Assignment2-Business Problem

Task 1: Demand-Supply Mismatch Analysis

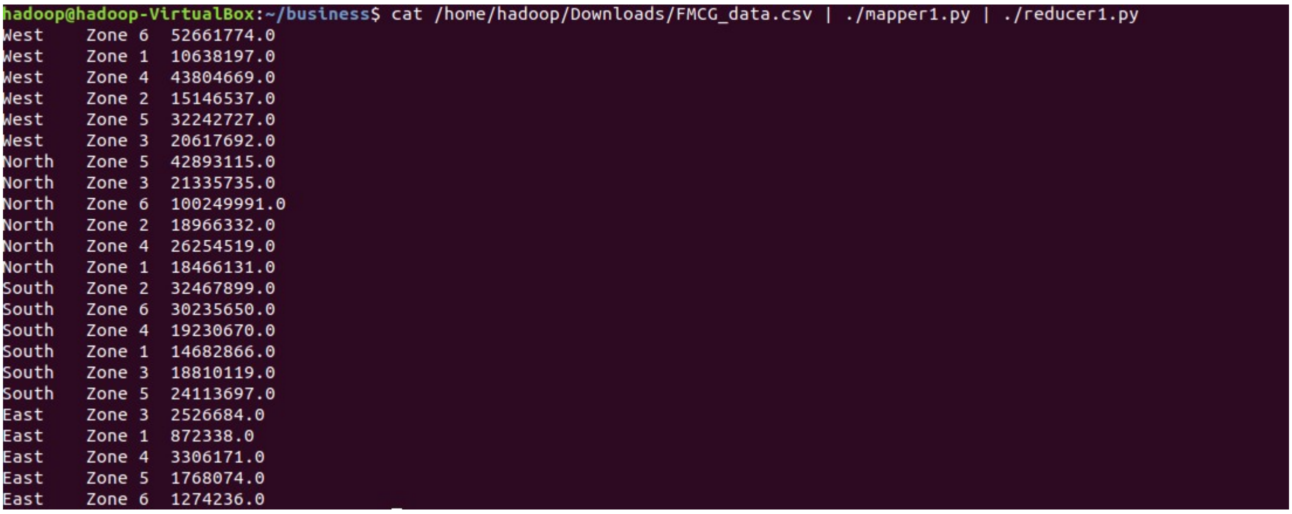
mapper.py

```
#!/usr/bin/env python3
"""mapper1.py"""
import sys
import csv
reader=csv.reader(sys.stdin)
next(reader)
for line in reader:
 print("%s\t%s\t%s"%(line[4],line[5],line[23]))
```

reducer1.py

```
#!/usr/bin/python3
"""reducer1.py"""
import sys
data = {}
for line in sys.stdin:
  zone, region, shipped = line.strip().split("\t")
 try:
  shipped = float(shipped)
 except ValueError:
  continue
 if zone not in data:
   data[zone] = {}
 if region in data[zone]:
  data[zone][region] += shipped
 else:
  data[zone][region] = shipped

for zone, regions in data.items():
 for region, total in regions.items():
   print(f"{zone}\t{region}\t{total}")
```



```
hadoop@hadoop-VirtualBox:~/business$ cat /home/hadoop/Downloads/FMCG_data.csv | ./mapper1.py | ./reducer1.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
```

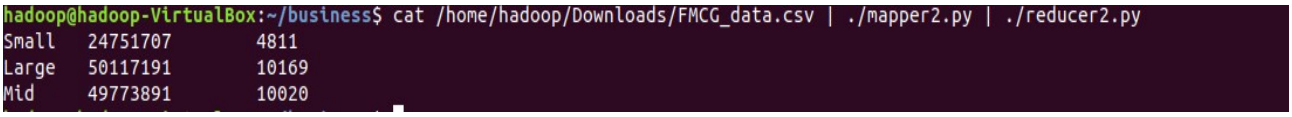
Task 2: Warehouse Refill Frequency Correlation

Mapper2.py

```
#!/usr/bin/python3
“””mapper2.py””””
import sys
import csv
reader=csv.reader(sys.stdin)
next(reader)
for line in reader:
 capacity_size = line[3]
 refill_req = int(line[9])
 print(f"{capacity_size}\t{refill_req}")
```

Reducer2.py

```
#!/usr/bin/python3
“””reducer2.py””””
import sys
from collections import defaultdict
data = defaultdict(list)
for line in sys.stdin:
 size, req = line.strip().split('\t')
 data[size].append(int(req))
for size, reqs in data.items():
 print(f"{size}\t{sum(reqs)}\t{len(reqs)}")
```



```
hadoop@hadoop-VirtualBox:~/business$ cat /home/hadoop/Downloads/FMCG_data.csv | ./mapper2.py | ./reducer2.py
Small   24751707        4811
Large   50117191        10169
Mid     49773891        10020
```

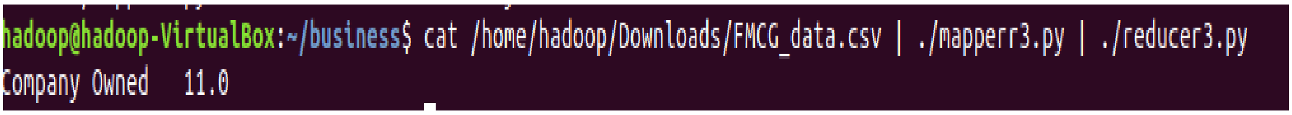
Task 3. Transport Issue Impact Analysis

Mapper3.py

```
#!/usr/bin/env python3
"""mapperr3.py"""
import sys
import csv
next(sys.stdin)
for line in csv.reader(sys.stdin):
 if len(line) > 22 and line[22].replace('.', '', 1).isdigit():
 issue = line[10]
 weight = float(line[22])
 print(f"{issue}\t{weight}")
```

Reducer3.py

```
#!/usr/bin/env python3
"""reducer3.py"""
import sys
data = {}
for line in sys.stdin:
 issue, weight = line.strip().split('\t')
 weight = float(weight)
 if issue in data:
  data[issue] += weight
 else:
  data[issue] = weight
for issue, total_weight in data.items():
 print(f"{issue}\t{total_weight}")
```



```
hadoop@hadoop-VirtualBox:~/business$ cat /home/hadoop/Downloads/FMCG_data.csv | ./mapperr3.py | ./reducer3.py
Company Owned   11.0
```

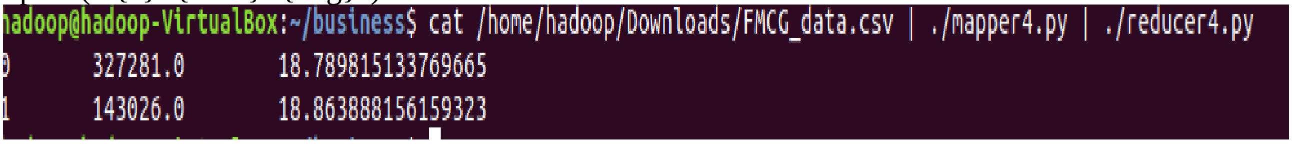
Task 4. Storage Issue Analysis

Mapper4.py

```
#!/usr/bin/python3
"""mapper4.py"""
import sys
import csv
reader=csv.reader(sys.stdin)
next(reader)
for row in reader:
 s = row[19] # storage issue
 w = row[22] # product weight
 print(f"{s}\t{w}")
```

Reducer4.py

```
#!/usr/bin/python3
"""reducer4.py"""
import sysd = {}
for line in sys.stdin:
s, w = line.strip().split("\t")
 try:
   w = float(w)
 except ValueError:
  continue
  if s in d:
   d[s].append(w)
 else:
   d[s] = [w]
for s, ws in d.items():
  total = sum(ws)
  avg = total / len(ws)
  print(f"{s}\t{total}\t{avg}")
```



```
hadoop@hadoop-VirtualBox:~/business$ cat /home/hadoop/Downloads/FMCG_data.csv | ./mapper4.py | ./reducer4.py
0       327281.0        18.789815133769665
1       143026.0        18.863888156159323
```