

## QUESTÕES COMENTADAS

Chegamos a nossa tradicional lista de questões comentadas.



### 7. Ano: 2018 Banca: CESPE Órgão: TCM-BA Cargo: Auditor de Contas Questão: 14

Acerca de big data, assinale a opção correta.

A A utilização de big data nas organizações não é capaz de transformar os seus processos de gestão e cultura.

B Sistemas de recomendação são métodos baseados em computação distribuída, que proveem uma interface para programação de clusters, a fim de recomendar os tipos certos de dados e processar grandes volumes de dados.

C Pode-se recorrer a software conhecidos como scrapers para coletar automaticamente e visualizar dados que se encontram disponíveis em sítios de navegabilidade ruim ou em bancos de dados difíceis de manipular.

D As ações inerentes ao processo de preparação de dados incluem detecção de anomalias, deduplicação, desambiguação de entradas e mineração de dados.

E O termo big data se baseia em cinco Vs: velocidade, virtuosidade, volume, vantagem e valor.

**Comentário:** Vamos comentar cada uma das alternativas acima:

A) Esse conceito de transformar processos de gestão e cultura está associado a outro conceito, o de gestão de processos de negócio, conhecido como BPM. Logo, temos uma alternativa **incorreta**.

B) Um **Sistema de Recomendação** combina várias técnicas computacionais para selecionar itens personalizados com base nos interesses dos usuários e conforme o contexto no qual estão inseridos. Tais itens podem assumir formas bem variadas como, por exemplo, livros, filmes, notícias, música, vídeos, anúncios, links patrocinados, páginas de internet, produtos de uma loja virtual, etc. Empresas como Amazon, Netflix e Google são reconhecidas pelo uso intensivo de sistemas de recomendação com os quais obtém grande vantagem competitiva. Veja que não existe nenhuma obrigação dos sistemas serem estruturados em um cluster e usarem computação distribuída.

C) Essa é a nossa resposta, e confesso desde já que não apresentamos esse conceito em nosso material. **Data scraping** (do inglês, raspagem de dados) é uma técnica computacional na qual um programa extrai dados de saída legível somente para humanos, proveniente de um serviço ou aplicativo. Os dados extraídos geralmente são minerados e estruturados em um formato padrão como CSV, XML ou JSON. Vejam que o objetivo é transformar os dados não estruturados em modelos mais fáceis de manipular.



D) **Data preparation** é o processo de **coletar, limpar, normalizar, combinar, estruturar e organizar dados para análise**. Ele é o passo inicial (e fundamental) para que o trabalho com Big Data seja bem-sucedido, uma vez que aumenta a qualidade dos dados – e, consequentemente, dos resultados com **data mining**. Dados “pobres”, de qualidade ruim, geram resultados incorretos e não-confiáveis ao fim do processo de uso das tecnologias de Data Science. Vejam que a mineração de dados é uma ação feita em uma etapa posterior a preparação dos dados. Logo, temos uma alternativa **incorreta**.

E) Observem que os v’s listados na alternativa não são os que está descrito na definição de big data, os conceitos corretos são: Volume, Velocidade, Variedade, Veracidade e Valor.

**Gabarito: C.**



**8. Ano: 2018 Banca: Cesgranrio Órgão: Petrobras Cargo: Analista de Processo de Negócio**  
**Questão: 42**

A principal definição de Big Data parte de três características, conhecidas como 3 V do Big Data, a saber: velocidade, variedade e volume.

O termo velocidade refere-se, principalmente, à

- (A) necessidade das aplicações de gerar respostas rapidamente, a partir de grandes massas de dados.
- (B) existência de um alto fluxo de dados na entrada.
- (C) necessidade de gerar aplicações rapidamente, em função da demanda do negócio.
- (D) importância da facilidade de manipular cubos de visualização de dados, rapidamente.
- (E) rapidez com que os dados se tornam inválidos com o tempo.

**Comentário:** Essa é a primeira questão de Big Data elaborada pela Cesgranrio. Acho que a falta de experiência nesse assunto fez com que as alternativas A e B pudessem ser consideradas a resposta correta. Vamos construir os conceitos para ver se chegamos a alguma conclusão sobre o pensamento do examinador ao elaborar essa questão. Começamos pela definição de big data.

**Big data** é um termo que descreve **grandes volumes de alta velocidade**, dados complexos e variáveis que exigem técnicas avançadas tecnologias para permitir a captura, armazenamento, distribuição, gestão e análise da informação.

Agora vamos voltar para os 3 V’s da definição original. **Volume** refere-se à magnitude dos dados. **Variedade** refere-se à heterogeneidade estrutural em um conjunto de dados. E o volume?

Velocidade refere-se à **taxa na qual os dados são gerados** e a **velocidade com que devem ser analisados e utilizados**. A proliferação de dispositivos digitais, como smartphones e sensores,



levou a uma taxa sem precedentes de criação de dados e está impulsionando a necessidade crescente de análise em tempo real e planejamento baseado em evidências.

Mesmo os varejistas convencionais estão gerando dados numa frequência extremamente alta. O Wal-Mart, por exemplo, processa mais de um milhão de transações por hora. Os dados provenientes de dispositivos móveis e fluindo por meio de aplicativos produzem um universo de informações que podem ser usadas para gerar ofertas personalizadas em tempo real para clientes comuns.

Esses dados fornecem informações sólidas sobre os clientes, como localização geoespacial, dados demográficos e padrões de compra anteriores, que podem ser analisados em tempo real para criar um valor real para o cliente.

Vejam que o fluxo de dados na entrada ou a taxa de geração de dados é uma medida relevante para a velocidade. Agora, se formos analisar a alternativa A tem dois pontos que me deixaram com uma pulga atrás da orelha.

O primeiro deles se refere ao fato dele afirmar que existe uma **necessidade** das aplicações (vamos supor que sejam as aplicações de Big Data) gerarem resposta rapidamente. Isso de fato tem relação com velocidade, mas não é uma obrigatoriedade. Já o segundo ponto refere-se a **grandes massas de dados** veja que esse termo está mais associado ao volume do que à velocidade. Assim, acabei me convencendo que a alternativa B era de fato a resposta para a questão.

Contudo, com uma boa referência, é possível refutar a banca mostrando a presença de duas afirmações corretas entre as alternativas.

**Gabarito: B**



**9. Ano: 2018 Banca: CESPE Órgão: TCE-PB Prova: Auditor de Contas Públicas - Demais Áreas**

Com referência a big data, assinale a opção correta.

- a) A definição mais ampla de big data restringe o termo a duas partes — o volume absoluto e a velocidade —, o que facilita a extração das informações e dos insights de negócios.
- b) O sistema de arquivos distribuído Hadoop implementa o algoritmo Dijkstra modificado para busca irrestrita de dados em árvores aglomeradas em clusters com criptografia.
- c) Em big data, o sistema de arquivos HDFS é usado para armazenar arquivos muito grandes de forma distribuída, tendo como princípio o write-many, read-once.
- d) Para armazenar e recuperar grande volume de dados, o big data utiliza bancos SQL nativos, que são bancos de dados que podem estar configurados em quatro tipos diferentes de armazenamentos: valor chave, colunar, gráfico ou documento.
- e) O MapReduce é considerado um modelo de programação que permite o processamento de dados massivos em um algoritmo paralelo e distribuído.



**Comentário:** Vejamos cada uma das afirmações acima.

a) a definição mais simples de big data já estrutura o termo em 3 conceitos: volume, variedade e velocidade. Logo, temos uma alternativa incorreta.

b) Essa questão tem diversos conceitos dispersos que não fazem sentido da forma como foi apresentada no texto. Veja que o Hadoop possui de fato um sistema de armazenamento distribuído: o HDFS. O HDFS possui uma arquitetura mestre/escravo. Um cluster HDFS consiste em um único **NameNode**, considerado o servidor mestre que gerencia o *namespace* do sistema de arquivos e regula o acesso a arquivos por clientes. Além disso, há vários **DataNodes**, geralmente um por nó do cluster, que gerenciam o armazenamento associado aos nós em que são executados. Não existe o uso do algoritmo de Dijkstra<sup>2</sup> para busca em registros, ele está preocupado em achar o menor caminho e um grafo.

Sobre criptografia, o HDFS implementa criptografia transparente de ponta a ponta. Depois de configurados, os dados lidos e gravados em diretórios HDFS especiais são criptografados e descriptografados de forma transparente, sem exigir alterações no código do aplicativo do usuário.

c) Em big data temos um princípio semelhante ao do data warehousing, os dados são escritos apenas uma vez, write-once e lido por diversas pessoas, read-many. Assim, a alternativa c também está errada.

d) Sabemos que os bancos de dados NoSQL não são SQL nativos, logo, não podemos marcar essa alternativa como resposta.

e) Sobrou apenas essa alternativa que apresentou uma definição consistente de MapReduce. Logo, marcamos nossa resposta nesta letra E.

**Gabarito: E**



**10. Ano: 2017 Banca: FCC Órgão: DPE-RS Prova: Analista - Banco de Dados**

Os sistemas de Big Data costumam ser caracterizados pelos chamados 3 Vs, sendo que o V de

a) Veracidade corresponde à rapidez na geração e obtenção de dados.

b) Valor corresponde à grande quantidade de dados acumulada.

c) Volume corresponde à rapidez na geração e obtenção de dados.

d) Velocidade corresponde à confiança na geração e obtenção dos dados.

e) Variedade corresponde ao grande número de tipos ou formas de dados.

<sup>2</sup> Uma aplicação de busca em um grafo comum é encontrar o caminho mais curto obtido entre um nó inicial e um ou mais nós de destino. Comumente feito em uma única máquina com o **Algoritmo de Dijkstra**.



**Comentário:** Analisando as alternativas acima, podemos aproveitar para relembrar os conceitos corretos associados a cada um dos termos. Os primeiros esforços para definir o significado do Big Data caracterizaram-no em termos dos Três V's: **Volume**, **Velocidade**, **Variedade**. À medida que mais organizações começaram a aproveitar o potencial do Big Data, a lista de V se expandiu, vejamos cada um deles:

**Volume:** refere-se à quantidade de dados. O Big Data geralmente tem milhares de entidades ou elementos em bilhões de registros.

**Velocidade:** refere-se à velocidade na qual os dados são **capturados**, **gerados** ou **compartilhados**. Big Data é **frequentemente gerado** e pode ser **distribuído** e até **analisado em tempo real**.

**Variedade/Variabilidade:** Refere-se às formas em que os dados são **capturados** ou **entregues**. Big Data requer armazenamento de vários formatos de arquivos e dados (estruturados e não estruturados).

**Volatilidade:** refere-se à frequência com que as alterações de dados ocorrem e, portanto, quanto tempo os dados são úteis.

**Veracidade:** refere-se ao grau de confiabilidade dos dados. Assim, podemos marcar nossa resposta na alternativa E.

**Gabarito: E**



**11. Ano: 2018 Banca: CESPE Órgão: CGM de João Pessoa – PB Prova: Auditor Municipal de Controle Interno - Desenvolvimento de Sistemas**

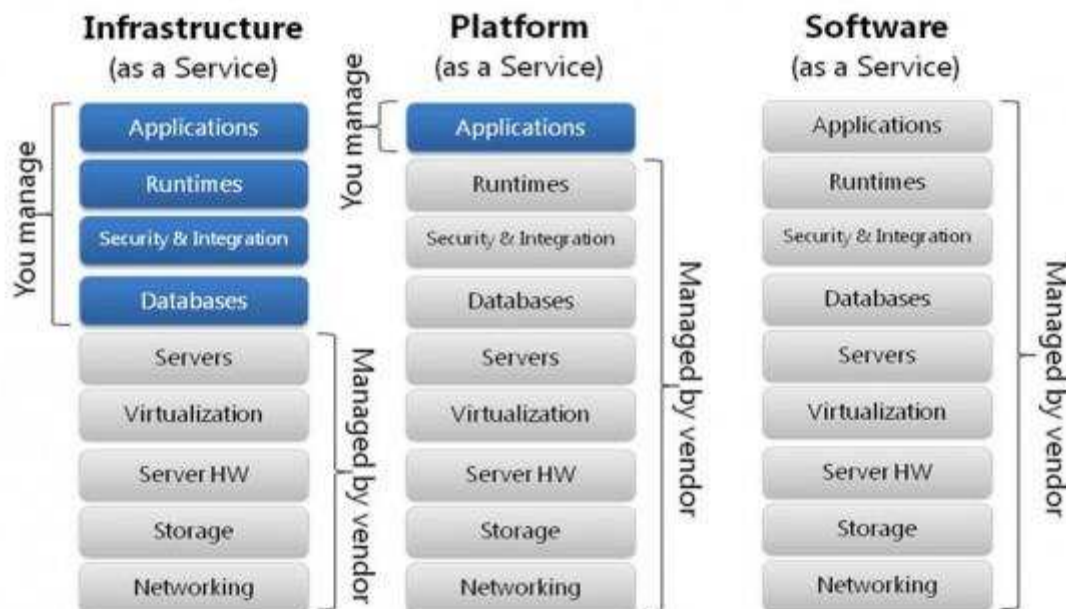
A respeito de business intelligence, julgue o próximo item.

**Situação hipotética:** Um órgão público pretende fazer uma análise de big data com o objetivo de realizar mineração em grandes conjuntos de dados corporativos para localizar ou encontrar padrões ocultos, o que exigirá uma grande quantidade de processamento.

**Assertiva:** Nesse caso, há necessidade de contratação de um SaaS (software como serviço), que acrescentará uma camada adicional de integração com frameworks de desenvolvimento de aplicativos, recursos de middleware e funções como bancos de dados.

**Comentário:** Para respondermos a essa questão precisamos olhar para os conceitos de Infraestrutura, Plataforma e Software como serviço, vamos, primeiramente olhar para a figura a seguir:





Gabarito: E



**12. Ano: 2017 Banca: FGV Órgão: IBGE Prova: Analista Censitário - Análise de Sistemas - Desenvolvimento de Aplicações - Web Mobile**

Bancos de Dados NoSQL podem armazenar dados em diversos formatos não relacionais, como documentos compostos por pares de campo-e-valor (field-and-value), conforme a estrutura exemplificada a seguir.

O Banco de Dados NoSQL utilizado para armazenar documentos compostos por pares campo-e-valor, no formato BSON (JSON-like), é o:

- a) OpenLink Virtuoso;
- b) Neo4j;
- c) Apache HBase;
- d) MongoDB;
- e) Titan.

**Comentário:** Veja que a questão exige apenas que você saiba qual desses nomes acima é um banco de dados NoSQL do tipo chave-valor. Vejam a figura abaixo para saber onde cada alternativa se enquadra e quais os outros exemplos de cada tipo de base de dados NoSQL.





### Chave-valor

Amazon  
DynamoDB (Beta)

ORACLE  
BERKELEY DB 11g

redis

### Grafo

Neo4j  
the graph database

InfiniteGraph

sones

### Tabular

HBASE

riak

Cassandra

### Documento

CouchDB  
relax

mongoDB

terastore

Desta forma, podemos marcar nossa resposta na alternativa D

**Gabarito: D**



### 13. Ano: 2015 Banca: CESPE Órgão: TJDF Cargo: Suporte em Tecnologia da Informação – Questões 92 e 93

A respeito de tipos de bancos de dados, julgue os itens que se seguem.

92 A capacidade de estender tipos de dados básicos é uma das características dos bancos de dados objeto relacional.

93 Sistemas de bancos de dados classificados como NoSQL permitem a inserção de dados sem que haja um esquema predefinido.

**Comentários:** Essa questão trata na alternativa 92 de banco de dados relacionais. A ideia descrita na alternativa trata dos itens conhecidos como *user defined type* – UDT. Eles aparecem dentro do padrão SQL e permitem que novos tipos possam ser estendidos dos tipos básicos.

A alternativa 93, por sua vez, refere-se ao assunto presente nesta aula. Vejam que a existência de esquema bem definido não é uma obrigatoriedade dentro do contexto de bancos de dados NoSQL. Essa propriedade é conhecida como *schema free* ou livre de esquema.



Pelo exposto acima, as duas alternativas estão corretas.

**Gabarito: C C**



**14. Ano: 2016 Banca: FGV Órgão: IBGE Cargo: WEB MOBILE Questão: 41**

Considere as seguintes características de um projeto de banco de dados.

- I. O modelo de dados é conhecido a priori e é estável;
- II. A integridade dos dados deve ser rigorosamente mantida;
- III. Velocidade e escalabilidade são preponderantes.

Dessas características, o emprego de bancos de dados NoSQL é favorecido somente por:

- (A) I;
- (B) I e II;
- (C) II;
- (D) II e III;
- (E) III.

**Comentário:** Veja que a questão trata das propriedades ou características de um projeto de banco de dados NoSQL. Sabemos que bancos de dados deste tipo substituem a sigla ACID (atomicidade, consistência, isolamento e durabilidade) que está relacionado com transações pela sigla BASE que está relacionada com os conceitos de BA – (Basically Available) disponibilidade é prioridade, S – (Soft-State) – Não precisa ser consistente o tempo todo e E – (Eventually Consistent) – Consistente em momento indeterminado. Outro ponto importante é a relação de NoSQL com o modelo de dados, geralmente o banco é considerado schema free, ou seja, livre de esquema ou de um modelo pré-determinado a priori.

Partindo desta exposição básica sobre NoSQL podemos analisar as alternativas I, II e III. Vejam que as duas primeiras são definições associadas a banco de dados relacionais ou transacionais. A alternativa III trata de outro aspecto que não vimos ainda, mas que também está relacionada ao conceito de NoSQL e Big Data. NoSQL é uma tecnologia que trouxe o foco de banco de dados para performance e escalabilidade. A pergunta é como analisar uma grande quantidade de dados.

**Gabarito: E**



**15. Ano: 2016 Banca: ESAF Órgão: ANAC Cargo: Analista de TI - QUESTÃO 62**

Big Data é:





- a) volume + variedade + agilidade + efetividade, tudo agregando + valor + atualidade.
- b) volume + oportunidade + segurança + veracidade, tudo agregando + valor.
- c) dimensão + variedade + otimização + veracidade, tudo agregando + agilidade.
- d) volume + variedade + velocidade + veracidade, tudo agregando + valor.
- e) volume + disponibilidade + velocidade + portabilidade, tudo requerendo - valor.

**Comentário:** Embora o termo "big data" seja relativamente novo, o ato de recolher e armazenar grandes quantidades de informações para eventual análise de dados é bem antigo. O conceito ganhou força no início dos anos 2000, quando um analista famoso deste setor, Doug Laney, articulou a definição de big data como os três Vs:

**Volume.** Organizações coletam dados de uma grande variedade de fontes, incluindo transações comerciais, redes sociais e informações de sensores ou dados transmitidos de máquina a máquina. No passado, armazenar tamanha quantidade de informações teria sido um problema – mas novas tecnologias (como o Hadoop) têm aliviado a carga.

**Velocidade.** Os dados fluem em uma velocidade sem precedentes e devem ser tratados em tempo hábil. Tags de RFID, sensores, celulares e contadores inteligentes estão impulsionando a necessidade de lidar com imensas quantidades de dados em tempo real, ou quase real.

**Variedade.** Os dados são gerados em todos os tipos de formatos - de dados estruturados, dados numéricos em bancos de dados tradicionais, até documentos de texto não estruturados, e-mail, vídeo, áudio, dados de cotações da bolsa e transações financeiras.

Outros Vs foram adicionados ao longo dos anos:

**Veracidade.** O quarto V do conceito está ligado à veracidade do conteúdo. Nem todos os dados gerados possuem consistência dentro do contexto. É preciso destacar o que é rico e correto em conteúdo no meio de tanta informação. Ao garantir essa separação, que é possível fazer a partir do Big Data, o que sobra são conhecimentos importantes para compreender melhor a entidade analisada ou o negócio em questão.

**Valor.** Para garantir que o trabalho dos outros Vs traga retorno, é preciso gerar valor para os resultados que retornam do Big Data. Assim, aparece o último V do conceito.

Juntando todos os conceitos acima, chegamos a nossa resposta na alternativa D.

**Gabarito: D**



**16. Ano: 2016 Banca: ESAF Órgão: ANAC Cargo: Analista de TI - QUESTÃO 70**

Para o processamento de grandes massas de dados, no contexto de Big Data, é muito utilizada uma plataforma de software em Java, de computação distribuída, voltada para clusters, inspirada no MapReduce e no GoogleFS. Esta plataforma é o(a)

- a) Yam Common.



- b) GoogleCrush.
- c) EMRx.
- d) Hadoop.
- e) MapFix

**Comentário:** Questão inspirada na WIKIPEDIA:

Hadoop é uma plataforma de software em Java de computação distribuída voltada para clusters e processamento de grandes massas de dados. Foi inspirada **no MapReduce e no GoogleFS (GFS)**. Trata-se de um projeto da Apache de alto nível, que vem sendo construído por uma comunidade de colaboradores utilizando em sua maior parte a linguagem de programação Java, com algum código nativo em C e alguns utilitários de linha de comando escrito utilizando scripts shell. Assim podemos marcar nossa resposta na alternativa D.

**Gabarito: D**



**17. ANO: 2015 BANCA: FGV ÓRGÃO: TJ-SC PROVA: ANALISTA JUDICIÁRIO - ANALISTA DE SISTEMAS**

Os termos Business Intelligence (BI) e Big Data confundem-se em certos aspectos. Uma conhecida abordagem para identificação dos pontos críticos de cada paradigma é conhecida como 3V, e destaca:

- A variedade, visualização, volume;
- B velocidade, virtualização, volume;
- C variedade, velocidade, volume;
- D virtualização, visualização, volume;
- E variedade, visualização, virtualização.

**Comentário:** Questão interessante! Trata dos 3 Vs que definem o primeiro conceito de Big Data. Sabemos que se trata da Variedade, Velocidade e Volume. O primeiro faz referência aos diferentes tipos de dados ou informações que devem ser processadas. O segundo se preocupa com o tempo para o processamento e armazenamento de forma consistente e acessível das informações geradas. Por fim, o volume consiste na percepção de que a quantidade dos dados aumenta a cada dia em grandes proporções.

**Gabarito: C**



**18. ANO: 2014 BANCA: CESPE ÓRGÃO: TJ-SE PROVA: ANALISTA JUDICIÁRIO - BANCO DE DADOS**

Julgue os itens que se seguem, no que se refere a Big Data.



[1] O processamento de consultas *ad hoc* em Big Data, devido às características de armazenamento dos dados, utiliza técnicas semelhantes às empregadas em consultas do mesmo tipo em bancos de dados tradicionais.

**Comentário:** O processamento de consultas *ad hoc* no contexto de *big data* traz desafios diferentes daqueles incorridos ao realizar consultas *ad hoc* em dados estruturados pelo fato de as fontes e formatos dos dados não serem fixos e exigirem mecanismos diferentes para recuperá-los e processá-los.

Embora as consultas *ad hoc* simples possam ser resolvidas pelos provedores de big data, na maioria dos casos, elas são complexas porque os dados, algoritmos, formatos e resoluções da entidade devem ser descobertos dinamicamente. O conhecimento dos cientistas de dados e dos usuários corporativos é necessário para definir a análise exigida para as seguintes tarefas:

- Identificar e descobrir os cálculos e algoritmos
- Identificar e descobrir as fontes de dados
- Definir os formatos necessários que podem ser consumidos pelos cálculos
- Executar os cálculos nos dados paralelamente

**Gabarito: E**



## 19. ANO: 2014 BANCA: CESPE ÓRGÃO: TJ-SE PROVA: ANALISTA JUDICIÁRIO - BANCO DE DADOS

Julgue os itens que se seguem, no que se refere a Big Data.

Em soluções Big Data, a análise dos dados comumente precisa ser precedida de uma transformação de dados não estruturados em dados estruturados.

**Comentário:** As soluções de big data são, em sua maioria, dominadas por sistemas Hadoop e tecnologias baseadas em MapReduce, que são soluções simples de instalar para processamento e armazenamento distribuídos. No entanto, a extração de dados a partir de dados não estruturados, como imagens, áudio, vídeo, feeds binários ou até mesmo texto, é uma tarefa complexa e precisa de técnicas como aprendizado de máquina e processamento de idioma natural, etc. O outro grande desafio é como verificar a precisão e a exatidão do resultado de tais técnicas e algoritmos.

Para executar a análise em quaisquer dados, eles devem estar em algum tipo de formato estruturado. Os dados não estruturados acessados de várias fontes podem ser armazenados como estão e, em seguida, transformados em dados estruturados (por exemplo, JSON) e novamente armazenados nos sistemas de armazenamento de big data. O texto não estruturado pode ser convertido em dados estruturados ou semiestruturados. Da mesma forma, os dados de imagem, áudio e vídeo precisam ser convertidos nos formatos que podem ser usados para análise. Além disso, a precisão e exatidão da analítica avançada que usa algoritmos preditivos e estatísticos dependem da quantidade de dados e algoritmos usados para treinar os modelos.



**Gabarito: C**



**20. ANO: 2014 BANCA: CESPE ÓRGÃO: TJ-SE PROVA: ANALISTA JUDICIÁRIO - BANCO DE DADOS**

Julgue os itens que se seguem, no que se refere a Big Data.

[1] Ao utilizar armazenamento dos dados em nuvem, a localização do processamento de aplicações Big Data não influenciará os custos e o tempo de resposta, uma vez que os dados são acessíveis a partir de qualquer lugar.

**Comentário:** Muito embora haja uma abstração quanto ao local do processamento, afinal, se contrata um serviço e pouco importa como e onde ele será prestado, essa escolha tem consequências.

Muitos provedores de serviço na nuvem, como a Amazon, possuem tarifas diferenciadas com relação à localização de cada servidor. Cada país tem seus custos e suas especificidades.

Além disso, quanto mais longe o servidor estiver do interessado, maior será o tempo de espera entre a requisição e a resposta.

**Gabarito: E**



**21. ANO: 2015 BANCA: FGV ÓRGÃO: TJ-BA PROVA: TÉCNICO DO JUDICIÁRIO - TECNOLOGIA DA INFORMAÇÃO**

Analise as afirmativas a respeito da classe de gerenciadores de bancos de dados, surgida em anos recentes, conhecida como NoSQL.

- I. Mesmo sem suportar tabelas relacionais, baseiam-se em esquemas de dados previamente definidos;
- II. Suas estruturas não permitem o uso de linguagens do tipo do SQL para recuperação de dados;
- III. Garantem operações com as propriedades conhecidas pela sigla ACID;
- IV. Privilegiam a rapidez de acesso e a disponibilidade dos dados em detrimento das regras de consistência das transações.

O número de afirmativas corretas é:

- A uma;
- B duas;
- C três;
- D quatro;



E cinco.

**Comentário:** Se voltarmos a parte da nossa aula teórica que trata de NoSQL veremos que a única alternativa correta é a presente na assertiva IV.

**Gabarito: A**



## 22. ANO: 2015 BANCA: CESPE ÓRGÃO: TCU PROVA: AUDITOR FEDERAL DE CONTROLE EXTERNO - TECNOLOGIA DA INFORMAÇÃO

Julgue os itens subsecutivos, a respeito de sistemas de bancos de dados.

Como forma de permitir as buscas em documentos semiestruturados, um banco de dados NoSQL do tipo orientado a documentos armazena objetos indexados por chaves utilizando tabelas de hash distribuída.

**Comentário:** Quando tratamos de bases de dados NoSQL podemos classifica-las em quatro diferentes tipos, são eles:

**Chave/valor (Key/Value):** conhecidos como tabelas de hash distribuídas. Armazenam objetos indexados por chaves, e facilita a busca por esses objetos a partir de suas chaves.

**Orientados a Documentos:** os documentos dos bancos são coleções de atributos e valores onde um atributo pode ser multivalorado. Em geral, os bancos de dados orientados a documento não possuem esquema, ou seja, os documentos armazenados não precisam possuir estrutura em comum. Essa característica faz deles boas opções para o armazenamento de dados semiestruturados.

**Colunar:** Bancos relacionais normalmente guardam os registros das tabelas contiguamente no disco. Por exemplo, caso se queira guardar id, nome e endereço de usuários em um banco de dados relacional, os registros seriam:

Id1, Nome1, Endereço1;

Id2, Nome2, Endereço2.

Essa estrutura torna a escrita muito rápida, pois todos os dados de um registro são colocados no disco com uma única escrita no banco. Também é eficiente caso se queira ler registros inteiros. Mas para situações onde se quer ler algumas poucas colunas de muitos registros, essa estrutura é pouco eficiente, pois muitos blocos do disco terão de ser lidos.

Para esses casos onde se quer otimizar a leitura de dados estruturados, bancos de dados de famílias de colunas são mais interessantes, pois eles guardam os dados contiguamente por coluna.

O exemplo anterior em um banco de dados dessa categoria ficaria:

Id1, Id2; Nome1, Nome2; Endereço1, Endereço2.



Os bancos de dados de famílias de colunas são mais interessantes para processamento analítico online (OLAP). Bigtable é uma implementação da Google dessa categoria de bancos de dados.

**Orientado a Grafos:** diferente de outros bancos de dados NoSQL, esse está diretamente relacionado a um modelo de dados estabelecido, o modelo de grafos. A ideia desse modelo é representar os dados e / ou o esquema dos dados como grafos dirigidos, ou como estruturas que generalizem a noção de grafos. O modelo de grafos é aplicável quando “informações sobre a interconectividade ou a topologia dos dados são mais importantes, ou tão importante quanto os dados propriamente ditos”. Possui três componentes básicos: os nós (são os vértices do grafo), os relacionamentos (são as arestas) e as propriedades (ou atributos) dos nós e relacionamentos.

Vejam que a afirmativa do CESPE confunde os conceitos de orientado a documentos com o conceito de chave valor. Quanto nós tratamos de tabelas hash distribuídas está associada ao tipo chave-valor e não ao conceito de orientado a documentos como está dito na questão. Sendo, portanto, alternativa **errada**.

**Gabarito: E**



**23. ANO: 2014 BANCA: CESPE ÓRGÃO: TJ-SE PROVA: ANALISTA JUDICIÁRIO - BANCO DE DADOS**

Acerca de bancos de dados semiestruturados e bancos de dados NOSQL, julgue os itens subsecutivos.

Bancos de dados NOSQL orientados a documentos são apropriados para o armazenamento de dados semiestruturados.

**Comentário:** Exatamente, neles é possível armazenamento de documentos que embora tenham uma estrutura básica, eles, geralmente não são definidos por nenhum modelo ou schema. Podemos, então, encontrar propriedades dentro dos documentos, mas não existe uma obrigatoriedade sobre a existência delas.

**Gabarito: C**



**24. ANO: 2014 BANCA: CESPE ÓRGÃO: TJ-SE PROVA: ANALISTA JUDICIÁRIO - BANCO DE DADOS**

Acerca de bancos de dados semiestruturados e bancos de dados NOSQL, julgue os itens subsecutivos.

Para garantir a eficiência das consultas a bancos de dados semiestruturados, é fundamental a adoção de técnica de indexação que leve em consideração, além das informações, as propriedades estruturais dos dados.





**Comentário:** A questão 87 envolve alguns conceitos interessantes. Começa falando sobre dados semiestruturados, por exemplo, XML ou JSON. Para viabilizar a indexação de dados semiestruturados é preciso conhecimento das propriedades dos objetos de dados. Consultas em bancos de dados semiestruturados consideram tanto a estrutura quanto os valores. Outra questão é a criação de índice sobre um conjunto de dados semiestruturados. Para avaliar se um índice deve ou não ser criado é importante usar as informações sobre a estrutura dos dados e os valores armazenados. Neste caso, considerando a necessidade de um espaço maior para armazenamento e do custo de manutenção, a criação do índice deve melhorar a performance para ser de fato implementado. Logo, a assertiva está **correta**.

**Gabarito: C**



**25. ANO: 2014 BANCA: CESPE ÓRGÃO: TJ-SE PROVA: ANALISTA JUDICIÁRIO - BANCO DE DADOS**

Acerca de bancos de dados semiestruturados e bancos de dados NOSQL, julgue os itens subsecutivos.

Devido à escalabilidade esperada para os bancos de dados NOSQL, a implementação desses bancos utiliza modelos de armazenamento de dados totalmente distintos dos utilizados em sistemas relacionais.

**Comentário:** Nem todos os modelos de armazenamentos são totalmente distintos dos sistemas relacionais. Se pensarmos no modelo de armazenamento colunar ele é semelhante a uma tabela, mas guardam os dados das colunas contiguamente no disco.

NoSQL são diferentes sistemas de armazenamento que vieram para suprir necessidades em demandas onde os bancos de dados tradicionais (relacionais) são ineficazes. Muitas dessas bases apresentam características muito interessantes, como alta performance, escalabilidade, replicação, suporte à dados estruturados e sub colunas.

O NoSQL surgiu da necessidade de uma performance superior e de uma alta escalabilidade. Os atuais bancos de dados relacionais são muito restritos a isso, sendo necessária a distribuição vertical de servidores, ou seja, quanto mais dados, mais memória e mais disco um servidor precisa. O NoSQL tem uma grande facilidade na distribuição horizontal, ou seja, mais dados, mais servidores, não necessariamente de alta performance. Um grande utilizador desse conceito é o Google, que usa computadores de pequeno e médio porte para a distribuição dos dados; essa forma de utilização é muito mais eficiente e econômica. Além disso, os bancos de dados NoSQL são muito tolerantes a erros.

No caso dos bancos NoSQL, toda a informação necessária estará agrupada no mesmo registro, ou seja, em vez de você ter o relacionamento entre várias tabelas para formar uma informação, ela estará em sua totalidade no mesmo registro.

**Gabarito: E**





**26. ANO: 2013 BANCA: CESPE ÓRGÃO: CNJ PROVA: ANALISTA JUDICIÁRIO - ANÁLISE DE SISTEMAS**

No que se refere ao desenvolvimento web de alto desempenho, julgue os itens subsequentes.

Uma característica de bancos de dados NoSQL é o suporte à replicação de dados. Entre as abordagens utilizadas para replicação, inclui-se a mestre-escravo.

**Comentário:** São dois tipos de replicação: mestre-escravo e ponto a ponto. A replicação mestre-escravo torna um nodo a cópia oficial, a qual lida com gravações, enquanto os escravos sincronizam-se com o mestre e podem lidar com as leituras. A replicação ponto a ponto permite gravações em qualquer nodo; os nodos são coordenados para sincronizar suas cópias de dados. A replicação mestre-escravo reduz a chance de conflitos de atualização, mas a ponto a ponto evita carregar todas as gravações em um único ponto de falha.

**Gabarito: C**



**27. ANO: 2013 BANCA: CESPE ÓRGÃO: CNJ PROVA: ANALISTA JUDICIÁRIO - ANÁLISE DE SISTEMAS**

No que se refere ao desenvolvimento web de alto desempenho, julgue os itens subsequentes.

A escalabilidade dos bancos de dados NoSQL é garantida pela ausência de um esquema (scheme free).

**Comentário:** A escalabilidade é garantida através da característica nativa de clusterização do banco de dados.

**Gabarito: E**



**28. ANO: 2013 BANCA: CESPE ÓRGÃO: CNJ PROVA: ANALISTA JUDICIÁRIO - ANÁLISE DE SISTEMAS**

No que se refere ao desenvolvimento web de alto desempenho, julgue os itens subsequentes.

[1] Apesar de implementarem tecnologias distintas, todos os bancos de dados NoSQL apresentam em comum a implementação da tecnologia chave-valor.

**Comentário:** Sabemos que são cada solução NoSQL possui um modelo de armazenamento diferente, os quais dividimos em quatro categorias amplamente utilizadas no ecossistema NoSQL: **chave-valor, documento, família de colunas ou colunar, e grafos.**

**Gabarito: E**





**29. ANO: 2016 BANCA: CESPE CONCURSO: FUNPRES CARGO: ESPECIALISTA ÁREA: TECNOLOGIA DA INFORMAÇÃO (TI)**

Com relação à forma como os dados são armazenados e manipulados no desenvolvimento de aplicações, julgue os itens a seguir.

[69] Em um banco de dados NoSQL do tipo grafo, cada arco é definido por um identificador único e expresso como um par chave/valor.

**Comentários:** Abaixo analisaremos cada uma das alternativas.

69. A alternativa vai exigir conhecimento sobre os modelos de armazenamento utilizados por bancos de dados NoSQL. Quando tratamos de bases de dados NoSQL podemos classifica-las em quatro diferentes tipos, dois deles são citados na questão:

**Chave/valor (Key/Value):** conhecidos como tabelas de hash distribuídas. Armazenam objetos indexados por chaves, e facilita a busca por esses objetos a partir de suas chaves.

**Orientado a Grafos:** diferente de outros bancos de dados NoSQL, esse está diretamente relacionado a um modelo de dados estabelecido, o modelo de grafos. A ideia desse modelo é representar os dados e/ou o esquema dos dados como grafos dirigidos, ou como estruturas que generalizem a noção de grafos. O modelo de grafos é aplicável quando “informações sobre a interconectividade ou a topologia dos dados são mais importantes, ou tão importante quanto os dados propriamente ditos”. Possui três componentes básicos: os nós (são os vértices do grafo), os relacionamentos (são as arestas) e as propriedades (ou atributos) dos nós e relacionamentos.

Vejam que a questão mistura os dois conceitos o que torna a alternativa incorreta.

**Gabarito:** E



**30. ANO: 2016 BANCA: CESPE CONCURSO: FUNPRES CARGO: ESPECIALISTA ÁREA: TECNOLOGIA DA INFORMAÇÃO (TI)**

A respeito dos fundamentos e das principais tecnologias relacionadas à computação em nuvem, julgue os próximos itens.

[90] Hadoop e Elasticsearch são exemplos de tecnologias que permitem a computação em nuvem.

[91] A computação em nuvem permite o processamento de dados de maneira distribuída em máquinas com diferentes arquiteturas físicas.

**Comentário:** Vamos comentar cada uma das alternativas acima. Começando pela definição de Hadoop e ElasticSearch.



O Hadoop é uma implementação de código aberto do paradigma de programação MapReduce. MapReduce é um paradigma introduzido pelo Google para processar e analisar grandes conjuntos de dados. Os programas que são desenvolvidos nesse paradigma realizam o processamento paralelo de conjuntos de dados e podem, portanto, ser executados em diferentes servidores sem muito esforço. A razão para a escalabilidade desse paradigma é a natureza intrinsecamente distribuída do funcionamento da solução. Uma grande tarefa é dividida em várias pequenas tarefas que são então executadas em paralelo em máquinas diferentes e então combinadas para chegar à solução. Os exemplos de uso do Hadoop são analisar padrões de usuários em sites de e-commerce e sugerir novos produtos para os mesmos.

ElasticSearch é uma engine de busca open-source, amplamente distribuível, facilmente escalável para uso empresarial ou pessoal. Acessível através de uma API extensa e bem elaborada, o ElasticSearch pode acionar pesquisas extremamente rápidas que suportam diversos tipos de integração.

90 Também conhecida no Brasil como computação nas nuvens ou computação em nuvem, a cloud computing se refere, essencialmente, à noção de utilizarmos, em qualquer lugar e independente de plataforma, as mais variadas aplicações por meio da internet com a mesma facilidade de tê-las instaladas em computadores locais. Dentro deste contexto, tanto Hadoop quanto ElasticSearch são ferramentas que contribuem para a infraestrutura para provimento de serviços de computação na nuvem. Sendo assim alternativa encontra-se correta.

91. O uso das ferramentas citadas no item anterior vai permitir um grande paralelismo no processamento por meio da distribuição entre os diferentes nós ou hosts do sistema. Mais uma alternativa correta.

**Gabarito:** C C

