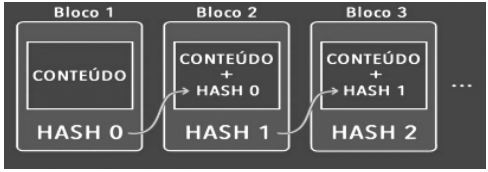
**141** [**Q1756824**](https://www.qconcursos.com/questoes-de-concursos/questoes/42ff92ac-d1) [Banco de Dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Desenvolvimento de Sistemas](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-desenvolvimento-de-sistemas)

Julgue o item a seguir, relativo a ***blockchain* e *smart contracts*.**

De acordo com a seguinte figura, *blockchain* corresponde a uma lista ordenada de blocos em que cada bloco em um *blockchain* é encadeado ao bloco anterior, de maneira a conter um *hash* da representação do bloco anterior, e, assim, as transações históricas no *blockchain* não podem ser excluídas ou alteradas sem se invalidar a cadeia de*hashes*.



O que é blockchain? **Uma rede que marca o tempo das transações, colocando-as em uma cadeia contínua no ‘hash’, formando um registro que não pode ser alterado sem refazer todo o trabalho.**

Fonte: Satoshi Nakamoto, 2008, ***Bitcoin: um sistema financeiro eletrônico peer-to-peer.***

Essa organização, em que parte da informação se encontra no dado anterior e parte no dado seguinte, faz com que a blockchain seja uma tecnologia segura.

Imaginem uma senha de papel, em que, além do número que representa a ordem de chegada, também exista informações sobre e data e a hora que a senha foi retirada. Será muito difícil inserir dados no meio dessa cadeia, já que cada uma dessas senhas apresenta informações sobre sua procedência, não podendo ser duplicadas. De forma similar é o funcionamento do blockchain.

**Hash =** é um algoritmo matemático que transforma qualquer bloco de dados em uma série de caracteres de comprimento fixo. É o hash que garante a integridade dos dados.

Algumas características do blockchain:

* Tecnologia de registro distribuído.
* Descentralização como medida de segurança.
* Sem intermédio de terceiros (peer-to-peer).
* Constantemente em crescimento.
* Blocos adicionados de modo linear e cronológico.
* Cada bloco em um *blockchain* é encadeado ao bloco anterior.
* Transações agrupadas e armazenadas em unidades encadeadas e interligadas por meio de códigos hash.
* Utiliza criptografia assimétrica.
* As transações realizadas no sistema são registradas em uma espécie de livro de registros, de acesso público, permitindo a rastreabilidade das transações na rede (Immutable Ledger).

Gabarito: Certo

A palavra Blockchain em tradução livre significa cadeia de blocos. E é exatamente isso que essa tecnologia representa. Calma! Você já vai entender! Trata-se de uma tecnologia que funciona como uma espécie de livro de registros, totalmente digital, e que pode ser auditado por vários usuários ao mesmo tempo. Introduzido pelo Bitcoin (sendo uma de suas principais contribuições), Blockchain é a tecnologia utilizada nas transações com criptomoedas. Comparando a um livro de registros de uma empresa (livro-razão) podemos traçar o seguinte paralelo: Enquanto o livro-razão registra todas as transações realizadas por uma empresa, o Blockchain registra todas as transações realizadas com uma criptomoeda, a exemplo do Bitcoin. Assim cada criptomoeda possui o seu “livro”.

O registro de informações é feito em ordem linear e cronológica, em uma espécie de lista de blocos interligados. Cada bloco registra um conjunto de informações, carregando seu próprio número de identificação e o número de identificação do bloco anterior – por isso trata-se de uma cadeia de blocos. Principais características do Blockchain:

• Sem servidor central: é um sistema descentralizado, o qual usuários em todo o mundo podem auditar ao mesmo tempo. Isso acontece, pois a rede é do tipo peer-to-peer (p2p) que se trata de um sistema para compartilhamento de dados sem a necessidade de um servidor central.

• O sistema funciona por meio de uma estrutura com dezenas de milhares de computadores espalhados no mundo todo. No entanto, é importante saber que tais computadores não estão interligados por meio de redes ou qualquer servidor centralizado (o que caracteriza um sistema descentralizado). Esses computadores possuem três principais atribuições: − Validação das transações na rede e verificação dos blocos; − Localização dos blocos e comunicação para o restante da estrutura − Lançamento de alertas a respeito de tentativas de fraudes.

• Peer-to-peer: significa “ponto a ponto”. Cada computador de um usuário é um “ponto”, os quais conectam-se entre si formando uma rede descentralizada. Essa tecnologia confere segurança e estabilidade às transações com criptomoedas– impedindo o gasto duplo (double spend). Gasto duplo seria o seguinte: já que a criptomoeda só existe digitalmente (ninguém a pega na mão), poderia existir “picaretas”, vendendo a mesma unidade de BTC para várias pessoas ao mesmo tempo.

• Impede a reprodutibilidade infinita. Garante que novas unidades da criptomoeda só sejam emitidas com base nos critérios definidos e até o limite estabelecido (Bitcoin = 21 milhões).

alguns dados que são registrados no Blockchain quando é realizada uma transação com criptomoeda:

• Data e hora de realização da transação.

• Quantidade transacionada.

• O usuário que enviou e o usuário que recebeu a transação. Atenção: a identificação é feita por meio de endereço de carteiras. Não são fornecidos nome e documentos dos usuários. Isso é algo importante, pois, ao mesmo tempo em que se dá publicidade às transações com criptomoedas, de modo que possam ser auditadas, preserva-se a identidade dos negociantes.

• Número do bloco, no qual foi registrada a transação.

• Número que confirma a autenticidade de cada transação (chamado de hash).

Apesar de ser muito presente no mercado de criptomoedas, a tecnologia do Blockchain está sendo utilizada em várias situações, como autenticação de documentos, uma vez que é possível sua utilização para a criação de assinaturas digitais. Outros usos são: transações financeiras em geral, registro de votos em eleições e até mesmo registro de imóveis.

Uma grande preocupação dos usuários de criptomoedas é a possibilidade de double spend. A esse respeito, marque a alternativa incorreta.

a) Não existe double spend para transações realizadas com moeda em espécie

b) Não existe double spend para transações realizadas através das redes bancárias.

c) A lógica do peer to peer permite que seja possível realizar double spend com criptomoedas.

d) A lógica do peer to peer não permite que seja possível realizar double spend com criptomoedas.

e) A transação realizada por meio das redes bancárias impossibilita o double spend.

A questão pede a alternativa incorreta. As transações com criptomoedas são protegidas por um sistema do tipo peer-to-peer (p2p) que significa “ponto a ponto”. Cada computador de um usuário é um “ponto”, os quais conectam-se entre si formando uma rede descentralizada. Essa tecnologia confere segurança e estabilidade às transações com criptomoedas – impedindo o gasto duplo (double spend). Gasto duplo seria o seguinte: um usuário vendendo a mesma unidade de BTC para várias pessoas ao mesmo tempo. Letra c.

**142** [**Q1756788**](https://www.qconcursos.com/questoes-de-concursos/questoes/4283cb9e-d1) [Linguagens de programação ,](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-programacao/linguagens-de-programacao)[Python](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-programacao/python) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Desenvolvimento de Sistemas](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-desenvolvimento-de-sistemas)

No que se refere ao **uso de Jenkins**, julgue o item seguinte.

É possível consumir recursos pela API disponibilizada pelo Jenkins, utilizando-se Python para buscar e instalar artefatos.

CERTO

Jenkins = é uma ferramenta de integração contínua de código aberto, fácil de instalar e configurar, ele oferece tbm APIs que pode utilizar os recursos utilizando a linguagem de programação Python.

**Jenkins**é um servidor de automação de código aberto. Ele ajuda a automatizar as partes do desenvolvimento de software relacionadas à construção, teste e implantação, facilitando a integração e entrega contínuas. É um sistema baseado em servidor que é executado em contêineres de servlet, como o Apache Tomcat.

Jenkins é uma ferramenta de integração contínua e entrega contínua.

O que é Jenkins e para que serve?

**Jenkins** é um sistema de servidor de automação de código aberto, com integração e entrega contínuas, que permite a criar, testar e implementar softwares, tornando, assim, o processo de desenvolvimento mais ágil e eficiente.

**143** [**Q1754056**](https://www.qconcursos.com/questoes-de-concursos/questoes/b97a0a2b-cd) [Administração de banco de dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/administracao-de-banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Em uma organização, foram elencadas as seguintes necessidades:

I identificar os *data stewards* como parte da governança de dados e do planejamento da gestão dos dados;

II implementar a gestão de qualidade de dados e a gestão de arquitetura de dados;

III gerenciar fatores ambientais tais como pessoas e tecnologias afetas à governança de dados.

A partir da situação hipotética apresentada, julgue os itens a seguir, relativos à governança de dados.

A necessidade II descreve duas **áreas de conhecimento do DMBOK** que, uma vez implementadas, ajudam a organização a entender as necessidades de informações da empresa, assim como a avaliar a adequação dos dados ao seu uso pretendido.

Gabarito: **CERTO**

Áreas de conhecimento na Gestão de Dados e suas definições:

**Gerenciamento de arquitetura de dados:**Trata-se de entender quais os requisitos do seu projeto de dados, ou seja, entender quais os dados que são necessários, de onde eles vêm e por onde terão de passar até chegar ao seu dashboard. É definir o caminho deste dado.

**Desenvolvimento dos dados:** Trata-se de analisar os requisitos dos dados, implantar o seu modelo de dados, definir como será a manutenção destes modelos de dados, projetar estruturas de bancos de dados para suportar suas necessidades, projetar como será o versionamento e integração de dados e modelo de dados, projetar planos de testes, projetar planos de migração entre outras atividades.

**Gerenciamento de operações de dados:** Trata-se do planejamento, controle, manutenção e suporte ao ativo dado, durante todo o seu ciclo de vida, ou seja, desde sua aquisição, passando por sua exibição, até a eliminação desse dado. Vale ressaltar que um planejamento de Recuperação de desastre é de extrema importância para qualquer projeto.

Gerenciamento de segurança de dados: É necessário planejar e projetar uma estrutura capaz de garantir a privacidade, confidencialidade e acesso apropriado ao dado. Consiste no planejamento, desenvolvimento e execução de políticas e procedimentos para assegurar a devida autenticação, autorização, acesso e auditoria nos ativos de dados e informações. (DMBOK, 2012).

**Gerenciamento de dados mestre e referência:** Trata-se de projetar uma estrutura que promova uma “Central Única da Verdade”, garantindo um versão consistente e confiável do dado, onde possa ser distribuído (ou compartilhado) para outros contextos.

**Gerenciamento de DW (Data Warehousing) e BI (Business Intelligence):** Trata-se de planejar e projetar modelos de dados que permitam a geração de informações para tomada de decisão sob várias perspectivas (dimensões).

**Gerenciamento da documentação e conteúdo:** Trata-se de planejar e projetar a implantação e a gestão a dados não estruturados, ou seja, que estão fora de um banco de dados. É importante que seja definido um plano para armazenamento, proteção e acesso a estes dados.

**Gerenciamento de Metadados:** Primeiro você precisa saber o que é metadados, correto? Pois bem, metadados é o dado a respeito de outro dado, ou seja, são informações que complementam um dado. Por exemplo: 2 maçãs(dado); foto destas mesmas 2 maçãs é(metadados).

**Gerenciamento de Qualidade de dados**: Trata-se de planejar e projetar o saneamento do dado, provendo qualidade ao mesmo, para que este dado possa gerar informações confiáveis para suportar a tomada de decisão.

**Gestão da Arquitetura de Dados:** é definida como uma função, responsável por gerenciar e manter os mecanismos necessários para gerir os conceitos, estruturas e demais metadados que compõe a Arquitetura de Dados da Empresa.

Fontes: <https://abracd.org/framework-dama-para-governanca-de-dados/>

<https://www.blrdata.com.br/arquitetura-de-dados-consultoria>

GABARITO CERTO

Gestão de Qualidade de dados

O objetivo é planejar, **implementar** **e controlar atividades que apliquem técnicas de gerência de qualidade**de dados para medir, avaliar, melhorar e garantir a adequação dos dados ao seu uso pretendido.

Gestão da arquitetura de dados

Segundo o DMBOK (2009), o objetivo da Gestão da Arquitetura de Dados é: **Entender as necessidades de informação da empresa.**

<https://www.fumsoft.org.br/comunica/arquivos/uma_visao_sintetica_e_comentada_do_dmbok_fumsoft_carlos_barbieri.pdf>

Gabarito: Certo para não assinante.

Além de concursanda, sou formada em Letras pela UERJ e corrijo redações e discursivas no valor de dez reais.Qualquer dúvida, basta entrar em contato pelo 21987857129.

**144** [**Q1754055**](https://www.qconcursos.com/questoes-de-concursos/questoes/b977176f-cd) [Administração de banco de dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/administracao-de-banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Em uma organização, foram elencadas as seguintes necessidades:

I identificar os *data stewards* como parte da governança de dados e do planejamento da gestão dos dados;

II implementar a gestão de qualidade de dados e a gestão de arquitetura de dados;

III gerenciar fatores ambientais tais como pessoas e tecnologias afetas à governança de dados.

A partir da situação hipotética apresentada, julgue os itens a seguir, relativos à governança de dados.

A implementação da gestão da necessidade III deve incluir a cultura organizacional e excluir os processos, uma vez que estes últimos não são abrangidos pelo DMBOK.

Através da **GD (Governança de Dados)**, as empresas hoje também definem mecanismos para analisar os processos que se abastecem ou produzem os dados, criando um sentido maior de qualidade conjunta entre esses dois elementos seminais (**dados e processos**) e contribuindo para a valorização desses ativos, através do pleno conhecimento da cadeia produtiva de informação e conhecimentos.

Prof. Thiago Cavalcanti (Estratégia Concursos)

*Rumo ao BRB.*

De forma resumida:

Para fazer um controle eficiente não pode excluir os processos, pois sem processos bem robustos e definidos não tem como fazer uma boa gestão dos dados.

**GAB: ERRADO**

a questão está tratando do HEXÁGONO DE FATORES AMBIENTAIS:

Os Fatores Ambientais , representados pelos hexágonos, sofreram algumas modificações. No DAMA-DMBoK® apareciam: Organização&Cultura; Atividades; Entregáveis; Papéis&Responsabilidades; Práticas&Técnicas e Tecnologia.            No DAMA-DMBoK®V2 permanecem Papéis&Responsabilidades; Atividades, surge *Ferramentas*; Organização&Cultura; Técnicas(sem práticas) e Entregáveis. A palavra Tecnologia entra com Processo e Pessoas, num nível de agregação acima, ficando Pessoas(Papéis e Responsabilidades); Processos(Atividades e Técnicas) e Tecnologia(Ferramentas e Entregáveis).

<https://pt.linkedin.com/pulse/vis%C3%A3o-comparativa-e-comentada-do-dama-dmbokv2-parte-ii-barbieri>

O tratamento ético dos dados obviamente inclui seguir a lei. Também influencia como os dados são analisados, interpretados e aproveitados interna e externamente. **Uma cultura organizacional**que valorize o comportamento ético não apenas terá um código de conduta, mas também garantirá a existência de controles claros de comunicação e governança para apoiar os usuários que tomam conhecimento de práticas ou riscos antiéticos. Os usuários precisam poder relatar essas circunstâncias sem medo de retaliação. Melhorar o comportamento ético de uma organização em relação aos dados geralmente requer um processo formal de Gerenciamento de Mudanças Organizacionais (Organizational Change Management - OCM).

**Integração e Interoperabilidade de Dados** A Integração e Interoperabilidade de Dados**inclui processos** relacionados à movimentação e consolidação de dados dentro e entre repositórios de dados, aplicativos e organizações.

**Data Warehousing e Business Intelligence** O Data Warehousing e o Business Intelligence **incluem os processos**de planejamento, implementação e controle para gerenciar os dados de suporte à decisão e permitir que os profissionais do conhecimento obtenham valor dos dados por meio de análise e relatório.

<https://repositorio.enap.gov.br/bitstream/1/5008/2/M%C3%B3dulo%202%20-%20Princ%C3%ADpios%2C%20import%C3%A2ncia%20e%20desafios%20do%20Gerenciamento%20de%20Dados.pdf>

**145** [**Q1754054**](https://www.qconcursos.com/questoes-de-concursos/questoes/b972cdf8-cd) [Administração de banco de dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/administracao-de-banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Em uma organização, foram elencadas as seguintes necessidades:

I identificar os *data stewards* como parte da governança de dados e do planejamento da gestão dos dados;

II implementar a gestão de qualidade de dados e a gestão de arquitetura de dados;

III gerenciar fatores ambientais tais como pessoas e tecnologias afetas à governança de dados.

A partir da situação hipotética apresentada, julgue os itens a seguir, relativos à governança de dados.

Atender a necessidade I é importante, pois os *data stewards* são responsáveis, dentro da área de negócios, pelo controle e uso dos dados, assim como pela gestão dos metadados, que são elementos que ajudam a entender os objetos, a sua composição e seu relacionamento.

GABARITO: **CERTO**

**DATA STEWARD**: significa Gerente Estratégico de Dados, antes que os mais desavisados pensem que Data Steward significa literalmente mordomo de dados.

Esse profissional surge em resposta ao cenário descrito acima e necessita ter já desenvolvidas algumas capacitações e potencialidades para que represente os interesses dos produtores e dos consumidores de dados; seja o curador dos dados; represente os interesses da organização quando o assunto for dados; seja responsável pela qualidade e pelo uso dos dados; saiba investir na preparação e no nivelamento dos recursos de dados sob sua responsabilidade; tenha certeza de que os recursos de dados estão de acordo com as necessidades do negócio, garantindo a qualidade dos dados e os seus metadados; trabalhe em parceria com os profissionais de gerenciamento de dados e, deste modo, funcione como elemento condutor das necessidades de dados da área de negócios junto à TI.

FONTE: <https://tiinside.com.br/01/08/2011/data-steward-e-tendencia-profissional/>

***Steward*** é um termo utilizado para designar aquele que é responsável por algo em nome de outra pessoa, ou de um grupo. Neste contexto, podemos dizer que os responsáveis pelos dados de uma determinada área ou assunto, sejam eles do negócio ou da área de tecnologia, são conhecidos como *Data Stewards*, ou administradores de dados.

O principal objetivo do *Data Stewardship* é garantir que os dados da organização sejam confiáveis, íntegros e de boa qualidade, assim como seus metadados.

<https://pt.linkedin.com/pulse/data-stewardship-o-que-%C3%A9-cec%C3%ADlia-garcia>

**146** [**Q1754053**](https://www.qconcursos.com/questoes-de-concursos/questoes/b96f50b4-cd) [Administração de banco de dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/administracao-de-banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Julgue o item seguinte, referente a **conceitos e especificidades de MDM (*master data management*).**

A descrição da arquitetura de dados de uma organização inclui uma coleção integrada de documentos de projeto mestre em diferentes níveis de abstração, excluindo-se dessa coleção os padrões que determinam o modo como os dados são coletados, armazenados, organizados, usados e removidos.

✅Gabarito(Errado)

Uma arquitetura de dados descreve como os dados são gerenciados, desde a coleta até a transformação, a distribuição e o consumo . Ela define o plano para os dados e a maneira como eles irão fluir pelos sistemas de armazenamento de dados.

Fonte: <https://www.ibm.com/br-pt/topics/data-architecture>

**147** [**Q1754050**](https://www.qconcursos.com/questoes-de-concursos/questoes/b964788f-cd) [Administração de banco de dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/administracao-de-banco-de-dados) **Ano:**2021 **Banca:**[CESPE / CEBRASPE](https://www.qconcursos.com/questoes-de-concursos/bancas/cespe-cebraspe) **Órgão:**[SERPRO](https://www.qconcursos.com/questoes-de-concursos/institutos/serpro) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Julgue o item seguinte, referente a **conceitos e especificidades de MDM (*master data management*).**

A fim de gerenciar a qualidade de dados, o MDM se serve de ferramentas de validação dos dados que ajudam na visualização de todo o fluxo de gestão dos dados mestres, o que torna possível, de maneira rápida, a identificação de quaisquer desvios em relação à política de dados da empresa.

O conceito Master Data Management (MDM) é um conjunto de processos que aponta qual é a informação mais importante dentro da estratégia de dados de uma empresa. Depois de identificar, \*cria uma única fonte, organiza e oferece melhorias nos procedimentos de dados em questão.

De fato, o MDM se serve de ferramentas de validação que ajudam nessa identificação. O que torna o item errado é que esse processo naturalmente, não é feito de maneira rápida para quaisquer desvios em relação a política da empresa. Vão existir análises complexas e incrementais que exigem um trabalho minucioso. É preciso de uma equipe de suporte qualificada para entender o uso dos dados pelos sistemas e prover mecanismos de melhoria.

fonte: tecconcursos

**148** [**Q1754048**](https://www.qconcursos.com/questoes-de-concursos/questoes/b95c6836-cd) [Administração de banco de dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/administracao-de-banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Julgue o item seguinte, referente a conceitos e especificidades de MDM (*master data management*).

Como o MDM envolve controle sobre os valores e identificadores de dados mestres, é correto afirmar que um de seus objetivos é reduzir o risco de identificadores ambíguos.

**Master Data Management (MDM)** compreende um conjunto de processos e ferramentas que definem consistentemente e gerenciam os  (i.e. ) de uma  (que pode conter ). MDM tem o objetivo de prover processos para coletar, agregar, combinar, consolidar, garantir a qualidade, persistir e distribuir estes dados a toda a organização para garantir a consistência e o controle nas futuras manutenções e uso desta informação.

Basicamente, MDM procura garantir que uma organização não utilize múltiplas (potencialmente inconsistentes) versões de um mesmo dado mestre em diferentes partes de sua operação, o que ocorre normalmente em grandes organizações.

Processos normalmente usados em soluções de MDM incluem identificação da origem dos dados, coleta, , , administração de regras, detecção e correção de erros, consolidação dos dados, armazenamento, distribuição, classificação, serviços de taxonomia, criação de itens mestre, mapeamento de esquema, codificação de produtos, enriquecimento dos dados e governança dos dados.

As ferramentas incluem , , um , , um , , , , federação de dados e . Uma das mais novas ferramentas, o gerenciamento virtual de dados mestres (também chamado virtual mdm) utiliza virtualização de dados e um servidor de metadados persistente para implementar uma hierarquia de MDM automatizada em vários níveis.

A seleção de entidades consideradas no MDM depende muito da natureza de uma organização. No caso comum de empresas comerciais, MDM pode aplicar a entidades como cliente (), produto (), empregado, fornecedor e. Processos MDM identificar as fontes das quais para coletar descrições dessas entidades.

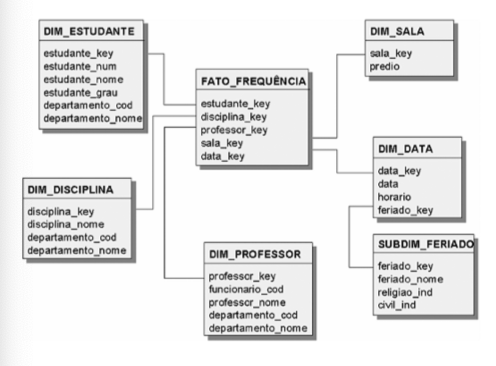
No decurso da transformação e normalização, os administradores adaptar descrições se conformar com formatos padrão e os domínios de dados, tornando possível a **remover**instâncias duplicadas de qualquer entidade. Esses processos geralmente resultam em um repositório MDM organizacional, a partir do qual todas as solicitações para uma instância determinada entidade produzir a mesma descrição, ***independentemente***das fontes originárias e os destinos requerentes.

MDM tem o objetivo de prover processos para coletar, agregar, combinar, consolidar, garantir a qualidade, persistir e distribuir estes dados a toda a organização para garantir a consistência e o controle nas futuras manutenções e uso desta informação.

CERTO

A questão usou "reduzir o risco de identificadores ambíguos", achei que deveria ser "eliminar o risco..."

**149** [**Q1754047**](https://www.qconcursos.com/questoes-de-concursos/questoes/b9583adc-cd) [DW - Data Warehouse ,](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/dw-data-warehouse)[Modelagem de dados ,](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/modelagem-de-dados)[BI (Business Intelligence)](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/bi-business-intelligence) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)



Internet: <sandrorgguimaraes.blogspot.com> (com adaptações).

Tendo como referência inicial esse modelo, e considerando que, para representar a quantidade de faltas em valores inteiros, seja inserida em **FATO\_FREQUÊNCIA** a métrica qtd\_faltas, em que se possam realizar *drill up* e a soma de seus valores ao longo do tempo, julgue o item a seguir, à luz dos conceitos afetos à modelagem dimensional.

Todos os elementos descritos em **FATO\_FREQUÊNCIA** fazem parte de uma única chave primária composta, sendo cada elemento uma chave estrangeira oriunda de cada dimensão a ela vinculada

**Gab: CERTO**

O enunciado da questão é autoexplicativo.

Tabelas **FATO** são um "compilado" das chaves estrangeiras das tabelas **DIMENSÃO** que estão vinculadas à ela.

Desta forma, as chaves estrangeiras de cada tabela dimensão compõe a chave composta da tabela **"FATO\_FREQUÊNCIA".**

Questão que talvez ajude a elucidar:

Q580222 - Para melhor manter o controle sobre identificadores de registro de ambientes de*data warehouse* (armazém de dados), em geral recomenda-se a geração de chaves substitutas (*surrogate keys*).

Assim, cada junção entre as tabelas de dimensão e tabelas fato em um ambiente de*data warehouse* deve se basear nessas chaves substitutas, e não nas chaves naturais existentes. **(GAB: CERTO)**

**INFORMAÇÕES ADICIONAIS:**

-Chave estrangeira é o atributo ou conjunto de atributos que se refere ou é relacionado com alguma chave primária ou única de uma tabela, podendo ser inclusive na mesma tabela.

-É possível ter mais de uma ou nenhuma em uma tabela.

Concordo que, na tabela fato, a "maioria" dos elementos contidos formam a chave composta (que são as chaves estrangeiras), porém afirmar que "todos" os elementos formam a chave composta é forçar a barra, uma vez que o comando da questão relata claramente que a métrica qtd\_faltas foi adicionada à tabela fato\_frequências, sendo assim, sabe-se que esta métrica não constitui chave primária. Portanto, entendo que tal questão caberia recurso, pois generalizou todos os elementos como chave primária, sendo que qtd\_faltas não é uma chave\_primária.

Olá colegas,

Acrescentando conhecimento, diferenças entre chave primária e chave estrangeira:

* **Chave Primária: identificação**

 Chave Primária (PRIMARY KEY) **identifica** uma tabela por meio da **unicidade** de um registro e por sua obrigatoriedade de **não** ser **nulo**.

* **Chave Estrangeira: relacionamento**

Chave Estrangeira nada mais é do que a Chave Primária de uma tabela ‘colocada’ em outra tabela.

Fonte: <https://omanualdofreelancer.com/diferenca-chave-primaria-e-chave-estrangeira/>

**gab. CORRETO**

CERTO

*Se analisar a entidade****FATO\_FREQUÊNCIA****as chaves presentes são estrangeiras de outras tabelas que por si só são chaves primarias das outra tabelas*

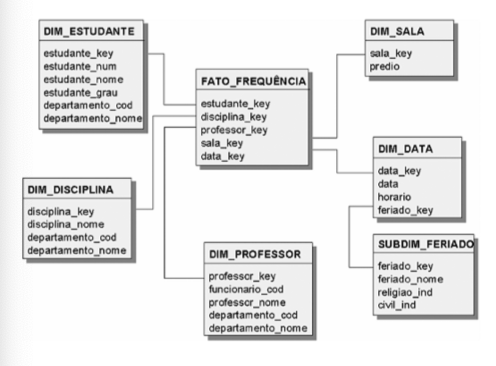
**Chave Primária**

* **é utilizada da identificar de forma única cada linha numa tabela**. Pode fazer parte do próprio registo atual ou pode ser um campo artificial (um que não tenha nada que ver com o registo atual). Uma chave primária pode ser composta por um ou mais campos numa tabela. Quando são utilizados vários campos como chave primária, são denominados por chave composta

**obs**:

* Em uma tabela de um banco de dados relacional, sua chave primária poderá se tornar uma chave estrangeira em outra tabela com a qual ela está se relacionando

**150** [**Q1754046**](https://www.qconcursos.com/questoes-de-concursos/questoes/b9545e0a-cd) [Modelagem de dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/modelagem-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)



Internet: <sandrorgguimaraes.blogspot.com> (com adaptações).

Tendo como referência inicial esse modelo, e considerando que, para representar a quantidade de faltas em valores inteiros, seja inserida em **FATO\_FREQUÊNCIA** a métrica qtd\_faltas, em que se possam realizar *drill up* e a soma de seus valores ao longo do tempo, julgue o item a seguir, à luz dos conceitos afetos à modelagem dimensional.

A métrica qtd\_faltas é aditiva

**Gabarito: CERTO**

**Medidas/Fatos Aditivos:**

Uma métrica é dita aditiva quando faz sentido sumarizá-la, somando seus valores ao longo de qualquer dimensão.

Por exemplo, o caso da questão:

Se eu quiser obter a quantidade de faltas de um estudante no ano inteiro, basta somar essas quantidades de faltas ao longo dos 12 meses.

Os 4 tipos de métricas são:

* Aditivas - métricas que permitem operações matemáticas como soma e subtração por todas as dimensões.
* Derivadas -São métricas que já estavam na fato e que são calculadas, criando uma nova métrica, que chamamos de derivada.
* Semi-aditivas - métrica semi-aditiva pode ser somada por todas as dimensões exceto a tempo.
* Não-aditivas - cálculo feito em tempo de execução, que não podem ser somadas por nenhuma dimensão.

*fonte: https://rafaelpiton.com.br/blog/data-warehouse-tipos-metricas/*

**151** [**Q1754043**](https://www.qconcursos.com/questoes-de-concursos/questoes/b948deb1-cd) [Banco de Dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Julgue o próximo item, relativo à **qualidade de dados**.

Uma boa prática em relação à qualidade de dados é verificar a precisão dos dados, validando-se se estão corretamente representados, e a sua consistência, avaliando-se se há integridade cruzada entre duas ou mais fontes que armazenem o mesmo dado.

O conceito de **qualidade de dados** indica o quanto as fontes de **dados** atendem às demandas de informação das áreas de negócio. Isto envolve conhecer a **qualidade de dados** atual, determinar a **qualidade** desses **dados** desejada e ajustar a **qualidade** atual em função do que se almeja para o negócio.

Algumas dimensões principais da qualidade dos dados incluem:

* ***plenitude*** indica se os dados coletados são suficientes para tirar conclusões. Isso pode ser avaliado assegurando que não haja informações ausentes em nenhum conjunto de dados.
* ***Consistência*** garante que os dados em todos os sistemas de uma organização sejam sincronizados e reflitam as mesmas informações. Um exemplo de dados consistentes inclui o registro da data de remessa no mesmo formato de data da planilha de informações de um cliente.
* ***Precisão*** implica se os dados que foram coletados representam com precisão o que deveriam. Isso pode ser medido contra  contra regras de negócios definidas pelo usuário.
* ***oportunidade*** significa que os dados estão disponíveis como e quando se espera para facilitar a tomada de decisão baseada em dados. Muitas empresas estão aproveitando ferramentas que oferecem suporte à integração de dados em tempo real para obter conhecimento de negócios atualizado. No entanto, é importante observar que a qualidade e a integração dos dados andam de mãos dadas. Antes de embarcar na integração de dados, os dados precisam ser perfilados e limpos, o que irá, por sua vez, acelerar o desenvolvimento de  e fluxos de trabalho.
* ***Singularidade*** envolve garantir que não haja duplicatas presentes nos dados. Por exemplo, a falta de dados exclusivos pode resultar no envio de vários emails para um único cliente devido a registros duplicados.
* ***Validade***mede se os dados atendem aos padrões ou critérios definidos pelo usuário de negócios. Por exemplo, uma empresa pode colocar uma verificação de qualidade de dados corporativos no campo de quantidade do pedido, ou seja, 'Quantidade do Pedido> = 0', pois a quantidade do pedido negativa implica em informações inválidas.

Deveria ter uma forma de salvar os comentários em uma pasta específica de cada matéria e assunto, seria mais dinâmico, ao invés de ficar copiando e colando em um word.

**152** [**Q1754042**](https://www.qconcursos.com/questoes-de-concursos/questoes/b945086e-cd) [Banco de Dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Julgue o próximo item, relativo à **qualidade de dados**.

Definir processos de qualidade para modelos de dados implica analisar as regras de negócios fundamentais dos processos, bem como observar a qualidade dos dados, com a finalidade de garantir a conformidade da regra com o processo.

"A análise de dados é uma etapa importante em qualquer processo decisório dentro de uma empresa. Mas, para que tais informações sejam relevantes e possam garantir retornos positivos para o negócio, é essencial que a organização se **certifique da qualidade de dados**.**Isso significa consultar a procedência dos fatos, verificar a composição dos elementos, avaliar a consistência das informações disponíveis, entre outros procedimentos."**

Recomendo a leitura:https://blog.gs1br.org/qualidade-de-dados-o-que-e-e-qual-a-importancia-para-negocios/

gabarito certo

**153** [**Q1754041**](https://www.qconcursos.com/questoes-de-concursos/questoes/b941900c-cd) [Banco de Dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Julgue o próximo item, relativo à **qualidade de dados**.

Executar o *profiling* dos dados é similar a fazer uma fotografia inicial do estado dos dados em um domínio ou em uma área de assunto da organização.

Gabarito: **CERTO**

Profiling = fazer um perfil

É um método revolucionário de análise e resposta comportamental: a análise de reações, respostas e microexpressões diz muito sobre a pessoa com quem estamos a intera

O *Profiling* é a resposta para as empresas que querem marcar a diferença no mercado cada vez mais competitivo!

As técnicas de vendas convencionais fazem parte do passado!

Os *Profiling* Specialists da SWAT vão ensinar a sua equipa a identificar os diferentes perfis de clientes.

Ao saberem exatamente com quem estão a lidar, os seus colaboradores vão poder adaptar a abordagem e proporcionar ao cliente, exactamente, aquilo que ele espera!gir!

No jargão tecnológico, “profiling” é a **utilização de algoritmos complexos para criar modelos ou perfis** – que tem caráter geral – a partir de dados específicos. Esses algoritmos processam enormes quantidades de informação e fazem correlações entre dados inicialmente desconexos (por exemplo, dados de conexão, biométricos, demográficos, entre outros), ..

**O que é Data Profiling?**

Em linhas gerais, Data Profiling é uma atividade de Qualidade de Dados que é caracterizada pelo uso de técnicas analíticas sobre os dados para conhecer o seu conteúdo, estrutura e qualidade. Em outras palavras, é uma espécie de exame (diagnóstico) multidisciplinar a respeito dos dados existentes.

Diversos motivos podem ser citados para justificar a criação de um perfil de dados, mas geralmente é parte de um processo definido que visa determinar a qualidade dos dados em um projeto maior, ou até mesmo de um Data Lake inteiro.

**Técnicas de data profiling**

Essencialmente, existem quatro métodos gerais pelos quais as ferramentas de Data Profiling realizam suas tarefas:

**Column profiling**: analisa uma tabela e conta o número de vezes que cada valor aparece em cada coluna. Uma utilização desse método é encontrar padrões e distribuição de frequência em uma coluna de dados.

**Cross-column profiling**: esse método possui dois processos. A análise de chave examina coleções de valores de atributos, procurando uma possível chave primária. A análise de dependência é um processo mais complexo que determina se há relacionamentos ou estruturas incorporadas em um conjunto de dados. Esses métodos, combinados ou não, auxiliam a expor dependências entre atributos de dados na mesma tabela, facilitando o mapeamento de dados em um Data Lake.

**Cross-table profiling**: utiliza a análise de chave estrangeira, que é a identificação de registros órfãos e a determinação de diferenças semânticas e sintáticas, para examinar os relacionamentos dos conjuntos de colunas em diferentes tabelas. Isso pode ajudar a reduzir a redundância, mas também identificar conjuntos de valores de dados que podem ser mapeados juntos.

Não menos importante, a **validação de regra de dados** utiliza a criação de perfil de dados de maneira proativa para verificar se as instâncias e conjuntos de dados estão em conformidade com as regras predefinidas.

Na  , o ***Profiling*** ("perfil de programa", "perfil de software") é uma forma de  que mede, por exemplo, o espaço (memória) ou  tempo  , o  ou a frequência e duração das chamadas de função. Mais comumente, as informações de perfil servem para auxiliar na  e, mais especificamente, na  .

**154** [**Q1754040**](https://www.qconcursos.com/questoes-de-concursos/questoes/b93e8f88-cd) [Banco de Dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Julgue o próximo item, relativo à **qualidade de dados.**

A acurácia na qualidade de dados está diretamente associada à de duplicação, uma vez que indica que há exclusividade da fonte de dados e de suas entidades, de forma a garantir a precisão dos dados na vida real.

**A acurácia e a veracidade avaliam o quão próximo ou distante está a informação da realidade**

**O que é acurácia:**

Exatidão e **precisão numa medição ou no resultad**o apresentado por um instrumento de medição. Física. **Proximidade ent**re o resultado de um instrumento de medida e o verdadeiro valor do que foi medido.

A acurácia na qualidade de dados está diretamente associada à de duplicação, uma vez que **indica que há exclusividade da fonte de dados e de suas entidades**, de forma a garantir a precisão dos dados na vida real.

A acurácia não garante que foi usado uma fonte exclusiva de dados, ela nos fornece a proximidade de um resultado com o seu valor de referência real.

Tomemos como exemplo um serviço de leitura automatizada de documentos. Se a solução é capaz de fornecer um grau de 90% de acurácia, isso quer dizer que as chances de que os dados extraídos sejam idênticos àqueles do documento real são de 90%.

Fonte: https://blog.idwall.co/o-que-e-acuracia/

Gabarito: **Errado**

**155** [**Q1754039**](https://www.qconcursos.com/questoes-de-concursos/questoes/b93aa945-cd) [Banco de Dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Julgue o próximo item, relativo à **qualidade de dados**.

O gerenciamento de qualidade de dados inclui a definição de padrões e métricas sobre os dados, porém dispensa o gerenciamento do ciclo de vida desses dados.

De um modo geral, a gestão da qualidade dos dados está focada na **avaliação de todos os conjuntos de dados** que chegam à organização e na aplicação de ações corretivas aos dados para assegurar que se enquadram para os fins para os quais foram originalmente destinados, para que sejam úteis e apropriados.

**Author:**Rúben Gonçalo Teixeira Pagaime

**Publish Year:**2019

O gerenciamento de qualidade de dados **NÃO** dispensa o gerenciamento do ciclo de vida desses dados, uma vez que o rejuvenescimento de dados pode estender a vida útil das informações históricas que anteriormente podiam ter sido deixadas no armazenamento.

Portanto, gabarito: ERRADO.

sacanagem tirar questão de artigo cientifico .... alguem tem que por limites nessa banca!

**156** [**Q1754036**](https://www.qconcursos.com/questoes-de-concursos/questoes/b92f2627-cd) [Administração de banco de dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/administracao-de-banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Acerca de ***dashboards***, julgue o item subsequente.

Relatórios do tipo *dashboard*visam apresentar uma visão integrada do sucesso numa organização, incluindo desempenhos financeiros e perspectivas sobre clientes, processos comerciais e aprendizado e crescimento.

O BSC (Balanced Scorecard), que pode ser traduzido como Indicadores Balanceados de Desempenho, pode ser definido como uma metodologia que possibilita **descrever e medir estratégias e processos de forma simples e clara** através de 4 perspectivas:

1. Perspectiva financeira;
2. Perspectiva dos clientes;
3. Perspectiva dos processos internos;
4. Perspectiva do aprendizado e crescimento.

**A dashboard tem uma aplicação de mensuração e monitoramento de resultados, enquanto o balanced scorecard tem um objetivo de gerenciamento de performance**.

**a estratégia é outro ponto em que estes conceitos divergem**. O dashboard, não precisa estar completamente alinhado com os objetivos, enquanto o balanced scorecard deve ter total sincronia com os objetivos e as estratégias da empresa.

Outra grande diferença entre ambos é em relação ao seu **escopo e ao seu propósito**. O dashboard, tem como objetivo dar uma visão mais ampla do que está acontecendo na empresa ou em um projeto, com um foco em questões operacionais no curto prazo.

Já o balanced scorecard visa planejar, executar e acompanhar uma estratégia, no longo prazo.

dashoard - monitora

BSC - verifica o desempenho

Gabarito: errado.

Um **dashboard** é uma ferramenta de gestão da informação que ajuda no acompanhamento e exibição de indicadores chave de performance (KPIs), métricas e dados que indicam a saúde de um negócio, setor ou processo. Eles são customizáveis para atender às demandas específicas de cada empresa e suas áreas.

Relatórios do tipo *dashboard*visam apresentar uma visão integrada **do sucesso numa organização**, incluindo desempenhos financeiros e perspectivas sobre clientes, processos comerciais e aprendizado e crescimento.

Acredito que o erro estar em limitar-se "do sucesso", pois nem sempre a empresa tem sucesso nos seus resultados

Questão um tanto complicada.

Relatórios do tipo *dashboard*visam apresentar uma visão integrada do sucesso numa organização, incluindo desempenhos financeiros e perspectivas sobre clientes, processos comerciais e aprendizado e crescimento.

Segue uma contribuição par análise segundo o que acabei de pesquisar.

Relatório não é a mesma coisa que Dashboard. Relatórios são gerados a partir de Dashboards.

**Por que a diferença entre dashboards e relatórios é uma questão importante?**

As primeiras ferramentas de BI não possuíam dashboards. Muito diferente de hoje em dia, em que, ao alcance de alguns cliques, você pode cruzar dados para ter (ainda mais) insights importantes. No início tudo se tratava de relatórios e esses relatórios exigiam uma pessoa com conhecimento técnico em programação SQL de banco de dados e mais conhecimento ainda para criar esses relatórios. Era um processo demorado do início ao fim e custava caro. Agora, as empresas e gestores têm acesso a ferramentas de BI que incluem dashboards e que podem gerar relatórios instantaneamente através de cliques.

**O que são relatórios?**

Os relatórios de dados são documentos que mostram um instantâneo das informações de uma questão específica, no mundo corporativo, podem ser vendedores específicos, setores específicos ou processo específico, por exemplo. Eles contêm tabelas de dados e geralmente alguns recursos visuais básicos, mas são essencialmente usados ​​para informar sobre um conjunto específico de dados.

**O que são dashboards?**

Por outro lado, um dashboard é uma ótima maneira de personalizar e ajustar a exibição dos dados escolhidos, como métricas específicas ou KPIs. Uma das maiores diferenças entre um relatório e um dashboards é que, embora um relatório seja normalmente focado em um tópico específico, um painel fornecerá uma visão geral, reunindo dados de vários tópicos e/ou fontes. Contrapondo aos relatórios, que tendem a ser estáticos, um dashboard é interativo e dinâmico.

**O que você precisa: dashboards ou relatórios?**

A resposta é: ambos. Depende de como você deseja usar os dados. Se você estiver fornecendo um conjunto de dados focado e específico para certa finalidade, um relatório pode funcionar.

Por outro lado, os dashboards são perfeitos para uma visão geral com indicadores-chave de desempenho com a qual os usuários podem interagir. Essencialmente, um relatório é um registro temporal que é particularmente útil ao apresentar informações e dados às partes interessadas, como CEOs ou um conselho de administração. Já um dashboard vai permitir que aqueles que estão trabalhando no momento vejam como as metas estão sendo cumpridas ou quando as métricas importantes mudam.

Fonte: <https://viasoft.com.br/blog/conheca-a-diferenca-entre-dashboards-e-relatorios-e-qual-deles-voce-precisa-27-09-2021/>

Não identifico o ero, dashboard pode estar estruturado por ester indicadores...

**157** [**Q1754031**](https://www.qconcursos.com/questoes-de-concursos/questoes/b91c7319-cd) [Banco de Dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Sobre **processamento de dados**, julgue o item que se segue.

Processamento em lote envolve entrada, processo e saída contínua de dados, em que a chave é o tempo de resposta.

O erro da questão é dizer que o processamento em LOTE tem saída continua, na verdade a saída é somente quando terminar de processar um determinado lote de tarefas. Não tem interação com o usuário.

*Processamento em****lote****envolve entrada, processo e saída contínua de dados, em que a chave é o tempo de resposta.*

A saída contínua de dados é uma característica do processamento em tempo real. Igualmente o tempo é outra característica do processamento em tempo real.

Logo, gabarito errado.

**158** [**Q1754030**](https://www.qconcursos.com/questoes-de-concursos/questoes/b9186a89-cd) [Banco de Dados ,](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/banco-de-dados)[Big Data](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/big-data) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Sobre **processamento de dados**, julgue o item que se segue.

MapReduce divide o conjunto de dados de entrada em blocos independentes que são processados pelas tarefas de mapa de uma maneira completamente paralela. Essa estrutura classifica as saídas dos mapas, as quais são, então, inseridas nas tarefas de redução.

**Gabarito: Certo**

* *O MapReduce e um modelo de programação que permite o processamento de dados massivos em um algoritmo****escalável****,****paralelo****e****distribuído****, geralmente utilizando um cluster de computadores.*

Para complementar: Mapreduce e OLAP são técnicas que processam grandes volumes de dados e permitem operar com grande quantidade de dados diferentes.

**COMPLEMENTANDO:**

MapReduce (Mapear e Reduzir):

» Processamento paralelo em ambiente distribuído

» Agrupamento por chave-valor

» Transforma dados maiores em dados menores

**MapReduce** divide o conjunto de dados de entrada em blocos **independentes** que são processados pelas tarefas de mapa de uma maneira completamente **paralela**. Essa estrutura**classifica as saídas dos mapas, as quais são, então, inseridas nas tarefas de redução.**

É um modelo de programação para processamento de dados de forma **paralela**. O modelo é simples, embora não seja tão simples escrever programas úteis o utilizando. Isto ocorre porque o MapReduce trabalha com duas primitivas de processamento de dados, *Mapper*e *Reducer*, o que torna a programação não muito trivial, além de ser diferente do que a maior parte dos desenvolvedores está acostumada. Mas o mais importante a se ressaltar é que os programas **MapReduce são inerentemente paralelos, o que o coloca como uma solução fantástica para problemas que envolvam análise de dados em larga escala**. Com isso, uma vez que a aplicação MapReduce está escrita, escalar a mesma para rodar em cima de centenas, milhares ou milhões de dados é apenas uma questão de mudança de configuração, ou seja, **independentes**.

No cerne do MapReduce estão duas funções: **Map** e **Reduce**

*"...divide o conjunto de dados de entrada em blocos independentes que são processados pelas tarefas de mapa de uma maneira completamente paralela"*

Os dados de entrada são primeiro divididos em blocos menores. Cada bloco é então atribuído a um mapeador para processamento.

Por exemplo, se um arquivo tiver 100 registros a serem processados, 100 mapeadores podem ser executados juntos para processar um registro cada. Ou talvez 50 mapeadores possam ser executados juntos para processar dois registros cada.

*"Essa estrutura classifica as saídas dos mapas, as quais são, então, inseridas nas tarefas de redução."*

Depois que todos os mapeadores concluem o processamento, a estrutura embaralha e classifica os resultados antes de passá-los para os redutores.

[1] https://www.talend.com/resources/what-is-mapreduce/ (traduzido).

**159** [**Q1754027**](https://www.qconcursos.com/questoes-de-concursos/questoes/b90ad688-cd) [Banco de Dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Acerca dos conceitos de **ingestão de dados**, julgue o item subsequente.

Quando a ingestão de dados ocorre em lotes, esses dados são movidos em intervalos programados, abordagem que é benéfica para os processos repetíveis.

**A ingestão de dados** pode ser realizada de diferentes maneiras, como em tempo real, lotes ou uma combinação de ambos (conhecida como arquitetura lambda), dependendo dos requisitos de negócios.

**Ingestão em tempo real:**A ingestão de dados em tempo real, também conhecida como **dados de streaming**, é útil quando os dados coletados são extremamente sensíveis ao tempo. Os dados são extraídos, processados ​​e armazenados assim que são gerados para a tomada de decisões em tempo real. Por exemplo, os dados adquiridos de uma rede elétrica devem ser supervisionados continuamente para garantir a disponibilidade de energia.

**Ingestão de lote:**Quando a ingestão ocorre em lotes, os dados são movidos em intervalos agendados de forma recorrente. Essa abordagem ***é benéfica para processos repetíveis***. **Por exemplo**, relatórios que precisam ser gerados todos os dias.

**Arquitetura Lambda:**A arquitetura lambda ***equilibra as vantagens dos dois métodos mencionados acima***, utilizando o processamento em lote para oferecer amplas visualizações dos dados em lote. Além disso, ele usa processamento em tempo real para fornecer visualizações de informações urgentes.

*@rotinaconcursos*

**160** [**Q1754025**](https://www.qconcursos.com/questoes-de-concursos/questoes/b9037666-cd) [Banco de Dados](https://www.qconcursos.com/questoes-de-concursos/disciplinas/tecnologia-da-informacao-banco-de-dados/banco-de-dados) **Prova:**[CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados](https://www.qconcursos.com/questoes-de-concursos/provas/cespe-cebraspe-2021-serpro-analista-especializacao-ciencia-de-dados)

Acerca dos conceitos de **ingestão de dados**, julgue o item subsequente.

O principal objetivo da ingestão de dados é organizar, armazenar e recuperar dados em tabelas relacionadas a entidades do mundo real.

O PRINCIPAL OBJETIVO DA INGESTÃO DE DADOS É:**absorver** dados de uma **variedade de fontes e transferi-los para um local de destino** onde podem ser **depositados e analisados**. De modo geral, os destinos podem ser u**m banco de dados, data warehouse, armazenamento de documentos, data mart, etc**. Por outro lado, existem várias opções de origem, como planilhas, extração de dados da web ou web scrapping, aplicativos internos e SaaS dados.

* QUESTÃO MUITO BOA, A INGESTÃO DE DADOS TAMBÉM PODE, APESAR DE NÃO SER O PRINCIPAL OBJETIVO:
* **ingestão de dados em tempo real**

A ingestão de dados em tempo real, também conhecida como dados de streaming, é útil quando os dados coletados são extremamente sensíveis ao tempo. Os dados são extraídos, processados ​​e armazenados assim que são gerados para a tomada de decisões em tempo real. Por exemplo, os dados adquiridos de uma rede elétrica devem ser supervisionados continuamente para garantir a disponibilidade de energia.

* **Ingestão de dados em lote**

Quando a ingestão ocorre em lotes, os dados são movidos em intervalos agendados de forma recorrente. Essa abordagem é benéfica para processos repetíveis. Por exemplo, relatórios que precisam ser gerados todos os dias.

* **Arquitetura Lambda**

A arquitetura lambda equilibra as vantagens dos dois métodos mencionados acima, utilizando o processamento em lote para oferecer amplas visualizações dos dados em lote. Além disso, ele usa processamento em tempo real para fornecer visualizações de informações urgentes.

O objetivo citado na questão tá errado. Inútil a pessoa ler toda a definição e não saber identificar o erro. Sendo que o rapaz primeira coisa que explica é o objetivo. É justamente o objetivo citado na questão que tá errado.

Há quatro etapas no ciclo de vida dos dados:

* **Ingerir**: a primeira etapa é extrair dados brutos, como de streaming de dispositivos, dados em lote no local, registros de aplicativos ou eventos e análises de usuários de aplicativos para dispositivos móveis.
* **Armazenar**: depois de recuperados, os dados precisam ser armazenados em um formato durável e facilmente acessível.
* **Processar e analisar**: nessa etapa, os dados são transformados de brutos em informações acionáveis.
* **Explorar e visualizar**: a etapa final é a de conversão dos resultados da análise em um formato que facilite a extração de informações e o compartilhamento com os colegas.

<https://cloud.google.com/architecture/data-lifecycle-cloud-platform>

Aplicação na questão:

Extrapolar o conceito de ingestão tornando a assertiva errada, estratégia típica do Cebraspe de análise de conhecimento do candidato.

Alguém conseguiu achar o erro?

"O principal objetivo da ingestão de dados é organizar, armazenar e recuperar dados em**tabelas relacionadas a entidades do mundo real.**"

Me parece estranho esse relacionamento entre tabelas e entidades do mundo real. Penso que as tabelas representam entidades do mundo real, conforme uma outra questão da CESPE:

"A modelagem de dados cria estruturas abstratas de dados **para representar entidades do mundo real**. Os quatro passos necessários para executar essa modelagem são:"