

# DESM4 - Desafio do Módulo 4

Entrega 17 jun em 23:59

Pontos 45

Perguntas 15

Disponível até 17 jun em 23:59

Limite de tempo Nenhum

## Instruções



**Desafio do Módulo**

É hora de mostrar que você aprendeu!  
O Desafio é um teste com 15 questões objetivas e **deve ser realizado com atenção!**

Reserve um tempo para realizar a atividade, leia as orientações e enunciados com atenção. Em caso de dúvidas utilize o Fórum de Dúvidas.

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

O gabarito será disponibilizado a partir de segunda-feira, **17/06/2024**, às 23h59.

- O arquivo abaixo contém o enunciado do trabalho prático. Confira agora:

### Enunciado do Desafio - Módulo 4 - Arquiteto(a) de Big Data.pdf

<https://online.igti.com.br/courses/7232/files/549863?wrap=1> ↓

[https://online.igti.com.br/courses/7232/files/549863/download?download\\_frd=1](https://online.igti.com.br/courses/7232/files/549863/download?download_frd=1)

**Bons estudos!**

**Atenciosamente,**

**Equipe XP Educação**



# Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	<a href="#">Tentativa 1</a>	64 minutos	45 de 45

⚠ As respostas corretas estarão disponíveis em 17 jun em 23:59.

Pontuação deste teste: **45** de 45  
Enviado 3 jun em 20:33  
Esta tentativa levou 64 minutos.

Pergunta 1

3 / 3 pts

Quantas linhas existem no dataset Case.csv?

☐ 170

☐ 200

☒ 174

☐ 164

Pergunta 2

3 / 3 pts

No dataset Case.csv, assinale a alternativa VERDADEIRA.

☐ province possui apenas 2 valores únicos

☒ group possui apenas 2 valores únicos

☐ city possui apenas 2 valores únicos

☐ infection\_case possui apenas 2 valores únicos

**Pergunta 3****3 / 3 pts**

No dataset Case.csv, a média de confirmed é:

- ☐ 11.89
- ☐ 31.50
- ☐ 101.99
- ☒ 65.49

**Pergunta 4****3 / 3 pts**

No dataset Case.csv, a cidade (city) que apresentou mais casos (confirmed) foi:

- ☐ Dalseong-gun
- ☐ from other city
- ☒ Nam-gu

**Pergunta 5****3 / 3 pts**

No dataset Case.csv, a média de casos (confirmed), na província (province) de Seoul foi de:

- ☐ 43.89
- ☐ 110.20
- ☐ 65.50

☒ 33.68

### Pergunta 6

3 / 3 pts

Quantas linhas existem no dataset PatientInfo.csv?

☐ 5265

☐ 10415

☒ 5165

☐ 10410

### Pergunta 7

3 / 3 pts

No dataset PatientInfo.csv, assinale a alternativa VERDADEIRA.

☐ sex possui apenas 2 valores únicos

☐ infection\_case possui apenas 3 valores únicos

☒ state possui apenas 3 valores únicos

☐ state possui apenas 2 valores únicos

### Pergunta 8

3 / 3 pts

No dataset PatientInfo.csv, pessoas do sexo (sex) feminino (female), da cidade (city) Jongno-gu, e do grupo de idade (age) 10s são:

☐ 1

☒ 2☐ 3☐ 0**Pergunta 9****3 / 3 pts**

Quantas linhas existem no dataset PatientRoute.csv?

☒ 10410☐ 5165☐ 5265☐ 10415**Pergunta 10****3 / 3 pts**

No dataset PatientRoute.csv, a maior latitude encontrada é de:

☒ 38.19317☐ 37.61525☐ 36.95589☐ 33.45464**Pergunta 11****3 / 3 pts**

No dataset PatientRoute.csv, a menor longitude encontrada é de:

☐ 127.434☒ 126.301☐ 129.476☐ 126.716**Pergunta 12****3 / 3 pts**

Pacientes detectados (type) no aeroporto (airport), e que tiveram caso de infecção (infection\_case) como contato com pacientes (contact with patient) retornam uma contagem de:

☐ 5 ocorrências☒ 20 ocorrências☐ Não é possível calcular☐ 10 ocorrências**Pergunta 13****3 / 3 pts**

Assinale a questão INCORRETA:

☐

O Spark apresenta velocidade de processamento de até 100x mais rápido que o MapReduce do Hadoop.

☐

O Spark não precisa utilizar o HDFS. Ele possui seu próprio modo de armazenamento, o RDD.



O Spark compartilha dados entre clusters através do Spark Dataframe e por isso, esse é o tipo recomendado a se trabalhar.



O Spark pode armazenar os dados apenas localmente ou com outros clusters também.

### Pergunta 14

3 / 3 pts

O PySpark suporta as seguintes linguagens, EXCETO:



Scala



Java



JavaScript



Python

### Pergunta 15

3 / 3 pts

Assinale a alternativa CORRETA:



O Spark é projetado para armazenar apenas dados acima de 100 GB



Não é possível testar o Spark em um único cluster apenas



O Spark não oferece suporte a machine learning, oferece apenas acesso, leitura e armazenamento de dados



O Spark oferece grande vantagem, processando em memória RAM e ainda podendo ser incorporado a notebooks

Pontuação do teste: **45** de 45