

CMPSC 380
Principles of Database Systems
Fall 2016

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 create, populate, and share your own repository with the course instructor. Specifically, you will use a text editor to prepare a brief report on both the features that data management tools should, and currently do, provide.

Creating and Populating a New Git Repository

First, you should create `cs380F2016/` directory that will store the course repository shared by the instructor and your own repository for the course. Now, clone the course repository titled `cs380F2016-share` into your `cs380F2016` directory. Next, make a new directory called `cs380F2016-<your user name>` inside your newly created `cs380F2016/` directory. 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 \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 \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., `cs380F2016-<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 “Oliver Bonham-Carter (oliverbc)”.

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 an electronic version of the following deliverables submitted via your Bitbucket repository:

1. A complete description of the features that data management tools should provide.
2. 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 `cs380F2016-<your user name>`. Please also be sure to *push* your repository out so that the instructor can collect your work. Please see the instructor if you have any questions about this assignment.