

## Autonomous Data Warehouse:

- It is cloud-based, self-managing data warehousing by the help of AI & ML.
- It can do such complex task by automation.
- It automates administration tasks like provisioning, scaling, backups, simplifying data warehouse management.

Core components:

i) Cloud Infrastructure:

Resource management

Computing resources:

The infrastructure is designed to be elastic - scaling resources up or down based on workload demands.

## storage:

- > The storage facility which is provided to different organizations for storing and managing data.
- > It provides a facility of extracting another resource if one of the resource fails as it keeps many copies of storage.

## Network:

- \* It is responsible for connecting cloud services over the Internet.
- \* For transmission of data and resources externally and internally.

## 2) DataBase Engine:

- > The engine processes queries on various data storage architectures such as local tables, partitioned tables, and tables distributed across multiple servers.

SQL Processing : It supports various data types and complex queries, making it suitable for analytics.

In-memory processing : It is used to speed up query execution.

3)

### AI & ML

> ADW leverages ml capabilities to generate insights & reports automatically from the data stored in the warehouse.

> Analyzing patterns, trends and relationships within the data.

4)

### Security Layer

Data Encryption : It converts data from a readable plaintext format into an unreadable, unencrypted format.

Access control : It is an essential element of security that determines who is allowed to access certain data, apps & resources.

Threat Detection : It is the process of identifying potential security risks to an organization.

## Data Integration :

It is a data management method that involves collecting and merging data from multiple sources.

### ETL (Extract, Transform, Load) :

Extract : Move data from a source system to a temporary staging data repository (as a area).

Transform : The extracted data is transformed into a format that is suitable for loading into the data warehouse.

Load : After the data is transformed it will be loaded into the data warehouse.

## Modern Data Warehouse :

- > It is a cloud-based data management system that collects, stores, transforms and applies data to support an organization.
- > It integrates & stored vast amount of data from various source.
- > It is mainly used for analysis and decision making.
- > It uses cloud-based infrastructure for scalability and flexibility.

Eg : Amazon Redshift,  
Google Big Query.

## Components :

Data source :

operation

Internal data : This data is generated within the company and is specific to it.

> The data come from areas such as operations, maintenance, personnel, and finance.

Eg: Customer data, Research and development data.

External data is the data originating from outside the company and is readily available to the public.

> It can come from areas such as statistics departments, government database.

Eg: market data, environmental data.

### Data Preparation:

> It is the process of transforming raw data into a format that can be analysed and used for business purpose.

Cleaning: Identifying data and removing outliers.

Reformatting: Making data more consistent.

Combining : merging datasets to enrich the data.

Labeling is preparing data for machine learning algorithms.

semantic Layer :

> It is a business-oriented interface that helps users access data from multiple sources by abstracting the underlying data architecture.

> It can help users to analyse, and visualize the complex data.

> It can help users make data-driven decisions.

> It is designed to support the advanced analytics, reporting, and business intelligence functions.

## Authoring:

> It refers to the process of creating and managing data model, report, dash board and other analytical content.

> This process involves defining how data is accessed, presented and analyzed with business requirements.

## Presentation:

> It is way to communicate information to an audience using words and pictures.

> The PL is made up of data mart that present data for a specific subject area in the form of a dimensional model.

> It is the process of comparing 2 or more data sets using visual aids, such as graphs,

Eg: Scatter plot,  
Histogram,  
Bar chart,  
Pie chart