

# Trabalho Prático do Módulo 2

**Entrega** 13 nov em 23:59

**Pontos** 25

**Perguntas** 15

**Disponível** até 13 nov em 23:59

**Limite de tempo** Nenhum

**Tentativas permitidas** 2

## Instruções

O Trabalho Prático do Módulo 4 está disponível!

### 1. Instruções para realizar o trabalho prático

Consulte a data de entrega no teste e em seu calendário.

Reserve um tempo para realizar a atividade, leia as orientações e enunciados com atenção. Em caso de dúvidas utilize o "Fórum de dúvidas do Trabalho Prático do Módulo 4".

Para iniciá-lo clique em "Fazer teste". Você tem somente **uma** tentativa e não há limite de tempo definido para realizá-lo. Caso precise interromper a atividade, apenas deixe a página e, ao retornar, clique em "Retomar teste".

Clique em "Enviar teste" **somente** quando você concluí-lo. Antes de enviar confira todas as questões.

Caso o teste seja iniciado e não enviado até o final do prazo de entrega, a plataforma enviará a tentativa não finalizada automaticamente, independente do progresso no teste. Fique atento ao seu teste e ao prazo final, pois novas tentativas só serão concedidas em casos de questões médicas.

O gabarito será disponibilizado a partir de domingo, **13/11/2022**, às 23h59.

Bons estudos!

### 2. O arquivo abaixo contém o enunciado do trabalho prático

**Enunciado do Trabalho Prático - Módulo 2 - Engenheiro(a) de Dados Cloud.pdf**

Este teste foi indisponível 13 nov em 23:59.

## Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	<u>Tentativa 1</u>	2.791 minutos	25 de 25

Pontuação desta tentativa: **25** de 25

Enviado 12 nov em 6:48

Esta tentativa levou 2.791 minutos.

### Pergunta 1

1,67 / 1,67 pts

Qual é a característica do Apache Spark em relação ao tratamento das computações intermediárias que faz com que ele seja mais rápido do que o Hadoop MapReduce?

Correto!

- ☒ Armazenamento dos resultados intermediários em memória.
- ☐ Utilização do lazy evaluation na realização das operações.
- ☐ Reorganização das computações no plano físico de execução.
- ☐ Realização de shuffles de dados.

### Pergunta 2

1,67 / 1,67 pts

Qual a principal diferença entre as transformações narrow e wide?

Correto!

- ☒ As transformações narrow não realizam shuffle (movimentação) de dados, enquanto as transformações wide realizam.
- ☐ As transformações narrow não ativam o histórico do lazy evaluation, enquanto as transformações wide ativam.
- ☐ As transformações wide não realizam shuffle (movimentação) de dados, enquanto as transformações narrow realizam.
- ☐ As transformações wide não ativam o histórico do lazy evaluation, enquanto as transformações narrow ativam.

**Pergunta 3****1,67 / 1,67 pts**

Qual dos blocos de código abaixo realiza o empilhamento de dois DataFrames df1 e df2?

**Correto!**

- ☒ df1.union(df2).
- ☐ df1.add(df2).
- ☐ df1.append(df2).
- ☐ df1.concat(df2).

**Pergunta 4****1,67 / 1,67 pts**

Quantos filmes (incluindo os da televisão) foram lançados no ano de 2015?

**Correto!**

- ☒ 19987.
- ☐ 3558.
- ☐ 16429.
- ☐ 356877.

**Pergunta 5****1,67 / 1,67 pts**

Qual o gênero de títulos mais frequente?

**Correto!**

- ☒ Drama.
- ☐ Comedy.

☐ Documentary.☐ War.**Pergunta 6****1,67 / 1,67 pts**

Qual o gênero com a melhor nota média de títulos?

**Correto!**☒ History.☐ Drama.☐ Documentary.☐ Crime.**Pergunta 7****1,67 / 1,67 pts**

Qual o vídeo game do gênero aventura mais bem avaliado em 2020?

**Correto!**☒ Half-Life: Alyx.☐ Final Fantasy VII Remake.☐ Omori.☐ Ghost of Tsushima.**Pergunta 8****1,67 / 1,67 pts**

Qual das seguintes operações NÃO é uma ação?

**Correto!**

- ☒ `filter()`.
- ☐ `count()`.
- ☐ `collect()`.
- ☐ `saveAsTextfile()`.

**Pergunta 9****1,67 / 1,67 pts**

Quantos títulos de filmes diferentes existem? Use `df_titles.select('primaryTitle').distinct().count()`.

**Correto!**

- ☒ 3931670.
- ☐ 8203690.
- ☐ 5641200.
- ☐ 1298121.

**Pergunta 10****1,67 / 1,67 pts**

Qual a duração média dos filmes com conteúdo adulto? Use uma combinação de `filter()` e `describe()`.

**Correto!**

- ☒ 92,79 minutos.
- ☐ 57,18 minutos.
- ☐ 1 minuto.
- ☐ 43,78 minutos.

**Pergunta 11****1,67 / 1,67 pts**

Quantos filmes têm o título atual (“primary”) diferente do título original? Use uma combinação de *filter*, e *count()*.

**Correto!**☒ 125056.☐ 8078634.☐ 1987121.☐ 256871.**Pergunta 12****1,67 / 1,67 pts**

Qual o filme que tem o nome mais longo?

Dica: consulte <https://sparkbyexamples.com/spark/spark-using-length-size-of-a-dataframe-column/>

(<https://sparkbyexamples.com/spark/spark-using-length-size-of-a-dataframe-column/>) e use algo como `df_titles.orderBy(length(col("primaryTitle")).desc())`.

**Correto!**☒ tt12985206.☐ tt13007080.☐ tt7727908.☐ tt0043766.**Pergunta 13****1,67 / 1,67 pts**

Qual filme tem a maior quantidade de votos? Dica: Use *describe()*.

**Correto!**☒ tt0111161.☐ tt0110570.☐ tt0110557.☐ tt0110613.**Pergunta 14****1,67 / 1,67 pts**Qual a principal estratégia adotada por plataformas de *big data*?**Correto!**☒ Distribuir a computação em várias máquinas.☐ Construir máquinas mais poderosas.☐ Amostrar os dados e trabalhar com menor quantidade de dados.☐ Paralelizar a computação na mesma máquina.**Pergunta 15****1,62 / 1,62 pts**Qual é a menor nota média de um filme? Use *describe()*.**Correto!**☒ 1.0☐ 2.0☐ 3.0☐ 10.0Pontuação do teste: **25** de 25

