Data Warehousing

Cloud Applications Development

211321205002-S.Arun Kumar

Data warehousing is a process and technology that involves collecting, storing, and managing large volumes of data from various sources in a centralized repository known as a "data warehouse." This data is stored in a structured and optimized manner to support analytical and reporting tasks. Data warehousing enables organizations to efficiently store historical data, integrate data from different sources, and provide a foundation for business intelligence and data analysis, helping in making informed decisions and gaining insights into business operations. It plays a crucial role in organizing and optimizing data for the purpose of querying, reporting, and data analysis.

import pandas as pd

import psycopg2

# Connect to the source database

connection = psycopg2.connect(

dbname="your\_db",

user="your\_user",

password="your\_password",

host="your\_db\_host"

)

# Extract data into a DataFrame

query = "SELECT \* FROM your\_table"

data = pd.read\_sql\_query(query, connection)

connection.close()

# Load data into Redshift

from sqlalchemy import create\_engine

engine = create\_engine('postgresql://username:password@redshift-host:5439/database')

data.to\_sql('your\_target\_table', engine, if\_exists='replace', index=False)

# Query data from Redshift

query = "SELECT column1, column2 FROM your\_target\_table WHERE condition"

result = pd.read\_sql(query, engine)