

Vaja 7

Postavitev in upravljanje računalniških oblakov

David Rubin

(david.rubin@student.um.si)

19. december 2018

1 Opis naloge

Vzpostaviti je bilo potrebno psevdo distribuirano postavitev Hadoop okolja in na njej pognati primer MapReduce programa.

2 Opis rešitve

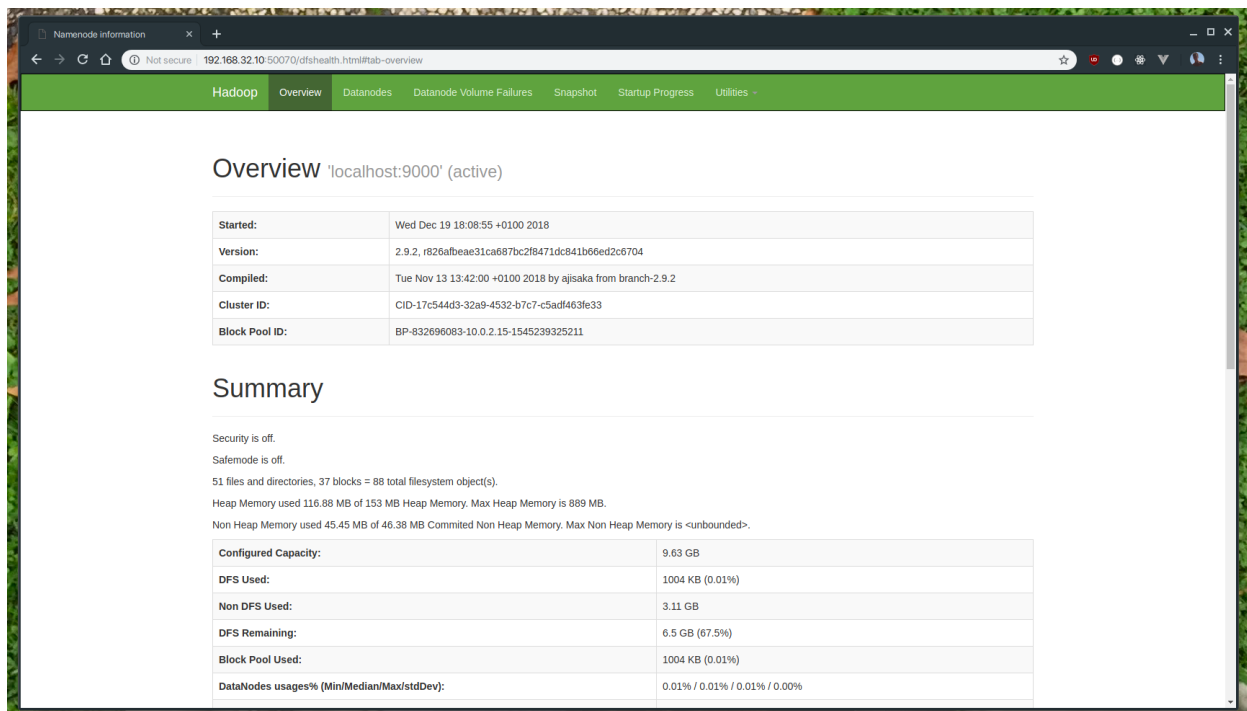
Rešitev sem implementiral v obliki Vagrantfile, ki zažene Ubuntu 18.04 in na njem vzpostavi Hadoop okolje z YARN upravljalcem. Na virtualko se namesti JAVA, zgenerira se nov ključ za povezovanje na *localhost* brez gesla in na virtualko se prenesejo 4 konfiguracijske datoteke: *core-site.xml*, *hdfs-site.xml*, *mapred-site.xml* in *yarn-site.xml*. Za tem se poženejo še ukazi za zagon HDFS in YARN ¹. Za ukaze glej navodila². Na sliki 1 Je prikazan spletni vmesnik Namenode, na sliki 2 je spletni vmesnik Resource managerja (YARN), na sliki 3 pa je prikazan izpis po zagonu primera MapReduce programa.

3 Izjava o izdelavi domače naloge

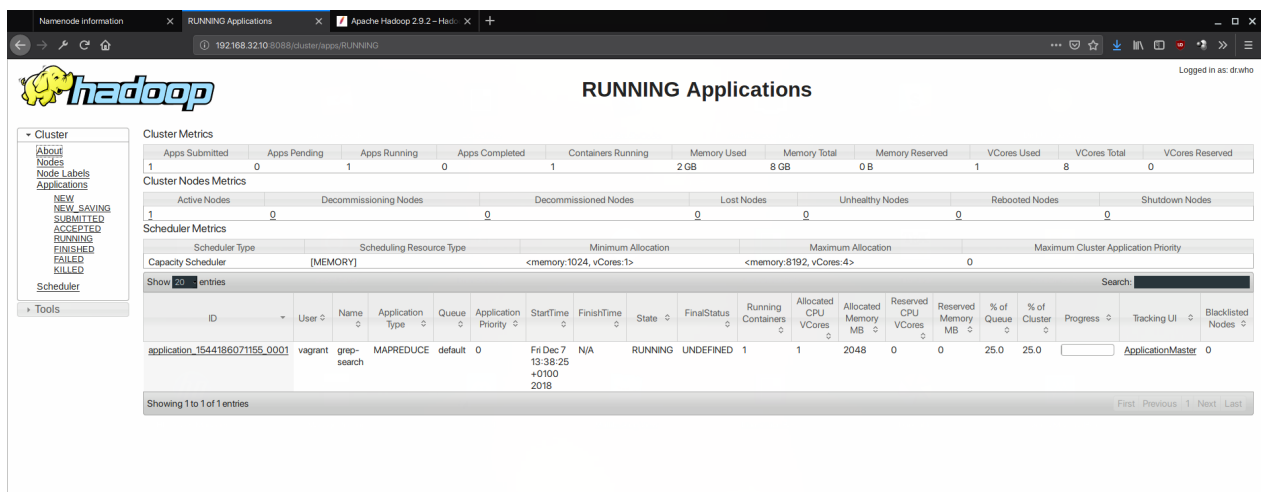
Domačo nalogo in pripadajoče programe sem izdelal sam.

¹Vagrantfile datoteka se nahaja na naslovu <https://github.com/rubinda/puro/blob/master/7-hadoop-pseudo/Vagrantfile>

²Navodila za vzpostavitev psevdo gruče se nahajajo na spletnem naslovu <https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleCluster.html>



Slika 1: Namenode na virtualki



Slika 2: YARN s tekočim MapReduce opravilom

```
vagrant@bionic: ~/hadoop-2.9.2      david@Ava: ~/Projects/PURO/7-hadoop-pseudo  +  -  □  ×
vagrant@bionic:~/hadoop-2.9.2$
vagrant@bionic:~/hadoop-2.9.2$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.9.2.jar grep inp
ut output 'dfs[a-z.]+'
18/12/19 17:12:43 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/12/19 17:12:43 INFO input.FileInputFormat: Total input files to process : 30
18/12/19 17:12:43 INFO mapreduce.JobSubmitter: number of splits:30
18/12/19 17:12:43 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprec
ated. Instead, use yarn.system-metrics-publisher.enabled
18/12/19 17:12:44 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1545239164763_0002
18/12/19 17:12:44 INFO impl.YarnClientImpl: Submitted application application_1545239164763_0002
18/12/19 17:12:44 INFO mapreduce.Job: The url to track the job: http://bionic:8088/proxy/application_154523916476
3_0002/
18/12/19 17:12:44 INFO mapreduce.Job: Running job: job_1545239164763_0002
18/12/19 17:12:51 INFO mapreduce.Job: Job job_1545239164763_0002 running in uber mode : false
18/12/19 17:12:51 INFO mapreduce.Job: map 0% reduce 0%
18/12/19 17:13:11 INFO mapreduce.Job: map 20% reduce 0%
18/12/19 17:13:29 INFO mapreduce.Job: map 37% reduce 0%
18/12/19 17:13:37 INFO mapreduce.Job: map 37% reduce 12%
18/12/19 17:13:40 INFO mapreduce.Job: map 40% reduce 12%
18/12/19 17:13:41 INFO mapreduce.Job: map 53% reduce 12%
18/12/19 17:13:43 INFO mapreduce.Job: map 53% reduce 18%
18/12/19 17:13:51 INFO mapreduce.Job: map 63% reduce 18%
18/12/19 17:13:52 INFO mapreduce.Job: map 70% reduce 18%
18/12/19 17:13:55 INFO mapreduce.Job: map 70% reduce 23%
18/12/19 17:13:58 INFO mapreduce.Job: map 73% reduce 23%
18/12/19 17:14:00 INFO mapreduce.Job: map 73% reduce 24%
18/12/19 17:14:02 INFO mapreduce.Job: map 87% reduce 24%
18/12/19 17:14:04 INFO mapreduce.Job: map 90% reduce 24%
18/12/19 17:14:07 INFO mapreduce.Job: map 90% reduce 30%
18/12/19 17:14:09 INFO mapreduce.Job: map 93% reduce 30%
18/12/19 17:14:10 INFO mapreduce.Job: map 100% reduce 100%
18/12/19 17:14:12 INFO mapreduce.Job: Job job_1545239164763_0002 completed successfully
18/12/19 17:14:12 INFO mapreduce.Job: Counters: 50
    File System Counters
        FILE: Number of bytes read=384
        FILE: Number of bytes written=6165009
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=94593
        HDFS: Number of bytes written=488
        HDFS: Number of read operations=93
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
    Job Counters
        Killed map tasks=1
        Launched map tasks=30
        Launched reduce tasks=1
        Data-local map tasks=30
        Total time spent by all maps in occupied slots (ms)=318984
        Total time spent by all reduces in occupied slots (ms)=54681
        Total time spent by all map tasks (ms)=318984
        Total time spent by all reduce tasks (ms)=54681
        Total vcore-milliseconds taken by all map tasks=318984
        Total vcore-milliseconds taken by all reduce tasks=54681
        Total megabyte-milliseconds taken by all map tasks=326639616
        Total megabyte-milliseconds taken by all reduce tasks=55993344
    Map-Reduce Framework
        Map input records=2387
        Map output records=28
        Map output bytes=663
        Map output materialized bytes=558
        Input split bytes=3684
        Combine input records=28
        Combine output records=15
        Reduce input groups=13
        Reduce shuffle bytes=558
        Reduce input records=15
        Reduce output records=13
        Spilled Records=30
        Shuffled Maps =30
        Failed Shuffles=0
        Merged Map outputs=30
        GC time elapsed (ms)=5184
        CPU time spent (ms)=14070
        Physical memory (bytes) snapshot=7740022784
        Virtual memory (bytes) snapshot=61382914048
        Total committed heap usage (bytes)=5036310528
    Shuffle Errors
        BAD_ID=0
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
        WRONG_REDUCE=0
    File Input Format Counters
        Bytes Read=90909
    File Output Format Counters
        Bytes Written=488
18/12/19 17:14:12 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/12/19 17:14:13 INFO input.FileInputFormat: Total input files to process : 1
18/12/19 17:14:13 INFO mapreduce.JobSubmitter: number of splits:1
18/12/19 17:14:14 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1545239164763_0003
18/12/19 17:14:14 INFO impl.YarnClientImpl: Submitted application application_1545239164763_0003
18/12/19 17:14:14 INFO mapreduce.Job: The url to track the job: http://bionic:8088/proxy/application_154523916476
3_0003/
18/12/19 17:14:14 INFO mapreduce.Job: Running job: job_1545239164763_0003
18/12/19 17:14:24 INFO mapreduce.Job: Job job_1545239164763_0003 running in uber mode : false
18/12/19 17:14:24 INFO mapreduce.Job: map 0% reduce 0%
```