

KDD E MINERAÇÃO DE DADOS:

Revisão em Bancos de Dados Relacionais e SQL

Prof. Ronaldo R. Goldschmidt

ronaldo@de9.ime.eb.br

rribeiro@univercidade.br

geocities.yahoo.com.br/ronaldo_goldschmidt

KDD E MINERAÇÃO DE DADOS

Revisão em Bancos de Dados Relacionais e SQL

Prof. Ronaldo Ribeiro Goldschmidt

REVISÃO DE BD RELACIONAIS E SQL

- **CONCEITOS BÁSICOS**
- **ETAPAS DE DESENVOLVIMENTO DE BANCO DE DADOS**
- **EXEMPLO DE UM BD RELACIONAL**
- **SQL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO EXCEL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS**

REVISÃO DE BD RELACIONAIS E SQL

- **CONCEITOS BÁSICOS**
- **ETAPAS DE DESENVOLVIMENTO DE BANCO DE DADOS**
- **EXEMPLO DE UM BD RELACIONAL**
- **SQL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO EXCEL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS**

CONCEITOS BÁSICOS

Definição:

Um **Banco de Dados** é uma **coleção de dados relacionados**. Onde **dados** significam **fatos conhecidos** que podem ser armazenados e que têm significado implícito.

Exemplo:

A coleção das dados sobre os clientes de uma Financeira que tenham recebido crédito para a compra de algum bem.

Que tipo de dados sobre estes clientes você acharia importante a Financeira manter armazenados ?

CONCEITOS BÁSICOS

Definição:

Um **Sistema Gerenciador de Banco de Dados (SGBD)** é um conjunto de programas que permite a criação e a manutenção de Bancos de Dados.

Exemplos:

- Oracle
- Sql Server
- Sybase
- Etc...

CONCEITOS BÁSICOS

Definição:

Um **Banco de Dados Relacional** é um tipo de banco de dados em que as informações são armazenadas em tabelas. O **Modelo de Banco de Dados Relacional** encontra-se fundamentado na **Teoria dos Conjuntos**.

Definição:

Uma **Relação** é o nome dado a **qualquer subconjunto** de um **produto cartesiano** entre dois ou mais conjuntos.

CONCEITOS BÁSICOS

Exemplo:

Sejam os seguintes domínios (conjuntos):

$D_1 = \{X/X \text{ é CPF de aluno da Disciplina Mineração de Dados}\}$

$D_2 = \{X/X \text{ é Nome de aluno da Disciplina Mineração de Dados}\}$

Forma Explícita:

$D_1 = \{97446262753, 94834568123, 52834567845\}$

$D_2 = \{\text{'José'}, \text{'Ana'}, \text{'Maria'}\}$

CONCEITOS BÁSICOS

Produto Cartesiano: $D_1 \times D_2$

$D_1 \times D_2 = \{$

$(97446262753, \text{'José'}), (97446262753, \text{'Ana'}),$

$(97446262753, \text{'Maria'}), (94834568123, \text{'José'}),$

$(94834568123, \text{'Ana'}), (94834568123, \text{'Maria'}),$

$(52834567845, \text{'José'}), (52834567845, \text{'Ana'}),$

$(52834567845, \text{'Maria'})\}$

Relação: $R \subseteq D_1 \times D_2$

$R = \{ (97446262753, \text{'José'}), (94834568123, \text{'Ana'}),$

$(52834567845, \text{'Maria'})\}$

CONCEITOS BÁSICOS

Relação: $R \subseteq D_1 \times D_2$

**$R = \{ (97446262753, \text{'José'}), (94834568123, \text{'Ana'}),$
 $(52834567845, \text{'Maria'}) \}$**

Aluno(CPF, Nome) é o esquema da Relação $R=ALUNO$

$(97446262753, \text{'José'})$ é uma *instância* da Relação $ALUNO$

CONCEITOS BÁSICOS

Representação Tabular da Relação ALUNO

CPF	NOME
97446262753	José
94834568123	Ana
52834567845	Maria

Aridade da Tabela ALUNO = 2 **atributos** (CPF, NOME)

Cardinalidade da Tabela ALUNO = 3 **instâncias** ou **tuplas**

CONCEITOS BÁSICOS

Tabela ALUNO

CPF	NOME
97446262753	José
94834568123	Ana
52834567845	Maria

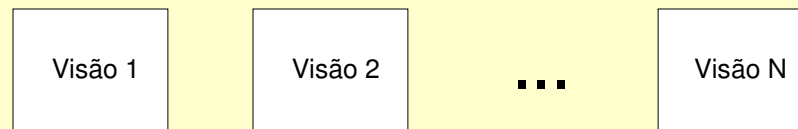
Operações Fundamentais sobre o BD:

- **Inclusão** de novas tuplas
- **Exclusão** de tuplas existentes
- **Alteração** de tuplas existentes
- **Consulta** a tuplas existentes

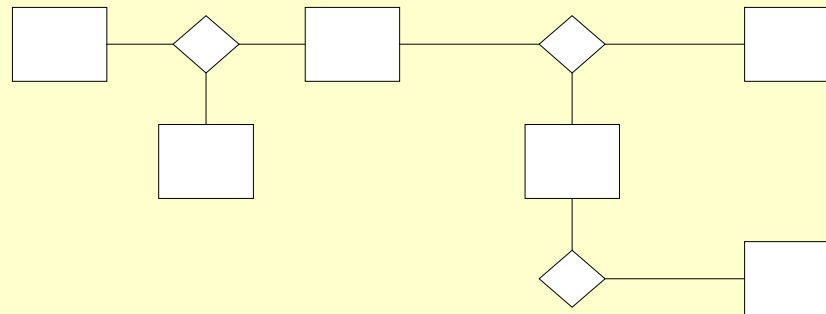
CONCEITOS BÁSICOS

Níveis de Abstração em Banco de Dados:

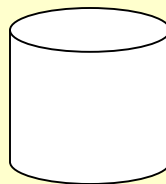
EXTERNO:



CONCEITUAL:



INTERNO



CONCEITOS BÁSICOS

Definição:

Um Data Warehouse é um conjunto de dados:

- Integrado**
- Baseado em assuntos**
- Não - volátil**
- Variável em relação ao tempo**
- De apoio a decisões gerenciais (SAD)**

CONCEITOS BÁSICOS

OBS:

- Ambiente operacional vs. Ambiente analítico**
- DW proporciona um ambiente analítico**
- Geração de um DW: dados operacionais → dados consolidados**
- DW é adequado para aplicações OLAP**
- DW permite visões multi-dimensionais em relação aos dados**

CONCEITOS BÁSICOS

Definição:

A granularidade de um DW refere-se ao nível de detalhe / resumo dos itens de dados deste DW.

Obs: Mais detalhe $\leftrightarrow$ menor granularidade

Mais consolidação $\leftrightarrow$ maior granularidade

REVISÃO DE BD RELACIONAIS E SQL

- **CONCEITOS BÁSICOS**
- **ETAPAS DE DESENVOLVIMENTO DE BANCO DE DADOS**
- **EXEMPLO DE UM BD RELACIONAL**
- **SQL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO EXCEL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS**

ETAPAS DE DESENVOLVIMENTO DE UM BD

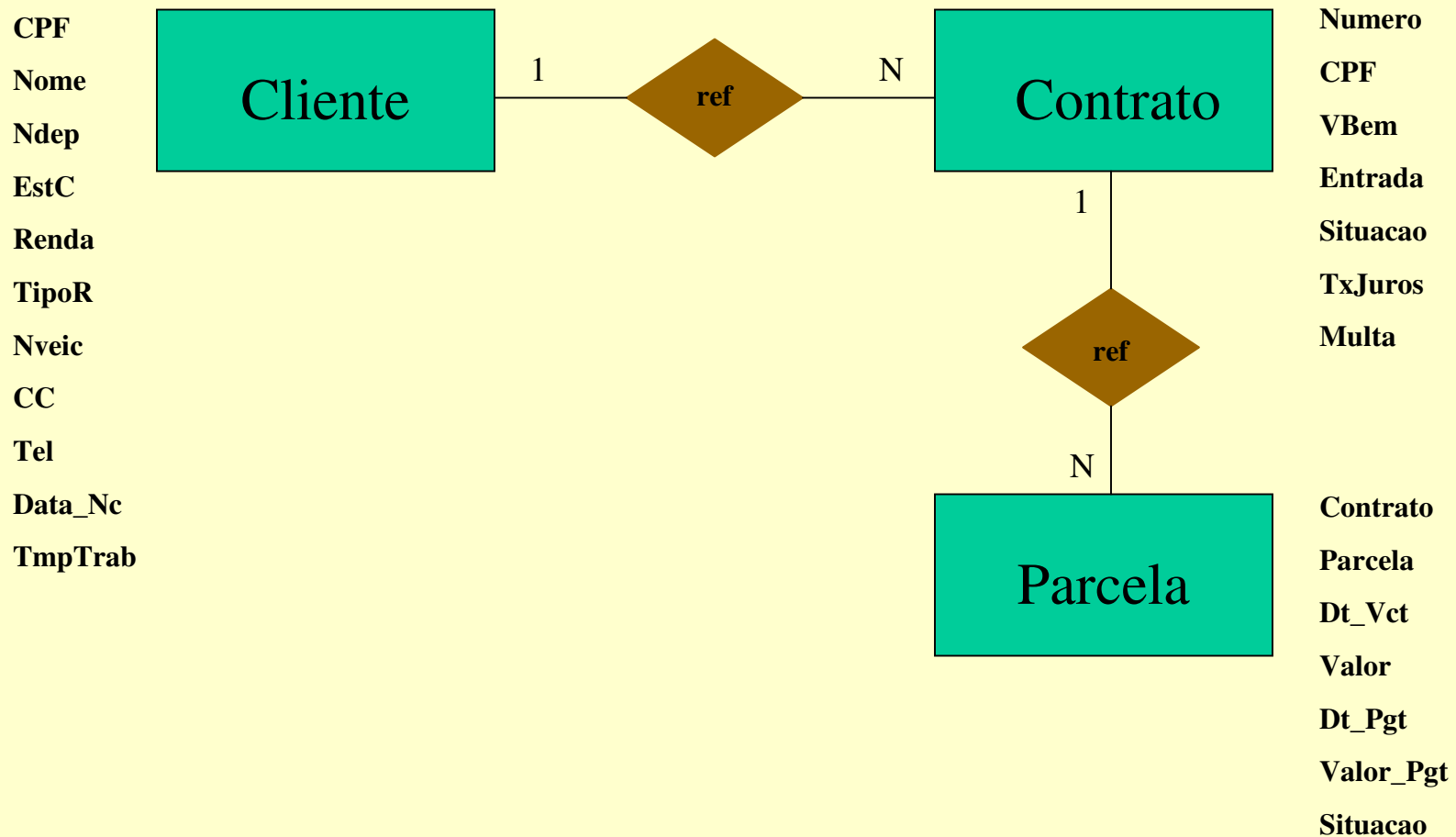
- **Projeto Lógico (Modelagem Conceitual)**
- **Projeto Físico**
- **Implementação**

REVISÃO DE BD RELACIONAIS E SQL

- **CONCEITOS BÁSICOS**
- **ETAPAS DE DESENVOLVIMENTO DE BANCO DE DADOS**
- **EXEMPLO DE UM BD RELACIONAL**
- **SQL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO EXCEL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS**

EXEMPLO DE UM BD RELACIONAL

Modelo de Dados (Produto do Projeto Lógico)



EXEMPLO DE UM BD RELACIONAL

Modelo Relacional

Cliente(CPF, Nome, Ndep, EstC, Renda, TipoR, Nveic, CC, Tel, DtNc, TmpTrab)

Contrato(Numero, CPF, Vbem, Entrada, Situacao, TxJuros, Multa)

Parcela(Contrato, Parcela, DtVct, Valor, DtPgt, ValorPgt, Situacao)

REVISÃO DE BD RELACIONAIS E SQL

- **CONCEITOS BÁSICOS**
- **ETAPAS DE DESENVOLVIMENTO DE BANCO DE DADOS**
- **EXEMPLO DE UM BD RELACIONAL**
- **SQL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO EXCEL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS**

SQL

- **Structured Query Language =**
Linguagem de Consulta Estruturada
- **“Padrão” em BD’s Relacionais**
- **Subdivide-se em:**
 - LDD – Linguagem de Definição de Dados**
 - LMD – Linguagem de Manipulação de Dados**
 - LCD – Linguagem de Controle de Dados**
- **Comandos SQL: embutidos em aplicações ou executados de**
forma interativa

SQL

Forma Geral de um Comando SQL para consulta:

SELECT <lista de Atributos>

FROM <lista de Tabelas>

WHERE <lista de Condições>

GROUP BY <lista de atributos>

HAVING <lista de condições “especiais”>

ORDER BY <lista de atributos>;

REVISÃO DE BD RELACIONAIS E SQL

- **CONCEITOS BÁSICOS**
- **ETAPAS DE DESENVOLVIMENTO DE BANCO DE DADOS**
- **EXEMPLO DE UM BD RELACIONAL**
- **SQL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO EXCEL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS**

EXERCÍCIOS PRÁTICOS UTILIZANDO EXCEL

- Abrir com o Excel o arquivo “**creditos.xls**” fornecido por uma Financeira
- Analisar o tipo de informação disponibilizada após o processo de seleção dos dados
- Pergunta-se: Qual o total de créditos pagos e o total de créditos não pagos ?

EXERCÍCIOS PRÁTICOS UTILIZANDO EXCEL

- **Verificar se existe necessidade de limpeza dos dados. Em caso positivo como proceder para limpar tais dados ?**

Sugestões:

- a) Verifique se há tempo de trabalho maior ou igual à idade do cliente**
- b) Verifique se há valor de entrada maior ou igual ao valor do bem**
- c) Verifique se há valor diferente de zero na coluna CC**
- d) Verifique se há registros em que o tipo de residência não esteja preenchido**

REVISÃO DE BD RELACIONAIS E SQL

- **CONCEITOS BÁSICOS**
- **ETAPAS DE DESENVOLVIMENTO DE BANCO DE DADOS**
- **EXEMPLO DE UM BD RELACIONAL**
- **SQL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO EXCEL**
- **EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS**

EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS

- **Para praticar SQL, crie um BD em Access para importar os dados da planilha em Excel. Nome da Tabela: “creditos”.**
- **Formule as seguintes consultas em SQL:**
 - a) **Obtenha todas as informações da tabela creditos**
 - b) **Recupere todos os dados dos créditos concedidos a clientes com renda acima de 1000 reais**

EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS

- c) Repita a consulta anterior ordenando o resultado em ordem decrescente de renda**
- d) Obter a renda e o valor do bem dos créditos concedidos com renda acima de 1000 reais. Apresentar resultado em ordem crescente de renda**
- e) Obter a idade e o tempo de emprego dos clientes com renda acima de 2000 reais que não pagaram o crédito recebido**
- f) Obtenha as quantidades de créditos concedidos e não pagos agrupadas por estado civil dos clientes**

EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS

- g) Recupere a soma dos valores de entrada dos créditos concedidos e não pagos**
- h) Qual a média de renda dos clientes que receberam crédito e pagaram**
- i) Qual a média da renda e da entrada dos créditos concedidos a clientes casados com idade entre 20 e 30 anos que pagaram**
- j) Obtenha todos os dados dos clientes que pagaram e que ou tenham renda entre 500 e 5000 reais ou tenham mais que 25 anos**

EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS

- **Construa as seguintes tabelas de apoio no BD do exemplo anterior:**

a) Estado_Civil

Codigo: número

Descricao: texto 20

Preencher com as seguintes tuplas:

**(0, 'casado'), (1, 'solteiro'), (2, 'separado'), (3, 'viúvo'), e
(4, 'outros')**

EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS

b) Tipo_Residencia

Codigo: número

Descricao: texto 20

Preencher com as seguintes tuplas:

(0, 'aluguel'), (1, 'própria'), (2, 'parente'), e

(3, 'funcional')

EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS

- Refaça a consulta anterior apresentando a descrição do estado civil ao invés do código

SELECT descricao, count(*)

FROM creditos, estado_civil

WHERE classe = 0 **and** creditos.estc = estado_civil.codigo

GROUP BY descricao;

Observações:

- Produto Cartesiano entre as tabelas;
- Junção Natural

EXERCÍCIOS PRÁTICOS UTILIZANDO ACCESS

- **Faça uma consulta que apresente a distribuição dos créditos pagos por tipo de residência. Mostre a descrição do tipo de residência ao invés do código.**
- **Investigue a base de dados fornecida na busca de regras que caracterizem os clientes que tenham pago os seus créditos.**