

Hadoop streaming

Uruchomienie joba hadoop streaming:

1. skopiuj skrypty mapper.py i reducer.py na vm-cluster-node2
2. Zaloguj się na vm-cluster-node2
3. uruchomienie joba

```
hadoop jar <sciezka_do_jara_hadoop_streaming> -files <lista_plikow_z_mapperem_i_reducerem> \  
-mapper <plik_z_mapperem> -reducer <plik_z_reducerem> -input <katalog_wejscowy> -output <katalog_wyjscowy>  
  
hadoop jar /opt/cloudera/parcels/CDH/jars/hadoop-streaming-2.6.0-cdh5.14.0.jar -files mapper.py,reducer.py \  
-mapper mapper.py -reducer reducer.py -input /user/xyz/loremipsum -output /user/xyz/outputs/output-2
```

Uwaga: Żeby polecenie się wykonało <katalog_wyjscowy> nie może istnieć!

Zadania:

1. policz literki w tekście loremipsum
2. posortuj policzone literki po ilości wystąpień
3. Zadanie dodatkowe: napisz mapper i reducer w C# i użyj w jobie hadoop streaming