

# Vaja 8

Postavitev in upravljanje računalniških oblakov

David Rubin

(david.rubin@student.um.si)

10. januar 2019

## 1 Opis naloge

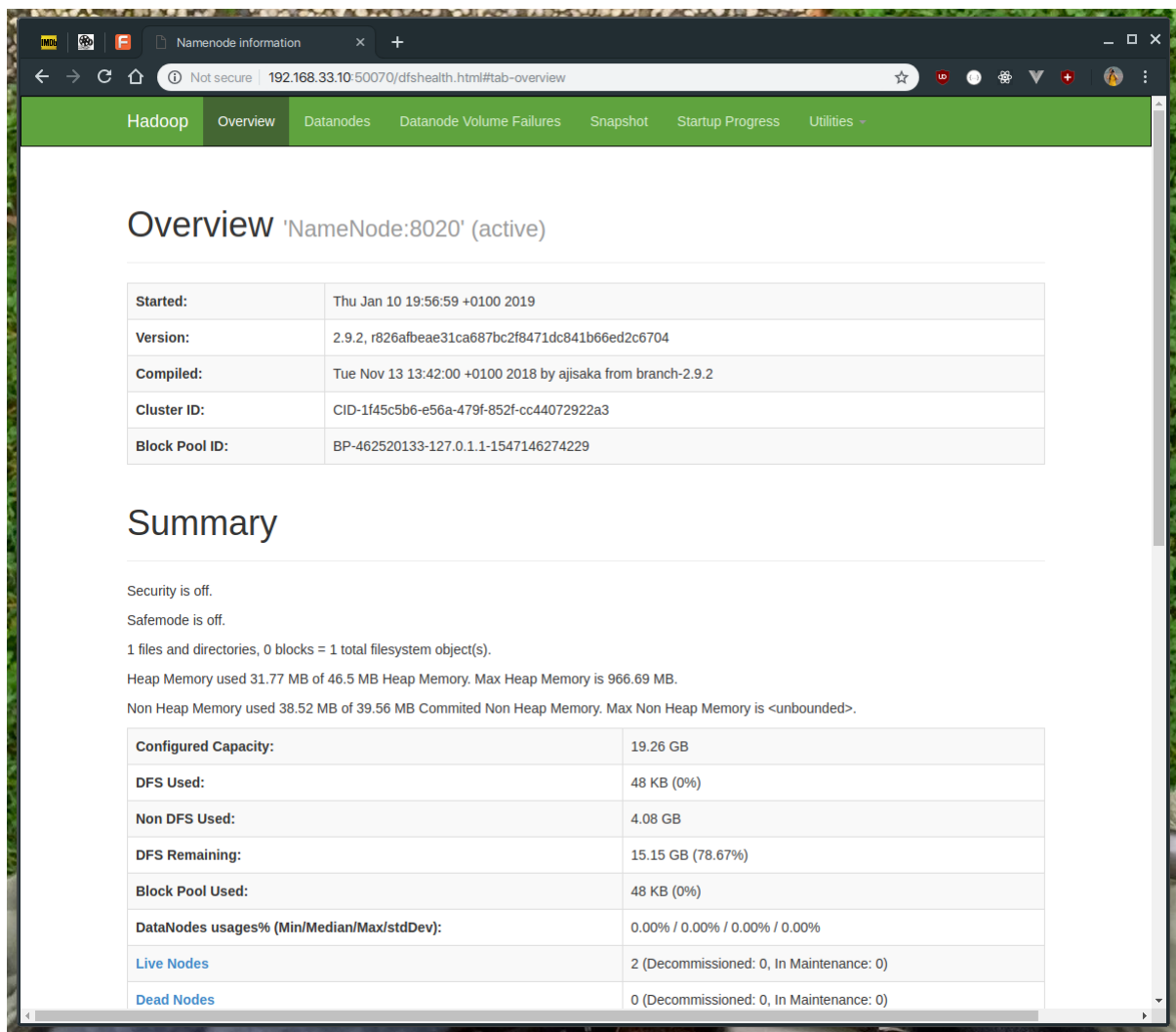
Vzpostavite Hadoop gručo za podporo Hadoopa sestavljenega iz več vozlišč.

## 2 Opis rešitve

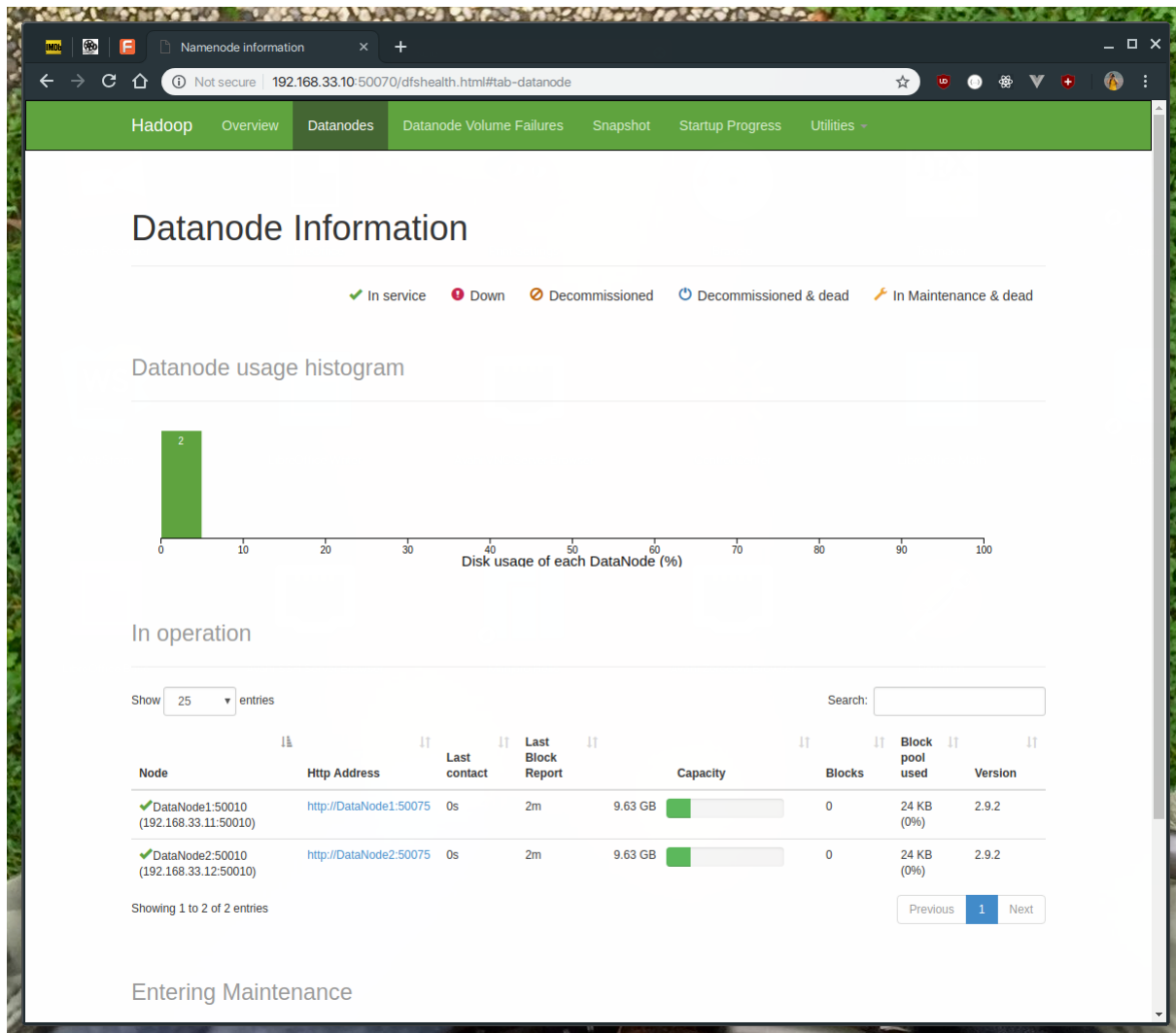
Rešitev sem implementiral v obliki Vagrantfile datoteke in še nekaj dodatnih ukazov potrebnih za delovanje. Vsebina datoteke Vagrantfile se nahaja v prilogi. Po ukazu *vagrant up*, je potrebno še izvesti nekaj korakov. Najprej se na NameNode vozlišču odstrani zapis v */etc/hosts*, kjer je vrstica *127.0.0.1 NameNode NameNode*. Slednja namreč povzroči, da ob zagonu Namenode vozlišča poslušamo le na localhost in ne na 0.0.0.0. Naslednji korak je zagon HDFS in YARN, torej skripti *start-dfs.sh* in *start-yarn.sh* znotraj */usr/local/hadoop/sbin*. Na HDFS ustvarimo mapo za vnosne podatke (*hdfs dfs -mkdir -p /user/vagrant*) in skopiramo nekaj vhodnih podatkov (*hdfs dfs -put [datoteke] input*). Sedaj lahko zaženemo primer MapReduce programa in spremljamo delovanje. Na sliki 1 vidimo pregled na Namenode vozliščem, na sliki 2 vidimo da v gruči tečeta 2 Datanode vozlišča, na sliki 3 pa je še izpis po uspešno zaključenem MapReduce programu.

## 3 Izjava o izdelavi domače naloge

Domačo nalogo in pripadajoče programe sem izdelal sam.



Slika 1: Pregled Namenode vozlišča



Slika 2: Pregled nad Datanode vozlišči, v našem primeru 2

```

vagrant@NameNode: ~
FILE: Number of bytes written=397143
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=218
HDFS: Number of bytes written=0
HDFS: Number of read operations=7
HDFS: Number of large read operations=0
HDFS: Number of write operations=2

Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots=2146
  Total time spent by all map tasks (ms)=2146
  Total time spent by all reduce tasks (ms)=213
  Total vcore-milliseconds taken by all map tasks=213
  Total vcore-milliseconds taken by all reduce tasks=213
  Total megabyte-milliseconds taken by all map tasks=213
  Total megabyte-milliseconds taken by all reduce tasks=213

Map-Reduce Framework
  Map input records=0
  Map output records=0
  Map output bytes=0
  Map output materialized bytes=6
  Input split bytes=132
  Combine input records=0
  Combine output records=0
  Reduce input groups=0
  Reduce shuffle bytes=6
  Reduce input records=0
  Reduce output records=0
  Spilled Records=0
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=163
  CPU time spent (ms)=860
  Physical memory (bytes) snapshot=353054720
  Virtual memory (bytes) snapshot=397326360
  Total committed heap usage (bytes)=137887744

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=86
  File Output Format Counters
  Bytes Written=0
vagrant@NameNode: ~

```

hadoop

Cluster

About Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used
2	0	2	0	0 B	16 GB

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
2	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers
application_1547147731129_0002	vagrant	grep-sort	MAPREDUCE	default	0	Thu Jan 10 20:16:43 +0100 2019	Thu Jan 10 20:17:02 +0100 2019	FINISHED	SUCCEEDED	N/A
application_1547147731129_0001	vagrant	grep-search	MAPREDUCE	default	0	Thu Jan 10 20:15:43 +0100 2019	Thu Jan 10 20:16:41 +0100 2019	FINISHED	SUCCEEDED	N/A

Showing 1 to 2 of 2 entries

Slika 3: Zagon primera MapReduce programa