CMPSC 312
Data Analytics
Spring 2020

Participation 1:
Setting Up Docker, Git and Atom, and An Analysis of Google Analyics

Figure 1: Maybe we cannot hack into a typewriter but we can still hack the text that typewriters have produced. For this type of hacking, we collect frequency information to determine the distribution of frequencies.

Summary

To learn how to set up a container-based platform, called Docker, for use in all class activities, labs and practicals. To configure Github and Github Classroom and a text editor called, Atom, for use in the class. Finally, you will be writing a page-long reflection piece about the challenges and solutions you encountered during the installation procedures.

1. Using a Container-based Platform, Docker

Figure 2: Docker

Docker is a platform for developers and system administrators to develop, deploy, and run software applications with containers. Docker is a new generation of virtualisation and it is revolutionizing the IT industry. Using Docker software developers no longer need to worry about installations and software versions on various operating systems their product may need to run on. All the necessary software is packaged into a container and software applications are run in it. Companies, such as Visa and PayPal, and universities, such as Cornell and Indiana University, have adopted the Docker platform since it simplifies the process of building, running, managing and distributing software applications.

The Department of Computer Science has adopted Docker in all computer science courses. The instructor in this course will deploy Docker containers with all the necessary software to run all class exercises, lab and practical assignments. Students will then be able to work on developing programs on their own machines and run them using instructor-provided Docker container without having to install a dozen of software languages and tools. Before being able to use Docker, students must first complete basic installations of Docker and Git, which will be used for code distribution and submission. These installations are specific to the operating system of your machine. As the first step of the Docker set-up process, please check the departmental website shown below to determine the possibility of running Docker on your machine.

`https://www.cs.allegheny.edu/canirundocker/`

We note here that there are some machines for which the "*Yes!*" or "*No!*" response from the link is inaccurate. Also, please make a note of the version of your operating system and check whether your laptop meets the requirements outlined below. If you are convinced that your machine will be able to run Docker Desktop, then proceed to the next step. If, however, you note a "*No!*" message when you clicked on the above link, or are a Windows 10 "Home" user, then you will not be able to run Docker Desktop. Unfortunately, the Department of Computer Science does not support Docker ToolBox.

Docker Desktop Specifications
Mac:

- 2010 model or newer with hardware support for MMU, EPT, and Unrestricted Mode
- macOS Sierra 10.12 or newer
- 4 GB of RAM
- VirtualBox prior to version 4.3.30 cannot be installed

Windows:

- Windows 10 64bit Pro, Enterprise, or Education (Build 15063 or later)
- Virtualization is enabled in BIOS
- CPU SLAT-capable feature
- 4 GB of RAM

Linux:

- 64bit
- Kernel 3.10 or later

Now, go to `https://docs.docker.com/install/` and from the menu on the left-hand side select and follow the installation tutorial for your operating system (Linux, MacOS, or Windows). Commonly used links for popular operating systems are offered below.

- MacOS: `https://docs.docker.com/docker-for-mac/`
- Binary Files for all OS's
  - Instructions: `https://docs.docker.com/install/linux/docker-ce/binaries/`
  - Files : `https://download.docker.com/`
  - Windows: `https://docs.docker.com/docker-for-windows/install/`
  - Ubuntu: `https://docs.docker.com/install/linux/docker-ce/ubuntu/`

Please note that the Docker Desktop for Mac and Windows requires you to create an account to download. To bypass this, use the following download links:

- Windows: `https://download.docker.com/win/stable/Docker%20for%20Windows%20Installer.exe`
- Mac: `https://download.docker.com/mac/stable/Docker.dmg`

In this class we will use a command-line interface, which allows the user to interact with the computer by typing in commands in a "terminal" window. 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-liner* command instead of dealing with lots of clicking on a gui.

Once your docker installation is complete, please locate a terminal window on your machine. In MacOS and Linux, you can search for a "terminal" and on Windows machines look for "cmd" (command prompt). If you are using Windows, we recommend using a Linux-based terminal, a step that you will complete when installing Git in Section 2, step 2.

Now in the terminal window type `docker run hello-world` command and press "Enter". If you see an output similar to the one below, feel free to high five a technical leader - your docker set up was successful!

```
docker run hello-world

docker : Unable to find image 'hello-world:latest' locally
...

latest:
Pulling from library/hello-world
ca4f61b1923c:
Pulling fs layer
ca4f61b1923c:
Download complete
```

```
ca4f61b1923c:
Pull complete
Digest: sha256:97ce6fa4b6cdc0790cda65fe7290b74cfebd9fa0c9b8c38e979330d547d22ce1
Status: Downloaded newer image for hello-world:latest

Hello from Docker!
This message shows that your installation appears to be working correctly.
...
```

Configuring Git and GitHub

Figure 3: Git and GitHub

During the subsequent practical and laboratory assignments, we will securely communicate with the GitHub servers that will host all of the project templates and your submitted deliverables. In this assignment, you will perform all of the steps to download Git and configure your account on GitHub, so that you are ready to start your first lab assignment using GitHub Classroom next week. You can also learn more about GitHub Classroom by visiting `https://classroom.github.com/`. As you will be required to use Git, an industry standard tool, in all of the laboratory and remaining practical assignments and during the class sessions, you should keep a record of all of the steps that you complete and the challenges that you face. You may see the course instructor or one of the teaching assistants if you are not able to complete a certain step or if you are not sure how to proceed.

1. If you do not already have a GitHub account, then please go to the GitHub website (`https://github.com/`) and create one, making sure that you use your "allegheny.edu" email address so that you can join GitHub as a student at an accredited educational institution. You are also encouraged to sign up for GitHub's "Student Developer Pack" at `https://education.github.com/pack`, qualifying you to receive free software development tools. Additionally, please add a description of yourself and an appropriate professional photograph to your GitHub profile. Unless your username is taken, you should also pick your GitHub username to be the same as Allegheny's Google-based email account.

2. If you have not done so already in the previous section (1), install Git for the operating system of your laptop. Please note that if you installed Docker Toolbox and checked the "Git for Windows", you do not need to install Git again. Please see suggestions below relevant to your operating system.

   For Windows: We recommend you complete the following steps.

- Open a terminal (command prompt) window in administrator mode. To do this, search for "cmd" in the start menu, then right-click and select "Run as Administrator".
- Then, copy and paste the following command into that terminal, and hit enter.

  ```
  $ @"%SystemRoot%\System32\WindowsPowerShell\v1.0\powershell.exe"
  -NoProfile -InputFormat None -ExecutionPolicy Bypass -Command
  "iex ((New-Object System.Net.WebClient).DownloadString
  ('https://chocolatey.org/install.ps1'))" &&
  SET "PATH=%PATH%;%ALLUSERSPROFILE%\chocolatey\bin"
  ```

- To test that this step was completed successfully, open a new terminal, and type the command "choco −−version". If 0.10.11 is printed (or some other higher version), you have completed this step. If not, ask for the instructor or a technical leader for assistance.
- Run the following commands in a terminal window in administrator mode.
  `choco install git -y --params "/GitAndUnixToolsOnPath /WindowsTerminal /NoShellIntegration"`
- To test that this step was completed successfully, open a new terminal and type:
  `git --version`.
  If git version followed by some version number is printed (or something similar), you have completed this step.

For Mac: In the terminal type: `git --version`. If Git is not installed on your machine already, then you will be prompted to install it with the Git installer.

For Linux:

- Type and run the following command in the terminal:
  `sudo apt-get install git`

3. If you have never done so before, you must use the "ssh−keygen" program to create secure-shell keys that you can use to support your communication with GitHub. These keys enable you to send your files to GitHub without having to type in a password each time. But, to start, this task requires you to type commands in a terminal.

4. Open the terminal, a window in which you can type commands, and now type the "ssh−keygen −t rsa −b 4096 −C your_email_used_to_create_github_account@allegheny.edu" command in it. Follow the prompts to create your keys and save them in the default directory. That is, you should press "Enter" after you are prompted to "Enter file in which to save the key..." and then type your selected passphrase whenever you are prompted to do so. Please note that a "passphrase" is like a password that you will type when you need to prove your identify to GitHub. What files does "ssh−keygen" produce? Where does this program store these files by default? Invite a technical leader to investigate the answers to these questions.

5. Once you have created your ssh keys, you can raise your hand to invite either a technical leader or the course instructor to help you with the next steps as needed. First, you must

log into GitHub and look in the right corner for an account avatar with a down arrow. Click on this link and then select the "Settings" option. Now, scroll down until you find the "SSH and GPG keys" label on the left, click to create a "New SSH key", and then upload your ssh key to GitHub. You can copy your SSH key to the clipboard by going to the terminal and typing "`cat ~/.ssh/id_rsa.pub`" command and then highlighting this output. When you are completing this step in your terminal window, please make sure that you only highlight the letters and numbers in your key—if you highlight any extra symbols or spaces then this step may not work correctly. Then, paste this into the GitHub text field in your web browser.

6. Again, when you are completing these steps, please make sure that you take careful notes about the inputs, outputs, and behavior of each command. If there is something that you do not understand, then please ask the course instructor or the teaching assistant about it.

3. Text Editor for Programming

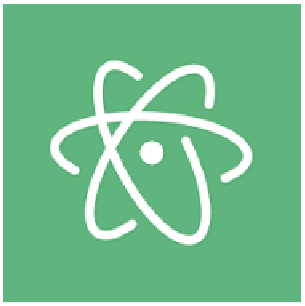


Figure 4: Atom

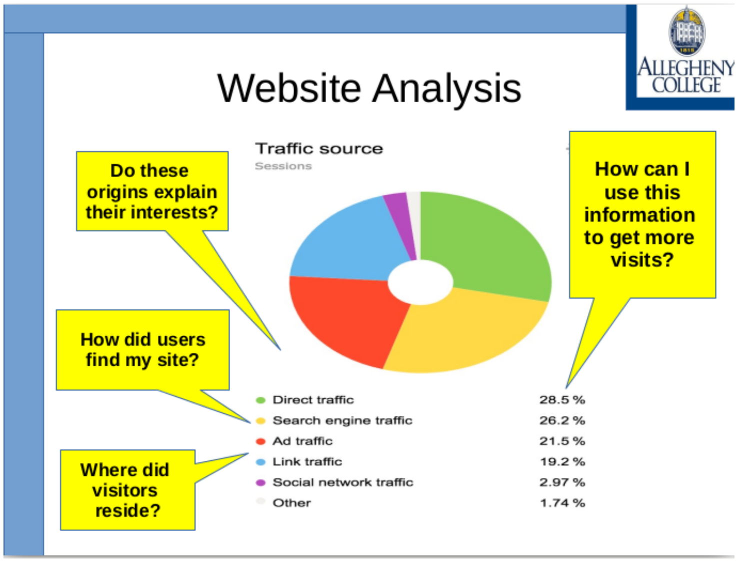
In this class we will use a text editor called, `atom`, to write and edit software. Please go to `https://atom.io`. You should see a Download button specific to your operating system. Use this link to install this editor onto your machine.

- **Experiment!** Practical sessions are for learning by doing without the pressure of grades or "right/wrong" answers. So try things! The best way to learn is by trying things out.
- **Complete *Something*.** Your grade for this assignment is a "checkmark" indicating whether you did or did not complete the work.
- **Practice Key Laboratory Skills.** As you are completing this assignment, practice using the Ubuntu terminal until you can easily use its most important features. Additionally, ask a teaching assistant or a course instructor to teach you some of the advanced features of the terminal, thereby helping you to work more effectively.
- **Try to Finish During the Class Session.** Practical exercises are not intended to be the equal of the laboratory assignments. If you are simply a slow typist, I've given you until the end of the day, but ideally you should complete the assignment by the end of the class period.

- **Help One Another!** Your instructor will likely be helping other members of your class. If you need help, please ask your peers who are sitting next to you; it is very likely that they have already dealt with the problem that you may be experiencing. The Technical Leaders (TL's) are also available for answering questions and their office hours can be found at the following link; `https://www.cs.allegheny.edu/teaching/technicalleaders/`.

  If your neighbor is struggling and you know what to do, offer your help. Please do not "do the work" for them, but advise them on what to type or how to handle things. If you are stuck on a part of this practical session and you could not find any insights in either your textbook or online sources, formulate an intelligent question to ask your neighbor, a teaching assistant, or a course instructor. Try to strike the right balance between asking for help when you cannot solve a problem and working independently to find a solution.

4. An *Analysis* of Google Analytics



Website Analysis

Do these origins explain their interests?

How did users find my site?

Where did visitors reside?

How can I use this information to get more visits?

GitHub Starter Link

https://classroom.github.com/a/SgQ6EJtN

To use this link, please follow the steps below.

- Click on the link and accept the assignment.
- Once the importing task has completed, click on the created assignment link which will take you to your newly created GitHub repository for this lab.
- Clone this repository (bearing your name) and work on the practical locally.
- As you are working on your practical, you are to commit and push regularly. You can use the following commands to add a single file, you must be in the directory where the file is located (or add the path to the file in the command):
  - `git commit <nameOfFile> -m ''Your notes about commit here''`
  - `git push`

  Alternatively, you can use the following commands to add multiple files from your repository:
  - `git add -A`
  - `git commit -m ''Your notes about commit here''`
  - `git push`

Summary

We have been studying Google Analytics by setting-up a basic website to which we connected our Google Analytics account. Using the Google Analytics dashboard, we have access to plots, statistics and other types of statistical information about the traffic to our site, as well as, the activity that is generated by the visitors.

In this participation work, you are to spend time to get to know your Analytics dashboard plots, statistics and tools. After you have spent some time to play with these resources, please answer some questions about your discoveries, see the questions in blue below. You will be working out of a GitHub Classroom repository that you will create by clicking on the starter link, above. This repository will provide you with your working files and you will use the repository to submit your work (see above commands for using Git) for a check-mark grade. This check-mark in the grade contributes to your overall grade in the course.

GatorGrader

We will be using GatorGrader in our course to help determine that the objectives of the participation, labs and other assignments have been correctly completed. Please see the `README.md` in the root directory of your GitHub Classroom repository. Important: GatorGrader can only work once Docker Desktop has been correctly installed on your laptop. Please ask questions as necessary.

Command to launch GatorGrader

With Docker Desktop installed, you will run the following code from your terminal or command prompt. The command is to be run from within the root of the GitHub Classroom repository and will function to check the that the objectives of your assignment have been achieved.

First, to ensure that the following command will work correctly, you must create the cache directory by running the command `mkdir $HOME/.dockagator` and push ENTER. Then, to see if your submission satisfies the minimal requirements, you can run the following command in the terminal:

```
bash
docker run --rm --name dockagator \
  -v "$(pwd)":/project \
  -v "$HOME/.dockagator":/root/.local/share \
  gatoreducator/dockagator
```

Please note, you may not need the word `bash` in your command and so if the above command does not run, try using the same command without the `bash` in it. Please see your `README.md` file for more details on this tool.

Questions

Below are the same questions that you will find in your reflection file in your working repository. These questions are designed to be thought-provoking and may not necessarily have one correct answer concerning Google Analytics and its dashboard, displaying its plots and information. Each question can be answered in a few sentences.

1. What types of information is given in the *Realtime* tab in Google Analytics?
2. Can you name four main types of information from this tab that could be used to get to know who your site's visitors are?
3. In the *Realtime* tab, explain the definitions and uses of the following items; *Pageviews*, *Top Active Pages* and *Top Referrals*.
4. What types of options exist in the *Audience* tab? Discuss the types of studies that can be made using this information.
5. What type of information can be found in the *Audience* → *UsersFlow*?
6. Describe what type of information can be found in the *Audience* → *Demographics* menu item. In a hypothetical situation, how could this information be helpful to a business that sells its products online?
7. In the *Audience* → *Behavior* → *New vs Returning* item, there are some plots and statistics available. Explain what this information is and how it could be used in the context of a social website whose goal is get visitors to read an extensive number of blog entries.
8. Describe how you could use the dashboard to investigate the geographic origins of the users of the web site.
9. Now that you have spent some time investigating the *Realtime* and *audience* tabs, explain why there cannot be just *one* (single) plot or *one* (single) major statistic that could accurately describe all there is to know about the traffic to web page.
10. **Ethical question**: Describe types of major decisions that could be made by only looking at the information available from Google Analytics, which could have unethical implications in terms of humanitarian or environmental interests. Please consider this answer carefully and be creative in the types of decisions which can be made in this informational arena.

Summery of Deliverables

1. Be sure that each of the above listed programs is has been installed, correctly configures and is able to run on your laptop.
2. Your completed reflection document (`writing/Reflection.md`) where you respond to each of the ten questions in blue (listed above and in your writing document). Note, please edit this document by removing the `TODO` tags for each question as you give your response. Be sure to push in your work by the due data which is shown in the header of the assignment sheet.