CI7320 - Databases and Data Management

Data Warehouse Assignment

Submitted by Nishit Gala

Contents

[1) Difference between OLTP and Data Warehouse 3](#_Toc72308014)

[2) Benefits of Having a Data Warehouse for the Wine Stopper Company 4](#_Toc72308015)

[3) Problems that Wine Stopper company may have by creating a Data Warehouse 6](#_Toc72308016)

[4) Warehouse using Star Schema 7](#_Toc72308017)

[5) CREATE TABLE statement for above STAR Schema 9](#_Toc72308018)

[6) Issues faced while loading the data from OLTP DB to the Staging Area 11](#_Toc72308019)

[7) Strategy to be used during ETL stage 12](#_Toc72308020)

[8) OLAP Cubes and its benefits 13](#_Toc72308021)

[9) Cube Design 15](#_Toc72308022)

[10) Conclusion 18](#_Toc72308023)

# Difference between OLTP and Data Warehouse

With the modern-day business, the Data Exploration and Data analytics leads a business to success.

The Wine Stopper company has retail warehouses across the UK and their business has a need to perform the analytics, forecasting, reporting based on data at their disposal. The management of the Wine Stopper company might be interested in quarterly sales, effectiveness of recent promotions, revenue generated from different regions, etc.

All the above can be done on an OLTP DB but by impacting its day-to-day operations and can be challenging for its business.

We cannot go with the existing database as pulling the analytical reports from the database that is in 3rd normal form is challenging due to complex queries and quite time consuming. This may also adversely impact the performance of the OLTP database resulting in the slowness and in some case complete shut down on the day-to-day operations of the company. This is a risky situation to the business and hence the company will be required to invest in building a data warehouse as a mean to fetch analytical insights from the data taken from the existing OLTP system.

The answer to above challenges in building a Data Warehouse that helps non techies like managers, business domain experts to utilize and explore the analytics side of the data.

The management teams of this company will require checking on analytical reports, gain business insights which can be beneficial for the company.

The below table summarizes the difference between OLTP and a Data Warehouse

|  |  |
| --- | --- |
| **OLTP DB** | **Data Warehouse** |
| Handles day to day operations of the company department like Insert, Update, Delete, select queries in real time to fulfill | A database that integrates data from multiple data sources (or source systems) using a star or snowflake schema concept to fulfill the analytical side of the business for gaining vital insights from the data. |
| Often multiple department (applications) within a same company has a DBMS system of their own. This is due to change in users of each application, structure of the DBMS | Data Warehouse usually collects the data from different source systems and bring it into a single umbrella using ETL technique |
| Need to maintain an uptime of ~ 100 % in production as it can result into chaos and complete halt of a business function | SLA is usually flexible as the downtime is expected for data load and transformation activities. The data warehouse is highly scalable. |
| Data is updated / inserted in the real time | Data is refreshed in the periodic intervals as per the business agreement |
| OLTP database should be able to deliver high performance and the query run time should ideally be in milli seconds. So this does not allow user to perform complex analytical queries as it might lock some critical tables. | A data warehouse is designed for supporting analytics and can handle high performance queries as it has large processors and disk bandwidth |
| The data is highly normalized and typically follow the 3rd Normal form | The data is in the form of star schema and in the denormalized form to enhance the response time for running complex analytical queries |

# Benefits of Having a Data Warehouse for the Wine Stopper Company

Today’s business usually thrives on competitive edge and staying ahead in the market by taking critical business decisions using the data analytical tools / technology available at their disposal. A data warehouse is a medium to gain significant business insights out of the data. The benefits of having a data warehouse is as below

1. **Assists in delivering the much-needed business intelligence**

Below are some examples that Wine Stopper company can benefit from by gaining enhanced business intelligence

1. Sales performance of each warehouse in terms of Profit, Total Sales across a month, quarter, year
2. Distribution of stocks across warehouse.
3. Profit earned from the warehouses across period range (Quarterly, Yearly, 3 Years, 5 years).
4. The promotion effectiveness with the customer base (which geographical region customers are utilizing the promotions well).
5. Order information placed by warehouses across month, quarter, year.
6. Special Order information placed by warehouses across month, quarter, year.
7. Checking on Under sold stocks with respect to warehouses in a quarterly, yearly report.
8. Tracking Sales record of High-Priced wines according to the warehouse for a given timeframe.
9. Tracking Sales record of Low-Priced wines according to the warehouse for a given timeframe.
10. Average monthly, yearly, quarterly transaction amount per warehouse on a yearly basis.
11. Average Sales forecast on a holiday in past year versus the average sales forecast on a normal day
12. Average Sales forecast over Weekend in the past year compared to the weekday
13. **Assists in improving time efficiency**

The data warehouse assists all the critical business functions in the organisation take informed business decisions based on the analytic inputs provided in a time efficient manner as user do not rely on support IT functions to help them with running the query and passing the results.

A UI layer on the top of the Data Warehouse will help the critical business users of the wine stopper company to interact and gain insights on key analytical parameters of business without having to write the complex business queries.

1. **Standardized Data at the Organization Level**

A data warehouse is the source of truth for the data in the organisation and since the data can be coming from multiple sources, a warehouse plays a key role in consolidation of data in a standardized format which assists the business user in making informed decision that are critical to the business.

There are high chances within the Wine Stopper company that different departments might have their respective DBMS systems and their own nomenclature. This might cause fields like name, dates, pricing etc. to be in different format from department to department. Hence the standard practice should be taken into consideration while creating a Data Warehouse so that the data is accurately reflected in the same format in the Dimension table of the Data Warehouse

1. **Enables organizations to forecast with confidence**

Data Warehouse assists in making the forecasting by looking at the historical data and can help in predicting key sales parameters for an organisation. This makes organisation take better business decision by planning in advance for some critical business situations.

Wine stopper company can broadly classify the KPI’s to be achieved for each warehouse, managers, employees and set their sales target to boost revenue and profits.

# Problems that Wine Stopper company may have by creating a Data Warehouse

Below are some problems that a Wine Stopper Warehouse may pose in the future

1. **Indexing**

Efficient indexing can easily go wrong can cause performance issues. Whilst an index can improve performance issues of select query, but the same indexing can cause overheads during insert operations. If there are too many tables with indexes in the data warehouse, it can degrade the load operations performance. So there has to be a trade off between select query performance vs data load time.

1. **User Expectation**

As the data is loaded into the data warehouse, the expectation from a business user to gain insights increases rapidly after using it sufficiently for making key business decision.

Sometimes, this may lead to high expectations that may result in complex SQL queries and may decrease the performance as the overall data in the warehouse is going to increase once the OLTP systems are stable and achieve business ready status. So it is important to have the users know about the query processing time by specifying the time range in the user documentation.

1. **Data Structuring**

A good amount of time should be spent in data warehouse model conception and should be information driven, structured driven. The data that enters warehouse should be highly standardized and all the metadata information needs to be properly documented including the common formats for dates, currency, address, phone numbers etc. It is highly necessary that all the stakeholders are informed about the structuring and standardization of data in the warehouse

1. **Access Management for Data Warehouse**

The access of the warehouse data should be only available to business users and should have a proper access control mechanism on users requesting the access and corresponding approvals. This is key as having multiple users accessing the warehouse simultaneously can degrade the performance and improve the efficiency of the system. There should be strict policy to review the access periodically and take actions in case user has permission without proper justification.

1. **Data Governance and Master Data**

Data going into a warehouse should be as per the master data configuration maintained across the organisation and to ensure that this happens on every time the data is loaded, it is highly necessary to ensure that there is an investment made by the business into the data governance and maintaining / updating the master data as and when required. This process requires investment for data governance and monitoring and business should not shy away from taking this responsibility.

The implementation of data governance also allows to define clear roles and responsibilities, ownership of data in the warehouse.

# Warehouse using Star Schema

The data warehouse grain to design the Fact table in the Star schema is as below

*“Wine purchase made by a customer in an invoice transaction taking into account the promotion offers and discount”*

We have considered a scenario of Wine Sale from different distribution center for building our fact table.

The Fact table has been linked to various dimensions table for WINE, CUSTOMER, INVOICE, DISTRIBUTION CENTER, DATE as per the setup in the Wine Stopper company.

The WINE\_SALE\_FACT table records all the transactions that have been processed in the system along with the foreign keys of dimension tables like Wine, Promotion, Invoice, Customer, Center and Date.

The WINE DIMENSION table keeps record of all wines that are listed for sale in Wine Stopper company along with their Wine Code, Wine Name, Color, Country, Type and prices.

The PROMOTION DIMENSION table keeps a record of all the promotions that the Wine Stopper company has offered in past as well as present (ongoing).

The INVOICE DIMENSION table keeps a record of all the invoices that are billed and paid to the customer keeping the account of total bill amount, savings (or discount amount), and delivery charges. It also tracks the invoice date and Delivery postcode.

The DATE DIMENSION table keeps a record of all the dates that wine stopper has been operating since the start and it stores information in form of day, month, year, day of Week, Quarter and Is Wine Stopper Holiday. This gives us a quick way to check and link all the dates of other tables with this table.

The CUSTOMER DIMENSION table keeps record of all customers.

The CENTER DIMENSION table keeps record of all distribution center along with the corresponding warehouse and the region where they belong.

The Star schema diagram is as shown below

Diagram

Description automatically generated

# CREATE TABLE statement for above STAR Schema

**WINE\_DIM:**

CREATE TABLE WINE\_DIM (

    WINE\_ID\_SK NUMBER NOT NULL,

    WINE\_CODE VARCHAR2(15) NOT NULL,

    WINE\_NAME VARCHAR2(50) NOT NULL,

    WINE\_CAT VARCHAR2(50) NOT NULL,

WINE\_COLOR VARCHAR2(20) NOT NULL,

    COUNTRY VARCHAR2(10) NOT NULL,

    CASE\_SIZE\_IN\_LTR NUMBER(4,2) ,

    BOTTLE\_SIZE\_IN\_ML NUMBER(4),

    NUM\_OF\_BOTTLE NUMBER(4) GENERATED ALWAYS AS (round(CASE\_SIZE\_IN\_LTR \* 1000 /BOTTLE\_SIZE\_IN\_ML)) VIRTUAL,

    PRICE\_PER\_CASE NUMBER(7,2) CHECK(PRICE\_PER\_CASE > 0 )NOT NULL,

    PRICE\_PER\_BOTTLE NUMBER(7,2) CHECK(PRICE\_PER\_BOTTLE > 0 ) NOT NULL,

    CONSTRAINT CHK\_PRICE\_DIM CHECK (PRICE\_PER\_BOTTLE < PRICE\_PER\_CASE),

    PRIMARY KEY(WINE\_ID\_SK)

)

**DATE\_DIM:**

CREATE TABLE DATE\_DIM (

    DATE\_ID NUMBER  NOT NULL,

    WS\_DAY NUMBER  CHECK(REGEXP\_LIKE (WS\_DAY,'0?[1-9]|[12][0-9]|3[01]')) NOT NULL,

    WS\_MONTH NUMBER   CHECK (REGEXP\_LIKE (WS\_MONTH,'0?[1-9]|1[012]')) NOT NULL,

    WS\_YEAR NUMBER  CHECK (REGEXP\_LIKE (WS\_YEAR,'19|20)[0-9][0-9]'))   NOT NULL,

    WS\_DOW NUMBER CHECK(REGEXP\_LIKE (WS\_DOW,'0?[1-7]'))      NOT NULL,

    WS\_QUARTER NUMBER  CHECK(REGEXP\_LIKE (WS\_QUARTER,'0?[1-4]'))    NOT NULL,

    WS\_IS\_HOLIDAY NUMBER CHECK(REGEXP\_LIKE (WS\_IS\_HOLIDAY,'[0-1]'))    NOT NULL,

    PRIMARY KEY(DATE\_ID)

)

**INVOICE\_DIM:**

CREATE TABLE INVOICE\_DIM (

    INVOICE\_NO\_SK NUMBER  NOT NULL,

    INVOICE\_NO VARCHAR2(25) NOT NULL,

    INVOICE\_DATE REFERENCES DATE\_DIM(DATE\_ID) NOT NULL,

    TOTAL\_AMOUNT NUMBER(7,2) NOT NULL,

    SAVINGS NUMBER(7,2),

    DELIVERY\_CHARGE NUMBER(5,2),

    DEL\_POSTCODE VARCHAR2(10),

    PRIMARY KEY(INVOICE\_NO\_SK)

)

**PROMOTION\_DIM:**

CREATE TABLE PROMOTION\_DIM (

    PROMOTION\_ID\_SK NUMBER NOT NULL,

    PTYPE\_METHOD VARCHAR2(25) CHECK (PTYPE\_METHOD IN ('PRICE\_DISCOUNT','QUANTITY\_DISCOUNT','REDUCED\_PRICE')) NOT NULL,

    PRIMARY KEY(PROMOTION\_ID\_SK)

)

**CENTER\_DIM:**

CREATE TABLE CENTER\_DIM (

    CENTER\_ID NUMBER  NOT NULL,

    CENTER\_NAME VARCHAR2(40) NOT NULL,

    LOCATION\_NAME VARCHAR2(40) NOT NULL,

    WAREHOUSE\_NAME VARCHAR2(40) NOT NULL,

    REGION\_NAME VARCHAR2(40) NOT NULL,

    PRIMARY KEY(CENTER\_ID)

)

**CUST\_DIM:**

CREATE TABLE CUST\_DIM (

    CUST\_ID\_SK NUMBER  NOT NULL,

    CUST\_ID NUMBER  NOT NULL,

    LNAME VARCHAR2(40) NOT NULL,

    FNAME VARCHAR2(40) NOT NULL,

    POSTCODE VARCHAR2(10) NOT NULL,

    PRIMARY KEY(CUST\_ID\_SK)

)

**WINE\_SALE\_FACT:**

CREATE TABLE WINE\_SALE\_FACT (

    SALE\_FACT\_ID NUMBER NOT NULL,

    CUST\_DIM\_ID REFERENCES CUST\_DIM(CUST\_ID\_SK) NOT NULL,

    WINE\_DIM\_ID REFERENCES WINE\_DIM(WINE\_ID\_SK) NOT NULL,

    CENTER\_ID REFERENCES CENTER\_DIM(CENTER\_ID) NOT NULL,

    INVOICE\_DIM\_ID REFERENCES INVOICE\_DIM(INVOICE\_NO\_SK) NOT NULL,

    DELIVERY\_DATE\_ID REFERENCES DATE\_DIM(DATE\_ID) NOT NULL,

    PROMOTION\_DIM\_ID REFERENCES PROMOTION\_DIM(PROMOTION\_ID\_SK) ,

    TYPE\_SOLD VARCHAR2(20) CHECK(TYPE\_SOLD in ('CASE','BOTTLE')) NOT NULL,

    QUANTITY\_SOLD NUMBER(4) NOT NULL,

    SELLING\_PRICE NUMBER(7,2)  NOT NULL,

    DISCOUNT\_AMOUNT NUMBER(5,2) NOT NULL,

    PRIMARY KEY(SALE\_FACT\_ID)

)

# Issues faced while loading the data from OLTP DB to the Staging Area

We might encounter below issues while loading the data to the staging area from the OLTP DB

1. Conflicting business rules used by various data sources.
2. Data validation issues like accepting non-numeric values in numeric fields
3. Too long data is truncated or trimmed due to column size mismatches
4. The address information might have invalid cities or zip codes due to poor data validations
5. The absence of an effective and centralized source metadata repository.
6. Errors in the transformation or substitution values for NULL values
7. Effectiveness of testing the data in the Staging area
8. Loss of data due to misinterpretation of importance of columns(fields)
9. No systematic way of reporting, metadata updates
10. Issues with data cleansing, reconciliation in the staging area
11. Misaligned primary and foreign key strategies for the same type of entity

# Strategy to be used during ETL stage

ETL performs the heavy lifting for your data integration workflows, efficiently collecting and centralizing vast quantities of information from a wide variety of sources. However, when things go wrong, ETL can be the source of serious pain points and bottlenecks that prevent you from gaining valuable data-driven insights. Below are some strategies to ensure smooth functioning during ETL phase

1. Need to have proper documentation of whole ETL process
2. This visual representation will demonstrate cardinal relationships between tables and how multiple sources are related to each other.
3. Extract only the necessary data - While it's good to have lots of information at hand for querying and analysis, too much data flowing through your ETL pipelines can slow things down considerably.
4. Avoid the use of SELECT \* and SELECT DISTINCT in SQL queries during the extraction phase
5. Reduce the usage of JOIN queries to as least as possible
6. Schedule ETL jobs to run overnight or outside peak (business) hours to avoid potential conflicts with other sessions and processes
7. Usage of ETL Logging and Monitoring like timing of data extraction, keep record of number of rows inserted, modified, or deleted.
8. Tracking the run time validation errors

# OLAP Cubes and its benefits

OLAP cubes are highly beneficial to analyze and view data in multiple dimension in an efficient and quick way compared to a traditional relational database. A relational database provides data in a tabular format comprising of rows and columns.

The OLAP cubes provides an extension of single table (fact) by additional information through layers (dimensions).

Benefits of OLAP Cubes:

1. **Improves performance**

Traditional RDBMS or OLTP database are used for supporting the day-to-day operations of a business and cannot facilitate running of high performance complex analytical business queries. Running these queries on OLTP might cause issue of locking and hence impact its performance, so OLAP cubes are used that provides an analytical edge by gaining data access through OLAP databases in the data Warehouse

The Wine stopper company has an OLTP DB to take care of the day-to-day operations and hence creating a warehouse and OLAP cubes will help the analytics side of the business.

1. **Drill Down Approach**

This is one of the most visible features of the OLAP Cube as a non-technical person can relate this to concept in the spreadsheet. The drill down mechanism provides user to have a more detailed view of the data by going down the hierarchy and adding another dimesion

In the Wine stopper company, the example of Drill down is usually getting the Region Level sales, then drilling down to distribution center level sales and followed by warehouse level sales.

1. **Roll Up Approach**

It’s exactly the opposite of the drill down as it aggregates data by moving up in the hierarchy thereby reducing the number of dimensions.

In case of Wine Stopper company, this can be Warehouse Requesting special Orders, then rolling up to Distribution Center processing the Special orders request and rolling up to check the Special-Order requirement as per region

1. **Slicing as per the dimension**

It helps in creating a sub-cube by slicing a single dimension from the whole cube and focusing on a particular piece of information within the data cube

In case of a Wine Stopper company, the slicing would refer to quarterly sales across all the distribution centers.

1. **Dicing**

The dice operation is to take out a single sub-cube by, this helps to extract all dimensions of the subset region (dice) in the entire cube.

For example, dice can be extracted by taking Wine Stopper’s quarterly sales (time dimension) for London Area (Location Dimension)

1. **Pivot**

The pivot is used for data visualization and helps in viewing the cube from different angles. It can be compared to the feature available in excel spreadsheet. However, OLAP is easier to use and requires less expertise. It also provides and improved performance and faster processing time

# Cube Design

1. Annual Sales , Quarterly Sales of a Distribution Center in the Wine Stopper company

The aim is to make a cube that will show the historical sales Center wise, year wise, drill down to quarter wise. So here we are showing the Total Sales amount of the sales made by Wine Stopper company through their centers on a yearly / Quarterly basis.

SELECT  COALESCE(CENTER.CENTER\_NAME,'ALL CENTERS') AS "CENTER\_NAME",

COALESCE(CAST(DT.WS\_YEAR AS VARCHAR2(25)),'ALL YEARS') AS "YEAR",

COALESCE(CAST(DT.WS\_QUARTER AS VARCHAR2(25)),'ALL QUARTERS') AS "QUARTER",

SUM (FCT.SELLING\_PRICE) AS "TOTAL SALES"

FROM WINE\_SALE\_FACT FCT, CENTER\_DIM CENTER, DATE\_DIM DT

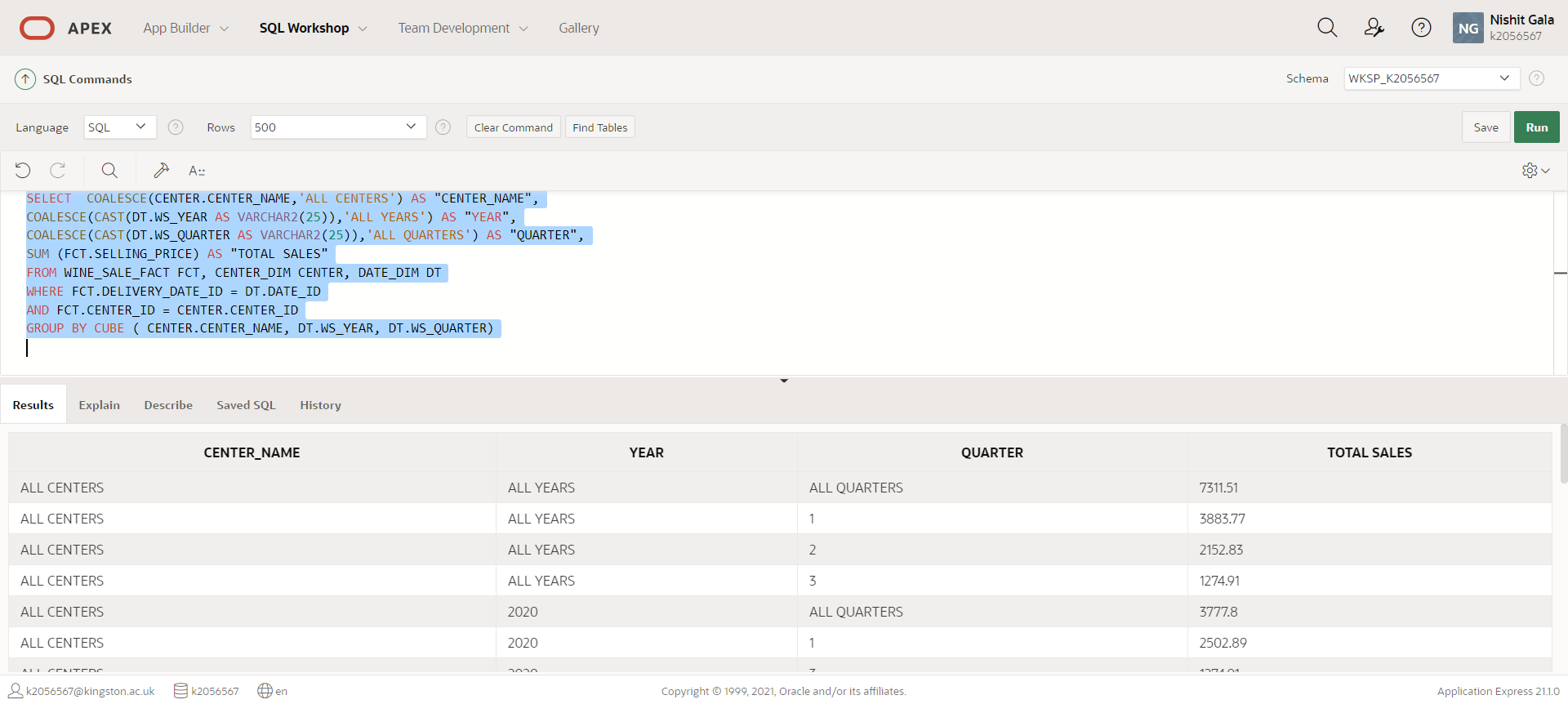
WHERE FCT.DELIVERY\_DATE\_ID = DT.DATE\_ID

AND FCT.CENTER\_ID = CENTER.CENTER\_ID

GROUP BY CUBE ( CENTER.CENTER\_NAME, DT.WS\_YEAR, DT.WS\_QUARTER)

The Output of above query is as attached in the CSV file





1. Sales of Wines as per the Type of Wine in the year / quarter as per the Center
2. Wine Sales in the year / quarter as per the Center and Wine Category (Sparkling vs Regular)

The aim is to make a cube that will show the historical sales Wine type wise (Regular vs Sparkling) , year wise, drill down to quarter wise. So here we are showing the Total Sales amount of the sales made by Wine Stopper company in terms the Wine Type on a yearly / Quarterly basis.

SELECT  COALESCE(CENTER.CENTER\_NAME,'ALL CENTERS') AS " CENTER\_NAME ",

COALESCE(WINE.WINE\_CAT,'ALL CATEGORIES') AS " WINE\_TYPE ",

COALESCE(CAST(DT.WS\_YEAR AS VARCHAR2(25)),'ALL YEARS') AS "YEAR",

COALESCE(CAST(DT.WS\_QUARTER AS VARCHAR2(25)),'ALL QUARTERS') AS "QUARTER",

SUM (FCT.SELLING\_PRICE) AS "TOTAL SALES",

SUM(FCT.QUANTITY\_SOLD) AS "TOTAL COUNT"

FROM WINE\_SALE\_FACT FCT, WINE\_DIM WINE, DATE\_DIM DT, CENTER\_DIM CENTER

WHERE FCT.DELIVERY\_DATE\_ID = DT.DATE\_ID

AND FCT.WINE\_DIM\_ID = WINE.WINE\_ID\_SK

AND FCT.CENTER\_ID = CENTER.CENTER\_ID

GROUP BY CUBE ( CENTER.CENTER\_NAME,WINE.WINE\_CAT, DT.WS\_YEAR, DT.WS\_QUARTER);

The Output of above query is as attached in the CSV file



1. Wine Sales in the year / quarter as per the Center and Wine Colour (Red / White / Rose)

The same logic can be extended to checking the sales of Wine Stopper company on the basis of Wine Colour (Red / White / Rose) for the historical records (Yearly / Quarterly)

SELECT  COALESCE(CENTER.CENTER\_NAME,'ALL CENTERS') AS " CENTER\_NAME ",

COALESCE(WINE.WINE\_COLOR,'ALL COLOURS') AS " WINE\_COLOR",

COALESCE(CAST(DT.WS\_YEAR AS VARCHAR2(25)),'ALL YEARS') AS "YEAR",

COALESCE(CAST(DT.WS\_QUARTER AS VARCHAR2(25)),'ALL QUARTERS') AS "QUARTER",

SUM (FCT.SELLING\_PRICE) AS "TOTAL SALES",

SUM(FCT.QUANTITY\_SOLD) AS "TOTAL COUNT"

FROM WINE\_SALE\_FACT FCT, WINE\_DIM WINE, DATE\_DIM DT, CENTER\_DIM CENTER

WHERE FCT.DELIVERY\_DATE\_ID = DT.DATE\_ID

AND FCT.WINE\_DIM\_ID = WINE.WINE\_ID\_SK

AND FCT.CENTER\_ID = CENTER.CENTER\_ID

GROUP BY CUBE ( CENTER.CENTER\_NAME,WINE.WINE\_COLOR, DT.WS\_YEAR, DT.WS\_QUARTER);

The Output of above query is as attached in the CSV file



1. Sales of Wines as per the Promotion offered by the Wine Stopper company in the year / quarter

The aim is to make a cube that will show the historical sales year wise, drill down to quarter wise for the promotions offered along with the total Sales Amount and total Savings Amount.

SELECT  COALESCE(CENTER.CENTER\_NAME,'ALL CENTERS') AS " CENTER\_NAME ",

COALESCE(PROMO.PTYPE\_METHOD,'ALL PROMOTION TYPES') AS " PRMOTION\_TYPE ",

COALESCE(CAST(DT.WS\_YEAR AS VARCHAR2(25)),'ALL YEARS') AS "YEAR",

COALESCE(CAST(DT.WS\_QUARTER AS VARCHAR2(25)),'ALL QUARTERS') AS "QUARTER",

SUM (FCT.SELLING\_PRICE) AS "TOTAL SALES",

SUM(FCT.DISCOUNT\_AMOUNT) AS "TOTAL\_SAVINGS",

SUM(FCT.QUANTITY\_SOLD) AS "TOTAL COUNT"

FROM WINE\_SALE\_FACT FCT, PROMOTION\_DIM PROMO, DATE\_DIM DT, CENTER\_DIM CENTER

WHERE FCT.DELIVERY\_DATE\_ID = DT.DATE\_ID

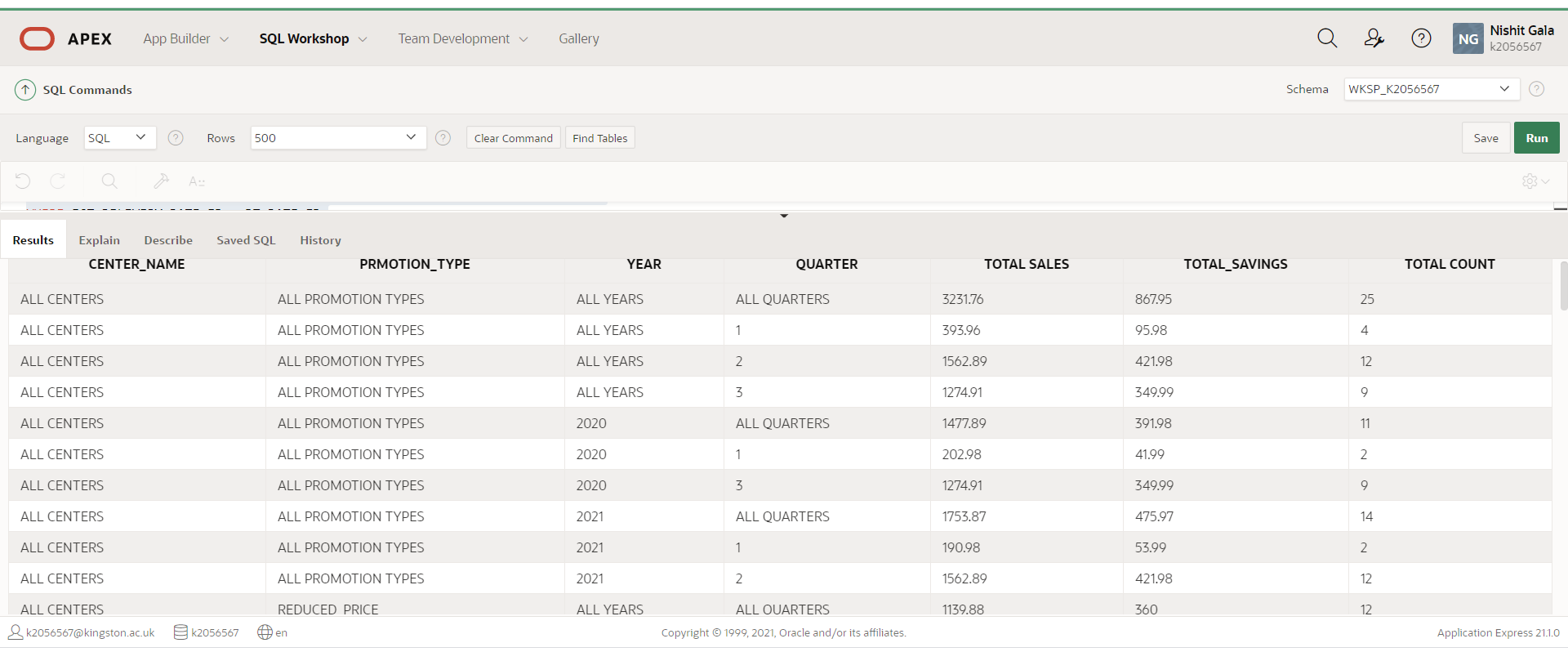
AND FCT.PROMOTION\_DIM\_ID = PROMO.PROMOTION\_ID\_SK

AND FCT.CENTER\_ID = CENTER.CENTER\_ID

GROUP BY CUBE ( CENTER.CENTER\_NAME,PROMO.PTYPE\_METHOD, DT.WS\_YEAR, DT.WS\_QUARTER);

The Output of above query is as attached in the CSV file





# 10) Conclusion

A data warehouse is a centralized repository that maintains an integrated stream of data from multiple sources or applications. There is a periodical load of historical data into the warehouse and this data is then used for analytics and reporting to take key business decisions. A data must go through an ETL (Extract, Transform and Load) mechanism to have the data cleansed, validated, standardized, structured and well formatted before it enters a warehouse. All these operations are performed in a stagging area, so that it does not impact the performance of OLTP database that is critical to day-to-day operations.

Data warehouse usually have consolidation of data from multiple tables that are connected based on the concept of a Star schema wherein there is a fact table that sits in the center surrounded by all the dimension tables. This architecture helps in denormalization of the data in the warehouse that assist in better performance as the run time performance of complex business queries used in analytics are highly optimized. Additionally, by taking these steps will empower the business users to use the BI systems independently for gaining business insights without taking help from data base administrator or any other technical resource like database developer or tester. This process also helps in taking fast business decision by reducing dependency on the technical team of the organisation.

Data warehouses are centralized and provide good access mechanisms to the user so that people of the organisation can access the historical data for business analysis from multiple office locations. It empowers the business users to gain critical insights and maintain a competitive edge over the business competitors. It provides a BI system layer that gives significant analytical insights in an efficient way by generate quick reports, visualizing the historical data in form of charts / pivots and data cubes. These reporting tools are directly accessing the warehouse and do not interfere with the OLTP database thereby ensuring the smooth functioning of day-to-day business operations. Also, majority of the data warehouse transactions focusses on reading the data and running complex analytical queries. These are done significantly faster than traditional RDMS as warehouse have support of best-in-class processors, memory, and disk availability.

The above pointers clearly indicate the importance of a data warehouse in a competitive digital world and ensuring the growth of an organisation. The future belongs to the organisation who can utilize the data at their disposal to its true potential for marketing purpose, setting up promotions, knowing the customers and segmenting the customers on basis of their choices / purchases. All these important aspects of business continuity can be achieved through a data warehouse.