

MapReduce Assignment

Task 1: Demand-Supply Mismatch Analysis

mapper.py

```
#!/usr/bin/python3
```

```
"""mapper.py"""
```

```
import sys
```

```
import csv
```

```
for row in csv.reader(sys.stdin):
```

```
    # Emit zone, regional zone, and product weight shipped
```

```
    print("%s\t%s\t%s" % (row[4], row[5], row[23]))
```

reducer.py

```
#!/usr/bin/python3
```

```
"""reducer.py"""
```

```
import sys
```

```
# Dictionary to store aggregated product weights by zone and regional zone
```

```
zone_regional_dict = {}
```

```
for line in sys.stdin:
```

```
    # Split input line into zone, regional zone, and product shipped
```

```

zone, wh_regional_zone, product_shipped = line.strip().split("\t")

try:

    product_shipped = float(product_shipped)

except ValueError:

    continue

# Update dictionary with aggregated product weights

if zone in zone_regional_dict:

    if wh_regional_zone in zone_regional_dict[zone]:

        zone_regional_dict[zone][wh_regional_zone] += product_shipped

    else:

        zone_regional_dict[zone][wh_regional_zone] = product_shipped

else:

    zone_regional_dict[zone] = {wh_regional_zone: product_shipped}

# Output aggregated results

for zone in zone_regional_dict:

    for regional in zone_regional_dict[zone]:

        print("%s\t%s\t%s" % (zone, regional, zone_regional_dict[zone][regional]))

```

```
cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py
```

```

South Zone 6 32057
West Zone 5 22068
East Zone 3 13119
South Zone 6 27142
North Zone 6 20147
West Zone 1 19098
South Zone 3 7088
South Zone 2 30078
South Zone 6 40121
South Zone 3 28071
West Zone 4 36058
North Zone 5 41079
South Zone 5 17087
South Zone 6 31136
West Zone 6 6102
West Zone 6 15129
South Zone 1 28106
West Zone 2 16091
South Zone 1 34098
West Zone 4 37065
West Zone 5 23101
South Zone 3 26091
North Zone 4 11083
North Zone 1 32093
West Zone 2 12114
South Zone 5 27080
North Zone 1 25093
West Zone 4 5058

```

hadoop@hadoop-VirtualBox:~/Documents/MapReduce/1\$ cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py

```
cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py | python3 reducer.py
```

```
hadoop@hadoop-VirtualBox:~/Documents/MapReduce/1$ cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py | python3 reducer.py
West Zone 6 52661774.0
West Zone 1 10638197.0
West Zone 4 43804669.0
West Zone 2 15146537.0
West Zone 5 32242727.0
West Zone 3 20617692.0
North Zone 5 42893115.0
North Zone 3 21335735.0
North Zone 6 100249991.0
North Zone 2 18966332.0
North Zone 4 26254519.0
North Zone 1 18466131.0
South Zone 2 32467899.0
South Zone 6 30235650.0
South Zone 4 19230670.0
South Zone 1 14682866.0
South Zone 3 18810119.0
South Zone 5 24113697.0
East Zone 3 2526684.0
East Zone 1 872338.0
East Zone 4 3306171.0
East Zone 5 1768074.0
East Zone 6 1274236.0
hadoop@hadoop-VirtualBox:~/Documents/MapReduce/1$
```

Activate Windows

Go to Settings to activate Windows.

Task 2: Warehouse Refill Frequency Correlation

mapper.py

```
#!/usr/bin/python3
```

```
"""mapper.py"""
```

```
import sys
```

```
import csv
```

```
for row in csv.reader(sys.stdin):
```

```
    # Emit warehouse capacity size and refill requests
```

```
    print("%s\t%s" % (row[3], row[6]))
```

reducer.py

```
#!/usr/bin/python3
```

```
"""reducer.py"""
```

```
import sys
```

```

# Dictionary to store warehouse capacity and list of refill requests
warehouse_dict = {}

for line in sys.stdin:

    # Split input line into capacity size and refill requests
    capacity, refill = line.strip().split("\t")

    try:

        refill = int(refill)

    except ValueError:

        continue

    # Update dictionary with capacity and refill data

    if capacity in warehouse_dict:

        warehouse_dict[capacity].append(refill)

    else:

        warehouse_dict[capacity] = [refill]

# Output aggregated results (total refill requests and count)

for warehouse in warehouse_dict:

    print("%s\t%s\t%s" % (warehouse, sum(warehouse_dict[warehouse]),
len(warehouse_dict[warehouse])))

```

```
cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py
```

```

Small 4
Large 1
Mid 1
Large 3
Large 8
Small 2
Mid 0
Mid 6
Large 2
Mid 8
Mid 7
Large 3
Large 4
Large 7
Small 6
Large 4
Small 8
Mid 1
Small 3
Mid 4
Large 3
Mid 5
Mid 7
Small 3
Mid 6
Large 7
Small 1
Mid 8

```

Activate Windows
Go to Settings to activate Windows.

hadoop@hadoop-VirtualBox:~/Documents/MapReduce/2\$ cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py

```
cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py | python3 reducer.py
```

```
hadoop@hadoop-VirtualBox:~/Documents/MapReduce/2$ cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py | python3 reducer.py
Small  19379  4811
Large  41630  10169
Mid     41217  10020
hadoop@hadoop-VirtualBox:~/Documents/MapReduce/2$
```

Activate Windows
Go to Settings to activate Windows.

Task 3. Transport Issue Impact Analysis

mapper.py

```
#!/usr/bin/python3
```

```
"""mapper.py"""
```

```
import sys
```

```
import csv
```

```
for row in csv.reader(sys.stdin):
```

```
    # Emit transport issue status and product weight shipped
```

```
    print("%s\t%s" % (row[10], row[22]))
```

reducer.py

```
#!/usr/bin/python3
```

```
"""reducer.py"""
```

```
import sys
```

```
# Dictionary to store aggregated product weights by transport issue status
```

```
transport_dict = {}
```

```
for line in sys.stdin:
```

```
# Split input line into transport issue status and product weight
```

```
transport, weight = line.strip().split("\t")
```

```
try:
```

```
    weight = float(weight)
```

```
except ValueError:
```

```
    continue
```

```
# Update dictionary with aggregated product weights
```

```
if transport in transport_dict:
```

```
    transport_dict[transport] += weight
```

```
else:
```

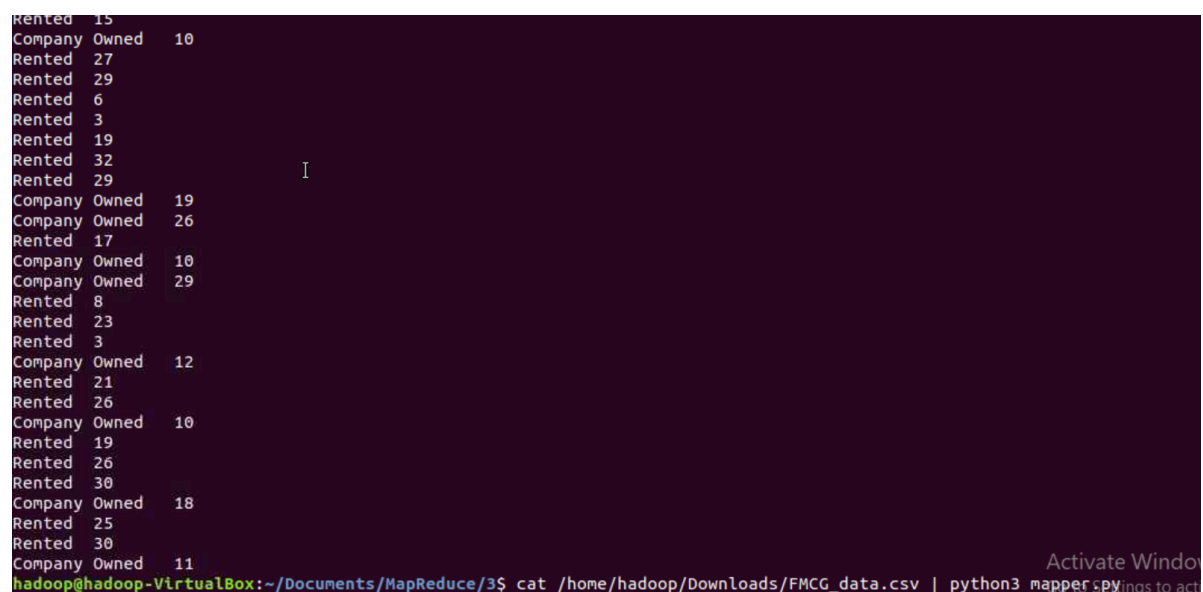
```
    transport_dict[transport] = weight
```

```
# Output aggregated results
```

```
for issue_status in transport_dict:
```

```
    print("%s\t%s" % (issue_status, transport_dict[issue_status]))
```

```
cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py
```



```
Rented 15
Company Owned 10
Rented 27
Rented 29
Rented 6
Rented 3
Rented 19
Rented 32
Rented 29
Company Owned 19
Company Owned 26
Rented 17
Company Owned 10
Company Owned 29
Rented 8
Rented 23
Rented 3
Company Owned 12
Rented 21
Rented 26
Company Owned 10
Rented 19
Rented 26
Rented 30
Company Owned 18
Rented 25
Rented 30
Company Owned 11
```

hadoop@hadoop-VirtualBox:~/Documents/MapReduce/3\$ cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py

```
cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py | python3 reducer.py
```

```
hadoop@hadoop-VirtualBox:~/Documents/MapReduce/3$ cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py | python3 reducer.py
Rented 216442.0
Company Owned 253865.0
hadoop@hadoop-VirtualBox:~/Documents/MapReduce/3$
```

Activate Windows
Go to Settings to activate Windows.

Task 4. Storage Issue Analysis

mapper.py

```
#!/usr/bin/python3
```

```
"""mapper.py"""
```

```
import sys
```

```
import csv
```

```
for row in csv.reader(sys.stdin):
```

```
    # Emit storage issue status and product weight shipped
```

```
    print("%s\t%s" % (row[19], row[22]))
```

reducer.py

```
#!/usr/bin/python3
```

```
"""reducer.py"""
```

```
import sys
```

```
# Dictionary to store aggregated product weights by storage issue status
```

```
storage_dict = {}
```

```
for line in sys.stdin:
```

```
# Split input line into storage issue status and product weight
```

```
storage, weight = line.strip().split("\t")
```

```
try:
```

```
    weight = float(weight)
```

```
except ValueError:
```

```
    continue
```

```
# Update dictionary with list of weights
```

```
if storage in storage_dict:
```

```
    storage_dict[storage].append(weight)
```

```
else:
```

```
    storage_dict[storage] = [weight]
```

```
# Output aggregated results (total weight and average weight per storage issue status)
```

```
for issue_status in storage_dict:
```

```
    total_weight = sum(storage_dict[issue_status])
```

```
    average_weight = total_weight / len(storage_dict[issue_status])
```

```
    print("%s\t%s\t%s" % (issue_status, total_weight, average_weight))
```

```
cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py
```

```
0      13
0      10
1      27
1      29
1      6
0      3
0      19
1      32
0      29
0      19
0      26
1      17
1      10
0      29
1      8
1      23
0      3
0      12
1      21
1      26
0      10
0      19
0      26
1      30
0      18
0      25
0      30
0      11
hadoop@hadoop-VirtualBox:~/Documents/MapReduce/4$ cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py
```



```
cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py | python3 reducer.py
```

```
hadoop@hadoop-VirtualBox:~/Documents/MapReduce/4$ cat /home/hadoop/Downloads/FMCG_data.csv | python3 mapper.py | python3 reducer.py
0      327281.0      18.789815133769665
1      143026.0      18.863888156159323
hadoop@hadoop-VirtualBox:~/Documents/MapReduce/4$
```

Activate Windows
Go to Settings to activate Windows.