

Assignment 4- Warm up Spark (10points)

For this assignment, you are to take the flight dataset for year 2001 and write a spark program which returns the top 10 flights with the highest departure delay in year 2001.

Your output should be in the following form:

Unique_Carrier	Source	Dest	year/month/day	depDelay
----------------	--------	------	----------------	----------

When I run my scala script on file 2001, I get the following output:

```
(NW,HNL,MSP,2001/1/19,1692)
(NW,SNA,DTW,2001/10/25,1592)
(NW,MCO,MSP,2001/1/6,1561)
(NW,HNL,MSP,2001/1/8,1499)
(NW,ORD,MEM,2001/9/19,1477)
(NW,HNL,MSP,2001/4/11,1466)
(NW,PHX,DTW,2001/2/9,1435)
(NW,PHX,DTW,2001/2/2,1435)
(NW,PHX,DTW,2001/2/14,1435)
(NW,PHX,DTW,2001/2/4,1435)
```

What you need to turn in:

Since we have not yet covered deploying spark applications on cluster, you can use spark-shell and/or pyspark on your master machine to write your scala or python script. Once you get the desired output, please copy and paste the script you used in a file and submit the file on blackboard.