HW_4_DataWarehouse

Overview

- This homework is worth 15 points (15%) toward your final grade.
- It is **due on Wednesday 7/10/2019, 11:59 pm**.
- Late policy applies (grace period of 3 days with 20% late penalty). After that time (7/13/2019), no late work will be accepted.
- Save your results in a document (such as a .txt file, MS Word or similar tool) and submit your assignment deliverable in a PDF on Canvas.

You must turn in BOTH your SQL and your ANSWER SET (unless otherwise specified).

This homework will give you hands-on practice in working with a sample Data Warehouse. In this homework you will use a data warehouse (built according to the dimensional model/star schema) and populated by using scripts provided. You will use the data warehouse answering query questions.

Objectives

1. Become familiar with the structure of a dimensional model / star schema data warehouse
2. Understand the unique nature of the date dimension
3. Successfully run the scripts necessary to create the sample data warehouse consisting of 5 dimension tables and one fact table. Run a script to verify that your data warehouse is correctly built.
4. Run SQL against your data warehouse to answer the assigned problems.

Step One: Ensure that your MySQL environment is working.

Make sure that MySQL is up and running on your device, and that you are able to run SQL queries against your database(s). Also, make sure Safe Undates is unchecked (Edit - SQL Editor - Uncheck Safe Undates). Then using your query editor, you must connect to the running MySQL instance prior to running the script.

Step Two: Creating the Data Warehouse Tables

Once you have selected your query editor, you need to download the "**HW_4_SQL_Sales_DW.sql**" script file from **Canvas**, and execute it. It will create your database and tables, and then load the tables with data. The script runs fine as-is without any modification.

The script begins with a command to DROP the table before it creates it. This allows you to run the script over and over as needed.

There is an ERD of this database in the Canvas called "**HW_4 DW_ERD.pdf**". You should download and print this diagram and keep it handy when you are doing the homework. It is very helpful to have table and column names in front of you when writing SQL queries.

Your **HW_4_SQL_Sales_DW** database consists of the following tables:

- Dim_Product
- Dim_Store
- Dim_Customer
- Dim_Date
- Dim_SalesPerson
- Fact_ProductSales

- Dates, Numbers, Numbers_Small – Created and used by the script to create the dim_date dimension, but NOT used for anything else.

After creating the five dimension and one fact tables from **HW_4_SQL_Sales_DW.sql** script, run the "verify" script. You should see the following tables and row counts for each.

Note: Please let instructor know if your tables/counts are different.

```
--  Verify row counts
```

```
USE HW_4_SQL_sales_dw;

SELECT 'Table', 'Rows' FROM dim_customer
UNION
SELECT 'dim_customer', COUNT(*) FROM dim_customer
UNION
SELECT 'dim_date', COUNT(*) FROM dim_date
UNION
SELECT 'dim_product', COUNT(*) FROM dim_product
UNION
SELECT 'dim_salesperson', COUNT(*) FROM dim_salesperson
UNION
SELECT 'dim_store', COUNT(*) FROM dim_store
UNION
SELECT 'fact_productsales', COUNT(*) FROM fact_productsales;
```

Table	Rows
dim_customer	5
dim_date	3653
dim_product	5
dim_salesperson	6
dim_store	3
fact_productsales	297

HW_4_DataWarehouse

Data Warehouse Query Problems

- You must turn in BOTH your SQL and your ANSWER SET.
- Make sure to use table/column aliases. Also, use derived column names, e.g. SalesPrice * Quantity = **Total Revenue** so derived column should call **'Total Revenue'**
- **All Questions are equally weighted.**

1. Display Sales Person and Total Revenue for a sales person who has produced the highest total revenue during 2012? (Revenue = SalesPrice * Quantity).

2. Display Top 3 Customer and their Total Revenue for 2013, sort by highest revenu 1st? (Revenue = SalesPrice * Quantity)

3. Display Year, StoreID and their 'Total Sales Price' for the year 2016 and 2017.

4. Which Top 3 Productnames yielded the highest profit in 2015, display by highest profit 1st? (Profit = (sales price - actual cost) * quantity)

5. In which Quarter in 2016 did the 'ValueMart Boulder' store show the highest revenue? (Revenue = SalesPrice * Quantity).

6 What is the total sales price for all items purchased by customer Melinda Gates?

7. What is the total revenue by storename for all items purchased in March 2013? (Revenue = SalesPrice * Quantity)

8. Who is the best performing SalesPerson? That is, the salesperson with the highest total revenue? (Revenue = SalesPrice * Quantity)

9. Which product shows the largest profit from sales?
(Profit = the difference between Revenue (SalesPrice * Quantity) and Cost (ProductCost * Quantity.))

10. Display Year, monthname and total revenue for the three months trend in total revenue (SalesPrice * Quantity) for January, February, and March 2017.