

Name: Manav Mehta

Roll No.: 41445

Batch: R4

Class: BE-IV

Course: Information Retrieval

Assignment 4

i. Code -

```
from collections import defaultdict
import re

# Sample dataset (you can replace this with your actual data)
data = ["Hello World!", "MapReduce in Python", "Counting characters like a pro."]

# Map function: Emits (key, value) pairs where key is a lowercase letter and value is 1
def map_function(dataset):
    intermediate = []
    for line in dataset:
        # Remove non-alphabetic characters and convert to lowercase
        letters = re.findall(r'[a-zA-Z]', line.lower())
        for letter in letters:
            intermediate.append((letter, 1))
    return intermediate

# Reduce function: Sums the counts for each letter
def reduce_function(intermediate_data):
    result = defaultdict(int)
    for key, value in intermediate_data:
        result[key] += value
    return result

# Map step
intermediate_data = map_function(data)

# Reduce step
result = reduce_function(intermediate_data)

# Output the result
for letter, count in result.items():
    print(f"{letter}: {count}")
```

ii. Output -

```
PS C:\Users\neils\OneDrive\Desktop\PICT Documents and Stuff\PICT_Coursework_BE_2024-25> python -u "C:\Users\neils\OneDrive\Desktop\PICT Documents and Stuff\PICT_Coursework_BE_2024-25\PICT_Coursework_BE_2024-25\LP-IV\IR\Assignment 4\Assignment4.py"
h: 3
e: 5
l: 4
o: 5
w: 1
r: 5
d: 2
m: 1
a: 4
a: 4
p: 3
u: 2
c: 4
i: 3
n: 4
y: 1
t: 3
g: 1
s: 1
k: 1
PS C:\Users\neils\OneDrive\Desktop\PICT Documents and Stuff\PICT_Coursework_BE_2024-25>
```