Linux, local filesystem, HDFS

Connection to Metacentrum cluster

```
ssh username@hador.ics.muni.cz
```

Local filesystem

1. Make a new directory `data` in your user directory.
2. On the `data` directory, set read/write/execute access rights for owner and group and read/execute for the others.
3. Copy files `stopwords.txt` and `bible-kjv.zip` from `/home/pascepet/fel_bigdata/data` directory into `data` directory inside your user directory. Switch to the `data` directory inside your user directory.
4. On the file `stopwords.txt`, set read/write access rights for owner, read for group and nothing for the others.
5. Write some first lines of the file `stopwords.txt` on the screen.
6. How many lines does the file `stopwords.txt` have? How many of them contain a string 'on'?
7. Unzip the file `bible-kjv.zip` (into the same directory). What file was inside the zip? Estimate the compression ratio.

HDFS

1. Make a new directory `data` in your user directory on HDFS.
2. On the `data` directory, set read/write/execute access rights for owner and group and read/execute for the others.
3. Copy the file `bible.txt` (unpacked from `bible-kjv.zip`) from your user directory (subdirectory `data`) on the local filesystem to the subdirectory `data` on HDFS (you have just created the subdirectory).
4. On the file `bible.txt` (on HDFS), set read/write access rights for owner, read for group and nothing for the others.
5. Find how many lines does the file `bible.txt` on HDFS have. Write some first lines of it on the screen.
6. Try to find the value of the replication factor HDFS.

Advanced Linux and regular expressions

We will work on the local filesystem, subdirectory `data` of your user directory.

1. Export to the file `a.txt` all lines of the file `stopwords.txt` starting with the letter 'a' followed by some letter.
2. Find how many lines of the file `bible.txt` contain the word 'Cain' or the word 'Abel'.
3. Find how many lines of the file `bible.txt` contain the word 'Jesus' but the line does not start with the word 'John'.
4. Find the longest word (string made of letters) in the Bible. (You may work iteratively.)