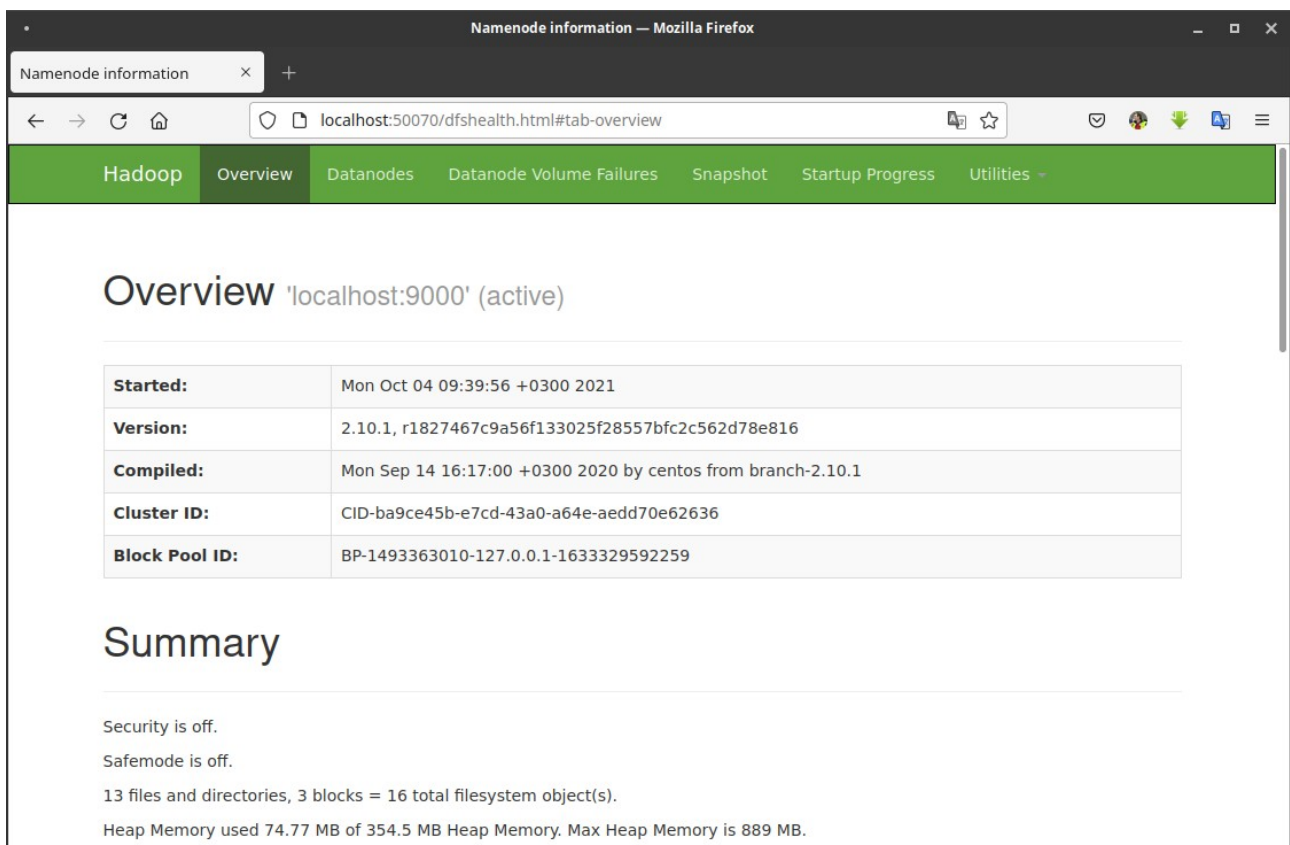


Домашнее задание к уроку 1 JVM, Apache Hadoop и Scala

1. Что означает *Stop-the-World* в терминах JVM?

Это полная остановка потоков программы для безопасной сборки мусора и других системных операций. Происходит в специальных местах программы, которые называются [safepoint](#).

2. Разверните кластер *hadoop*, соберите *WordCount* приложение, запустите на датасете *prkm_sentiment* и выведите 10 самых редких слов



The screenshot shows the Hadoop NameNode web interface in a Mozilla Firefox browser. The address bar shows `localhost:50070/dfshealth.html#tab-overview`. The interface has a green navigation bar with tabs: Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The 'Overview' tab is selected, showing the 'Overview' page for 'localhost:9000' (active). Below the navigation bar, there is a table with the following information:

Started:	Mon Oct 04 09:39:56 +0300 2021
Version:	2.10.1, r1827467c9a56f133025f28557bfc2c562d78e816
Compiled:	Mon Sep 14 16:17:00 +0300 2020 by centos from branch-2.10.1
Cluster ID:	CID-ba9ce45b-e7cd-43a0-a64e-aedd70e62636
Block Pool ID:	BP-1493363010-127.0.0.1-1633329592259

Below the table, there is a 'Summary' section with the following text:

Security is off.
Safemode is off.
13 files and directories, 3 blocks = 16 total filesystem object(s).
Heap Memory used 74.77 MB of 354.5 MB Heap Memory. Max Heap Memory is 889 MB.

GBMapReduce - WordCount.java [GBMapReduce.main]

File Edit View Navigate Code Refactor Build Run Tools VCS Window Help

GBMapReduce > src > main > java > WordCount > TokenizerMapper > m

GBMapReduce [build]

Project

build.gradle (GBMapReduce) WordCount.java

1 import org.apache.hadoop.conf.Configured;
2 import org.apache.hadoop.util.Tool;
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.io.Text;
5 import org.apache.hadoop.mapreduce.Mapper;
6 import java.io.IOException;
7 import org.apache.hadoop.mapreduce.Reducer;
8 import org.apache.hadoop.mapreduce.Partitioner;
9 import org.apache.hadoop.fs.Path;
10 import org.apache.hadoop.mapreduce.Job;
11 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
12 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
13 import org.apache.hadoop.conf.Configuration;
14 import org.apache.hadoop.util.ToolRunner;
15
16 public class WordCount extends Configured implements Tool {
17
18 public static class TokenizerMapper extends Mapper<Object, Text, Text, IntWritable> {
19 private final IntWritable one = new IntWritable(value: 1);
20 private final Text text = new Text();
21
22 @Override
23 public void map(Object key, Text value, Context context)
24 throws IOException, InterruptedException {
25 String line = value.toString().toLowerCase();
26 String[] words = line.split(regex: "\\s");
27 for (String word: words) {
28 text.set(word);
29 context.write(text, one);
30 }
31 }
32 }
33
34 public static class IntSumReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
35 private final IntWritable result = new IntWritable();
36
37 public void reduce(Text key, Iterable<IntWritable> values, Context context)
38 throws IOException, InterruptedException {
39 int sum = 0;
40 for (IntWritable val : values) {
41 sum += val.get();
42 }
43 result.set(sum);
44 context.write(key, result);
45 }
46 }
47 }
48 }

Structure Favorites Bookmarks

Run

TODO

Problems

Terminal

Build

Dependencies

Event Log

Shared indexes are downloaded for Maven library in 2 sec, 595 ms (4,14 MB) (today 13:30) 24:2 LF UTF-8 4 spaces

```
Терминал - hduser@localhost: ~
Файл  Правка  Вид  Терминал  Вкладки  Справка
rs      1
sm      1
to      1
tp      3
tu      1
w       1
wi      1
y       1
ya      6
yg      56
|       1
hduser@localhost:~$ ls
GBMapReduce-1.0-SNAPSHOT.jar  archive.zip  hadoop  ppkm_dataset.csv
hduser@localhost:~$
```

```
Терминал - hduser@localhost: ~
Файл  Правка  Вид  Терминал  Вкладки  Справка
dan    88
ppkm   100
hduser@localhost:~$ hdfs dfs -cat /user/hduser/ppkm_out/* | sort -nk2 | head
41
| 1
""kalian      1
""peraturan   1
""pilihan     1
#adaptasikebiasaambaru 1
#anggaran"    1
#aniesbaswedan 1
#astagfirallah 1
#atasi 1
hduser@localhost:~$
```