Åbo Akademi University

Cloud Computing

Assignment 5

Luis Araújo(2004624)

May 21, 2021

Contents

Chapter 1

Introduction

Github repository: https://github.com/it-teaching-abo-akademi/assignment-5-LAraujo7

Chapter 2

Problem 1: Word counting

First, I started by creating a bucket named "cc-assignment5-bucket".

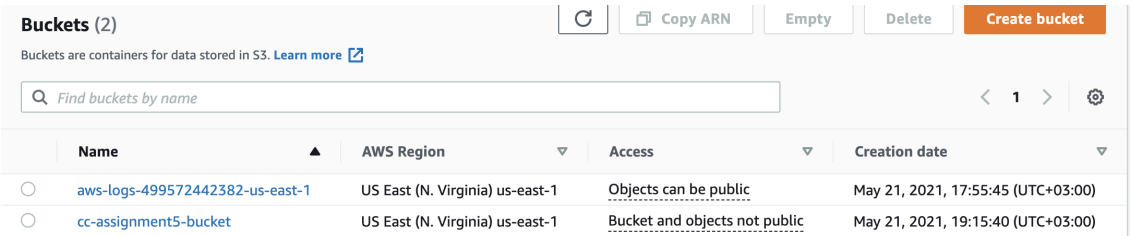

Figure 2.1: Creating a bucket

Then, the reduce, the map and the input files were uploaded to the bucket

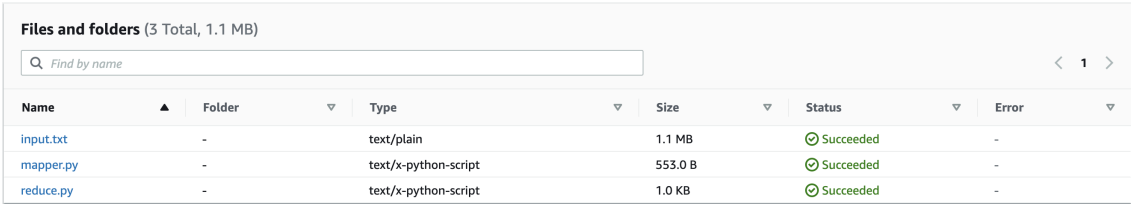

Figure 2.2: Files uploaded

The Mapper file:

```python
#!/usr/bin/env python
"""mapper.py"""

import sys

# input comes from STDIN (standard input)
for line in sys.stdin:
    # remove leading and trailing whitespace
    line = line.strip()
    # split the line into words
    words = line.split()
    # increase counters
    for word in words:
        # write the results to STDOUT (standard output);
        # what we output here will be the input for the
        # Reduce step, i.e. the input for reducer.py
        #
        # tab-delimited; the trivial word count is 1
        print '%s\t%s' % (word, 1)
```

Figure 2.3: Mapper file

Reduce file of the 100 most frequent words:

```python
#!/usr/bin/python

import sys
import collections

counter = collections.Counter()

for line in sys.stdin:
    k, v = line.strip().split("\t", 2)

    counter[k] += int(v)

print counter.most_common(100)
```

Figure 2.4: 100 Most Frequent Words

Reduce file for the words of length to 3 and 5:

```python
#!/usr/bin/env python
"""reducer_length.py"""

import sys
import collections

counter = collections.Counter()

for line in sys.stdin:
    k, v = line.strip().split("\t", 2)

    if len(k) == 3 or len(k) == 5 :
        counter[k] += int(v)

print counter.most_common(100)
```

Figure 2.5: Words of length to 3 and 5

Once uploaded the files, it was the time to create the cluster

Network and hardware

Availability zone:	us-east-1f
Subnet ID:	subnet-37517339
Master:	Running 1 m5.xlarge
Core:	Running 4 m5.xlarge
Task:	--
Cluster scaling:	Not enabled

Figure 2.6: Cluster

Finally, I added the step.

s-16BO3E2LV4W42	Streaming program	Completed	2021-05-21 20:03 (UTC+3)	38 seconds	View logs

Figure 2.7: Step completed