



Big Data

Dashboard ► Mineração de Dados Complexos ► 2018 ► Venturus ► INF-0617-Venturus ►
Avaliações ► Trabalho 3: Hadoop Streaming

Trabalho 3: Hadoop Streaming

Trabalho 3: Apuração Eleitoral com Hadoop Streaming

Introdução

Neste trabalho utilizaremos os dados dos boletins de urna do TSE para contabilizar os votos na eleição para governador do Estado de São Paulo.

Preparação

Faça o download dos dados de votação por seção eleitoral no repositório do TSE, relativos ao primeiro turno do ano de 2014. Extraia o arquivo no diretório data do seu container, e leia a documentação do formato de dados.

Objetivo

Você deve implementar uma solução MapReduce para totalizar os votos para candidato a governador, considerando todas as seções eleitorais do estado. O resultado final deverá mostrar em cada linha de saída o número do candidato e o total de votos obtidos. A sua solução deve ser executada no container Spark da disciplina, usando Hadoop Streaming.

Entrega

Para a entrega, você deve fazer upload de um único arquivo .zip contendo o seu código Python, instruções sobre como executa-lo no container Docker, e um breve relatório em formato PDF ou TXT descrevendo resultados. O seu relatório deve incluir:

- Um parágrafo descrevendo a sua estratégia para o Mapper
- Um parágrafo descrevendo a sua estratégia para o Reducer
- O resultado final da apuração (saída do seu programa)

Observações

- Você deve usar o modelo de programação MapReduce, e executar a sua solução com Hadoop Streaming
- O trabalho pode ser feito em dupla.
- Teste o seu programa com `cat | map | reduce` **dentro** do container
- Para corrigir problemas com codificação dos arquivos, você pode usar o seguinte código

```
with io.open(sys.stdin.fileno(), 'r', encoding='latin-1') as sin:
    for line in sin:
        # seu código
```

Submission status

Submission status	No attempt
Grading status	Not graded
Due date	domingo, 21 outubro 2018, 11:55
Time remaining	1 day 10 hours
Last modified	terça, 9 outubro 2018, 10:15
Submission comments	► Comments (0)

Add submission

Make changes to your submission

NAVIGATION



Dashboard

■ Site home

Site pages

Current course

INF-0617-Venturus

Participants

Badges

Apresentações

Tutoriais

Aulas

Avaliações

 Notas e Frequências

 Trabalho 1: Processamento Paralelo Básico

 Trabalho 2: MapReduce Local

 **Trabalho 3: Hadoop Streaming**

 Trabalho 4: Complexidade de Vocabulário

My courses

ADMINISTRATION



Course administration