

Those of you who work with Windows operating system should download Cygwin. A very comprehensive deck with detail by detail instruction on how to install Cygwin is attached, courtesy of Diane Howard. As you follow installation instructions please make sure that you include OpenSSH tools. Cygwin download presents you with a list of Unix (Linux) packages. Under Net packages, look for openssh and then toggle the selection to install. OpenSSH will provide you with both SSH and SCP (secure copy) tools. Cygwin creates a Linux environment on your PC. Mac and Linux users have all those tools already installed.

Please register for the Free Tier of Amazon AWS. Please use Micro Instances only. As soon as you are done with any of your homework tasks, stop your instance and terminate it. In that way you will accumulate no or only minimal expenses.

To those of you who have taken Cloud Computing course, please do the first two problems. They are trivial.

Problem 1. Create an S3 bucket. Enable versioning. Upload a single text document. A few minutes later modify the document a tiny bit. Upload modified document.

Demonstrate that you can retrieve both versions of the document. Give permission to everyone to see your bucket. Upload an image. Demonstrate that you can observe that image in your browser by providing a proper URL.

Problem 2. In AWS Management Console, please navigate to EC2 Service (Dashboard). In the top right corner, next to your name, please select the region that is closest to you.

That is not all that essential for this simple problem but is a good practice. Launch a Linux instance. If you insist and you know what you are doing, you are welcome to launch a Windows instance as well. Otherwise, please launch **Amazon Linux AMI 2013.09.2** - ami-bba18dd2 (64-bit) / ami-d7a18dbe (32-bit). Demonstrate that you can use `ssh` protocol to connect to your instance. Demonstrate that you can use `scp` (secure copy protocol) to copy files from your local machine to the instance in the Cloud as well as copy a file from the instance in the Cloud to your local machine. Use Linux command `man scp` to read Linux manual on command `scp`.

Problem 3. Please write a small R program that will estimate the value of π (pi) using Monte Carlo technique discussed in class. Please estimate how many random points you need to generate in order to obtain standard 2 decimal points precision ($\pi \approx 3.14$).

Problem 4. Use any language of your choice to implement a map function, a reduce function and a master program for a simple word count problem as described on slides 45-48 of lecture notes. Take a single page of a text, treat imaginary line numbers (1, 2, 3 . . . 50) as the initial keys and the content of the corresponding lines as the values. Demonstrate that you are getting proper word counts. You can use Linux (Unix) `egrep` and `wc` commands to verify results of your program. Run your code on your local machine, not on a Hadoop cluster.

SUBMISSION INSTRUCTIONS:

Your main submission should be an MS Word document containing your code, images and other results produced by that code and brief textual descriptions of what you did and why. Typically, you just copy your code and results from the R console and past them into the Word document. Start with this text of homework assignment as the template. Please add your code (R, Java, Ruby, Python) into a single. It is more convenient for us to open one or two files than a large number of files. If we recognize from your Word document what your code is doing and the results it is producing we will not run your code. If we have doubts we will run your code. In order to be able to do that it is convenient if your code is in a txt document. In special cases we might request more convenient formats.

Package your submission into an archive called E63_LastNameFirstNameHW03.zip. Naming your file properly is important. We download many files and if they are all named Assignment01.zip it becomes hard not to overwrite and lose them. Please do not use archiving tools like RAR or TAR which do not produce ZIP files. If you are using a Mac, please make sure that your files are READABLE to users of Windows. You are welcome to save your work as a PDF file, but please, always submit a Word document, as well. You can use Open source imitations of Microsoft Office as well. Upload your ZIP archive to the course web site. Every assignment has its drop box. If you miss the deadline, please submit your solution into the 00_AnyHW_WayLate Drop Box. Those assignments will be graded as well. **We will chop 5% of your grade for every day you are late.** Your grade for every assignment will be entered as a comment next to your submission.

If you have issues with the formulation of the assignment or the software you are using, please FIRST go to the Discussion Forum on the class web site: <http://isites.harvard.edu/icb/icb.do?keyword=k102025> and check whether someone else raised the same issue and whether the answer is already there. If not, raise the issues yourself. A person from the class or a member of the teaching stuff will respond. The discussion forum is a very important tool. We all learn from the discussions on the forum.

If the issue is not address for a while, please send an inquiry to cscie63@fas.harvard.edu. A member of the teaching stuff will respond.

If we respond to your inquiry to class email address or to the inquiry to any email address of the teaching stuff, PLEASE DO NOT RESPOND WITH A "THANK YOU NOTE". This is not a joke. We will take 2% of your grade for that week's assignment for every "thank you note". We know that you are both polite and thankful for the effort of the volunteers who make up the teaching stuff. We all have limited time. Please let us use it for something else rather than opening and closing emails.

We will apply the same penalty to any trivial email. Please do not complain when you lose a few points on your assignment.