

Final Assignment - Introduction to Data Warehousing

Estimated time needed: 90 minutes

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

Screenshots

Throughout this lab you will be prompted to take screenshots and save them on your own device. These screenshots will need to be uploaded for peer review in the next section of the course. You can use various free screengrabbing tools or your operating system's shortcut keys (Alt + PrintScreen in Windows, Command + Shift + 5 on Mac, Shift + Ctrl + Show windows on Chromebook) to capture the required screenshots. The screenshots can be either jpg or png.

About the dataset

The dataset you would be using in this assignment is not a real life dataset. It was programmatically created for this assignment purpose.

Prerequisites

- You need access to a cloud instance of IBM DB2 to proceed with this assignment. If you do not have an instance of IBM DB2 on cloud, follow the instructions in this [lab](#) to create one.
- You need access to Cognos Analytics. This [lab](#) will guide to get your access to Cognos Analytics, and also get you started with how to use it to analyze the data.

Exercise 1 - Design a Data Warehouse

The solid waste management company has provided you the sample data they wish to collect.

Trip number	Waste Type	Waste Collected in tons	Collection Zone	City	Date
1	Dry	45.23	South	Sao Paulo	23-Jan-20
2	Wet	43.12	Central	Rio de Janeiro	24-Jan-20
3	Electronic	40.19	South	Sao Paulo	23-Jan-20
4	Plastic	34.87	West	Rio de Janeiro	24-Jan-20
5	Wet	45.34	West	Rio de Janeiro	23-Jan-20

You will start your project by designing a Star Schema warehouse by identifying the columns for the various dimension and fact tables in the schema.

Task 1 - Design the dimension table MyDimDate

Write down the fields in the MyDimDate table in any text editor one field per line. The company is looking at a granularity of day. Which means they would like to have the ability to generate the report on yearly, monthly, daily, and weekday basis.

Here is a partial list of fields to serve as an example:

dateid

month

monthname

...

...

Take a screenshot of the fieldnames for the table MyDimDate.

Name the screenshot 1-MyDimDate.jpg. (Images can be saved with either the .jpg or .png extension.)

Task 2 - Design the dimension table MyDimWaste

Write down the fields in the MyDimWaste table in any text editor one field per line.

Take a screenshot of the fieldnames for the table MyDimWaste.

Name the screenshot 2-MyDimWaste.jpg. (Images can be saved with either the .jpg or .png extension.)

Task 3 - Design the dimension table MyDimZone

Write down the fields in the MyDimZone table in any text editor one field per line.

Take a screenshot of the fieldnames for the table MyDimZone.

Name the screenshot 3-MyDimZone.jpg. (Images can be saved with either the .jpg or .png extension.)

Task 4 - Design the fact table MyFactTrips

Write down the fields in the MyFactTrips table in any text editor one field per line.

Take a screenshot of the fieldnames for the table MyFactTrips.

Name the screenshot 4-MyFactTrips.jpg. (Images can be saved with either the .jpg or .png extension.)

Exercise 2 - Create schema for Data Warehouse on DB2

In this exercise you will create the tables, you have designed in the previous exercise.

Task 5 - Create the dimension table MyDimDate

Create the MyDimDate table.

Take a screenshot of the sql statement you used to create the table MyDimDate.

Name the screenshot 5-MyDimDate.jpg. (Images can be saved with either the .jpg or .png extension.)

Task 6 - Create the dimension table MyDimWaste

Create the MyDimWaste table.

Take a screenshot of the sql statement you used to create the table MyDimWaste.

Name the screenshot 6-MyDimWaste.jpg. (Images can be saved with either the .jpg or .png extension.)

Task 7 - Create the dimension table MyDimZone

Create the MyDimZone table.

Take a screenshot of the sql statement you used to create the table MyDimZone.

Name the screenshot 7-MyDimZone.jpg. (Images can be saved with either the .jpg or .png extension.)

Task 8 - Create the fact table MyFactTrips

Create the MyFactTrips table.

Take a screenshot of the sql statement you used to create the table MyFactTrips.

Name the screenshot 8-MyFactTrips.jpg. (Images can be saved with either the .jpg or .png extension.)

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.

Task 9 - Load data into the dimension table DimDate

Download the data from <https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB0260EN-SkillsNetwork/labs/Final%20Assignment/DimDate.csv>

Load this data into DimDate table.

Take a screenshot of the first 5 rows in the table DimDate.

Name the screenshot 9–DimDate.jpg. (Images can be saved with either the .jpg or .png extension.)

Task 10 - Load data into the dimension table DimTruck

Download the data from <https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB0260EN-SkillsNetwork/labs/Final%20Assignment/DimTruck.csv>

Load this data into DimTruck table.

Take a screenshot of the first 5 rows in the table DimTruck.

Name the screenshot 10–DimTruck.jpg. (Images can be saved with either the .jpg or .png extension.)

Task 11 - Load data into the dimension table DimStation

Download the data from <https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB0260EN-SkillsNetwork/labs/Final%20Assignment/DimStation.csv>

Load this data into DimStation table.

Take a screenshot of the first 5 rows in the table DimStation.

Name the screenshot 11–DimStation.jpg. (Images can be saved with either the .jpg or .png extension.)

Task 12 - Load data into the fact table FactTrips

Download the data from <https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB0260EN-SkillsNetwork/labs/Final%20Assignment/FactTrips.csv>

Load this data into FactTrips table.

Take a screenshot of the first 5 rows in the table FactTrips.

Name the screenshot 12–FactTrips.jpg. (Images can be saved with either the .jpg or .png extension.)

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.

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.

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.

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.

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 or Google Looker Studio

Download the data from <https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/UKqcgqXi00W1OdpXC1JHTw/DataForCognosLooker.csv>

Use the DataForCognosLooker.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.)

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.

Authors

Ramesh Sannareddy

Other Contributors

Rav Ahuja

© IBM Corporation 2023. All rights reserved.