INF-0617: Ex03, Hadoop Streaming

Students: Carlos Eduardo Fernandes and Yakov Nae

In this exercise we implement the electoral forecasting for the second round of elections in São Paulo state in 2014 using Hadoop Streaming. The database for this exercise can be downloaded here.

Running instructions:

This zip file contains a run.sh file with the following two fields:

```
#INITIALIZATION
INPUT_FOLDER="/tmp/data/votacao2014/"
WORK_FOLDER="/tmp/data/INF-0617_Ex03"
```

In order to run our code, please modify these two reference accordingly. The `INPUT_FOLDER` variable is the foder where the `votacao_secao_2014_SP.txt` file is at within the container. The `WORK_FOLDER` variable is where our code is at. Also, output files will be writen to this folder. After modifing these variables, please execute `$ ./run.sh`.

Mapper strategy

The input file contain election records comma separated where the general elections (12ve entry iquals to 3) is of following form:

```
"27/10/2014";"17:38:00";"2014";"1";"ELEIÇÕES GERAIS 2014";"SP";"SP";"63134";"CARAP
ICUÍBA";"388";"197";"3";"GOVERNADOR";"95";"40"
"27/10/2014";"17:38:00";"2014";"1";"ELEIÇÕES GERAIS 2014";"SP";"SP";"63134";"CARAP
ICUÍBA";"388";"197";"3";"GOVERNADOR";"96";"35"
```

The 14th entry in each line reffers to the voted party and the 15th is the number of votes. Our mapper maps the input file into `Map(Key,Value)` where the key is the party number and the value is the number of votes.

Reducer strategy

Our reducer strategy is to combine all keys (party number) by summing their values (number of votes). Shuffeling would try to direct the same key to the same reducer worker. However, even if two workers were summing the same key, they can later be reduced by this same strategy.

Results:

Here we show the finnel results of our program. Remembering that the 95 party code was converted to blank vote and 96 as Null vote. There were no 97 ("Voto Anulado e Apurado em Separado") party codes in the database:

```
13 - 3888584
15 - 4594708
21 - 12958
28 - 22822
29 - 11118
31 - 132042
43 - 260696
45 - 12230807
50 - 187487
Voto Branco - 2020613
Voto Nulo - 2374946
```