Quick Tips for Using the Hadoop Ecosystem: Hands On with Big Data!

Using VirtualBox

You'll need at least 8 GB of RAM in order to run HDP (Hortonworks Data Platform) on your PC – more is better. If you don't have 8GB – consider upgrading; RAM is pretty cheap these days. But, you can always just watch the videos and observe how I work with HDP without following along yourself if you need to.

Logging Into Your Sandbox with a Terminal

Throughout the course, we'll be logging into your virtual machine via SSH. Make sure you have started your virtual machine for Hortonworks using VirtualBox first, and it has finished booting up.

In my videos, we log in from Windows using a program called PuTTY, available from http://www.putty.org/. Refer to Lecture 6 on how to set this up; you need to connect to 127.0.0.1 on port 2222.

On **MacOS** or **Linux**, you can just bring up your Terminal application, and connect to your sandbox with:

```
ssh 127.0.0.1 -p 2222
```

Log in as maria_dev, with password maria_dev. So when you see me launching PuTTY in my videos, Mac and Linux users should launch your Terminal instead, and type the above command.

If Your Sandbox Seems Hosed…

If you get into a situation where you can no longer successfully boot up the Hortonworks Sandbox environment in VirtualBox or log into it, you can always delete the Hortonworks image from Virtual Box, re-download it from Hortonworks (be sure to get the sandbox version for VirtualBox,) and open up a fresh image in VirtualBox. You'll need to reset any passwords you had set after doing this, and be aware that data you may have set up in earlier lectures may be needed for future ones.

Dealing with Passwords

We'll walk through all of this in the course, but this is here for reference if you do need to delete and recreate your Hortonworks sandbox virtual machine image.

The user "maria_dev" can be used to log into Ambari and also into your Sandbox using SSH or Putty. The password for this account is "maria_dev".

Make sure you are able to connect as "root" while in SSH or Putty. Type:

```
su root
```

And from that point on, your prompt will change to a # indicating you are logged in as root with full privileges. The first time you do this on your image, you will be prompted to change the password. The default password is "hadoop", and you should change it to something you'll remember.

To manage services with Ambari, you need to use the "admin" user instead. But first, you need to set a password for admin. After opening an SSH session on your sandbox, you can do this via:

```
su root
```

```
ambari-admin-password-reset
```

(At this point you'll be prompted to enter your password for the Ambari admin user)

```
ambari-agent restart
```

Command Line Basics

If you're new to Linux, the commands I type in while connected to the Sandbox via PuTTY or SSH may be confusing. Here's a quick primer:

- **cd** – This command changes your current directory that you are working within.
- **ls** – This lists the files within the directory we're currently in.
- **less** – This is a way to quickly view the contents of a file. Press the "Q" key to exit less
- **tar** – This command is used to decompress zipped-up files that we download from the Internet. It's like unzipping.
- **wget** – This retrieves a file that's hosted on a web server. Most of the course materials are obtained used wget.
- **vi** – This is a very basic text editor included with Linux, that we'll use for things like editing configuration files. When you're in vi, you need to hit the "I" key to enter "insert mode", which lets you actually edit things. When you're done editing, press ESC to leave insert mode. Then, you can type commands such as `:wq` to write your changes and quit vi.

If you're following along, you might see me typing file names at what seems like impossible speeds. The trick is to hit the TAB key once you've typed enough of the file name for the computer to figure out what you mean; then it will "auto-complete" the file name for you.

You might also see me using the "less" command to view files, and then exiting that view in a mysterious way. Just hit the "Q" key to get out of "less."

Remember – pay attention to little things while following along! Case matters – what's uppercase and lowercase will make the difference between a command working and not working. Watch out for dashes in commands as well; sometimes you'll see a single dash (-) sometimes double dashes (--) or sometimes no dashes at all. You must transcribe what I'm typing exactly, unless I say otherwise.

Getting the Course Materials

The slides for the course are available in PDF format at http://media.sundog-soft.com/hadoop/HadoopSlides.zip

Code, configuration files, and data are downloaded directly to your sandbox using the wget command as needed throughout the course. These files won't be of much use outside of that context. However, if you really want them – they're all at http://media.sundog-soft.com/hadoop/HadoopMaterials.zip

Getting Help

In Udemy, please use the Q&A feature on individual lectures if you have any questions or problems. Myself, a teaching assistant, or fellow students will help you out if we can.