

## Orientações para elaboração do Trabalho Interdisciplinar Big Data + ML

2º semestre 2021 – Prof. Bianca

### Detalhamento da parte do trabalho referente à Big Data

Processamento de algoritmos de ML em ambiente distribuído, formado por um cluster de máquinas rodando hadoop map reduce/spark. Estes ambientes podem ser encontrados na AWS e Google, entre outros.

#### 1. Organização do conteúdo do trabalho

Os trabalhos devem ter os seguintes tópicos:

- Introdução, que deve ser uma breve contextualização.
- Descrição dos dados (fonte, período, volume, formato, amostras dos dados)
- Workflow/pipeline de arquitetura
- Infraestrutura (Configuração do cluster como nro de nós, tipo de máquinas, permissões)
- Análise do Processamento (análise do desempenho das operações ETL, segundo critérios como tempo de processamento, formas intermediárias de armazenaneto dos dados (partições), comparação do desempenho com diferentes configurações de cluster)
- Funções definidas para processamento distribuído (map, reduce, outras)
- Conclusões (limitações e vantagens)
- Referências

#### 2. Forma de entrega e apresentação

O trabalho deve ser publicado em um notebook ou no github. No github organizar pastas com os mesmos nomes do item 1.

Enviar o link para este notebook/GIT via **MOODLE até 12 de dezembro**, véspera da data em que o grupo fará uma breve apresentação para a turma.

##### Quanto à apresentação do trabalho

- Todos os membros do grupo devem estar presentes no momento da apresentação.
- Cada apresentação terá duração de 15 minutos

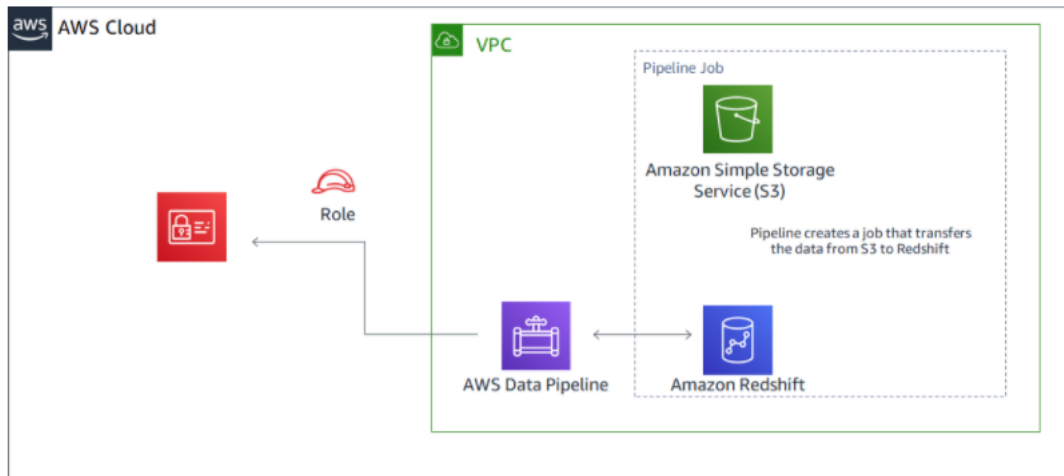
Critérios de Avaliação da parte específica de Big Data, correspondente a 60% da nota final

Parte Específica de Big Data		
Introdução	0,5	6,0
Dados	0,5	
Workflow de Arquitetura	0,5	
Infraestrutura	1,0	
Análise do Processamento	1,0	
Funções	1,0	
Conclusões	1,0	
Referências	0,5	
Parte Geral	Parte geral	
Demo	2,0	4,0
Arguição	1,0	
Slides	1,0	

Exemplos de pipeline (google cloud) ou workflow de arquitetura (AWS).

O Workflow de Arquitetura mostra quais as ferramentas/recursos AWS serão usados e em qual sequencia.

O site <https://online.visual-paradigm.com/> oferece templates para criação de workflows da AWS.



Referências:

Seguem alguns notebooks para inspiração de README(observar mais a forma q o conteúdo)

<https://github.com/Mgosi/Big-Data-Analysis-using-MapReduce-in-Hadoop#readme>

<https://github.com/faiderfl/Big-Data--architecture-aws-spark#readme>