CMPSC 390
Data Analytics
Fall 2017

Lab 1: Laboratory Assignment One: Data Analytics Applications and BitBucket Setup.
Save this lab assignment to: `labs/lab1`

Part 1: Bitbucket Repository Set Up

Objectives

To learn how to navigate the directories within Ubuntu operating system using command line interface. To establish and configure Bitbucket repositories that will be used in this course throughout the semester; learn basic commands for downloading course materials from the instructor and for submitting material to the instructor using Git.

Reading Assignment

Please review the handout on "Tips on Using Linux and the Command Line Interface" and "Bitbucket Commands". Throughout the assignment, you should also refer to the following Web site for additional information about the Bitbucket: `https://confluence.atlassian.com/display/BITBUCKET/Bitbucket+101` and to find information about Git commands, please go to: `https://confluence.atlassian.com/stash/basic-git-commands-278071958.html` (bookmark this page).

Navigating using the Command Line Interface

A command-line interface allows the user to interact with the computer by typing in commands. Computing professionals prefer to use the command line interface, built into operating systems like Linux, instead of using the graphical user interface. In many situations command line interface tends to be very efficient and effective, for example, it allows you to complete some tasks with a simple one line command instead of using the "pumping" motion of the mouse!

1. Read through the supplemental handout on using Linux and the command line interface. As you read through the handout, follow along by trying the commands on your machine. Remember to execute a command, you should press "Enter" after typing a command. Check with your neighbors to see if they are able to open the terminal window, and use commands such as `cd, cd .., ls`, etc.

2. Open a terminal window on your workstation and create a directory called `cs390f2017/` in your home directory, by typing `mkdir cs390f2017/` command in your terminal.

3. Go to the newly created `cs390f2017/` directory. Remember, the "`cd`" command followed by the name of the directory allows you to change into a directory. Note: After you have typed in enough of the filename to specify this file from another, you can push the `tab` key to automatically complete the rest of the filename.

4. Type `ls` or `ls -l` command to list all of the items inside that directory. It should be empty, so nothing should appear after you type the `ls` command.

5. You can now close the terminal window by typing the `exit` command.

Configuring Git and Bitbucket

Practicing software developers normally use a version control system to manage most of the artifacts produced during the phases of the software development life cycle. In this course, we will always use the Git distributed version control system to manage the files associated with our class, laboratory and practical sessions. In particular, we will securely communicate with the Bitbucket.org servers that will host all of our projects. In this laboratory assignment, we will perform all of the steps to configure the accounts on the departmental servers and the Bitbucket service. As you will be required to use Git in the remaining laboratory and practical assignments and during the class sessions, please be sure to keep a record of all of the steps that you complete and the challenges that you face. You are also responsible for working with a partner to ensure that each of you is able to successfully complete each of the steps outlined in this assignment.

1. If you do not already have a Bitbucket account, please go to the Bitbucket Web site and create one—make sure that you use your `allegheny.edu` email address so that you can create an unlimited number of free Bitbucket repositories while you are a student.

2. The instructor has created the course's Bitbucket repository (repository called `cs390f2017-share/`) that will be shared with you once you have created your Bitbucket account. Once the course's repository has been shared with you, Bitbucket will give you permission to clone it to have the file stored locally. To do this, open a terminal window on your workstation and change into the directory called `cs390f2017/` that you created in the first portion of this assignment.

3. Go to your browser where you have completed your login to Bitbucket. In the middle of the screen, under "Overview" you will see a command for downloading or cloning this repository/folder starting with "git clone". Select HTTPS from the drop-down menu and then copy the text that is located in the text box. Back in the terminal, once you have changed into `cs390f2017/` directory, please paste the bash command line code into the terminal using your mouse. If everything has worked correctly, you should be able to download all of the files that you will need for this laboratory assignment. Please resolve any problems that you encountered by first reviewing the Bitbucket documentation and then discussing the matter with the instructor. If you are still not able to run `git clone`, then please ask a peer or see the instructor.

4. Using your terminal window, you should be able to browse the files that are in this shared in your `Git` repository.

Creating a New Repository

Now that you have learned how to clone an existing Git repository, you should make another (completely new) repository in the `cs390f2017/` directory that you previously created. This new directory will be used to share your work with the instructor.

If your name were *Buffalo Bill*, then your student name would be *bbill*. Then inside the `cs390f2017/` directory, create a new directory called `cs390f2017-bbill/` using the `mkdir` command in your terminal window (`mkdir cs390f2017-bbill/`. If you opened a new terminal window,

then you could type the following commands to create the needed directory; again, make sure that you understand each of these steps, discussing them with your neighbor, or the course instructors if you have a question.

```
cd cs390f2017/
mkdir cs390f2017-bbill/
cd cs390f2017-bbill/
```
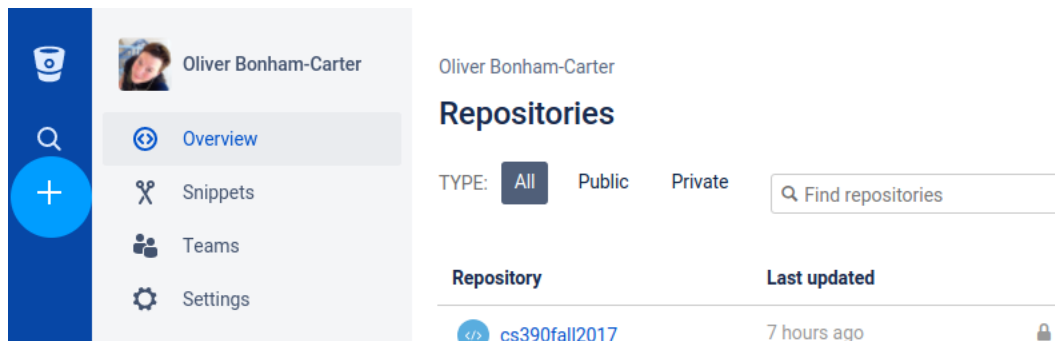

Figure 1: The button to create a repository using Bitbucket's website.

In Figure 2 you will note the screen after you have clicked on the "create" command (shown in Figure 1 on Bitbucket's website. In the blank, yuo are to enter the directory name, `cs390f2017-bbill/` (where you replace Buffalo Bill's ID with your own). Follow Bitbucket's instructions which will give you code for the building the repository locally and have it associated with Bitbucket's servers. Use the, **I am starting from scratch**-mode for this task. Next, give access privileges to the course instructor, (whose account is "oliverbc") to be able to share you work.
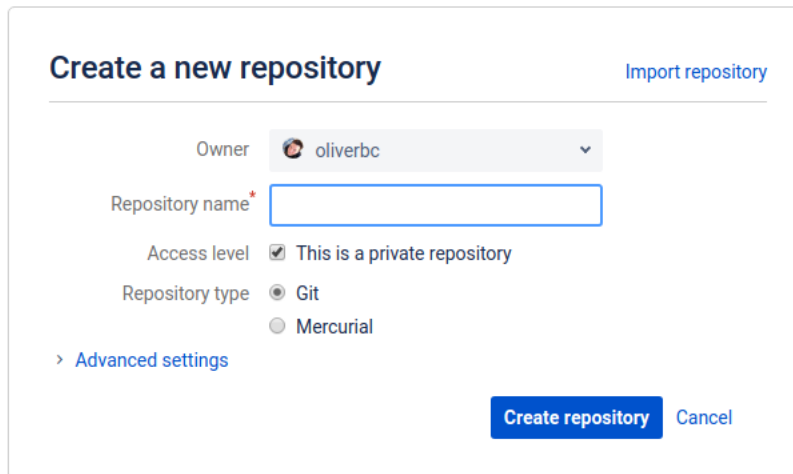
You can learn more about Git by consulting Web sites like `http://try.github.io/` and `http://gitimmersion.com/`. After discussing them with a class member and the instructor, you should ensure that you have a basic understanding of the following Git commands: `git init, git status, git commit, git push, git pull (used to pull over all updates from the server to your local repository`

Adding your updates to the Bitbucket server

Once you have copied in files for your homework submission, use the following commands to push these files up to Bitbucket's cloud.

```
git add -A
git commit -m "Your details about the pushed files: what you added and why."
git push
```

Please note: Do not resubmit the data files or others, which originated in our class-shared directory. These files will likely take up massive amounts of space on the instructor's machine and departmental servers. Please locate the hidden file, `.gitignore`, in the class shared directory to copy it into your own submission directory to prevent extraneous files from accidentally being pushed with the submissions of your work. Read over the `.gitignore` file to determine which files are not included with each push.

HANDED OUT ON: 1st SEPTEMBER 2017

Figure 2: Fill in the name of the repository that you will create and then follow the instructions for a repository made from "scratch".

Part 2: Data Case Studies

Data misuse is a very serious problem. Although a data scientist may have the means to process data for particular types of results, such a task may not always be ethically sound. For instance, some of the misuses of data may take the following forms, as shown below.

Note: The included examples were taken from:
https://www.observeit.com/blog/importance-data-misuse-prevention-and-detection/

- A recent high profile case of data misuse occurred when an employee at one of the world's fastest growing companies, Uber, violated the company's policy by using its *God View* tool to track a journalist who was late for an interview with an Uber exec. If you haven't heard, *God View* allows the company's staff to track both Uber vehicles and customers. It is not open to drivers at all, but it is apparently "widely available" at a corporate level. Tracking the journalist obviously flies in the face of even Uber's latest privacy policy, which states that employees are prohibited to look at customer rider histories except for "legitimate business purposes."

- In 2012, state auditors found that 88 police officers in departments across the state of Minnesota misused their access to personal data in the state driver's license database to look up information on girlfriends, family, friends, or others without authorization or relevance to any official investigation. Auditors said that this is not uncommon and that more than half of the police officers in the state made questionable searches in the database.

- The Florida Supreme Court heard a case in which a lower court found that Broward County, police officers misused data by conducting real-time GPS tracking on the location of a man's cellphone, using undisclosed techniques in collaboration with the cellphone carrier.

- A Chicago police officer responsible for administrating the department's criminal history database used the system to look up his girlfriend's record. Similar cases have shown up in other states, resulting in cases involving stalking, harassment, and identity theft.

- AT&T will pay $25 million to the Federal Communications Commission as a result of an investigation that discovered that employees at international call-centers illegally disclosed the personal information of upwards of 280,000 customers. The workers sold U.S. AT&T customer names and Social Security numbers to third parties who used it to unlock mobile phones so the devices would work on networks other than AT&T's, said Wednesday in a news release by the Federal Communications Commission.

- ... And plenty of other examples exist.

Your Task For Part 2

As part of an introductory lab exercise, you will research a similar example of the misuse of data to those from above. Your task is to find an article from a reputable news source (Washington Post, New York Times, Los Angeles Times, and etc) where some form of crime has been committed due to the misuse of data. **Write up four main points of argument to explain why the scenario of your news article is in violation of some public trust.** You are to submit a page-long response (via Bitbucket) to the instructor where your four reasons are clearly displayed and supported by a short discussion. Your submission should be one page.

Important Details

Lab directory structure: You are to create a labs directory (`mkdir labs` in which you are to add directories for each of your weekly labs (use this command`mkdir labs/labsx`, where x is the lab number). For example, your first and second labs should be located in the paths, `labs/lab1` and `labs/lab2`, respectively.

   Note: Please remember to include your name on everything you submit for the class. Although the instructor collects your work from Bitbucket, each work must be graded outside of the Bitbucket directory and so without adding your name, the instructor will be unable to award you credit for your work.

Required Deliverables

This portion of the assignment invites you to submit an electronic version of the following deliverable through your Bitbucket repository (`cs390f2017-bbill`.)

1. A document fulfilling the part 2 requirements placed into a directory called `lab1`.

2. Share your assignment files with the instructor through your Git repository by correctly using using appropriate Git commands, such as `git add`, `git commit -m ''your message''` and `git push` to send your screen shot images to the Bitbucket's server. When you have finished, please ensure that the Bitbucket Web site has your pushed work. Please see the instructor if you have any questions about assignment submission.