

ARA0168 – TÓPICOS DE BIG DATA EM PYTHON

Exercício 1 de NOTA FINAL

1. (Sala de Aula Virtual) Informe quais são os principais elementos que compõe a **Arquitetura da Big Data** e escolha um desses elementos e descreva-o com detalhes. (0,2 pontos)
2. Descreva quais são os 5's características da *Big Data* e a definição de cada um deles. (0,2 pontos)
3. Escolha uma área (fora da área de tecnologia) em que o uso do *Big Data* tem um papel fundamental na sua evolução no cenário atual (0,3 pontos).
4. Descreva o que se trata um **Data Lake** e sobre seus principais benefícios. (0,2 pontos)
5. Apresente (além dos conceitos vistos em Sala de Aula) as principais definições de **metadados** e sua importância em *Big Data* e *Machine Learning*. (0,2 pontos)
6. Do que se trata a sigla **HDFS** dentro do Ecossistema *Hadoop*? Elenque também suas principais características. (0,2 pontos)
7. Seja o seguinte script Python. Descreva com detalhes o que está sendo executado (0,3 pontos):

```
import json
from collections import Counter

arquivo = open("file1.txt", "r")
aux1 = arquivo.read()
aux1 = json.loads(aux1)

arquivo = open("file2.txt", "r")
aux2 = arquivo.read()
aux2 = json.loads(aux2)

aux_reduce = Counter(aux1) + Counter(aux2)
```

8. Faça um script Python que **leia um arquivo** (pode ser *txt*, *cvs*, *json*, etc) que, tendo o nome e sobrenome de alunos de uma turma, insira em um dicionário a contagem de nomes desses alunos (0,4 pontos). Exemplo:

Se o arquivo lido contém:

```
maria silva
jose Souza
joão Antunes
maria nogueira
```

Então deve ser inserido no dicionário:

```
{ 'maria': 2, 'jose': 1, 'joao': 1 }
```

Observações:

1. Este é o 1º Exercício que irá compor a Nota Final da disciplina ARA0168 e vale 2,0 pontos.
2. O exercício deve ser respondido em arquivo formato pdf e suas respostas enviadas para o campo de trabalho da Sala Virtual da disciplina ARA0168, no dia 19/10/2022, no horário de aula.
3. A interpretação e organização das questões faz parte do processo avaliativo.