

**Aim:-** To design, create, and populate a data warehouse using sample data from multiple sources in Google Firestore.

**Software Used:** IDLE

**Theory :**

## **Steps -**

First, install the required libraries.

```
pip install pandas google-cloud-firestore
```

- Set up Firebase Firestore:
- Go to the Firebase console: <https://console.firebase.google.com/>
- Create a new project.
- Enable Firestore in the project.
- Set up authentication and get your service account key (JSON file). Save this file as serviceAccountKey.json.
- Create CSV files for regional sales data.

### **Follow this steps for Google Firebase Account:**

Step 1: Search on Google "Firebase"

Step 2: Click on "Firebase Console"

Step 3: Click on "Create a Project"

Step 4: "Enter your Project Name"

Step 5: Go to Left Panel and Click on "Firestore Database"

Step 6: "Create Database"

Step 7: Go to "Project Overview Setting" -> "Project Setting" -> "Service Accounts"

Step 8: Click on "Generate Private Key"

Step 9: Save this private key in Practical 6 folder.

Step 10: Save dataset1 in Practical 6 folder.

Step 11: Save dataset2 in Practical 6 folder.

Step 12: Create a Code and save in Practical 6 folder.

### **Dataset1 (region1\_sales.csv)**

```
OrderID,Product,Quantity,Price,Date
1,Widget,5,25.00,2024-01-01
2,Widget,3,25.00,2024-01-03
3,Gadget,7,50.00,2024-01-04
```

### **Dataset2 (region2\_sales.csv)**

```
OrderID,Product,Quantity,Price,Date
4,Widget,2,25.00,2024-01-02
5,Gadget,1,50.00,2024-01-05
6,Gizmo,4,75.00,2024-01-06
```

### **Code:**

```
import pandas as pd
from google.cloud import firestore

# Initialize Firestore DB
import os
os.environ["GOOGLE_APPLICATION_CREDENTIALS"] = "serviceAccountKey.json"
db = firestore.Client()

# Extract data from CSV files
df_region1 = pd.read_csv('region1_sales.csv')
df_region2 = pd.read_csv('region2_sales.csv')

# Transform data: Standardize column names and data types if necessary
df_region1['Region'] = 'Region1'
```

```

df_region2['Region'] = 'Region2'

# Load data into Firestore
def load_data_to_firestore(df, collection_name):
    for _, row in df.iterrows():
        doc_ref = db.collection(collection_name).document(str(row['OrderID']))
        doc_ref.set(row.to_dict())

load_data_to_firestore(df_region1, 'sales')
load_data_to_firestore(df_region2, 'sales')

# Verify the data loading
sales_ref = db.collection('sales')
docs = sales_ref.stream()

print("Data in Firestore:")
for doc in docs:
    print(f'{doc.id} => {doc.to_dict()}')

```

## Output-

```

Data in Firestore:
1 => {'Product': 'Widget', 'OrderID': 1, 'Price': 25.0, 'Region': 'Region1', 'Quantity': 5, 'Date': '2024-01-01'}
2 => {'Product': 'Widget', 'OrderID': 2, 'Price': 25.0, 'Region': 'Region1', 'Quantity': 3, 'Date': '2024-01-03'}
3 => {'Product': 'Gadget', 'OrderID': 3, 'Price': 50.0, 'Region': 'Region1', 'Quantity': 7, 'Date': '2024-01-04'}
4 => {'Product': 'Widget', 'OrderID': 4, 'Price': 25.0, 'Region': 'Region2', 'Quantity': 2, 'Date': '2024-01-02'}
5 => {'Product': 'Gadget', 'OrderID': 5, 'Price': 50.0, 'Region': 'Region2', 'Quantity': 1, 'Date': '2024-01-05'}
6 => {'Product': 'Gizmo', 'OrderID': 6, 'Price': 75.0, 'Region': 'Region2', 'Quantity': 4, 'Date': '2024-01-06'}

```