KJSCE/IT/SYBTech/SEM IV/AD/2022-23



**Experiment No.7**

**Title: Data Warehouse simulation**

**Batch: B2 Roll No.: 16010421119 Experiment No.:7**

**Aim:** To run Data Warehouse simulation

# Resources needed: Different RDBMS such as MySQL, Postgres and Excel, CSV, Rapidminer 5.3

**Theory**

# Data Warehouse :

A data warehouse is a type of data management system that is designed to enable and support business intelligence (BI) activities, especially analytics. Data warehouses are solely intended to perform queries and analysis and often contain large amounts of historical data. The data within a data warehouse is usually derived from a wide range of sources such as application log files and transaction applications.

A data warehouse centralizes and consolidates large amounts of data from multiple sources. Its analytical capabilities allow organizations to derive valuable business insights from their data to improve decision-making. Over time, it builds a historical record that can be invaluable to data scientists and business analysts. Because of these capabilities, a data warehouse can be considered an organization’s “single source of truth**.**

# ETL :

Extract, Transform, Load (ETL) refers to a process in database usage and especially

in data warehousing. Data extraction is where data is extracted from homogeneous or heterogeneous data sources; data transformation where the data is transformed for storing in the proper format or structure for the purposes of querying and analysis; data loading where the data is loaded into the final target database, more specifically, an operational data store, data mart, or data warehouse.

By using an established ETL framework, one may increase one's chances of ending up with better connectivity and scalability. A good ETL tool must be able to communicate with the many different relational databases and read the various file formats used throughout an organization. ETL tools have started to migrate into Enterprise Application Integration, or even Enterprise Service Bus, systems that now cover much more than just the extraction, transformation, and loading of data. A common use case for ETL tools include converting CSV files to formats readable by relational databases. A typical translation of millions of records is facilitated by ETL tools that enable users to input csv-like data feeds/files and import it into a database with as little code as possible. ETL tools in most cases contain a GUI that helps users conveniently transform data, using a visual data mapper, as opposed to writing large programs to parse files and modify data types.

# Activities:

**For Data Warehouse:**

1. Visit [*https://www.oracle.com/webfolder/s/assets/demo/adw-quicktour-na/index.html#step1*](https://www.oracle.com/webfolder/s/assets/demo/adw-quicktour-na/index.html#step1)
2. Go through the demo of Autonomous Data Warehouse for different businesses
3. Prepare a report with following points
   1. The nature of analytics for different businesses given
   2. Comparison between traditional analysis and analysis with Data Warehouse
   3. For any two business type given, specify at least two different scenarios where the tool can be useful.

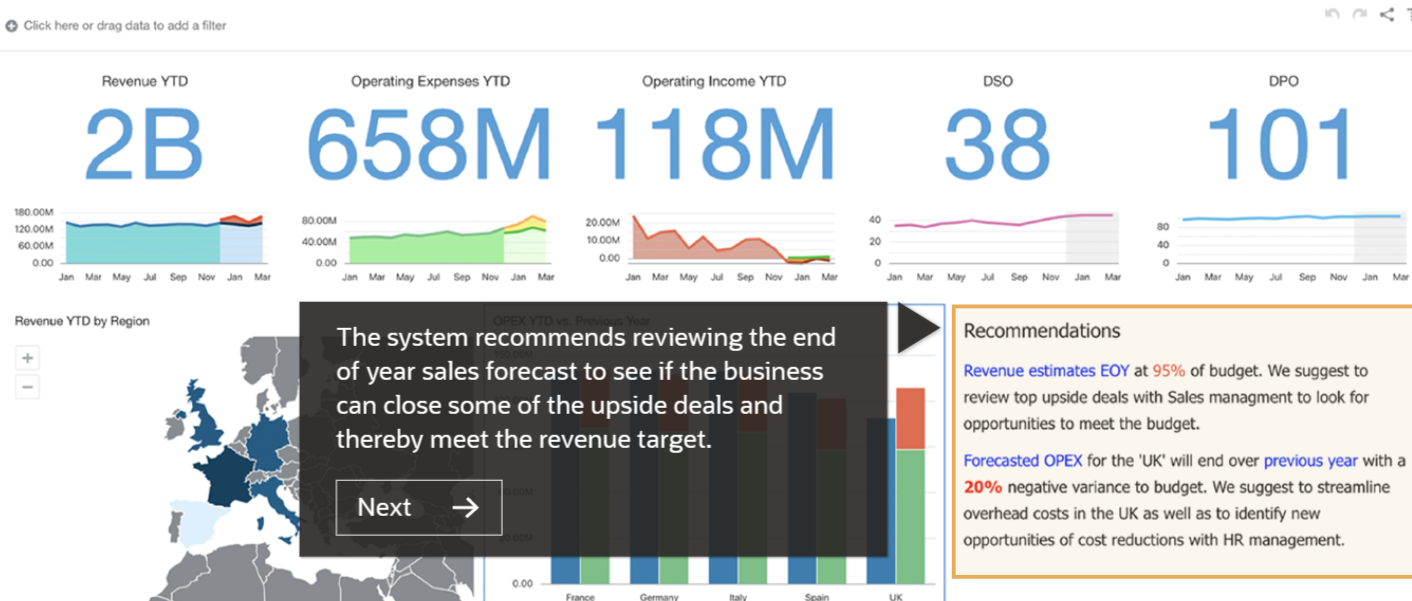
# Results:

**Report for Data Warehouse**

* **Nature of Analytics**

1. **Finance :-**

**Objective 1: Discover how the CFO came to her conclusions**

****

**We are assigned the role of lead finance analyst by the CFO to analyze why the revenue generated from UK is negatively affect. The Sales Forecast reveals that we need to discuss the reliability of the forecast and the possibility to close additional deals with sales management.**

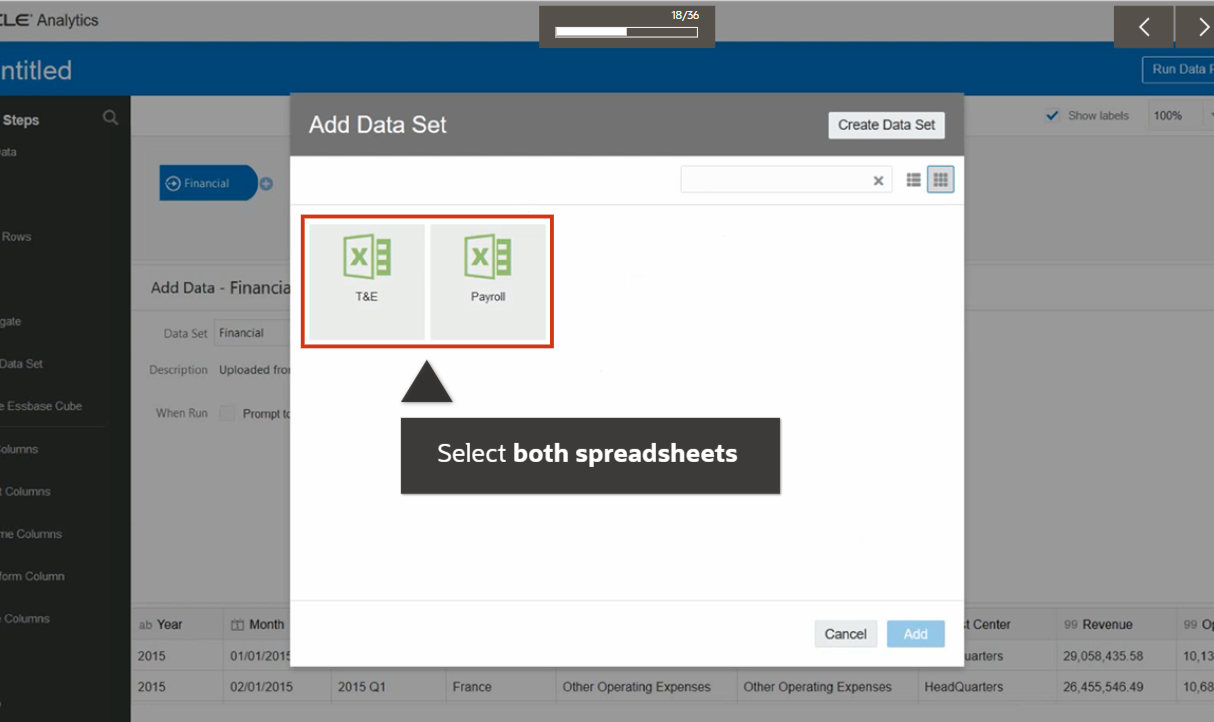
**With this the first objective is complete.**

**Objective 2:**

**Load your Autonomous Database and prepare data for analysis**

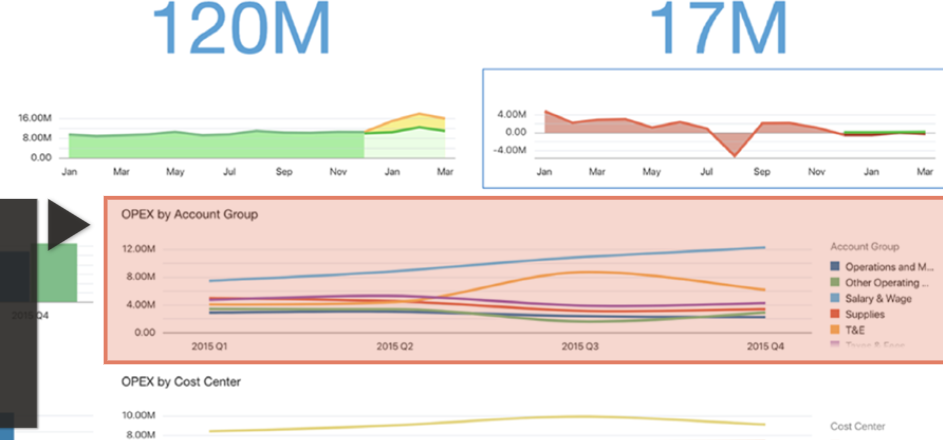
**We are asked to add our own data related to the financial information from the autonomous Data Warehouse . From 2 Spreadsheets the combined data is used.**

**Now we have uploaded a new dataset to Autonomous Data Warehouse and joined it with two others to create a new multi-dimensional dataset for your analysis.**

****

**Objective 3: Analyze UK data to understand the situation**

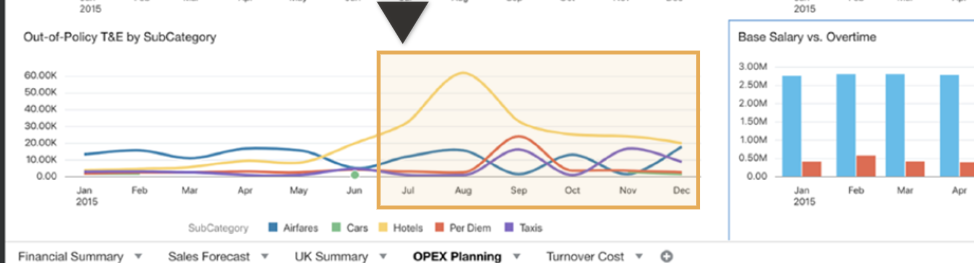
**The UK finance dashboard reveals rising troublesome trends in “OPEX by Account Group”. We have to examine this trend with the given graph.**

****

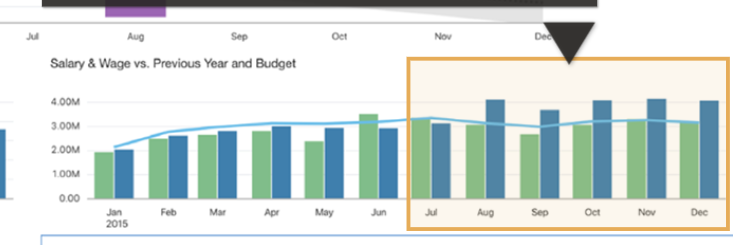
**After analyzing we find that “Salary & Wages” and “Travel and Expense (T&E)” as key culprits in budget loss.**

****

**For Travel and Expenses the budget has begun to increase around July/August.**

****

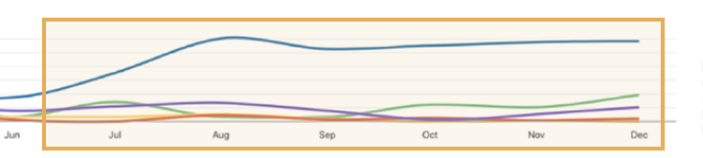
**Investigating “Out of Policy T&E” , hotel expenses spike around July/August and it continues to be a problem.**

****

**As for the “Salary & Wages” issue, starting in August we’ve exceeded both the budget and last year’s numbers**

**Now that you have a better understanding of the situation, We can move o to the final Objective.**

**Objective 4: Dig deeper to discover the why behind the issues**

****

**Drilling into payroll data, overtime hours match the same July/August spike, where we see a trend up in the call center**

**Also the call center had a huge staff turnover in July and has had a hard time filling the open job positions.**

**Combined data from various sources in the HCM system reveals many young, low salary employees have left the business around that same timeframe.**

****

**A word cloud, using data collected from an employee survey, sheds light on why employees may be leaving.**

**Better share with the hiring managers and HR, who can brainstorm ideas to address retention and hiring issues**

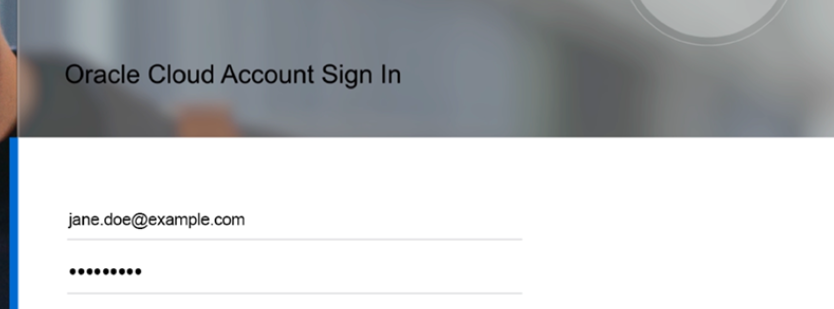
**With this final objective complete we have enabled us to understand the problem with operating profits from all angles. By loading and combining multiple data sources, you’ve diagnosed the root cause and you’re well prepared for a successful board room meeting to share a powerful storyboard of your analysis.**

1. **IT Sector: -**

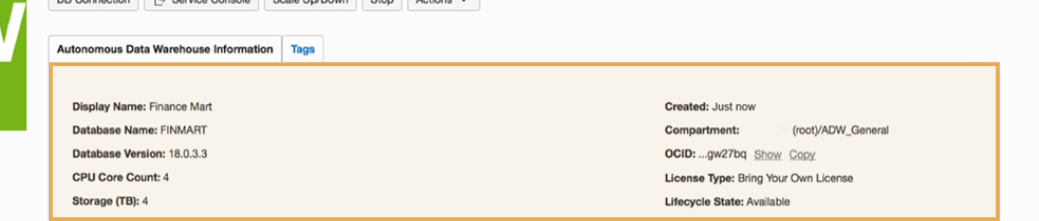
**Scnario : - We are assigned to be an IT professional, Our goal being to support our business colleagues and the objectives of the enterprise from all technical aspects.**

**Objective 1: Set up an Autonomous Data Warehouse**

**We have to begin creating a new autonomous data warehouse, let’s log into Oracle Cloud**

****

**Next step being to name our database, choosing the number of CPUs, editing storage amounts for the data warehouse, specifying the admin password, and hit “Create.”**

****

**Our database is now up and running**

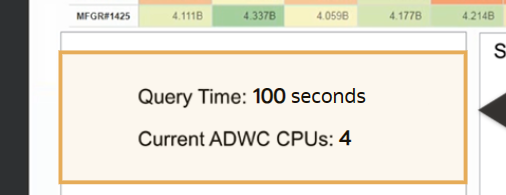
**Oracle auto-configures your new database with optimal settings, provisioning a high-performance database based on Oracle’s Exadata architecture—meaning the system is fully encrypted and secure, with high availability and automatic backups.**

**This completes our first objective.**

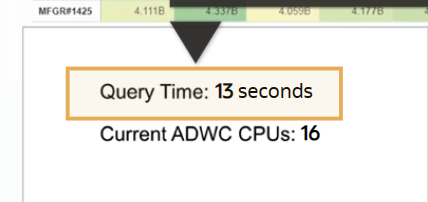
**Objective 2: Scale your warehouse up and down as needed**

**Running against 6 billion records, a typical business analyst report (Profit by Brand and Regions) could take 100 seconds with a 4 CPU data warehouse.**

**Sometimes 100 seconds is way too long as time is precious so we simply and quickly adjust the number of CPUs via a custom call to a REST interface.**

****

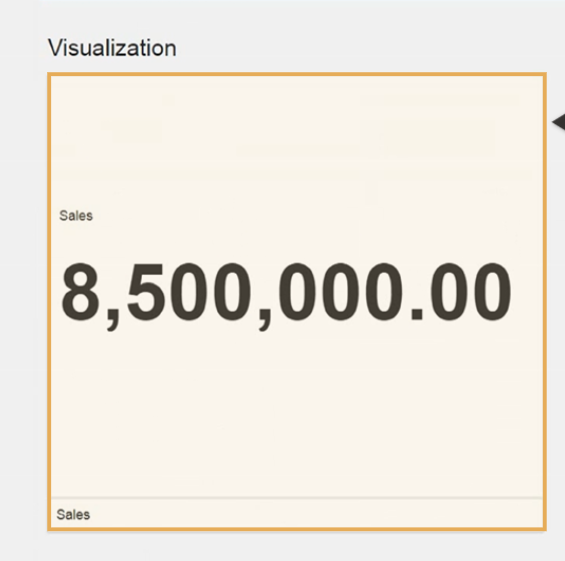
**Once we increase the number of CPUs on one task it decreases the running query time from 100s o 13s.**

****

**With this the objective 2 is complete by scaling the number of CPUs as per accordance.**

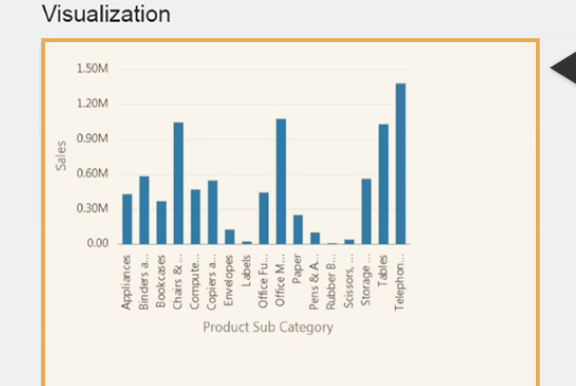
**Just as we are exploring easy scaling, a request for data visualizations comes in from the sales manager**

**Objective 3: Build visualizations for the sales manager**

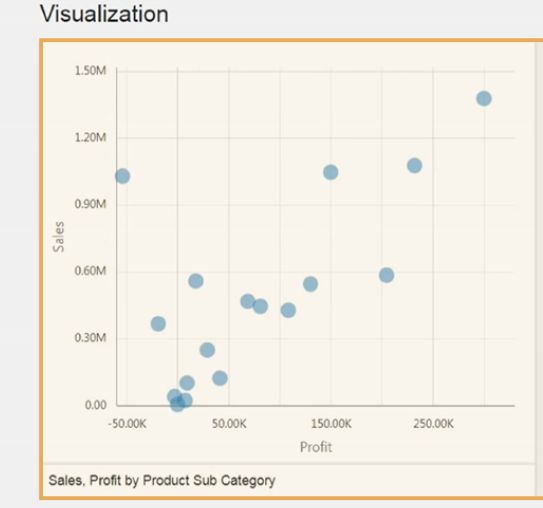
****

**We asked for a visualization of the sales .**

**But sales has various sub categories which has a graph shown below**

****

**We can now analyze the profits from each subcategory In the graph below.**

****

**By adding some customer data we have the updated data in the mass graph**

****

**Thanks to the ease of use and power of Oracle Autonomous Data Warehouse you’ve successfully embraced the power of autonomous with Oracle.**

**With this our Final Objective is complete.**

* **Comparison between traditional analysis and analysis with Data Warehouse**

1. **First, only Autonomous Data Warehouse fully automates database administration.**
2. **Second, the solution is simpler to deploy and use with extensive integrated capabilities that eliminate the need for separate, standalone services—making data warehousing easy to consume for both data analysts and business users.**
3. **Third, Oracle costs less due to its built-in features and autoscaling that closely matches costs to overall workload requirements.**
4. **Fourth, Autonomous Data Warehouse provides stronger built-in security to protect data against both internal and external threats.**
5. **Finally, Oracle makes it easy for customers to comply with data sovereignty requirements.**

* **Some Real Life Scenarios where Autonomous Data Warehouse might be useful.**

1. **Finance**

**Suppose We are working with Ghadi - the up and coming detergent brand. The sales have begun to decline with the rise of competitors like Nirma and Vinn etc. We are assigned the role of lead analyst to inspect these declines in sales and introspect on the current sales approach and refine it to keep up with the competitors in the market.**

1. **IT**

**Suppose we are working for Amazon as an IT professional. We want data related to newer and brighter innovations in the market which are quite underrated. Our role is to**

**Create a new autonomous database and add customer data to it and help the business partners in analyzing the sales of some new innovations and balance the buying and selling trends and approach appropriately.**

# Outcomes: Understanding data warehouse and multi-dimensional modeling

**Conclusion: (Conclusion to be based on the outcomes achieved)**

**We can conclude that we have learnt about Autonomous Data Warehouses.**

# Grade: AA / AB / BB / BC / CC / CD /DD

**Signature of faculty in-charge with date**

# References:

* https:/[/www.ora](http://www.oracle.com/in/database/what-is-a-data-warehouse)c[le.com/in/database/what-is-a-data-warehouse](http://www.oracle.com/in/database/what-is-a-data-warehouse)
* Paulraj Ponniah, “Data Warehousing: Fundamentals for IT Professionals”, Wiley India