

Lógica de Banco de Dados

Prof. Leandro Colevati

Dados e Informação

Os dados são medições, números, valores, enfim, constatações.

Você pode definir um dado como um conteúdo quantificado. Se eu digo para você “5 pessoas leram este post”, o que isso significa? Exatamente: que 5 pessoas leram o post, não há nenhum conhecimento além do que foi dito que não seja suposição.

Sendo assim, os dados são caracterizados por conteúdos que por si só não agregam nenhum valor ou trazem algum conhecimento. Os dados, sozinhos, não são capazes, por exemplo, de orientar a gestão à tomada de decisão.

Você teve 200 pacientes esse mês no laboratório: o que isso, isoladamente, significa? Pois é. Nada. Aqui entra a importância da informação.

Dados e Informação

Informação é o real valor da coleta. Quando o dado é tratado de tal forma que consegue transmitir uma mensagem, ele já se tornou uma informação. Em outras palavras, organizar e ordenar os dados é o que faz a informação existir.

Quando você coloca o dado sob uma certa ótica e dentro de um contexto determinado, ele passa a significar alguma coisa e se converte em informação. Ou seja, a informação é o processamento dos dados. A tecnologia, como os sistemas de gestão, são altamente capazes de agrupar dados para oferecer informações estratégicas.

Voltando ao exemplo do laboratório: você teve 200 pacientes este mês. Mês passado, teve 180. Houve um aumento de 20 pacientes. Entende como isso caracteriza uma informação? Se dos 200 pacientes, 50 são novos, significa que houve 30 pacientes do mês passado que não retornaram.

Nesse meio de caminho, é válido comentar que houve uma campanha de vacinação, por exemplo, voltada à terceira idade. Quantos desses pacientes vieram graças à ação? Esse tipo de informação — proveniente dos dados que o sistema de gestão pode agrupar — pode colaborar para medir resultados e aplicar recursos no que realmente faz a diferença.

Dados

No mundo dos dados, há três formas de classificação, baseadas na maneira como eles são estruturados, classificadas em: estruturados, semiestruturados e não estruturados.

Para que se entenda melhor cada uma delas, traçando uma lógica de comparação entre ambas, podemos partir de uma avaliação por:

- Características específicas;
- Complexidade para análise;
- Representatividade em nosso dia-a-dia;
- Abrangência de uso.

Dados Estruturados

Os dados estruturados são aqueles organizados e representados com uma estrutura rígida, a qual foi previamente planejada para armazená-los, por exemplo um banco de dados, que é a representação mais típica e comum de dados estruturados.

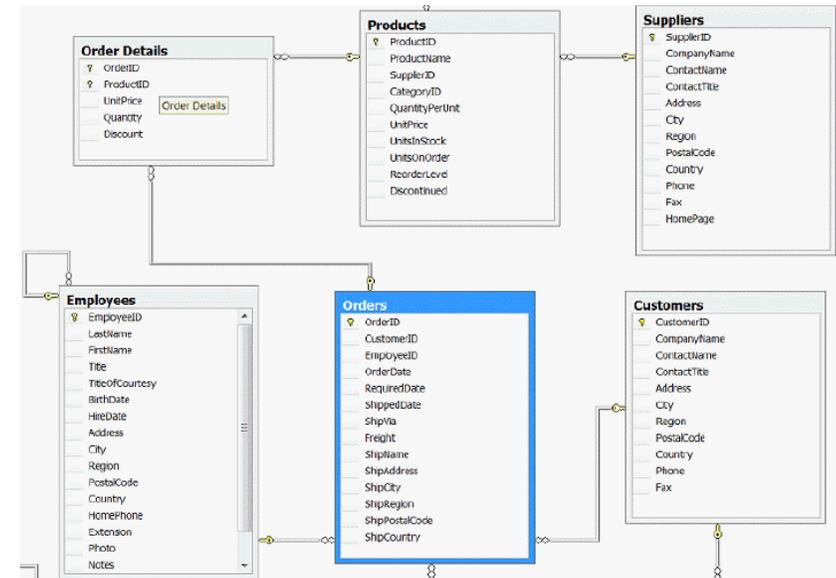
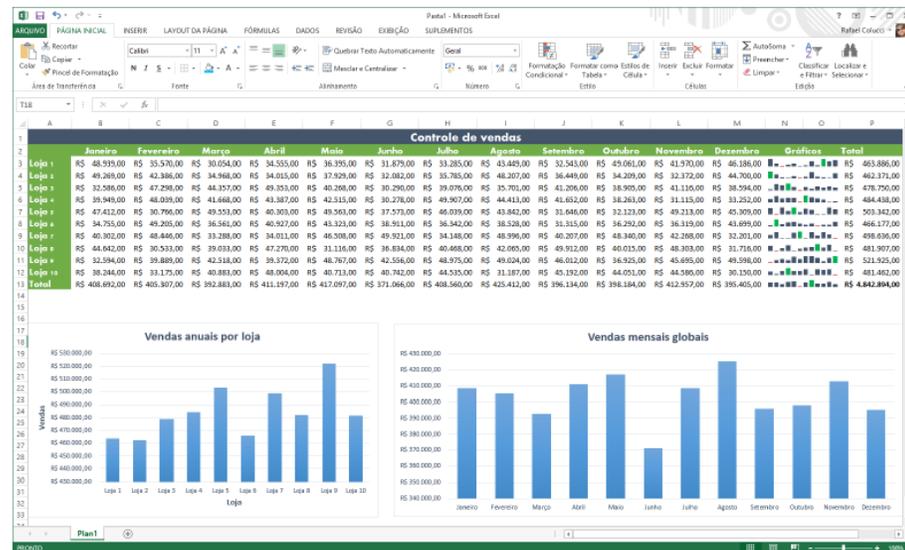
Em um banco de dados, os dados são estruturados conforme a definição de um esquema, que define como as tabelas e suas respectivas linhas e colunas serão armazenadas. Podemos conceituar o esquema de um banco de dados como sendo uma descrição sobre uma organização, ou sobre o minimundo que se deseja representar, definindo quais dados que serão armazenados.

Podem ser representados como uma tabela, por exemplo, dentro de um banco de dados, considera-se uma estrutura bem definida e que sua alteração não é tão simples, por isso dizemos que os dados estruturados são rígidos e não-flexíveis.

Dados Estruturados

Id_Ciente	Nome	Data_Nascimento	Salario	CPF
1	Judas Iscariotes	2066-06-06 00:00:00.000	10	NULL
2	Marcos	1975-01-07 00:00:00.000	1626.438	NULL
3	Simão	1991-12-18 00:00:00.000	2584.97	NULL
4	Pedro	1986-11-20 00:00:00.000	987.52	NULL
5	Paulo	1974-08-04 00:00:00.000	6259.14	NULL
6	José	1979-09-01 00:00:00.000	5272.13	NULL
7	Bartolomeu	1975-05-28 00:00:00.000	9238.11	11111111111

Id	Dt_Atualizacao	Login	Hostname	Operacao	Id_Ciente	Nome	Data_Nascimento	Salario	CPF
1	2016-07-17 10:45:23.937	VM-DBA@droeu.resende	VM-DBA	INSERT	8	Bartolomeu	1975-05-28 00:00:00.000	6158,74	NULL
2	2016-07-17 10:45:23.943	VM-DBA@droeu.resende	VM-DBA	UPDATE	8	Bartolomeu	1975-05-28 00:00:00.000	9238,11	NULL
3	2016-07-17 10:45:23.943	VM-DBA@droeu.resende	VM-DBA	DELETE	3	André	1962-11-11 00:00:00.000	7151,45	NULL
4	2016-07-17 10:45:23.950	VM-DBA@droeu.resende	VM-DBA	DELETE	3	Marcos	1975-01-07 00:00:00.000	1626,438	NULL



Dados Semiestruturados

Seus dados nem sempre podem ser estruturados ou não estruturados. Dados semiestruturados são outra categoria entre esses dois que está parcialmente estruturada. Esses dados são definidos como semiestruturados.

Os dados semiestruturados são um tipo de dado que possui algumas características

consistentes e definidas e não se confina a uma estrutura rígida como a necessária para bancos de dados relacionais. Propriedades organizacionais, como metadados ou tags de semântica, são usadas com dados semiestruturados para torná-los mais gerenciáveis; no entanto, ainda contém alguma variabilidade e inconsistência.

```
<bibliography>  
  <book> <title> Foundations... </title>  
    <author> Abiteboul </author>  
    <author> Hull </author>  
    <author> Vianu </author>  
    <publisher> Addison Wesley </  
  publisher>  
    <year> 1995 </year>  
  </book>  
  ...  
</bibliography>
```

Dados Não Estruturados

Ao contrário dos dados estruturados, os dados não estruturados possuem uma estrutura totalmente inversa e não podem ser organizados em tabelas.

Esse tipo de dados não tem uma estrutura bem definida, não tem um padrão pré-estabelecido. São dados flexíveis e dinâmicos, podendo ser compostos por diversos elementos diferentes dentro um todo.

Um exemplo simples de dados não estruturados são os dados de redes sociais. Esse é um típico exemplo, pelo grande número de textos, imagens, vídeos e diversos outros que são criados diariamente a partir do uso das tecnologias.

Na verdade, a maior parte dos dados gerados no mundo todo são dados não estruturados, tendo aproximadamente o percentual de 80%. Esses dados têm uma complexidade um pouco maior para análise, já que são informações de difícil processamento e recuperação, pois não contam com componentes necessários para a sua identificação. Porém, com a disponibilização de tecnologias que nos ajudam nesse quesito, essa complexidade da análise de dados tem diminuído.

Dados Não Estruturados

A maioria dos dados que trocamos todos os dias **são dados não estruturados**, sem padrão como: e-mails, fotos, vídeos, áudios.

Todos eles são indicadores que categorizam informações sobre pessoas e torna possível levar uma oferta certa para aquele consumidor.

Empresas gigantes como Facebook vivem de acumular dados, sobretudo dados não estruturados.

Eles cobram de quem quiser anunciar por eles **a estruturação desses dados**, desta forma ele reúne todos os dados registrados **através da movimentação dos usuários** dentro da rede social para categorizar esses usuários e cobra para que pessoas atinjam seus públicos.

É por isso que os dados que circulam são **ativos valiosos para as empresas**, porque permitem conhecer melhor os clientes e com isso fazer a melhor oferta, no momento mais propício, o que aumenta muito as chances de conversão.

Noções Preliminares

Um **Banco de Dados** ou Base de Dados (BD) é uma coleção de **dados** relacionados entre si.

Por dados entenda-se fatos conhecidos que podem ser armazenados e possuem um significado implícito.

Exemplo (simples): coleção de nomes das pessoas que você conhece, juntamente com o endereço e telefone de cada uma delas (podem ser armazenados em uma agenda ou em um computador).

Outro exemplo (um pouco mais complexo): coleção de todas as fotos recolhidas ao longo dos anos pelo programa espacial americano.

Noções Preliminares

Um BD possui as seguintes propriedades:

1. representa algum aspecto do mundo real;
2. Correponde a um coleção logicamente coerente de dados com um significado inerente;
3. é projetado, construído e preenchido (instanciado) com dados para um propósito específico. A ele está relacionado um grupo de usuários e aplicações de interesse a esses usuários.

Em outras palavras, um BD possui alguma fonte de onde os dados são obtidos, um grau de iteração com eventos do mundo real e um grupo de pessoas interessadas no seu conteúdo.

Noções Preliminares

Exemplo (extremamente simples) de um BD para a universidade:

Aluno	Nome	RGA	Disciplinas
	João	2000.0213.001-1	111 e 453
	José	2001.0213.001-2	128

Disciplina	Nome	Código	Horas	Sala
	Banco de Dados	111	64	70516
	Compiladores	453	64	70517
	Teoria dos Grafos	128	64	70516

Essa base de dados é organizada em dois arquivos distintos, cada um contendo registros do mesmo tipo. Para defini-la, necessitamos especificar a estrutura dos registros em cada um dos arquivos. Ou seja, necessitamos definir os elementos a serem armazenados em cada um dos registros. Cada registro de estudante, por exemplo, guarda o nome, código e disciplinas sendo cursadas pelo estudante. Além disso, o tipo de cada um dos elementos da estrutura precisa ser definido (nome: seqüência de caracteres; RGA: inteiro sem sinal; Disciplina: inteiro sem sinal).

Sistemas de Arquivos x Banco de Dados

Em um sistema baseado em arquivos, o que temos é um coleção de arquivos cujos conteúdos são dados a serem consultados ou modificados de acordo com as necessidades do usuário detentor desses arquivos. Essa manipulação é feita por meio de programas específicos, desenvolvidos com o formato dos arquivos em mente.

Exemplo: a secretária do departamento tem a sua disposição um arquivo com os nomes dos estudantes e as disciplinas sendo cursadas por cada um deles. Já a tesouraria do departamento precisaria de um arquivo com os nomes dos estudantes e informações sobre pagamento de bolsas.

Apesar da validade dessa abordagem, ela possui um problema: a redundância de dados. A principal consequência desse problema está no gasto de espaço e na necessidade de esforços consideráveis para manter consistência.

Sistemas de Arquivos x Banco de Dados

Diferente de um sistema baseado em arquivos, um sistema baseado em uma base de dados possui um único repositório de dados. Isso elimina o problema da redundância. Outros aspectos de um sistema baseado em uma base de dados são:

1. Um sistema baseado em banco de dados inclui uma descrição detalhada de sua base. Essa descrição corresponde ao **catálogo** do banco de dados, que guarda, dentre outras coisas, informações sobre a estrutura de cada arquivo, e sobre o tipo de cada item de dado. O catálogo permite que um mesmo conjunto de programas acesse diferentes tipos de bases de dados, ao contrário dos programas que manipulam os arquivos em um sistema baseado em arquivos.
2. Por meio do catálogo, qualquer mudança na base de dados é prontamente reconhecida pelos programas que as acessam (independência entre dados e programa). Além disso, em alguns sistemas de banco de dados, existe também uma independência entre o programa e operação.

Sistemas de Arquivos x Banco de Dados

3. Em um sistema baseado em banco de dados, certas operações sobre os dados devem ser feitas de forma conjunta e indivisível com o intuito de preservar a consistência da base de dados. Isso mesmo na presença de falhas no equipamento de armazenamento ou na comunicação com a base de dados (atomicidade).
4. Um sistema baseado em banco de dados inclui programas de controle de concorrência. Ou seja, programas que asseguram um acesso adequado a uma base de dados quando vários usuários a acessam ao mesmo tempo (consistência).
5. Um sistema baseado em banco de dados é suficientemente robusto para se recuperar de falhas de *hardware* ou *software* (durabilidade).

Sistemas de Arquivos x Banco de Dados

A construção e manutenção de um sistema que atende aos requisitos acima requer o auxílio de programa bem elaborados. Daí temos as seguintes definições:

Sistema Gerenciador de Base de Dados: coleção de programas que permitem a criação e manutenção de qualquer base de dados, independentemente da aplicação pretendida.

Sistema de Banco de dados: Sistema gerenciador juntamente com a base de dados.

Modelo de Dados

Um modelo de dados é um conjunto de conceitos lógicos (objetos, suas propriedades e o inter-relacionamento entre eles) compreensíveis aos usuários finais do sistema e usados para descrever a estrutura de uma base de dados. Por estrutura de uma base de dados entenda-se os tipos de dados, relacionamentos e restrições associadas aos dados. Vários modelos de dados definem também um conjunto de operações para recuperação e modificação da base de dados.

De acordo com os tipos de conceitos que fornecem para descrever a estrutura da base de dados, os modelos de dados podem ser classificados em:

Modelo de Dados

- **Conceituais:** provê conceitos muito próximos àqueles percebidos pelos usuários finais do sistema. Esses modelos utilizam conceitos entidades, atributos, e relacionamentos. Uma entidade representa objetos ou conceitos do mundo real (exemplos: funcionários, vãos). Um atributo representa alguma propriedade de interesse relacionada a uma entidade (nome e salário dos funcionários, por exemplo). Um relacionamento entre duas (ou mais entidades) representa uma interação entre entidades (relacionamento entre funcionários e projetos, por exemplo);
- **Lógicos:** provê conceitos que podem ser entendidos pelos usuários finais do sistema e que estão próximos da forma como os dados são fisicamente armazenados. São os mais utilizados pelos SGBD atuais e incluem o modelo de dados relacional, de rede e hierárquicos;
- **Físicos:** provê conceito que descrevem os detalhes de como os dados encontram-se armazenados no computador. Representam informações como formato dos registros, ordenação dos registros e caminhos de acesso.

Esquema x Instância

Em qualquer modelo de dados é importante distinguir entre a descrição da base de dados e base de dados propriamente dita. A descrição de uma base de dados é chamada de **esquema** de uma base de dados. Esse esquema é especificado durante o projeto da base de dados e não costuma sofrer mudanças freqüentes. A forma de visualização de um esquema é chamada Diagrama do Esquema. Muitos modelos de dados têm certas convenções para, diagramaticamente, mostrar esquemas especificados no modelo. Tomando como exemplo o BD (extremamente simples) para a universidade, temos o seguinte diagrama do esquema para essa base:

Aluno

Nome	RGA	Disciplinas
------	-----	-------------

Disciplina

Nome	Código	Horas	Sala
------	--------	-------	------

Os dados em uma base de dados, esses sim costumam mudar freqüentemente. Os dados armazenados em um banco em um determinado instante é chamado de estado ou **instância** da base de dados.

Linguagens de Banco de Dados

- Linguagem de Definição de Dados (*DDL - Data Definition Language*): utilizada pelo administrador e projetistas da base de dados para definição dos esquemas. O SGBD possui um compilador para que traduz descrições em DDL em descrições dos esquemas em catálogo;
- Linguagem de Manipulação de Dados (*DML - Data Manipulation Language*): utilizada pelos usuários para manipulação (recuperação, inserção, remoção e modificação) dos dados.

Sistema Gerenciador de Banco de Dados

Um sistema gerenciador de base de dados é composto de vários módulos distintos:

- gerenciador de dados armazenados: controla o acesso às informações do sistemas que estão armazenadas no disco. Esse módulo utiliza-se de serviços básicos do sistema operacional para transferir dados do disco para a memória principal. Uma vez transferidos, esse dados podem então ser usados por outros módulos do sistema (ou por aplicações do usuário).
- compilador DDL : processa as descrições do esquema (escritas em DDL) e a armazena no catálogo do sistema (que se encontra no disco rígido).
- processador *run-time*: manipula acessos à base de dados em tempo de execução. Ele recebe operações de atualizações ou recuperações de dados.
- compilador de consultas: manipula consultas em mais alto nível. Ele compila ou interpreta consultas e então gera chamadas ao processador *run-time*.
- O pré-compilador extrai comandos DML de um programa de aplicação escrito em uma linguagem de alto nível qualquer. Esses comandos são então enviados ao compilador DML, que os transformam em códigos para acesso à base. O resto do código é enviado ao compilador da linguagem. Os códigos para os comandos DML e o resto do programa são *linkados*, gerando um código executável que inclui chamadas ao processador *run-time*.