FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

1. Domaća zadaća - ROVKP

Vinko Kolobara
22. ožujka 2017.

1 ZADATAK

2. `$ start-dfs.sh`

3. `$ hdfs dfs -ls /user/rovkp`

4. `$ wget http://svn.tel.fer.hr/gutenberg.zip`

5. `$ hdfs dfs -copyFromLocal gutenberg.zip /user/rovkp`

6. `$ hdfs fsck /user/rovkp/gutenberg.zip`

   a. datoteka se sastoji od 2 bloka

   b. replikacijski faktor je 1

   c. hdfs je napravljen da jako dobro radi sa velikim datotekama, što datoteka od 151MB nije

7. `$ mv gutenberg.zip gutenberg_backup.zip`

8. `$ hdfs dfs -copyToLocal /user/rovkp/gutenberg.zip .`

9.
```
$ echo `md5sum gutenberg.zip` | \
  awk '{print $1 " gutenberg_backup.zip";}' | \
  md5sum -c
```

2 ZADATAK

```
package hr.vinko.rovkp.dz1.zad2;

import java.io.BufferedReader;
import java.io.BufferedWriter;
import java.io.FileReader;
import java.io.FileWriter;
import java.io.IOException;
import java.nio.file.FileVisitResult;
import java.nio.file.Files;
import java.nio.file.Path;
import java.nio.file.Paths;
import java.nio.file.SimpleFileVisitor;
import java.nio.file.attribute.BasicFileAttributes;
import java.util.concurrent.atomic.AtomicInteger;

public class GutenbergToTxtLocal {

  public static void main(String[] args) throws IOException {

    final AtomicInteger lineCounter = new AtomicInteger(0);

    long startTime = System.currentTimeMillis();
    try (BufferedWriter out = new BufferedWriter(new FileWriter("
        gutenberg_books.txt"))) {
      Files.walkFileTree(Paths.get("gutenberg"), new SimpleFileVisitor<
          Path>() {
        @Override
        public FileVisitResult visitFile(Path file, BasicFileAttributes
            attrs) throws IOException {

          try (BufferedReader in = new BufferedReader(new FileReader(
              file.toFile()))) {
            String line;

            while ((line = in.readLine()) != null) {
              out.write(line + "\n");
              lineCounter.getAndIncrement();
            }

          }
```

```
            return FileVisitResult.CONTINUE;
        }
      });
    }

    System.out.println("DURATION: " + (System.currentTimeMillis() -
        startTime) + "ms.");
    System.out.println("LINE NUMBERS: " + lineCounter);

  }

}
```

Konačna veličina datoteke gutenberg_books.txt je 418.242.611 byte-ova.

Ukupno je pročitano 8481553 redaka.

Za datoteku su potrebna 4 bloka, a uz faktor replikacije 3 na HDFS-u bi se stvorilo 12 blokova.

Program se izvodio 16879ms, a očekivano vrijeme na HDFS-u je bar četiri puta brže.

3 ZADATAK

```
package hr.vinko.rovkp.dz1.zad3;

import java.io.IOException;
import java.net.URI;
import java.net.URISyntaxException;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.LocalFileSystem;
import org.apache.hadoop.fs.Path;

public class HadoopJavaTest {

  public static void main(String[] args) throws IOException,
      URISyntaxException {
    Configuration conf = new Configuration();
    LocalFileSystem lfs = LocalFileSystem.getLocal(conf);
    FileSystem hdfs = FileSystem.get(new URI("hdfs://localhost:9000"),
        conf);
    Path localPath = new Path("/home/rovkp/ROVKP_DZ1/");
    Path hdfsPath = new Path("/user/rovkp/gutenberg.zip");

    System.out.println("Na putanji \"" + hdfsPath + "\" se nalazi "
        datoteka: " + hdfs.isFile(hdfsPath));
    System.out.println("Na putanji \"" + localPath + "\" se nalazi "
        direktorij: " + lfs.isDirectory(localPath));

    lfs.close();
    hdfs.close();
  }

}
```