Datawarehousing With IBM Cloud Db2 Warehouse

# ETL Process in Data warehouse:

* ETL stands for Extract, Transform, Load and it is a process used in data warehousing to extract data from various
* sources, transform it into a format suitable for loading into a data warehouse, and then load it into the warehouse.

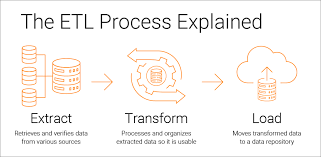
**Extract:** The first stage in the ETL process is to extract data from various sources such as transactional systems, spreadsheets, and flat files. This step involves reading data from the source systems and storing it in a staging area.

1. **Transform:** In this stage, the extracted data is transformed into a format that is suitable for loading into the data

warehouse. This may involve cleaning and validating the data, converting data types, combining data from multiple sources, and creating new data fields.

1. **Load:** After the data is transformed, it is loaded into the data warehouse. This step involves creating the physical data

structures and loading the data into the warehouse.



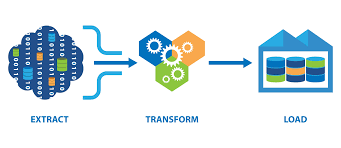
# ETL Tools :

Most commonly used ETL tools are Hevo, Sybase, Oracle Warehouse builder, CloverETL, and MarkLogic.

# Data Warehouse :

Most commonly used Data Warehouses are Snowflake, Redshift, BigQuery, and

Firebolt.



Advantages of ETL process in data **warehousing:**

1. **Improved data quality :** ETL process

ensures that the data in the data warehouse is accurate, complete, and up-to-date.

1. **Better data integration :** ETL process helps to integrate data from multiple sources and systems, making it more accessible and usable.
2. **Increased data security :** ETL process can help to improve data security by controlling access to the data warehouse and ensuring that only authorized users can access the data.
3. **Improved scalability :** ETL process can help to improve scalability by providing a way to manage and analyze large amounts of data.
4. **Increased automation :** ETL tools and

technologies can automate and simplify the ETL process, reducing the time and effort

required to load and update data in the warehouse.

# Disadvantages of ETL process in data warehousing:

1. **High cost :** ETL process can be expensive to implement and maintain, especially for organizations with limited resources.
2. **Complexity :** ETL process can be complex and difficult to implement, especially for organizations that lack the necessary

expertise or resources.

1. **Limited flexibility :** ETL process can be

limited in terms of flexibility, as it may not

be able to handle unstructured data or real-

time data streams.

**4. Limited scalability :** ETL process can be

limited in terms of scalability, as it may not be able to handle very large amounts of data.

**5. Data privacy concerns :** ETL process can raise concerns about data privacy, as large amounts of data are collected, stored, and analyzed.

**SELECT column1, column2** FROM source\_table WHERE condition;

**Transform:** Apply transformations to the extracted data.

**Extract:** Retrieve data from the source.

# UPDATE target\_table

SET transformed\_column =

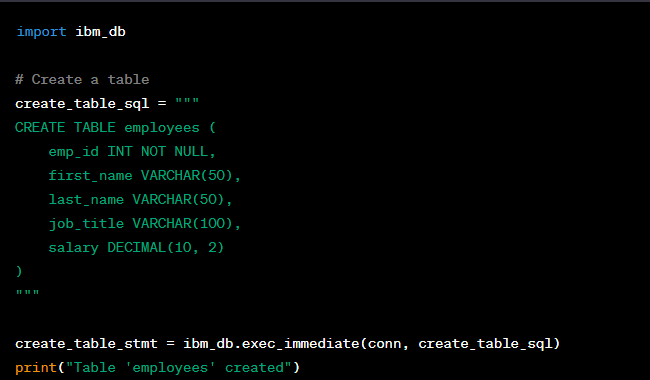
some\_transformation\_function(original\_col umn);

**Load:** Load the transformed data into the target database.

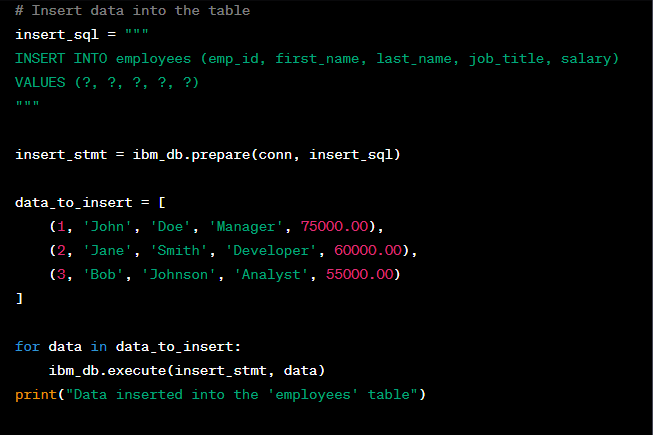
# INSERT INTO target\_table (column1, column2)

VALUES (value1, value2);

**ETL process:**



# Data Exploration:



**SQL Queries :**

**SELECT Customer id** , Customer Name FROM Customer ;

**INSERT INTO Customers** (CustomerName, ContactName, Initial , Phone no, Date of

Birth, Email id)

# UPDATE Customers

SET CustomerName = ‘Afra’, Sex= ‘Female’ WHERE CustomerID = 19;

**DELETE FROM Customers** WHERE CustomerName=’Midhuna’;

**CREATE TABLE Customers**( Customer Id Int,

Customer Name Varchar(20), Initial Varchar(10),

Sex Varchar (20),

Email id Varchar(20),

Phone no Number, Date of Birth Number,

);