Lab Sheet: Day 1 - Hadoop and HDFS

Installing Hadoop

1. Open VirtualBox
2. Open the image **Sandbox_HDP_2.3** and proceed to installation

Getting started

We can first login to the virtual machine (pressing Alt + F5), and login with the username **root** and password **hadoop**

Monitor Hadoop Server using Ambari

The Ambari is a suite of tools for users to manage Hadoop server monitor. Enter http://127.0.0.1:8080 as the url.

Click on HDFS to see the status of HDFS server.

Task 1: Turning on and off services

1. Click HDFS
2. Click Service Actions on the right
3. Click stop

You should realize the **ops** (i.e. operations), beside the Sandbox on the top bar. After a while, the icon on the left should turn into a red exclamation mark. You have now stopped the service.

HDFS

Task 1: Login as "*HDFS*" user

Remembering your username and password

For *root*,
username: *root* password: *hadoop*

For *hdfs*
username: *hdfs* and there is no password.

As we want to perform task related to HDFS, we need to use a HDFS account, we can do this by switching user
`su – hdfs` to switch to hdfs user. If performed successfully, you will see the prompt as below:

```
[hdfs@sandbox ~]$
```

p/s: You will have to login as *root* before you are able to switch to the other users (e.g. HDFS).

```
su – hdfs
```

Task 2: Creating a directory in HDFS

Let's create a user directory using `hadoop fs –mkdir /user/hdfs`.

Let's create a temporary directory using `hadoop fs –mkdir /user/hdfs/tmp`.

We can then list the directory using `hadoop fs –ls /user` and this will give you a list of users who has an account on our HDFS.

Task 3: Create a simple file

Let's create a local file by `touch sample.txt`

We then upload the file to Hadoop using **put** command

Usage: `hadoop fs -put sample.txt /user/hdfs/sample.txt`

You can create another local file and compare. So now you can see, if you list the directory using ls.

Usage: `hdfs dfs -ls /user/hdfs/`

You will see a file sample.txt that is uploaded to our server, but two files in our local folder.

Task 4: Find out disk utilization in Hadoop

Usage: `hadoop dfs -du /user/hdfs`

Task 5: Browsing HDFS

You can now browse the HDFS by browsing:

`http://127.0.0.1:50070/`

Task 6: Download a file from HDFS

We can download a file from HDFS to local folder

Usage: `hadoop fs -get sample.txt /user/hdfs/sample.txt`

Task 7: Modify a file in HDFS

As HDFS only supports append operation, we can only add content but not removing content from a file.

Usage `hadoop fs -appendToFile sample-append.txt /user/hdfs/sample.txt`

Task 8: Delete a file in HDFS

To remove a file from HDFS

Usage `hadoop fs -rm /user/hdfs/sample.txt`