

Development phase-2

Data warehousing using IBM Cloud Db2 warehouse

Implementing ETL processes to extract, transform, and load data into the data warehouse :

The ETL (Extract, Transform, Load) process is a crucial component in data warehousing and plays a significant role in the collection, preparation, and delivery of data for analytical and reporting purposes. Here's an overview of the ETL process in data warehousing:

1. **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.
2. **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.
3. **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.

```
datawarehouse.py - C:\Users\shafiq\AppData\Local\Programs\Python\Python311\datawarehouse.py (3.11.0)*
File Edit Format Run Options Window Help
import pandas as pd
from sqlalchemy import create_engine
source_data = pd.read_csv('source_data.csv')
def transform_data(data):
    transformed_data = data.groupby('product_category')['sales_amount'].sum().reset_index()
    return transformed_data
transformed_data = transform_data(source_data)
def load_data_to_warehouse(data):
    db_url = "postgresql://username:password@hostname/database"
    engine = create_engine(db_url)
    data.to_sql('data_warehouse_table', engine, if_exists='replace', index=False)

    engine.dispose()

load_data_to_warehouse(transformed_data)
```

Analysing data within Db2 Warehouse using SQL queries and analysis techniques :

```
queries.sql x + 3zs7urh6u NEW MYSQL RUN
1 create table customers(
2 customer_id INT PRIMARY KEY,
3 customer_name VARCHAR(100), email VARCHAR(100));
4
5 create table orders(
6 order_id INT PRIMARY KEY,
7 customer_id INT,
8 order_date DATE,
9 total_amount DECIMAL(10,2));
```

STDIN

Output:

Program did not output anything!

-- Connect to the Db2 Warehouse database :

```
CONNECT TO <datawarehousing> USER <admin > USING <admin >;
```

-- Select data from a table :

```
SELECT *  
  
FROM customers  
  
LIMIT 10;
```

-- Calculate statistics on numeric data :

```
SELECT  
  
    AVG(total_amount) AS avg_total,  
    MIN(total_amount) AS min_total,  
    MAX(total_amount) AS max_total,  
FROM your_table;
```

-- Filter data based on conditions :

```
SELECT *  
  
FROM orders  
  
WHERE order_date BETWEEN '2023-01-01' AND '2023-12-31';
```

Summary :

Data architects can explore and analyze data within IBM Db2 Warehouse using SQL queries and various analysis techniques. Db2 Warehouse is a powerful data warehousing solution that supports SQL, and it provides various tools and capabilities for data analysis. Here's a step-by-step guide on how data architects can explore and analyze data in Db2 Warehouse.