

Task 6. Extra Credit 3:

I used time command with the original command to get the runtimes below. For example “time hadoop jar Covid19_2.jar Covid19_2 /cse532/input/covid19_full_data.csv 2020-01-01 2020-03-31 /cse532/output/”

Task 2:

Hadoop Run Time:

```
real    0m18.878s
user    0m6.676s
sys     0m0.226s
```

Task 2:

Spark Run Time:

```
real    0m18.776s
user    0m17.004s
sys     0m1.440s
```

Task 3:

Hadoop Run Time:

```
real    0m19.105s
user    0m6.446s
sys     0m0.251s
```

Task 3:

Spark Run Time:

```
real    0m18.554s
user    0m17.090s
sys     0m1.619s
```

Both implementations give very similar run times because the input size isn't big enough to make big differences. However, we can still see that Spark run time is lower than Hadoop run time relatively. As the input size grow, Spark implementation will be more efficient than the Hadoop implementation.