



Big Data

Dashboard ► Mineração de Dados Complexos ► 2018 ► Venturus ► INF-0617-Venturus ►
Avaliações ► Trabalho 2: MapReduce Local

Trabalho 2: MapReduce Local

Trabalho 2: MapReduce Local

Introdução

Neste trabalho você deve desenvolver uma solução em Python 3, baseada no modelo de programação MapReduce, para encontrar a maior variação de temperatura observada em uma única hora, dentre uma série de medições climáticas reportadas por estações climáticas nos Estados Unidos em 2017.

Cada arquivo no pacote de dados armazena medições coletadas em uma estação. Cada linha é um registro climático da estação em um certo dia e horário.

Por exemplo, a linha

```
04126 20170103 1700 20170103 1000 2.423 -113.56 43.46 -12.9 -13.2 -12.9 -13.9
9 0.0 44 0 69 0 18 0 C -12.8 0 -12.1 0 -13.9 0 85 0 -99.000 -9
9.000 -99.000 -99.000 -99.000 -1.0 -0.8 -9999.0 -9999.0 -9999.0
```

Indica que na estação com ID 04126, localizada em (-113.56, 43.46), às 17:00 UTC do dia 2017-01-03, a temperatura média da hora foi -13.2 graus C, a temperatura mínima foi -13.9 graus C, e a temperatura máxima foi -12.9 graus C.

A documentação do formato de dados, incluindo todas as colunas apresentadas, pode ser encontrada neste link. Note que valores -99 e -9999 são usados para indicar dados indisponíveis.

Objetivo

A sua solução deve imprimir o local e hora com a maior variação de temperatura no período de uma hora, definida como a diferença entre a temperatura máxima e temperatura mínima registrada para aquela hora. Como exemplo, a saída deve ter formato

```
2017-01-03 1700, (-113.56, 43.46), 1
```

indicando que a maior diferença de temperatura foi de um grau, observada no dia 2017-01-03, às 17:00 UTC, na estação localizada em (-113.56, 43.46).

Entrega:

Para a entrega, você deve fazer upload de um único arquivo .zip contendo o seu código Python, instruções sobre como executá-lo, e um breve relatório em formato PDF ou TXT descrevendo resultados. No relatório você deve indicar:

- Qual foi o resultado final.
- Qual foi a sua estratégia para implementação do Mapper.

- Qual foi a sua estratégia para implementação do Reducer.
- Como o desempenho poderia ser melhorado no futuro.

Observações

- Você deve usar o modelo de programação MapReduce. Veja os slides de aula para sugestões sobre como estruturar a sua solução.
- O trabalho pode ser feito em dupla.
- Você deve processar todos os arquivos climáticos fornecidos.
- Ignore registros indisponíveis (-99, ...)
- Os dados registrados apresentam erros, como por exemplo valores incorretos. Encontre maneiras de filtrar registros inválidos no seu Mapper.

Submission status

Submission status	No attempt
Grading status	Not graded
Due date	domingo, 14 outubro 2018, 11:55
Time remaining	5 days 1 hour
Last modified	quinta, 4 outubro 2018, 9:39
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

My courses

ADMINISTRATION



Course administration

You are logged in as Yakov Nae (Log out)
INF-0617-Venturus