Date __/__/__
Page _____

Create a Map Reduce program to

a) find average temperature for each year from NCDC data set

Average Driver

```
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text
import org.apache.hadoop.mapreduce.Job

public
```

Mapper.py

```
from mrjob.job import MRJob
from mrjob.protocol import TextProtocol

class AverageMapper(MRJob):
    output_protocol = TextProtocol
    def mapper(self, _, line):
        missing = 9999
        year = line[15:19]

        if line[87] == '+':
            temperature = int(line[88:92])
        else:
            temperature = int(line[87:92])
        quality = line[92:93]
```

Reducer.py

```
from mrjob.job import MRJob
from mrjob.protocol import TextProtocol

Class AverageReducer(MRJob):
    Output_protocol = Text.Protocol

    def reducer(self, key, values):
        max_temp = 0;
        count = 0
        for value in values:
            max_tip = value
            count += 1
        yield key, max, temp
```

b) Find the mean max temp for every month.

Mapper.py

```
from mrjob.job import MRJob
from mrjob.protocol import TextProtocol

Class MeanMaxMapper(MRJob):
    Output_protocol = TextProtocol

    def mapper(self, line):
        missing = 999
        month = line[19:21]
```

Date __/__/__
Page _____

```
if line[87] == '+'
    temperature = int(line[88:92])
else
    temperature = int(line[87:92])

quality = line[92:93]

if temperature != missing & quality in
0,4,5,8: yield month, temperature
```

Reducer.py

```
from mrjob.job import MRJob
from mrjob.protocol import TextProtocol
class MeanMaxReducer(MRJob):
    output_protocol = TextProtocol

def reducer(self, key, values)
    max_temp = 0
    total_temp = 0
    count = 0
    days = 0

for value in values
    temp = value
    if temp > max_temp:
        max_temp = temp
        count += 1

if count == 3:
    total_temp += max_temp
```

Date __/__/__
Page

```
max_temp = 0
count = 0
days += 1

yield key, total temp
```