

1

Marcar para revisão

O MapReduce é uma técnica clássica de programação e é bastante utilizada por diversos frameworks, como o Spark, por exemplo. Em relação à técnica MapReduce, selecione a opção correta.

A

É uma técnica de computação distribuída

B

Só pode ser aplicada para grandes volumes de dados

C

Foi substituída no Spark por acesso direto à memória

D

É uma técnica lenta para processamento de grandes volumes de dados

E

Consiste em uma técnica de programação sequencial

00 : 44 : 12

hora min seg



Ocultar

Questão 1 de 10

1

2

3

4

5

6

7

8

9

10

● Respondidas (10) ● Em branco (0)

Finalizar prova



Feedback

2

Marcar para revisão

Selecione a opção correta que contenha a operação responsável por retornar um subconjunto aleatório dos dados de entrada.

☒ A sample☐ B distinct☐ C random☐ D mapRandom☐ E filter

3

Marcar para revisão

Atualmente, o Apache Spark é o mais bem-sucedido framework para Big Data. Selecione a opção correta que contenha apenas funções que podem ser categorizadas como transformações no Spark.

A map, sample e collect.

B count, collect e take.

C reduce, first e map.

D map, filter e union.

E map, take e reduce.

4

Marcar para revisão

Eduardo quer mostrar aos gestores da empresa como as ações da carteira de investimentos recomendada que ele e o time dele propuseram estão valorizando e gerando muitos dividendos aos clientes. Para isso, Eduardo construiu um gráfico de \_\_\_\_\_, pois está lidando com dados \_\_\_\_\_ e quer mostrar o quão bem as recomendações geradas por seu time estão evoluindo.

Respectivamente a resposta que completa as duas lacunas é:



☐ A Linha; Categóricos

☒ B Linha; Temporais

☐ C Linha; Numéricos

☐ D Pizza; Numéricos

☐ E Pizza; Categóricos

5

Marcar para revisão

No mundo de Big Data, temos dados os mais variados possíveis, e sua classificação é fundamental para um cientista de dados. As notas que compõem o histórico escolar de um aluno são classificadas como que tipo de dados?

☐ A Atemporais

☐ B Numéricos

☐ C Semi-Estruturados



D Temporais

E Categóricos

6

Marcar para revisão

Gabriel entrou recentemente em uma empresa de gestão de fundos de investimento. No seu primeiro dia de trabalho, lhe é conferida a tarefa de buscar os relatórios do pessoal da contabilidade para fazer suas análises exploratórias. O departamento de contabilidade extrai os dados de seus relatórios de um SAP que retorna valores brutos de compras, quantidades de produtos etc. Então os contadores organizam, por agregação, informação relevante no relatório em planilhas por eles desenvolvidos. Por questões de acesso e segurança, Gabriel só pode lidar com essas planilhas. Este tipo de coleta feita por Gabriel e pelos Contadores é respectivamente dos tipos:



**A** Secundária; Primária.

B Primária; Primária.

C Secundária; Terciária.

D Primária; Secundária.

E Secundária;  
Secundária.

7

Marcar para revisão

As técnicas de aprendizado de máquina para aprendizado supervisionado são:

I - Classificação

II - Agrupamento

III - Regressão

**A** Apenas I e III

B Apenas II e III

C Apenas I



☐ D Apenas I e II

☐ E Apenas II

8

Marcar para revisão

Dentre as inúmeras bibliotecas da linguagem Python, qual destas é a que abstrai conceitos de engenharia de aprendizado de máquina e implementa modelos de ML?

☐ A Pandas

☒ B Scikit-Learn

☐ C BeautifulSoup

☐ D OS

☐ E Numpy



9

Marcar para revisão

O PySpark caracteriza-se como uma biblioteca Spark responsável por executar programas usando recursos do Apache Spark. Selecione a

opção correta que contenha itens que são pré-requisitos de instalação para utilização do PySpark.

☐ A Java e R

☐ B Python e Escala

☐ C Hadoop e Spark

☐ D Casandra e Spark

☒ E Java e Python

10

Marcar para revisão

A linguagem Python fornece para o cientista de dados uma grande variedade de funções e bibliotecas. Qual o método que faz a projeção do Dataframe orientado pelos rótulos das colunas do conjunto?

☐ A iloc

☐ B shuffle

☐ C sum





D loc

E reset\_index

