

# Lab 2

## Student information

---

- Full name: Qicheng Hu
- E-mail: [qhu027@cs.ucr.edu](mailto:qhu027@cs.ucr.edu)
- UCR NetID: qhu027
- Student ID: X675102

## Answers

---

- (Q1) Verify the file size and record the running time.

The command line output is:

```
From file [/Users/qhu/Downloads/AREAWATER.csv] to file [/Users/qhu/Downloads/copy_AREAWATER.csv], bytes=2271210910, time=7.202588884 second(s)
```

The file is 2271210910 bytes in total and time is 7.203 seconds.

- (Q2) Record the running time of the copy command.

Run the Linux copy command and time it:

```
> time cp AREAWATER.csv copy.txt
```

, time consumed is:

```
real    0m3.789s
user    0m0.006s
sys     0m1.552s
```

, which is 5.347 seconds in total.

- (Q3) How do the two numbers compare? Explain in your own words why you see these results.

The system command is faster than Hadoop, which is expected. The hadoop copy needs to call some libraries before calling the underlying system copy, therefore it is slower.

- (Q4) Does the program run after you change the default file system to HDFS? What is the error message, if any, that you get?

It seems that the default file system is changed to HDFS because the error is "Input file does not exist".

- (Q5) Use your program to test the following cases and record the running time for each case.

1. Copy a file from local file system to HDFS

Copy the AREAWATER from local to HDFS, console output:

bytes=2271210910, time=5.196631706 second(s)

2. Copy a file from HDFS to local file system.

bytes=2271210910, time=9.22202627 second(s)

3. Copy a file from HDFS to HDFS.

bytes=2271210910, time=8.065695134 second(s)

Case	Time
Program, Local->Local	7.203s
Command, Local->Local	5.347s
Program, Local->HDFS	5.197s
Program, HDFS->Local	9.222s
Program, HDFS->HDFS	8.066s