

# Mapreduce

Run mapreduce job:

1. Build jar and copy it on hadoop machine
2. Login on hadoop machine
3. Run jar:

```
hadoop jar <path_to_jar> <package.Class> <arguments>  
hadoop jar mapreduce-Jobs-1.0-SNAPSHOT.jar pl.isa.hadoop.WordCount /user/xyz/loremipsum /user/xyz/outputs/output-1
```

Attention: Job will not execute if <output-dir> exists!

## Tasks

1. count letters in loremipsum
2. sort counted letters by occurrence (you will need another job for that)
3. count how many transfers there were from each account (in 'transfers' file)
4. join summed tranfers with client name from file 'clients'. You can use MultipleInputs (reduce side join), or DistributedCache (map side join).

## Extra tasks

1. sort transfers by amount using many reducers (you will have many result files). Write partitioner which will assign transfer to specific reducer. You can assume that amount is a number in range <1, 1000000>
2. build inverted index from multiple files (word -> list of files containing this word). You can fetch name of the file using:

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