Final Assignment using PostgreSQL

Scenario

You are a data engineer hired by a solid waste management company. The company collects and recycles solid waste across major cities in the country of Brazil. The company operates hundreds of trucks of different types to collect and transport solid waste. The company would like to create a data warehouse so that it can create reports like

- total waste collected per year per city
- total waste collected per month per city
- total waste collected per quarter per city
- total waste collected per year per trucktype
- total waste collected per trucktype per city
- total waste collected per trucktype per station per city

You will use your data warehousing skills to design and implement a data warehouse for the company.

Objectives
In this assignment you will:

- Design a Data Warehouse
- Load data into Data Warehouse
- Write aggregation queries
- Create MQTs
- Create a Dashboard

Exercise 1 - Design a Data Warehouse

- https://github.com/craigtrupp/ibm-deng/tree/main/DataWarehousing_BIAnalytics/Labs/FinalProject/TxtTables

- https://github.com/craigtrupp/ibm-deng/tree/main/DataWarehousing_BIAnalytics/Labs/FinalProject/Screenshots
    - https://dbdiagram.io/d/WasteDiagram-652485feffbf5169f05a37eb
    - Use this to more easily create the postgresql script and visualize the relationship between the dimensions and the fact of interest (just seems like waste tons is our sole fact)

Exercise 2 - Create schema for Data Warehouse on PostgreSQL

- https://github.com/craigtrupp/ibm-deng/blob/main/DataWarehousing_BIAnalytics/Labs/FinalProject/WasteDiagram.sql

Exercise 3 - Load data into the Data Warehouse

In this exercise you will load the data into the tables.

After the initial schema design, you were told that due to operational issues, data could not be collected in the format initially planned.

- This implies that the previous tables (MyDimDate, MyDimWaste, MyDimZone, MyFactTrips) and their associated attributes are no longer applicable to the current design. The company has loaded data using CSV files per the new design

You will load the data provided by the company in csv format.

- Quick download of the scripts with wget

```
theia@theiadocker-craigtrupp8:/home/project$ wget
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB02
60EN-SkillsNetwork/labs/Final%20Assignment/DimDate.csv
--2023-10-09 20:04:29--
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB02
60EN-SkillsNetwork/labs/Final%20Assignment/DimDate.csv
Resolving cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud
(cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud)...
169.63.118.104
Connecting to cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud
(cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud)|169.63.118.104
|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 16508 (16K) [text/csv]
Saving to: 'DimDate.csv'

DimDate.csv
100%[=======================================================>]  16.12K
--.-KB/s    in 0s
```

```
2023-10-09 20:04:29 (101 MB/s) - 'DimDate.csv' saved [16508/16508]

theia@theiadocker-craigtrupp8:/home/project$ wget
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB02
60EN-SkillsNetwork/labs/Final%20Assignment/DimTruck.csv
--2023-10-09 20:04:43--
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB02
60EN-SkillsNetwork/labs/Final%20Assignment/DimTruck.csv
Resolving cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud
(cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud)...
169.63.118.104
Connecting to cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud
(cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud)|169.63.118.104
|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 767 [text/csv]
Saving to: 'DimTruck.csv'

DimTruck.csv
100%[=======================================================>]     767
--.-KB/s    in 0s

2023-10-09 20:04:43 (79.8 MB/s) - 'DimTruck.csv' saved [767/767]

theia@theiadocker-craigtrupp8:/home/project$ wget
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB02
60EN-SkillsNetwork/labs/Final%20Assignment/DimStation.csv
--2023-10-09 20:04:53--
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB02
60EN-SkillsNetwork/labs/Final%20Assignment/DimStation.csv
Resolving cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud
(cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud)...
169.63.118.104
Connecting to cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud
(cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud)|169.63.118.104
|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 282 [text/csv]
Saving to: 'DimStation.csv'

DimStation.csv
100%[=======================================================>]     282
--.-KB/s    in 0s
```

```
2023-10-09 20:04:53 (28.2 MB/s) - 'DimStation.csv' saved [282/282]

theia@theiadocker-craigtrupp8:/home/project$ wget
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB02
60EN-SkillsNetwork/labs/Final%20Assignment/FactTrips.csv
--2023-10-09 20:05:20--
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB02
60EN-SkillsNetwork/labs/Final%20Assignment/FactTrips.csv
Resolving cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud
(cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud)...
169.63.118.104
Connecting to cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud
(cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud)|169.63.118.104
|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2433579 (2.3M) [text/csv]
Saving to: 'FactTrips.csv'

FactTrips.csv
100%[=============================================================>]   2.32M
--.-KB/s    in 0.02s

2023-10-09 20:05:20 (138 MB/s) - 'FactTrips.csv' saved [2433579/2433579]
```

Exercise 4 - Write aggregation queries and create MQTs

In this exercise you will query the data you have loaded in the previous exercise.

- Task 13 - Create a grouping sets query

Create a grouping sets query using the columns stationid, trucktype, total waste collected.

```
-- Create a grouping sets query using the columns stationid, trucktype,
total waste collected.
SELECT
      ft.stationid, dt.trucktype, SUM(ft.wastecollected) AS total_waste
FROM FACTTRIPS AS ft
LEFT JOIN DIMTRUCK AS dt
      ON ft.truckid = dt.truckid
GROUP BY GROUPING SETS(ft.stationid, dt.trucktype)
ORDER BY total_waste DESC;
```

Take a screenshot of the sql and the output rows.

Name the screenshot 13-groupingsets.jpg. (Images can be saved with either the .jpg or .png extension.)

- Task 14 - Create a rollup query

Create a rollup query using the columns year, city, stationid, and total waste collected.

```
-- Create a rollup query using the columns year, city, stationid, and total
waste collected.
SELECT
       dd.year, ds.city, ft.stationid, SUM(ft.wastecollected) AS
total_waste
FROM FACTTRIPS AS ft
LEFT JOIN DIMDATE AS dd
      USING(dateid)
LEFT JOIN DIMSTATION AS ds
      USING(stationid)
GROUP BY ROLLUP(dd.year, ds.city, ft.stationid)
ORDER BY dd.year, ft.stationid;
```

Take a screenshot of the sql and the output rows.

Name the screenshot 14-rollup.jpg. (Images can be saved with either the .jpg or .png extension.)

- Task 15 - Create a cube query

Create a cube query using the columns year, city, stationid, and average waste collected.

```
-- Create a cube query using the columns year, city, stationid, and average
waste collected.
SELECT
       dd.year, ds.city, ft.stationid, ROUND(AVG(ft.wastecollected), 3) AS
avg_waste
FROM FACTTRIPS AS ft
LEFT JOIN DIMDATE AS dd
      USING(dateid)
LEFT JOIN DIMSTATION AS ds
      USING(stationid)
GROUP BY CUBE(dd.year, ds.city, ft.stationid)
ORDER BY dd.year, ft.stationid;
```

Take a screenshot of the sql and the output rows.

Name the screenshot 15-cube.jpg. (Images can be saved with either the .jpg or .png extension.)

- Task 16 - Create an MQT

Create an MQT named max_waste_stats using the columns city, stationid, trucktype, and max waste collected.

- https://www.ibm.com/docs/en/db2-for-zos/13?topic=rewrite-creating-materialized-query-table

```
-- Create an MQT named max_waste_stats using the columns city, stationid,
trucktype, and max waste collected.
DROP TABLE IF EXISTS cds03902.max_waste_stats;
CREATE TABLE max_waste_stats(city, stationid, trucktype, max_waste) AS
(SELECT dst.city AS city, dst.stationid AS stationid, dt.trucktype,
MAX(ft.wastecollected) AS max_waste
FROM FACTTRIPS AS ft
LEFT JOIN DIMSTATION AS dst
      ON ft.stationid = dst.stationid
LEFT JOIN DIMTRUCK AS dt
      ON ft.truckid = dt.truckid
GROUP BY dst.city, dst.stationid, dt.trucktype)
DATA INITIALLY DEFERRED
REFRESH DEFERRED
MAINTAINED BY SYSTEM;

REFRESH TABLE max_waste_stats;
```

```
SELECT * FROM max_waste_stats;
```

Take a screenshot of the sql.

Name the screenshot 16-mqt.jpg. (Images can be saved with either the .jpg or .png extension.)

Exercise 5 - Create a dashboard using Cognos Analytics

Download the data from DataForCognos_date

Use the DataForCognos_date.csv file to generate the following charts.

Task 17 - Create a pie chart in the dashboard

- Create a pie chart that shows the waste collected by truck type.

Take a screenshot of the pie chart.

Name the screenshot 17-pie.jpg. (Images can be saved with either the .jpg or .png extension.)

Task 18 - Create a bar chart in the dashboard

- Create a bar chart that shows the waste collected station wise.

Take a screenshot of the bar chart.

Name the screenshot 18-bar.jpg. (Images can be saved with either the .jpg or .png extension.)

Task 19 - Create a line chart in the dashboard

- Create a line chart that shows the waste collected by month wise.

Take a screenshot of the line chart.

Name the screenshot 19-line.jpg. (Images can be saved with either the .jpg or .png extension.)

Hint: Use Extract function in Create Calculation to collect the data month wise.

Task 20 - Create a pie chart in the dashboard

- Create a pie chart that shows the waste collected by city.

Take a screenshot of the pie chart.

Name the screenshot 20-pie.jpg. (Images can be saved with either the .jpg or .png extension.)

End of the assignment.

- https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FWarehouse_FProject_Dash&action=view&mode=dashboard&subView=model0000018b1a40a092_00000000

Final Quiz - Section

Congratulations! You passed!

Grade received 91.66%

Latest Submission Grade 91.67%

To pass 80% or higher

Go to next item

1. Which of the following is a total cost of ownership consideration when choosing a data warehouse? **1 / 1 point**

- ○ Data privacy
- ○ Ease of use
- ◉ Data migration
- ○ Location

Expand

Correct
Correct! Data migration and integration costs for moving data into the warehouse and pruning and purging as required are considerations when looking at total cost of ownership.

2. Which of the following are the data sources for data marts? **1 / 1 point**

- ○ Hierarchical databases
- ◉ Data warehouses
- ○ Data lakes
- ○ File data storage systems

Expand

Correct
Correct! Data marts use transactional databases or data warehouses as data sources.

3. Which of the following statements is true regarding Snowflake? **1 / 1 point**

- ○ It is a hybrid data warehouse provider.
- ○ It uses Amazon Web Services-specific hardware for accelerated data compression and encryption.
- ○ It is a multicloud data platform that unifies data lakes, data warehouses, analytics, and new data sources and types.
- ◉ It offers data encryption for both data in transit and at rest.

Expand

Correct

Correct! Up-and-coming Snowflake offers a multicloud analytics solution that complies with GDPR and CCPA data privacy regulations. Snowflake advertises its always-on encryption of data in transit and at rest.

4. Who is a typical user of data lakes? **1 / 1 point**

- ○ Product manager
- ○ Business analyst
- ○ Data analyst
- ◉ Machine learning engineer

Expand

Correct

Correct! Data scientists, data developers, and machine learning engineers are the typical users of data lakes.

5. Which of the following is an example of a “fact” in the context of data warehousing? **1 / 1 point**

- ○ Zip code
- ○ Order ID
- ○ Make of a car
- ◉ Daily sales total

Expand

Correct
Correct! Facts are quantities that can be measured, such as temperature, number of sales, or millimeters of rainfall.

6. Which of the following is true regarding a star schema? **1 / 1 point**

- ○ They can be represented by nodes on a graph.
- ◉ They contain at most one fact table.
- ○ They contain at most one dimension table.
- ○ They are normalized snowflake schemas.

Expand

Correct
Correct! The fact table is at the center of a star schema with dimension tables linked to the fact table’s primary key.

7. What operations easily generate fact summaries that may be required by management? **1 / 1 point**

- ○ Filters
- ◉ CUBE and ROLLUP
- ○ ELT
- ○ Staging tables

Expand

Correct
Correct! CUBE and ROLLUP provide summary reports.

8. Which statement is true in regard to staging areas? **1 / 1 point**

- ◉ They decouple data processing from source systems to minimize the risk of data corruption.
- ○ They help normalize star schemas in order to create snowflake schemas.
- ○ They assist in aggregating data from one or more sources into a single, central, consistent data store to support various data analytics requirements.
- ○ They serve as a way to organize materialized views.

Expand

Correct
Correct! Staging areas decouple data processing from the source systems and thus help minimize the risk of data corruption.

9. Which of the following URLs would you use to sign up for a trial of Cognos Analytics? **1 / 1 point**

- IBM.BIZ/COGNOS_TRIAL
- IBM.COM/TRIAL_COGNOS
- IBM.BIZ/TRY_COGNOS
- IBM.COM/TRY_COGNOS

Expand

Correct
Correct! This is the correct URL to sign up for a trial of Cognos Analytics.

10. Which statement best describes prescriptive analytics? **1 / 1 point**

- Prescriptive analytics provide insight into actions an organization should take to create a specific outcome.
- Prescriptive analytics provide insight into the past.
- Prescriptive analytics describe the basic features of data in a study and provide simple summaries about the sample and its measures.
- Prescriptive analytics provide insight in what could happen in the future.

Expand

Correct
Correct! Analytical outcomes can be descriptive, prescriptive, or predictive. Prescriptive analytics provide insight into actions an organization should take to create a specific outcome.

11. What are the three methods you can use to create visualizations in Cognos? **1 / 1 point**

- [x] Automatically

  Correct
  Correct! One of the methods you can use to create a visualization in Cognos is automatically.

- [x] Manually

  Correct
  Correct! One of the methods you can use to create a visualization in Cognos is manually.

- [ ] Cognos Help
- [x] Cognos Assistant

  Correct
  Correct! One of the methods you can use to create a visualization in Cognos is using the Cognos Assistant.

Expand

Correct
Great, you got all the right answers.

12. When connecting Cognos to a data repository with IBMdb2, what must you remember to do before entering the JDBC URL into Cognos? (Select two correct answers.) **0 / 1 point**

- [x] Replace the username and password in the JDBC URL with actual credentials.

  Correct
  Correct! Before entering the JDBC URL into Cognos, you must replace the username and password with actual values.

- [ ] Name the service credential.
- [ ] Name the connection.
- [ ] Remove angle brackets from the JDBC URL