

1. Domaća zadaća – ROVKP

Gregory Tomadin

JMBAG: 0036496597

1. Zadatak

2. `sbin\start-dfs.cmd`
3. `hdfs dfs -ls /user/gtoma`
4. `wget https://gitlab.tel.fer.hr/gutenberg.zip`
5. `hdfs dfs -copyFromLocal gutenberg.zip /user/gtoma`
6. `hdfs fsck /user/gtoma/gutenberg.zip`
 - a. datoteka se sastoji od 2 bloka
 - b. replikacijski faktor je 1
 - c. datoteka je premala za rad s HDFS-om
7. `move gutenberg.zip gutenberg2.zip`
8. `hdfs dfs -copyToLocal /user/gtoma/gutenberg.zip`
9. `powershell If((Get-FileHash gutenberg.zip -Algorithm MD5).Hash -eq (Get-FileHash gutenberg2.zip -Algorithm MD5).Hash) {echo "true"} else {echo "false"} exit`

2. Zadatak

```
package rovkp;

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.FileVisitor;
import java.nio.file.Files;
import java.nio.file.Path;
import java.nio.file.Paths;
import java.nio.file.attribute.BasicFileAttributes;

public class task2 {

    protected static int bookCounter = 0;
    protected static int lineCounter = 0;

    public static void main(String[] args) {

        long StartTime = System.currentTimeMillis();

        Path gutenbergrg =
            Paths.get("C:\\Code\\NetBeans\\ROVKP_DZ.1\\ROVKP_1DZ\\gutenbergrg_books.txt");

        try {
            BufferedWriter bw = new BufferedWriter( new
                FileWriter("C:\\Code\\NetBeans\\ROVKP_DZ.1\\ROVKP_1DZ\\gutenbergrg_books.txt"));

            Files.walkFileTree(gutenbergrg, new FileVisitor<Path>() {

                @Override
                public FileVisitResult preVisitDirectory(Path dir,
                    BasicFileAttributes attrs) throws IOException {
                    return FileVisitResult.CONTINUE;
                }
            });
        }
    }
}
```

```

@Override
public FileVisitResult visitFile(Path file,
BasicFileAttributes attrs) throws IOException {

    ++bookCounter;

    BufferedReader br = new BufferedReader(new
    FileReader(file.toFile()));

    String line;

    while(br.readLine() != null) {
        line = br.readLine();
        ++lineCounter;
        bw.write(line + "\n");
    }

    return FileVisitResult.CONTINUE;
}

@Override
public FileVisitResult visitFileFailed(Path file, IOException
exc) throws IOException {
    return FileVisitResult.CONTINUE;
}

@Override
public FileVisitResult postVisitDirectory(Path dir,
IOException exc) throws IOException {
    return FileVisitResult.CONTINUE;
}
});

System.out.println("Number books: " + bookCounter);
System.out.println("Number lines: " + lineCounter);

long endTime = System.currentTimeMillis();
long duration = endTime - StartTime;
System.out.println("Time: " + duration);

bw.close();

```

```
        } catch (IOException e) {  
            // TODO Auto-generated catch block  
            e.printStackTrace();  
        }  
    }  
}
```

1. Veličina datoteke je 208.815.183 bajta.
2. Broj redaka je 4240922.
3. Potrebna su 2 bloka za datoteku, uz replikaciju bilo bi potrebno 6 blokova.
4. Program se izvodi 1764 ms.

3. Zadatak

```
package rovkp;

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 task3 {
    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("C:/Code/NetBeans/ROVKP DZ.1/gutenberg.zip");
        Path hdfsPath = new Path("/user/gtoma/gutenberg.zip");

        System.out.println("HDFS direktoriji: " + hdfs.getFileStatus(hdfsPath));
        System.out.println("Lokalni direktoriji: " + lfs.getFileStatus(localPath));

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

Ispis:

```
HDFS direktoriji:
HdfsLocatedFileStatus{path=hdfs://localhost:9000/user/gtoma/gutenberg.zip;
isDirectory=false; length=158790992; replication=1; blockSize=134217728;
modification_time=1616667866678; access_time=1616675301851; owner=gtoma;
group=supergroup; permission=rw-r--r--; isSymlink=false; hasAcl=false;
isEncrypted=false; isErasureCoded=false}
Lokalni direktoriji:
DeprecatedRawLocalFileStatus{path=file:/C:/Code/NetBeans/ROVKP
DZ.1/gutenberg.zip; isDirectory=false; length=158790992; replication=1;
blockSize=33554432; modification_time=1616676963155; access_time=1616777608745;
owner=; group=; permission=rw-rw-rw-; isSymlink=false; hasAcl=false;
isEncrypted=false; isErasureCoded=false}
```