

Mapreduce

Uruchomienie joba mapreduce:

1. Zbuduj jara i wrzuć na maszynę adminsk01
2. Zaloguj się na adminsk01
3. uruchomienie joba:

```
hadoop jar <sciezka_do_jara> <paket.Klasa> <katalog_wejscowy> <katalog_wyjscowy>  
hadoop jar mapreduce-jobs-1.0-SNAPSHOT.jar pl.isa.hadoop.WordCount /user/xyz/loremipsum /user/xyz/outputs/output-1
```

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

Zadania

1. policz literki w tekście loremipsum
2. posortuj policzone literki po ilości wystąpień (za pomocą kolejnego joba)
3. policz ile było przelewów z każdego z rachunków
4. posortuj przelewy po kwocie korzystając z wielu reduerów. Napisz Partitioner, który odpowiednio przydzieli przelew do reducera. Możesz założyć, że kwota przelewu to liczba z przedziału <1, 1000000>
5. połącz posumowane przelewy z danymi klientów z pliku 'clients'. Warto skorzystać z klasy MultipleInputs (reduce side join), lub DistributedCache (map side join)
6. zbuduj indeks odwócony (słowo -> lista plików w których się znajduje) z wybranych danych. Nazwę pliku można wydobyć za pomocą:

```
((FileSplit) context.getInputSplit()).getPath().getName();
```