

Friday, March 20

Spring Break !!

Homework # 4

Homework # 5

See you next time on Monday, March 30

HW5 – Data Warehouse

Overview:

- 100 points (out of 1000)
- Due on Sunday April 5th, at 11:59 p.m.
- Submit as a PDF file via the link found in the assignment section of “Week Twelve” in Moodle

This assignment gives students hands-on practice in working with a realistic data warehouse.

HW5 – Data Warehouse

Objectives

- Create a connection to the Adventureworks Data Warehouse (“aw”) on a virtual shared server in our CU CS Private Cloud
- Examine the Adventureworks Data Warehouse to become familiar with the structure of a realistic star schema data warehouse
- Using a SQL query editor (such as MySQL Workbench)
 1. Study the database and answer questions about its structure
 2. Create and run queries against the warehouse to answer questions about the content of the warehouse
 3. Create and run analytical queries against the warehouse to analyze Adventureworks’ business

HW5 – Data Warehouse

Why?

Challenge you to get a glimpse of what it feels like when you step into an organization and are required to figure out their databases

Study and understand an unfamiliar database

Using the tools and skills you've learned in class,

- figure out how their database is structured
- what data is in what tables
- how a realistic data warehouse is actually constructed

HW5 – Data Warehouse

Submission Requirements:

- Answers must be clearly identified and numbered according to the questions
- Where the question requires SQL, submit **both** your SQL code and your answer set.
- Where the question requires a written answer, submit proper sentences (spelling and grammar)
- Save the document as a PDF and submit your PDF using the Homework # 5 link

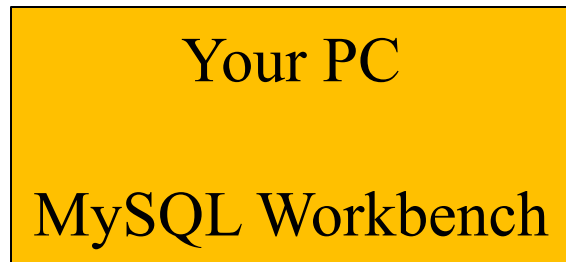
HW5 – Data Warehouse

“Pair Programming”

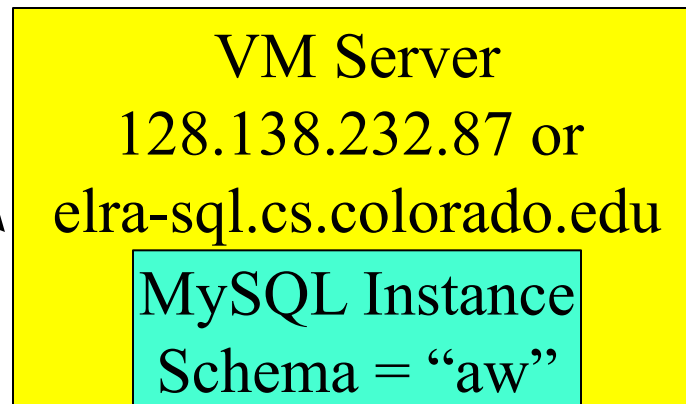
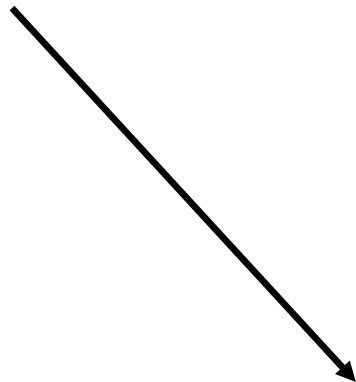
- You may work with **one** partner on completing this assignment.
- However, this is an individual assignment: each one must submit your own final deliverable
- Be sure to specify your partner's name on the document you submit.

HW5 – Data Warehouse

Connecting to the Data Warehouse



1. Connect to the VM via SSH
2. Connect to the MySQL instance



HW5 – Data Warehouse

This database is managed by MySQL running on a Linux VM server.

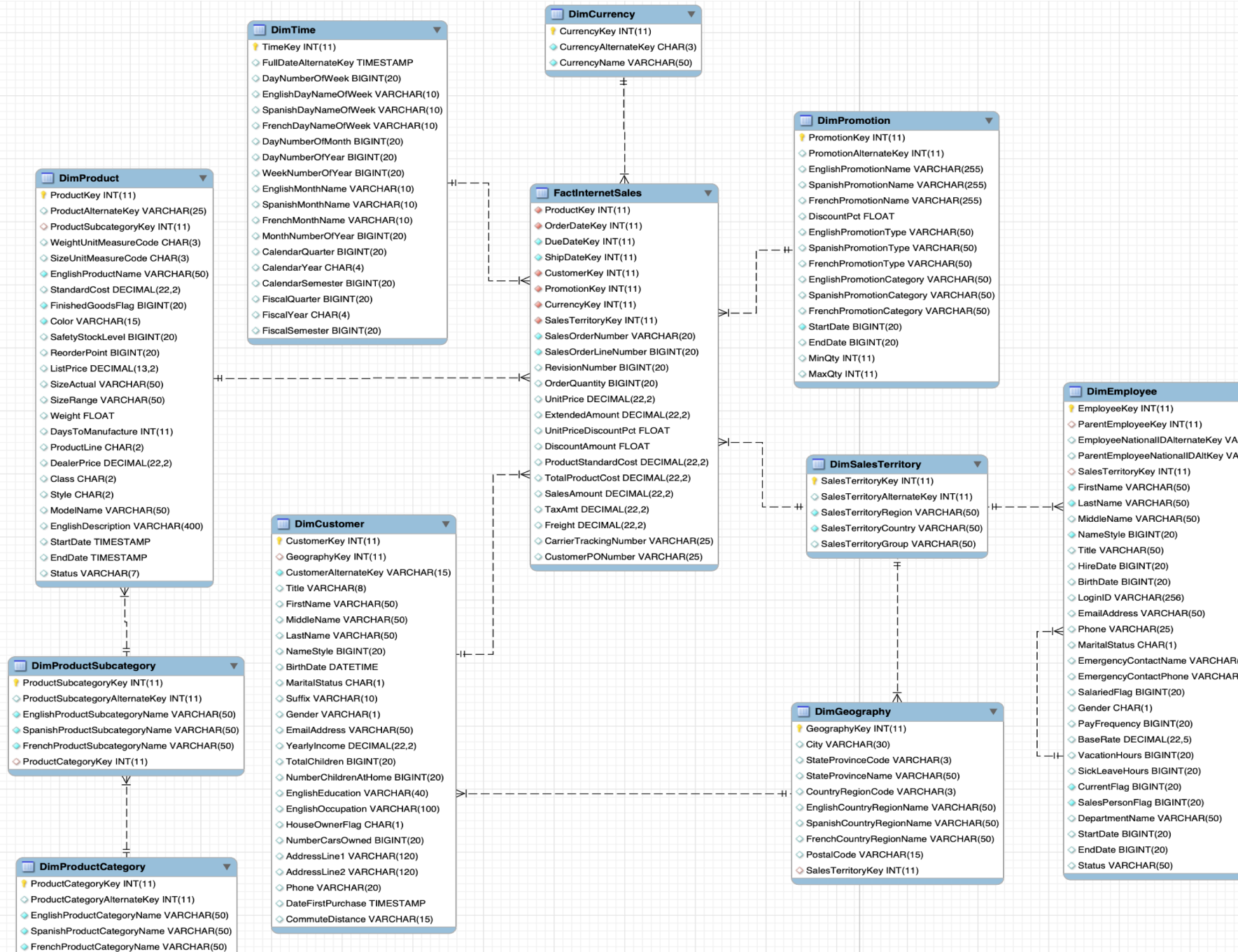
Because it is Linux, the names of database objects are case sensitive:

DimProduct \neq Dimproduct

HW5 – Data Warehouse

Adventureworks

- The AdventureWorks data warehouse is based on a fictitious bicycle manufacturing company named Adventure Works Cycles.
- Microsoft created this company and its databases (an OLTP database *and* a Star Schema Data warehouse) to assist their customers in learning about database technologies (in order to sell more MS SQL Server!!)
- It was built for MS SQL Server, but I converted part of it to MySQL
- Multinational: descriptions in several languages



HW5 – Data Warehouse

Adventureworks Data Model

- FactInternetSales Granularity: one row represents the online sale of one item by AdventureWorks to a customer.
- One sale = bike, article of clothing, accessory (like a helmet), or a repair/replacement part for a bike.
- We will only look at **bike sales**.

UnitPrice = the dollar amount of the sale.

OrderQuantity = number of items sold.

OrderDateKey = date of the sale

ProductStandardCost = Adventureworks' cost

UnitPrice - ProductStandardCost = profit or margin

Adventureworks Data Model

There is a hierarchy within the product dimensions

A specific **product** sold belongs to a **subcategory**

A **subcategory** belongs to a **category**

A category consists of many subcategories

A subcategory consists of many products

Adventureworks Data Model

- `DimProductSubcategory.EnglishProductSubcategoryName`
= type of item sold (Bike, Component, Accessory, etc.)
- `DimProduct.EnglishProductName`
= more detailed product/model information
- `DimProduct.ModelName`
= specific model of bike

Information_schema

The MySQL “catalog”

```
use information_schema;  
show tables;  
desc tables;  
Select table_name, table_rows  
      from tables  
     where table_name like 'nw%';
```

HW5 – Data Warehouse

Information_schema

```
desc columns;
```

If a column is a Primary Key,
then `columns.column_key` contains “PRI”

Information_schema

For more information:

<https://dev.mysql.com/doc/refman/8.0/en/information-schema-introduction.html>