## README.md

## Teammates

Joshua Tyndale jtyndale@rams.colostate.edu

Stephen Harayda sharayda@colostate.edu

Dziugas Butkus dziugas@rams.colostate.edu

## Project

CS455 Term project - Using Spark for Scalable Analytics

## Files

indeedProg.py - PySpark code to break down and analyze indeed job data

kaggleProg.py - PySpark code to break down and analyze housing data from craigslist

##start master and workers

tick:~/spark/latest$ ./sbin/start-all.sh

##submit jobs to master

tick:~/pySpark$ $SPARK\_HOME/bin/spark-submit --master spark://tick:30155 indeedProg1.py

## stop masters and workers

tick:~/spark/latest$ ./sbin/stop-all.sh

##config setup

Spark-env.sh

export SPARK\_MASTER\_IP=tick

export SPARK\_MASTER\_PORT=30155

export SPARK\_MASTER\_WEBUI\_PORT=30166

export SPARK\_WORKER\_CORES=2

export SPARK\_WORKER\_MEMORY=2g

export SPARK\_WORKER\_INSTANCES=2

Slaves

maggot

mosquito

lice

preying-mantis

moth

wasp

weevil

deer-fly

dragonfly

dung-beetle

spark-defaults.conf

spark.master spark://tick:30155