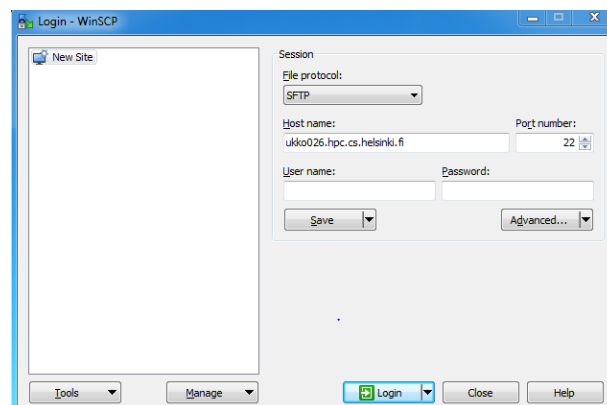


**General Instruction:** The computer science department has a high performance cluster, **Ukko**, which is available to all department's users. You can check the list of nodes and their status from [www.cs.helsinki.fi/ukko/hpc-report.txt](http://www.cs.helsinki.fi/ukko/hpc-report.txt). Public nodes are available using CS user accounts and work exactly like CS department's interactive Linux servers. In this course, you can run your Hadoop Mapreduce codes on Ukko to answer the exercise questions.

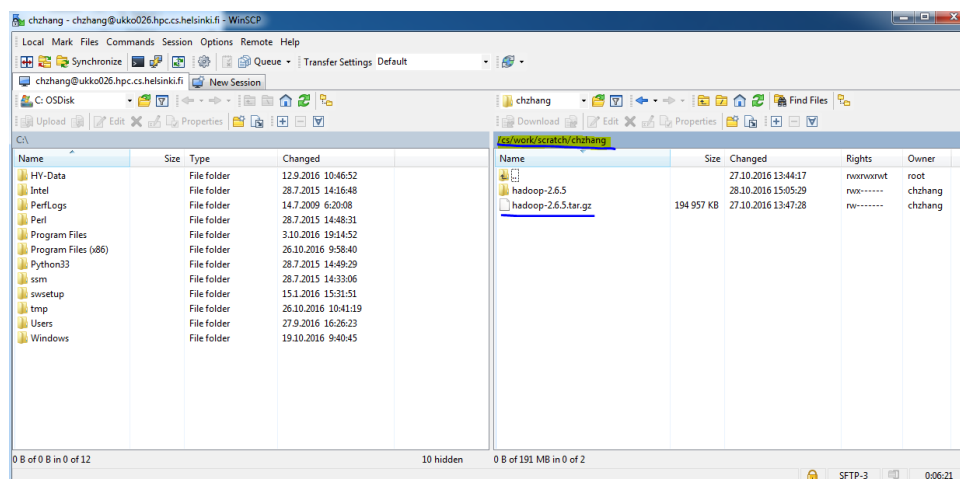
**Step 1 :** If your machine runs Windows, then use WinSCP (<https://winscp.net/eng/index.php>) to connect to the Ukko server. If you are using linux or Mac OS, please use the built-in SSH command



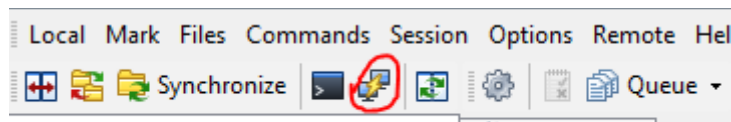
If you can't login in the Ukko nodes, please either

- (1) login to melkinpaasi.cs.helsinki.fi via SSH, then connect to Ukko using SSH from this server, or
- (2) Use HY-VPN OpenVPN (this is important, Pulse Secure is NOT able to connect Ukko): <https://helpdesk.it.helsinki.fi/en/instructions/logging-and-connections/networks/connections-outside-university>

**Step 2 :** Access to the folder `/cs/work/scratch`, build a new folder for yourself, move the Hadoop binary file here (drag it to the right-part window). In this document, we use *Hadoop.2.6.5.tar.gz*



**Step 3 :** Open the Putty embedded in the WinSCP, and unzip the Hadoop file.



The command is `tar xvf Hadoop.2.6.5.tar.gz`

If Putty is not installed in your machine, please download and install it.

<http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html>

**Step 4 :** Edit the Hadoop configure file **Hadoop-env.sh** under Hadoop-2.6.5/etc/hadoop (Right-click the file in the WinSCP and edit directly)

Add following the lines:

```
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64
export PATH=${JAVA_HOME}/bin:${PATH}
export HADOOP_CLASSPATH=${JAVA_HOME}/lib/tools.jar
```

**Step 5 :** Now the Hadoop environment is ready, next we show how to compile a WordCount MapReduce program and create a jar in the server. Download the source code of MapReduce here:

<https://hadoop.apache.org/docs/stable/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>

Move the MapReduce source codes into the folder Hadoop-2.6.5/.

Then use putty to input the following commands

**Step 6 :** Compile the java file

```
$ bin/hadoop com.sun.tools.javac.Main WordCount.java
```

Create the jar file

```
$ jar cf wc.jar WordCount*.class
```

Create the input folder

```
$ mkdir input
```

Create two files f1 and f2 under input folder

```
Hello World Bye World
```

and

```
Hello Hadoop Goodbye Hadoop
```

Run WordCount code

```
$ bin/hadoop jar wc.jar WordCount wordcount/input wordcount/output
```

Assume that the output folder does not exist here.

View the result

```
$ bin/hadoop fs -cat wordcount/output/part-r-00000
```

If success, you should see the following output:

Bye 1

Goodbye 1

Hadoop 2

Hello 2

World 2