DEPARTMENT OF COMPUTER SCIENCE
6-B BIG DATA SYSTEMS & ANALYTICS (BDS&A)

AUTHORS: THOMAS BYRNE, AHMAD TAMBUWAL
CONTRIBUTIONS FROM COMPUTING: BASHIR MOHAMMED
CONTRIBUTIONS FROM IT SERVICES: TARIQ MAHMOOD

Hadoop tutorial and Lab Exercises

All teams now have their username and password which give them admin privileged to login into their respective Cloud Virtual Machine (VM) instances. Access the cloud server using the address: **bigdata<team number>.inf.brad.ac.uk** where "team number" refers to your individual group numbers. For example, team1 can access the server using the address: **bigdata1.inf.brad.ac.uk. In today's lab:**

- Team leaders will access the above server from University Campus by using ssh which available in Linux OS of all labs computers.

- This exercise offers a number of steps that will guide you through setting up your Virtual Machine (VM) and start Hadoop.

Note: Each team is working on one cloud VM, so no multiple team members can work with it at the same time.

Accessing VM using Linux OS

Step 1.

Open Linux > Open Terminal

Step 2.

    Type: `ssh bigdata<team number>.inf.brad.ac.uk`

Note: Enter <team password>

Step 3.

Navigate to local file.

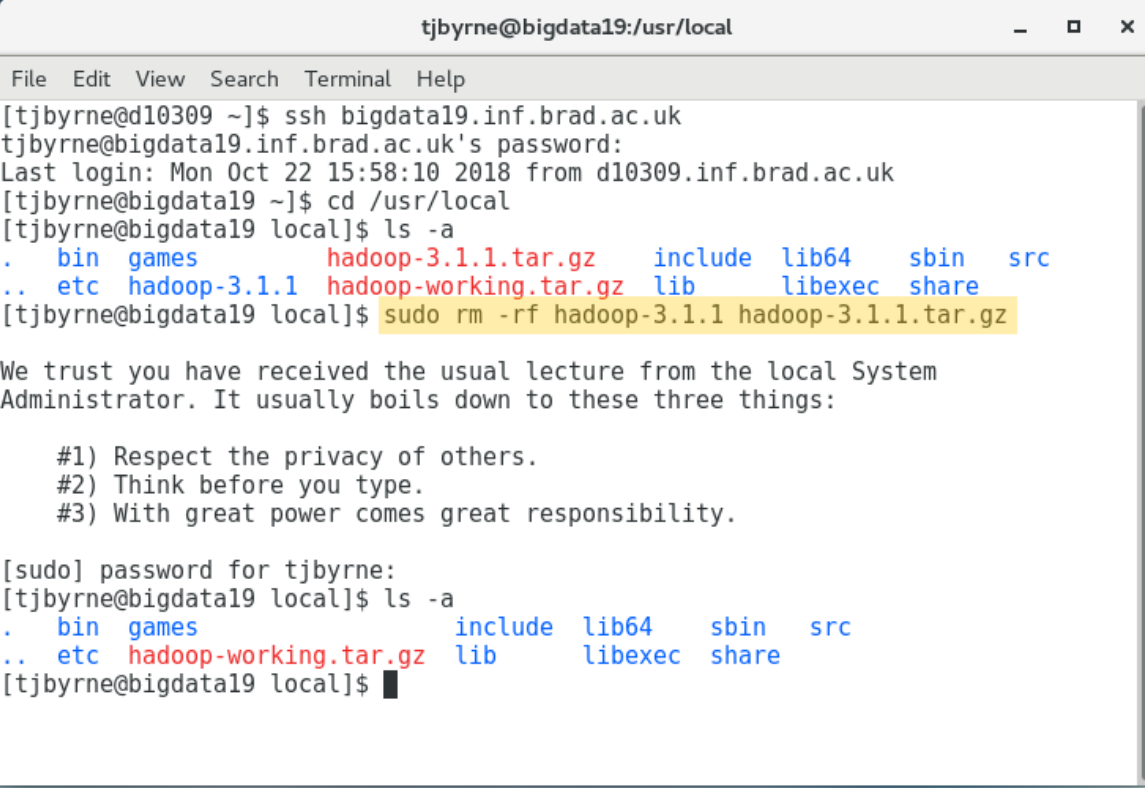   Type: `cd /usr/local`

Look inside local file.

    Type: `ls -a`

Step 4.

Delete **Hadoop-3.1.1.tar.gz** and **Hadoop-3.1.1**.

> Type: `sudo rm -rf hadoop-3.1.1 hadoop-3.1.1.tar.gz`

```
                          tjbyrne@bigdata19:/usr/local          _  □  ×

File  Edit  View  Search  Terminal  Help
[tjbyrne@d10309 ~]$ ssh bigdata19.inf.brad.ac.uk
tjbyrne@bigdata19.inf.brad.ac.uk's password:
Last login: Mon Oct 22 15:58:10 2018 from d10309.inf.brad.ac.uk
[tjbyrne@bigdata19 ~]$ cd /usr/local
[tjbyrne@bigdata19 local]$ ls -a
.    bin  games          hadoop-3.1.1.tar.gz    include  lib64    sbin    src
..   etc  hadoop-3.1.1   hadoop-working.tar.gz  lib      libexec  share
[tjbyrne@bigdata19 local]$ sudo rm -rf hadoop-3.1.1 hadoop-3.1.1.tar.gz

We trust you have received the usual lecture from the local System
Administrator. It usually boils down to these three things:

    #1) Respect the privacy of others.
    #2) Think before you type.
    #3) With great power comes great responsibility.

[sudo] password for tjbyrne:
[tjbyrne@bigdata19 local]$ ls -a
.    bin  games                  include  lib64    sbin    src
..   etc  hadoop-working.tar.gz  lib      libexec  share
[tjbyrne@bigdata19 local]$ █
```

Look inside again to confirm that **hadoop-3.1.1.tar.gz** and **hadoop-3.1.1** are delete.

> Type: `ls -a`

Step 5.

Create a Hadoop group and user.

> Type: `sudo groupadd hadoop`

> Type: `sudo adduser hduser -g hadoop`

Step 6.

Extract the zip folder called **hadoop-working.tar.gz**.

> Type: `sudo tar zxpf hadoop-working.tar.gz`

Step 7.

Open the firewall.

Type: *sudo   firewall-cmd   --permanent--zone=public   --add-port=9870/tcp*

Type: *sudo   firewall-cmd   --permanent   --zone=public   --add-port=8088/tcp*
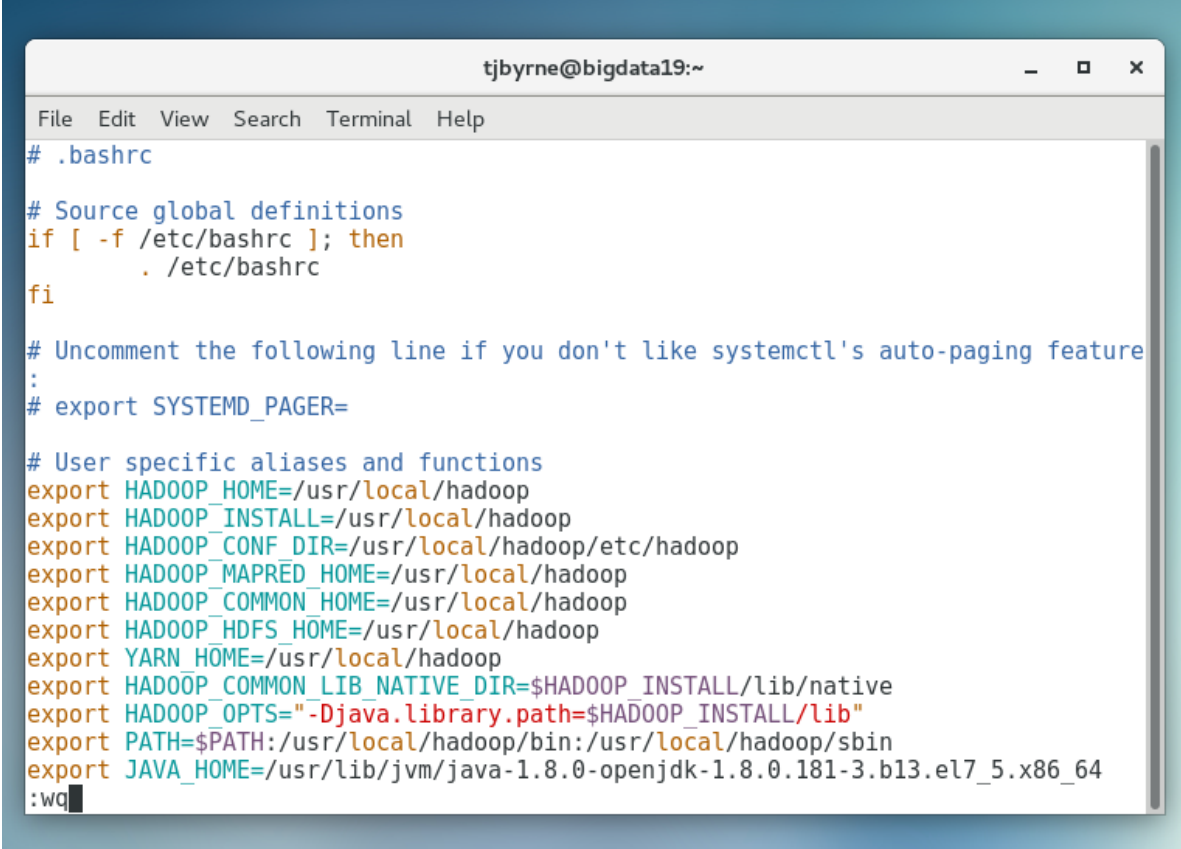
Type: *sudo firewall-cmd --reload*

Step 8.

The Hadoop folder is in */usr/local* and the following variables are used in user *~/.bashrc* file

Type: `vi ~/.bashrc`

Type:
```
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_INSTALL=/usr/local/hadoop
export HADOOP_CONF_DIR=/usr/local/hadoop/etc/hadoop
export HADOOP_MAPRED_HOME=/usr/local/hadoop
export HADOOP_COMMON_HOME=/usr/local/hadoop
export HADOOP_HDFS_HOME=/usr/local/hadoop
export YARN_HOME=/usr/local/hadoop
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib"
export PATH=$PATH:/usr/local/hadoop/bin:/usr/local/hadoop/sbin
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.181-3.b13.el7_5.x86_64
```


To save these changes.

Press: *Esc*

Type: `:wq`

To exit.

Press: *Esc*

Type: `:q!`

Step 9.

Create a data folder structure.

Type: `sudo mkdir –p /hadoop_store/hdfs`

Type: `sudo mkdir /hadoop_store/hdfs/namenode`

Type: `sudo mkdir /hadoop_store/hdfs/datanode`

Change permissions to hduser:Hadoop and hdfs format

Type: `sudo chown -R hduser:hadoop /hadoop_store`

Navigate to *hadoop* folder

Type: `cd hadoop`

Format the file system to run a MapReduce locally.

Type: `sudo bin/hdfs namenode –format`

Step 10.

Change to hduser.

Type: `sudo su – hduser`

Type: `ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa`

Type: `cat ~/.sh/id_rsa.pub >> ~/.ssh/authorized_keys`

Type: `chmod 0600 ~/.ssh/authorized_keys`

Step 11.

Change back to your own username.

Type: `exit`

Start the NameNode and DataNode daemon.

Type: `sudo sbin/start-all.sh`

Step 12.

Browse through Mozilla Firefox web browser to access the NameNode.

Type: `http://bigdata<team number>.inf.brad.ac.uk:9870`

Browse through Mozilla Firefox web browser to access the ResourceManager.

Type: `http://bigdata<team number>.inf.brad.ac.uk:8088`

All done now we are ready to start Hadoop operations

Note: If you are working on your personal computer outside the University Campus, then you need install PuTTY on your windows computer or login to your Linux remote desktop using: **ssgd2.inf.brad.ac.uk.** For those that have Linux OS on their computer, they need to access the cloud server using the ssh client (terminal command) built into their Linux desktop and connecting to: **<your university username>@ssh.inf.brad.ac.uk**

Accessing VM using '*ssgd2.inf.bradford.ac.uk*'

Step 1

To login to the Virtual Machine (VM) first connect to the remote server. To do this, type: or copy/paste *'ssgd2.inf.bradford.ac.uk'* into the URL. This will direct you to a privacy error page shown in Figure 1.

Figure 1

Step 2

To proceed click ADVANCED followed by Proceed to ssgd2.inf.bradford.ac.uk (unsafe).
This will direct you to the ORACLE Secure Global Desktop Web Server shown in Figure 2.

Figure 2

Step 3

From the *ORACLE Secure Global Desktop* page (shown above) click **'Login'** (use your
university username/password to login - see Figure 3). This will direct you to the ORACLE
Secure Global Desktop workspace home page show in Figure 4. From there click **'MY
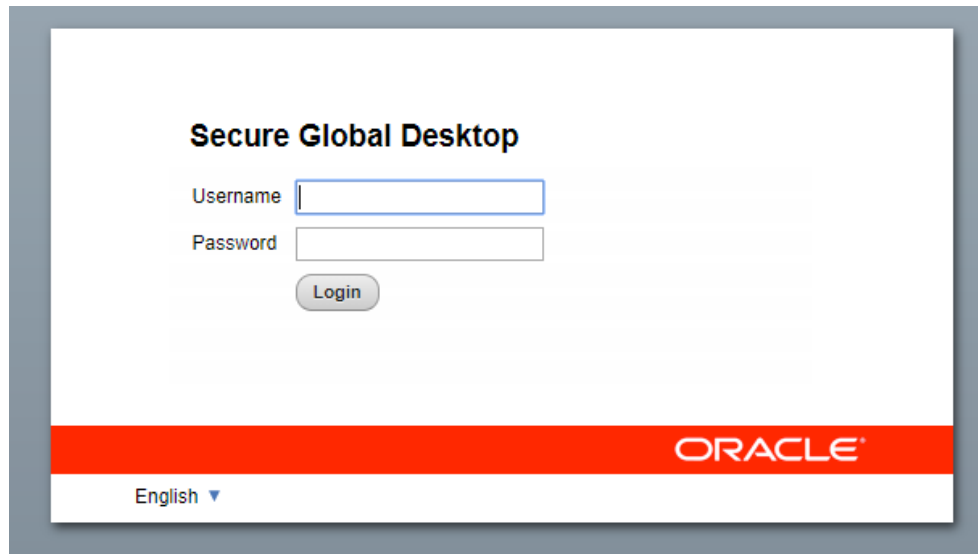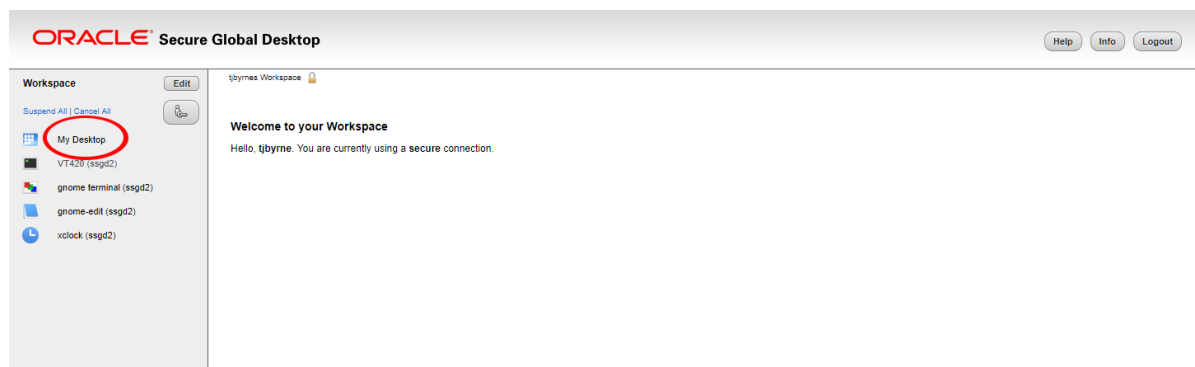Desktop'**. This will direct you to the remote server shown in Figure 5.

Figure 3

Figure 4

Step 4

Right click to reveal the menu shown in Figure 5, and click **'*Open in Terminal'***. This will open a command terminal (see Figure 6). Once the command terminal is open type: **'*ssh bigdata2.inf.brad.ac.*uk'** as shown in the first line in Figure 6, then press Enter.
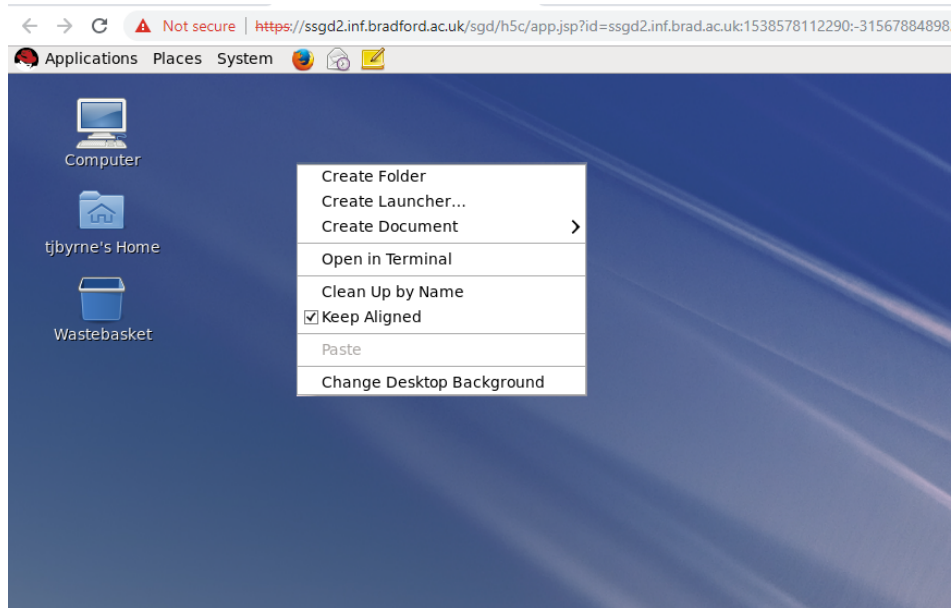
Figure 5

Note: If a message: "*The authenticity of host 'bigdata2.inf.brad.ac.uk (###.###.###.###.)' can't be established… Are you sure you want to continue connecting (yes/no)?*" - type: '*yes*' and press Enter.
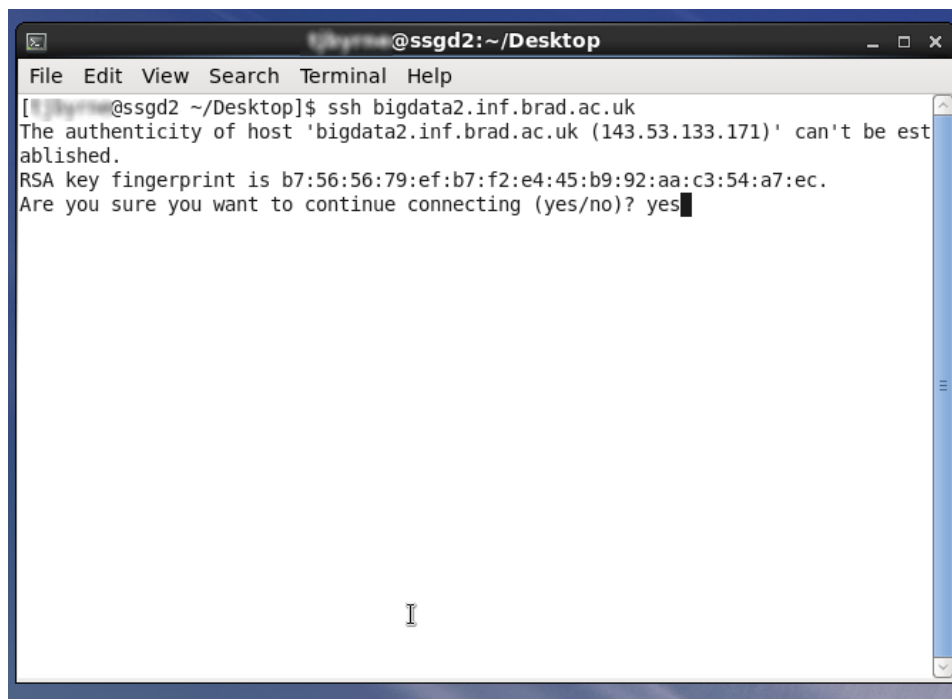

Figure 6

Step 5

Each team has been provided with a unique username (e.g. Team x) and password. Once prompted, enter the password as shown in Figure 7 and press Enter.

Note: Don't worry if you can't see the password, just type: (case sensitive) and press Enter.
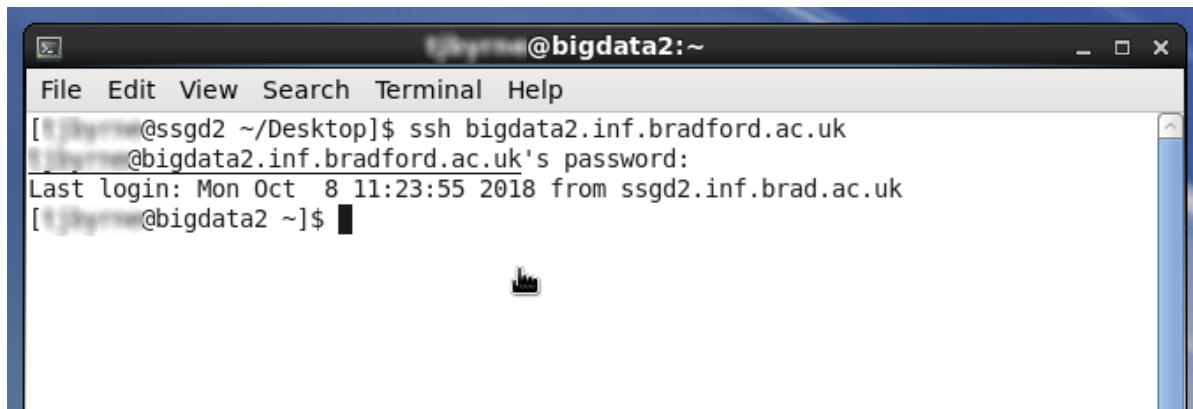

Figure 7

Step 6: Navigate to your local drive.

To navigate to the local drive where Hadoop is located type: **'cd /usr/local'** and press Enter. Then type **'ls'** to see whether you have folder name: **'hadoop'** as shown in Figure 8.
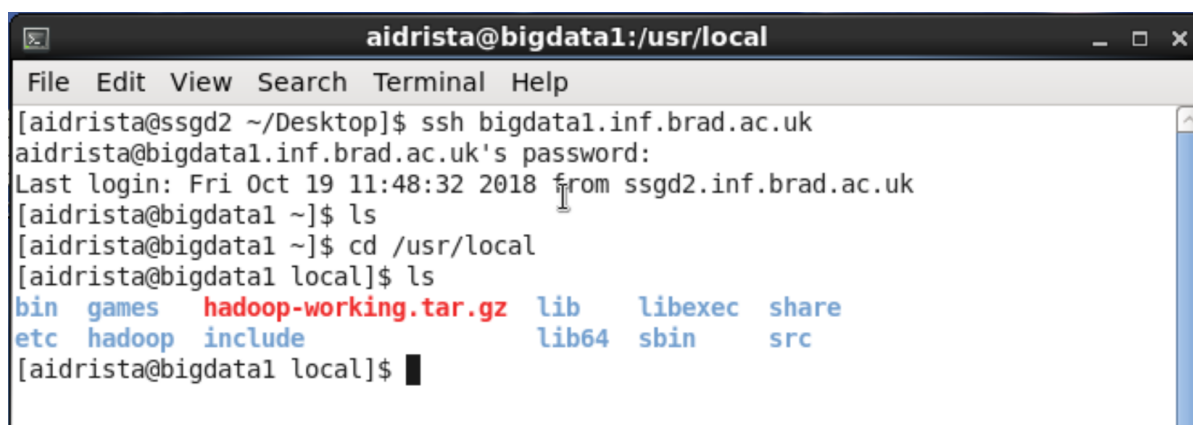

Figure 8

Accessing VM using PuTTY

Step 1:

Download PuTTY from http://www.chiark.greenend.org.uk/~sgtatham/putty/ or another PuTTY download source. The "putty.exe" download is good for basic SSH.

Step 2:

Save the download to your C: Drive Folder. If you want to make a link to PuTTY on your desktop:

- ➢ Open the C:\ folder in Windows Explorer.
- ➢ Right click on the putty.exe file and select Send To > Desktop

Step 3:

Double-click on the putty.exe program or the desktop shortcut to launch the application.

Note: You may receive a warning that the publisher cannot be verify. If you have downloaded this program from a good source, select Run. The link provided above is generally a good source for PuTTY

Step 4: Enter your connection settings as follows:

- ➢ Host Name: bigdata<team number>.inf.brad.ac.uk
- ➢ Port: 22 (leave as default)
- ➢ Connection Type: SSH (leave as default)
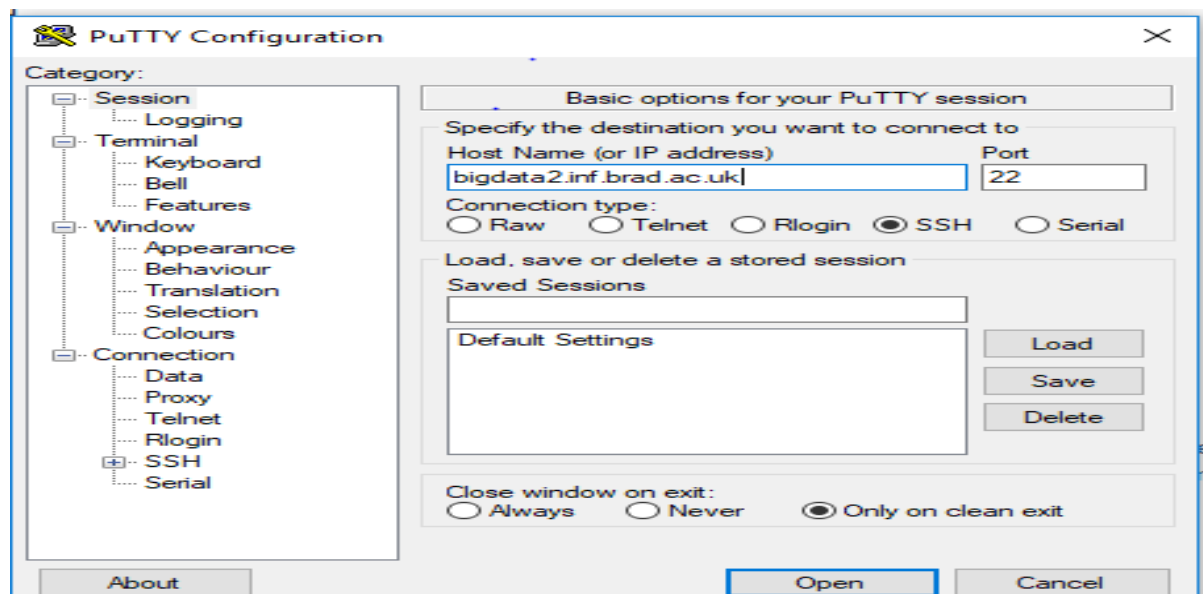
- ➢ Click Open to start the SSH session.

➢ If this is your first time connecting to the server from this computer, you will see the following output. Accept the connection by clicking Yes.

➢ Once the SSH Connection is open, you should see a terminal prompt asking for your username:

Step 5:

Type or enter your username (The team username) and your password. Please note that you will NOT see your cursor moving, or any characters typed (such as ******), when typing your password. This is a standard PuTTY security feature. Lastly, hit enter to log into your server with SSH and should see output like this:

Step 6:

Navigate to the local drive where Hadoop is located type: '*cd /usr/local*' and press Enter. Then type *'ls'* to see whether you have folder name: '*hadoop*' and continue working