

1. Why I cannot connect to my VM?

There are several things you need to check:

- a) Check your ssh client setup
- b) Check your connection
- c) Check your ssh key pair:
 - i. If you use mac or linux, you need to make sure the private key file "key_student" can be only modified by the current user by the command "chmod 600 key_sutdent". Then try to connect by "ssh" again.
 - ii. If you use windows putty client, please make sure you choose the "student.ppk" file in the putty setting

2. Which Hadoop version should I choose?

Any version you satisfy. But the latest stable version can be a good choice.

3. What should I do if the VMs meet a memory issues when running some MapReduce examples?

You can either reduce the number of slaves or try an example with smaller input.

4. Should I do the performance test for the n-gram Part 3 with a randomly generated dataset?

No, you have not to do that. This part is recommended but not mandatory.

5. If I have messed up, what should I do if I want to start from the beginning?

The steps you need to do are:

- i. Stop all the Hadoop services
- ii. Delete the Hadoop directory
- iii. Delete the HDFS directory
- iv. Delete the template files in /tmp directory for the SecondaryNameNode
- v. Re-check any configuration files you may edited before, they can be the files as below but not restricted to:
 - a) /etc/hosts
 - b) /etc/hostname
 - c) ~/.bachrc
 - d) /etc/environment
 - e) /etc/ssh/sshd_config

f) /etc/profile

g) ...

vi. Re-extract the Hadoop package to the directory

vii. Re-do the Hadoop cluster setup from the beginning

6. In the Docker setup, what should I submit?

You need to submit the Dockerfile and the configuration files(e.g. core-site.xml, hdfs-site.xml, ...) for a pseudo-distributed (single node) Hadoop, which is much simpler than what you did in the real Hadoop cluster setup. The test of it can be done on your own computer.

If you have no idea of how to start, you can refer to other Hadoop docker image such as:

<https://github.com/sequenceiq/hadoop-docker>

7. In Dockerfile, how can we type a “yes” when builds the Docker image with some installation steps like “ apt-get install ssh”?

You just need an automatical acceptance installation command like :

```
apt-get install -y ssh
```