Group members + contributions

- Jasmine Beale
  - Helped with hadoop configurations described in lecture notes
- Kaylie Nguyen
  - helped test code for running in hadoop cluster
- Justin Morton
  - Coded, tested, and debugged hw2-mapper.py and hw2-reducer.py
- Blake Powell
  - Helped debugging and testing code
- Matthew Rodriguez
  - Helped double-check hadoop configurations and debugged code for running in a hadoop cluster

What we learned

- Jasmine Beale
  - Learned how the various xml files (core-site, mapred-site, hdfs-site) are involved in the setup of the overall hadoop cluster
- Kaylie Nguyen
  - Learned MapReduce programming and has file in HDFS
- Justin Morton
  - Learned how to design a program to utilize Hadoop for processing large amounts of data efficiently. Also improved skills of exporting info to stdout, sorting the info, and piping into another program using stdin.
- Blake Powell
  - Learned how to work with large datasets using Hadoop
- Matthew Rodriguez
  - Learned the locations of the config files for hadoop and how to find jobs and their respective inputs and outputs

Console output

```
cc@group10-1:~$ $HADOOP_PREFIX/bin/hadoop jar $HADOOP_PREFIX/contrib/streaming/hadoop-streaming-*.jar  -input /hw
2-input  -output /hw2-output8  -file /home/cc/hw2-mapper.py  -mapper /home/cc/hw2-mapper.py  -file /home/cc/hw2-r
educer.py  -reducer /home/cc/hw2-reducer.py  -cmdenv LC_CTYPE=en_GB.UTF-8
packageJobJar: [/home/cc/hw2-mapper.py, /home/cc/hw2-reducer.py, /app/hadoop/tmp/hadoop-unjar6733572948779960392/
] [] /tmp/streamjob3087239014721853459.jar tmpDir=null
21/07/16 13:50:41 INFO util.NativeCodeLoader: Loaded the native-hadoop library
21/07/16 13:50:41 WARN snappy.LoadSnappy: Snappy native library not loaded
21/07/16 13:50:41 INFO mapred.FileInputFormat: Total input paths to process : 1
21/07/16 13:50:41 INFO streaming.StreamJob: getLocalDirs(): [/app/hadoop/tmp/mapred/local]
21/07/16 13:50:41 INFO streaming.StreamJob: Running job: job_202107161259_0005
21/07/16 13:50:41 INFO streaming.StreamJob: To kill this job, run:
21/07/16 13:50:41 INFO streaming.StreamJob: /usr/local/hadoop-1.2.1/libexec/../bin/hadoop job  -Dmapred.job.track
er=10.56.2.193:54311 -kill job_202107161259_0005
21/07/16 13:50:41 INFO streaming.StreamJob: Tracking URL: http://group10-1:50030/jobdetails.jsp?jobid=job_2021071
61259_0005
21/07/16 13:50:42 INFO streaming.StreamJob:  map 0%  reduce 0%
21/07/16 13:50:46 INFO streaming.StreamJob:  map 100%  reduce 0%
21/07/16 13:50:54 INFO streaming.StreamJob:  map 100%  reduce 33%
21/07/16 13:50:56 INFO streaming.StreamJob:  map 100%  reduce 100%
21/07/16 13:50:57 INFO streaming.StreamJob: Job complete: job_202107161259_0005
21/07/16 13:50:57 INFO streaming.StreamJob: Output: /hw2-output8
cc@group10-1:~$
```

- Screenshot of program running in hadoop cluster

```
cc@group10-1:~$ cat NYPD_Complaint_Data_Current_YTD.csv | python hw2-mapper.py | sort | python hw2-reducer.py
BRONX   80267
BROOKLYN        106202
MANHATTAN       87339
QUEENS  71373
STATEN ISLAND   16521
BROOKLYN has the highest crime rate with 106202 crimes committed
Crimes include ['ADMINISTRATIVE CODE', 'AGRICULTURE & MRKTS LAW-UNCLASSIFIED', 'ARSON', 'ASSAULT 3 & RELATED OFFENSES', "BURGLAR'S TOOLS", 'BURGLARY', 'CRIMIN
AL MISCHIEF & RELATED OF', 'CRIMINAL TRESPASS', 'DANGEROUS DRUGS', 'DANGEROUS WEAPONS', 'DISORDERLY CONDUCT', 'ENDAN WELFARE INCOMP', 'FELONY ASSAULT', 'FORGE
RY', 'FRAUDS', 'FRAUDULENT ACCOSTING', 'GRAND LARCENY', 'GRAND LARCENY OF MOTOR VEHICLE', 'HARRASSMENT 2', 'INTOXICATED & IMPAIRED DRIVING', 'KIDNAPPING & REL
ATED OFFENSES', 'MISCELLANEOUS PENAL LAW', 'MURDER & NON-NEGL. MANSLAUGHTER', 'NYS LAWS-UNCLASSIFIED FELONY', 'OFF. AGNST PUB ORD SENSBLTY &', 'OFFENSES AGAIN
ST PUBLIC ADMINI', 'OFFENSES AGAINST PUBLIC SAFETY', 'OFFENSES AGAINST THE PERSON', 'OFFENSES INVOLVING FRAUD', 'OFFENSES RELATED TO CHILDREN', 'OTHER OFFENSE
S RELATED TO THEF', 'OTHER STATE LAWS (NON PENAL LA', 'PETIT LARCENY', 'POSSESSION OF STOLEN PROPERTY', 'PROSTITUTION & RELATED OFFENSES', 'RAPE', 'ROBBERY',
'SEX CRIMES', 'THEFT-FRAUD', 'UNAUTHORIZED USE OF A VEHICLE', 'VEHICLE AND TRAFFIC LAWS']
cc@group10-1:~$
```

- Screenshot of program running locally in master VM