

**CMPSC 380**  
**Principles of Database Systems**  
**Fall 2014**

**Laboratory Assignment One: Introducing Data Management Concepts and Tools**

## Introduction

Practicing data management professionals use a wide variety of tools to manage data sets. When the task at hand requires the management of files, many people will use a version control system like Git. In this course, we will always use the Git distributed version control system to manage the files associated with our laboratory assignments and the final project. In this laboratory assignment, you will learn how to use the Bitbucket service for managing Git repositories and the `git` command-line tool in the Ubuntu operating system. After creating, populating, and regularly using your own repository and sharing it with the course instructor, you will use a text editor to prepare a brief report on both the features that data management tools should, and currently do, provide.

## Configuring Git and Bitbucket

During this laboratory assignment and subsequent assignments, 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. Throughout the assignment, you should refer to the following Web site for additional information: <https://confluence.atlassian.com/display/BITBUCKET/Bitbucket+101>. As you will be required to turn in a report describing each step in this assignment, please be sure to keep a record of all of the tasks that you complete and the challenges that you face.

1. If you have never done so before, you must use the `ssh-keygen` program to create secure-shell keys that support your communication with the Bitbucket servers. Type `man ssh-keygen` and talk, at a high level, with the members of the class to learn more about how to use this program. What files does `ssh-keygen` produce? Where does this program store these files?
2. 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. Then, upload your ssh key to Bitbucket.
3. At this point, you can learn more about Git both by discussing it, at a high level, with your colleagues and consulting Web sites like <http://try.github.io/> and <http://gitimmersion.com/>. You should ensure that you fully understand how to use the following Git commands:
  - (a) `git init`
  - (b) `git status`
  - (c) `git add`
  - (d) `git commit`
  - (e) `git push`
  - (f) `git pull`

## Creating and Populating a New Git Repository

Now that you have learned more about using Git, you should make a new repository in the `cs380F2014/` directory that you previously created. First, make a new directory called `cs380F2014-<your user name>`. Now, you can use the `git init` command to turn this directory into a local Git repository. After completing this step, you should make a `lab1/` directory and then create a plain-text file (e.g., a `.txt`, `.md`, or `.tex` file) with a name containing the word “features”. For instance, if you choose to write this report in  $\text{\LaTeX}$ , then you would name your file `features.tex`. Finally, you must create an additional plain-text file with a name including the word “tools”. Again, if you decide to write this report in  $\text{\LaTeX}$ , then you could create a file called `tools.tex`.

Next, you should use the Bitbucket Web site to create a repository that has the same name as the local directory and local repository (i.e., `cs380F2014-<your user name>`). You must follow Bitbucket’s instructions to push the code and tags in your local repository to the remote one. When you are finished with this step, you should see in your Web browser that the Bitbucket servers are storing the two text files. Once the Git repository contains the correct files, you should share your Bitbucket repository with the course instructor, whose Bitbucket user name is “gkapfham”.

## Investigating Data Management Tools

Using your “features” file in the Git repository, please write a short one-page document that explains the features that you think data management tools should provide. In advance of listing and explaining these features, your document should clearly define the term “data management tool”. Whenever possible, you should rank the features in the order of their importance.

In the “tools” file, you should prepare a comprehensive listing of ten free and open-source tools that provide data management facilities. This report should give the name of the tool, the Web site(s) and/or papers that you referenced to learn more about it, and a detailed description of the features that it provides. As you explain each tool, you should comment on whether or not it furnishes any of the features that you mentioned in your “features” document. Whenever possible, the report should comment on the type of data that the tool aims to manage.

When writing these two documents, you must adhere to the Honor Code statement as articulated in the syllabus. In particular, you must take care to ensure that you properly cite your sources and that you use your own words to explain both the desired features and the ten tools.

## Summary of the Required Deliverables

This assignment invites you to submit one printed version of the following deliverables:

1. A description of the steps that a user must take to configure Git and Bitbucket.
2. A description of the inputs, outputs, and behavior of the six aforementioned Git commands.
3. A complete description of the features that data management tools should provide.
4. A comprehensive listing of ten free and open-source data management tools.

Before you turn in this assignment, you also must ensure that the course instructor has read access to your Bitbucket repository that is named according to the convention `cs380F2014-<your user name>`. Please see the instructor if you have any questions about this assignment.