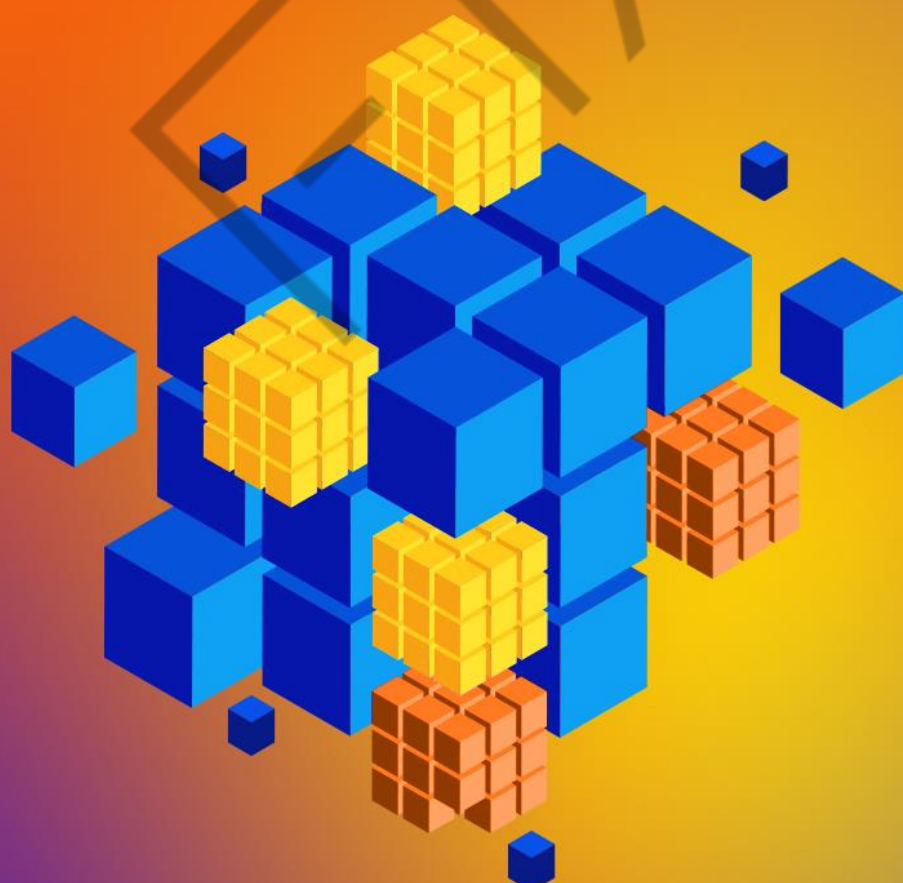


GESTÃO DE INFRAESTRUTURA DE TI

INTRODUÇÃO AO BANCO DE DADOS

ADRIANO BREVIOLIERI E RITA DE CÁSSIA



3

LISTA DE FIGURAS

Figura 3.1 – Máquina de escrever manual	6
Figura 3.2 – Arquivos de aço.....	6
Figura 3.3 – Escaninho de pesquisa	6
Figura 3.4 – Computador “pessoal”	7
Figura 3.5 – Transporte de computador	Er
ror! Bookmark not defined.	
Figura 3.6 – Disco Hard Drive	Er
ror! Bookmark not defined.	
Figura 3.7 – Computador pessoal integrado	Er
ror! Bookmark not defined.	
Figura 3.8 – Estrutura do SGBD.....	15
Figura 3.9 – Fabricantes de BD.....	16
Figura 3.10 – SGBD	18
Figura 3.11 – Níveis de BD	22
Figura 3.12 – Dicionário de dados.....	23
Figura 3.13 – RAID 0.....	25
Figura 3.14 – RAID 1.....	25
Figura 3.15 – RAID 2.....	26
Figura 3.16 – RAID 3.....	26
Figura 3.17 – RAID 4.....	27
Figura 3.18 – RAID 5.....	27
Figura 3.19 – RAID 10.....	28
Figura 3.20 – Banco de dados Relacionais x Não Relacionais	33

LISTA DE QUADROS

Quadro 3.1 – Composição de um dado.....	13
-----------------------------------------	----

EMSE

LISTA DE TABELAS

Tabela 3.1 – Tabela prontuário	8
--------------------------------------	---

EMSE

SUMÁRIO

3 INTRODUÇÃO AO BANCO DE DADOS	6
3.1 Histórico	6
3.1.1 Década de 1960	6
3.1.2 Década de 1970	7
3.1.3 Década de 1980	8
3.1.4 Fim da década de 1980 e início da década de 1990	10
3.1.5 Década de 1990	10
3.1.6 Século XXI	11
3.2 Conceituando Banco de Dados	11
3.3 Mas o que é um Dado?	12
3.4 Diferenciando Dado e Informação	12
3.5 Propriedades de um Banco de Dados	13
3.6 Qual a importância do Banco de Dados para as empresas?	14
3.7 Armazenamento sem Banco de Dados	14
3.8 Conceituando Sistema Gerenciador de Banco de Dados	15
3.9 Conceituando Sistema de Banco de Dados	17
3.10 Visualizando os componentes de um Sistema de Banco de Dados	18
3.11 Caracterizando um Sistema de Banco de Dados	18
3.12 Classes de usuários em um Sistema de Banco de Dados	19
3.13 Vantagens de utilizar um Sistema de Banco de Dados	20
3.14 Arquitetura de três esquemas e a independência de dados	21
3.15 Capacidade de um Sistema de Gerenciamento de Banco de Dados	22
3.16 Armazenamento	24
3.16.1 RAID – Redundant Array of Independent Drives (Conjunto Redundante de Discos Independentes)	24
3.16.2 NoSQL – Bancos não relacionais para Armazenamento de Dados na web	29
3.16.2.1 Bancos de Dados NoSQL	29
3.16.2.2 Natureza dos Dados da web	30
3.16.2.3 Bancos de Dados orientados a documentos	30
3.16.2.4 Escalabilidade	31
3.16.3 Big Data	32
3.16.3.1 Fatores de influência para Big Data	33
3.16.3.2 Como o Big Data muda o modelo de negócios?	34
3.16.3.3 E a ética?	35
REFERÊNCIAS	37

3 INTRODUÇÃO AO BANCO DE DADOS

3.1 Histórico

Por meio da história, é possível entender o momento atual da tecnologia voltada para banco de dados e indicar as tendências que estão se consolidando e construindo a nova história.

É necessário voltar aos registros feitos em consultórios médicos, hospitais, cartórios, bibliotecas, grandes empresas e todas as informações armazenadas e pesquisadas frequentemente.



Figura 3.1 – Máquina de escrever manual
Fonte: Google Imagens (2015)



Figura 3.2 – Arquivos de aço
Fonte: Google Imagens (2015)



Figura 3.3 – Escaninho de pesquisa
Fonte: Google Imagens (2015)

3.1.1 Década de 1960

Os computadores começam a ser utilizados nas empresas juntamente com o crescimento da capacidade de armazenamento.

Detalhes de armazenamento dependiam do tipo de informação a ser guardada e, dessa forma, para um campo extra, necessitava-se de uma reescrita dos fundamentos de acesso/modificação do esquema. Os usuários precisavam conhecer a estrutura física do banco de dados para poder realizar uma consulta.



Figura 3.4 – Computador “pessoal”
Fonte: Google Imagens (2017)

3.1.2 Década de 1970

Edgar Frank Codd propõe o modelo de dados relacional, que se tornou um marco em como pensar em banco de dados.

Ele desconectou a estrutura lógica do banco de dados do método de armazenamento físico. E esse sistema passou a ser padrão (BALIERO, 2015).

MÉDICO

Cod_med	Nome_med	Especialidade
---------	----------	---------------

PACIENTE

Cod_pac	Nome_pac
---------	----------

CONSULTA

Cod_med	Cod_pac	Dia	Hora
---------	---------	-----	------

MÉDICO

Cod_med	Nome_med	Especialidade
M1	Leon	Pediatra
M2	Carlos	Cardiologia
M3	Maria	Neurologia

PACIENTE

Cod_pac	Nome_pac
P1	Lisandra
P2	Gerson
P3	Matias

CONSULTA

Cod_med	Cod_pac	Dia	Hora
M1	P2	2/1	14
M1	P3	7/1	15
M2	P1	2/1	14

Tabela 3.1 – Tabela prontuário
Fonte: Google Imagens (2017)

O termo Sistema de Gerenciamento de Banco de Dados Relacional (SGBDR – RDBMS em inglês) foi definido durante esse período.

3.1.3 Década de 1980

A comercialização de sistemas relacionais começa a virar uma febre entre as organizações.

A Linguagem Estruturada de Consulta – SQL (Structured Query Language) torna-se um padrão mundial.

A IBM transforma o DB2 em carro-chefe da empresa de produtos para BD.

Os modelos em rede e hierárquicos passam a ficar em segundo plano praticamente sem desenvolvimentos, utilizando seus conceitos, porém, vários sistemas legados continuam em uso.



Figura 3.2 – Transporte de computador
Fonte: Google Imagens (2017)

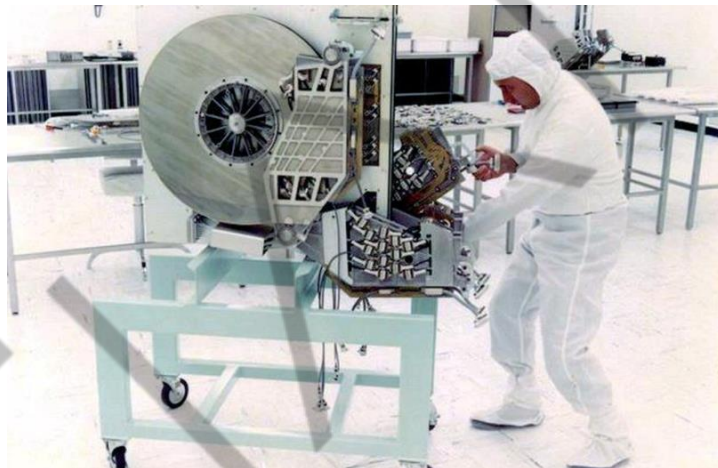


Figura 3.3 – Disco Hard Drive
Fonte: Google Imagens (2017)



Figura 3.4 – Computador pessoal integrado
Fonte: Google Imagens (2017)

3.1.4 Fim da década de 1980 e início da década de 1990

O desenvolvimento do IBM PC desperta muitas empresas e produtos de BD, como: RIM, RBASE 5000, PARADOX, OS/2 Database Manager, Dbase III e IV (mais tarde, transformado em FoxBase e, depois, como Visual FoxPro), Watcom SQL, entre outros.

Muitos desenvolvimentos foram realizados em ferramentas para o desktop e aplicações (client tools), tais como: PowerBuilder (Sybase), Oracle Developer, Visual Basic (Microsoft), entre outros.

O modelo cliente-servidor (client-server) passa a ser uma regra para futuras decisões de negócio e vê-se o desenvolvimento de ferramentas de produtividade, como Excel/Access (Microsoft) e ODBC.

3.1.5 Década de 1990

É quando vemos a explosão da Internet/WWW e uma louca corrida para prover acesso remoto a sistemas de computadores com dados legados. Percebe-se um crescimento exponencial na tecnologia Web/BD.

Aumenta o uso de soluções de código aberto (*open source*) por meio de GCC, CGI, Apache, MySQL etc.

Processos de Transação em Tempo Real (OLTP – Online Transaction Process) e Processos Analíticos em Tempo Real (OLAP – Online Analytical Process) atingem maturidade por intermédio de muitos negócios, utilizando os PDVs (Pontos de Venda).

O grande investimento em empresas de Internet impulsiona as vendas de ferramentas para a conexão Web/Internet/BD. Active Server Pages, Front Page, Java Servlets, JDBC, Enterprise Java Beans, ColdFusion, Dream Weaver, Oracle Developer 2000 são exemplos dessas ferramentas.

3.1.6 Século XXI

Sólidos crescimentos em aplicações para BD continuam.

Aparecem mais aplicações que interagem com PDAs (Personal Digital Assistant), transações em PDVs, consolidação de vendas etc.

Três companhias predominam no amplo mercado de BD: IBM (que comprou a Informix), Microsoft e Oracle.

Devido ao aprimoramento da área de Telecom e à explosão de aplicações para Web, o acesso a dispositivos móveis passa a ser massivo.

Redes sociais e empresariais e portais colaborativos se tornam padrões de comunicação entre as pessoas.

Big Data começa a se tornar tendência em companhias que manipulam grande volume de dados.

Banco de Dados NoSQL começa a ser utilizado visando facilitar o uso de aplicações comerciais.

3.2 Conceituando Banco de Dados

Elmasri e Navathe (2011) descrevem que os bancos de dados e sua tecnologia têm um importante impacto sobre o uso crescente dos computadores. Além disso, segundo eles, desempenham papel crítico em quase todas as áreas nas quais os computadores são usados, incluindo negócio, comércio eletrônico, engenharia, medicina, genética, direito, educação, biblioteconomia, entre outros exemplos.

Podemos dizer que um **Banco de Dados** é uma coleção de dados persistentes, usada pelos sistemas de aplicação de uma determinada empresa.

Mas qual é a ideia de persistência? Entende-se por dados persistentes aqueles que, uma vez aceitos por um SGBD (Sistema Gerenciador de Banco de Dados) para a entrada no banco de dados, somente poderão ser removidos por uma requisição explícita ao SGBD.

Bancos de dados são utilizados por empresas. Chamamos de empresa uma organização comercial, científica, técnica ou qualquer outra organização. Podendo ser um único indivíduo, uma corporação ou grande empresa. Uma empresa necessita manter muitos dados sobre sua operação, seguem exemplos de “dados persistentes”:

- Dados sobre produtos.
- Dados sobre contas.
- Dados sobre pacientes.
- Dados sobre alunos.
- Dados sobre planejamento.

Uma coleção de dados representa dados armazenados e inter-relacionados, que atendem às necessidades de vários usuários dentro das organizações.

3.3 Mas o que é um Dado?

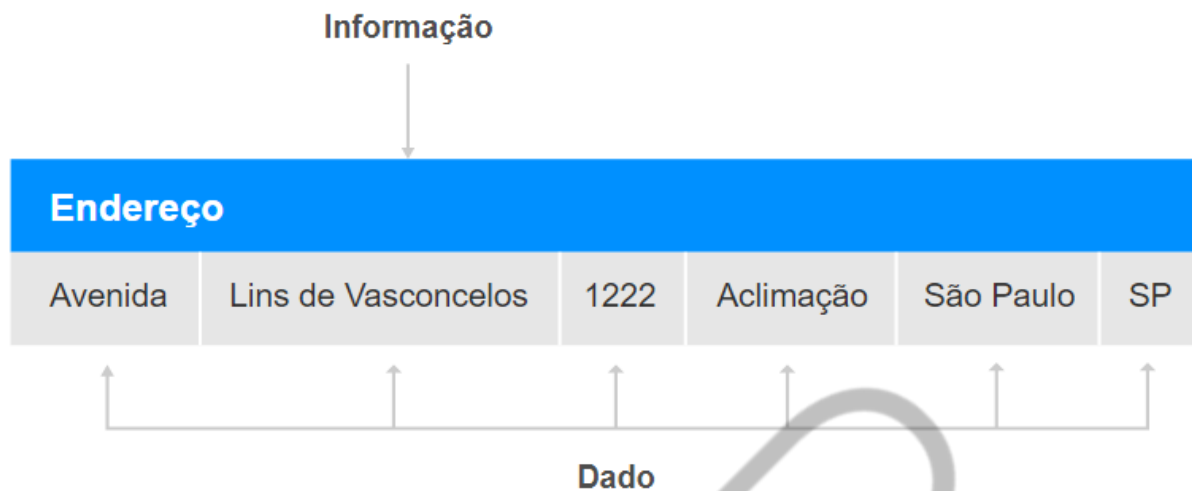
Dados são fatos conhecidos que registrados que possuem significado implícito, podendo ser armazenados (exemplos: nomes, telefones e endereços).

3.4 Diferenciando Dado e Informação

Qual a diferença entre DADO e INFORMAÇÃO?

Podemos dizer que Dado é utilizado para se referir ao que realmente está armazenado. (exemplo: endereço – logradouro, número, complemento etc.).

Informação é utilizada para se referir ao significado dos dados para um determinado usuário. (exemplo: endereço).



Quadro 3.1 – Composição de um dado
Fonte: FIAP (2017)

Concluindo: Dado é aquilo guardamos por meio de uma estrutura de armazenamento, e Informação, o que extraímos de uma estrutura de armazenamento. A informação sempre tem um significado para o usuário, podendo ser constituída por vários dados.

3.5 Propriedades de um Banco de Dados

Um banco de dados possui propriedades. Podemos destacar, a seguir, algumas delas:

- Coleção lógica e coerente de dados (dados dispostos de forma desordenada não podem ser referenciados como banco de dados).
- É projetado, construído e populado com dados para um propósito específico.
- Possui um conjunto predefinido de usuários e aplicações.
- Representa algum aspecto do mundo real, porção da realidade, o qual é chamado de “minimundo”; qualquer alteração efetuada no minimundo é automaticamente refletida no banco de dados.

3.6 Qual a importância do Banco de Dados para as empresas?

Gestão de conhecimento é o grande desafio empresarial dos dias de hoje, sendo a estratégia central para desenvolver a competitividade de empresas e países. Discute o investimento em pesquisa e desenvolvimento os avanços da tecnologia gerencial, tecnologias de informática e de telecomunicações e as conclusões das teorias sobre criatividade e aprendizado individual e organizacional.

Assim sendo, o banco de dados da empresa permite acesso ao conhecimento tácito angariado pelos colaboradores da empresa no momento do cadastro das informações sistêmicas, guardando um ativo valioso para as empresas em forma de documentos, itens, cadastros, campos de informação, entre outros. Eles podem ser acessados e ordenados conforme novas necessidades, sem precisar que tudo seja reaprendido a cada nova situação.

O banco de dados traz para a empresa um repositório de materiais de referência, um conhecimento explícito que pode ser facilmente acessado e que evita duplicações de esforços. Possibilita também a criação de listas e descrições das competências de indivíduos dentro e fora da organização, por exemplo, facilitando o compartilhamento de conhecimento tácito. Essa ação tornaria a gestão de pessoal muito mais fácil e rápida para as tomadas de decisões de alocação de pessoas dentro de suas competências, ainda seguindo no mesmo exemplo.

3.7 Armazenamento sem banco de dados

Dados de diferentes aplicações não estão integrados. Dados estão projetados para atender a uma aplicação específica.

O mesmo objeto é representado múltiplas vezes na base de dados, o que acarreta uma redundância de dados que não pode ser controlada (exemplo: guarda de fotos de viagem).

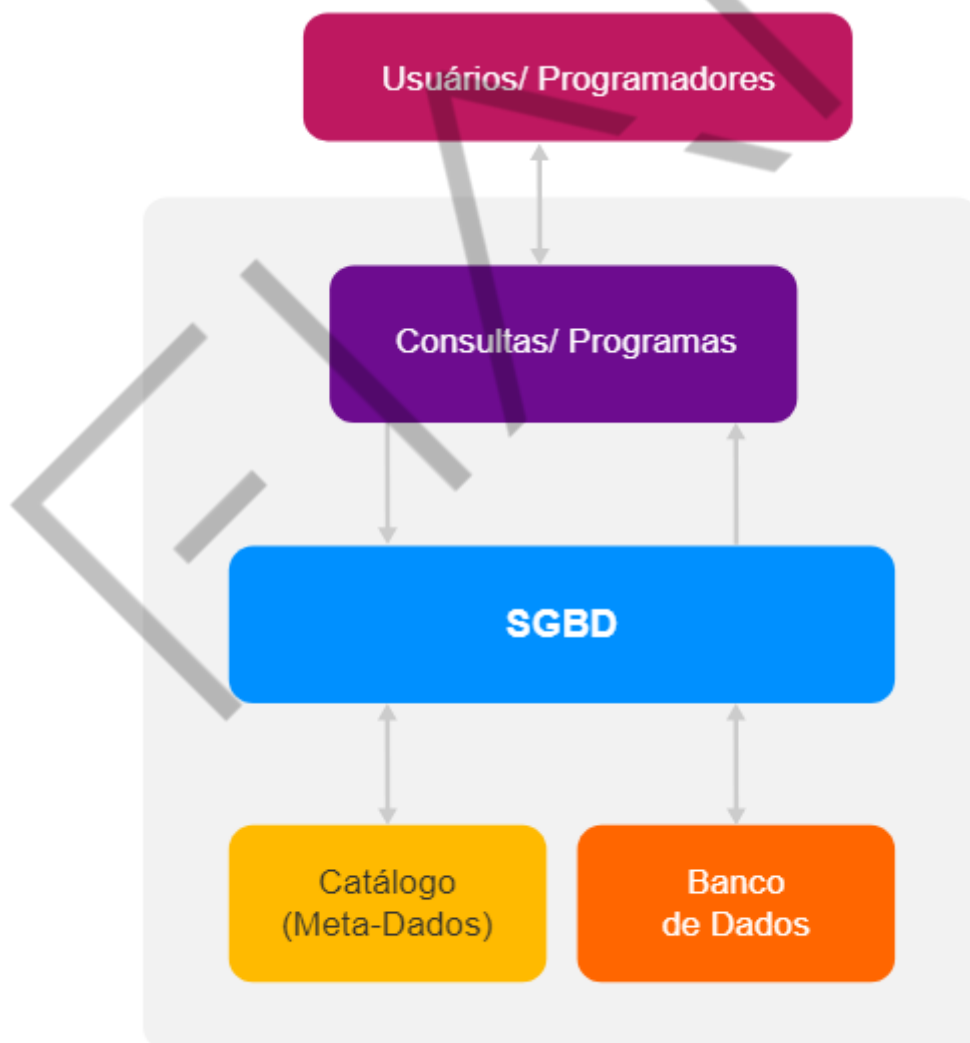
Isso traz uma inconsistência dos dados (devido à não sincronização temporal dos diversos bancos) e trabalho repetitivo de redigitação de informações, podendo levar a erros de dados e, conseqüentemente, a dificuldades de extração de

informações. Sem levar em consideração que o conteúdo que consta de um cadastro no banco de dados pode não constar em outro.

Elmasri e Navathe (2011) descrevem que um banco de dados tem alguma fonte da qual o dado é derivado, algum grau de interação com eventos do mundo real e um público que está ativamente interessado em seu conteúdo.

3.8 Conceituando Sistema Gerenciador de Banco de Dados

Sistema Gerenciador de Banco de Dados, SGBD, ou DBMS (Database Management System), refere-se a um sistema de software genérico para manipular bancos de dados.



Sistema de Banco de Dados

Figura 3.5 – Estrutura do SGBD

Fonte: FIAP (2017)

É um software com recursos específicos para facilitar o processo de definição, construção, manipulação e compartilhamento das informações dos bancos de dados e o desenvolvimento de programas aplicativos.

O Sistema Gerenciador de Banco de Dados propicia um ambiente conveniente e eficiente para a recuperação e armazenamento das informações do banco de dados. (exemplos: Oracle, Sybase, DB2, Informix, SQL Server, MySQL, PostgreSQL, Interbase, Caché, entre outros).



Figura 3.6 – Fabricantes de BD
Fonte: Google Imagens (2017)

Quando falamos que o SGBD é um software que facilita alguns processos associados aos bancos de dados, podemos dizer, segundo Elmasri e Navathe (2011) que:

- Definição de um banco de dados relaciona-se à especificação de tipos, estruturas e restrições associados aos dados que serão armazenados.
- Construção de um banco de dados refere-se ao processo de armazenar os dados em algum meio controlado pelo SGBD.
- Manipulação de um banco de dados vincula-se à inclusão de funções, como consulta ao banco de dados para recuperação de dados específicos, atualização do banco de dados e geração de relatórios com base nos dados.

- Compartilhamento de um banco de dados permite que seja acessado por diversos usuários e programas simultaneamente.

Quando não usar um SGBD?

Se a base de dados e as aplicações forem simples, bem definidas e sem perspectivas de mudanças e não necessitarem de acesso concorrente aos dados.

3.9 Conceituando Sistema de Banco de Dados

Podemos dizer que Sistema de Banco de Dados é um conjunto formado por um Banco de Dados (coleção de dados persistentes), mais as aplicações (SGBD) que o manipulam.

É um sistema de manutenção de registros por computador, envolvendo quatro componentes principais:

- a) dados;
- b) hardware;
- c) software;
- d) usuários.

Os sistemas de banco de dados são projetados para gerir grandes volumes de informações.

Gerir as informações implica definição das estruturas de armazenamento e também dos mecanismos para a manipulação dessas informações.

Um sistema de banco de dados deve garantir a segurança das informações armazenadas contra eventuais problemas com o sistema, além de impedir tentativas de acesso não autorizadas.

3.10 Visualizando os componentes de um Sistema de Banco de Dados

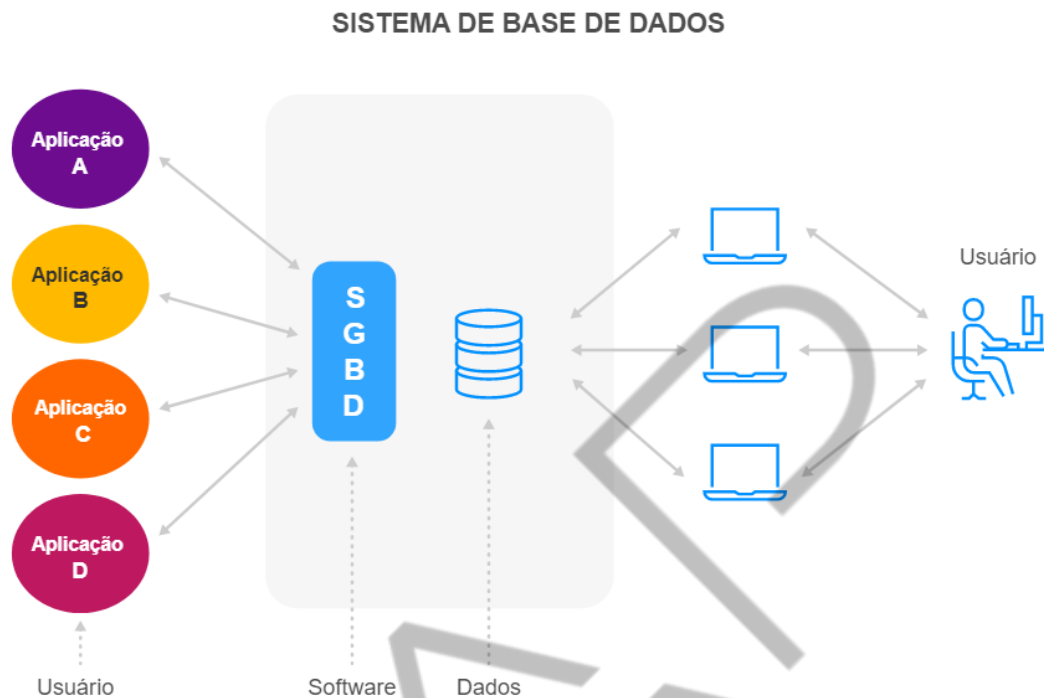


Figura 3.7 – SGBD
Fonte: Google Imagens (2017)

O Sistema Gerenciador de Banco de Dados (SGBD) permite que um banco de dados tenha a característica MULTIUSUÁRIO, ou seja, vários usuários podem acessá-lo ao mesmo tempo.

De modo geral, os dados de um banco de dados estarão integrados e compartilhados.

Integrado: unificação de vários arquivos, eliminação de redundância (Exemplo: dados do aluno e boletim do aluno).

Compartilhado: vários usuários podem ter acesso aos mesmos dados, possivelmente ao mesmo tempo (Exemplo: consulta da ficha cadastral).

3.11 Caracterizando um Sistema de Banco de Dados

São características de um sistema de banco de dados:

- Autodescrição de um banco de dados. Em seu sistema, há o banco de dados e uma definição completa de sua estrutura e restrições. Essas informações, chamadas de METADADOS, são armazenadas em uma espécie de catálogo, que contém a: estrutura de cada arquivo, tipo e formato de armazenamento e um conjunto de restrições sobre os dados.
- Independência de dados do programa, ou seja, a estrutura dos arquivos de dados é armazenada no catálogo do SGBD, separadamente dos programas de acesso. Caso alguma descrição do catálogo seja alterada, nenhum programa associado ao SGBD precisa ser alterado, automaticamente, as alterações são refletidas e acessadas pelos programas de SGBD.
- Suporte a várias visões dos dados. Normalmente, um banco de dados possui muitos usuários, cada um com necessidades diferentes. Tais necessidades refletem em visões diferentes dos dados, que podem ser subconjuntos dos dados armazenados, ou informações derivadas das informações armazenadas. Para o usuário final, não há a necessidade de conhecer a origem da informação (armazenada ou derivada). O SGBD precisa fornecer funcionalidades que permitam a criação dessas visões.
- Compartilhamento de dados e processamento de transação multiusuário referem-se à característica de que um SGBD deve permitir que múltiplos usuários acessem o banco de dados ao mesmo tempo. Para isso, o SGBD precisa de um software para controle de concorrência. Isso posto, é possível garantir que vários usuários possam realizar alterações simultaneamente de uma forma controlada e que essas transações concorrentes operem de maneira correta e eficiente.

3.12 Classes de usuários em um Sistema de Banco de Dados

Podemos considerar três classes de usuários, são elas:

Programadores de Aplicações: responsáveis pela escrita de programas de aplicações de banco de dados em alguma linguagem de programação. Esses programas acessam o banco de dados emitindo uma requisição apropriada (instrução SQL).

Usuários Finais: um usuário pode acessar o banco de dados através de uma aplicação desenvolvida pelos programadores de aplicações.

Administrador de Banco de Dados (DBA – Data Base Administrator) e Administrador de Dados (DA – Data Administrator): decidem os dados que devem ser armazenados e estabelecem normas para manter e tratar esses dados.

3.13 Vantagens de utilizar um Sistema de Banco de Dados

Utilizar sistemas de banco de dados nos traz várias vantagens, a seguir, listamos as principais:

Os dados podem ser compartilhados: várias aplicações podem acessar os mesmos dados de um banco de dados.

A redundância pode ser reduzida: cada aplicação tem seus próprios arquivos, ou seja, os mesmos dados são armazenados várias vezes. Quando a redundância ocorre, temos alguns problemas como: o desperdício de espaço em disco devido à repetição da informação em vários lugares, a dificuldade para a atualização das informações repetidas em lugares diferentes e a inconsistência em função da falha durante a atualização das informações repetidas.

Para exemplificar, imagine que, em uma escola particular, as áreas Pedagógica, Secretaria e Financeiro tivessem seus próprios arquivos de alunos e fosse necessária a alteração do endereço de um aluno. Normalmente, a secretaria faz a recepção de documentos e atualização cadastral do aluno, portanto, é ela quem receberá a solicitação de alteração de endereço.

- Por terem arquivos distintos, a Secretaria precisa informar às demais áreas sobre a alteração cadastral do aluno, para que os arquivos de dados permaneçam consistentes. Se, por algum motivo, a Secretaria não repassar a informação, o Departamento Financeiro acabará enviando um boleto de pagamento para um endereço que não é mais válido. Conclusão; o aluno não receberá o boleto e, por consequência, haverá falta de pagamento, caso ele não perceba que o boleto não chegou em tempo hábil.

A inconsistência pode ser evitada: considerando o exemplo anterior, se os departamentos de uma escola particular tiverem seus próprios arquivos dos

alunos e ocorrer uma alteração nos dados cadastrais de um aluno e ele não for replicada às áreas que necessitam dessa informação, teremos uma inconsistência. É necessário que a propagação das informações seja realizada.

Suporte a transações: transação é uma unidade lógica de trabalho de banco de dados, envolvendo diversas operações de banco de dados (exemplo: transferência de dinheiro entre contas). Neste caso, temos duas atualizações: uma para retirar o dinheiro da conta A e outra para depositá-lo na conta B. Se o usuário declarar que as duas atualizações fazem parte da mesma transação, o sistema poderá garantir que ambas serão realizadas ou nenhuma delas.

A integridade pode ser mantida: é assegurar que os dados no banco de dados estão corretos. Podemos aplicar restrições aos dados para que a integridade seja garantida. (Exemplos: podemos inserir restrições para o dado referente ao tipo de uma nota fiscal (1 = Entrada, 2 = Saída), ao sexo de uma pessoa (F ou M). O tipo de dado atribuído também é considerado uma restrição, pois limita o conteúdo que será armazenado pelo dado).

Segurança: o DBA pode assegurar que o único meio de acesso ao banco de dados seja através dos canais apropriados. Com isso, define restrições de segurança a serem verificadas sempre que houver uma tentativa de acesso a dados confidenciais. As restrições podem ser para cada tipo de acesso (busca, inclusão, exclusão etc.) e cada item de informação do banco de dados.

3.14 Arquitetura de três esquemas e a independência de dados

O objetivo desta arquitetura é separar as aplicações do usuário de banco de dados.

Podemos dividir em três níveis:

Nível externo ou de visão do usuário: representa as visões de usuários individuais. Descreve partes do banco de dados, de acordo com as necessidades de cada usuário, individualmente.

Nível conceitual: representa a visão da comunidade de usuários. Descreve quais dados estão armazenados e seus relacionamentos.

Nível interno ou físico: representa a visão do meio de armazenamento. Nível mais baixo de abstração. Descreve como os dados estão realmente armazenados.

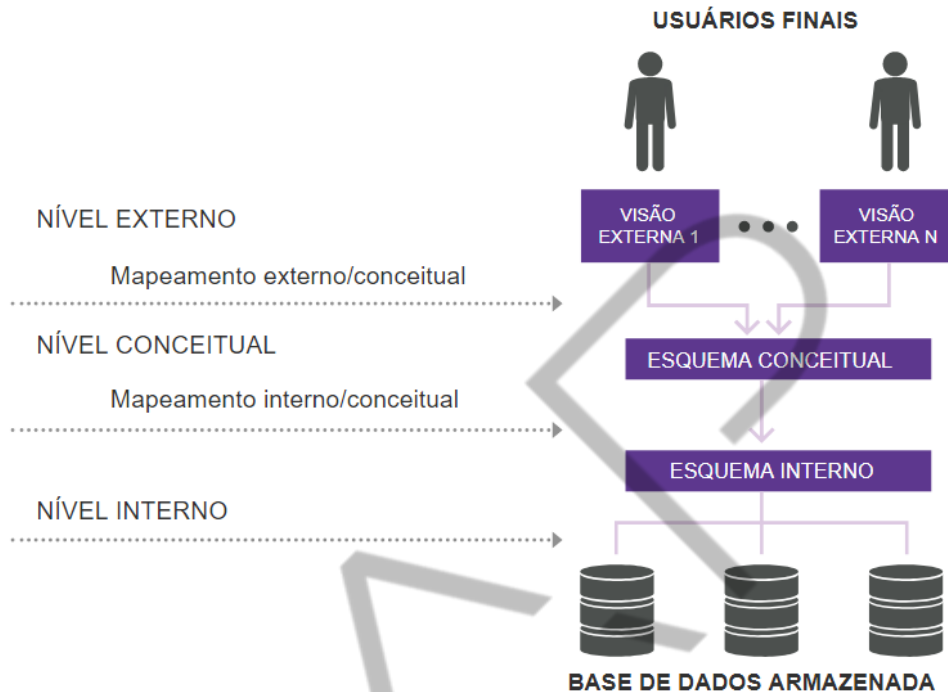


Figura 3.8 – Níveis de BD
Fonte: Google Imagens (2017)

3.15 Capacidade de um Sistema de Gerenciamento de Banco de Dados

O Sistema de Gerenciamento de Banco de Dados deve ser capaz de:

Definição de dados: o SGBD deve incluir componentes para processar e compilar DDL – Data Definition Language (Definição de dados: tabelas, campos etc.).

Manipulação de dados: o SGBD deve lidar com as requisições do usuário para buscar, atualizar ou excluir dados existentes no banco de dados, para isso, deve incluir componentes para processar e compilar DML.

O que ocorre conceitualmente quando uma requisição é enviada ao SGBD?

- O usuário faz um pedido de acesso usando determinada sublinguagem de dados (normalmente SQL).
- O SGBD intercepta o pedido e o analisa.

- O SGBD verifica as características da solicitação do usuário.
- O SGBD executa as operações necessárias sobre o banco de dados armazenado.

DML – Data Manipulation Language (Comandos para manipulação de dados: select, insert, update e delete).

Otimização e execução: determinar um modo eficiente de implementar a requisição feita pelo usuário.

Segurança e integridade de dados: deve monitorar as requisições de usuários e rejeitar toda tentativa de violar as restrições de segurança e integridade definidas pelo DBA.

Recuperação de dados e concorrência: gerenciadores de transação devem impor certos controles de recuperação e concorrência.

Dicionário de dados: contém os dados sobre os dados (chamados de metadados ou descritores). Um dicionário completo inclui os programas que utilizam partes do banco de dados, usuários que exigem determinados relatórios etc.

Exemplos de Dicionário de Dados:

TABELA Vet01_Cliente				
Campo	Tipo de Dado	Descrição	Observação	
fk011t_idCliente	INTEGER	Chave Primária	Não nulo	
at01st_nomeCliente	VARCHAR(60)	Nome do Cliente	Não nulo	
at01dt_dtNasc	DATE	Data de nascimento do Cliente	Não nulo	
at01st_cpf	VARCHAR(11)	CPF do Cliente	Não nulo	
at01st_logradouro	VARCHAR(60)	Endereço (Rua/Av./Travessa e nº)		
at01st_bairro	VARCHAR(40)	Bairro onde o cliente reside		
at01st_cidade	VARCHAR(40)	Cidade onde o cliente reside		
at01sc_uf	CHAR(2)	Sigla do estado onde o cliente reside		
at01st_cep	VARCHAR(8)	Código postal do endereço do cliente		
at01sc_sexo	CHAR(2)	Sexo do cliente	Não nulo	
at01st_email	VARCHAR(30)	Conta de e-mail do cliente		
fk031t_animal	INTEGER	Foreign Key da tabela Vet03_Animal		

Entidade: Cliente				
Atributo	Classe	Domínio	Tamanho	Descrição
Codig_cliente	Determinante	Númérico		
Nome	Simples	Texto	50	
Telefone	Multivalorado	Texto	50	Valores sem as máscaras de entrada
Cidade	Simples	Texto	50	
data_nascimento	Simples	Data		Formato dd/mm/aaaa

TABELA: ARTISTAS				
CAMPO	DESCRIÇÃO	TIPO	TAM	DEC
PK ART_CODIGO	Código do Cliente	INTEIRO	6	-
ART_NOME	Nome do Cliente	CHARACTER	100	-
ART_EMAIL	E-mail de Contato do Artista	CHARACTER	100	-
FK GEN_CODIGO	Código do Gênero	INTEIRO	6	-
ART_IMAGEM	Imagem do Artista	CHARACTER	100	-

TABELA: SUBGÊNEROS				
CAMPO	DESCRIÇÃO	TIPO	TAM	DEC
PK SUB_CODIGO	Código do Subgênero	INTEIRO	6	-
SUB_NOME	Nome do Subgênero	CHARACTER	100	-

TABELA: MÚSICAS				
CAMPO	DESCRIÇÃO	TIPO	TAM	DEC
PK MUS_CODIGO	Código da Música	INTEIRO	6	-
MUS_NOME	Nome da Música	CHARACTER	100	-
FK ART_CODIGO	Código do Artista	INTEIRO	6	-
FK SUB_CODIGO	Código do Subgênero	INTEIRO	6	-

Figura 3.9 – Dicionário de dados
Fonte: Google Imagens (2015)

3.16 Armazenamento

3.16.1 RAID – Redundant Array of Independents Drives (Conjunto Redundante de Discos Independentes)

- A função do RAID é a de criar um sistema alternativo para armazenar dados, composto por mais de um disco individual, tendo como objetivo potencializar o desempenho e garantir mais segurança.
- Seu funcionamento pode ser caracterizado como dois ou mais discos, os quais administram o sistema ao mesmo tempo, mantendo um espelhamento de um a outro, com a finalidade de realizar um *backup* automático. Dentre as vantagens propostas pelo sistema, podem-se citar as principais: a facilidade na recuperação de dados, utilização de diversas unidades de disco, expressivo aumento de desempenho no acesso e a duplicação de armazenamento de dados. Porém, mesmo com a máxima segurança e a premissa de ausência de falhas dos discos, o RAID não é eficiente contra eventuais interrupções de energia, pois, uma vez desligado, o erro operacional pode causar danos irreparáveis.
- A implementação do sistema de RAID pode ser realizada por um software, que terá a sua operação feita por meio da controladora de discos e todo o processo executado pelo sistema operacional. Já a instalação do sistema RAID em hardware é capaz de compactar diversos tipos de discos de diferentes fabricantes e configurações. Seu funcionamento é baseado pelo BIOS (Basic Input/Output System) e possui uma integração com o driver do dispositivo capaz de oferecer mais eficiência no diagnóstico de falhas.
- Em RAID 0 (striping), também chamado de distribuição, os dados são divididos e colocados em dois discos separados. Quando ocorre a leitura ou a gravação, a velocidade, teoricamente, duplica.

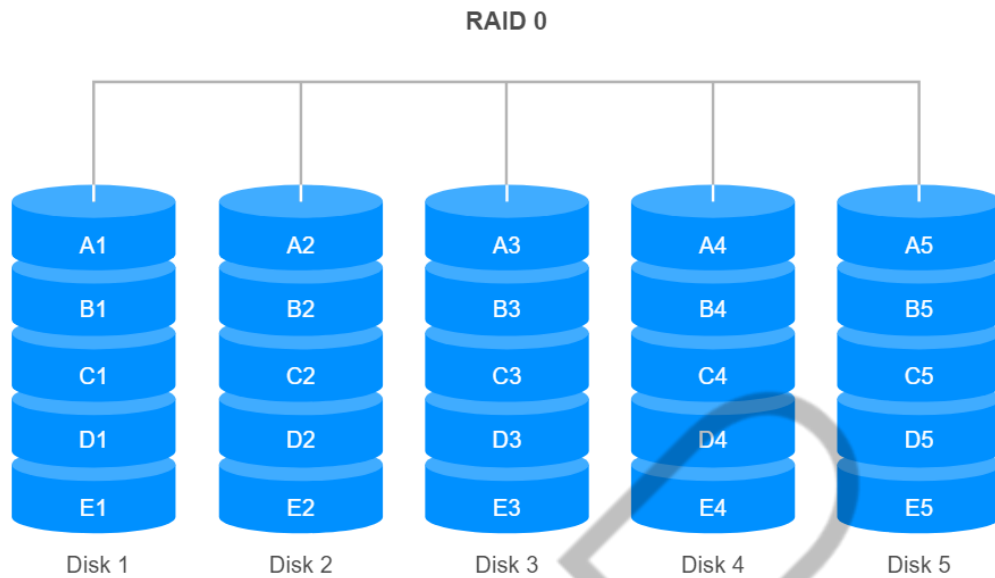


Figura 3.10 – RAID 0
Fonte: Google Imagens (2017)

- O RAID 1 é a implementação denominada mirror, que significa o espelhamento ou a duplicação do sistema em dois discos. Neste caso, os dados são armazenados (copiados) em ambos, garantido mais segurança.

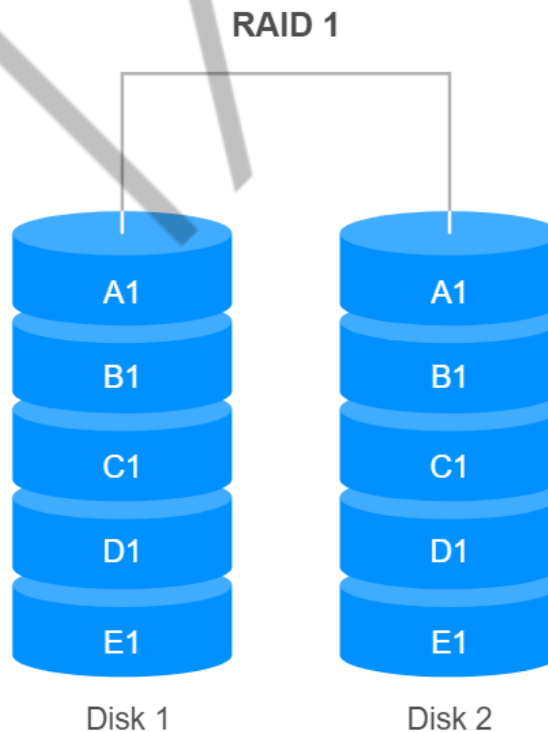


Figura 3.11 – RAID 1
Fonte: Google Imagens (2017)

- O RAID 2 possui uma unidade especialmente desenvolvida para a reparação de eventuais erros, chamado ECC (Error Correcting Code).

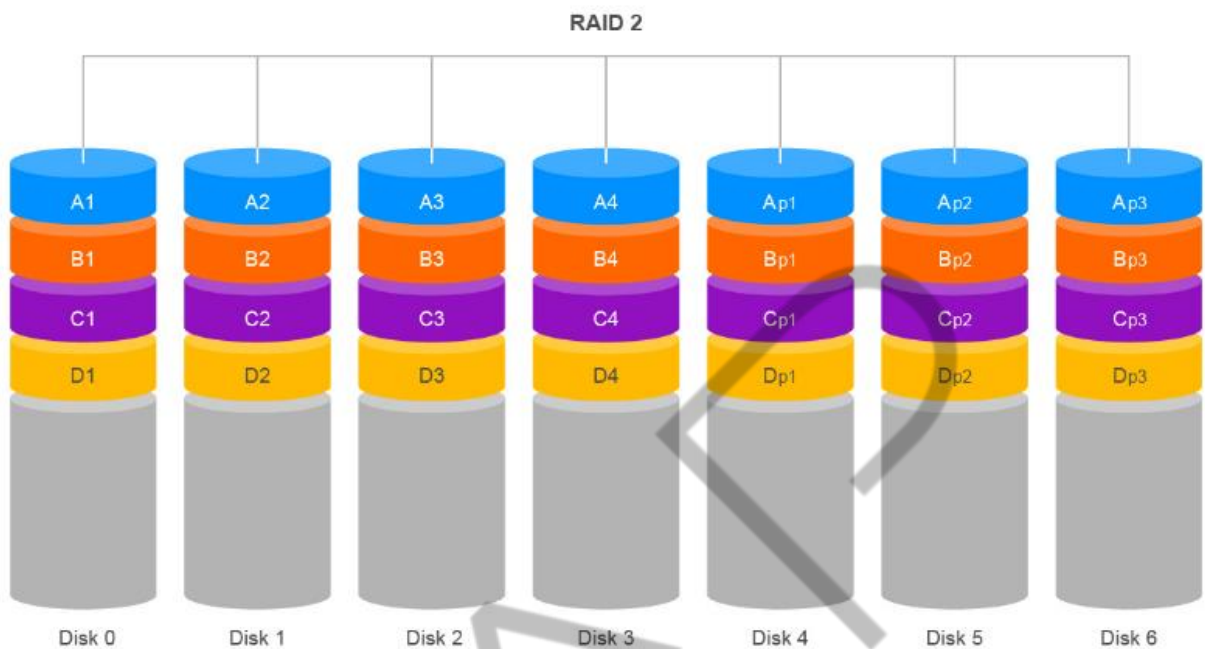


Figura 3.12 – RAID 2
Fonte: Google Imagens (2017)

- O RAID 3 é uma versão melhorada do RAID 2 e oferece em um único bit de paridade a capacidade de detectar o erro e rapidamente corrigi-lo.

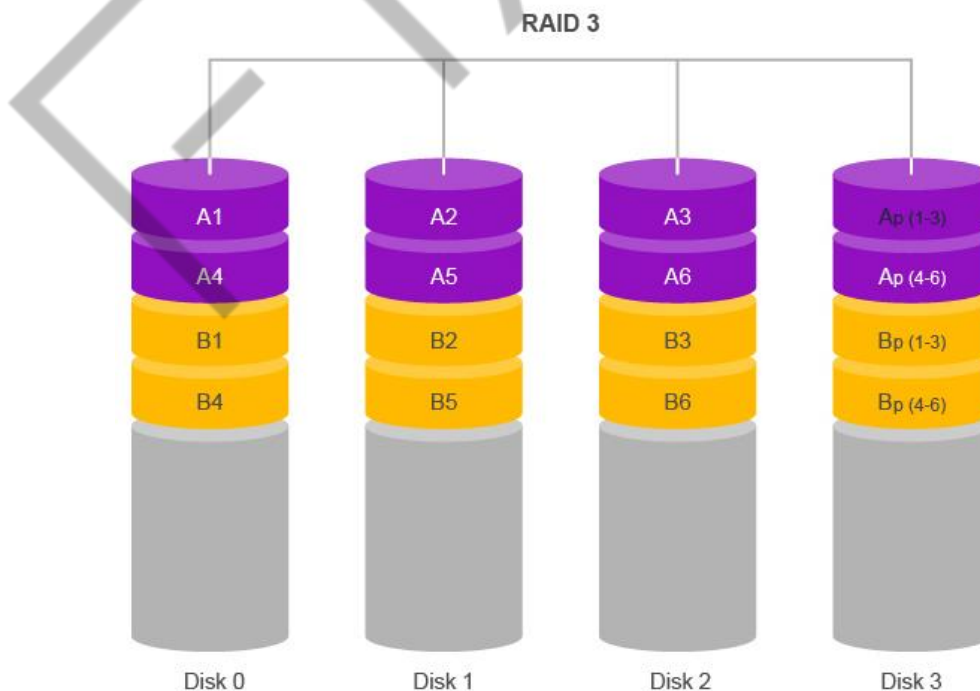


Figura 3.13 – RAID 3
Fonte: Google Imagens (2017)

- O RAID 4 utiliza o processo de recuperação mais eficiente, com componentes denominados arrays, os quais têm a capacidade de criar discos de grande dimensão.

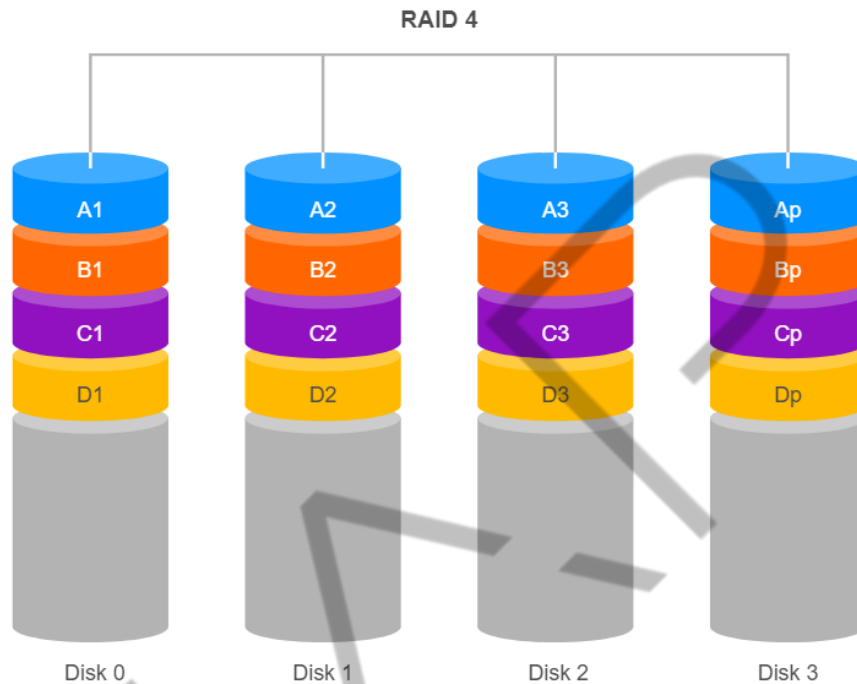


Figura 3.14 – RAID 4
Fonte: Google Imagens (2017)

- O RAID 5 usa um sistema similar ao RAID 4, no entanto, o disco oferece mais desempenho e segurança a possíveis falhas.

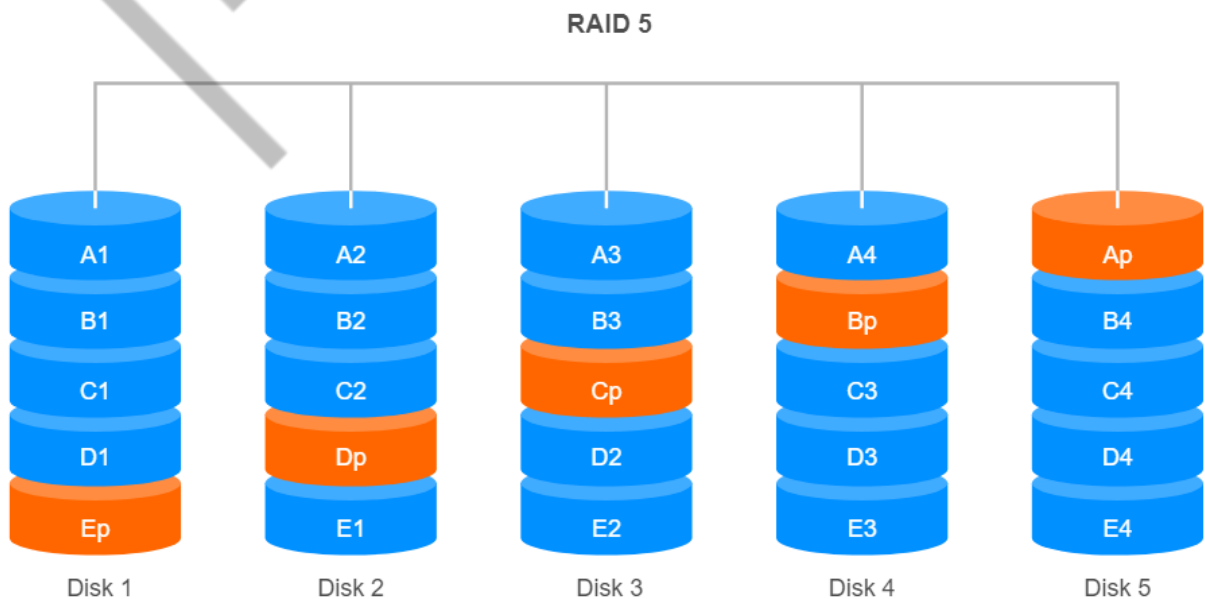


Figura 3.15 – RAID 5

Fonte: Google Imagens (2017)

- Em uma implementação RAID 1+0, os dados são segmentados através de grupos de discos espelhados, isto é, os dados são primeiro espelhados para, depois, serem segmentados, como demonstrado na Figura RAID 10:

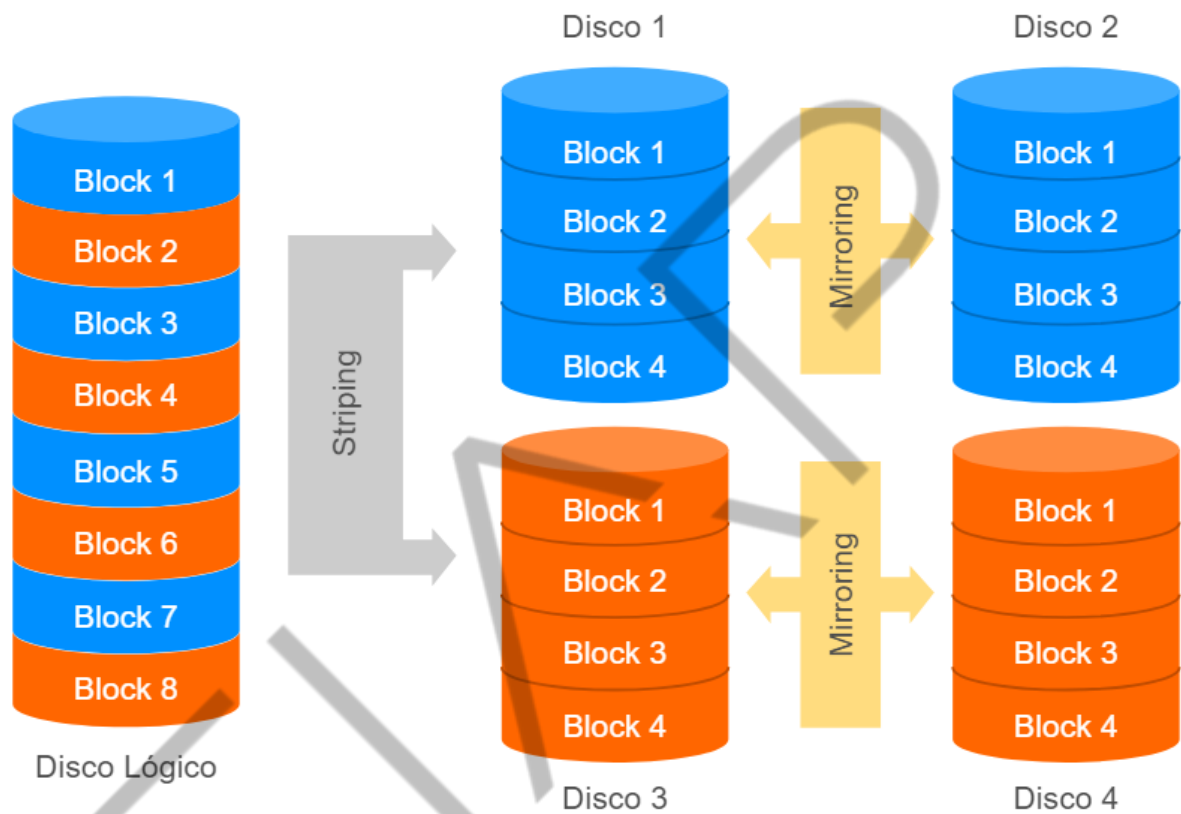


Figura 3.16 – RAID 10
Fonte: Google Imagens (2017)

Há a possibilidade de combinar dois tipos de RAIDs, como o RAID 0+1, em que é utilizado um múltiplo de quatro discos que garantem mais segurança (proveniente do RAID 1, já que dois discos serão cópias dos outros dois) e desempenho (do RAID 0, pois os dados são divididos em dois discos).

Nos dois casos (0+1 ou 1+0), a perda de um único disco não resultará na falha do sistema RAID. A diferença aparece no caso da perda de um segundo disco que, dependendo do disco, colocaria o sistema RAID 0+1 em desvantagem sobre o sistema RAID 1+0. Outra diferença é na velocidade de recuperação, porque, caso ocorra uma falha de disco, no sistema RAID 1+0, será necessário apenas reespelhar um disco, ao contrário do sistema RAID 0+1, em que será preciso espelhar todo um

conjunto segmentado. Portanto, não se esqueça de que RAID 01 é diferente de RAID 10.

3.16.2 NoSQL – Bancos não relacionais para Armazenamento de Dados na web

O mundo nunca lidou com volumes de dados tão grandes, em parte, graças à Web 2.0. No artigo em que define Web, Tim O'Reilly fala sobre a necessidade de se aproveitar a Inteligência Coletiva e sobre dados como diferencial competitivo (O'REILLY, 2005). Não por acaso, muitas inovações na área de gerenciamento de dados vieram de algumas empresas pioneiras da Web 2.0. Encontrando limites nas técnicas e ferramentas disponíveis naquele momento, elas criaram suas próprias soluções, em sua maioria não relacionais. Boa parte da motivação por trás dessas soluções estava ligada a novos requisitos de escalabilidade e disponibilidade, que fizeram com que outros requisitos tidos como indiscutíveis fossem revistos, como a consistência dos dados, por exemplo [VOGELS, 2009 e PRITCHETT, 2008].

3.16.2.1 Bancos de Dados NoSQL

Os bancos de dados NoSQL surgiram como uma solução para a questão da escalabilidade no armazenamento e processamento de grandes volumes de dados na Web 2.0. No início, grandes empresas enfrentando esse tipo de problema criaram suas próprias soluções e publicaram alguns artigos científicos descrevendo diversas soluções ligadas ao gerenciamento de dados distribuído em larga escala, mas sem usar ainda o nome NoSQL [CHANG et al., 2006, DECANDIA et al., 2007, COOPER et al., 2008].

O nome só surgiu alguns anos depois, em 2009, quando algumas novas empresas da Web 2.0 e a comunidade de software livre e código aberto começaram a desenvolver novas opções de bancos de dados, inspiradas nas ideias que apareceram naqueles artigos.

Assim como o termo não relacional, o termo NoSQL não ajuda a definir o que esses bancos são de fato. Além do problema da falta de precisão, esse termo também tem contribuído para uma grande confusão em torno dessa categoria de

bancos de dados, já que, a princípio, a linguagem SQL não é sinônimo de bancos de dados relacionais nem representa as limitações desses bancos de dados. Devido a isso, o termo NOSQL tem sido usado com o significado de “Não apenas SQL” 1 numa tentativa da comunidade de reconhecer a utilidade dos modelos tradicionais e não divergir as discussões.

No fim, NoSQL não define precisamente esses bancos de dados, mas, no geral, cada um deles apresenta a maioria das seguintes características: não relacional, distribuído, de código aberto e escalável horizontalmente, ausência de esquema ou esquema flexível, suporte à replicação nativo e acesso via APIs simples [NOSQL 2010]. Entre os principais fatores que favoreceram seu surgimento, estão a natureza dos dados da web, a importância de se atingir altos graus de paralelismo no processamento de grandes volumes de dados e a distribuição de sistemas em escala global.

3.16.2.2 Natureza dos Dados da web

A web é composta por uma grande quantidade de dados semiestruturados e crus, como as páginas web (cuja estrutura descrita no documento HTML expressa muito pouco sobre o significado do conteúdo do documento) e conteúdo multimídia (imagens, sons e vídeos). Ao reconhecer a natureza particular dos dados, é possível criar soluções otimizadas para eles, ao invés de tentar estruturá-los, como proposto por alguns autores [ATZENI et al. 1997]. A constatação de que os dados na web não são estruturados é um dos fatores que favoreceram o surgimento de tecnologias de gerenciamento de dados diferentes das tradicionais.

3.16.2.3 Bancos de Dados orientados a documentos

Os documentos de bancos de dados orientados a documentos são coleções de atributos e valores, nos quais um atributo pode ser multivalorado. Em geral, os bancos de dados orientados a documento não têm esquema, ou seja, os documentos armazenados não precisam possuir estrutura em comum. Essa característica faz deles boas opções para o armazenamento de dados semiestruturados.

É comum que os bancos de dados orientados a documento usem algum formato de armazenamento capaz de suportar que uma estrutura seja embutida em outra (ou seja, um documento embutido em outro), como o JSON. Embutir um documento em outro, muitas vezes, leva à duplicação de dados. Duplicação de dados, por sua vez, pode criar problemas de consistência no banco de dados, causados por anomalias de atualização e deleção, por exemplo.

A normalização em bancos de dados relacionais visa evitar esses problemas. Mas, considerando que se perde parte das garantias de integridade ao se fazer sharding em um banco de dados relacional, a duplicação de dados em bancos de dados orientados a documentos não é um impeditivo. Além disso, a duplicação de dados facilita a distribuição do sistema, já que a quantidade de nós a serem consultados em uma busca envolvendo várias entidades relacionadas é menor, caso elas estejam próximas – no caso em que todas estejam embutidas umas nas outras, a consulta ocorre em um único nó.

Esse fato aponta para uma característica presente não só nos bancos de dados orientados a documentos, mas em bancos de dados NOSQL em geral, que é a criação de modelos de dados que favoreçam o seu uso para leitura. Bancos de dados populares nessa categoria são o MongoDB e o CouchDB.

3.16.2.4 Escalabilidade

Uma propriedade importante nos produtos NoSQL é o poder de escalar horizontalmente de forma não compartilhada (i.e., replicando e particionando os dados em diferentes servidores). O que permite que um grande volume de operações de leitura/escrita possa ser executado de forma muito eficiente. Além dos conceitos de particionamento e distribuição serem bem definidos, não existe a aplicação do conceito de bases de dados federados em produtos NoSQL.

Em produtos NoSQL, toda a base é considerada uma só, enquanto em bases de dados federados é possível administrar e usar separadamente cada base de dados e, em alguns momentos, utilizar todas as bases como se fossem uma única. O conceito de escalabilidade vertical está relacionado com o uso de vários núcleos/CPU que compartilham memória e discos.

Portanto, mais núcleos e/ou memórias podem ser adicionados para aumentar o desempenho do sistema, porém, essa abordagem é limitada e, normalmente, cara. Já a escalabilidade horizontal se refere à funcionalidade de distribuição de dados e de carga por diversos servidores, sem o compartilhamento de memória ou disco. Esta última abordagem permite o uso de máquinas mais baratas e comuns (i.e., hardware commodity).

3.16.3 Big Data

O termo Big Data é bem amplo e ainda não existe um consenso em sua definição. Porém, Big Data pode ser resumidamente definido como o processamento (eficiente e escalável) analítico de grandes volumes de dados complexos produzidos por (diversas) aplicações. Exemplos de aplicações no contexto Big Data variam bastante, como aplicações científicas e de engenharias, redes sociais, redes de sensores, dados de Web Click, dados médicos e biológicos, transações de comércio eletrônico e financeiro, entre inúmeras outras.

As semelhanças entre os dados desses exemplos de aplicações incluem: grande quantidade de dados distribuídos, características de escalabilidade sob demanda, operações ETL (Extract, Transform, Load [27]) de dados “brutos” (raw) semi ou não estruturados para dados estruturados e a necessidade de extrair conhecimento da grande quantidade de dados.

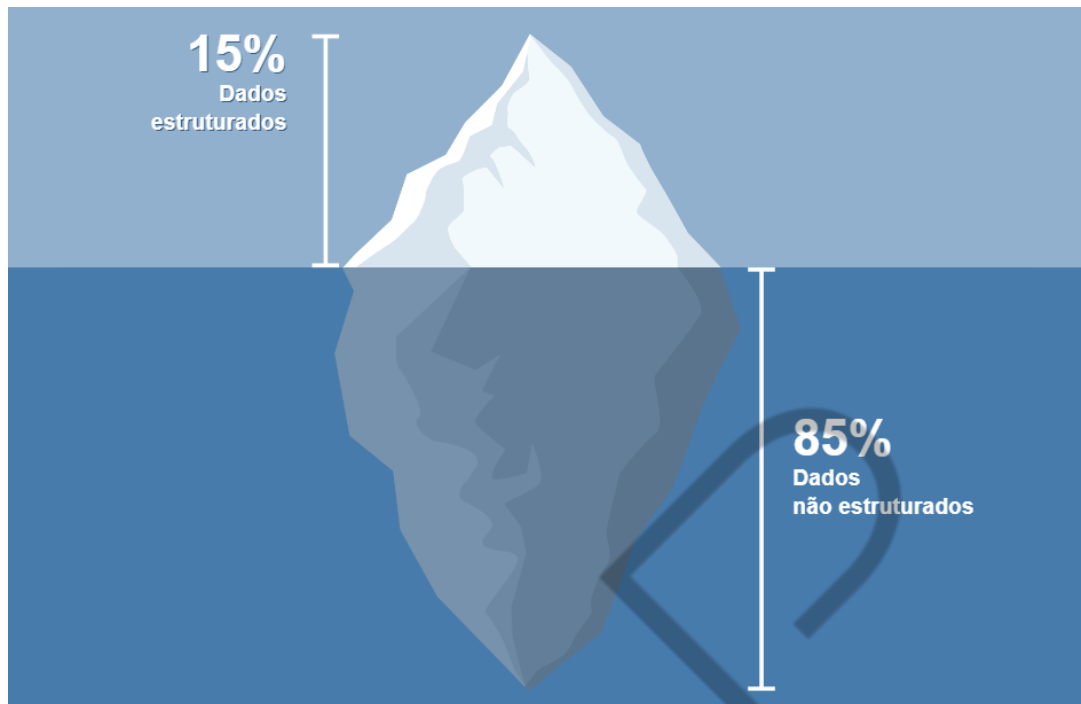


Figura 3.17 – Banco de dados Relacionais x Não Relacionais
Fonte: Google Imagens (2017)

3.16.3.1 Fatores de influência para Big Data

Três fatores influenciaram o grande aumento de volume de dados sendo coletados e armazenados para posterior análise: difusão dos dispositivos de captação de dados, dispositivo com armazenamento na ordem de Terabytes e aumento de velocidade de transmissão nas redes. Os dispositivos de aquisição e os dispositivos de armazenamento de grande escala se difundiram, principalmente, pelo seu barateamento (e.g., redes de sensores, GPS, smartphones), enquanto as redes aumentaram sua velocidade e abrangência geográfica.

Outro fator importante é a facilidade de geração e aquisição de dados gerados digitalmente, como máquinas fotográficas digitais, smartphones, GPS etc. Como consequência, novas demandas estão surgindo, como a demanda por análise de grande volume de dados em tempo real (data analytics), o aumento do detalhamento das informações e as plataformas escaláveis e eficientes de baixo custo.

Atualmente, existe uma grande adoção e difusão de tecnologias NoSQL nos mais diversos domínios de aplicação no contexto de Big Data. Esses domínios envolvem, em sua maioria, quais segmentos os SBGDs tradicionais ainda são

fortemente dominantes, como, por exemplo, instituições financeiras, agências governamentais e comércio de produtos de varejo.

Isso pode ser explicado pela existência de uma demanda muito grande para soluções que tenham alta flexibilidade, escalabilidade, performance e suporte a diferentes modelos de dados complexos. Podemos, basicamente, resumir as características de Big Data em quatro propriedades:

- (a) dados na ordem de dezenas ou centenas de Terabytes;
- (b) poder de crescimento elástico horizontal;
- (c) fácil distribuição dos dados e/ou processamento;
- (d) tipos de dados variados, complexos e/ou semiestruturados.

A característica de manipulação de dados na ordem (ou maior) de Terabytes envolve, entre outros aspectos, o requisito de alto poder computacional de processamento, manipulação e armazenamento de dados. O poder de crescimento elástico está relacionado ao fato de que a quantidade de dados pode variar de alguns Megabytes a várias centenas de Terabytes (e vice-versa) em um espaço de tempo relativamente curto, fazendo com que a estrutura de hardware/software demandada tenha que se adaptar, e. seja alocada/desalocada sob demanda da aplicação.

A distribuição significa que os dados devem ser distribuídos de forma transparente em vários nós de processamento, o que demanda armazenamento e processamento distribuído. E a quarta característica se refere à adoção de modelos mais apropriados, flexíveis e eficientes para o armazenamento de tipos de dados complexos, variados e semiestruturados.

Vale ressaltar que o modelo relacional tradicional não é o mais adequado para as propriedades citadas, pois não possui suficiente flexibilidade para o armazenamento de dados e a evolução do modelo de dados.

3.16.3.2 Como o Big Data muda o modelo de negócios?

Como exemplo, os bancos estão mais atentos a conhecer e responder rapidamente aos acontecimentos da vida dos indivíduos. Por exemplo, se um cliente

fizer um depósito significativo, o banco pode responder, no dia seguinte, com uma oferta específica para um depósito mais rentável, ou, se o cliente solicitar um grande adiantamento do seu cartão de crédito, a entidade financeira pode oferecer uma via mais econômica para obter dinheiro físico, da próxima vez.

3.16.3.3 E a ética?

Não podemos deixar de lembrar que Big Data também traz à tona novos problemas éticos.

O que empresas e governos fazem com tantos dados privados?

Que informação pode ser deduzida a partir dos dados?

É famosa a história de um pai que descobriu, por meio da propaganda direcionada de uma rede norte-americana de lojas de varejo, que sua filha adolescente estava grávida.

A filha comprou diferentes produtos para teste de gravidez, fato que havia sido detectado pelos algoritmos da empresa, a qual, logo, começou a enviar informativos e propagandas de assuntos relacionados para ela, o que, eventualmente, foi percebido pelo pai.

“Se as forças do Império de Star Wars usassem Big Data, elas venceriam.”

Daniel Solove, CEO da TeachPrivacy, aponta que uma análise simples de informações conhecidas revelaria onde Luke Skywalker estava vivendo. Uma coleção mais eficiente e a agregação de registros teriam localizado os droides imediatamente. Um cruzamento de dados simples teria revelado que Ben Kenobi foi realmente Obi Wan Kenobi, e uma busca de registros de nascimento teria revelado que a Princesa Leia era irmã de Luke.

EMSE

REFERÊNCIAS

BALIERO, Ricardo. **Banco de dados**. Rio de Janeiro: SESES, 2015.

DIANA, M. de.; GEROSA, M. A. NOSQL na Web 2.0: Um Estudo Comparativo de Bancos Não-Relacionais para Armazenamento de Dados na Web 2.0. Disponível em: <https://www.ime.usp.br/~mdediana/nosql_wtddb10.pdf>. Acesso em: 11 abr. 2020.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados: fundamentos e aplicações**. 6. ed. São Paulo: Pearson Addison-Wesley, 2011.

GARTNER. **Why Data and Analytics are key to digital transformation**. Disponível em: <<https://www.gartner.com/smarterwithgartner/why-data-and-analytics-are-key-to-digital-transformation/>>. Acesso em: 12 abr. 2020.

HINDENBURGO, F. Pires. **O surgimento dos primeiros computadores**. Disponível em: <<http://www.educacaopublica.rj.gov.br/biblioteca/geografia/0016.html>>. Acesso em: 1º set. 2014.

IBM. **Edgard Frank Codd**. Disponível em: <http://www-03.ibm.com/ibm/history/exhibits/builders/builders_codd.html>. Acesso em: 1º set. 2014.

REZENDE, Ricardo. **Histórico dos banco de dados**. Devmedia. Disponível em: <<https://www.devmedia.com.br/a-historia-dos-banco-de-dados/1678>>. Acesso em: 11 abr. 2020.

SILBERSCHATZ, A; KORTH, H. F.; SUDARSHAN, S. **Sistema de Banco de Dados**. 5. ed. Rio de Janeiro: Campus, 2006.

TERRA, J. C. C. **Gestão do conhecimento: o grande desafio empresarial [s.d]**. Disponível em: <https://edisciplinas.usp.br/pluginfile.php/263603/mod_folder/content/0/Gestao%20do%20Conhecimento%20Desafio%20Empresarial.pdf?forcedownload=1>. Acesso em: 25 abr. 2019.

TIINSIDE. **Iniciativas de big data e analytics são fundamentais na transformação digital, defende Gartner**. Disponível em: <<https://tiinside.com.br/17/05/2019/iniciativas-de-big-data-e-analytics-sao-fundamentais-na-transformacao-digital-defende-gartner/>>. Acesso em: 12 abr. 2020.