



HOMework WEEK 7

HILDA MEIRANITA PRASTIKA DEWI

MapReduce

Data customer orders terdiri dari customer, item dan order amount, terdapat dalam file customer-orders.csv

	A	B
1	44,8602,37.19	
2	35,5368,65.89	
3	2,3391,40.64	
4	47,6694,14.98	
5	29,680,13.08	
6	91,8900,24.59	
7	70,3959,68.68	
8	85,1733,28.53	
9	53,9900,83.55	
10	14,1505,4.32	

Problem:

- Berapa banyak item yang dibeli oleh setiap customer
- Sortir total item

MRTotalItem.py

#MapReduce Job Python dengan sortir

```
from mrjob.job import MRJob
from mrjob.step import MRStep

class MRTotalItem(MRJob):

    def steps(self):
        return[
            MRStep(mapper=self.mapper_get_customer,
                  reducer=self.reducer_sum_item),
            MRStep(reducer=self.reduce_sort_counts)
        ]

    def mapper_get_customer(self, _, line):
        (customerID, itemID, orderAmount)=line.split(',')
        yield customerID, int(itemID)

    def reducer_sum_item(self, customerID, itemID):
        yield None, (sum(itemID), customerID)

    def reduce_sort_counts(self, _, pair):
        sort_pairs=sorted(pair, reverse=True)
        for pair in sort_pairs:
            yield pair

if __name__=='__main__':
    MRTotalItem.run()
```

ItemByCustomer.py

#MapReduce Job Python tanpa sortir

```
from mrjob.job import MRJob
from mrjob.step import MRStep

class ItemByCustomer(MRJob):

    def mapper(self, _, line):
        (customerID, itemID, orderAmount)=line.split(',')
        yield customerID, int(itemID)

    def reducer(self, customerID, totalItem):
        yield customerID, sum(totalItem)

if __name__=='__main__':
    ItemByCustomer.run()
```

Running MapReduce Job Python

```
!pip install mrjob
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/  
Collecting mrjob  
  Downloading mrjob-0.7.4-py2.py3-none-any.whl (439 kB)  
    |██████████████████████████████████████| 439 kB 4.9 MB/s  
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.7/dist-packages (from mrjob) (3.13)  
Installing collected packages: mrjob  
Successfully installed mrjob-0.7.4
```

```
[2] import mrjob
```

```
[22] !python ItemByCustomer.py customer-orders.csv > customer-item.txt #Sebelum di sortir
```

```
No configs found; falling back on auto-configuration  
No configs specified for inline runner  
Creating temp directory /tmp/ItemByCustomer.root.20220627.150331.882368  
Running step 1 of 1...  
job output is in /tmp/ItemByCustomer.root.20220627.150331.882368/output  
Streaming final output from /tmp/ItemByCustomer.root.20220627.150331.882368/output...  
Removing temp directory /tmp/ItemByCustomer.root.20220627.150331.882368...
```

```
[23] !python MRTotalItem.py customer-orders.csv > customer-item_sorting.txt #Sesudah di sortir
```

```
No configs found; falling back on auto-configuration  
No configs specified for inline runner  
Creating temp directory /tmp/MRTotalItem.root.20220627.151003.547442  
Running step 1 of 2...  
Running step 2 of 2...  
job output is in /tmp/MRTotalItem.root.20220627.151003.547442/output  
Streaming final output from /tmp/MRTotalItem.root.20220627.151003.547442/output...  
Removing temp directory /tmp/MRTotalItem.root.20220627.151003.547442...
```

Result

Tanpa di sortir

```
1 "0" 519786
2 "1" 526479
3 "10" 457973
4 "11" 533563
5 "12" 435462
6 "13" 458904
7 "14" 481914
8 "15" 531736
9 "16" 561911
10 "17" 468239
11 "18" 476420
12 "19" 486154
13 "2" 633531
14 "20" 447942
15 "21" 437292
16 "22" 459825
17 "23" 450260
18 "24" 499501
19 "25" 505979
20 "26" 606536
```

Setelah di sortir

```
1 633531 "2"
2 618634 "68"
3 606536 "26"
4 606010 "39"
5 594044 "71"
6 584738 "8"
7 580333 "63"
8 572736 "54"
9 569298 "46"
10 567571 "70"
11 567210 "73"
12 566011 "40"
13 561911 "16"
14 561524 "41"
15 553412 "29"
16 552536 "33"
17 550388 "82"
18 549473 "56"
19 548933 "9"
20 543990 "55"
```