

## Assignment 05

Handed out: 03/01/2014

Due by 11:59PM on Friday, 03/07/2014

**Problem 1)** Please, download and install VMware Workstation 10 on your 64 bit Windows PC or VMWare Fusion 6 if you are on a 64 bit MAC. If you only have access to a 32 bit Windows or MAC, please download and install Workstation 7 and Fusion 3 respectively. Create a virtual machine based on CentOS 6.3 or 6.5 operating system. If you have a 32 bit machine, please pay attention and install 32 bit version of CentOS. If you have a 64 bit machine, please download 64 bit CentOS and create a 64 bit VM. If you work with another flavor of Linux, please be free to create a virtual machine based on your favorite flavor. Name the main user of your VM cloudera. Do not use name hadoop, as I did in my notes. “hadoop” is a bad name for a user, since Hadoop framework has an executable called hadoop and it creates many directories with that same name and those are not necessarily owned by the VM user called hadoop. On that VM create yet another user called joe. Make both users sudo users. Once your CentOS is fully installed, please shut the VM down and make another copy. Name the folder containing that copy differently. Two VMs are identical and you could even run them simultaneously if your machine has enough memory. In the folder of each VM add a text file describing OS on your VM, usernames and passwords of important users. This little file will make your VMs useful long into the future.

**Problem 2)** Use one of two VMs and follow closely steps in the attached CDH4.6 Quick Start Guide, or my notes. PDF and PPT formats and characters do not always map well into Unix (Linux) characters. If you want to copy commands you are better off doing it from the HTML version of the CDH Quick Start Guide, which you could find at: <http://www.cloudera.com/content/cloudera-content/cloudera-docs/CDH4/latest/CDH4-Quick-Start/CDH4-Quick-Start.html>. You can actually open that page inside your VM and avoid any issue caused by different character sets on Windows and Linux boxes. Both my notes and the Quick Start Guide will lead you through a “semi-automated” process of installing Hadoop. Please install YARN version of Hadoop first on one of VMs. My notes add a few explanations beyond what you can see in the Cloudera’s guide. Read the notes and the guide very carefully. If you are working with 64/32 bit CentOS VM, use only commands for 64/32 bit CentOS. Do not execute commands for other flavors of Linux. You will know that you have successfully installed Hadoop if all of tests described in the guide work properly.

**Problem 3)** Please repeat above steps for MRv1 on the other VM. We need both versions since they have slightly different MapReduce Java/C++ APIs.

Capture all steps of your implementation with comments indicating what are accomplishing with every step. Place those in an MS Word document. Please place all files you want to submit in a folder named: HW05. Compress that folder into an archive named E63\_LastNameFirstNameHW05. ZIP. Upload the archive to the course drop box. Please send comments and questions to the Discussion Forum on the

class site. If some questions are not clarified on the discussion forum please submit them to [cscie63@fas.harvard.edu](mailto:cscie63@fas.harvard.edu)