

BIG DATA

ATIVIDADE 11 - HADOOP MAPREDUCE

- Imagem hadoop funcionando com Docker:

```
hadoop@2b6e3fe16032:~$ /opt/hadoop/sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
localhost: Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Starting datanodes
Starting secondary namenodes [2b6e3fe16032]
2b6e3fe16032: Warning: Permanently added '2b6e3fe16032,10.40.0.2' (ECDSA) to the list of known hosts.
Starting resourcemanager
Starting nodemanagers
hadoop@2b6e3fe16032:~$ jps
816 ResourceManager
1330 Jps
551 SecondaryNameNode
360 DataNode
219 NameNode
1116 NodeManager
```

- Código Python usado:

```
from mrjob.job import MRJob
from mrjob.step import MRStep

class MoviesCount(MRJob):
    def steps(self):
        return [
            MRStep(mapper=self.mapper_count,
                  reducer=self.reducer_count),
            MRStep(reducer=self.reducer_sort)
        ]

    def mapper_count(self, _, line):
        user_id, movie_id, rating, timestamp = line.split('\t')
        yield movie_id, 1

    def reducer_count(self, movie_id, counts):
        yield None, (sum(counts), movie_id)

    def reducer_sort(self, _, count_movie_pairs):
        for count, movie_id in sorted(count_movie_pairs, reverse=True):
            yield movie_id, count

if __name__ == '__main__':
    MoviesCount.run()
```

- Partes do resultado final (saída no terminal):

```
"1197" 45
"1079" 45
"984" 44
"975" 44
"896" 44
"846" 44
"761" 44
"694" 44
"688" 44
"656" 44
"638" 44
"575" 44
"539" 44
"224" 44
"1188" 44
"1142" 44
"1067" 44
"1025" 44
"1010" 44
"944" 43
"890" 43
"809" 43
"620" 43
"589" 43
"540" 43
"407" 43
"379" 43
"336" 43
"261" 43
"1091" 43
"1050" 43
"902" 42
"900" 42
"872" 42
"844" 42
"769" 42
"1098" 42
"107" 42
```

"503"	39
"373"	39
"370"	39
"16"	39
"1221"	39
"1217"	39
"1060"	39
"646"	38
"637"	38
"1210"	38
"1134"	38
"1089"	38
"1021"	38
"1013"	38
"937"	37
"783"	37
"583"	37
"41"	37
"364"	37
"30"	37
"1090"	37
"1008"	37
"856"	36
"796"	36
"670"	36
"1244"	36
"979"	35
"644"	35
"567"	35
"522"	35
"266"	35
"1059"	35
"1020"	35
"971"	34
"961"	34
"925"	34

"1411"	28
"1152"	28
"1129"	28
"1109"	28
"1042"	28
"978"	27
"905"	27
"842"	27
"805"	27
"734"	27
"681"	27
"645"	27
"613"	27
"564"	27
"464"	27
"46"	27
"457"	27
"389"	27
"348"	27
"1170"	27
"1070"	27
"1034"	27
"966"	26
"836"	26
"800"	26
"775"	26
"6"	26
"422"	26
"398"	26
"352"	26
"253"	26
"1284"	26
"1240"	26
"1218"	26
"1160"	26