Questão 1/10 - Big Data

O Kafka consiste e um uma plataforma distribuída e, dessa forma, é executado através de um cluster. Um Cluster Kafka armazena fluxos de eventos que são divididos em categorias denominadas tópicos.  
  
A respeito dos conceitos de eventos e tópicos em Kafka avalie as seguintes afirmações:  
  
I – Cada evento é constituído de um par chave-valor.

II – Os servidores Kafka são capazes de armazenar a indicação de qual a última mensagem que cada cliente inscrito em um tópico recebeu.

III – As aplicações cliente podem se inscrever em um ou mais tópicos.

IV – O Kafka é capaz de implementar sistemas de mensagens baseados em publicação-inscrição, mas não é capaz de entregar a implementação de uma fila de mensagens, uma vez que os dados dos eventos são armazenados de forma permanente e permanecem disponíveis.

V – O Kafka permite a implementação de um sistema de armazenamento a partir do armazenamento dos dados em disco. Outra característica é que os dados armazenados no Kafka são replicados para garantir o princípio da tolerância a falhas.

Dentre essas afirmações estão CORRETAS:

Você não pontuou essa questão

|  |  |  |
| --- | --- | --- |
|  | A | II, III e V |
|  | B | I, II e III |
|  | C | Todas |
|  | D | I, III e IV |
|  | E | Todas, exceto IV  Você assinalou essa alternativa (E) |

Questão 2/10 - Big Data

Uma vez que conhecemos os conceitos que envolvem a arquitetura big data assim como as aplicações que implementam suas características, podemos pensar em como combinar todas essas tecnologias e ideias para o desenvolvimento de um produto visando atender a alguma necessidade existente.  
  
Avalie as seguintes afirmações a respeito do desenvolvimento e design de soluções Big Data:  
  
I – O desenvolvimento de aplicações Big Data possui grandes desafios em como combinar todas as suas tecnologias e ideias para desenvolver novos produtos. Porém a integração com tecnologias tradicionais é uma questão trivial atualmente.

II – Em muitos casos as soluções de Big Data têm como objetivo otimizar processos de negócios, adquirir vantagens competitivas ou otimizar operações. Dessa forma podem representar muitas vantagens para uma empresa.

III – O entendimento das necessidades que nossa aplicação busca atender é uma das maneiras mais importantes de se obter as informações necessárias para guiar o desenvolvimento

IV – A documentação de requisitos é capaz de evitar desperdícios e garantir que a implementação está alinhada com as necessidades do usuário.

V – Working Backwards é uma abordagem de desenvolvimento de produtos que prioriza o entendimento dos requisitos e o feedback do usuário antes mesmo do início do desenvolvimento.

Dentre essas afirmações estão CORRETAS:

|  |  |  |
| --- | --- | --- |
|  | A | II, III e IV |
|  | B | I, IV e V |
|  | C | I, II e IV |
|  | D | Todas, exceto I  Você assinalou essa alternativa (D) |
|  | E | Todas |

Questão 3/10 - Big Data

Analisar os requisitos de um projeto é um dos pontos mais importantes no desenvolvimento de aplicações, uma vez que tais requisitos devem refletir todas as necessidades e críticas dos usuários.  
  
A respeito dos requisitos de aplicações de Big Data, avalie as seguintes afirmações:  
  
I – Deve-se utilizar uma solução com tecnologias de Big Data para qualquer tamanho de dados, uma vez que as vantagens da aplicação de tais tecnologias são vantajosas em diversos cenários.

II – Pode ser muito caro mover os dados armazenado em um determinado sistema de armazenamento para outro lugar, uma vez que o local de armazenamento foi definido. Por isso deve-se tentar prever cenários futuros da melhor forma possível.

III – Escalar a complexidade de um sistema para atender uma demanda além do previsto é uma boa estratégia que pode auxiliar na antecipação de custos de manutenção futuros. É sempre uma boa estratégia, avaliar qual é o máximo de infraestrutura necessária para operar a aplicação.

IV – Em muitos casos pode ser necessário manter os dados armazenados por um tempo maior por motivos de auditoria ou por necessidades jurídicas. Por isso é importante a definição de uma política de retenção de dados.

V – É importante avaliar a relação do sistema com o teorema CAP. É sempre necessário saber se devemos priorizar a disponibilidade ou a consistência dos dados.

Dentre essas afirmações estão INCORRETAS:

Você não pontuou essa questão

|  |  |  |
| --- | --- | --- |
|  | A | I, II e V |
|  | B | I e III |
|  | C | Nenhuma  Você assinalou essa alternativa (C) |
|  | D | II, III e IV |
|  | E | II, IV e V |

Questão 4/10 - Big Data

Para que o valor dos dados contidos em um Data Lake não seja perdido, se faz necessário criar uma solução de Data Lake que inclua as noções de gerenciamento, acessibilidade e governança.  
  
Avalie as seguintes afirmações a respeito do gerenciamento de Data Lake:  
  
I – Um Data Lake oferece o mesmo nível de complexidade que outras estratégias de armazenamento como Data Mart e Data Warehouse.

II – Uma das características que podem ser agregadas aos dados armazenados em uma estratégia de Data Lake utilizando bons esquemas de governança é a transparência no uso dos dados.

III – Entre as informações que se podem obter dos dados em um Data Lake podemos destacar os metadados técnicos que fornecem informações sobre a qualidade, perfil, origem, e a sua linhagem.

IV – Os metadados técnicos de um Data Lake podem fornecer informações a respeito da forma ou estrutura dos dados tais como tamanho, tipo de dado ou esquema.

V – Os metadados de um Data Lake permitem buscar, localizar e aprender sobre os dados armazenados.

Dentre essas afirmações estão CORRETAS:

Você não pontuou essa questão

|  |  |  |
| --- | --- | --- |
|  | A | II, IV e V |
|  | B | Todas, exceto IV |
|  | C | I, II e V |
|  | D | III, IV e V |
|  | E | Todas  Você assinalou essa alternativa (E) |

Questão 5/10 - Big Data

Recomendações se baseiam nas previsões de preenchimento das avaliações desconhecidas. Existem duas principais abordagens para isso: recomendações baseadas em conteúdo (*content-based recommendations*) e filtragem colaborativa (*collaborative filtering*).  
  
Avalie as seguintes avaliações sobre as duas principais abordagens de recomendações:  
  
I – O sistema de recomendações baseadas em conteúdo possuem como estratégia a recomendação de itens mais parecidos aos que o usuário avaliou.

II – Antes de recomendar um novo item, o sistema de recomendações baseadas em conteúdo precisa que algum usuário avalie o novo item.

III – O cálculo de similaridade entre os usuários é utilizado por sistemas de filtragem colaborativa para medir a chance de um par de usuários avaliar determinado item de forma parecida.

IV – Métodos de filtragem colaborativa pode ser aplicado apenas para medir a semelhança entre usuários.

V – É possível utilizar métodos de diferentes abordagens de recomendação de uma forma híbrida para tentar melhorar a qualidade das predições.

Dentre essas afirmações estão INCORRETAS:

|  |  |  |
| --- | --- | --- |
|  | A | Todas, exceto V |
|  | B | I e III |
|  | C | Nenhuma |
|  | D | II e IV  Você assinalou essa alternativa (D) |
|  | E | II e V |

Questão 6/10 - Big Data

Avalie as seguintes afirmações sobre os componentes do Hadoop: Impala e Accumulo:  
  
I – Impala é um motor de consultas SQL capaz de realizar consultas de baixa latência em HDFS ou Hive.

II – O Impapa foi projetado para consultas analíticas em Haddop utilizando SQL ou ferramentas de *business inteligence* (BI).

III – Accumulo é um sistema de armazenamento distribuído baseado em chave-valor que implementa segurança e nível de célula.

IV – Todos os dados armazenados no Accumulo devem possuir os mesmos requisitos de segurança em uma mesma tabela.

V – O Accumulo permite realizar operações em pares de chave-valor assim que são inseridos.

Dentre essas afirmações estão CORRETAS:

Você não pontuou essa questão

|  |  |  |
| --- | --- | --- |
|  | A | I, II e III  Você assinalou essa alternativa (A) |
|  | B | Todas |
|  | C | I, II e IV |
|  | D | II, III e V |
|  | E | Apenas III |

Questão 7/10 - Big Data

O aprendizado de máquina é parte do campo da Inteligência Artificial que se refere ao estudo de modelos estatísticos para resolver problemas específicos com padrões e inferências.  
  
Sobre as diferentes categorias de aprendizado de máquina avalie as seguintes afirmações:  
  
I – Os algoritmos de classificação e de regressão fazem parte da categoria de algoritmos de aprendizado supervisionado.

II – Os algoritmos de aprendizado supervisionado buscam identificar estruturas inerentes ao conjunto de dados.

III – O que diferencia os algoritmos de aprendizado por reforço dos algoritmos de aprendizado não-supervisionado é que no aprendizado por reforço os dados nunca são explicitamente relacionados com os seus rótulos.

IV – Os algoritmos de aprendizado por reforço se baseiam nos conceitos de reforço e punição para maximizar parâmetros através da tomada de ações.

V – Os algoritmos de clustering fazem parte da categoria de aprendizagem não-supervisionados.

Dentre essas afirmações estão CORRETAS:

Você não pontuou essa questão

|  |  |  |
| --- | --- | --- |
|  | A | I, II, e III |
|  | B | Todas  Você assinalou essa alternativa (B) |
|  | C | I, III e IV |
|  | D | I, IV e V |
|  | E | Todas, exceto III |

Questão 8/10 - Big Data

GraphX é o componente do Spark para computação de grafos em sistemas distribuídos de larga escala. Ele foi desenvolvido através de um projeto de pesquisa como forma de unificar o processamento de grafos e de sistemas paralelos até se tornar uma parte integral do projeto Spark.   
  
Avalie as seguintes afirmações a respeito do GraphX:  
  
I – A API GraphX implementa abstrações de arestas e vértices baseados em extensões de RDDs

II – A computação de grafos é importante em casos onde as relações entre os dados são mais importantes que os dados em si.

III – Grafos são uma forma muito eficiente de representar dados de tabelas esparsas.

IV – O algoritmo *Connected Components* é um tipo de algoritmo paralelo de grafos que é capaz de medir a coesão entre os seus dados (componentes).

V – O algoritmo *PageRank* desenvolvido pelo Google é capaz de classificar a importância de sites através da quantidade de referências apontadas para um site.

Dentre essas afirmações estão CORRETAS:

Você não pontuou essa questão

|  |  |  |
| --- | --- | --- |
|  | A | Todas, exceto IV |
|  | B | I, II e III |
|  | C | II, III e IV |
|  | D | IV e V  Você assinalou essa alternativa (D) |
|  | E | Todas |

Questão 9/10 - Big Data

A API GraphX oferece uma forma de armazenar grafos na forma de tabelas e utilizar operações de tabelas para expressar operações de grafos.  
  
A respeito da API do GraphX avalie as seguintes afirmações:  
  
I – Um grafo de propriedades em GraphX é uma abstração capaz de definir um multigrafo dirigido com objetos definidos para cada vértice e aresta.

II – Um grafo de propriedade pode ser definido como um par de conjuntos de RDDs que codificam as características de vértices e arestas.

III – Grafos de propriedades são distribuídos, tolerantes a falhas e seus valores e estruturas podem ser modificados facilmente.

IV – O construtor da classe Edge recebe os parâmetros que correspondem aos identificadores dos vértices de origem e destino.

V – O membro *vertices* da classe Graph retorna um *VertexRDD[(String,String)]* que é uma extensão do objeto *RDD[(VertexId,(String,String)]*. Por outro lado, o membro *edge* retorna um *EdgeRDD* que contém um objeto *Edge[String]*.

Dentre essas afirmações estão INCORRETAS:

Você não pontuou essa questão

|  |  |  |
| --- | --- | --- |
|  | A | I, II e III |
|  | B | I, III e IV  Você assinalou essa alternativa (B) |
|  | C | Apenas III |
|  | D | Todas |
|  | E | II, IV e V |

Questão 10/10 - Big Data

O funcionamento estrutural de um fluxo de aprendizado de máquina pode ser descrito através das seguintes etapas: Utilizamos um conjunto de dados de teste e de treino para escolher os modelos de aprendizado de máquina mais adequados, treiná-los, e avaliá-los com o objetivo de alcançar um modelo que seja capaz de realizar predições úteis a partir de dados do mundo real.

Avalie as seguintes afirmações a respeito do fluxo de aprendizado de máquina:

I – Na etapa de obtenção de dados são adquiridos os dados que serão utilizados tanto para treinamento dos modelos quanto para avaliar a eficiência dos modelos.

II – Os problemas de dados faltantes, dados com ruído ou dados inconsistentes devem ser resolvidos na etapa de obtenção dos dados.

III – Os dados devem ser formatados da maneira mais adequada durante a etapa de preparação dos dados.

IV – A etapa de treinamento é quando utilizamos os dados de treino na expectativa de que o modelo aprenda a processar os dados da forma esperada.

V – Durante a etapa de avaliação o desempenho do modelo é avaliado e caso não seja considerado suficiente, o modelo passa por um novo ciclo de ajuste de parâmetros, treinamento e avaliação.

Dentre essas afirmações estão INCORRETAS:

Você não pontuou essa questão

|  |  |  |
| --- | --- | --- |
|  | A | I, IV e V |
|  | B | Apenas II |
|  | C | II e III |
|  | D | Todas, exceto IV |
|  | E | Nenhuma  Você assinalou essa alternativa (E) |