**Task 4**

Write MapReduce codes to perform the tasks using the files you’ve downloaded on your EMR Instance:

We downloaded the csv files into the input folder in Hadoop and

ran the corresponding mrjob files with the following command

python mrtask\_a.py input > output\_a.txt

python mrtask\_b.py input > output\_b.txt

python mrtask\_c.py input > output\_c.txt

python mrtask\_d.py input > output\_d.txt

python mrtask\_e.py input > output\_e.txt

python mrtask\_f.py input > output\_f.txt

1. Which vendors have the most trips, and what is the total revenue generated by that vendor?
2. Which pickup location generates the most revenue?
3. What are the different payment types used by customers and their count? The final results should be in a sorted format.

**Answer**: The payment types used by the customers in descending order of their counts are –

1,2,3,4 and 5.

[hadoop@ip-172-31-63-115 ~]$ cat output\_c.txt

"1" 39754212

"2" 18832370

"3" 306912

"4" 88794

"5" 3

1. What is the average trip time for different pickup locations?  
    **Answer**: The average trip time for different pickup locations are as follows. Full output is in attached file output\_d.txt

[hadoop@ip-172-31-63-115 ~]$ cat output\_d.txt

"1" "0.0hours 8.0minutes 11.18705207835643seconds"

"10" "0.0hours 49.0minutes 46.1142231548215seconds"

"100" "0.0hours 15.0minutes 27.10363236544333seconds"

"101" "0.0hours 18.0minutes 31.17156862745105seconds"

"102" "0.0hours 22.0minutes 6.7421875seconds"

"104" "0.0hours 23.0minutes 37.0seconds"

"105" "0.0hours 19.0minutes 58.66666666666674seconds"

…

1. Calculate the average tips to revenue ratio of the drivers for different pickup locations in sorted format.
2. How does revenue vary over time? Calculate the average trip revenue per month - analysing it by hour of the day (day vs night) and the day of the week (weekday vs weekend).
3. How does revenue vary over time? Calculate the average trip revenue per month - analysing it by hour of the day (day vs night) and the day of the week (weekday vs weekend).