

## TD2 – MapReduce Solutions (Python)

All solutions are written in simple Python, compatible with Hadoop Streaming API.

### Exercise 1 – Word Frequency Counter

mapper.py:

```
#!/usr/bin/env python3
import sys

for line in sys.stdin:
    words = line.strip().split()
    for word in words:
        print(word.lower(), 1)
```

reducer.py:

```
#!/usr/bin/env python3
import sys

current_word = None
current_count = 0

for line in sys.stdin:
    word, count = line.strip().split()
    count = int(count)

    if current_word == word:
        current_count += count
    else:
        if current_word:
            print(current_word, current_count)
        current_word = word
        current_count = count

if current_word:
    print(current_word, current_count)
```

### Exercise 2 – Average Temperature by Station

mapper.py:

```
#!/usr/bin/env python3
import sys

for line in sys.stdin:
    parts = line.strip().split(',')
    if len(parts) == 3:
        station = parts[0]
        temp = parts[2]
        print(station, temp)
```

reducer.py:

```
#!/usr/bin/env python3
import sys

current_station = None
total_temp = 0
count = 0
```

```

for line in sys.stdin:
    station, temp = line.strip().split()
    temp = float(temp)

    if current_station == station:
        total_temp += temp        count += 1    else:
        if current_station:
            print(current_station, total_temp / count)
        current_station = station        total_temp = temp        count = 1

if current_station:
    print(current_station, total_temp / count)

```

### Exercise 3 – Total Sales per Product

mapper.py:

```

#!/usr/bin/env python3
import sys

for line in sys.stdin:
    parts = line.strip().split(',')
    if len(parts) == 4:
        product = parts[1]
        quantity = int(parts[2])
        price = int(parts[3])
        revenue = quantity * price        print(product, revenue)

```

reducer.py:

```

#!/usr/bin/env python3
import sys

current_product = None total_revenue = 0

for line in sys.stdin:
    product, revenue = line.strip().split()
    revenue = int(revenue)

    if current_product == product:
        total_revenue += revenue    else:
        if current_product:
            print(current_product, total_revenue)
        current_product = product        total_revenue = revenue

if current_product:
    print(current_product, total_revenue)

```

### Exercise 4 – Average Movie Rating

mapper.py:

```

#!/usr/bin/env python3
import sys

for line in sys.stdin:
    parts = line.strip().split(',')
    if len(parts) == 3:
        movie = parts[0]

```

```
rating = parts[2]
print(movie, rating)
```

reducer.py:

```
#!/usr/bin/env python3
import sys

current_movie = None total_rating = 0 count = 0

for line in sys.stdin:
    movie, rating = line.strip().split()
    rating = float(rating)

    if current_movie == movie:
        total_rating += rating        count += 1    else:
        if current_movie:
            avg = total_rating / count        print(current_movie, avg)
            current_movie = movie            total_rating = rating        count = 1

if current_movie:
    avg = total_rating / count    print(current_movie, avg)
```

## Exercise 5 – Word Count by First Letter

mapper.py:

```
#!/usr/bin/env python3
import sys

for line in sys.stdin:
    words = line.strip().split()
    for word in words:
        first = word[0].upper()
        print(first, 1)
```

reducer.py:

```
#!/usr/bin/env python3
import sys

current_letter = None current_count = 0

for line in sys.stdin:
    letter, count = line.strip().split()
    count = int(count)

    if current_letter == letter:
        current_count += count    else:
        if current_letter:
            print(current_letter, current_count)
            current_letter = letter        current_count = count

if current_letter:
    print(current_letter, current_count)
```