

Course Name: Business Intelligence Lab

Course Code: CSP-421

Experiment: 2.1

Aim: Implement OLAP (Online Analytical Processing) for Business Insights

IDE USED: - Python, R

Software Required:

- Python Jupyter Notebook / R

Description: The experiment involves setting up an OLAP system to analyze a dataset and extract meaningful insights. The dataset can be sourced from various data repositories within the organization, such as transactional databases or data warehouses. The OLAP system will enable users to explore the data from multiple dimensions or perspectives, facilitating deeper analysis and visualization.

Pseudo code/Algorithms/Flowchart/Steps:

- 1.Dataset Selection: Choose a suitable dataset that represents a relevant business scenario. The dataset should contain structured data with multiple dimensions, such as time, geography, product categories, or customer segments.
- 2.OLAP System Setup: Set up an OLAP system, which could be a software platform or a cloud-based service. Configure the system to connect to the dataset and establish the necessary data connections.
- 3.Dimensional Modeling: Design the dimensions and hierarchies for the dataset. Define the relationships between the dimensions, such as time hierarchies (year, quarter, month), geographic hierarchies (country, region, city), or product hierarchies (category, subcategory, product).
- 4.Cube Creation: Create a multidimensional cube based on the dataset and dimensional model. Define the measures and aggregations required for analysis. Aggregations could include sum, average, count, minimum, maximum, etc.
- 5.Data Loading: Load the dataset into the OLAP cube. This process involves extracting, transforming, and loading the data to ensure it is in the appropriate format for analysis.

Implementation :

```
import numpy as np
```

```
import sqlite3
```

```
# Create a SQLite database
```

```
conn = sqlite3.connect("sales_data.db")
```

```
cursor = conn.cursor()
```

Course Name: Business Intelligence Lab

Course Code: CSP-421

Create tables (sales and products) and insert sample data

```
cursor.execute("""
```

```
CREATE TABLE IF NOT EXISTS sales (  
    transaction_id INTEGER PRIMARY KEY AUTOINCREMENT,  
    product_id INTEGER,  
    sale_date DATE,  
    quantity INTEGER,  
    revenue REAL )
```

```
""")
```

```
cursor.execute("""
```

```
CREATE TABLE IF NOT EXISTS products (  
    product_id INTEGER PRIMARY KEY,  
    product_name TEXT )
```

```
""")
```

Insert some sample data

```
cursor.executemany("""
```

```
INSERT INTO products (product_id, product_name) VALUES (?, ?)
```

```
""", [(1, 'Product A'), (2, 'Product B')])
```

```
cursor.executemany("""
```

```
INSERT INTO sales (product_id, sale_date, quantity, revenue) VALUES (?, ?, ?, ?)
```

```
""", [(1, '2023-01-01', 10, 100.0), (2, '2023-01-01', 5, 50.0)])
```

```
cursor.executemany("""
```

```
INSERT INTO sales (product_id, sale_date, quantity, revenue) VALUES (?, ?, ?, ?)
```

```
""", [(1, '2003-09-21', 50, 10.0), (3, '2003-09-21', 5, 50.0)])
```

```
conn.commit()
```

```
conn.close()
```

Load data into a DataFrame

```
conn = sqlite3.connect("sales_data.db")
```

```
query = ""
```

```
SELECT p.product_name, strftime('%Y-%m', s.sale_date) as month,
```

Course Name: Business Intelligence Lab**Course Code:** CSP-421

```
SUM(s.quantity) as total_quantity, SUM(s.revenue) as total_revenue

FROM sales s

JOIN products p ON s.product_id = p.product_id

GROUP BY p.product_name, month

ORDER BY p.product_name, month

'''

df = pd.read_sql_query(query, conn)

conn.close()

# Pivot the data for analysis

pivot_table = pd.pivot_table(df, values=['total_quantity', 'total_revenue'],

                              index='product_name', columns='month')

# Display the result

print(pivot_table)
```

Output:

Requirement already satisfied: pandas in c:\users\abhis\appdata\local\programs\python\python311\lib\site-packages (2.1.0)Note: you may need to restart the kernel to use updated packages.

Requirement already satisfied: numpy>=1.23.2 in c:\users\abhis\appdata\local\programs\python\python311\lib\site-packages (from pandas) (1.25.2)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\abhis\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\abhis\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: tzdata>=2022.1 in c:\users\abhis\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: six>=1.5 in c:\users\abhis\appdata\local\programs\python\python311\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)

	total_quantity	total_revenue
month	2023-01	2023-01
product_name		
Product A	10.0	100.0
Product B	5.0	50.0

LEARNING OUTCOME:

- Understanding of OLAP
- Technical Skills: It help in acquiring hands-on experience in setting up and configuring an OLAP system, designing dimensional models, creating cubes, and loading data for analysis.
- Data Analysis Techniques: Participants will learn various OLAP analysis techniques, such as slicing and dicing, drill-down, and roll-up operations, to explore data from different dimensions and levels of detail.
- Business Insights: By analyzing the dataset using OLAP, participants will extract meaningful insights and understand the potential impact of such insights on business operations, performance, and strategic decision-making.