

## STSCI 5065 Pop Quiz 2

(3/1/2019)

Your Name and NetID:

1. What is URL to go to Ambari in Hortonworks Sandbox? (15 pts)

<http://127.0.0.1:8080> or <http://localhost:8080>

However, [127.0.0.1:8080](http://127.0.0.1:8080) or [localhost:8080](http://localhost:8080) is also fine.

2. What is URL to go to the CentOS shell web client in Hortonworks Sandbox? (15 pts)

<http://127.0.0.1:4200> or <http://localhost:4200>

However, [127.0.0.1:4200](http://127.0.0.1:4200) or [localhost:4200](http://localhost:4200) is also fine.

3. When you use Files View in Ambari to upload a file into HDFS, from what OS do you upload the file? (15 pts)

From the very bottom OS, Windows OS is you are using the Windows-based computer or the Mac OS if you are using a Mac.

4. What are the commonly used filesystems in HDFS? What are the maximum files sizes of the filesystems? (20 pts)

ext3 (third extended file system) with a maximum file size of 32 TB  
ext4 (fourth extended file system) with a maximum file size of 1 EB

5. What are the two default block sizes that people normally for HDFS files? Which size is now more commonly used? Why do people use such a block size? (35 pts)

The two default block sizes are 64 MB or 128 MB. Nowadays 128MB block size is more commonly used. The reason to use such big file block sizes is to minimize the seek time (or the time taken for a hard disk controller to locate a specific piece of stored data). A large block size like this can make the time to transfer the data from the disk be significantly greater than the time to seek to the start of the block. Thus, the data transfer can operate at near the disk transfer rate.