**DATASET 1: Real\_Estate Map\_Reduce code:**

Dataset source: <https://nycopendata.socrata.com/data?browseSearch=DOF__Condominium_comparable_rental_income&type=&agency=&cat=housing+%26+development&scope=>

Since the dataset are scattered and arranged Borough wise we have merged the datasets as per years . Please download merged dataset from following link.

**Final Dataset Source:** <https://drive.google.com/drive/folders/0B44zAiU3fsFjNVZQUDdHZ3gwRG8?usp=sharing>

**RealMapper.java :** This code checks all the neighbourhood present in input dataset. We have introduce Hashmap to map the neighbourhood to precinct. This code maps neighbourhood present at 4th position in input csv file and map to precinct which will be Mapper output key. Also it sends Market per sq ft value present at 12th position in input dataset as a mapper output value.

**RealReducer.java :** This code will have Precinct number as key and Market sq ft price as value as a input.This reducer will calculate average market value for all the precinct . Output of this code will be Precinct number as key and average market value as value.

**Real.java :** This code calls RealMapper and RealReducer codes.

Run Instruction:

Run following instruction for all the datasets and save the result into separate csv files.

hadoop jar Real40.jar Real /user/cloudera/Project1/Combined\_2008-2009.csv(input file name) /user/cloudera/Project1/output1(output file name)

Output can be checked by running following instruction:

hdfs dfs -cat /user/cloudera/Project1/output1/part-r-00000

**DATASET 2: Major\_Crime Map\_Reduce code:**

Dataset source:

[https://data.cityofnewyork.us/Public-Safety/NYPD- 7-Major- Felony-Incidents/hyij- 8hr7](https://data.cityofnewyork.us/Public-Safety/NYPD-%207-Major-%20Felony-Incidents/hyij-%208hr7)

Or (NYCOpenData requires to register to get this dataset. You can use below link directly.)

<https://drive.google.com/open?id=0BwnB6-6Il8cSMjMwei1meFBzd3M>

**CrimeDataMapper.java :** This code checks precinct and the type of crime for each entry in the dataset. The code is ran for data from 2009 to 2012.

**CrimeDataReducer.java :** This code will have Precinct number as key and type of crime as value as a input. This reducer will calculate weighted average for each precinct. Again it is calculated from 2009 to 2012.

**Real.java :** This code calls CrimeDataMapper and CrimeDataReducer codes.

Run Instruction:

Run following instruction for all the datasets and save the result into separate csv files.

hadoop jar crimeData.jar Real /user/cloudera/project/NYPD\_7\_Major\_Felony\_2009.csv(input file name) /user/cloudera/Project1/output\_2009(output file name)

Output can be checked by running following instruction:

hdfs dfs -cat /user/cloudera/project1/output\_2009/part-r-00000

**DATASET 3: 311\_Map\_Reduce code:**

Dataset source:

<https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9>

And

<https://data.cityofnewyork.us/Social-Services/311-Service-Requests-for-2009/3rfa-3xsf>

**ThreeOneOneDataMapper.java :** The mapper code maps the precinct data with the 311 complaint type(agency responsible) that is registered in that area(present at column 4). As precinct data is not available in 311 data set. A mapping is made from zipcode(present at column 9) to precinct using a HashMap into which mapping is hardcoded.

**ThreeOneOneDataReducer.java :** This code takes the input as precinct : complaint type(11 types – including null) and filters out irrelevant complaint types and counts number of complaints in each of the precincts.

**ThreeOneOneDataDriver.java :** Is the driver class.

Run Instruction:

Run following instruction for all the datasets and save the result into separate csv files.

hadoop jar final\_analytic.jar ThreeOneOneDataDriver /user/cloudera/project/311\_Service\_Requests\_2011.txt /user/cloudera/project/output10\_2011

Output can be checked by running following instruction:

hdfs dfs -cat /user/cloudera/project/output10\_2011/part-r-00000