



CWI SOFTWARE

Modulo 2 | banco de dados

Crescer 2017-1

André Luís Nunes

Mestrado em Computação Aplicada, UNISINOS (2017)

Bacharel em Sistema de Informações, FACCAT (2008)

Desde dez/2007 na CWI Software.

Gestor Técnico

Núcleo de Tecnologia

Principais Funções:

- ❑ Programa Crescer;
- ❑ Aplicações internas;
- ❑ Apoio em projetos/arquiteturas

De 2001 a 2007 Calçados Azaléia S.A.

- Estagiário;
- Programador;
- Analista de Sistema/Programador;
- DBA.



André Luís Nunes



E-mail andre.nunes@cwi.com.br

Telephone (51)98175-1020

Skype live:andre_298

@local São Leopoldo, 4º andar

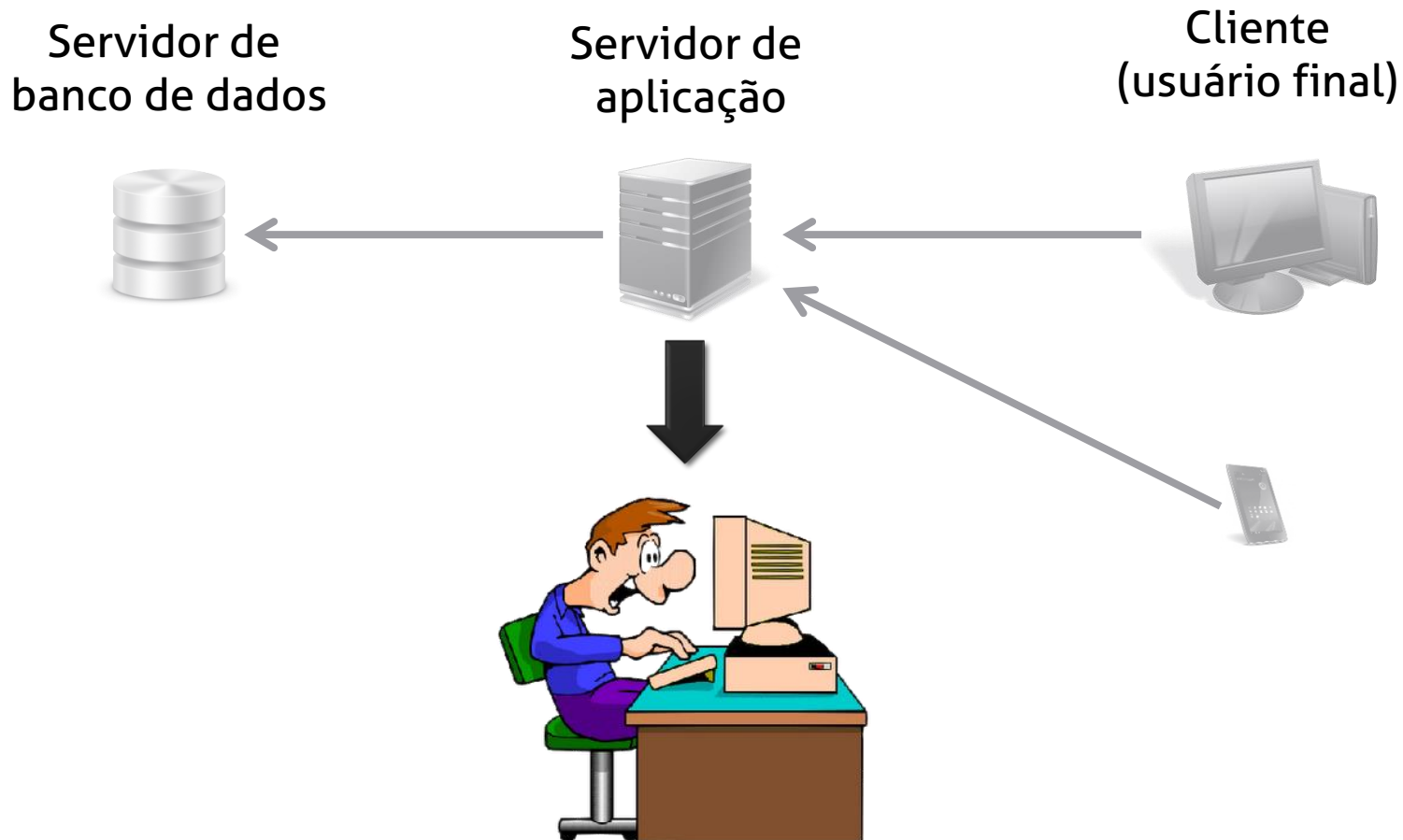
Tópicos

- Definição de banco de dados;
- Conceitos
- Exemplos
- Prática

Como é utilizado?



Como é utilizado?



Banco de dados



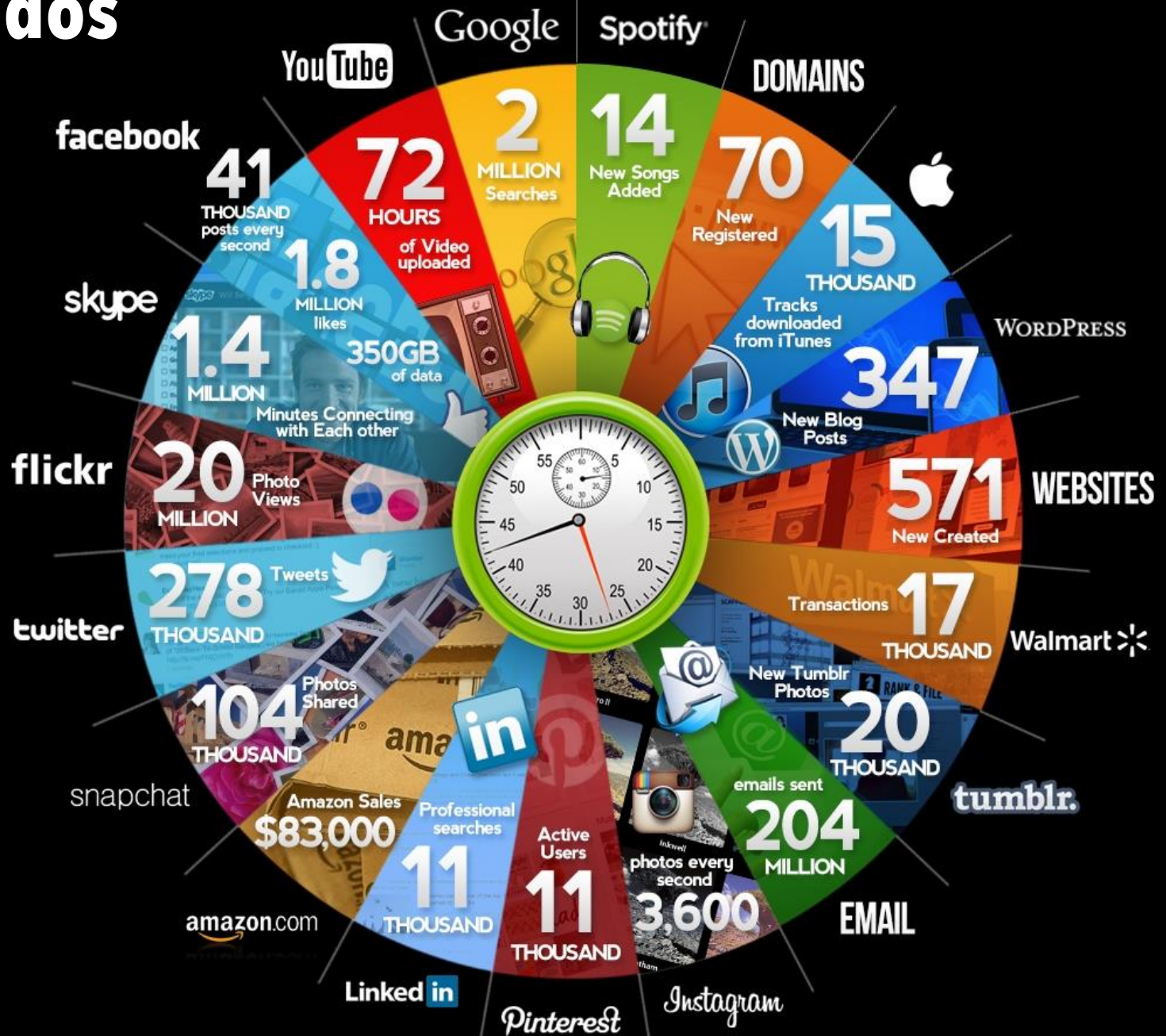
estatísticas



Alguns números

—

60 segundos



Banco de dados | projeção



Volume de dados

- » Em 2009: **800 exabytes** (1 exabyte = 1 bi GB)
- » Em 2020*: **35000 exabytes** (35 zettabytes)

Aumento de 44x em 10 anos

Banco de dados | unidade de medida

Quanto é 1 zettabyte ?

- » 1.000.000.000.000 gigabytes
- » 1.000.000.000.000 terabytes
- » 1.000.000.000.000 petabytes
- » 1.000.000.000.000 exabytes
- » 1.000.000.000.000 zettabyte

Banco de dados | unidade de medida

Quanto é 1 zettabyte ?

» 1.000.000.000.000 gigabytes

» 1.000.000.000.000 terabytes

» 1.000.000.000.000 petabytes

» 1.000.000.000.000 exabytes

» 1.000.000.000.000 zettabyte

Yotabytes	= 1000 ZB
Brontobytes	= 1000 YB
Geopbyte	= 1000 BB

Banco de dados | evolução

Anos 1990-2000

1992 HP C3013A rompe os 2 GB	2,1 GB
1997 IBM (GMR)	16,8 GB
2006 Seagate Barracuda	750 GB
2007 Hitachi GST	1 TB
2011 /12 Seagate, Western e Toshiba	4 TB

Banco de dados | evolução

2013 - ADATA DashDrive Air AE800

- » Dispositivo de armazenamento **wireless**
- » Capacidade de 500 GB
- » Cerca de R\$ 500,00



Banco de dados | evolução

2013 - novas tecnologias

- » Western Digital faz experimentos com o uso de **gás hélio** nos discos (10 anos de pesquisa).
- » Aumento de 50% a capacidade
- » Economia de 23% em consumo de energia
- » 38% mais leve



6 TB

Banco de dados | evolução

Tendências

2014

- » Seagate's SMR permitirá discos com a capacidade de Shingled Magnetic Recording

5 TB

2020

- » Seagate's HAMR permitirá discos com a capacidade de Head Assisted Magnetic Recording

20 TB

Banco de dados | novos databases

Dados não estruturados ...



Banco de dados | novos databases



Banco de dados | novos databases

Explosão de dados até final de 2014

Não estruturados

» Fotos, redes sociais, dados geográficos

78 EB

Estrutura menos rígida, menos ordenada e com maior inter-relação

Estruturados

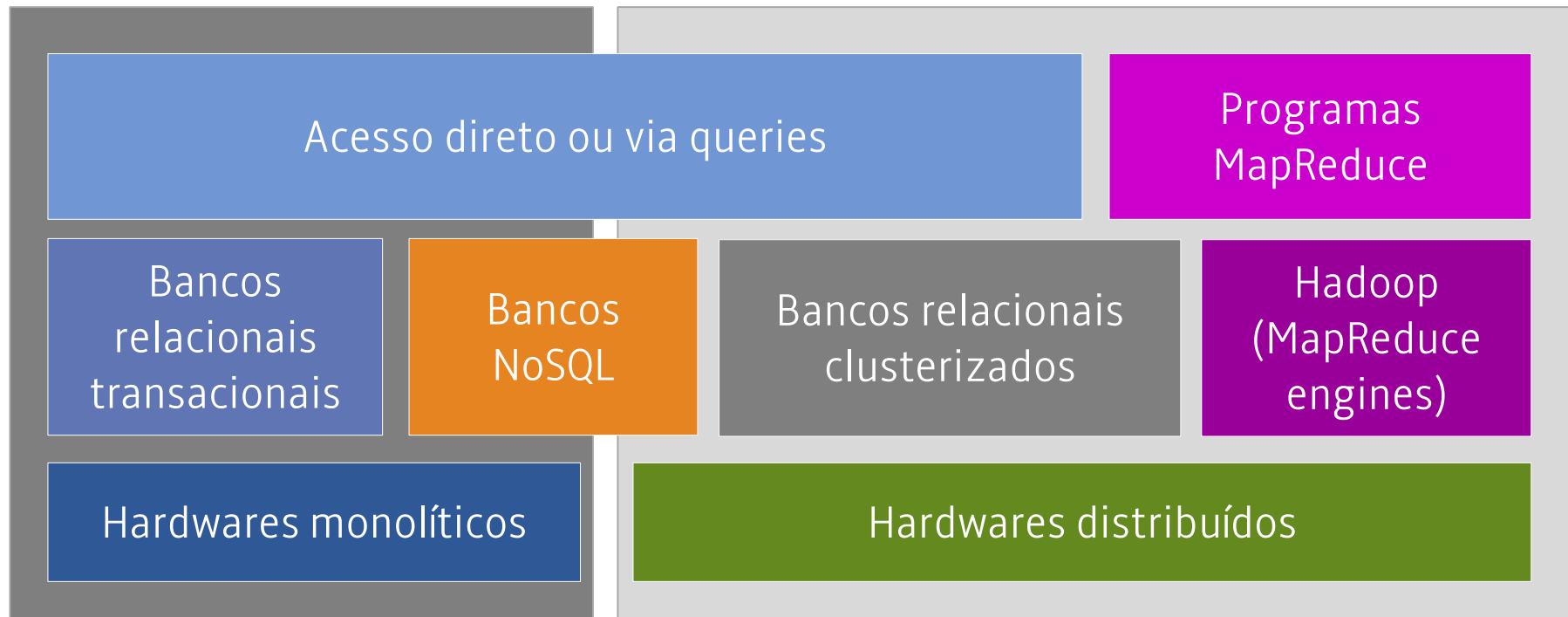
» Sistemas atuais (tradicionais)

15 EB

Banco de dados | novos databases

Distribuição dos tipos de ambiente

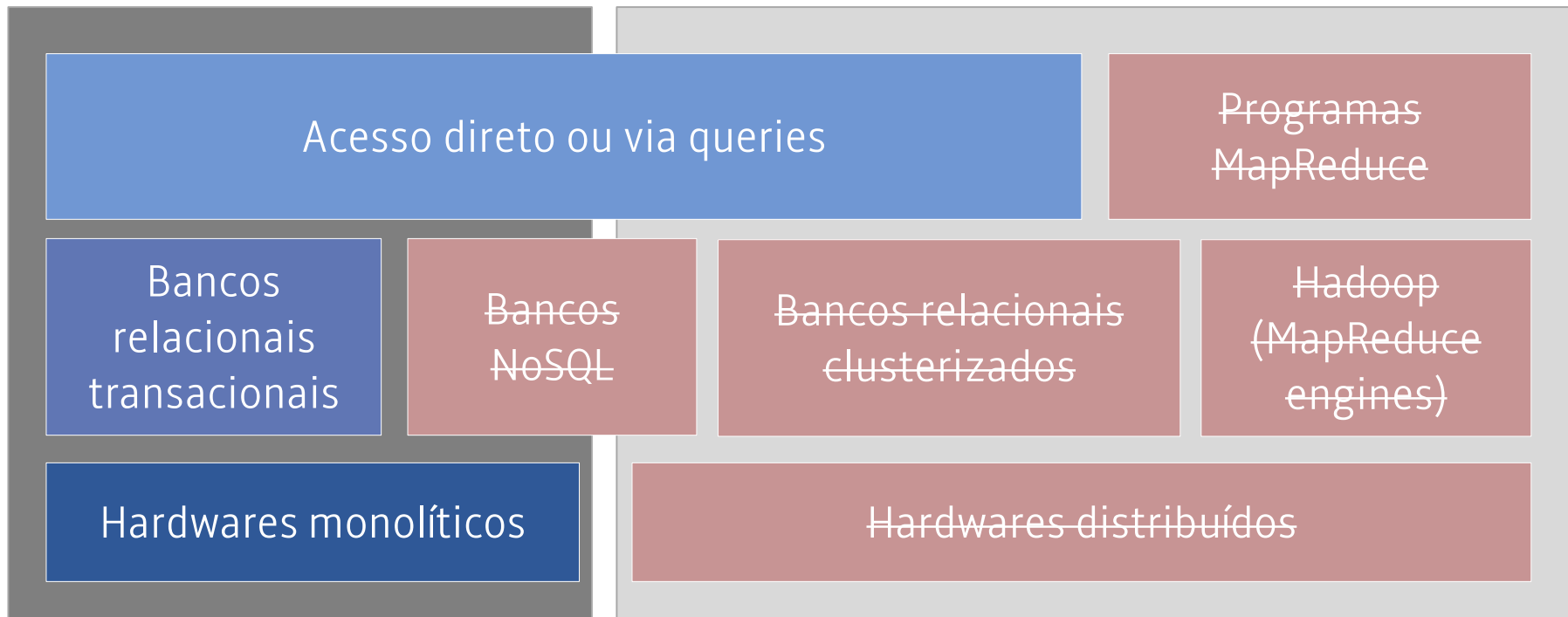
» Novas tecnologias, estruturas, arquiteturas...



Banco de dados | novos databases

Distribuição dos tipos de ambiente

» O que iremos trabalhar diretamente no Crescer

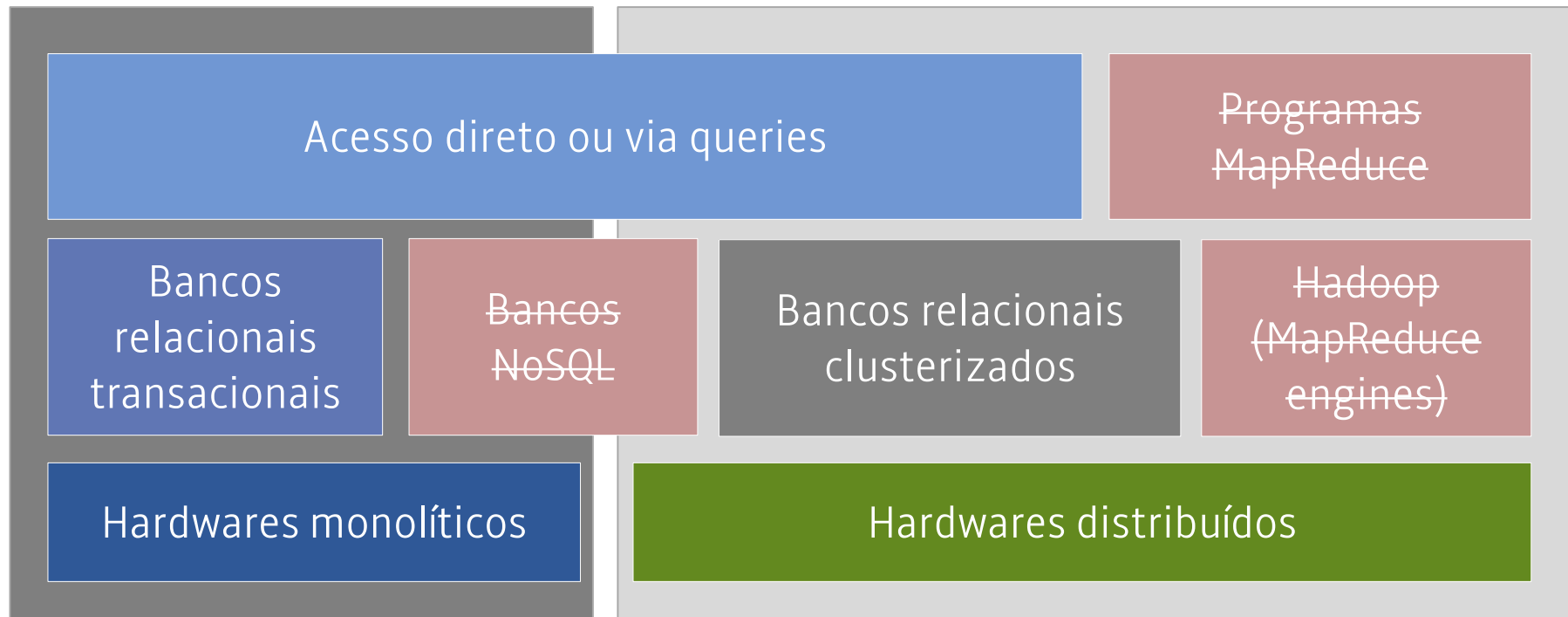


ambiente centralizador

Banco de dados | novos databases

Distribuição dos tipos de ambiente

» O que poderemos trabalhar



ambiente centralizador

Banco de dados | análise

Novas perspectivas de análise de dados

» Em em todas as áreas praticamente



Banco de dados | análise

Exemplo

» Pesquisa com genomas

- 42.000 amostras de sangue
- 50.000 genes por amostra
- 1 - 2 mi de “partes” de informação genética por amostra

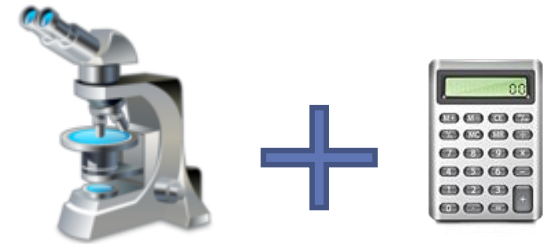


Banco de dados | análise

Exemplo

» Pesquisa com genomas

- 42.000 amostras de sangue
- 50.000 genes por amostra
- 1 - 2 mi de “partes” de informação genética por amostra



4,2 quintilhão de possibilidades

Banco de dados | otimização

Exemplo



» Pesquisa do Google

- 10 bilhões de páginas
- Média de tamanho de cada página 20 KB
- $10 \text{ bi} \times 20\text{KB} = 200 \text{ TB}$
- Taxa de leitura em disco = 50 MB/seg

Banco de dados | otimização

Exemplo



» Pesquisa do Google

- 10 bilhões de páginas
- Média de tamanho de cada página 20 KB
- $10 \text{ bi} \times 20\text{KB} = 200 \text{ TB}$
- Taxa de leitura em disco = 50 MB/seg

4 milhões de segundos (>46 dias)

BIG DATA IN FORMULA ONE



Formula One cars generate **terabytes of data** during a race. Dozens of engineers at the track and as far away as the U.K. comb over the data during a race in near real-time, looking for any adjustment that could **win or lose** a race.

RACE TEAMS COMBINED TO GENERATE **243 TERABYTES** OF DATA FROM THEIR VEHICLES AT THE 2014 U.S. GRAND PRIX IN AUSTIN, TX.

BIG DATA ?

243 TERABYTES OF DATA COMPARED TO ...



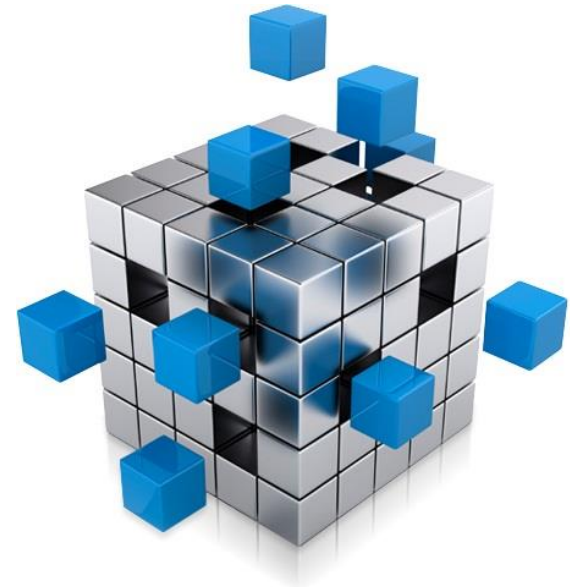
EQUIPPED WITH **HUNDREDS OF SENSORS**, F1 CARS PROVIDE A STREAM OF DATA THAT'S ANALYZED **THOUSANDS OF MILES** AWAY IN NEAR REAL-TIME

Banco de dados | visões

Novas visões são necessárias

- Considerando redes sociais
- Localização geográfica / GPS
- Informações de sensores físicos

MUITAS informações



Voltando...

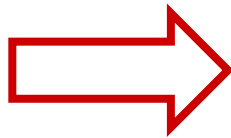


...no Crescer

É simples trabalhar com esta diversidade?

Bancos mais utilizados

- Oracle
- SQL Server

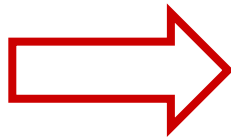


Representam a maioria dos bancos nos clientes.

É simples trabalhar com esta diversidade?

Bancos mais utilizados

- Oracle
- SQL Server



Representam a maioria dos bancos nos clientes.

Novos tipos de bancos

- noSQL (mongoDB, Cassandra, Neo4J, HBase...)

Novos bancos substituirão os relacionais?

Resposta simples: NÃO!

Novos bancos substituirão os relacionais?

Resposta simples: NÃO!

***Sistemas já desenvolvidos dificilmente
serão migrados para outra arquitetura***

Novos bancos substituirão os relacionais?

Resposta simples: NÃO!

*Sistemas já desenvolvidos dificilmente
serão migrados para outra arquitetura*

Resposta mais completa: Em alguns casos sim!

Novos bancos substituirão os relacionais?

Resposta simples: NÃO!

*Sistemas já desenvolvidos dificilmente
serão migrados para outra arquitetura*

Resposta mais completa: Em alguns casos sim!

*Porém, é preciso conhecer as características de cada um.
Mas não há “**BALA DE PRATA**” em bancos de dados!*

Conceitos

- DADOS;
- INFORMAÇÃO;
- BANCO DE DADOS;
- SISTEMA GERENCIADOR DE BANCO DE DADOS (SGBD)

Conceitos

- **Dados:** são códigos, nomes, siglas que constituem a matéria-prima para a informação. São os dados não tratados.

Exemplo:

Codigo	Ocorrencia	ValorX	ValorY
1	20120227	10,342.00	15,024.00
2	20120301	4,952.00	4,400.00
3	20000130	13,409.23	13,429.20
15	20101111	9,340.45	10,345.97

Tabela 1: lista de registros (dados apenas).

Conceitos

- **Informações:** são os dados tratados, é o significado dos dados. É o resultado do processamento dos dados.

Exemplo:

Tabela de movimentação de Caixa

identificador	Data (aaaa mm dd)	Valor Entrada (em reais)	Valor Saída (em reais)
1	2012 02 27	10,342.00	15,024.00
2	2012 03 01	4,952.00	4,400.00
3	2000 01 30	13,409.23	13,429.20
15	2010 11 11	9,340.45	10,345.97

Tabela 1: Deixou de ser apenas “números”.

Conceitos

- Banco de dados
- Este termo é normalmente utilizado para se referir a um Sistema Gerenciador de Banco de Dados - **SGBD**. Porém existem diferenças nos termos, vejamos a seguir:
- **Base de Dados (BD)**: um SGBD pode possuir várias, é uma coleção de dados interligados, representando informações sobre um domínio específico.

SGBD x GA

- As diferenças entre o sistema tradicional de gerenciamento de dados, denominado Gerenciador de Arquivos (GA) e um SGBD são muitas.
- Um SGBD deve obedecer integralmente as seis regras. Caso contrário, será um GA.

SGBD

- Seis regras:
 - ❑ Autocontenção;
 - ❑ Independência dos dados;
 - ❑ Abstração dos dados;
 - ❑ Visões;
 - ❑ Transações;
 - ❑ Acesso automático.

SGBD – regra 1

- Autocontenção

Um SGBD não contém apenas os dados em si, mas armazena completamente toda a descrição dos dados, seus relacionamentos e formas de acesso.

Conhecido como METADATA (ou metadados).

Em um GA, os programas aplicativos declaram estruturas ou geram os relacionamentos entre os arquivos.

Quando é necessário a definir a forma do registro em no código do programa, não é um SGBD.

SGBD – regra 2

- Independência dos dados

Quando as aplicações estiverem realmente imunes a mudanças na estrutura de armazenamento ou na estratégia de acesso aos dados, podemos dizer que esta regra foi atingida.

Nenhuma alteração na definição (DDL) deve ser realizada a nível de aplicação.

Portanto, nenhuma definição dos dados deverá estar contida nos programas da aplicação. Quando é preciso criar uma nova forma de acesso, um novo índice, se precisar alterar o código de seu aplicativo, não estamos lidando com um SGBD.

SGBD – regra 3

- Abstração dos dados

Em um SGBD real é fornecida ao usuário somente uma **representação conceitual dos dados**, o que não inclui maiores detalhes sobre sua forma de armazenamento real. Informações como regras, do chamado “Modelo de Dados” ficam abstratas à aplicação.

Exemplo: se ao inserir um valor definido como único (CPF de um estudante) o SGBD deve responder automaticamente que já existe o valor.

Assim como a relação entre as tabelas (integridade).

SGBD – regra 4

- Visões

Um SGBD deve permitir que cada usuário visualize os dados de forma diferente daquela existente previamente no Banco de Dados.

Uma visão consiste de um subconjunto de dados do Banco de Dados, necessariamente derivados dos existentes no Banco de Dados, porém estes **não deverão estar explicitamente armazenados**. Portanto, toda vez que somos obrigados a replicar uma estrutura, para fins de acesso de forma diferenciada por outros aplicativos, não estamos lidando com um SGBD.

SGBD – regra 5

- Transações

Um SGBD deve gerenciar completamente a integridade referencial definida em seu esquema, sem precisar em tempo algum, do auxílio do programa aplicativo.

Desta forma exige-se que o banco de dados tenha ao menos uma instrução que permita a gravação (**COMMIT**) de uma série modificações simultâneas e uma instrução capaz de cancelar (**ROLLBACK**) um série modificações.

Exemplo: controle da quantidade de produtos em estoque.

SGBD – regra 6

- Acesso automático

A responsabilidade de evitar um *dead-lock* deve ser do SGBD, se esta situação precisar ser controlada pela aplicação isto caracteriza um GA ou qualquer outra estrutura de armazenamento de dados diferente de um SGBD.

Exemplo:

Horario	Usuario Maria	Usuario Julia
8:00	Conecta no sistema.	
8:05	Altera nome do Cliente 10 para JOAO.	Conecta no sistema.
8:10		Altera o valor do Produto 125 para R\$ 13.45
8:15	Altera o valor do Produto 125 para R\$ 14.19 (usuario Maria está esperando Julia concluir a transação)	
8:20		Altera o nome do Cliente 10 para JOÃO. (usuario Julia estaria esperando Maria concluir a transação)
	LOCK	DEAD LOCK

SGBD – regra 6

- Exemplo de dead-lock

Um exemplo prático:



SGBD – regra 6

- Exemplo de dead-lock

Um exemplo prático:

CWI



SGBD – regra 6

- Exemplo de dead-lock

Um exemplo prático:



SGBD (definição)

- Sistema Gerenciador de Banco de Dados

É um software projetado para auxiliar a **definição**, **construção** e **manipulação** de um banco de dados.

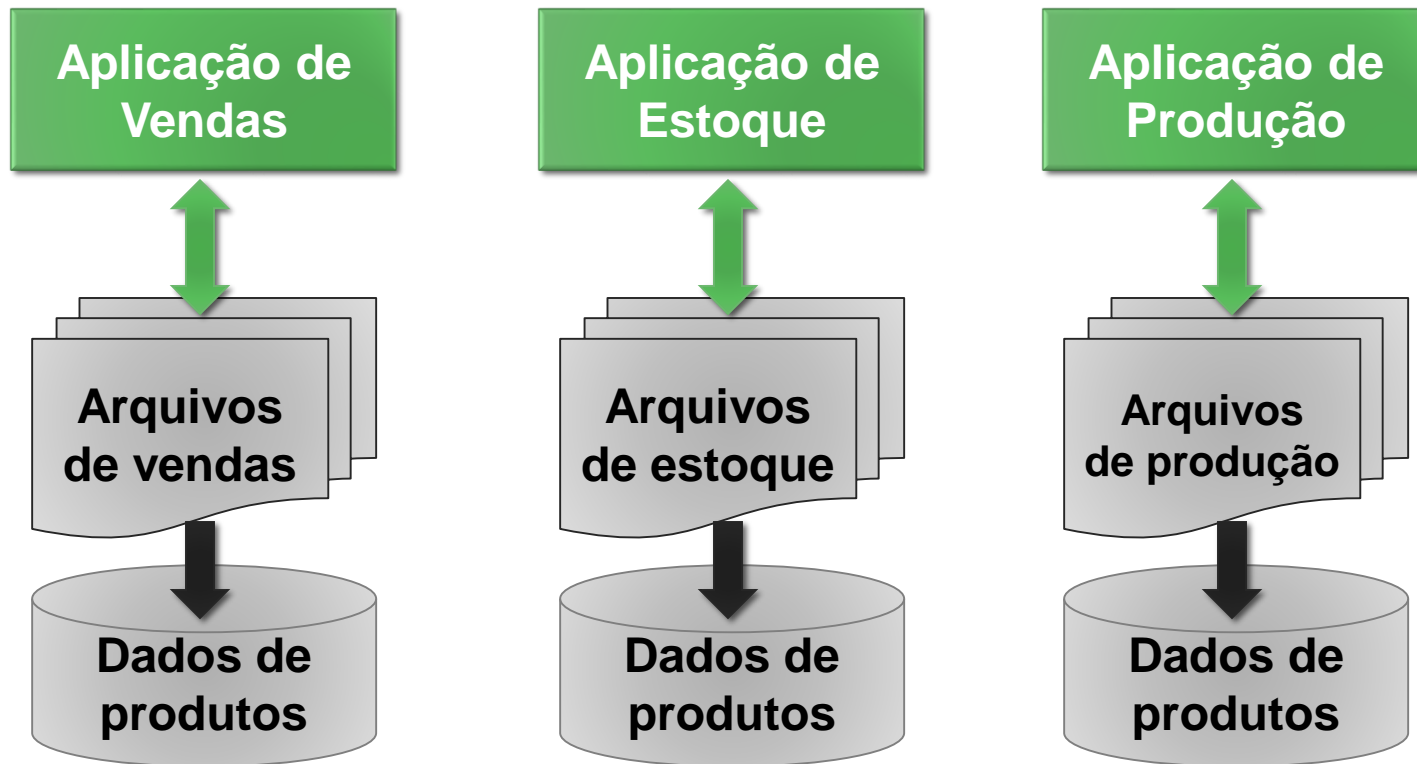
Em inglês: *DataBase Management System - DBMS*

Mais utilizados:

- ☐ Oracle
- ☐ SQL Server
- ☐ DB2
- ☐ MySQL
- ☐ PostgreSQL

SGBD - importância

- Armazenamento de dados de uma indústria baseado em arquivos (GA).



Desvantagens sem SGBD

- Redundância de dados não controlada

- ❑ Os dados das 3 aplicações podem não estar integradas.

- ❑ Isto gera um problema sério, INCONSISTÊNCIA DOS DADOS.

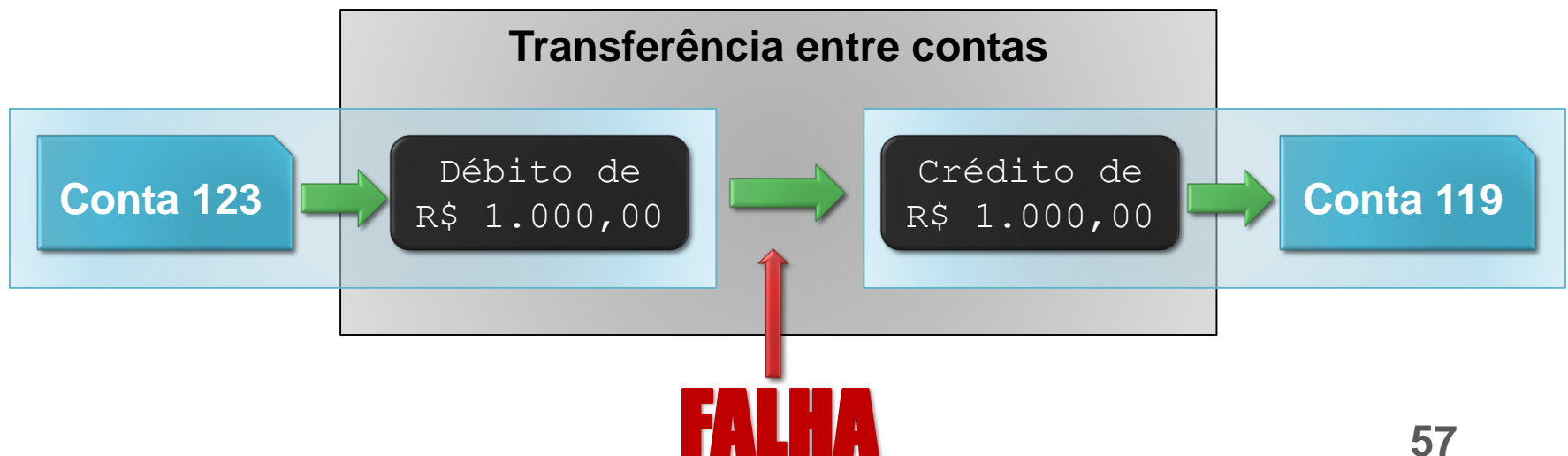
- **Aplicação Vendas** : COURO XYZ – **CÓDIGO** 10
 - **Aplicação Estoque** : COURO X.YZ – **CÓDIGO** 351
 - **Aplicação Produção:** Couro Marrom XYZ – **CÓDIGO** 1293

- ❑ Como cruzar estas informações para geração de relatórios?

- ❑ A redundância controlada ?

Desvantagens sem SGBD

- Proteção contra falhas
 - ❑ Problemas de **transações** (devem ser atômicas)
 - Um conjunto de operações precisa ser tratado de UMA única OPERAÇÃO ATÔMICA.
 - ❑ Perda de informações.



Desvantagens sem SGBD

- Problema de acesso concorrente
 - ❑ Interações entre atualizações concorrentes podem resultar em inconsistência dos dados.
 - ❑ Exemplo:
 - Dois vendedores estão vendendo o mesmo produto

Vendedor A	Vendedor B
Lê Quantidade em estoque (10)	
$Ea = \text{Saldo}(10) - \text{Qtde_VendidaA}(6)$	
	Lê Quantidade em estoque (10)
	$Eb = \text{Saldo}(10) - \text{Qtde_VendidaB}(8)$
Grava (Saldo= Ea)	
	Grava (Saldo= Eb)

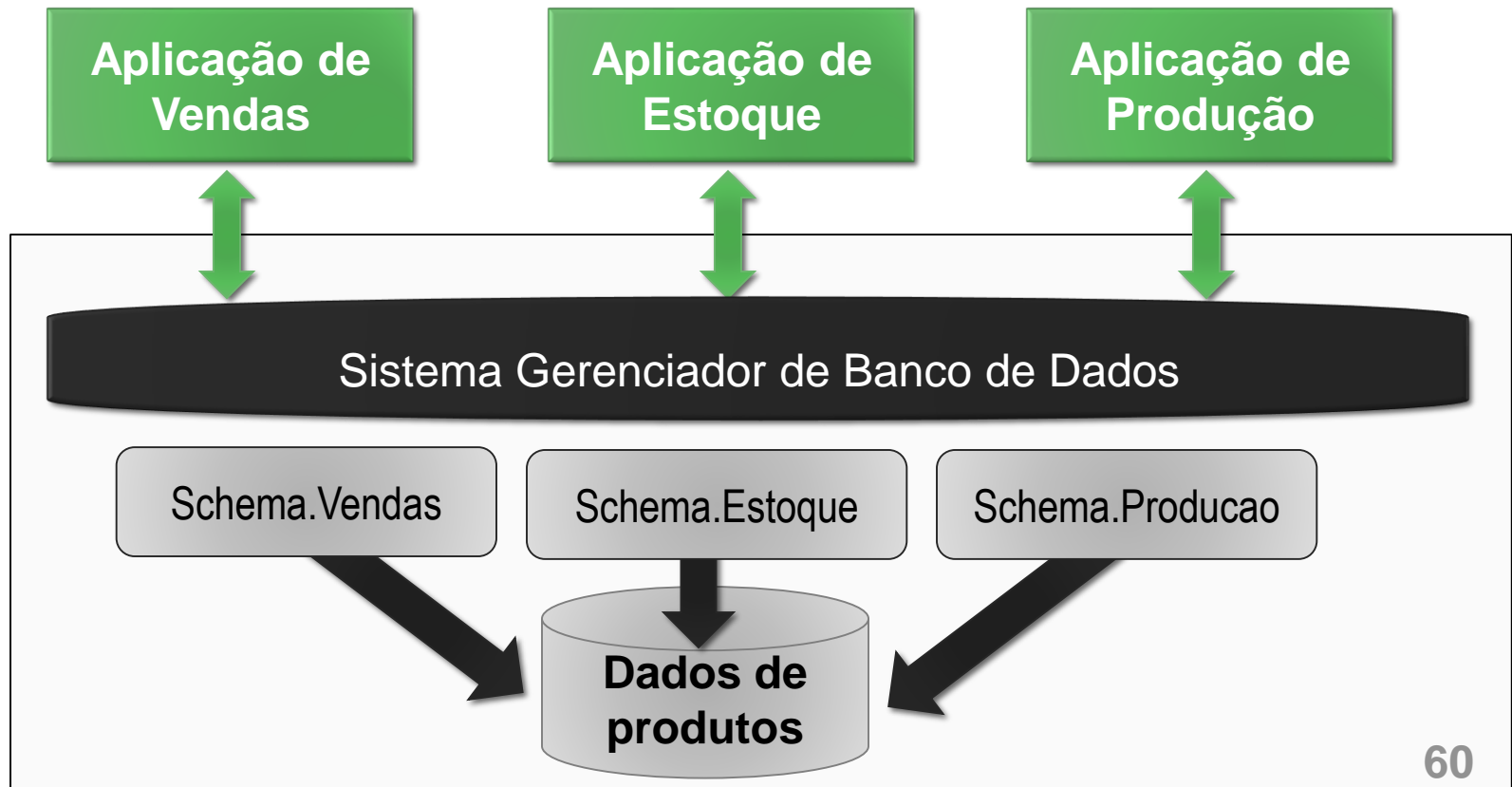
Desvantagens sem SGBD

- Problema de segurança

- ❑ Nem todos os usuários e aplicações devem/podem acessar as mesmas informações.
- ❑ Cada aplicação e cargo deve ter acesso a um tipo de informação.
 - Exemplo: Dados da Folha de Pagamento não podem ser acessadas por funcionários da área Industrial.
- ❑ Com SGBD é possível controlar o acesso por schema/database e por usuário.

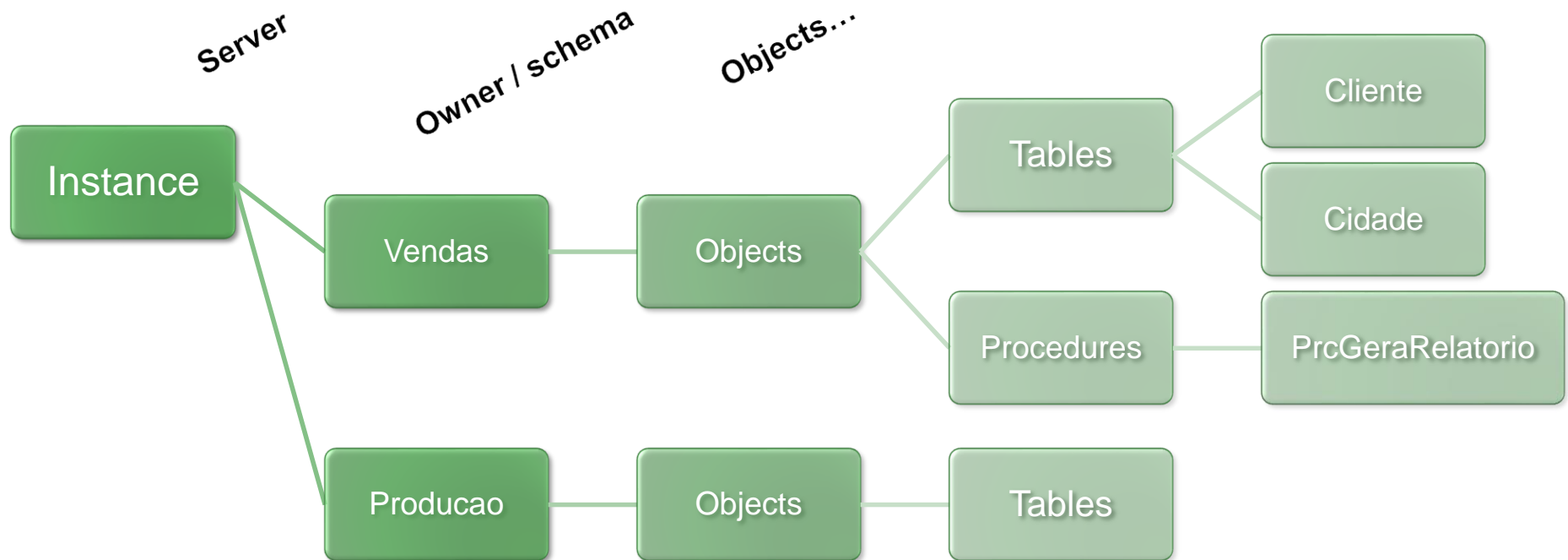
SGBD - exemplo

- Armazenamento de dados de uma indústria baseado em SGBD.



SGBD – arquitetura lógica

- Exemplo de uma visão da arquitetura lógica de um banco de dados (SQL Server).



Quando não precisa de SGBD

- Aplicações muito *simples*
 - Manipula poucos dados operacionais
 - Dados podem ser mantidos em poucos arquivos (ou 1 apenas)
- Aplicações *CPU-Bound*
 - Processamento pesado e pouca gerência de dados (sem necessidade de armazenamento).
 - Aplicações científicas, engenharia e outras aplicações que demandam de alta performance computacional.

→ Nestes casos pode surgir algo como SQLite.

Linguagem

- A linguagem utilizada para desenvolvimento e administração de um banco de dados é SQL (Structured Query Language), foi criado na década de 70 pela IBM.
- O padrão para a linguagem foi criado pela American National Standards Institute (ANSI) em 1986 e ISO em 1987.
- O SQL foi revisto em 1992 e a esta versão foi dado o nome de SQL-92. Foi revisto novamente em 1999 e 2003 para se tornar SQL:1999 (SQL3) e SQL:2003, respectivamente.
- Cada fabricante de SGBD costuma ter variações em cima do SQL ANSI.


Tipos de comandos (SQL)

- **DDL**: Data Definition Language
 - ❑ Comandos que permitem definir a estrutura de objetos.
- **DML**: Data Manipulation Language
 - ❑ Comandos que permitem alterar os dados de uma tabela.
 - ❑ Inserir, alterar, excluir e consultar registros.
- **DCL**: Data Control Language
 - ❑ Comandos que permitem conceder ou revogar permissões à determinados objetos de uma base dados.

SGBD – abstração dos dados

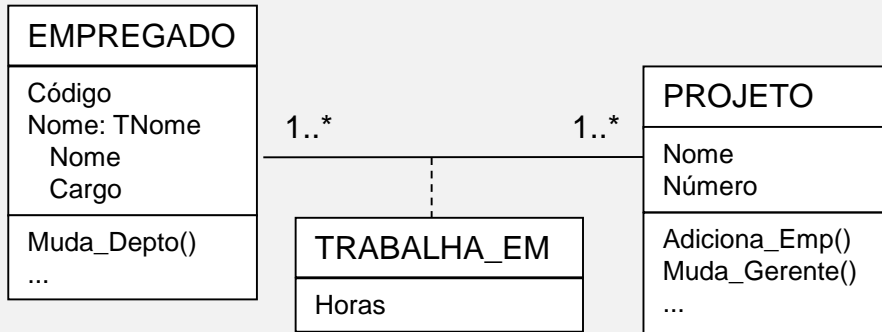
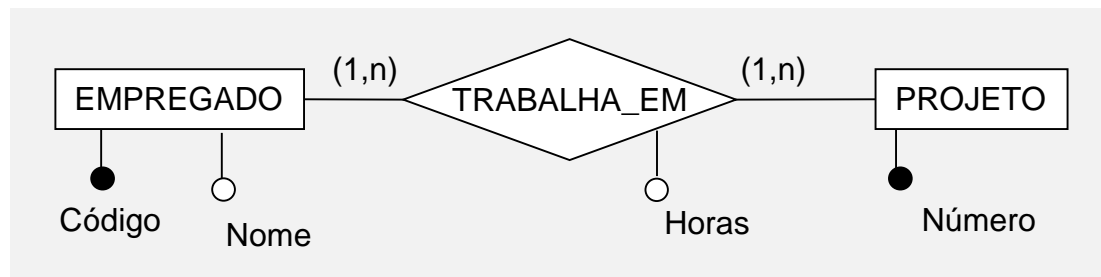
- Cada SGBD oferece um ambiente com as seguintes características:

Banco de Dados = Dados + Descrição dos dados

- 
- Modelo de dados
 - Esquema e Instância da base de dados
 - Independência de dados
 - Dicionário ou catálogo de dados

SGBD – modelo de dados

- O que é um modelo de dados?
 - ❑ Coletânea de conceitos utilizados para descrever a estrutura de um banco de dados



EMPREGADO (Código, Nome, Sobrenome)
TRABALHA_EM (Código, Número, Horas)
Código **Referencia** EMPREGADO
Número **Referencia** PROJETO
PROJETO (Número, Nome)

SGBD – modelo conceitual

- Modelo Conceitual
 - ❑ Descrição mais abstrata da base de dados
 - ❑ Não contém detalhes de implementação
 - ❑ Independente de tipo de SGBD (genérico)
 - ❑ Início do projeto da base de dados

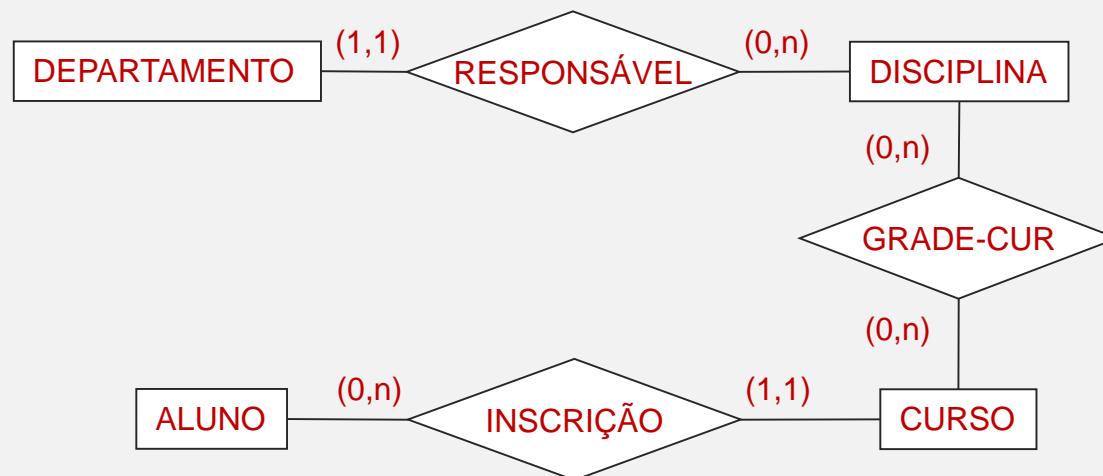


Diagrama Entidade Relacionamento

SGBD – modelo lógico

- Modelo Lógico – níveis de abstração
 - ❑ Descrição da base de dados como vista pelos usuários do SGBD (programadores, usuários que tem acesso ao banco diretamente)
 - ❑ Dependente de SGBD
 - ❑ Não contém detalhes físicos de implementação (índices, etc)

```
EMPREGADO (Código, Nome, Cargo)
TRABALHA_EM (Código, Número, Horas)
           Código Referencia EMPREGADO
           Número Referencia PROJETO
PROJETO (Número, Nome)
```

SGBD – modelo físico

- Modelo Físico – níveis de abstração
 - ❑ Descrição da base de dados como armazenada.
 - ❑ Ordem e forma de armazenamento dos registros (blocos físicos).
 - ❑ Índices disponíveis
 - ❑ Tendência é cada vez mais esconder o modelo físico.



André Luís Nunes

andre.nunes@cwj.com.br