##### INF 2920 – Computação em Nuvem

## Trabalho 2 – WordCount

Eduardo de Oliveira Ferreira

**Tecnologias Utilizadas**

- Amazon Elastic MapReduce

- Python 2.7.2

- boto 2.0 - [Python interface to Amazon Web Services](http://code.google.com/p/boto/)

**Testes**

O trabalho foi executado em três testes, cada um com um número diferente de instâncias. O número de instâncias utilizadas foram: 1, 10 e 20. Durante o desenvolvimento do trabalho, foi observado que entrado do programa (as obras de Shakespeare) era relativamente pequena para que se pudesse observar uma diferença nítida. Devido a isso, esses testes foram rodados com entradas sessenta vezes maiores que todas as obras de Shakespeare. Dessa forma pode-se observar os seguintes tempos de execuções:

* 1 instância = 11 minutos
* 10 instância = 6 minutos
* 20 instancia = 6 minutos

Podemos perceber que por aumentarmos o número de instâncias, tivemos um ganho significativo de tempo. Porém o tempo de execução dos Jobs com 10 e 20 instâncias foram semelhante. Acredito que a diferença de tempo pudesse ficar mais definida com entradas maiores que a utilizado. Pois, sem sombra de dúvidas, quanto maior for o número de instâncias utilizadas em um job, menor será seu tempo de execução.

**Resultados**

- As 10 palavras mais usadas foram: *the* (27843), *and* (26847), *i* (22538), *to* (19883), *of* (18307), *a* (14800), *you* (13928), *my* (12489), *that* (11563), e *in* (11183).

- Shakespeare escreveu mais sobre o amor do que ódio. Foram 2198 ocorrências de *love* contra 179 de *hate.*

- Em relação a comentar sobre vida ou morte, Shakespeare escreveu 900 a palavra *death* e 863 *life*.

- Em relação sobre vida ou morte, Shakespeare escreveu 8760 a palavra *not* e 7147 *be*.

**Repositorio GitHub**

O código fonte e o relatório do trabalho estão em https://eduadroeof@github.com/eduadroeof/WordCount.git