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



**Experiment No.7**

**Title: Data Warehouse simulation**

**Batch:A2 Roll No.:16010421073 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:**

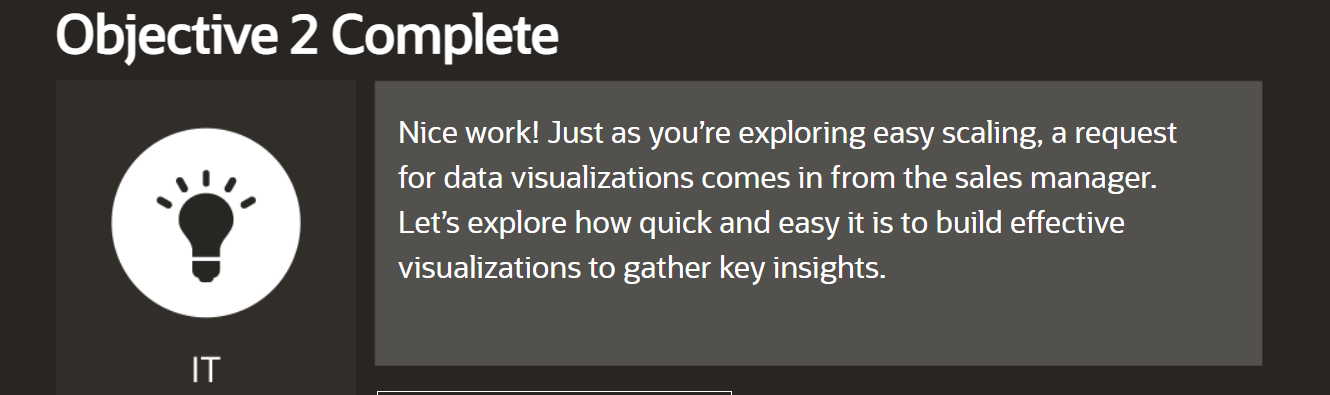
1. Visit [*https://www.oracle.com/webfolder/s/assets/demo/adw-quicktour-na/index.html#step1*](https://www.oracle.com/webfolder/s/assets/demo/adw-quicktour-na/index.html#step1)
2. Go through the demo of Autonomous Data Warehouse for different businesses
3. Prepare a report with following points
   1. 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:

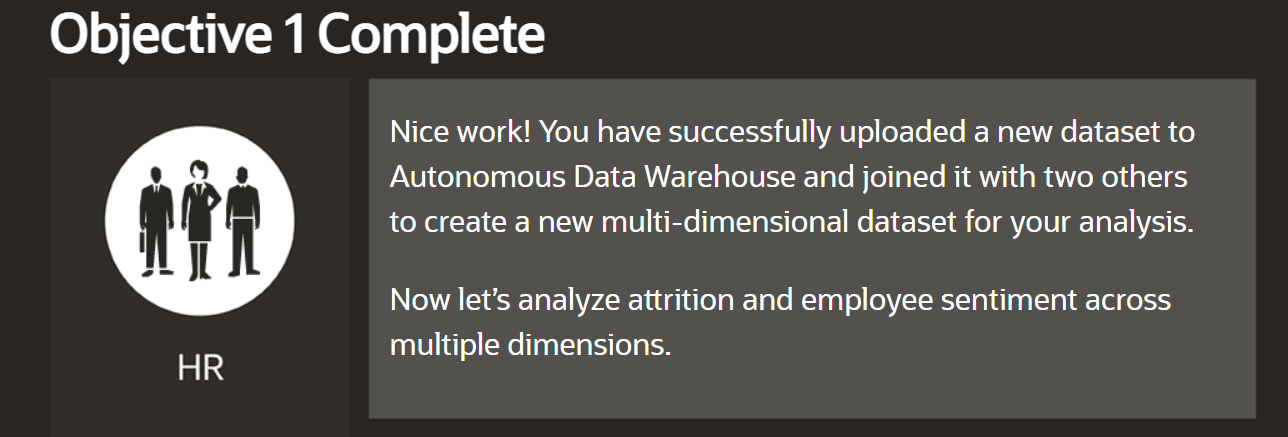
# IT

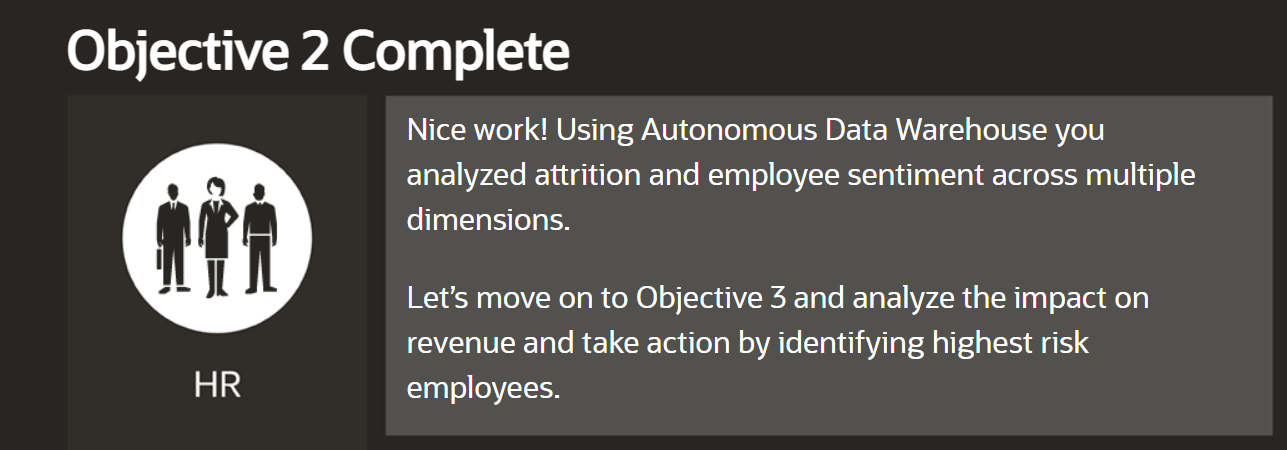
# 

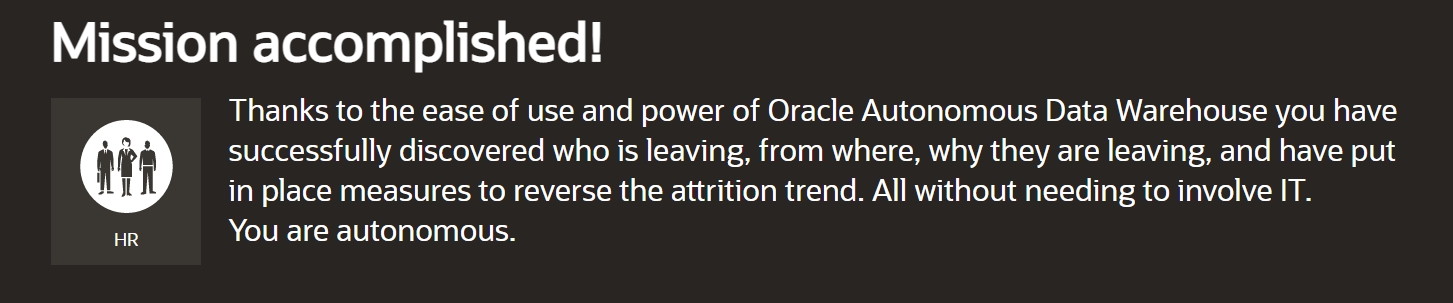
# 



**HR**

****

****

****

1. **Report for Data Warehouse**

**(A)**The nature of analytics for IT and HR in the context of data warehouse can vary depending on the specific needs and goals of each department. Generally speaking, analytics for IT and HR within a data warehouse involve using data to gain insights into various aspects of the organization.

For IT, analytics within a data warehouse can include:

1. **Performance monitoring:** Analyzing system usage patterns, detecting bottlenecks, predicting resource requirements, and optimizing system performance.
2. **Security analytics:** Monitoring security events, identifying potential vulnerabilities, and detecting and responding to security threats.
3. **User behavior analytics**: Analyzing user behavior patterns to identify anomalies or suspicious activities that may indicate a security threat.
4. **Service level analytics:** Analyzing service level data to identify trends and patterns, measure performance against SLAs, and identify areas for improvement.

For HR, analytics within a data warehouse can include:

1. **Talent management:** Analyzing employee data to identify high-performing employees, track employee performance over time, and identify areas for improvement in talent management strategies.
2. **Diversity and inclusion:** Analyzing employee demographics, performance evaluations, and employee engagement surveys to track diversity metrics over time, identify areas for improvement, and measure the impact of diversity and inclusion initiatives.
3. **Recruitment analytics:** Analyzing recruitment data to identify trends in the hiring process, measure the effectiveness of recruitment strategies, and identify areas for improvement.
4. **Employee engagement analytics**: Analyzing employee feedback and engagement data to identify areas of improvement in employee satisfaction, measure the effectiveness of engagement initiatives, and track progress over time.

In summary, analytics for IT and HR within a data warehouse involve using data to gain insights into various aspects of the organization, such as system performance, security, talent management, and employee engagement. By analyzing this data, organizations can make data-driven decisions to optimize their processes and improve their performance.

**(B) Traditional analysis and analysis with Data Warehouse (DW) differ in several ways. Here are some of the main differences:**

* **Data sources**:

In traditional analysis, data is often sourced from multiple systems and applications, and it can be difficult to consolidate and integrate this data. In contrast, DW allows data from different sources to be integrated into a single, centralized repository, which makes analysis much easier.

* **Data structure:** Traditional analysis typically works with raw data in its original format, which can be difficult to work with and analyze. In contrast, DW structures data in a way that is optimized for analysis, with consistent definitions, standardized formats, and efficient query performance.
* **Querying:** Traditional analysis often requires custom queries and ad hoc reporting to get the information needed for analysis. In contrast, DW allows for more flexible querying and reporting, with the ability to slice and dice data in many different ways.
* **Data quality:** Traditional analysis is often limited by data quality issues, such as incomplete, inconsistent, or inaccurate data. DW provides mechanisms for data cleaning and validation, which can help ensure data quality and consistency.
* **Scalability:** Traditional analysis can be limited in its ability to scale to handle large volumes of data. DW is designed to handle large volumes of data, and can easily scale as the amount of data grows.

**(C)**

**IT Business Scenario**

**Scenario:** A retail company has a variety of data sources, including sales data from its point-of-sale (POS) systems, inventory data from its warehouse management system, and customer data from its CRM system. The company wants to create a centralized data repository for analysis and reporting purposes.

**In this scenario, an ETL tool can be useful in the following ways**:

* Extracting data: The ETL tool can extract data from the different data sources, including the POS system, warehouse management system, and CRM system. This process involves identifying the relevant data fields, tables, and schemas for each data source.
* Transforming data: Once the data has been extracted, the ETL tool can transform it into a format that can be easily integrated into the centralized data repository. This may involve standardizing data formats, removing duplicates, or merging data from different sources.
* Loading data: The ETL tool can load the transformed data into the centralized data repository. This involves mapping the data fields from the different data sources to the appropriate fields in the repository.

**HR Business Scenario:**

**An ETL (Extract, Transform, Load) tool can be useful in many scenarios for an HR business. Here's an example:**

**Scenario:** A large company has multiple HR systems, including a payroll system, an employee management system, and a benefits system. The company wants to create a centralized HR data warehouse that consolidates data from these systems to provide a comprehensive view of HR-related data.

**In this scenario, an ETL tool can be useful in the following ways:**

* **Extracting data:** The ETL tool can extract data from the different HR systems, including employee data, payroll data, and benefits data. This process involves identifying the relevant data fields, tables, and schemas for each system.
* **Transforming data:** Once the data has been extracted, the ETL tool can transform it into a format that can be easily integrated into the centralized HR data warehouse. This may involve standardizing data formats, removing duplicates, or merging data from different sources.
* **Loading data:** The ETL tool can load the transformed data into the centralized HR data warehouse. This involves mapping the data fields from the different HR systems to the appropriate fields in the data warehouse.

# Outcomes:

**CO3** : Understanding of data warehouse and its multi-dimensional modeling.

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

**Thus we successfully simulated Data Warehouse simulation and analysed the nature of different businesses.**

# 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