

Avaliação - Big Data Foundations - Prazo Final dia 21/02 as 17h

Iniciado: 17 fev em 8:38

Instruções do teste

Pergunta 1

0,5 pts

O que é o Apache Hadoop?

- ☐ Hardware para alto desempenho
- ☐ Software para computação distribuída
- ☐ Hardware para Big Data
- ☐ Software para analytics

Pergunta 2

0,5 pts

Como é a arquitetura de armazenamento do HDFS?

- ☐ Principal/Secondary
- ☐ Namenode/Datanode
- ☐ Jobtracker/Tasktracker
- ☐ ResourceManager/NodeManager

Pergunta 3

0,5 pts

A arquitetura do HDFS assume que os datanodes **não** são 100% confiáveis

☐ Verdadeiro

☐ Falso

Pergunta 4

0,5 pts

Quem desenvolveu o modelo de programação MapReduce?

☐ Apache Foundation

☐ Google

☐ Microsoft Research

☐ IBM Research

Pergunta 5

0,5 pts

Quais os componentes do MapReduce para processamento dos jobs?

☐ ResourceManager/NodeManager

☐ Namenode/Datanode

☐ Principal/Secondary

☐ Jobtracker/Tasktracker

Pergunta 6**0,5 pts**

Não faz parte do ecossistema Hadoop:

☐ HBase

☐ Hive

☐ Sqoop

☐ Python

Pergunta 7**0,5 pts**

Qual node pode ser instalado em um cluster Hadoop para diminuição do uso de recursos do NameNode?

☐ Secondary Node

☐ Standby Node

☐ Edge Node

☐ Slave Node

Pergunta 8**0,5 pts**

Quais são os componentes do YARN para coordenar e executar os jobs?

☐ Principal/Secondary

- ☐ ResourceManager/NodeManager
- ☐ Jobtracker/Tasktracker
- ☐ Namenode/Datanode

Pergunta 9**0,5 pts**

O que é Data Lake?

- ☐ Repositório de armazenamento e engine para processamento de Big Data
- ☐ Repositório de armazenamento para dados Estruturados e Semi estruturados
- ☐ Ingestão de dados distribuídos e escalável para Big Data
- ☐ Visualização de insight para Big Data

Pergunta 10**0,5 pts**

Qual ferramenta não é responsável pela ingestão de dados?

- ☐ HBase
- ☐ Kafka
- ☐ Sqoop
- ☐ Flume

Pergunta 11**0,5 pts**

Caso exista um arquivo no diretório do hdfs

"/user/hive/warehouse/db_test/employees", o que irá acontecer com a importação do Sqoop abaixo:

```
sqoop import --table employees --connect jdbc:mysql://database/employees --username root --password secret --warehouse-dir /user/hive/warehouse/db_test -m 4 -delete-target-dir --fields-terminated-by "\t" --lines-terminated-by ";"
```

- ☐ A importação irá falhar
- ☐ A importação irá deletar o diretório existente db_test
- ☐ A importação irá deletar o diretório existente employees
- ☐ A importação irá adicionar os arquivos juntamente com os arquivos já existentes

Pergunta 12

0,5 pts

O que irá acontecer com a importação do Sqoop abaixo em relação a tabela titles no hdfs existente:

```
sqoop import --table titles --connect jdbc:mysql://localhost/sakila --username=root --password=secret --warehouse-dir db_test -m 1 --incremental lastmodified --merge-key id --check-column date --last-value '2021-01-20'
```

- ☐ A importação irá anexar apenas os dados do campo id da tabela titles com os dados novos no período de 20/01/2021 até hoje
- ☐ A importação irá atualizar apenas o campo id da tabela titles com os dados novos no período de 20/01/2021 até hoje
- ☐ A importação irá anexar os dados novos no diretório db_test no período de 20/01/2021 até hoje
- ☐ A importação irá atualizar a tabela titles com os dados novos no período de 20/01/2021 até hoje

Pergunta 13

0,5 pts

O Hive é um sistema de gerenciamento de banco de dados?

☐ Verdadeiro

☐ Falso

Pergunta 14

0,5 pts

Onde são armazenados os metadados do Hive?

☐ WebHCat

☐ Warehouse

☐ Metastore

☐ HCatalog

Pergunta 15

0,5 pts

Quando apagamos uma tabela externa no Hive,

☐ os metadados não são deletados

☐ os dados são deletados

☐ os dados são deletados

☐ os dados não são deletados

Pergunta 16

0,5 pts

Qual das opções abaixo melhora o tempo de uma consulta no hive de um arquivo .csv?

- ☐ Converter o arquivo para Sequence File
- ☐ Comprimir o arquivo para Snappy
- ☐ Nenhuma das opções
- ☐ Converter o arquivo para Parquet

Pergunta 17

0,5 pts

Para qual tipo de dados foi criado o armazenamento do HBase?

- ☐ Dados semi estruturados
- ☐ Dados relacionais
- ☐ Dados não relacionais
- ☐ Dados estruturados

Pergunta 18

0,5 pts

O que o comando do HBase abaixo está realizando na tabela cliente:

```
get 'cliente','5',{COLUMN=>['pedido:num_pedido'],VERSIONS=>3,LIMIT=>10}
```

- ☐ Pesquisar os valores das 3 últimas versões da família de coluna num_pedido da chave 5
- ☐ Pesquisar o valor da 3 versão da coluna num_pedido da chave 5
- ☐ Pesquisar o valor da 3 versão da família de coluna num_pedido da chave 5
- ☐ Pesquisar os valores das 3 últimas versões da coluna num_pedido da chave 5

Pergunta 19**0,5 pts**

Qual opção está **incorreta** sobre o Spark?

- ☐ Usado para criação de gráficos de Big Data
- ☐ Usado para Machine Learning
- ☐ Usado para ETL
- ☐ Usado para processamento batch e stream

Pergunta 20**0,5 pts**

Segue uma parte do arquivo "teste.csv" abaixo:

id, nome, valor

1, mouse, 20.50

2, teclado, 50.60

...

5000, cabo USB, 5.00

Com uso do spark para leitura do arquivo "teste.csv", como o schema será apresentado com o código abaixo:

```
val data = spark.read.option("inferSchema", "true").option("header",  
"true").csv("teste.csv").printSchema()
```

- ☐ id: integer (nullable = true)
nome: string (nullable = true)
valor: integer (nullable = true)
- ☐ id: string (nullable = true)
nome: string (nullable = true)

valor: string (nullable = true)

☐ id: integer (nullable = true)

nome: string (nullable = true)

valor: double (nullable = true)

☐ id: integer (nullable = true)

nome: string (nullable = true)

valor: float (nullable = true)

Não salvo

Enviar teste