# Introduction

1. You will need sample data for this lab.
2. In order to save time, I have generated the necessary data files for you. They are available for download on Blackboard.
3. However, if you want you can generate the files yourself. Please see the following:\
   1. The simplest method for this is to run the python script supplied with the text. You will need to install python on your windows machine. Python comes standard on Ubuntu Linux. Get **Python 2.7** for Windows at the following URL: <https://www.python.org/downloads/release/python-2711/>
   2. Download the data package from Blackboard at this URL: <https://blackboard.temple.edu/bbcswebdav/pid-4945796-dt-content-rid-51276613_2/xid-51276613_2>
   3. Unzip the file to a convenient location.
      1. Note: If you are using Ubuntu Linux you will need to have a writeable file system for the generated files. The default location in the script are owned by the user Splunk and the user root.
      2. Carefully edit the location for these files by editing **line 66** in the generate script:

files = [ "/opt/log/BigDBBook-www1/access.log", "/opt/log/BigDBBook-www2/access.log", "/opt/log/BigDBBook-www3/access.log"]

* + 1. Change **line 66** to this:

files = [ "/users/[YOURLOGINNAME]/opt/log/BigDBBook-www1/access.log", "/users/[YOURLOGINNAME]/opt/log/BigDBBook-www2/access.log", "/users/[YOURLOGINNAME]/opt/log/BigDBBook-www3/access.log"]

* + 1. For example, change **line 66** to:

files = [ "/users/eferrara/opt/log/BigDBBook-www1/access.log", "/users/eferrara/opt/log/BigDBBook-www2/access.log", "/users/eferrara/opt/log/BigDBBook-www3/access.log"]

* + 1. You will import the necessary data files from this location.
  1. Ensure that you have a working version of python on your Ubuntu or Windows system.
  2. Open a terminal session
     1. On Windows this is **cmd**.
     2. In Ubuntu this is **term**.
  3. Now run the script
     1. On Windows at the C> prompt (or whatever your prompt is) in the directory where you copied the script:

C> Generate\_Apache\_logs.py

* + 1. On Ubuntu

$ ./Generate\_Apache\_logs.py

* 1. Running these commands will generate the data necessary to do the rest of the tasks described in Chapter 2 of the text.

1. Loading Data
   1. Now that you have sources of data – please load the data into Splunk
   2. Follow the instructions on page 19 of the text:
2. On successful execution of the Python script, you will be able find the generated log files in the /opt/log directory if it is a Unix operating system and in case of a Windows operating system the files are placed in the directory c:\opt\log.
3. As we mentioned earlier, a typical ecommerce web site runs on various servers, thus our data generator simulates log entries for three different hosts that are named BigDBook-www1, BigDBook-www2, and BigDBook-www3.
4. The location of generated files can be customized in the script file. For this chapter the log files were generated in the following directories: /opt/log/BigDBook-www1/access.log /opt/log/BigDBook-www2/access.log /opt/log/BigDBook-www3/access.log.
5. Type the following Unix command **ls –Rla**in the /opt/log directory (or whatever directory you created/copied the log files)
6. This command will list the files as seen in Figure 2-9 of the text and we can see that the access log files were created for the three different hosts in separate directories. The same command in windows is C> dir /r

**Do this on your system and submit the results.**

# 

# File & Directories

Follow these steps to load data into Splunk.

1. You can load data from a static file as a one-time operation, also known as a one shot, or you can ask Splunk to monitor a set of directories for certain types of files.
2. We start by loading a single file. In order to make this easy we have generated an access log for MyGizmoStore.com that has approximately 250 log entries, which represent user activity over a period of two days in the life of the store.
3. The file access.log is part of the download package of the book. – See the section above.
4. Splunk will give you the option of adding data based on the type or the source of the data. For this initial example, we will work with a source, the access log file. Once you have logged into the Splunk instance, go to the Splunk home page and click on “Add data” button in the “Do more with Splunk” section.
5. In the Add Data to Splunk page you will see different options are available under two categories.
6. Choose a Data Type—allows to select a pre-determined type of logs such as access logs, sys logs etc. Choose a Data Source—allows to bind determined type of sources such as windows registry data or get an output from a script which will be a data input into Splunk.
7. Click on the “From files and directories” link under “Choose a Data Source” section. The difference between this option and a similar option in the “Choose a Data Type” section is that we get an additional option to make use of a forwarder to send the data to a Splunk Server in the data type s category.



1. This will take you into the Data preview page, as seen in Figure 2-2. Because this is the first time that we are loading the data into Splunk, it makes sense to get a preview of the data.
2. The preview option provides an opportunity to see how the log entries are going to get processed before they are committed into the Splunk data store, which is called index.
3. Next you select the “Preview data before indexing” radio button and choose the access.log file under \opt directory if you are on Linux or C:\opt if you are in Windows environment.  Then click on the Continue button.



1. The dialog box that comes up gives you options to set the source type. By default Splunk parses the data files provided as input and determines what the source type for the file is. It then loads or indexes the file according to the source type characteristics. In this case, because we are loading an access log file, which has been identified as a combined access log file, we accept the default option of “Use auto-detected source type” as seen in Figure 2-3 and click the Continue button.



1. **Once you have the data loaded – preview the data and submit your results.** Your results should look something like this:

