

CSCI3287 Database Systems

Homework # 3 – Data Warehouse Lab

Overview

This homework is worth 10 points (out of 100) toward your final grade. It is due on Sunday, April 22, at 11:55 p.m. Late submissions will be penalized 50% during a 3-day grace period up until Wednesday, April 25, 11:55 p.m. After that time, no late work will be accepted. Your homework submission should be a document saved and submitted as a PDF file via the link found in the Homework section of the “Week 13 Moodle April 9 – 13” -- which is the same place where you found this file.

This homework will give you hands-on practice in working with a sample Data Warehouse. In this homework you will create a data warehouse (built according to the dimensional model/star schema) and populate it using scripts provided. You will use the data warehouse you create for running some analytical queries and answering a few questions.

Objectives

1. Become familiar with the structure of a dimensional model / star schema data warehouse
2. Understand the unique nature of the data 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.

This assignment follows on Homework Assignment # 2 in which you created a MySQL environment on your personal computer and executed queries against it. For Homework # 3, 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).

If your MySQL environment (DBMS Engine and Query Editor) isn't working properly, please refer to the instructions for Homework #2 to get the MySQL environment set up and working.

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.

CSCI3287 Database Systems

Homework # 3 – Data Warehouse Lab

Creating the Data Warehouse:

The SQL statements to create the tables for your Homework # 3 data warehouse can be found on the Moodle site “Week 13 Moodle April 9 – 13” under the Homework assignments heading. There are 9 scripts for you to run.

Script # 1: Create the Database

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.

To complete this homework, 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.

CSCI3287 Database Systems

Homework # 3 – Data Warehouse Lab

table_name	table_rows
dim_customer	5
dim_date	3514
dim_product	5
dim_salesperson	6
dim_store	3
fact_productsales	25

Preparing Your Homework Submission

Your results for this homework 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 Homework Assignment section of “Week 13 Moodle April 9 – 13” Moodle site to submit your work for grading. If you are doing PAIR PROGRAMMING on this assignment, please be sure to identify the name of your programming partner on your submission. You must EACH submit your own results document for this homework.

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 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.

1. What is the total sales price for all items purchased by customer Melinda Gates?
2. What is the total revenue by store for all items purchased in March 2013? (Total Revenue = SalesPrice * Quantity)
3. Who is the best performing SalesPerson? (That is, the salesperson with the highest total revenue amount?)
4. Which product shows the largest profit from sales? (Profit = the difference between Total Revenue (SalesPrice * Quantity) and Total Cost (ProductCost * Quantity.))
5. Describe the three month trend in total sales revenue comparing January, February, and March 2013.