

## Mapreduce

Uruchomienie joba mapreduce:

1. Zbuduj jara i wrzuć na maszynę vm-cluster-node2
2. Zaloguj się na vm-cluster-node2
3. uruchom joba poleceniem `'hadoop jar ścieżka_do_jara pakiet.Klasa katalog_wejściowy katalog_wyjściowy'`  
Żeby polecenie się wykonało *katalog\_wyjściowy* nie może istnieć

Zadania:

1. Napisz joba mapreduce, który policzy literki w tekście loremipsum
2. Napisz joba mapreduce, który posortuje policzone literki
3. Napisz joba mapreduce, który posumuje content\_length z apache\_logs pogrupowane po metodzie HTTP
4. Napisz joba mapreduce, który usunie zduplikowane linijki z loga
5. Napisz joba mapreduce, który wypisze połączone Imie, Nazwisko i useragent z plików apache\_logs i dump.sql. Warto skorzystać z klasy MultipleInputs