



APRENDIZAGEM EM FOCO

BANCO DE DADOS RELACIONAL E BIG DATA



APRESENTAÇÃO DA DISCIPLINA

Autoria: Cassio Rodolfo Aveiro da Silva

Leitura crítica: Flavio Fiuza

Tudo ao seu redor gira em torno da comunicação entre pessoas, entre sistemas e pessoas, e entre os próprios sistemas de informação. Seu computador pessoal, as listas de filmes em *stream*, os arquivos bancários, os check-ins de aeroportos ou de hospedagens e até mesmo os caixas de supermercados são alguns dos muitos exemplos em seu cotidiano que possuem em comum a presença de bancos de dados para armazenar essa grande variedade de arquivos e fornecer informações importantes quando necessário.

É interessante que você esteja por dentro das concepções, ferramentas e curiosidades por trás dos bancos de dados, uma vez que os dados são, hoje, o bem mais valioso de uma empresa. Com os avanços tecnológicos, muitos sistemas, técnicas e procedimentos no contexto da informação possibilitaram às pessoas, e sobretudo às empresas, mudarem suas estratégias profissionais e de negócio. Utilizar-se de recursos tecnológicos ultrapassados e desatualizados é sinônimo de perda de lucratividade.

Portanto, a disciplina *Banco de Dados Relacional e Big Data* será uma oportunidade para você adquirir conhecimento sobre atualizações, metodologias, ferramentas utilizadas, novos conceitos e a importância do armazenamento correto e adequado de dados, o que pode representar uma vantagem competitiva para empresas que, por exemplo, utilizam-se de um banco de dados relacional para otimizar suas operações de negócio.

Para tal, esta disciplina apresentará a você definições sobre estruturação e modelos de desenvolvimento de um banco de dados; relações entre entidades de um banco; princípios de cardinalidade e da linguagem-padrão utilizada. Além disso, será estudado o que é um sistema responsável pela gestão de bancos de dados. Você entenderá o que significa armazenamento em nuvem, tratamento e manipulação de dados, aplicações e quais as perspectivas futuras para os bancos de dados.

Por fim, para enriquecer ainda mais sua prática profissional, você estudará sobre Big Data, a tecnologia de informação presente no mundo todo e em crescente desenvolvimento, mostrando-se extremamente eficiente na organização, no armazenamento e na manipulação de enormes volumes de dados originados das mais variadas fontes, por exemplo, as redes sociais. Além, é claro, de aprender sobre as principais ferramentas utilizadas em Big Data para transformar com velocidade e veracidade meros dados brutos em informações de valor inestimável, como os *datas warehouses*, Hadoops e os chamados NoSQL (bancos de dados não relacionais).

INTRODUÇÃO

Olá, aluno (a)! A *Aprendizagem em Foco* visa destacar, de maneira direta e assertiva, os principais conceitos inerentes à temática abordada na disciplina. Além disso, também pretende provocar reflexões que estimulem a aplicação da teoria na prática profissional. Vem conosco!

TEMA 1

Banco de dados e Big Data: união necessária

Autoria: Clarissa Fernanda Correia Lima Loureiro

Leitura crítica: Flavio Fiuza





DIRETO AO PONTO

Os dados são o novo petróleo, o combustível mais valioso do mundo. As ferramentas utilizadas para organizar esses dados são os bancos de dados. Nesta disciplina, vamos aprender sobre um tipo de banco de dados bastante eficiente que é o banco de dados relacional.


Este tipo de banco é baseado no modelo relacional.

A representação dos dados se dá por tabelas, em que cada linha desta é um registro com o seu endereço que pode também ser chamado de ID ou pode também ter o nome de chave. Na tabela, suas colunas têm as características que são também chamadas de atributos dos dados. As relações entre os dados é que são a razão pela qual esse tipo de banco tem o seu nome, pois para cada atributo terá um valor registrado que se relaciona com os respectivos dados.

Por causa da relação entre as tabelas e, conseqüentemente, a relação entre os dados das tabelas, podemos ter um processamento de dados mais rápido e confiável. Quando o usuário procura uma informação, conseguirá chegar à solução de maneira mais eficiente e otimizada.

O modelo de banco de dados relacional tem a sua característica principal na utilização de tabelas para armazenar de forma eficiente e também acessar as informações procuradas. Foi assim criado um modelo mais eficiente de acesso aos dados.

Na tabela, as colunas têm as características que são também chamadas de atributos dos dados. As relações entre os dados são a razão pela qual esse tipo de banco tem o seu nome, pois, para cada atributo, terá um valor registrado que se relaciona com os respectivos dados. O usuário pode ter, por exemplo, duas tabelas,



sendo que existe uma coluna que é comum a ambas as tabelas. Desta forma, existe uma relação entre as tabelas em que o banco de dados relacional pode estabelecer uma ID que é, neste caso, uma chave.


O modelo de banco de dados relacional representa um banco de dados como uma coleção de tabelas, cada uma pode ser armazenada como um arquivo separado. A maior parte dos bancos de dados relacionais utiliza a linguagem de consulta de alto nível chamada SQL e admite uma forma limitada de visões do usuário.

Sobre Big Data, temos que a quase totalidade dos dados que existem hoje foi gerada nos últimos dois anos. Isto significa que a maior parte dos dados que temos hoje em dia foi gerada entre 2018 e 2020. A qualidade de câmeras, vídeos e capacidade de armazenamento de informações nos dispositivos, aliada aos meios de comunicação, como mídias sociais, e aos novos hábitos de compartilhamento digitais, fizeram com que cerca de 8 zetabytes (10^{21} bytes) fossem gerados em 2015. Em 2020, foram gerados cerca de 350 zetabytes de dados, ou seja, uma quantidade 43,75 vezes maior.

Essa nova realidade faz com que o Big Data esteja entre os assuntos mais importantes em termos de soluções globais. Analisar situações, gerenciar empresas e tomar decisões pode se tornar mais eficiente com uma análise em Big Data.

Para trabalhar com Big Data, o processamento utilizado é basicamente paralelo. É possível analisar dados não estruturados e é comum o uso de banco de dados NoSQL.

Algumas tecnologias dão suporte ao Big Data, como Hadoop e MapReduce, Cassandra, todos sistemas *open source*, muito utilizados pelas redes sociais na parte de análise e de



infraestrutura para armazenar e processar os dados. O Hadoop é um projeto *open source* desenvolvido pelos projetos Hadoop MapReduce (HMR), utilizado para processamento paralelo; Hadoop Distributed File System (HDFS), utilizados para trabalho com dados não estruturados.

A razão pela qual o processamento é muito acelerado é que os dados são pesquisados paralelamente. Os dados são distribuídos em blocos, que são divididos em diferentes servidores. Os blocos têm um volume maior do que o que normalmente é usado, são de 64 MB. A razão do tamanho grande é que cada servidor pode trabalhar paralelamente com um volume maior de dados.

A área visual de imagens e vídeos vai se desenvolver cada dia mais e a tendência é que a velocidade de compartilhamento e quantidade dos dados se multipliquem a cada ano. Por este motivo, torna-se tão importante aprender sobre a análise e o tratamento de um grande volume de dados como a matéria de banco de dados e Big Data. Um ponto fundamental é que os dados se transformem em informações úteis e inteligíveis, das quais os usuários consigam extrair um importante valor.

Já falamos sobre algumas tecnologias que aprofundaremos ao longo da disciplina, mas é importante falar também dos profissionais e da importante atuação e do conhecimento necessário do assunto da disciplina, assim como novas profissões atuais e do futuro. O novo cargo chamado cientista e analista de dados é exemplo de profissionais do presente e futuro. Eles estão entre os profissionais mais desejados pelas empresas nacionais e internacionais. Precisam ter conhecimentos de estatística, matemática, familiaridade com linguagem de programação como NoSQL, Python, R, entre outras, e saber lidar com o modelo relacional, suas regras e técnicas.

Referências bibliográficas


NAVATHE, E. **Sistemas de banco de dados**. 6. ed. São Paulo: Pearson, 2011.
TAURION, C. **Big Data**. Rio de Janeiro, Brasport, 2013.

PARA SABER MAIS

Para entender melhor sobre um banco de dados relacional, precisamos entender de forma leve e simples o que é um banco de dados. Podemos fazer analogias com o nosso dia a dia. Então, se pensarmos em dados como se fossem atributos, ou itens, como se fossem objetos, esses precisam ser organizados com uma certa estrutura, ser registrados, por exemplo, para saber o que se tem quando se deseja usar. Desta maneira, um conjunto de registros desses itens, nós chamamos arquivos, e eles precisam ser acessados por programas. Assim, denominados o conjunto de arquivos como banco de dados.

Esse grupo de arquivos, que na verdade estão armazenando dados para serem acessados de forma mais rápida e eficiente, pode estar relacionado entre si. Surge, então, alguns conceitos sobre os tipos de acesso e de organização dos bancos de dados. Todos gostam de ter suas informações de forma rápida e correta, para isso acontecer, temos uma certa lógica para organizar e estruturar os bancos de dados. Além disso, temos também um conjunto de regras e princípios para que o tempo de aquisição das informações estruturadas seja curto, e o processo, confiável.

Os princípios são: redundância, inconsistência e integração. Hoje em dia, sabemos que guardar itens, ou dados, não é uma tarefa fácil quando temos um volume muito grande deles, certo? Então, quando falamos em redundância, estamos tratando de ter os mesmos dados em áreas diferentes de uma mesma empresa.



Portanto, se tivermos os mesmos dados armazenados em áreas diferentes, a empresa incorrerá em maiores custos e queremos evitar isso. É por isso que falamos que temos de ter cuidado com o princípio da redundância.

Sobre a inconsistência, temos exemplos de dados que podem ter alterações ao longo do tempo, então, normalmente, tais dados precisam de uma atualização para que não ocorra uma inconsistência. Então, quando temos importantes decisões a serem tomadas em grandes empresas, podemos ter grandes erros caso ocorra a inconsistência nos dados.

Em relação ao princípio da integração, podemos entender com exemplos de compartilhamento de dados. Temos um exemplo bastante utilizado hoje em dia que são as redes sociais. Normalmente, compartilhamos muitas imagens, vídeos e informações por meio do WhatsApp, Facebook, Instagram e outras redes ou até e-mails, então, às vezes, facilita se tivermos uma integração na forma como compartilhamos os dados. Esse é o princípio da integração, em que, para evitar erros, é necessário ter uma comunicação eficiente dos dados e a necessidade de uma integração de dados com qualidade e confiabilidade.

Podemos fazer várias analogias com os princípios, regras e linguagens utilizadas para coletar, adquirir, tratar, analisar os dados e extrair importantes informações que serão utilizadas em tomadas de decisão em empresas, e até mesmo ter subsídios para ter uma boa política pública, baseadas em dados das cidades.

Referências bibliográficas

MEDEIROS, F. L. **Banco de dados:** princípios e práticas. Curitiba: Intersaberes, 2013.



TEORIA EM PRÁTICA

Refleta sobre a seguinte situação: você é o diretor da empresa Google, existem muitas áreas e vários programas que estão sob a mesma empresa, como os e-mails Gmail, sites de busca, agendas virtuais, Google Maps, vários aplicativos; a empresa multinacional, que atua mundialmente, possui um volume de dados gigantesco, na faixa de zetabytes. Com a maior parcela de seu lucro vindo da publicidade, é uma das maiores empresas do mundo, embora seja recente. Foi fundada em 4 de setembro de 1998. O crescimento da empresa foi muito rápido e, com isso, culminou em diversos produtos inovadores e tecnológicos no mercado. O Google é executado em mais de um milhão de servidores e processa mais de 5 bilhões de solicitações de pesquisa em vários países diferentes. Já foi tido como o site mais visitado do mundo e considerado uma das melhores empresas para se trabalhar. Com todas essas informações sobre o Google, é possível perceber, com o conteúdo estudado, que estamos tratando do maior volume de dados do mundo. Quais as características de profissionais que você contrataria para sua empresa? Quais os princípios aplicados aos dados que você recomendaria ter para que sua empresa continuasse a ter sucesso?

Para conhecer a resolução comentada proposta pelo professor, acesse a videoaula deste *Teoria em Prática* no ambiente de aprendizagem.



LEITURA FUNDAMENTAL

Indicações de leitura

Indicação 1

O capítulo tem como objetivo compreender conceitos básicos de bancos de dados, seus esquemas e instâncias. Dentro do conteúdo, é discutido também sobre a arquitetura e alguns tipos de linguagens de programação necessárias para trabalhar com bancos. Algumas características do modelo relacional e de banco de dados relacional são explicadas.

Para realizar a leitura, acesse a Biblioteca Virtual da Kroton e busque pelo título da obra.

NAVATHE, E. **Sistemas de banco de dados**. 6. ed. São Paulo: Pearson, 2011. Cap. 2, p. 19- 36.

Indicação 2

O crescimento do volume de dados é muito grande e a velocidade de compartilhamento de informações, em forma de imagens, áudio e vídeo, vem crescendo muito rápido. Os dados são importantes na economia, no trabalho. A análise de Big Data vem trazer novas soluções e profissões no mundo atual e futuro.

Para realizar a leitura, acesse a Biblioteca Virtual da Kroton e busque pelo título da obra.

TAURION, C. **Big Data**. Rio de Janeiro, Brasport, 2013. Cap. 2, p. 37-46.

Indicação 3

O capítulo destaca a importância da programação nos dias de hoje, seu papel nas empresas e os conceitos iniciais sobre algoritmos, quais as formas de sua representação e como realizar a entrada, o processamento e a saída de dados.

Para realizar a leitura, acesse a Biblioteca Virtual da Kroton e busque pelo título da obra.

LEAL, L. C. G. **Linguagem, programação e banco de dados**. Curitiba: Intersaberes, 2015. Cap. 1, p. 15-48.

Indicação 4

Este capítulo mostra a importância dos bancos de dados, a sua representação eficiente, que possibilita acesso a informações corretas, em tempo hábil. São explicados princípios que devem ser levados em consideração para obtenção de um banco de dados eficiente, como redundância, inconsistência e integração.

Para realizar a leitura, acesse a Biblioteca Virtual da Kroton e busque pelo título da obra.

MEDEIROS, F. L. **Banco de dados: princípios e práticas**. Curitiba: Intersaberes, 2013. Cap. 1, p. 11-32.



QUIZ

Prezado aluno, as questões do Quiz têm como propósito a verificação de leitura dos itens *Direto ao Ponto, Para Saber Mais, Teoria em Prática e Leitura Fundamental*, presentes neste *Aprendizagem em Foco*.

Para as avaliações virtuais e presenciais, as questões serão elaboradas a partir de todos os itens do *Aprendizagem em Foco* e dos slides usados para a gravação das videoaulas, além de questões de interpretação com embasamento no cabeçalho da questão.

1. Quais os princípios importantes na área de banco de dados que foram estudados?
 - a. Redundância, inconsistência e integração.
 - b. Reconstrução, consistência e integração.
 - c. Reconstrução, inconsistência e união.
 - d. Reordenação, consistência e integração.
 - e. Reconstrução, consistência e integração.

2. Assinale a alternativa correta que possui algumas tecnologias que dão suporte ao Big Data:
 - a. Hadoop e MapReduce, Lisandra.
 - b. Hadoop e MapRoad, Cassandra.
 - c. Hapop e MapReduce, Cassandra.
 - d. Hadoop e MapReduce, Cassandra.
 - e. Hadoop e NoteReduce, Cassandra.



GABARITO

Questão 1 - Resposta A

Resolução: Os princípios são: redundância, inconsistência e integração, que são um conjunto de regras e princípios para que o tempo de aquisição das informações estruturadas seja curto, e o processo, confiável.

Questão 2 - Resposta D

Resolução: Algumas tecnologias dão suporte ao Big Data, como Hadoop e MapReduce, Cassandra (sistema *open source*), muito utilizados pelas redes sociais, na parte de análise e de infraestrutura, para armazenar e processar os dados. O Hadoop é um projeto *open source*, desenvolvido pelos projetos Hadoop MapReduce (HMR), utilizado para processamento paralelo; Hadoop Distributed File System (HDFS), utilizados para trabalho com dados não estruturados.

TEMA 2

ACID e benefícios do banco de dados relacional

Autoria: Clarissa Fernanda Correia Lima Loureiro

Leitura crítica: Flavio Fiuza





DIRETO AO PONTO

Um dos conceitos básicos que são sempre usados quando se fala de banco de dados é a transação. Ela é usada para uma série de atividades no banco de dados, tais como: atualizações, inserções, exclusões e recuperações dos dados. Estas transações podem ser executadas por aplicativos que estão em vários tipos de dispositivos, como o smartphone, por exemplo.

Quando as transações são usadas para recuperar dados, são chamadas de transação leitura, se for usada para inserir, excluir e atualizar dados, é chamada de transação de leitura e gravação. Alguns comandos para esta operação são usuais, como *Begin Transaction* e *End Transaction*, o que temos entre esses comandos será considerado transação.

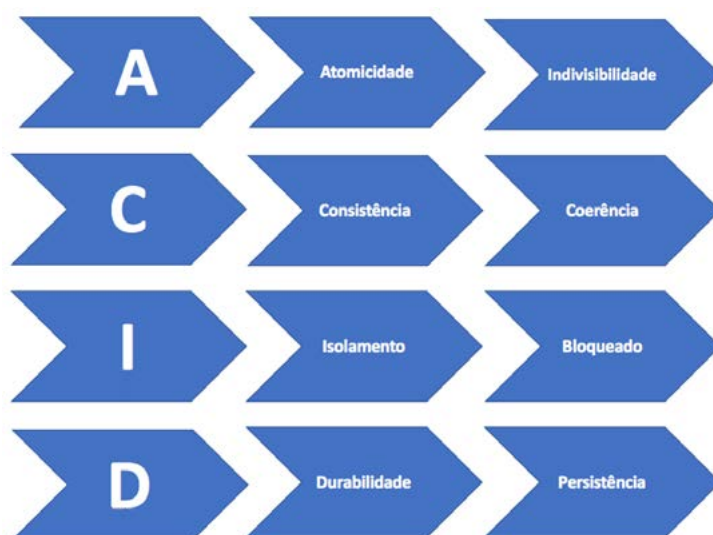
Para evitar que sejam interrompidas as transações em bancos de dados, é necessário implementar técnicas de controle de concorrência para manter o banco em um bom estado, assim como ter propriedades desejáveis e necessárias para as transações. Tais técnicas são chamadas de ACID:

- **Atomicidade:** uma transação deve ser realizada em sua totalidade, caso não seja, não deve ser realizada.
- **Consistência:** uma transação deve ser realizada do começo ao fim, preservar sua consistência, sem interferência de outras transações.
- **Isolamento:** uma transação deve ser executada isoladamente, ainda que ao mesmo tempo que outras. Mesmo acontecendo simultaneamente, não deve sofrer interferência de outras transações.

- **Durabilidade:** quando houver mudanças aplicadas pelas transações ao banco de dados, estas não podem ser perdidas por causa de falhas, as mudanças devem ser mantidas.

Quando a transação é bem executada e finalizada, ela é confirmada (*committed*), quando existe falha, pode-se dizer que a transação é abortada (*abort*). Para que sejam evitadas as falhas, o banco de dados relacional deve ter e impor as propriedades ACID nas transações para que o controle e a recuperação após as falhas sejam garantidos, e o controle de concorrência, realizado.

Figura 1 – Propriedades ACID



Fonte: elaborada pela autora.

Um conceito também importante no banco de dados é a cardinalidade, que pode ser entendida como o grau de relacionamento das ocorrências de determinadas entidades. Para entender melhor esse conceito, precisamos entender também os tipos de relacionamentos entre as entidades e suas ocorrências:

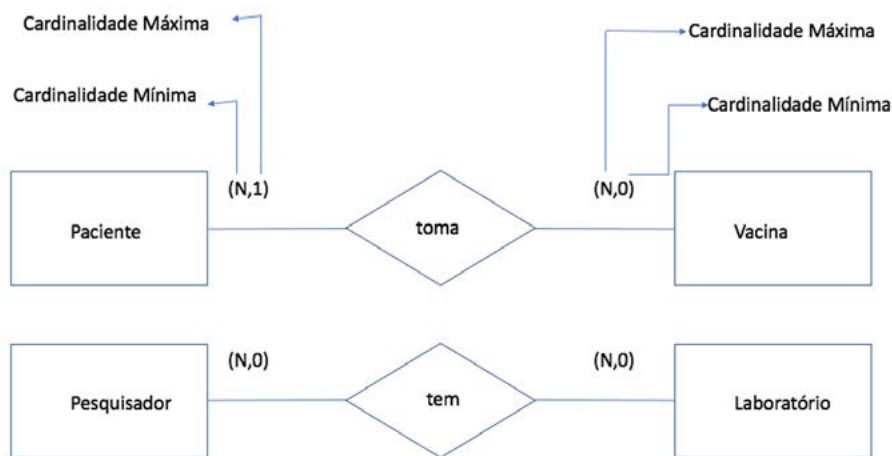
- Relacionamento de um para um (1x1).
- Relacionamento de um para muitos ou muitos para um (1xN) ou (Nx1).
- Relacionamento de muitos para muitos (NxN).

Em relação à cardinalidade, existem dois tipos:

- **Cardinalidade máxima:** refere-se ao número máximo de vezes em que pode haver ocorrência de uma entidade A em B.
- **Cardinalidade mínima:** refere-se ao número mínimo de vezes que pode haver ocorrência de uma entidade A em uma entidade B. Os valores possíveis são zero ou um.

Nos relacionamentos em que temos a possibilidade de ocorrência N vezes, temos então uma cardinalidade máxima para esses casos.

Figura 2 – Exemplo de cardinalidade e relacionamento com entidades e número de ocorrências



Fonte: elaborada pela autora.


Referências bibliográficas

HARVARD-MIT. **MIT-BIH Database Distribution**. 2005. Disponível em: <https://ecg.mit.edu/>. Acesso em: 16 nov. 2020.



PARA SABER MAIS

É importante entendermos bem as vantagens, os benefícios e importância de um banco de dados relacional confiável. Podemos considerar um investimento que as empresas fazem



que pode ter um impacto positivo entre os funcionários e na produtividade, sendo estes alguns dos bons resultados que a empresa pode alcançar:

- **Comunicação entre os setores das empresas:** relacionamento entre as diversas áreas e dados da empresa. Um bom banco de dados pode manter as várias áreas diferentes da empresa se comunicando de forma eficiente, isso melhora a produtividade da empresa.
- **Reduzir os riscos de operação:** a redução de riscos ocorre porque a transparência das informações é maior, dado que as diferentes equipes terão acesso confiável aos dados. Isso faz com que as equipes estejam alinhadas com as necessidades do empreendimento.
- **Aumentar a segurança:** um sistema de gestão de informações, precisa ser seguro. A confiança na aquisição e armazenamento de dados para pessoas que precisam ter acesso, é fundamental. Como solução, é possível limitar e controlar as pessoas com acesso e assim ter uma maior segurança dos dados.
- **Melhorar a tomada de decisão:** basear em informações valiosas de forma eficiente é a melhor forma de diminuir os riscos e tomar decisões de forma coerente com as características da empresa. É mais provável ter uma visão geral do todo e identificar os principais problemas e melhorias a serem alcançadas no negócio.

Referências bibliográficas

LEAL, L. C. G. **Linguagem, programação e banco de dados**. Curitiba: Intersaberes, 2015. Cap. 1, p. 15-48.



TEORIA EM PRÁTICA

Refleta sobre a seguinte situação: uma grande empresa de produção e venda de produtos eletrônicos, atualmente, realiza o controle de seus dados de maneira totalmente manual, ou seja, por meio de anotações referentes às atividades desenvolvidas. Ao ter um aumento significativo de seus processos de negócio, a empresa se vê na obrigação de criar um banco de dados relacional (BDR). Sendo responsável pelo desenvolvimento deste BDR, quais as relações e cardinalidades que você poderia estabelecer nas relações entre fornecedores, clientes, produtos e vendas?

Para conhecer a resolução comentada proposta pelo professor, acesse a videoaula deste *Teoria em Prática* no ambiente de aprendizagem.

LEITURA FUNDAMENTAL

Indicações de leitura

Indicação 1

Ler o capítulo introdutório do livro *Banco de Dados: princípios e práticas*, publicado em 2013 pela editora Intersaberes.

Para realizar a leitura, acesse a plataforma Biblioteca Virtual 3.0_Pearson, na Biblioteca Virtual da Kroton, e busque pelo título da obra.

MEDEIROS, F. L. **Banco de dados:** princípios e práticas. Curitiba: Intersaberes, 2013.

Indicação 2

Ler o capítulo introdutório do livro *Sistemas de banco de dados*, publicado em 2011 pela editora Pearson.

Para realizar a leitura, acesse a plataforma Biblioteca Virtual 3.0_Pearson, na Biblioteca Virtual da Kroton, e busque pelo título da obra.

NAVATHE, E. **Sistemas de banco de dados**. 6. ed. São Paulo: Pearson, 2011.

QUIZ

Prezado aluno, as questões do Quiz têm como propósito a verificação de leitura dos itens *Direto ao Ponto, Para Saber Mais, Teoria em Prática e Leitura Fundamental*, presentes neste *Aprendizagem em Foco*.

Para as avaliações virtuais e presenciais, as questões serão elaboradas a partir de todos os itens do *Aprendizagem em Foco* e dos slides usados para a gravação das videoaulas, além de questões de interpretação com embasamento no cabeçalho da questão.

1. É o conceito que dá a ideia de indivisibilidade, ou seja, as transações que acontecem com partes de informações discretas devem ser totalmente executadas, caso contrário, se forem divididas, não serão executadas. Então este conceito, assim como o átomo, garante a indivisibilidade de suas partes.

Estamos falando do conceito de:

- a. Atomicidade.
- b. Consistência.
- c. Isolamento.
- d. Durabilidade.
- e. Coerência.

2. Este conceito é como se fosse uma proteção, quando uma transação não foi validada, mas está em andamento, o conceito garante que deve permanecer separada, ou seja, não haverá interferência por outra ação que acontece concomitantemente.

Estamos falando do conceito de:

- a. Atomicidade.
- b. Consistência.
- c. Isolamento.
- d. Durabilidade.
- e. Coerência.



GABARITO

Questão 1 - Resposta A

Resolução: Atomicidade: uma transação deve ser realizada em sua totalidade, caso não seja, não deve ser realizada.

Isolamento ou isolação: uma transação deve ser executada isoladamente, ainda que ao mesmo tempo que outras.

Mesmo acontecendo simultaneamente, não deve sofrer interferência de outras transações.

Questão 2 - Resposta C

Resolução: O isolamento garante que transações concorrentes não sofram mútua interferência. Noutras palavras, uma transação em andamento que ainda não esteja validada deve permanecer isolada de qualquer outra operação, garantindo que a transação não será interferida por nenhuma outra.

TEMA 3

A SQL e os sistemas de gestão de banco de dados

Autoria: Clarissa Fernanda Correia Lima Loureiro

Leitura crítica: Flavio Fiuza

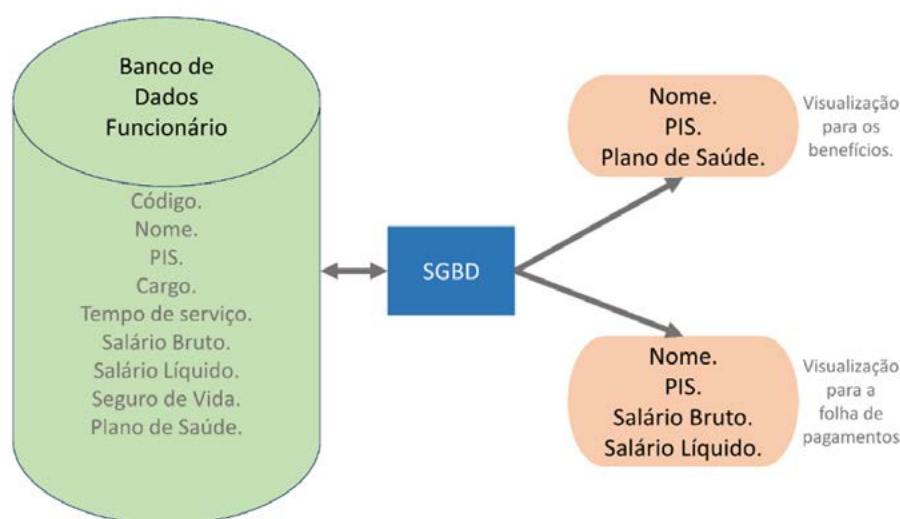


► DIRETO AO PONTO

Para que o uso de um banco de dados (BD) relacional e suas transações de dados necessárias faça sentido, é preciso que você tenha em mãos um adequado sistema de gestão de banco de dados (*database management system* – DBMS), responsável por criar, armazenar, organizar e promover acesso aos dados de um determinado BD (LAUDON; LAUDON, 2014).


Para Machado (2020), atualmente, os DBMS possuem uma estruturação em módulos com a função básica de gerar variadas visualizações lógicas a partir de um único banco de dados. Tais visualizações são especificadas de acordo com os interesses dos usuários que as utilizarão, ou seja, são dados ou informações selecionados e disponibilizados para uma finalidade. Um exemplo típico é a aplicação do DBMS no BD de recursos humanos (Figura 1):

Figura 1 – Ação do DBMS (SGBD) no banco de dados de recursos humanos



Fonte: adaptada de Laudon e Laudon (2014).

Os DBMS são constituídos por dois módulos principais: módulo cliente e módulo servidor. O primeiro, por meio de interfaces intuitivas, promove a interação com os usuários sem exigir deles




uma linguagem complexa. Já o módulo servidor está relacionado às operações técnicas relacionadas ao armazenamento de dados de modo geral, como acessos e consultas (MACHADO, 2020).

Em relação às interfaces dos DBMS, Laudon e Laudon (2014) e Machado (2020) trazem seis tipos de interfaces mais disponibilizadas ao usuário:

- Interface baseada em menus (MBI – Menu Based Interface).
- Interface baseada em formulários (FBI – Forms Based Interface).
- Interface gráfica para usuários (GUI).
- Interface de linguagem natural (NLI – Natural Language Interface).
- Interface para usuários parametrizáveis (PUI – Parameterizable User Interface).
- Interface para os administradores dos bancos de dados (DBAI – Data Base Administrator Interface).

Os DBMS, segundo Laudon e Laudon (2014), são excelentes para a gestão de banco de dados relacionais (BDR), pois dispõem aos administradores três funções essenciais aos BDRs: *select* – seleção de um conjunto de tuplas importantes de uma entidade; *join* – combinação de tabelas para obtenção de mais informações; e *project* – conjunto de atributos importantes para a geração de uma nova tabela apenas com as informações pertinentes.

Destarte, para compreender melhor o funcionamento de um DBMS, é necessário também entender a linguagem computacional por trás deste sistema. O sucesso do BDR ocorreu graças à utilização da Structure Query Language (SQL – Linguagem Estruturada de Consulta), desenvolvida em



1970 e, em virtude de sua expansão em utilização mundial, tornou-se a linguagem-padrão dos DBMS pela ANSI em 1982 e pela ISO em 1986 (MACHADO, 2020).

Para Machado (2020), a SQL, que pode ser utilizada por meio de linguagens hospedeiras (C, C++, Fortran etc.), isenta usuários e programadores da preocupação relacionada à migração de dados entre diferentes DBMS. Além disso, a linguagem estruturada em consulta age como um mediador na troca de dados e informação entre o banco de dados e processos (consultas interativas; acessibilidade a um banco de dados específico; otimização da gestão dos DBAs; relacionamento cliente-servidor; comunicação entre computadores e máquinas; e a interação com outros bancos de dados).

Laudon e Laudon (2014) trazem a observação de que, por meio dos comandos da SQL, é possível promover um compartilhamento de dados entre usuários concorrentes sem interferir na atividade do outro, muito menos fornecer dados sigilosos por engano. Para o BDR, essa linguagem permite otimizar a integralidade dos dados, um dos princípios desse tipo de banco.

A SQL pode, por exemplo, ser utilizada em um DBMS apenas para definição lógica dos dados exigidos pelo usuário. Neste caso, utiliza-se de sua sublinguagem Data Definition Language (DDL – linguagem de definição de dados) juntamente a um compilador DDL para leitura e escrita de comandos (MACHADO, 2020).

Outra utilização muito importante é na manipulação dos dados de um BD, principalmente se estiverem já compilados, exigindo alterações, recuperações, exclusões ou mesmo realocações. Para essa situação, a SQL faz uso de sua outra sublinguagem, a Data Manipulation Language (DML – linguagem de manipulação de dados) segundo Laudon e Laudon (2014).

Indo mais a fundo, a DML pode ainda ser subdividida em linguagem de alto e baixo nível. A primeira, também conhecida como não processual, tem a capacidade de entender a consulta solicitada pelo usuário por meio de uma linguagem de programação geral. Já a de baixo nível, ou processual, exige ser embutida em uma linguagem hospedeira por meio de códigos do tipo *loop*, para que consiga realizar as tarefas solicitadas, porque tem a capacidade de processar apenas um dado de cada vez (MACHADO, 2020).

Os mais conhecidos DBMS baseados em SQL são o MySQL, um dos softwares de gestão de banco de dados mais populares de código aberto, e o pacote comercial Microsoft® Access® (LAUDON; LAUDON, 2014).

Referências bibliográficas

LAUDON, K.; LAUDON, J. **Sistemas de informações gerenciais**. Tradução Célia Taniwaki. 11. ed. São Paulo: Pearson, 2014.

MACHADO, F. N. R. **Projeto e implementação de banco de dados**. 4. ed. São Paulo: Érica, 2020.

PARA SABER MAIS

Ao estudar sobre sistemas de gestão de banco de dados, é muito importante que você conheça seus principais componentes. Para começar, o armazenamento, como você já deve imaginar, é realizado em discos rígidos normalmente. O controle de acesso a esse armazenamento físico é gerenciado por um sistema operacional (SO), por meio de uma ferramenta de gerenciamento de dados armazenados, conforme apresentado por Machado (2020).

O gerenciador utiliza-se do SO para promover as transações de dados entre o disco rígido e a memória do PC, além da manipulação dos *buffers* da memória em questão para que

todos os módulos do DBMS possam ter acesso ao banco de dados. Existem sistemas que contam com um módulo próprio para gerenciamento de *buffers*, dispensando o uso desnecessário do SO (ELMASRI; NAVATHE, 2009).

O catálogo de dados também é um componente do DBMS, o qual armazena informações e detalhes referentes aos dados de um banco por meio da DDL. O catálogo serve de suporte para outros componentes e módulos do sistema, segundo Elmasri e Navathe (2009).

Já o processador *runtime* é o componente responsável por controlar o acesso ao banco de dados em tempo de execução de comandos, recuperando-os ou atualizando-os. Esse acesso passa pelos gerenciadores de armazenamento e de *buffers*. Esse componente está associado a outro destinado a manipular as pesquisas de alto nível. Trata-se do compilador de consulta, o qual analisa os comandos e linguagens utilizados na consulta requerida e os compila para gerar um código de acesso ao banco de dados e solicitações ao processador *runtime*. Existe também um pré-compilador com a função de extrair os comandos DML da linguagem hospedeira (ELMASRI; NAVATHE, 2009).

Nos DBMS, é prática comum disponibilizarem um programa do cliente para o acesso ao banco de dados por um computador (computador cliente) que não está fisicamente ligado ao referido banco, como no caso dos servidores de banco de dados. Há ainda o servidor de aplicação, um computador intermediário na relação cliente-servidor, segundo Elmasri e Navathe (2009).

Além dos componentes e módulos, os DBMS disponibilizam em sua estrutura utilitários de banco de dados. Segundo Machado (2020), estes utilitários auxiliam os DBAs em suas tarefas relacionadas ao banco de dados. Para Elmasri e Navathe (2009), as funções básicas desses recursos são:

- **Loading:** carregamento de arquivos de dados. É necessário fornecer ao utilitário a fonte e a estrutura do dado que se deseja, formatando ou convertendo os arquivos e os disponibilizando no BD.
- **Backup:** criação de cópia de todos os arquivos do BD como medida de segurança contra falhas inesperadas. Podem ser realizados backups pontuais em apenas parte do banco que sofre recente modificação.
- **File reorganization:** reorganização de arquivos no BD visando melhor performance do DBMS (semelhante ao desfragmentador de disco de um SO).
- **Performance monitoring:** monitoramento do uso do banco de dados para geração de relatórios estatísticos que servirão de apoio à tomada de decisões pelos DBAs.

Existem muitos outros utilitários além desses básicos, que são disponibilizados conforme o grau de necessidade dos usuários dos DBMS.

Referências bibliográficas

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. Tradução Marília G. Pinheiro. 4. ed. São Paulo: Pearson, 2009.

MACHADO, F. N. R. **Projeto e implementação de banco de dados**. 4. ed. São Paulo: Érica, 2020.

TEORIA EM PRÁTICA

Sua empresa, do ramo de venda de produtos importados, possui uma alta demanda de pedidos, o que exige um relacionamento estrito com fornecedores para garantir o atendimento de seus clientes. Atualmente, a empresa conta com sistemas de gestão

da cadeia de suprimento e de relacionamento com o cliente, sistemas de informação que já seriam suficientes para garantir o sucesso dos processos de negócios. Porém, o banco de dados que alimenta tais sistemas não possui uma gestão adequada, causando atrasos e informações desconexas. Você recebe a missão de resolver esse problema. Como você poderia utilizar um sistema de gestão de banco de dados estruturado em SQL nesta situação? E quais utilitários e interfaces são recomendados para esse sistema?

Para conhecer a resolução comentada proposta pelo professor, acesse a videoaula deste *Teoria em Prática* no ambiente de aprendizagem.



LEITURA FUNDAMENTAL

Indicações de leitura

Indicação 1

Esta dissertação de mestrado teve como objetivo apresentar a integração possível das funcionalidades do DBMS SQL Server R Services 2016, uma ação que permite potencializar a performance desse sistema na gestão de banco de dados.

Para realizar a leitura, acesse a Biblioteca Virtual da Kroton e busque pelo título da obra.

FREITAS, C. M. A. **Integração das funcionalidades do programa SQL Server R Services 2016 na empresa Quidgest**. Dissertação (Mestrado em Métodos Quantitativos para a Decisão Econômica e Empresarial) – Instituto Superior de Economia e Gestão, Universidade de Lisboa, Lisboa, 2017.

Indicação 2

Este trabalho de conclusão de curso tem como objetivo apresentar práticas para limitar ou impedir ataques de hacker em sistemas de informação, por meio da inserção de SQL em aplicações web, uma aplicação desta linguagem em DBMS para restringir acessos não autorizados no banco de dados relacional.

Para realizar a leitura, acesse a Biblioteca Virtual da Kroton e busque pelo título da obra.

FREITAS, C. M. A. **Integração das funcionalidades do programa SQL Server R Services 2016 na empresa Quidgest**. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre: 2009.

QUIZ

Prezado aluno, as questões do Quiz têm como propósito a verificação de leitura dos itens *Direto ao Ponto, Para Saber Mais, Teoria em Prática e Leitura Fundamental*, presentes neste *Aprendizagem em Foco*.

Para as avaliações virtuais e presenciais, as questões serão elaboradas a partir de todos os itens do *Aprendizagem em Foco* e dos slides usados para a gravação das videoaulas, além de questões de interpretação com embasamento no cabeçalho da questão.

1. Em relação aos dois módulos principais do sistema de gestão de banco de dados, qual das alternativas apresenta corretamente tais módulos?

- a. Módulo automático e módulo manual.
 - b. Módulo empresarial e módulo pessoal.
 - c. Módulo integrado e módulo dividido.
 - d. Módulo cliente e módulo servidor.
 - e. Módulo de escrita e módulo de leitura.
2. Em relação ao Structure Query Language (SQL) e suas estruturas de linguagem empregadas na gestão de um banco de dados, qual alternativa contém as duas principais sublinguagens da SQL?
- a. IML e BDR.
 - b. DLL e DML.
 - c. BDR e DDL.
 - d. DML e DVL.
 - e. DDL e DML.



GABARITO

Questão 1 - Resposta D

Resolução: Os sistemas de gestão de banco de dados são baseados na relação cliente-servidor. Consequentemente, os dois principais módulos são o módulo cliente (com interfaces relacionais com o usuário) e o módulo servidor (para a realização das operações físicas do banco de dados).

Questão 2 - Resposta E

Resolução: A Data Definition Language (DDL – Linguagem de Definição de Dados) e a Data Manipulation Language (DML – Linguagem de Manipulação de Dados) são as principais sublinguagens da SQL utilizadas para consultas e/ou manipulação de arquivos do banco de dados, respectivamente.

TEMA 4

Big Data e bancos de dados não relacionais

Autoria: Clarissa Fernanda Correia Lima Loureiro

Leitura crítica: Flavio Fiuza





DIRETO AO PONTO

Com os avanços tecnológicos, o fluxo de dados e informações cresceu de forma exponencial, o que impossibilitou a utilização eficiente dos tradicionais bancos de dados relacionais e sistemas de informação. Com o advento da *web*, surgiram enormes volumes de dados gerados em curtos períodos de tempo. Com isso, o uso de planilhas, por exemplo, tornou-se demasiadamente obsoleto. Como alternativa eficiente para o gerenciamento desses imensos volumes, utiliza-se a concepção de Big Data.

Big Data excede as tradicionais atividades de coletar, armazenar e analisar dados. Não necessita de grandes servidores, softwares ou hardwares potentes, pois seu uso pode ser desenvolvido com armazenamento em nuvem. Benefícios convertidos em redução de custos e visualização de novas oportunidades empresariais (LAUDON; LAUDON, 2014).

Taurion (2013) descreve Big Data em cinco “Vs”, palavras-chave que definem a concepção:

1. Volume.
2. Velocidade.
3. Veracidade.
4. Variedade.
5. Valor.

Segundo Taurion (2013), Big Data é um conjunto de tecnologias, práticas e processos voltados à análise de dados que anteriormente não eram nem identificados, o que possibilita uma gestão muito mais eficiente e de alto desempenho.

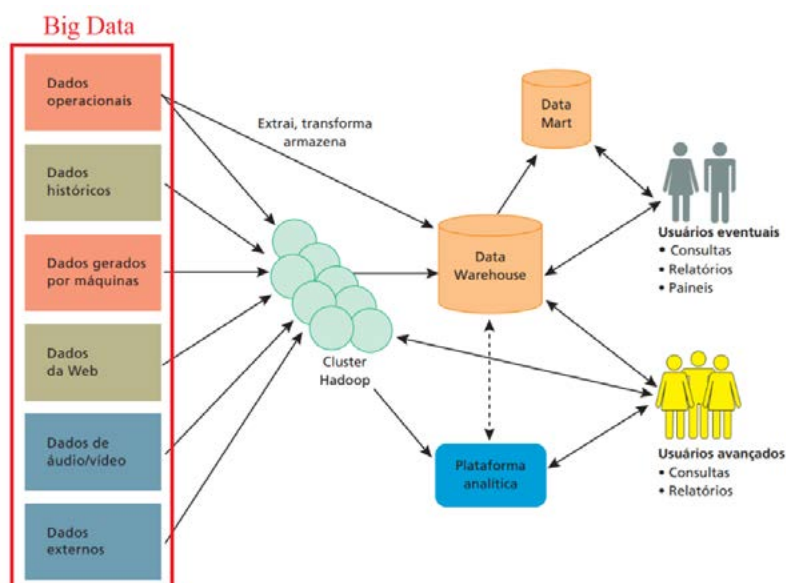
Big Data exige mudanças organizacionais no que tange a processos e infraestrutura de tecnologia de informação. Assim sendo, sua

utilização deve, obrigatoriamente, passar por três etapas básicas (LAUDON; LAUDON, 2014). Primeiramente, realiza-se o processo de coleta de dados das mais variadas fontes possíveis. Esses dados devem, então, ser tratados, agregados e integrados entre si para que, ao final, possam ser utilizados em ações analíticas que descobrirão seus padrões e relações, segundo Taurion (2013).


A utilização de Big Data é tão surpreendente que traz impactos significantes em empresas, governos e sociedade como um todo. Segundo Taurion (2013), o acesso facilitado a um grande volume de dados antes inacessíveis permite uma maior transparência de informações; uma segmentação profunda da sociedade, chegando ao nível do indivíduo; previsões de eventos como pandemias e tempestades, por exemplo; além de contribuir para a ocorrência de tomadas de decisões automatizadas precisas e com alta velocidade. Pontos que no ramo empresarial influenciaram o desenvolvimento de novos processos de negócios.

É necessária toda uma infraestrutura tecnológica direcionada à utilização de Big Data. Essa infraestrutura pode ser visualizada na Figura 1 a seguir.

Figura 1 – Infraestrutura de utilização de Big Data



Fonte: adaptada de Laudon e Laudon (2014).



Como observado, Big Data compreende enormes quantidades de dados das mais variadas fontes, encaminhados aos clusters Hadoop, para posteriormente serem retrabalhados pelos *datas warehouses* e *marts* e plataformas analíticas, antes de chegar aos interessados em processar informações importantes (LAUDON; LAUDON, 2014).

Datas warehouses compreendem bancos de dados direcionados aos dados operacionais de uma empresa relacionados a dados externos, recombinao dados e os armazenando novamente. O *data mart* é semelhante, mas gera grupos menores de dados selecionados de acordo com os interesses finais (TAURION, 2013).

As plataformas analíticas são as responsáveis por oferecer processamento de grandes grupos de dados com rapidez e eficiência, otimizando as consultas solicitadas. Processamento equivalente à computação em memória, que cria banco de dados armazenados na própria memória RAM de um computador, facilitando cálculos em poucos segundos, conforme exposto por Laudon e Laudon (2014).

Completando as ferramentas e os sistemas utilizados em Big Data está o Hadoop, composto por clusters de servidores tratando os dados de duas formas:

- Processamento paralelo por meio do Hadoop MapReduce (HMR), que mapeia os dados e os fragmenta em tuplas a serem combinadas, a fim de fornecer o resultado esperado por uma consulta.
- Segmentação dos dados em pequenos blocos de dados por meio do Hadoop Distributed File System (HDFS), que garante a integridade contra falhas, realizando cópias desses blocos em três repositórios diferentes, em média.

Para desenvolver um projeto de Big Data utilizando Hadoop e obter sucesso, Taurion (2013) acredita que a ferramenta deva realmente ter utilidade para um setor empresarial, por exemplo, se a empresa possui a capacidade e qualificação necessária e que a fornecedora de uma distribuição Hadoop tenha essa ferramenta como seu principal produto e assim garante-se o suporte.

Por fim, os bancos NoSQL possuem diferentes formas e tamanhos, dependendo das peculiaridades e necessidades exigidas, direcionados a gerenciar volumes consideráveis de dados em computadores dispersos, aumentando ou reduzindo máquinas se for necessário (LAUDON; LAUDON, 2014).

Referências bibliográficas


LAUDON, K.; LAUDON, J. **Sistemas de informações gerenciais**. Tradução Célia Taniwaki. 11. ed. São Paulo: Pearson, 2014.

TAURION, C. **Big Data**: velocidade, volume, variedade, veracidade, valor. Rio de Janeiro: Brasport, 2013.

PARA SABER MAIS

Ao lidar com Big Data, é muito importante que você saiba quais as perguntas-chaves para uma consulta eficiente de dados, um reflexo do conhecimento aprofundado da informação que você está procurando. Se não há a formulação correta desses questionamentos, a pesquisa fica comprometida. Para resolver esse problema, existem ferramentas tais como o *data mining* (mineração de dados), destinado a identificar padrões de relacionamento ocultos em enormes bancos de dados (LAUDON; LAUDON, 2014).

Como o próprio nome sugere, o *data mining* visar extrair conhecimento dos dados por meio de métodos de análise de dados com algoritmos específicos de processamento de grandes volumes e com o uso de inteligência artificial, segundo Tan *et al.* (2014).



Para Tan *et al.* (2014), o processamento do *data mining* surge no momento em que o usuário define um problema, escolhendo os dados e ferramentas para analisá-lo. Basicamente, são utilizadas três técnicas para tal situação: estatística, que é a essência do *data mining*; inteligência artificial, como forma de imitar a expertise humana na resolução de problemas; e *machine learning* (aprendizado de máquina), com a confecção de algoritmos capazes de aprender e reconhecer padrões em dados que auxiliarão a tomada de decisão.

Os tipos de informações oriundas do *data mining* incluem, segundo Laudon e Laudon (2014):

- **Associativas:** informações referentes a ocorrências relacionadas com um único evento. Amplamente utilizada para vendas cruzadas.
- **Sequenciais:** originadas durante a realização de um evento, como a compra de eletrodomésticos e móveis durante o processo de compra de um imóvel.
- **Classificativas:** informações que classificarão e definirão o perfil de clientes que, por exemplo, se mantêm fiéis ou que tendem a abandonar a empresa. Com tais informações, os gestores podem desenvolver estratégias para a retenção e manutenção de clientes.
- **De aglomeração:** informações originadas antes das classificativas, agrupando perfis ou dados em categorias de clientes a serem analisadas.
- **De prognósticos:** informações que preveem valores a partir da análise de outros valores. Auxiliam, por exemplo, em previsões de vendas ou até mesmo de ações em uma bolsa de valores.

As práticas do *data mining* visam analisar os dados em alto nível frente à necessidade de obter tendência e padrões. Logo, o *data mining* pode ser aplicado em todos os setores funcionais de uma empresa, em organizações governamentais ou mesmo na área acadêmica (pesquisas científicas). Evidentemente que a aplicação mais usual é na identificação de padrões de consumo e direcionar as informações obtidas às empresas, que podem, então, promover campanhas de marketing direcionadas ou simplesmente visualizar quais clientes são potencialmente mais lucrativos (LAUDON; LAUDON, 2014).

Referências bibliográficas

LAUDON, K.; LAUDON, J. **Sistemas de informações gerenciais**. Tradução Célia Taniwaki. 11. ed. São Paulo: Pearson, 2014.

TAN, P. *et al.* **Introduction to data mining**: Pearson new international edition. 1 ed. Londres, 2014.

TEORIA EM PRÁTICA

Uma grande empresa com filiais espalhadas por todo o país é a mais conhecida fabricante de molhos prontos (do tipo maionese) para utilização em estabelecimentos alimentícios e vendas no varejo em supermercados. Procurando lançar um novo produto no mercado, composto por um ingrediente de sabor típico do Brasil, o gerente-geral da organização delegou aos departamentos de marketing e tecnologia que promovessem consultas e pesquisas boca a boca nas regiões de cada uma das filiais para identificar qual é o sabor que, para os consumidores, reflete a sua nação. Pensando nisso, como a concepção de Big Data poderia contribuir no processo de pesquisa? Qual ferramenta pode ser utilizada?

Para conhecer a resolução comentada proposta pelo professor, acesse a videoaula deste *Teoria em Prática* no ambiente de aprendizagem.



LEITURA FUNDAMENTAL

Indicações de leitura

Indicação 1

Este artigo traz uma reflexão do conceito Big Data no contexto do conhecimento, aplicado em pesquisas em ciências sociais, além de apresentar perspectivas sobre privacidade e vigilância social.

Para realizar a leitura, acesse a Biblioteca Virtual da Kroton e busque pelo título da obra.


SCHROEDER, R. Big Data: moldando o conhecimento, moldando a vida cotidiana. **Matrizes**, São Paulo, v. 12, n. 2, p. 135-163, 2018.

Indicação 2

Este artigo propõe uma análise crítica sobre os impactos epistemológicos e metodologias de pesquisas em Big Data do ponto de vista científico. Ou seja, aplicações reais da concepção referentes à comunicação.

Para realizar a leitura, acesse a Biblioteca Virtual da Kroton e busque pelo título da obra.

SOARES, A. T. N. Epistemologia, métodos e teorias da comunicação na era do Big Data: panorama crítico da pesquisa em mídias sociais. **Comunicação e Sociedade**, n. 33, p. 151-166, Brasília, 2018.



QUIZ

Prezado aluno, as questões do Quiz têm como propósito a verificação de leitura dos itens *Direto ao Ponto, Para Saber Mais, Teoria em Prática e Leitura Fundamental*, presentes neste *Aprendizagem em Foco*.

Para as avaliações virtuais e presenciais, as questões serão elaboradas a partir de todos os itens do *Aprendizagem em Foco* e dos slides usados para a gravação das videoaulas, além de questões de interpretação com embasamento no cabeçalho da questão.

1. Pensando na definição de Big Data, qual das alternativas a seguir traz corretamente as palavras-chave dessa concepção?
 - a. Integração; agregação; segurança; dispersão.
 - b. Valor; criação; avaliação; consultas; processamento.
 - c. Distribuição; transparência; otimização; análises.
 - d. Estratégia; oferta; objetivos; velocidade; distribuição.
 - e. Volume; velocidade; veracidade; variedade; valor.

2. Qual das alternativas a seguir apresenta corretamente os tipos de informações obtidas por meio do *data mining* (mineração de dados)?
 - a. De incrementação; associativas; colaborativas; introdutória; de seleção.
 - b. Retroativas; corporativas; funcionais; estruturais; de visualização.
 - c. Associativas; sequenciais; classificativas; de aglomeração; de prognósticos.

- d. De incrementação; sequenciais; de aglomeração; colaborativas; de seleção.
- e. De incrementação; funcionais; emergenciais; valiosas; de descarte.



GABARITO

Questão 1 - Resposta E

Resolução: Conhecidos como “5 Vs” o volume, a velocidade, a veracidade, a variedade e o valor são palavras que definem o que é Big Data e sua concepção.

Questão 2 - Resposta C

Resolução: Os tipos de informações obtidas do *data mining* são: associativas, referentes a ocorrências de um mesmo evento; sequenciais, que ocorrem durante o evento; classificativas, que definem o perfil de clientes; de aglomeração, agrupando perfis identificados; e de prognósticos, que preveem valores a partir da análise de outros valores.



BONS ESTUDOS!