Overview

This Project is worth 100 points (out of 1000) toward your final grade. It is due on Sunday April 5, at 11:59 p.m.

Your assignment submission should be a document saved and submitted as a PDF file via the link found in the assignment section of "Week Twelve" in Moodle which is the same place where you found this file.

This assignment will give you hands-on practice in working with a realistic data warehouse.

Objectives

- Create a connection to the Adventureworks Data Warehouse ("aw") on a shared server
- 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 queries against the warehouse to analyze Adventureworks' business.

Submission Requirements:

1. Compose answers to the questions assigned below in a document. Answers must be clearly identified and numbered according to the questions below.
2. Where the question requires SQL, submit both your SQL code and your answer set.
3. Where the question requires a written answer, please submit proper sentences explaining your answer to the question.
4. Save the document as a PDF and submit your PDF using the Homework # 5 link in the Moodle page for assignment section of "Week Twelve" in Moodle which is the same place where you found this file.

   NOTES:

   You may work with a partner on completing this assignment. However, this is an **individual** assignment; each one must submit your own final deliverable for this assignment. If you did work with a partner, be sure to specify your partner's name on the document you submit.

Assignment Problems:

NOTE: This is a Linux server and Linux commands are typically case sensitive. If you attempt a query and it looks right but doesn't work, be sure that the case used in naming database objects in your SQL exactly matches object names on the server.

NOTE: Be sure to read "Hints regarding the data" on page 4 of this document prior to attempting to answer these questions.

1. Use the information_schema to find out how many rows there are in each table in the adventureworks data warehouse. Show the table name and its row count.

   Hints:
   - `use information_schema;`
   - There is a table within `information_schema` called `TABLES`.

2. Use the information_schema to list out each table in the adventureworks data warehouse and its primary key.

   Hints:
   - `use information_schema;`
   - There is a table within `information_schema` called `COLUMNS`.
   - There is a column within the `COLUMNS` table called `column_key`

3. What standard table naming convention did the AdventureWorksDW database designers use to differentiate dimension tables from fact tables in this star schema data warehouse?

4. What do you think is the purpose of the recursive relationship on `DimEmployee`?

5. What are the three types of models of bikes sold by AdventureWorks?

6. Compare and rank the total counts of the bikes sold by AdventureWorks for each of the years 2001 – 2004 by color. What was the most popular color of bikes sold in each of these 4 years? Provide your SQL query, and your answer set along with your answer to the question. You can assume that one row in the fact table equals one sale.

   HINT: Since the fact table contains sales for all kinds of products, you should include only fact rows where the sale is for a **bike**. One easy way to do this is a WHERE clause selecting only rows where EnglishProductSubcategoryName contains the string "bikes".

7. List and compare the total sales volume (in dollars) of bikes sold (all model types) by customer state/province by year. Which 4 states/provinces showed the highest sales volume for each of the years from 2001 – 2004? Provide your SQL query, and a text-based answer to the question. You do not need to provide your SQL answer set (it may be rather long.)

HINT: Bikes sales only!

HINT: You can list all the states/regions in descending order by year and just provide in your answer the top four for each year.

8. For the year 2002, which model of bike yielded the highest margin for AdventureWorks? Provide your SQL query, and your answer set along with your answer to the question.

HINT: Bikes sales only!

Appendix A - Background on the Company:

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 people in learning about database technologies.

Adventure Works produces and distributes metal and composite bicycles to North American, European, and Asian commercial and consumer markets. The base of operations is located in Bothell, Washington with about 500 employees, and several regional sales teams are located throughout their market base.

Adventure Works sells products wholesale to specialty shops and to individuals through the Internet. For the data warehouse exercises, you will work with the **AdventureWorksDW** Internet sales tables, which contain realistic patterns that work well for data warehousing exercises.

The Internet Sales star schema contains a fact table with data regarding customer purchases of bicycles via the web.

The Internet Sales star schema contains information on several thousand customers. These customers live in several countries, which are combined into three regions:

- North America (83%)
- Europe (12%)
- Australia (7%)

The database contains sales data covering several fiscal years.

The products in the database are broken down by subcategory, model, and product.

AdventureWorks does business in multiple countries, so some attributes in the data warehouse contain descriptions in multiple languages.

In 2000, Adventure Works Cycles bought a small manufacturing plant, Importadores Neptuno, located in Mexico. Importadores Neptuno manufactures several critical subcomponents for the Adventure Works Cycles product line. These subcomponents are shipped to the Bothell location for final product assembly. In 2001, Importadores Neptuno, became the sole manufacturer and distributor of the touring bicycle product group.

Coming off a successful fiscal year, Adventure Works Cycles is looking to broaden its market share by targeting their sales to their best customers, extending their product availability through an external Web site, and reducing their cost of sales through lower production costs.

As a bicycle manufacturing company, Adventure Works Cycles has the following four product lines:

- Bicycles that are manufactured at the Adventure Works Cycles company.
- Bicycle components that are replacement parts, such as wheels, pedals, or brake assemblies.
- Bicycle apparel that is purchased from vendors for resale to Adventure Works Cycles customers.
- Bicycle accessories that are purchased from vendors for resale to Adventure Works Cycles customers.

At Adventure Works Cycles, the purchasing department buys raw materials and parts used in the manufacture of Adventure Works Cycles bicycles. Adventure Works Cycles also purchases products for resale, such as bicycle apparel and bicycle add-ons like water bottles and pumps.

The AdventureWorks Data Warehouse is updated by an ETL process that periodically pulls data from the OLTP database and loads it into the data warehouse.

Hints regarding the data:

Granularity: A row in the `FactInternetSales` table represents the sale of one item by AdventureWorks. The item sold might be a bike, an article of clothing, a biking accessory (like a helmet), or a repair/replacement part for a bike. For this exercise, we will only look at bike sales.

For analysis of sales in `FactInternetSales`, use the column `UnitPrice` to reflect the dollar amount of sales. Use the column `OrderQuantity` to reflect the number of items sold. Use the `OrderDateKey` to determine the date of the sale.

The column `ProductStandardCost` holds the dollar amount that it cost AdventureWorks to build or obtain a bike. `UnitPrice` minus `ProductStandardCost` represents how much profit or margin AdventureWorks made on the sale.

`DimProductSubcategory.EnglishProductSubcategoryName` contains a description that identifies which type of item was sold. This columns holds a grouping of model types.

`DimProduct.ModelName` and `DimProduct.EnglishProductName` further identify more detailed product/model information regarding the item sold. `DimProduct.ModelName` holds the specific model of bike sold.

Appendix B – Creating your connection to the aw data warehouse

Steps to Connect the aw MySQL instance to your query editor.

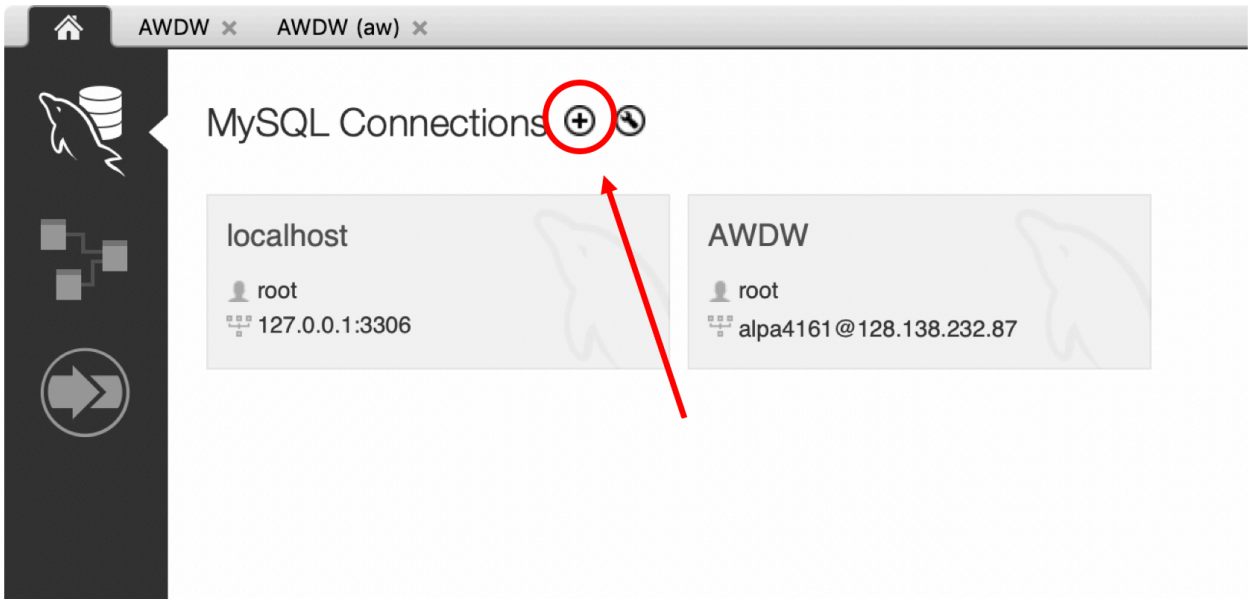
To complete this assignment, you must connect your MySQL Workbench (or query editor of your choice) to an instance of MySQL running on a VM within the CU Boulder Computer Science department. The server name is **elra-sql.** You will connect to the server using the SSH protocol. Then while connected to that server, you will establish a connection to the instance of MySQL running on that server.

If you can't connect using the DNS name, try connecting using the IP address.

Prerequisites:

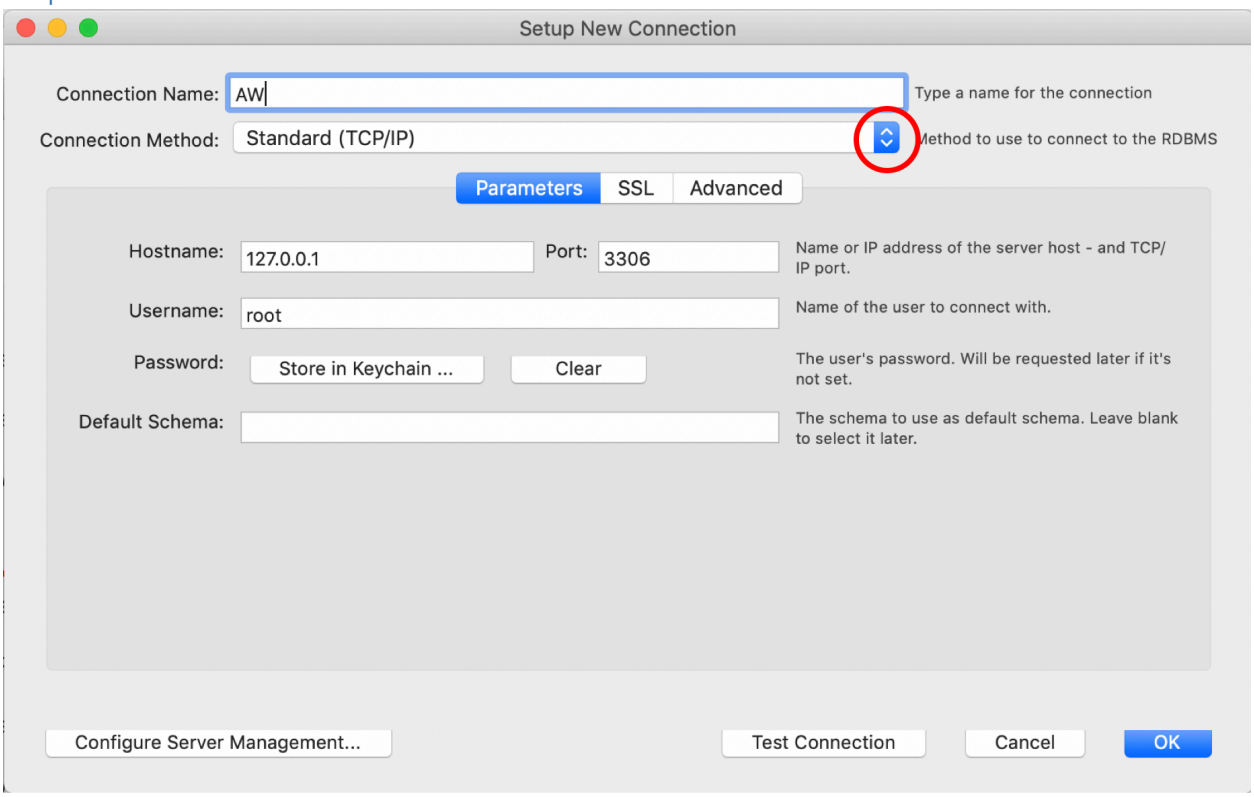
SSH Hostname:	elra-sql.cs.colorado.edu OR 128.138.232.87
SSH Username:	your CU Identikey username (ex. **niri0478**)
SSH Password:	your CU Identikey password
MySQL Username:	your CU Identikey username (ex. **niri0478**)
MySQL Password:	"password"

Step 1:



This is the first screen you will see when you open MySQL Workbench. Navigate to the upper left corner and click on "+" icon to create a new connection.

Step 2:

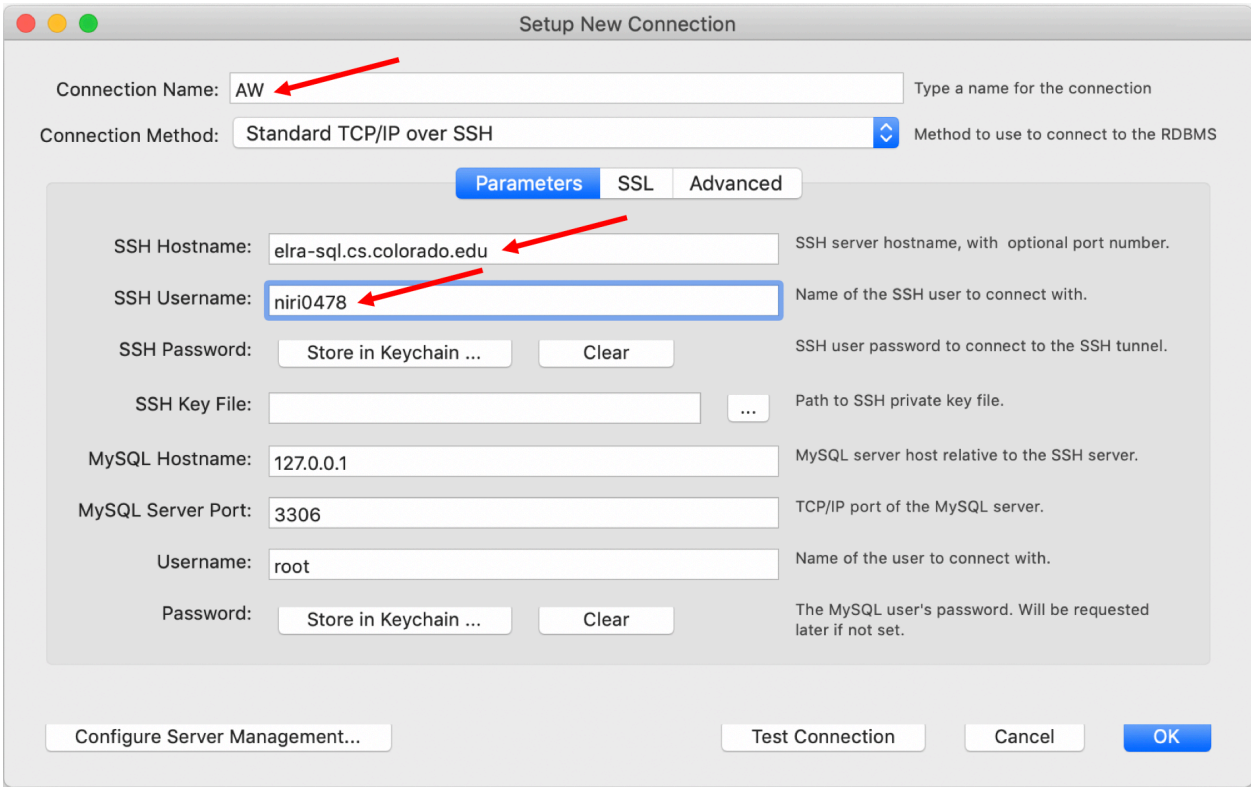


The setup new connection dialog window will pop up.

First, give your new connection a name, like “AW” or “HW5” or whatever you like.

Then navigate to the dropdown menu titled “Connection Method” and select **Standard TCP/IP over SSH.**

Step 3:

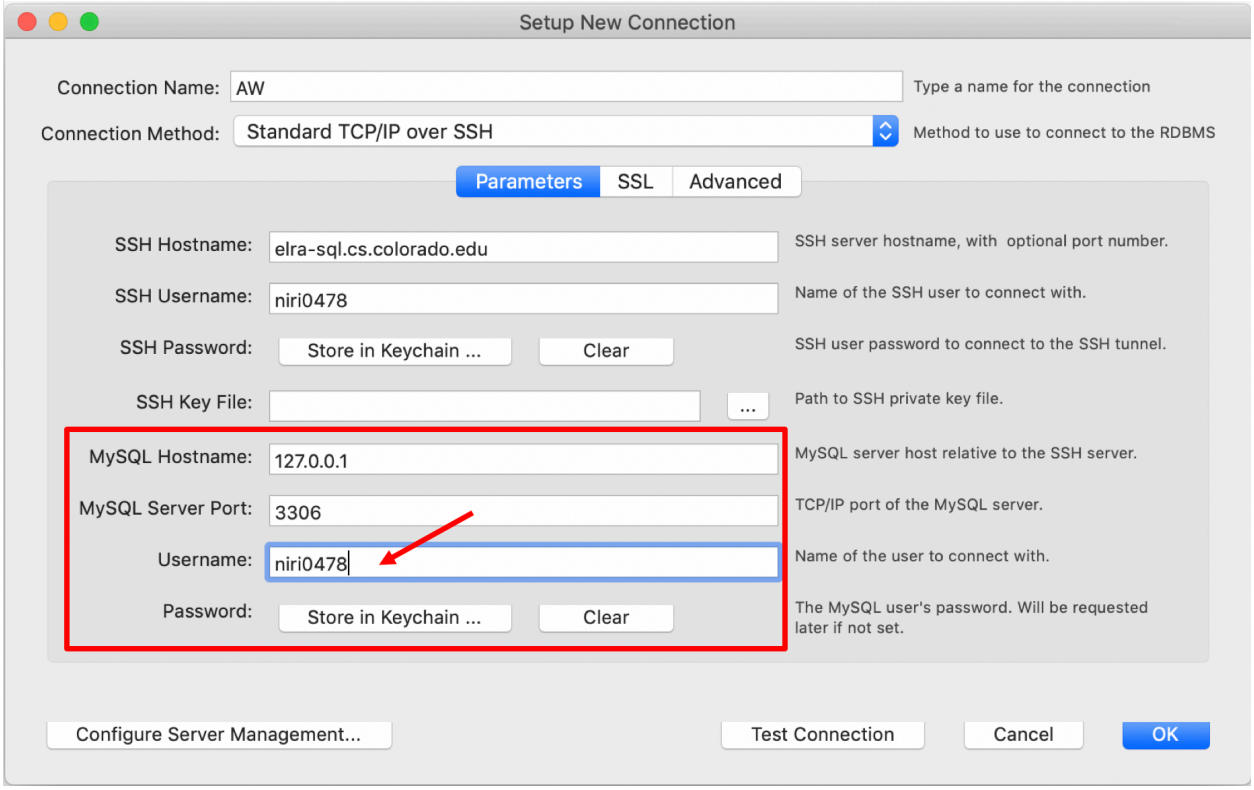


Fill in the **SSH Hostname** as **elra-sql.cs.colorado.edu** and enter your identikey as your **SSH Username**.

If you can't connect using the DNS name, try connecting using the IP address **128.138.232.87**.

Then click the button labeled "Store in Keychain" for **SSH Password** and type in your identikey password when prompted.

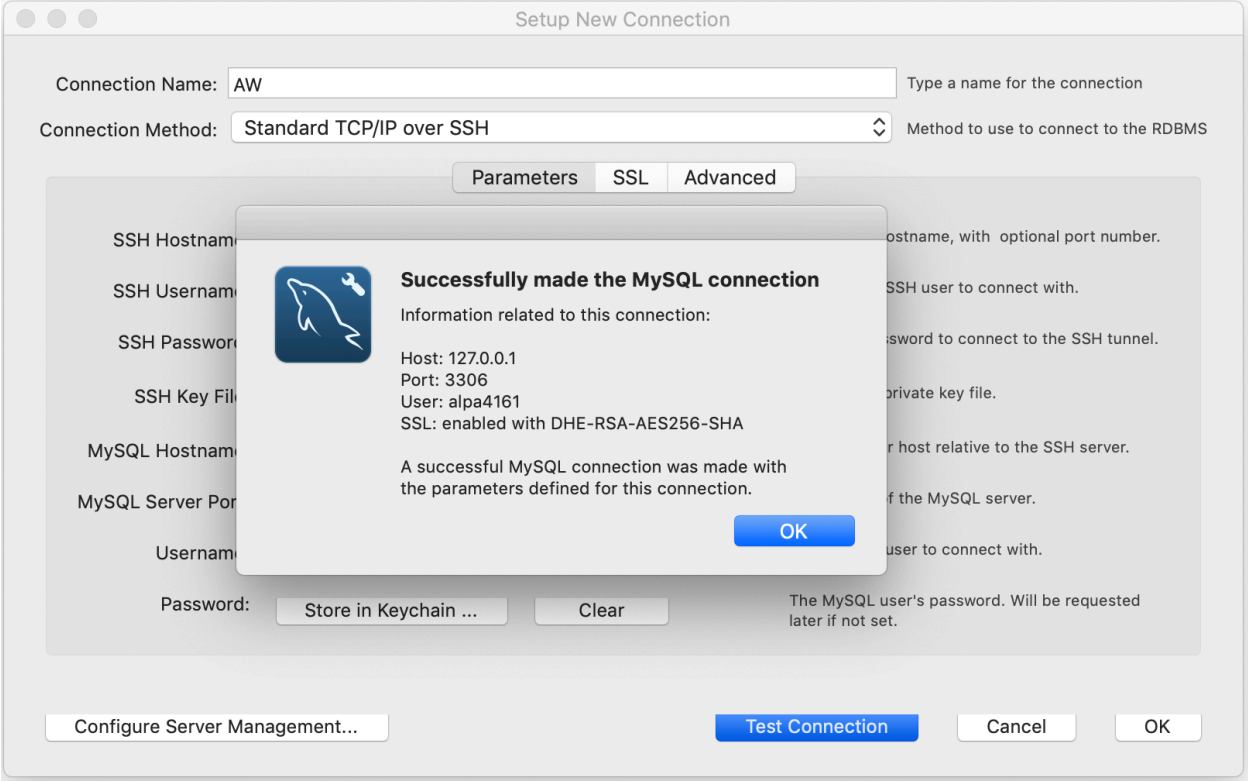
Step 4



Leave the first two fields in the red box as is. Then enter your MySQL login credentials:

Enter your CU Identikey name in the username. Then, select “Store in Keychain” and when prompted, type in the password “password”.

Step 5

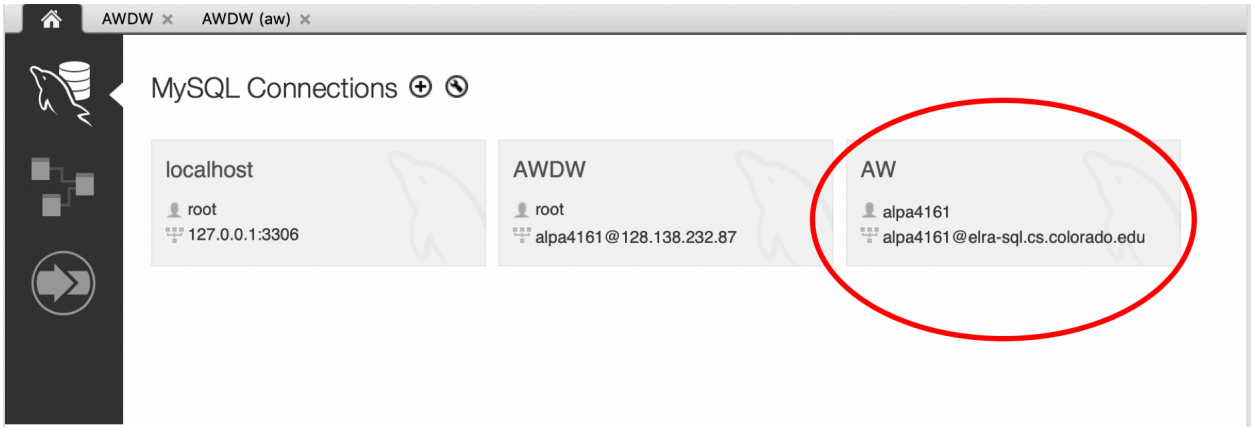


Test the connection by clicking "Test Connection". If you have followed the steps correctly the window shown above will appear.

If you get a message about the server's fingerprint, just continue.

Once the success window appears, click "OK" on the popup window, then "OK" on the configuration window.

Step 6:



Your list of connections will now display the new one you just created.

Click on the new connection to launch a MySQL Workbench query editor session against your connection to the "aw" instance of MySQL on elra-04.

Appendix C – AdventureWorks Data Warehouse Data Model