



Mineração de Dados

GRIMALDO OLIVEIRA

Sobre Grimaldo

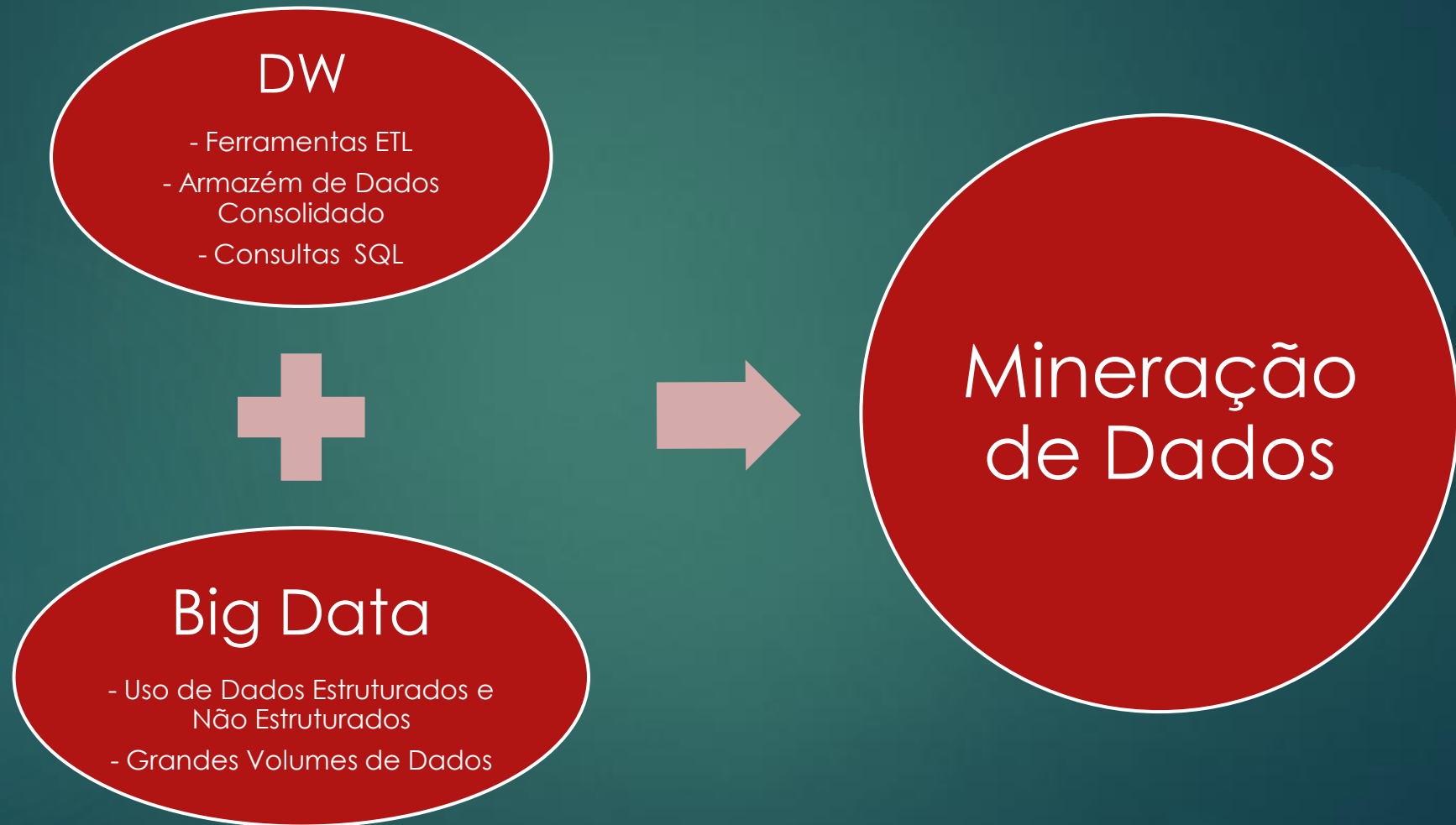


- Grimaldo Oliveira
 - grimaldo_lopes@hotmail.com
- Formação
 - Mestre em Tecnologias Aplicadas a Educação Universidade do Estado da Bahia.
 - Especialização em Análise de Sistemas pela Faculdade Visconde de Cairu.
 - Estatístico pela Universidade Federal da Bahia.
- Atividades
 - Mais de 10 anos atuando como Consultor de Business Intelligence.
 - Projetos Governos Maranhão, Mato Grosso e Bahia.
 - Idealizador do Blog : BI com Vatapá – bicomvatapa.blogspot.com.
 - Livro: BI Como Deve Ser – bicomodeveser.com.br

PROGRAMA

- ▶ Motivação para utilizar Mineração de Dados
- ▶ Processo KDD
- ▶ Aplicações da Mineração de Dados
- ▶ O que é um Padrão nos Dados

Tudo leva a Data Mining



Como Diferenciar DW X Big Data e Data Mining

Pergunta: Armazenar todos os bons e maus pagadores do Banco HSBC no Mundo



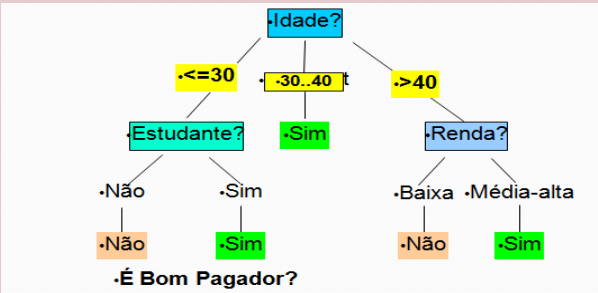
Big Data

Pergunta: Gerar Análise com os maus pagadores apenas nas capitais

Nome	Idade	Renda	Profissão	Bom Pagador
Daniel	≤ 30	Média	Estudante	Sim
João	31..50	Média-Alta	Professor	Sim
Carlos	31..50	Média-Alta	Engenheiro	Sim
Maria	41..50	Baixa	Vendedora	Não
Paulo	≤ 30	Baixa	Porteiro	Não
Otavio	> 60	Baixa	Aposentado	Não

DW

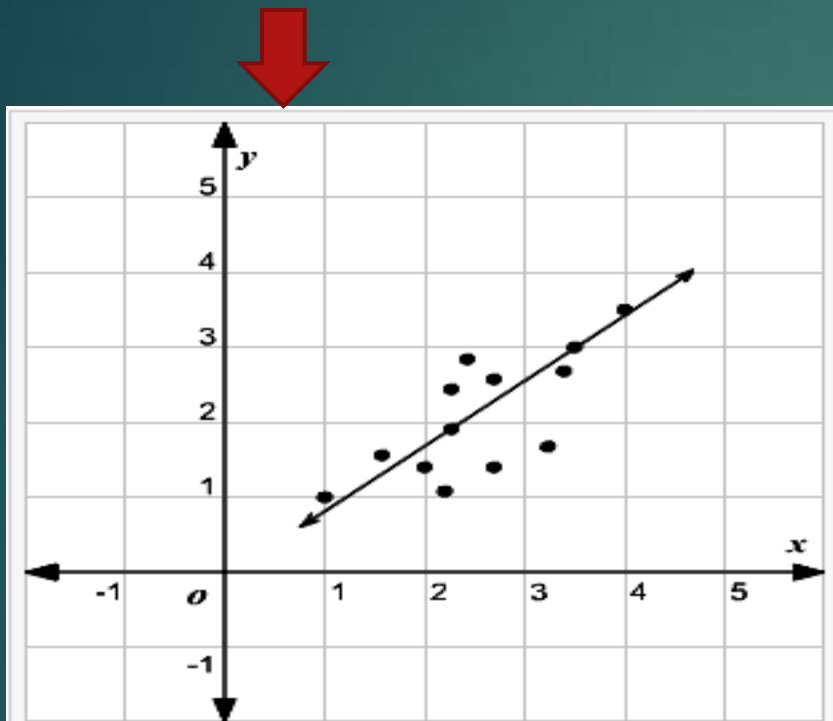
Pergunta: João dos Santos pediu crédito em São Paulo, ele será um bom pagador?



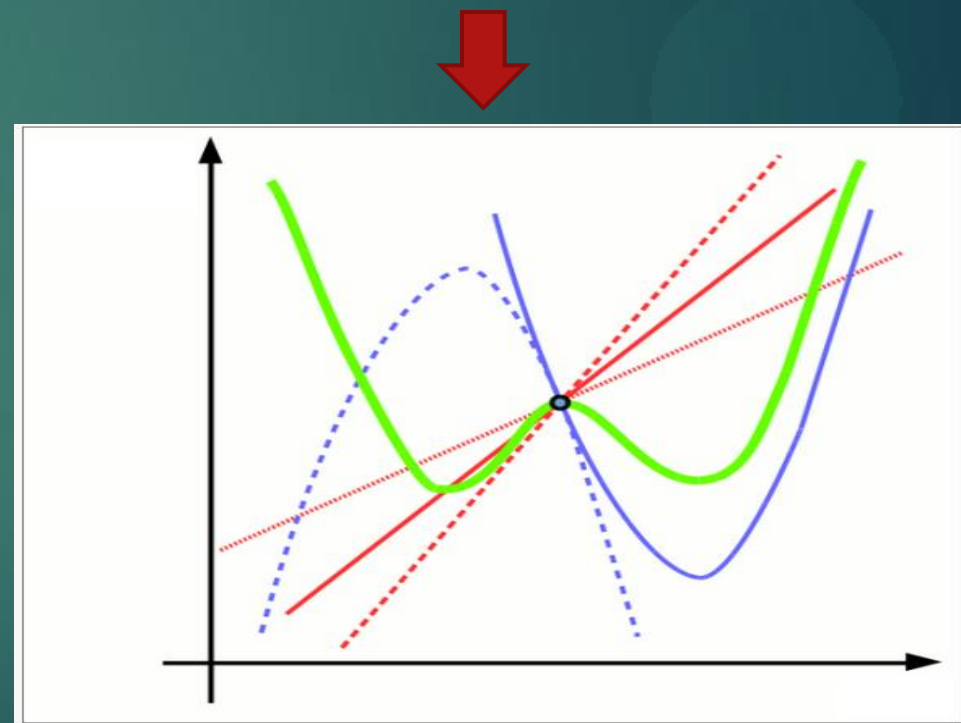
Mineração de Dados

Seria Fácil, se não minerássemos

Como os gestores gostariam que fosse



Como possivelmente seria!



Motivação

- ⌘ A informatização dos meios produtivos permitiu a geração de grandes volumes de dados:
 - ▶ Transações eletrônicas;
 - ▶ Novos equipamentos científicos e industriais para observação e controle;
 - ▶ Dispositivos de armazenamento em massa;
- ⌘ Aproveitamento da informação permite ganho de competitividade: “*conhecimento é poder (e poder = \$\$!)*”

Motivação

- ⌘ Os recursos de análise de dados tradicionais são inviáveis para acompanhar esta evolução
- ⌘ *“Morrendo de sede por conhecimento em um oceano de dados”*
- ⌘ Gigantismo do problema de análise de dados para tomada de decisão:
 - ▶ BD da Wal-Mart: 20 milhões de transações por dia
 - ▶ Data Warehouse da Mobil: 100 TB
 - ▶ BD da NASA: recebe de satélites 50 GB por hora

Motivação

⊕ Solução:

- ▶ ferramentas de automatização das tarefas repetitivas e sistemática de análise de dados
- ▶ ferramentas de auxílio para as tarefas cognitivas da análise
- ▶ integração das ferramentas em sistemas apoiando o processo completo de descoberta de conhecimento para tomada de decisão

Exemplo Preliminar

- ⊕ Um problema do mundo dos negócios: entender o perfil dos clientes
 - ▶ desenvolvimento de novos produtos;
 - ▶ controle de estoque em postos de distribuição;
 - ▶ propaganda mal direcionada gera maiores gastos e desestimula o possível interessado a procurar as ofertas adequadas;
- ⊕ Quais são meus clientes típicos?

COMO DESCUBRO ESTES DADOS ?????



Descoberta de Conhecimento em Bancos de Dados = KDD

- ⌕ “O processo não trivial de extração de informações implícitas, anteriormente desconhecidas, e potencialmente úteis de uma fonte de dados”;
- ⌕ “Torture os dados até eles confessarem”;
- ⌕ O que é um padrão interessante ? (válido, novo, útil e interpretável)

Etapas do KDD



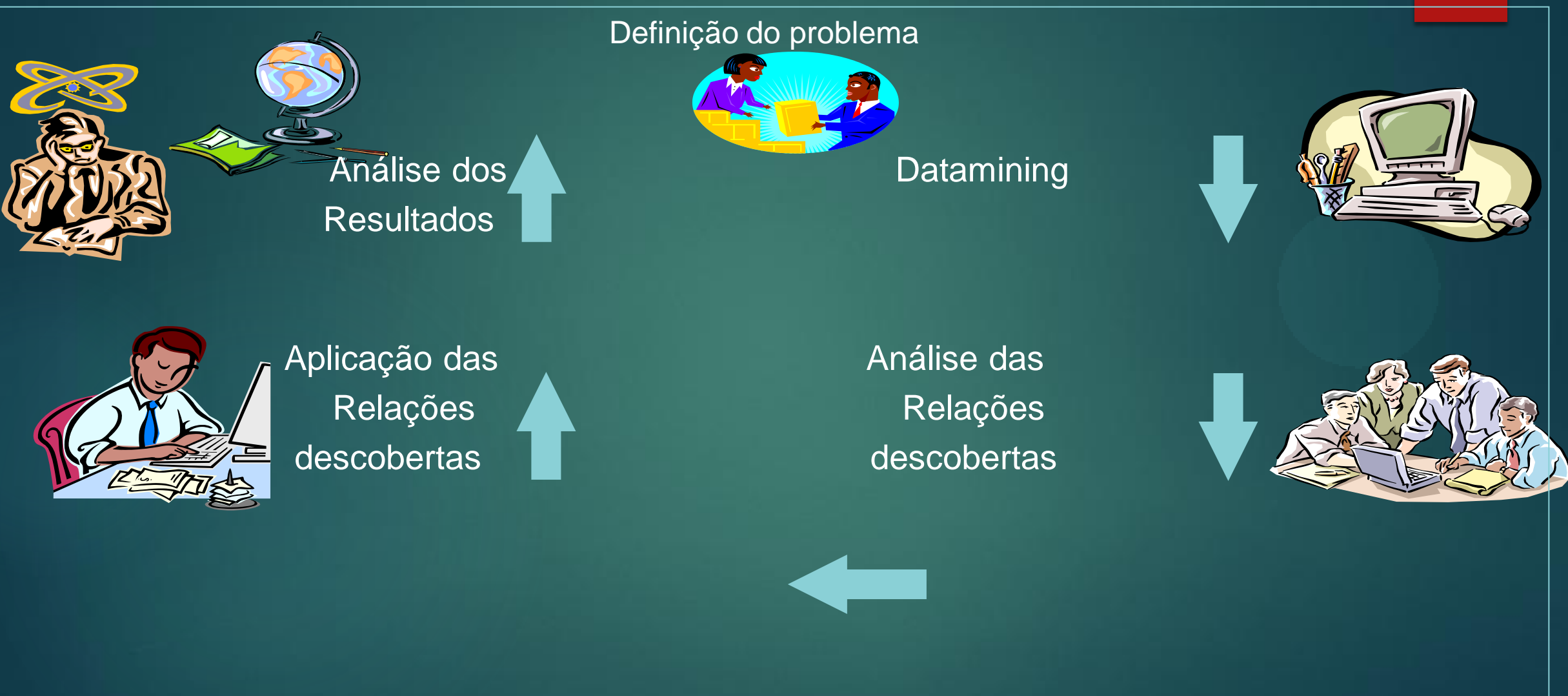
KDD x Data Mining

- ⊕ Mineração de dados é o passo do processo de KDD que produz um conjunto de padrões sob um custo computacional aceitável;
- ⊕ KDD utiliza algoritmos de *data mining* para extrair padrões classificados como “conhecimento”. Incorpora também tarefas como escolha do algoritmo adequado, processamento e amostragem de dados e interpretação de resultados;

Mineração de Dados

A mineração de dados também conhecida como “**garimpagem os dados**” é o processo de extração de informações, sem conhecimento prévio em um grande banco de dados, das características dos dados e seu uso são para tomada de decisões

Fases da Mineração de Dados



Aplicações da Mineração de dados

- ▶ Comércio
 - ▶ Real
 - ▶ Virtual
- ▶ Medicina
- ▶ Detecção de Fraudes
- ▶ Inteligência Competitiva
 - ▶ Concorrentes
 - ▶ Tendências do Mercado

Exemplos

✦ Áreas de aplicações potenciais:

▶ Vendas e Marketing

- ▶ *Identificar padrões de comportamento de consumidores*
- ▶ *Associar comportamentos à características demográficas de consumidores*
- ▶ *Campanhas de marketing direto (mailing campaigns)*
- ▶ *Identificar consumidores “leais”*

Exemplos

✦ Áreas de aplicações potenciais:

- ▶ Bancos

- ▶ *Identificar padrões de fraudes (cartões de crédito)*

- ▶ *Identificar características de correntistas*

- ▶ *Mercado Financeiro (\$\$\$)*

Exemplos

✦ Áreas de aplicações potenciais

▶ Médica

- ▶ *Comportamento de pacientes*
- ▶ *Identificar terapias de sucessos para diferentes tratamentos*
- ▶ *Fraudes em planos de saúde*
- ▶ *Comportamento de usuários de planos de saúde*

Localizando Padrões

Veja se você consegue localizar um padrão na estrutura abaixo:

ABCVYABCVKABDKCABCVUABEWLABCVO

Localizando Padrões

Passos:

- 1-A primeira etapa é perceber que existe uma seqüência de letras que se repete bastante;
- 2-Após as seqüências determinadas, verificamos que elas segmentam o padrão original em diversas unidades independentes

Localizando Padrões

Veja se você encontrou estas possíveis
respostas:

"ABCXY"

"ABCVK"

"ABDKC"

"ABCVU"

"ABEWL"

"ABCVO"

Generalize os
padrões

"ABC??"

"ABD??"

"ABE??"

"AB???",

Substitua

'A' -> "aquisição de pão"

B' -> "aquisição de leite"

Localizando Padrões

Regras Sugeridas

- ▶ Quer dizer, na prática, que toda vez que alguém comprou **pão(A)**, também comprou **leite(B)**. Esses dois atributos estão associados e isto foi revelado pelo processo de descoberta de padrões.

Próximos vídeos...

Finalidade: Coleta de dados com os gestores para a construção do BI.

Fatos		Diária
Dimensões		
Hóspede		✓
Tipo Quarto		✓
Código Tipo Quarto		
Tipo Quarto	HISTÓRICO	
Classe Quarto		✓
Tempo (Data Registro Primeira Diária)		✓

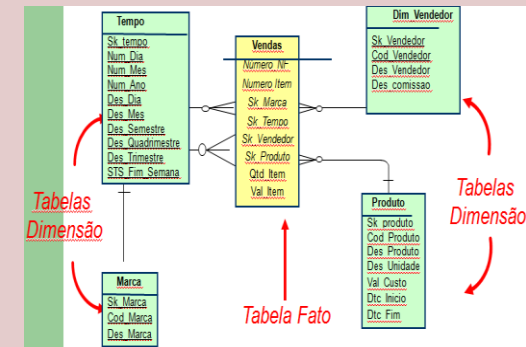
Tarefas de
Mineração de
Dados -Parte01

Finalidade: Levantamento dos relacionamentos e objetos que armazenam os dados da empresa.

DIMENSÕES	ORIGEM	
	TABELA/VISÃO	CAMPO
Hóspede		
Nome Hóspede	HOSPEDE	NOM_HOSPEDE
Cidade Hóspede	CIDADE_ORIGEM	NOM_CIDADE
País Hóspede	PAIS_ORIGEM	NOM_PAIS
Aeroporto Hóspede	AEROPORTO_SAIDA	DES_AEROPORTO
Local Aeroporto Saída	AEROPORTO_SAIDA	NOM_LOCALIDADE
Código Hóspede	HOSPEDE	COD_HOSPEDE

Tarefas de
Mineração de
Dados -Parte02

Finalidade: Modelo adequado para realizar as consultas nas bases que servirão ao BI



Mineração
Visual

contato@bicomodeveser.com