CSCI3287 Database Systems

Extra Credit # 2 – Data Warehouse Lab

Overview

This extra credit assignment has been created to allow students to increase their total number of points earned in this course in order to boost your final grade.

This extra credit assignment is worth up to 50 points (out of 1000) toward your final grade. It is due on Wednesday, December 18th at 11:59 p.m.   No extension can be allowed for this assignment. (Grades must be turned in to the registrar by Dec. 23, and we will need a few days to do the grading.) Your assignment submission should be a document saved and submitted as a PDF file via the link found in the assignment section of "Week Sixteen" in Moodle which is the same place where you found this file.

This assignment will give you hands-on practice in working with a sample Data Warehouse. In this assignment you will run scripts to create a data warehouse in MySQL, load the DW with data, and then answer some questions about the data in the warehouse you have created.

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. Run SQL against your data warehouse to analyze the data contained in the data warehouse
4. 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.
5. Run SQL against your data warehouse to answer the assigned problems.

Submission Requirements:

Your results for this extra credit assignment should be captured in a document (such as a .txt file, MS Word or similar tool.) Please then save your final deliverable document as a PDF. Use the link found in the Extra Credit section of the Week 16 Moodle page.

Step One: Ensure that your MySQL environment is working.

This assignment follows on Homework # 3 in which you created a MySQL environment on your personal computer and executed queries against it. For Homework # 5, you must first ensure that MySQL is up and running on your device, and that you are able to run SQL queries against your database(s).

Step Two: Creating the Data Warehouse Tables

To get started, you need to download some scripts from Moodle and create your database and tables, and then load the tables with data.

Before you can create your database, you need to make sure that your instance of MySQL is running. Then using your query editor, you must connect to the MySQL instance.

Creating the Data Warehouse:

The SQL statements to create the tables for your Extra Credit # 2 Data Warehouse can be found on the Week 16 Moodle Page under the Extra Credit assignments heading. There are 9 scripts for you to run.

Script # 1: Create the Database, turn off SAFE MODE if it is "on"

Scripts # 2-6: Create and Load the Five Dimension Tables

Script # 7: Create and Load the Fact Table

Script # 8: Create the Foreign Key Constraints needed for the Fact Table

Script # 9: Run the "verify" script to ensure that the data warehouse is built properly

Your 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.

CSCI3287 Database Systems

Extra Credit # 2 – Data Warehouse Lab

To complete this extra credit assignment, you must open up each of these 9 script files, copy the SQL statements, paste the SQL statements into your query editor and execute the script.

Note: Most scripts begin with a command to DROP the table before it creates it. This allows you to run the script over and over as needed.

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

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

Data Warehouse Problems

For this homework you must answer the questions below stating the results of your analysis of the data in the data warehouse. Each answer should be stated in a sentence or two providing the requested analysis. For these questions, you must create and execute one or more SQL Queries against the sales_dw data warehouse to answer the question. In addition to the answer to the question, you must turn in your **SQL code** AND your **answer set** from the query you used to determine the answer.

Question 1. Which salesperson produced the most total revenue for this organization during 2012? (Total Revenue = SalesPrice * Quantity)

Question 2. Which customer's revenue increased the most from 2012 to 2013? (Total Revenue = SalesPrice * Quantity)

Question 3. Rank all stores by their total revenue for each year represented in the data warehouse. In other words, for a given year rank the three stores by revenue for that year.

Question 4. Which product yielded the highest profit in 2015? (Profit = (sales price – actual cost) * quantity)

Question 5. In which calendar quarter in 2016 did the Boulder store show the highest revenue?