

How to run a program on the
HDFS?

Step

1. Upload your program
2. Enter server by ssh
3. Upload your input file from local
4. Compile your Java program
5. Package your Java program
6. Upload your program to HDFS and run

Upload your program

- Upload by FTP client software
 - Filezilla, cuteftp....etc
- Server IP: 140.122.184.99
- Account: your Student No. (e.g. 60047006S)
- Password: the same as account (temp)
 - You can edit your own password by “passwd” command at you first sign in by ssh.

Enter server by ssh

- Use pietty [官網](#)
- ssh mode
- Server IP: 140.122.184.99
- Account: your Student No. (e.g. 60047006S)
- Password: the same as account (temp)
 - You can edit your own password by “passwd” command at you first sign in by ssh.

Before we go on...

Following the command is run at
home directory (~)

Upload your input file from local

- `/opt/hadoop/bin/hadoop fs -put <local-input> <HDFS-input>`
- `/opt/hadoop/bin/hadoop fs -put input input`
- [hadoop fs reference](#)

Compile your Java program

- `javac -classpath /opt/hadoop/hadoop-0.20.2-core.jar [-d <your-class-put-here>] <your-java-program>`
- `javac -classpath /opt/hadoop/hadoop-0.20.2-core.jar -d MyJava WordCount.java`
- [javac reference](#)

Package your Java program

- `jar -cvf <your-jar-name> [-C <your-classfile-dir> .] [<your-class-file>]`
- `jar -cvf wordcount.jar -C MyJava .`
- [jar reference](#)

Upload your program to HDFS and run

- `/opt/hadoop/bin/hadoop jar <your-jarfile>
<mainclass-name> [other parameter]`
- `/opt/hadoop/bin/hadoop jar wordcount.jar
WordCount input output`
- [hadoop command jar reference](#)