CS 163 SQL Lab

Fall 2019

August 29, 2019

Make sure to put the name of each team member on this sheet.

1   Log In

Open up a web browser and navigate to:

`http://neilklingensmith.com/phpmyadmin`

This is a web application called phpMyAdmin which is a front end to a database called MySQL. MySQL is a popular database used by lots of big tech companies. Its default user interface is command line only, so not very user friendly. The phpMyAdmin web app makes it easier to work with.

When you get to this web page, log in using the following creds:

username: `cs163`
password: `XXXXXXXX`
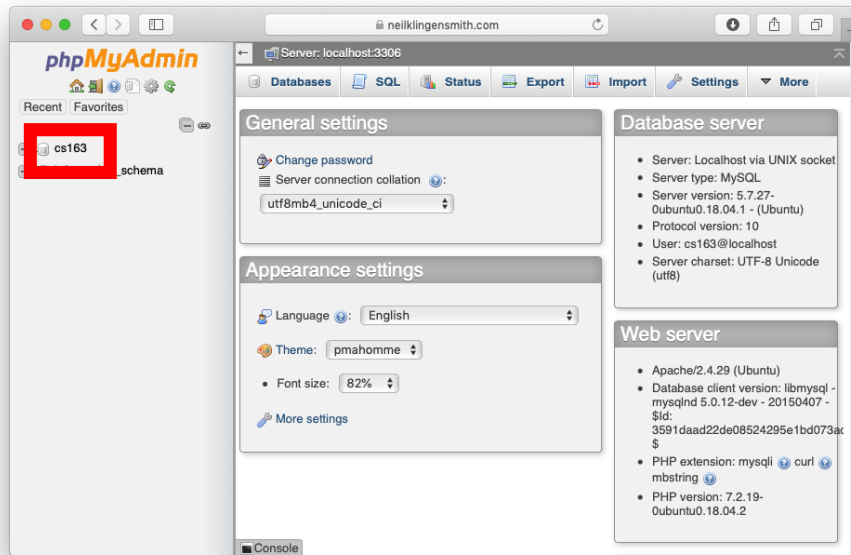
Once you're logged in, you'll see two databases on the left side of the screen. Click on the one called `cs163` to select it. This will show you the list of tables in the database—there is only one called `water`. We will be using water for this activity. If you click on the name of the table, you will see all the rows in the table.

The water table contains a bunch of data gathered from a water treatment system in a hotel. The `timestamp` column tells us the time at which each datapoint was collected. The `flow` column tells us the cumulative amount of water that has been used by the building. The `hardness` column gives a metric of the mineral content of the water.

The information contained in this table is a set. Each row is an element of the set. The two main operations people usually perform on databases are adding new data and viewing existing data. We use a simple programming language called SQL (**S**tructured **Q**uery **L**anguage) to operate on the databse. SQL is useful because it allows us to programmatically modify the data in a database. This is one reason why programmers prefer to store data in databases rather than in spreadsheets or text files.

Example Suppose we are building an industrial IoT system to monitor and control large water treatment systems, and we need to gather a lot of data about many different water treatment systems all in one place. The IoT devices themselves will be located in a bunch of different buildings. Each will have an internet connection that it will use to send the data it collects to a cloud-based server. The server will collect data from these IoT devices and record it in a database, and the data can be displayed on a web app for building managers to

2

view the status of their water treatment equipment.

In this scenario, the IoT devices are adding new data to the database and the web app is reading from the database, using the data to create charts. We can also write scripts that can access the database to automatically validate the data that is being collected.

All of these database accesses are done programmatically using SQL. We can embed SQL commands inside programs written in other languages (python, C, go, etc) so those programs have access to the database. In this lab, we are just going to issue some SQL commands to the database using the phpMyAdmin web app.

2   Issuing a SQL Query
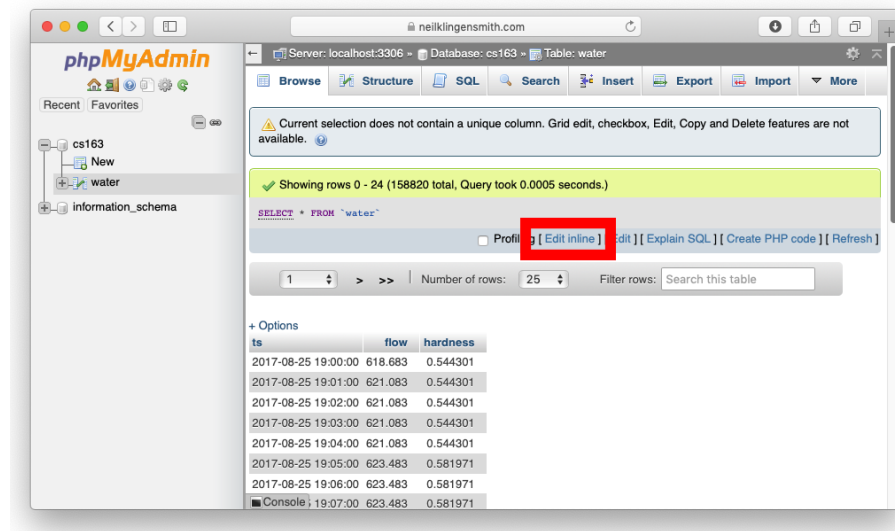
Click on the link that says *Edit Inline*. This will allow you to issue a SQL query. Delete the text that says `SELECT * FROM 'water'` and type in the following query.

```
select * from water where timestamp < '2017-08-25 19:10'
```

And press go. How many rows are returned?

Functions in SQL   We can get the same answer by using the `count()` function in SQL:

```
select count(*) from water where timestamp < '2017-08-25 19:10'
```

Enter the following SQL statement:

```
select max(hardness) from water
```

What is the max hardness value in the table?

Sorting Results   What if we want to see the three largest hardness values? We can also sort results by a particular column:

```
select flow,hardness from water order by hardness desc limit 0,3
```

What is the third largest hardness value in the table?

Cost Functions   The notion of a cost function is that there is some cost (not necessarily in dollars) associated with taking some action. In the case of a hotel water treatment system, the cost is related to allowing high volumes of mineral-rich water to flow through the plumbing system. This can cause buildup of lime on faucets and require them to be cleaned or replaced.

$$c(t) = h(t)f(t)$$

where $h(t)$ is the hardness of the water at time t and $f(t)$ is the flowrate of the water at time t. This cost function has large values when both $h(t)$ and $f(t)$ are big. If either the hardness is small or the flow rate is small, then the cost is also small. To see times when the cost function was large:

```
select timestamp,hardness*flow as cost from water order by cost
                              desc
```

What is the approximate date and time of the highest cost?

Example   Set Cardinality

What is the cardinality of the set for which flow is larger than 6500 gallons?

Are the elements of this set contiguous in time, or are they scattered around at different times?

Example: Graphing Data   We can graph data within this web interface. To make a graph of the data, enter the following SQL query in the "Edit Inline" box:

```
select * from water order by ts asc
```

4

What is this query doing?

Now scroll down to the bottom of the page and click on the link that says "Display chart" This will display a chart of the rows that were selected from your last query. By default it only displays the first 26 rows of the table. There should be a box above the chart with two text boxes in it: one for the start row, and one for the number of rows to display. Leave the start row at 0, and change the number of rows to display to 2500. Click go. Draw a picture of the chart in the box below.
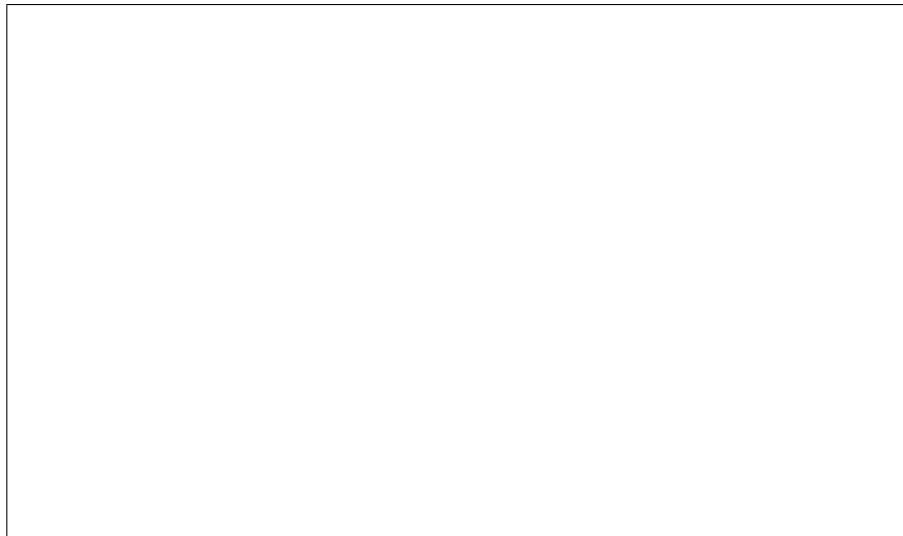
Draw a second chart also with 2500 points where the timestamp starts at `2017-08-26 16:53:00`