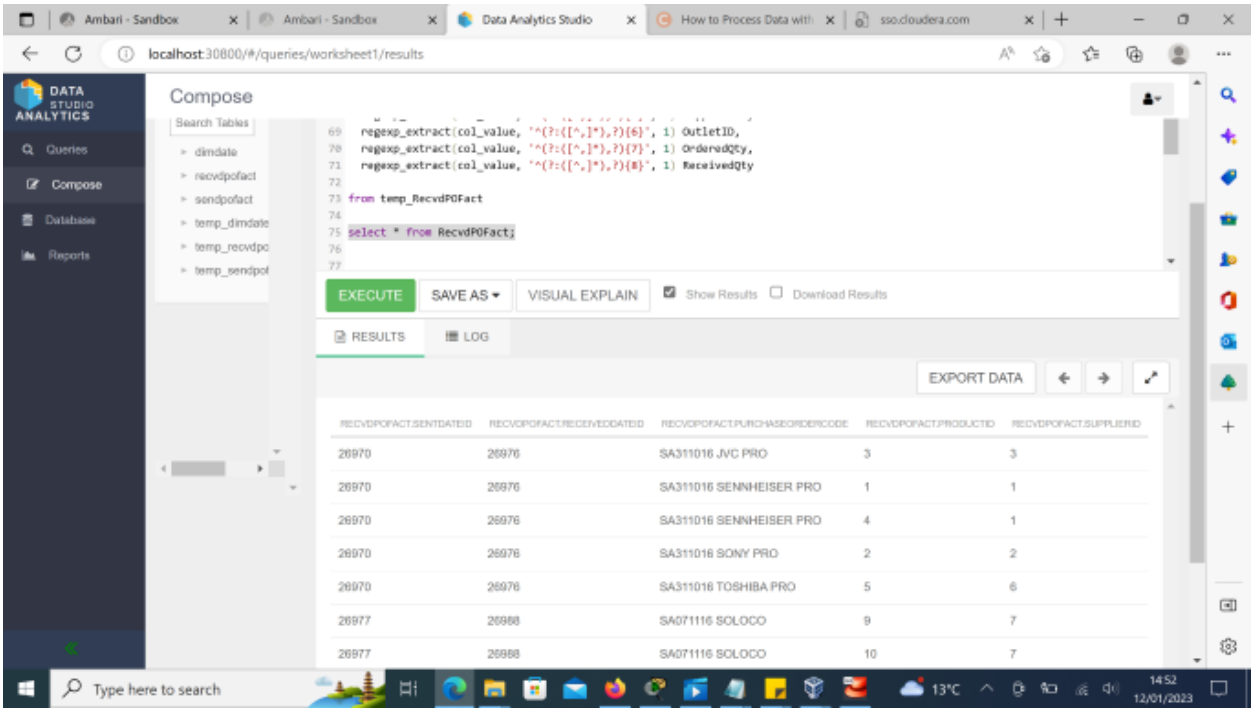
8. Migrating Data file to Apache HDFS

This was done by using the Files View and browsing to a preferred location or folder to upload the data files to. In this case, the path '/user/maria_dev' was navigated to and write permission access was granted to the maria_dev user.



9. Creating a data structure in hive





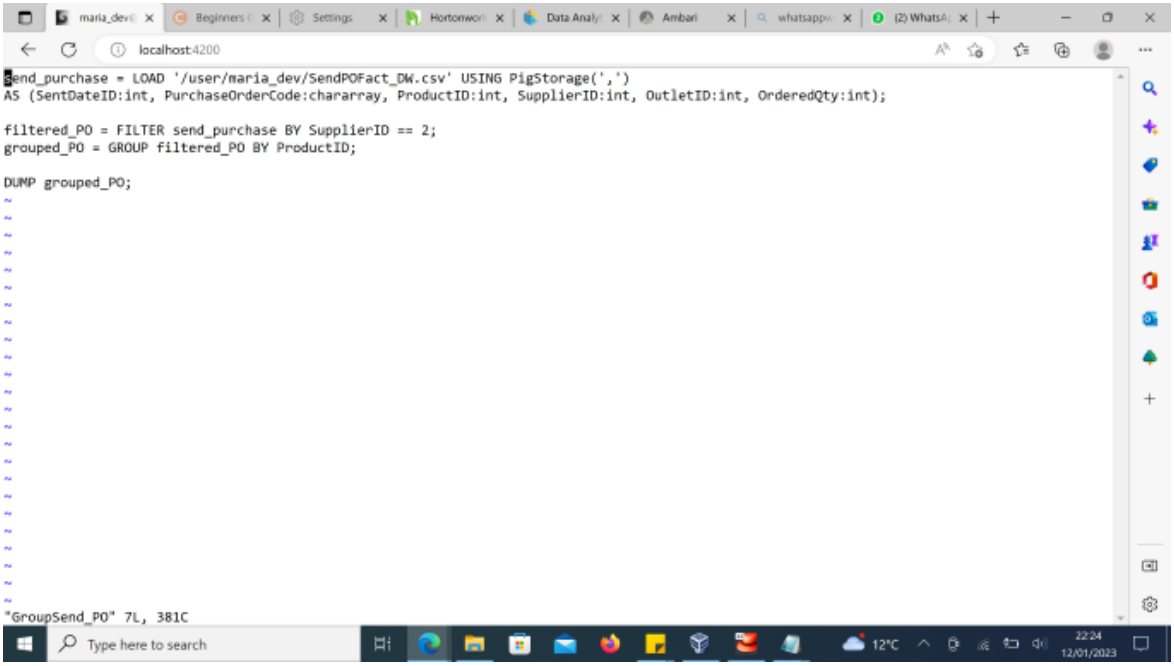
10. Demonstrating the use of Pig

The figure below shows how to load data and create data structure in pig and also using the DUMP command to view data.

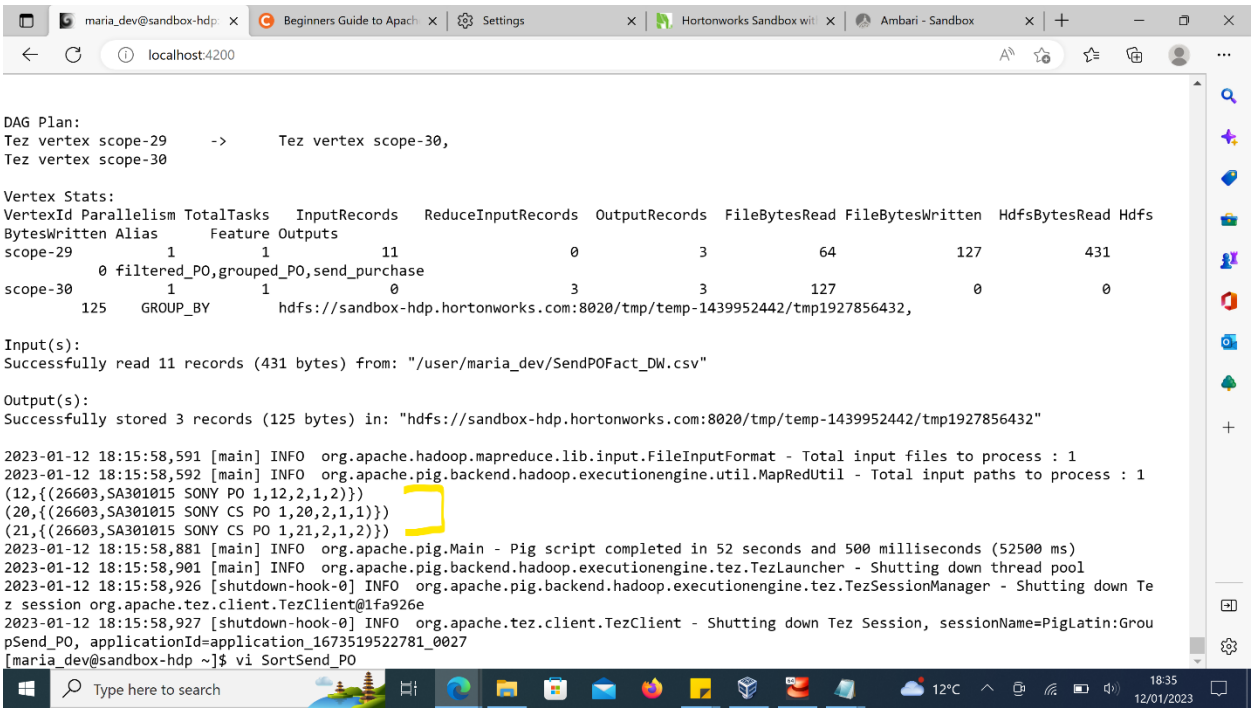
10.1. DUMP

10.2. FILTER and GROUP.

The below script shows how the loaded data was filtered by SupperID and grouped by the ProductID.

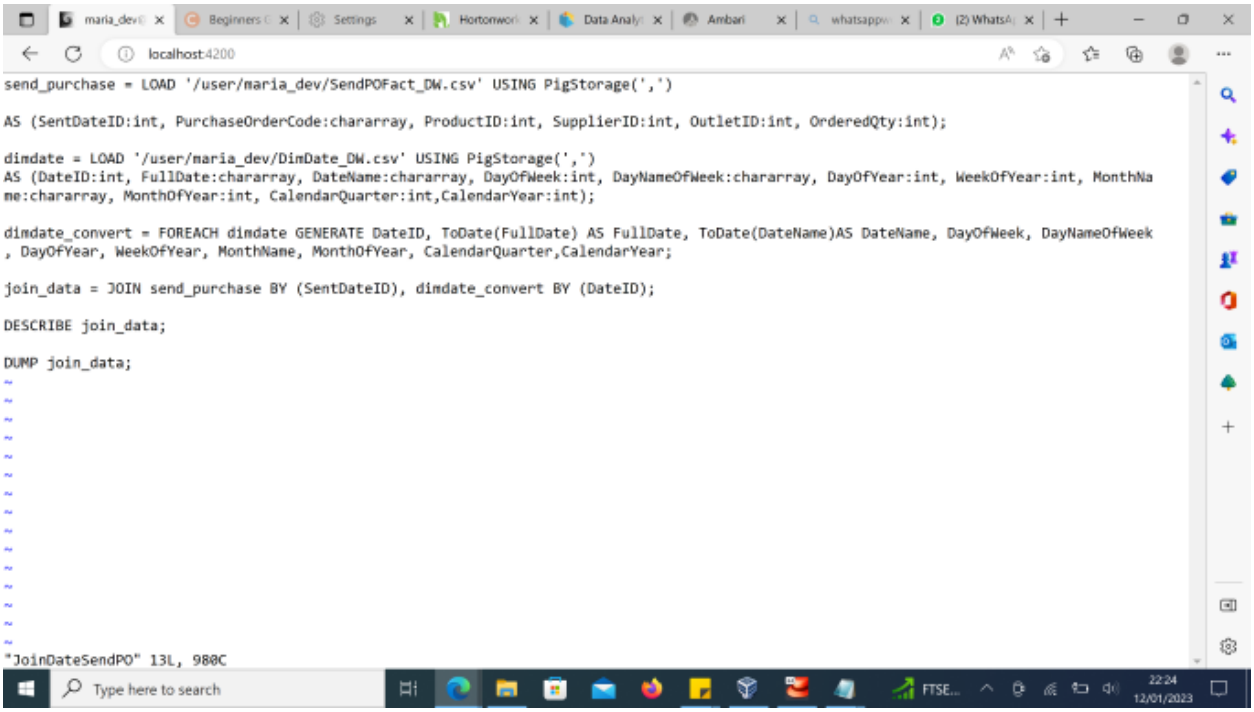


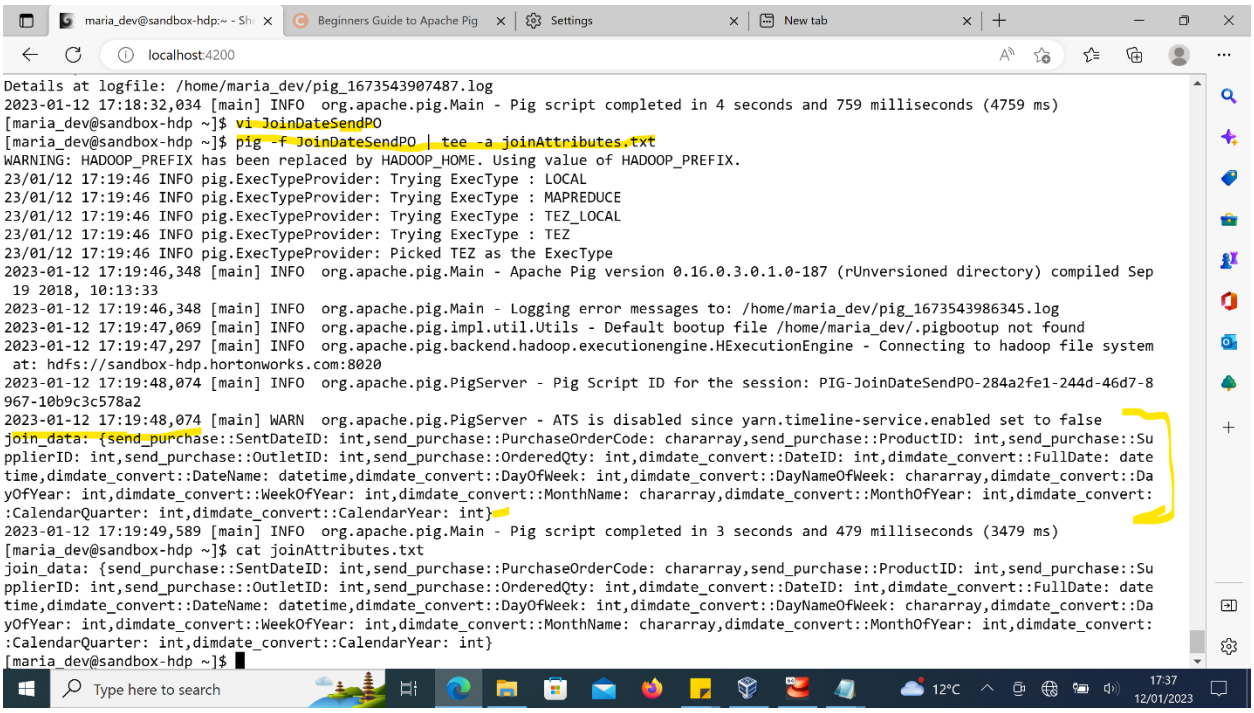
The result of the above script is as follows showing the result of filtering only the supplier whose ID is 2 and grouped all the products.



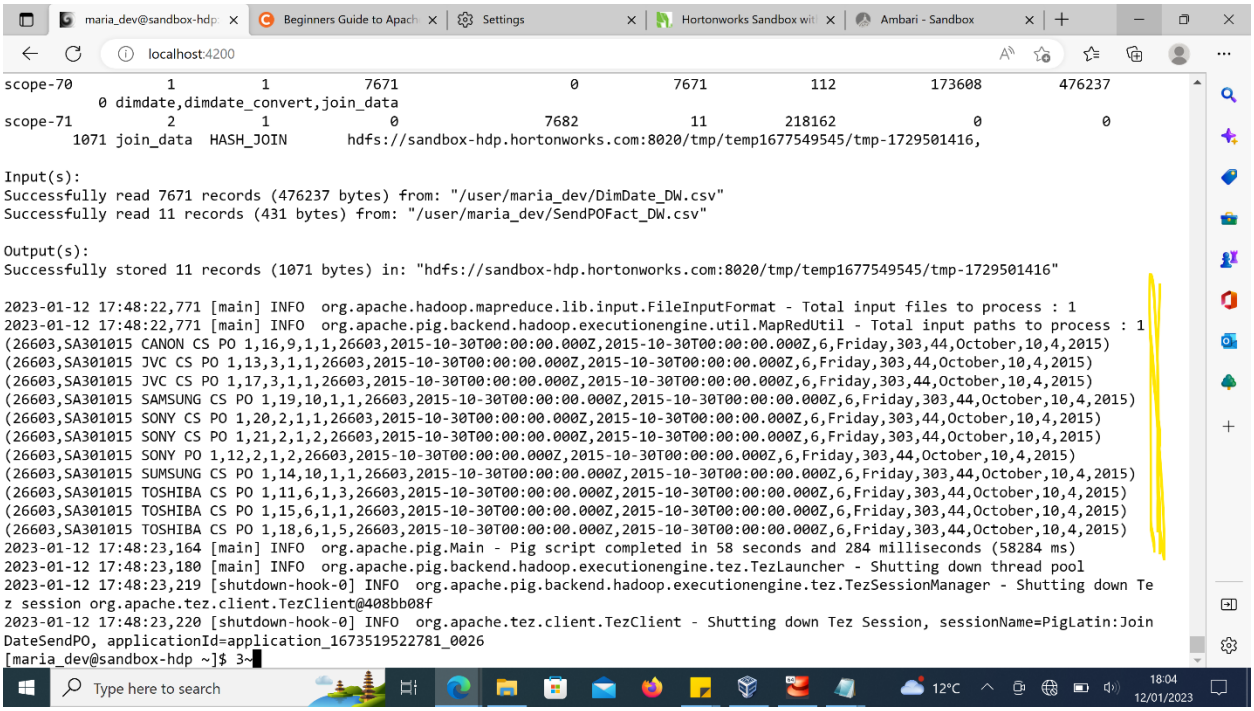
10.3. Join Data

The script below shows how to join two tables together and display the result



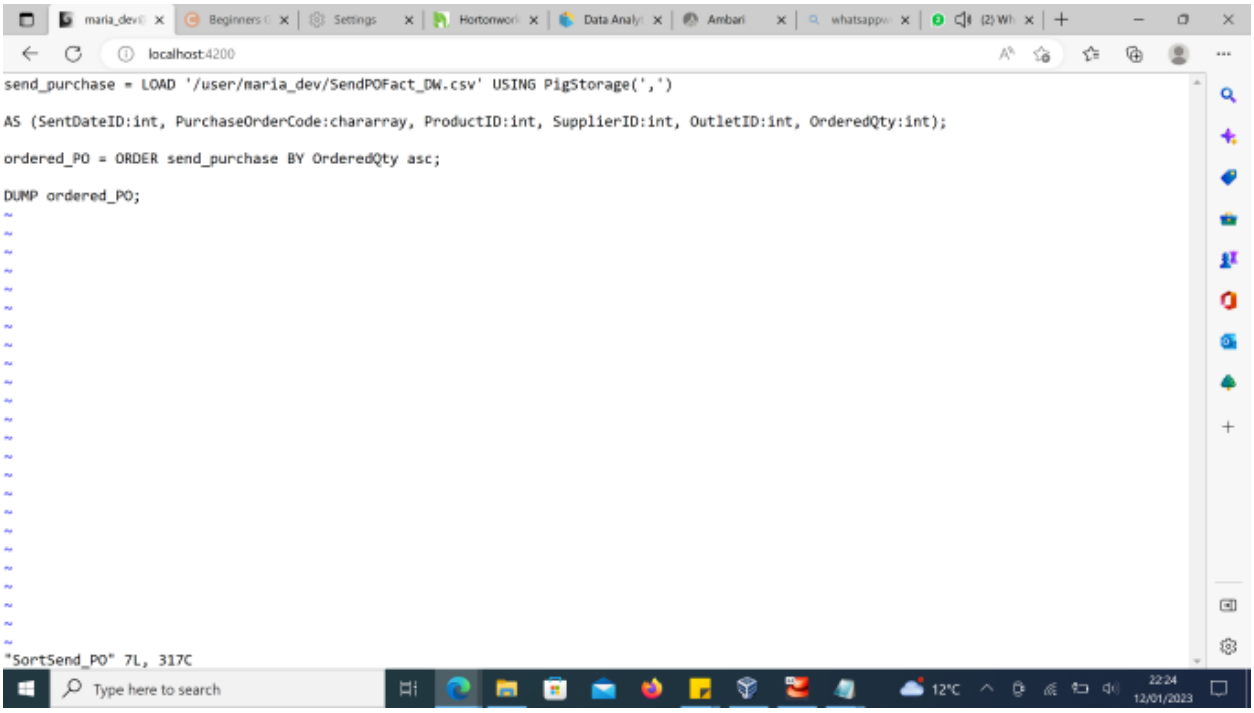


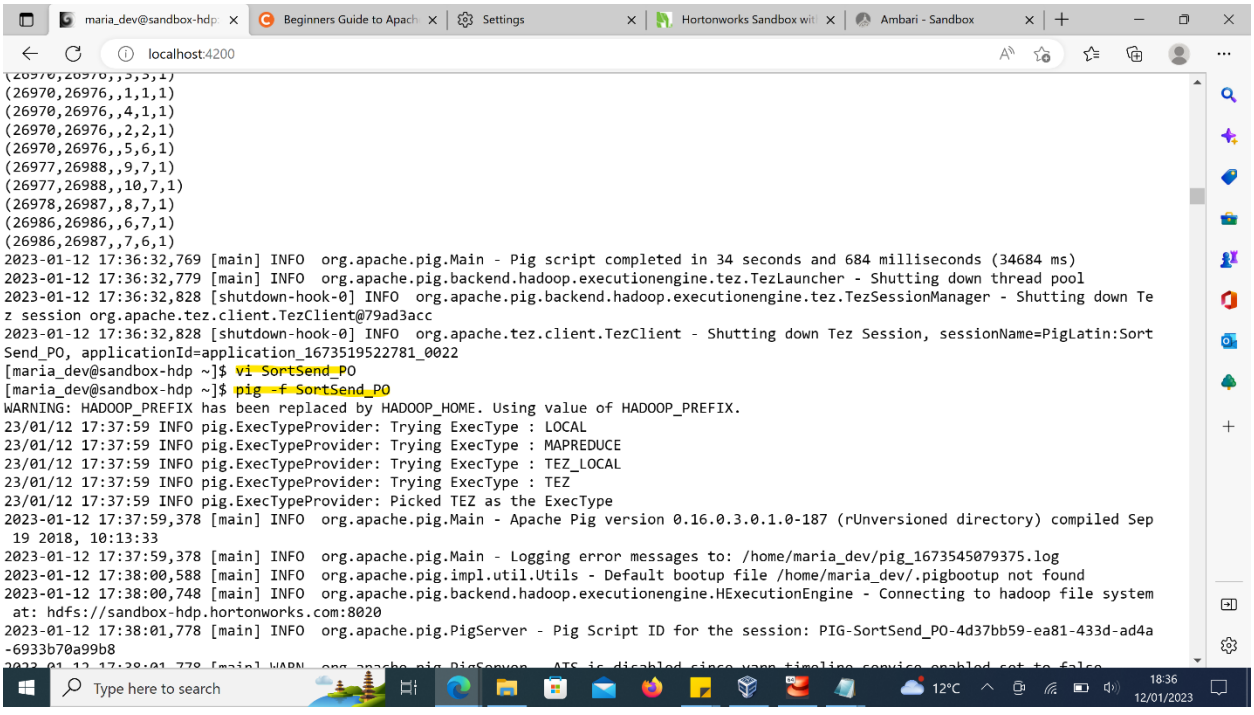
The result of the join tables – Dimdate and SendPurchaseOrder is as follows:



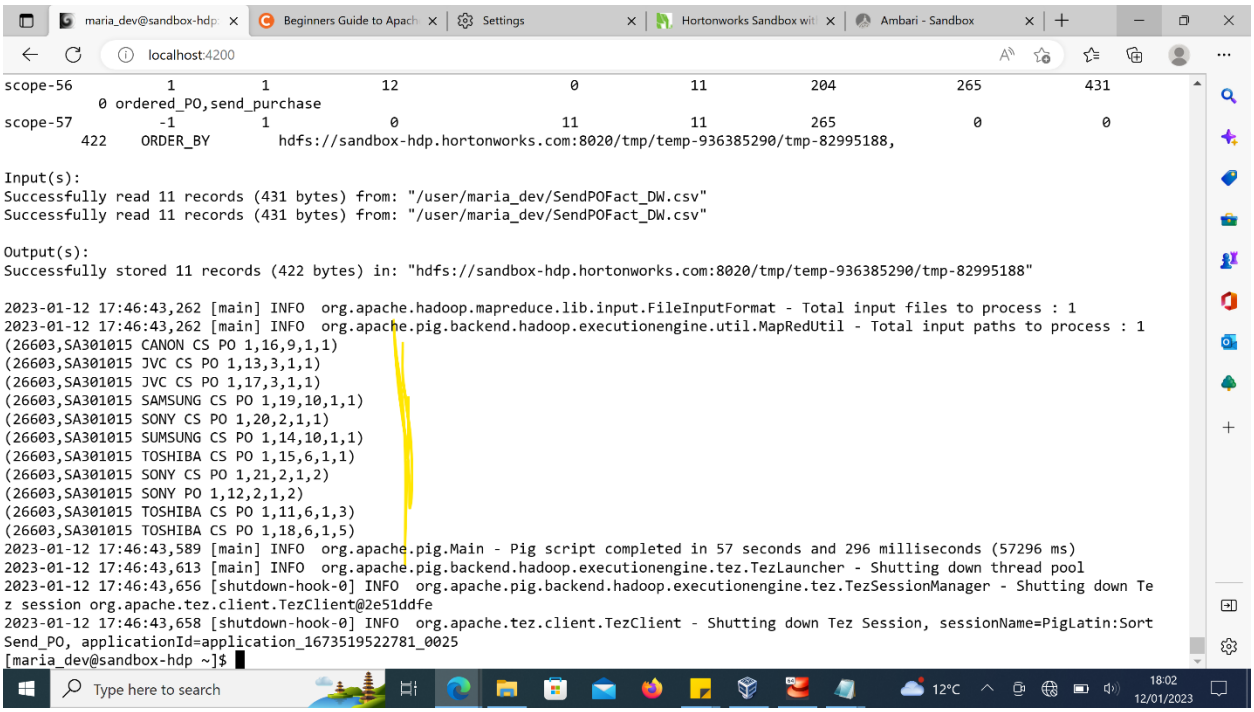
10.4. SORT

Loaded data can be ordered or sorted by using the ORDER command. The script below shows how to order a table.





The result below shows an ordered table by OrderedQty which is the last column inn ascending order.



11. REFLECTIVE CONCLUSION

Overall, I have learnt the following:

- How to effectively define and translate business and reporting requirements.
- How to identify and define the grain level of a data warehouse project.
- How to define dimensions and their attribute.
- How to define measures of a fact table.
- How to design a data warehouse schema to conform to the identified business requirements.
- How to effectively use SSMS to create database and tables.
- How to effectively populate dimension tables with test data using SSMS.
- How to effectively create lookup tables and populate fact tables using SSMS.
- How to use Apache HDFS to upload external file.
- How to use the Data Analytics Studio to create data structure and load data in Hive
- How to use Pig to load and manipulate data.