

# Cloud Computing - Project Report

Totoro Group

Zahra Omrani  
z.omrani@studenti.unipi.it

Paolo Palumbo  
p.palumbo3@studenti.unipi.it

Ettore Ricci  
e.ricci32@studenti.unipi.it

June 13, 2024

## Abstract

## 1 Introduction

## 2 Mapreduce

### 2.1 Mapper

### 2.2 Reducer

## 3 Dataset

## 4 Experiments

## 5 Results

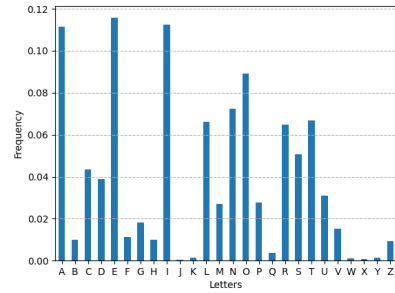


Figure 2: Italian letter frequency

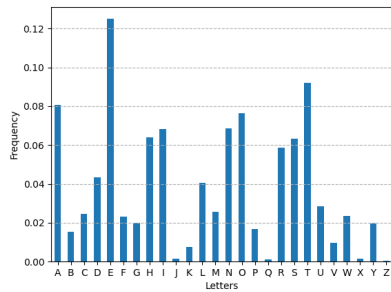


Figure 1: English letter frequency

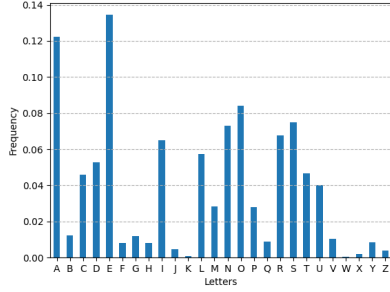


Figure 3: Spanish letter frequency

Test ID	Arguments
0	-i english.txt -r 1
1	-i english.txt -r 2
2	-i english.txt -r 4
3	-i english.txt -r 8
4	-i english.txt -r 1 -no-combiner
5	-i english.txt -r 1 -no-in-mapper-combiner
6	-i english.txt -r 1 -no-in-mapper-combiner -no-combiner
7	-i english.txt -r 4 -no-combiner
8	-i english.txt -r 4 -no-in-mapper-combiner
9	-i english.txt -r 4 -no-in-mapper-combiner -no-combiner

Table 1: General tests

Test ID	Arguments
10	-i part_100MB.txt -r 1
11	-i part_200MB.txt -r 1
12	-i part_300MB.txt -r 1
13	-i part_400MB.txt -r 1
14	-i part_500MB.txt -r 1
15	-i part_600MB.txt -r 1
16	-i part_700MB.txt -r 1
17	-i part_800MB.txt -r 1
18	-i part_900MB.txt -r 1
19	-i part_1000MB.txt -r 1
20	-i part_1100MB.txt -r 1

Table 2: Input split tests

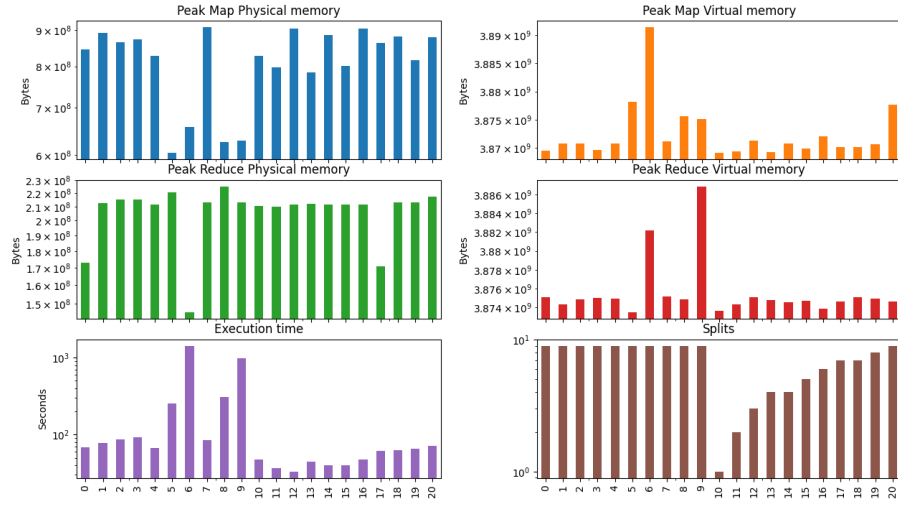


Figure 4: Tests results, the x-axis represents the test ID.

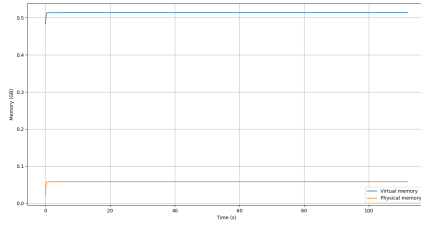


Figure 5: Python baseline memory usage (GB) over time (s)

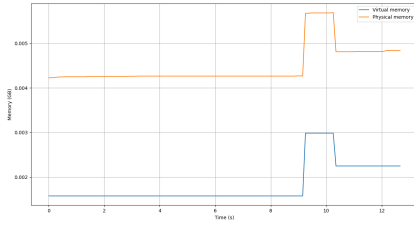


Figure 6: Rust baseline memory usage (GB) over time (s)

## 6 Conclusions