

Mapreduce

Uruchomienie joba mapreduce:

1. Zbuduj jara i wrzuć na maszynę vm-cluster-node2
2. Zaloguj się na vm-cluster-node2
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
3. posumuj kwotę z pliku 'transfers' pogrupowaną po rachunku źródłowym
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