

Sistemas de Informação Gerencial

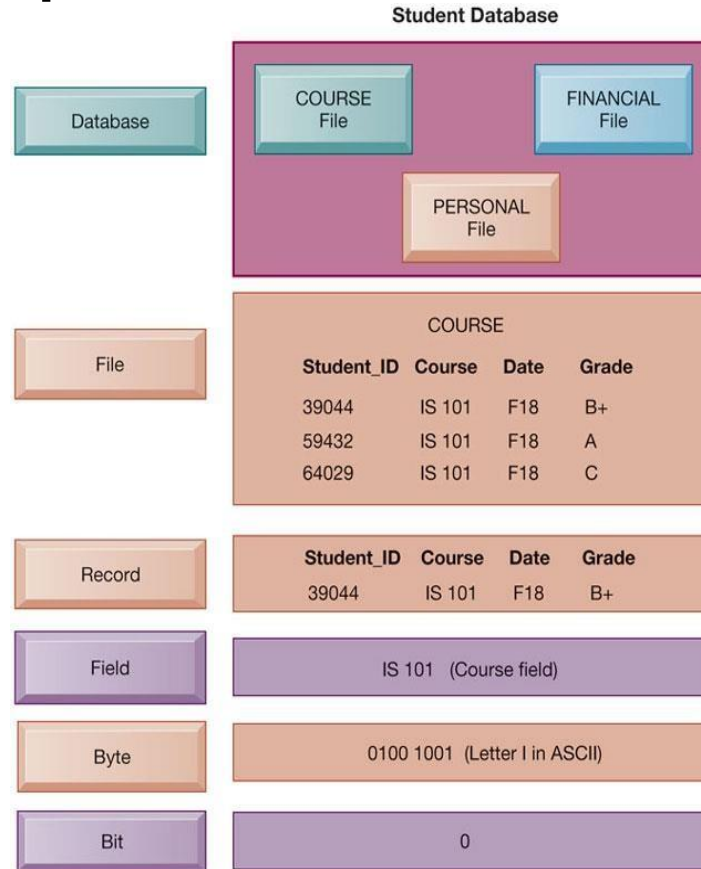
Fundamentos da Inteligência de Negócios: Bases de dados e Gestão da Informação

Prof. Dr. Tiago Araújo

Termos e Conceitos de Organização de Arquivos

- Base de dados: Grupo de arquivos relacionados
- Arquivo: Grupo de registros do mesmo tipo
- Registros: Grupo de campos relacionados
- Campo: Grupo de caracteres como palavra(s) ou número(s)
- Entidade: Pessoa, lugar, coisa sobre a qual armazenamos informações
- Atributo: Cada característica, ou qualidade, que descreve a entidade

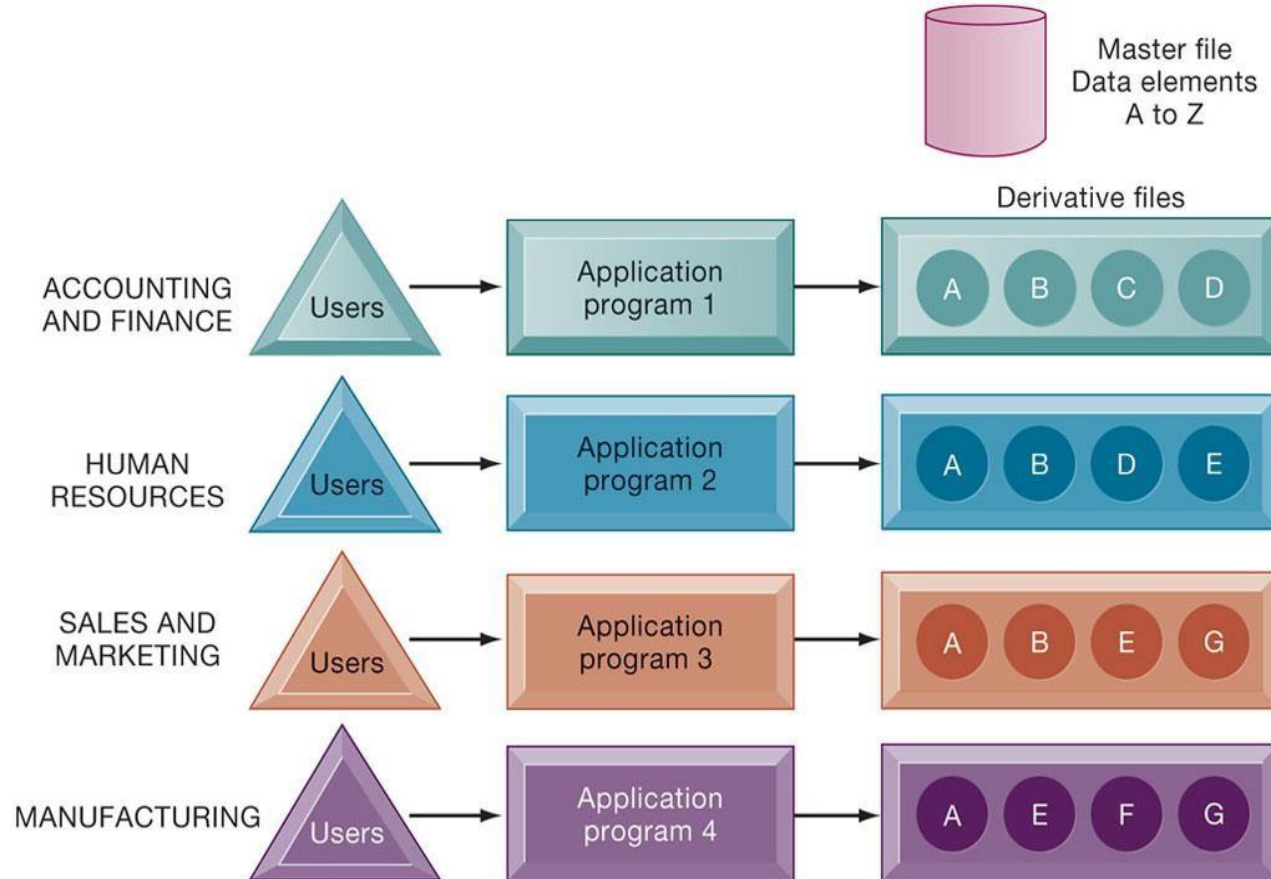
Figura 6.1: A Hierarquia de Dados



Problemas com o Ambiente de Arquivo Tradicional

- Arquivos mantidos separadamente por diferentes departamentos
- Redundância de dados
- Inconsistência de dados
- Dependência de programas e dados
- Falta de flexibilidade
- Má segurança
- Falta de compartilhamento e disponibilidade de dados

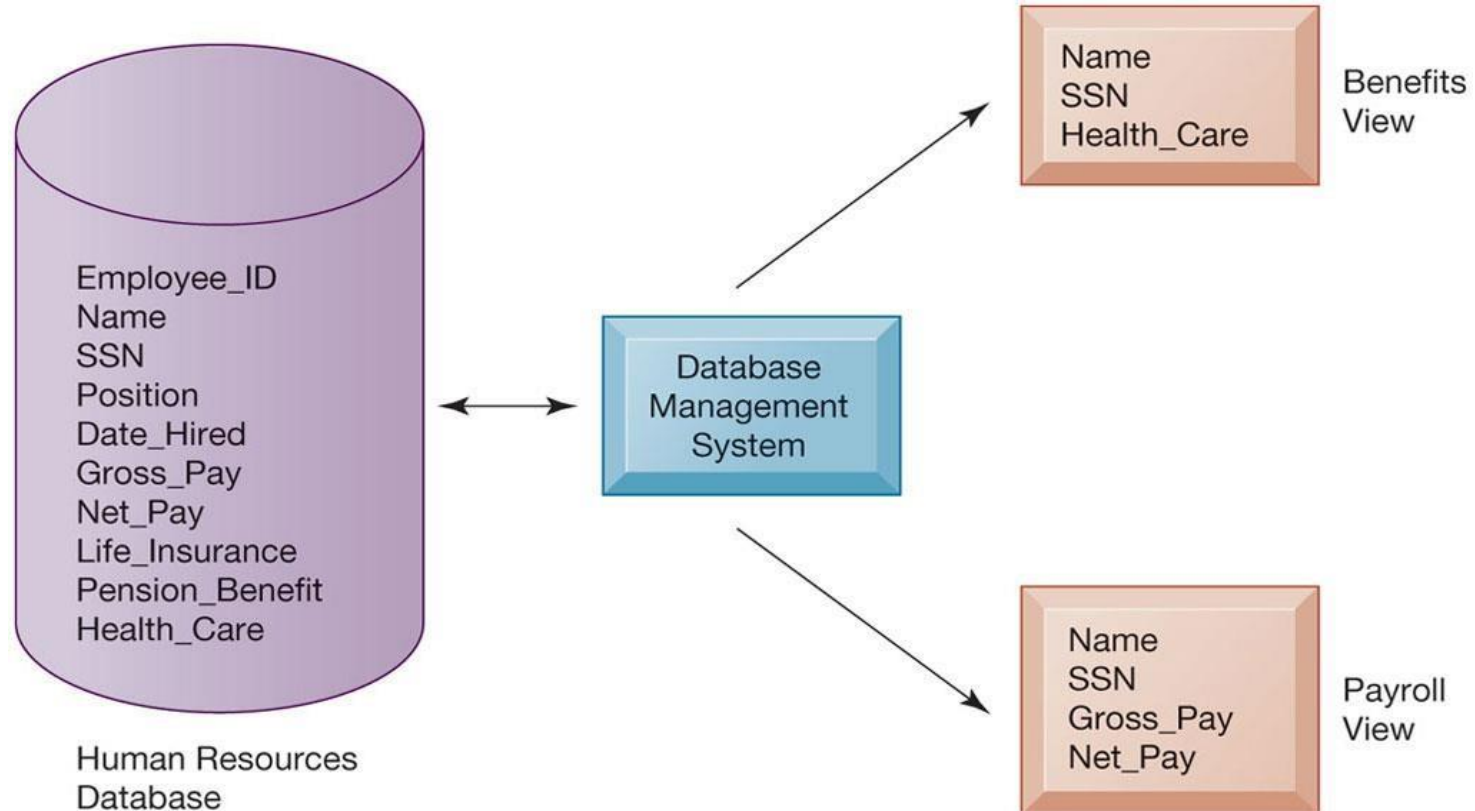
Figura 6.2: Processamento de Arquivos Tradicionais



Sistemas de gerenciamento de banco de dados

- **Base de dados**
 - Atende muitas aplicações centralizando dados e controlando dados redundantes
- **Sistema de gerenciamento de banco de dados (SGBD)**
 - Interfaces entre aplicações e arquivos de dados físicos
 - Separa as visões lógica e física dos dados
 - Resolve problemas de ambiente de arquivos tradicionais
 - Controla a redundância
 - Elimina a incoerência
 - Desacoplar programas e dados
 - Permite a organização gerenciar centralmente os dados e a segurança dos dados

Figura 6.3: Banco de dados de recursos humanos com múltiplas visões



SGBD relacional

- **Representar dados como tabelas bidimensionais**
- **Cada tabela contém dados sobre a entidade e os atributos**
- **Tabela: grade de colunas e filas**
 - Linhas (tuplas): Registros para diferentes entidades
 - Campos (colunas): Representa atributo para entidade
 - Campo chave: Campo usado para identificar de forma única cada registro
 - Chave primária: Campo em tabela utilizado para campos-chave
 - Chave estrangeira: Chave primária utilizada na segunda tabela como campo de observação para identificar registros da tabela original

Figura 6.4: Tabelas Relacionais de Bancos de Dados

SUPPLIER

Columns (Attributes, Fields)

Supplier_Number	Supplier_Name	Supplier_Street	Supplier_City	Supplier_State	Supplier_Zip
8259	CBM Inc.	74 5 th Avenue	Dayton	OH	45220
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345
8263	Jackson Composites	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344

Rows (Records, Tuples)

Key Field (Primary Key)

PART

Part_Number	Part_Name	Unit_Price	Supplier_Number
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

Primary Key

Foreign Key

Operações de um SGBD Relacional

- **Três operações básicas usadas para desenvolver conjuntos úteis de dados**

- **SELECT**

- Cria um subconjunto de dados de todos os registros que atendem aos critérios estabelecidos

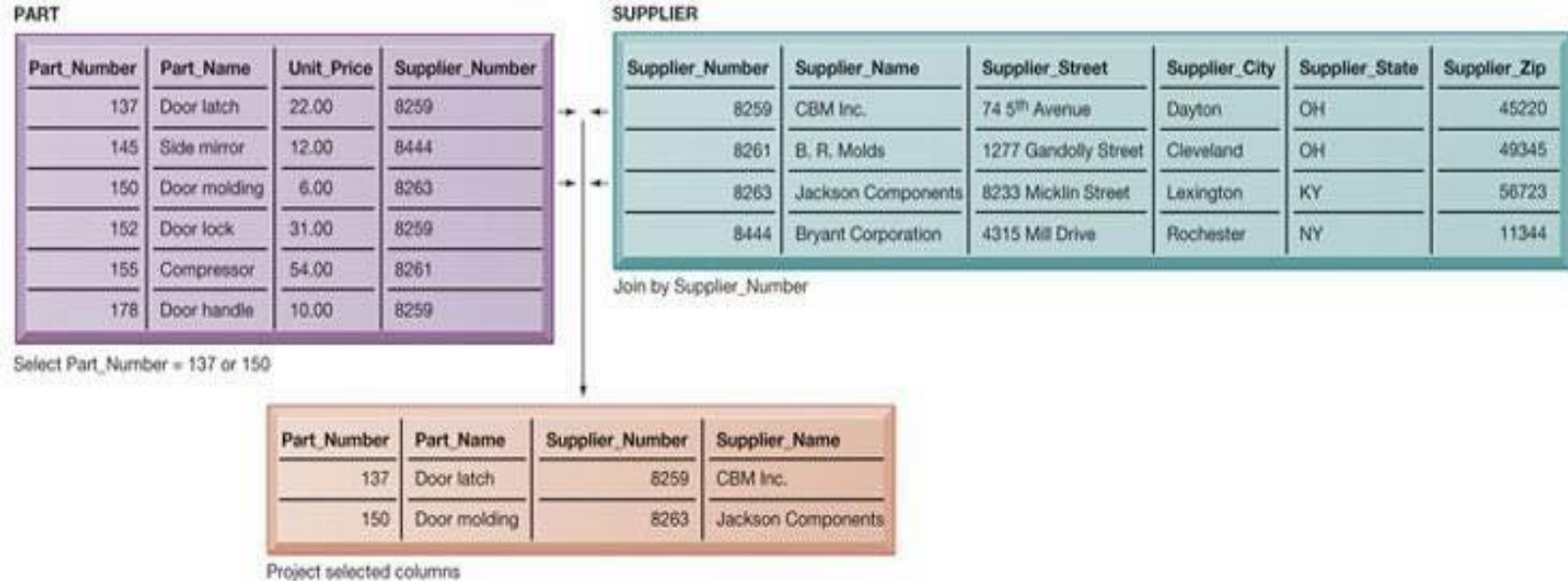
- **JOIN**

- Combina tabelas relacionais para fornecer ao usuário mais informações do que as disponíveis em tabelas individuais

- **PROJECT**

- Cria um subconjunto de colunas em tabela, criando tabelas apenas com as informações especificadas

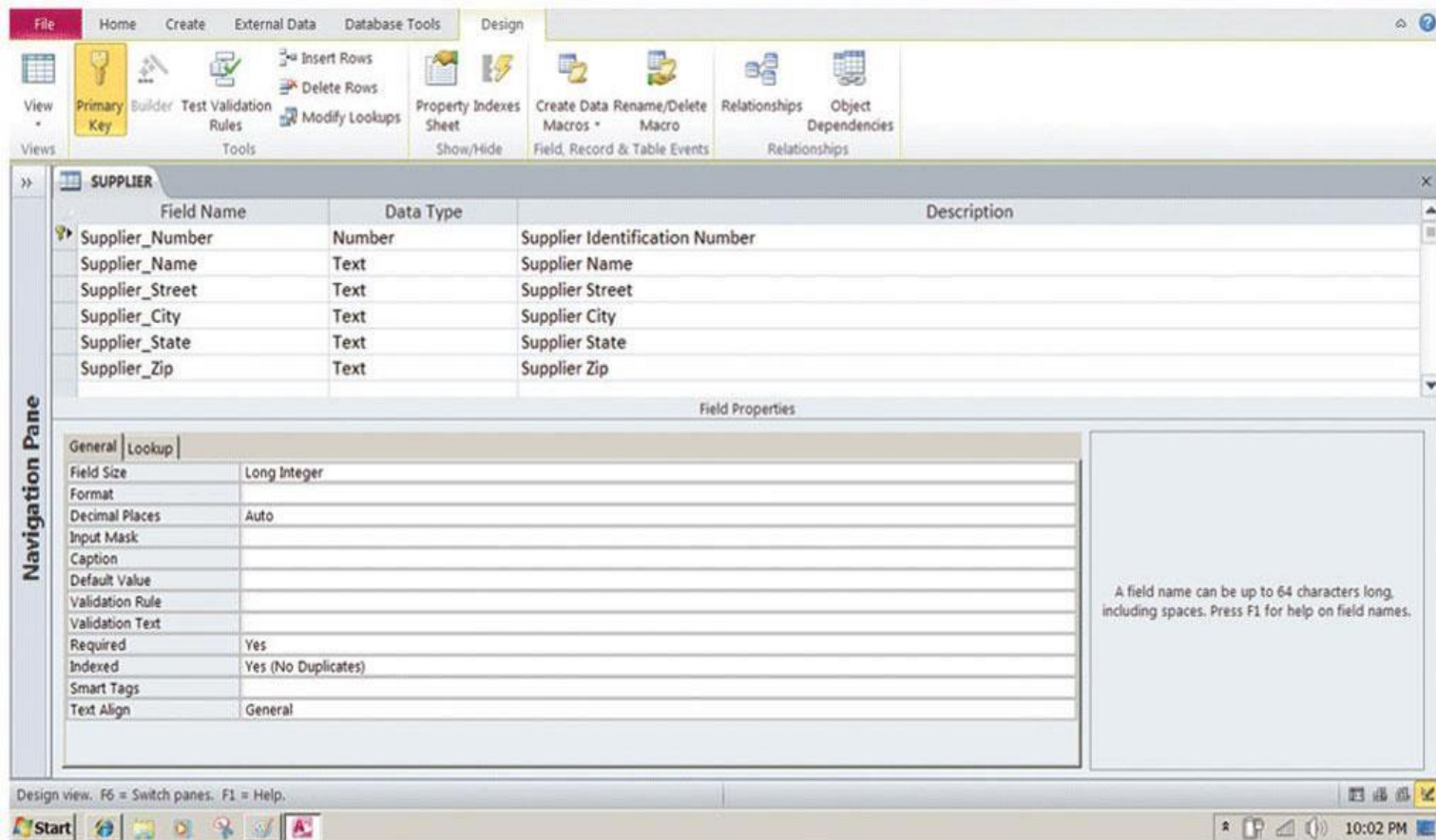
Figura 6.5: As Três Operações Básicas de um SGBD Relacional



Capacidades dos sistemas de gerenciamento de banco de dados

- **Capacidade de definição de dados**
- **Dicionário de dados**
- **Consultas e relatórios**
 - Linguagem de manipulação de dados
 - Linguagem de Consulta Estruturada (SQL)
- **Muitos SGBD têm capacidade de geração de relatórios para a criação de relatórios polidos (Microsoft Access)**

Figura 6.6: Características do Dicionário de Dados de Acesso



Navigation Pane

SUPPLIER

Field Name	Data Type	Description
Supplier_Number	Number	Supplier Identification Number
Supplier_Name	Text	Supplier Name
Supplier_Street	Text	Supplier Street
Supplier_City	Text	Supplier City
Supplier_State	Text	Supplier State
Supplier_Zip	Text	Supplier Zip

Field Properties

General | Lookup

Field Size	Long Integer
Format	
Decimal Places	Auto
Input Mask	
Caption	
Default Value	
Validation Rule	
Validation Text	
Required	Yes
Indexed	Yes (No Duplicates)
Smart Tags	
Text Align	General

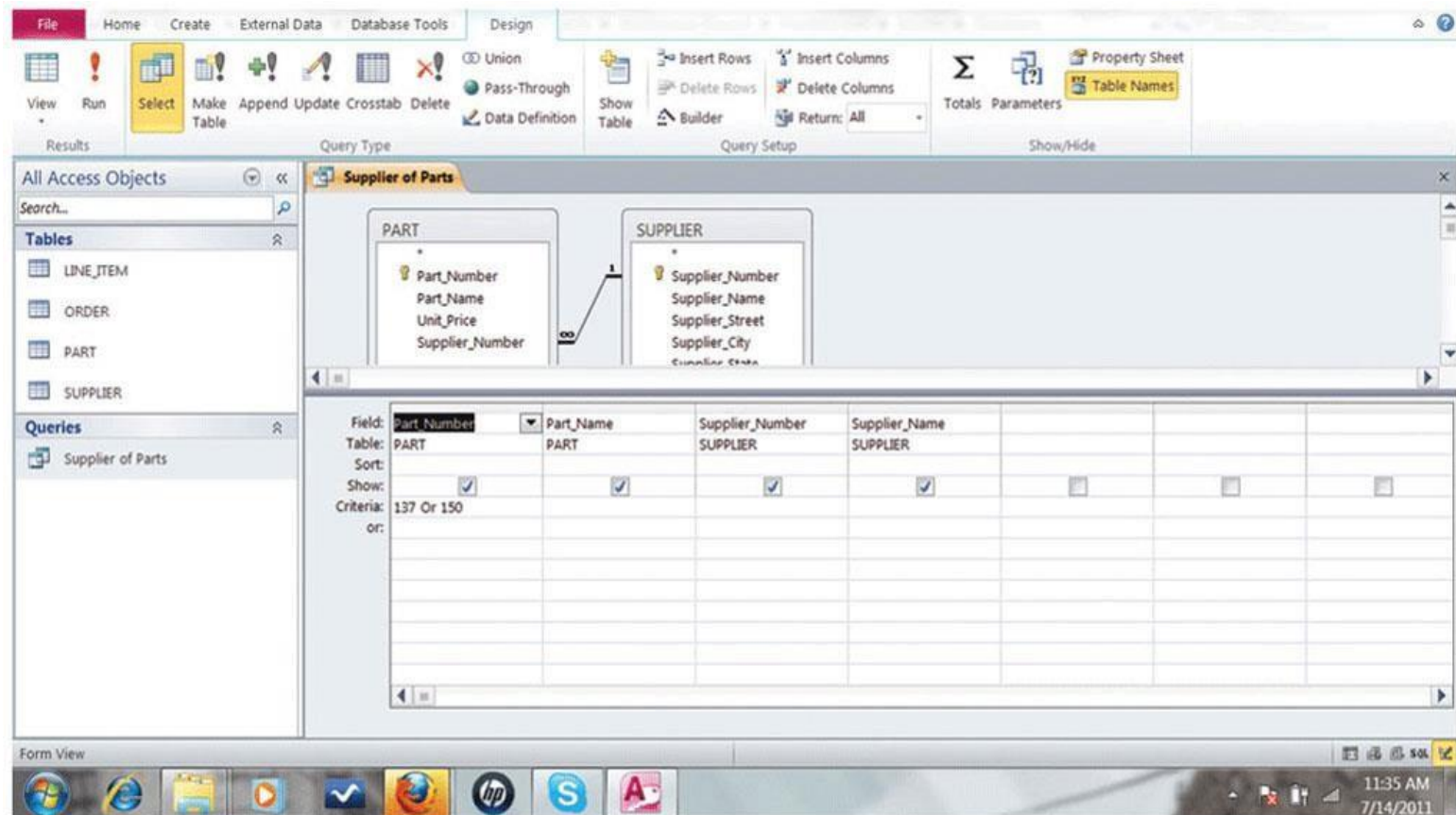
A field name can be up to 64 characters long, including spaces. Press F1 for help on field names.

Design view. F6 = Switch panes. F1 = Help.

Figura 6.7: Exemplo de uma consulta SQL

```
SELECT PART.Part_Number, PART.Part_Name, SUPPLIER.Supplier_Number,  
SUPPLIER.Supplier_Name  
FROM PART, SUPPLIER  
WHERE PART.Supplier_Number = SUPPLIER.Supplier_Number AND  
Part_Number = 137 OR Part_Number = 150;
```

Figura 6.8: Uma consulta de acesso



Projetando bancos de dados

- **Projeto conceitual vs. projeto físico**
- **Normalização**
 - Simplificação de agrupamentos complexos de dados para minimizar redundâncias de elementos de dados e relacionamentos difíceis de muitos para muitos
- **Integridade referencial**
 - As regras usadas pelo RDBMS para garantir que as relações entre as tabelas permaneçam consistentes
- **Diagrama de relacionamento entidade-relacionamento**
- **Um modelo de dados correto é essencial para um sistema que serve o bem de negócios**

Figura 6.9: Uma relação não normalizada para a ordem

ORDER (Before Normalization)

Order_ Number	Order_ Date	Part_ Number	Part_ Name	Unit_ Price	Part_ Quantity	Supplier_ Number	Supplier_ Name	Supplier_ Street	Supplier_ City	Supplier_ State	Supplier_ Zip
------------------	----------------	-----------------	---------------	----------------	-------------------	---------------------	-------------------	---------------------	-------------------	--------------------	------------------

Figura 6.10: Tabelas Normalizadas Criadas a partir da Ordem

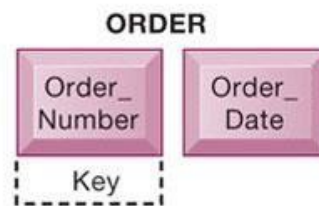
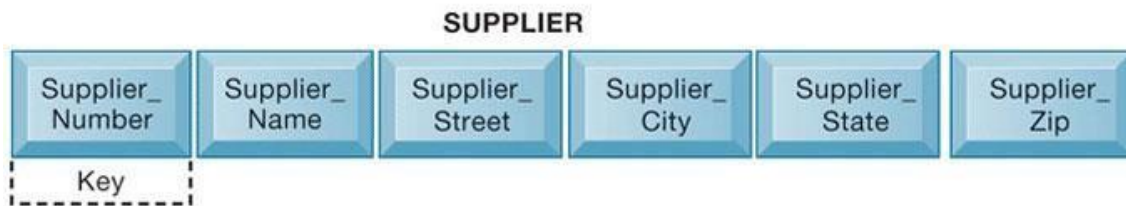
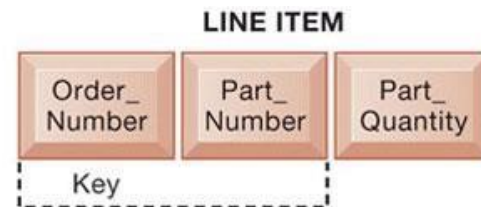
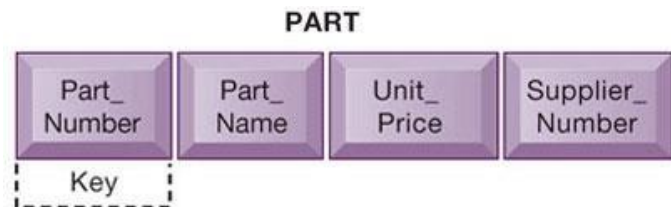
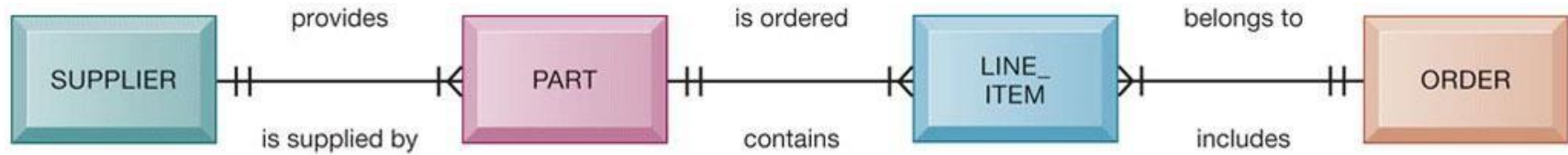


Figura 6.11: Diagrama de Relacionamento Entidade-Relacionamento



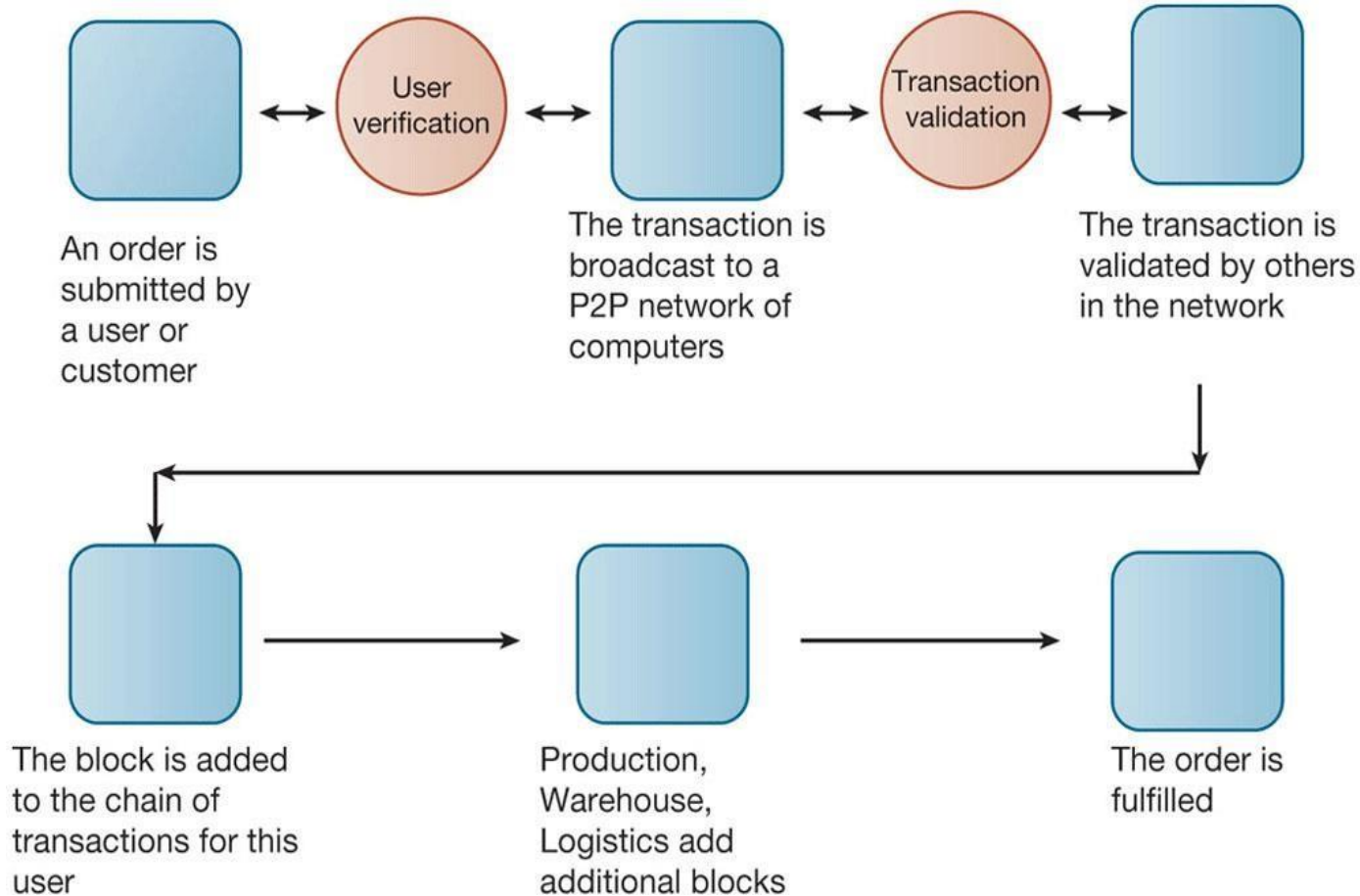
Bancos de Dados não-relacionais e Bancos de Dados na Nuvem

- **Bases de Dados não-relacionais: "Sem SQL"**
 - Modelo de dados mais flexível
 - Conjuntos de dados armazenados em máquinas distribuídas
 - Mais fácil de escalar
 - Manusear grandes volumes de dados não estruturados e estruturados
- **Bases de Dados na Nuvem**
 - Apelo às empresas iniciantes, empresas menores
 - Serviço de Banco de Dados Relacionais da Amazon, Microsoft SQL Azure
 - Nuvens privadas

Blockchain (Cadeia de Bloqueio)

- Listas distribuídas em um banco de dados distribuído ponto-a-ponto
- Mantém uma lista crescente de registros e transações compartilhados por todos
- Criptografia utilizada para identificar participantes e transações
- Utilizado para transações financeiras, cadeia de suprimentos e registros médicos
- Fundação da Bitcoin, e outras moedas criptográficas

Figura 6.12 Como funciona a Blockchain



O desafio do Big Data

- **Big Data**

- Conjuntos maciços de dados não estruturados/semiestruturados do tráfego web, mídia social, sensores, e assim por diante

- **Volumes muito grandes para o SGBD típico**

- Petabytes, exabytes de dados

- **Requer novas ferramentas e tecnologias para gerenciar e analisar**

Infraestrutura de Inteligência de Negócios (1 de 3)

- **Conjunto de ferramentas para obter informações de sistemas separados e de grandes dados**
- **Armazenamento de dados**
 - Armazena dados atuais e históricos de muitos sistemas centrais de transações operacionais
 - Consolida e padroniza as informações para uso em toda a empresa, mas os dados não podem ser alterados
 - Fornece ferramentas de análise e relatórios

Infraestrutura de Inteligência de Negócios (2 de 3)

- **Data marts**

- Subconjunto de armazenamento de dados
- Tipicamente focado em um único assunto ou linha de negócios

- **Hadoop**

- Possibilita o processamento paralelo distribuído de grandes dados através de computadores baratos
- Principais serviços:
 - Hadoop Distributed File System (HDFS): armazenamento de dados
 - MapReduce: divide os dados em grupos para o trabalho
 - Hbase: Sem banco de dados SQL
- Usado Yahoo, NextBio

Infraestrutura de Inteligência de Negócios (3 de 3)

- **Computação na memória**

- Usado em grandes análises de dados
- Utiliza memória principal (RAM) dos computadores para armazenamento de dados para evitar atrasos na recuperação de dados do armazenamento em disco
- Pode reduzir horas/dias de processamento a segundos
- Requer hardware otimizado

- **Plataformas analíticas**

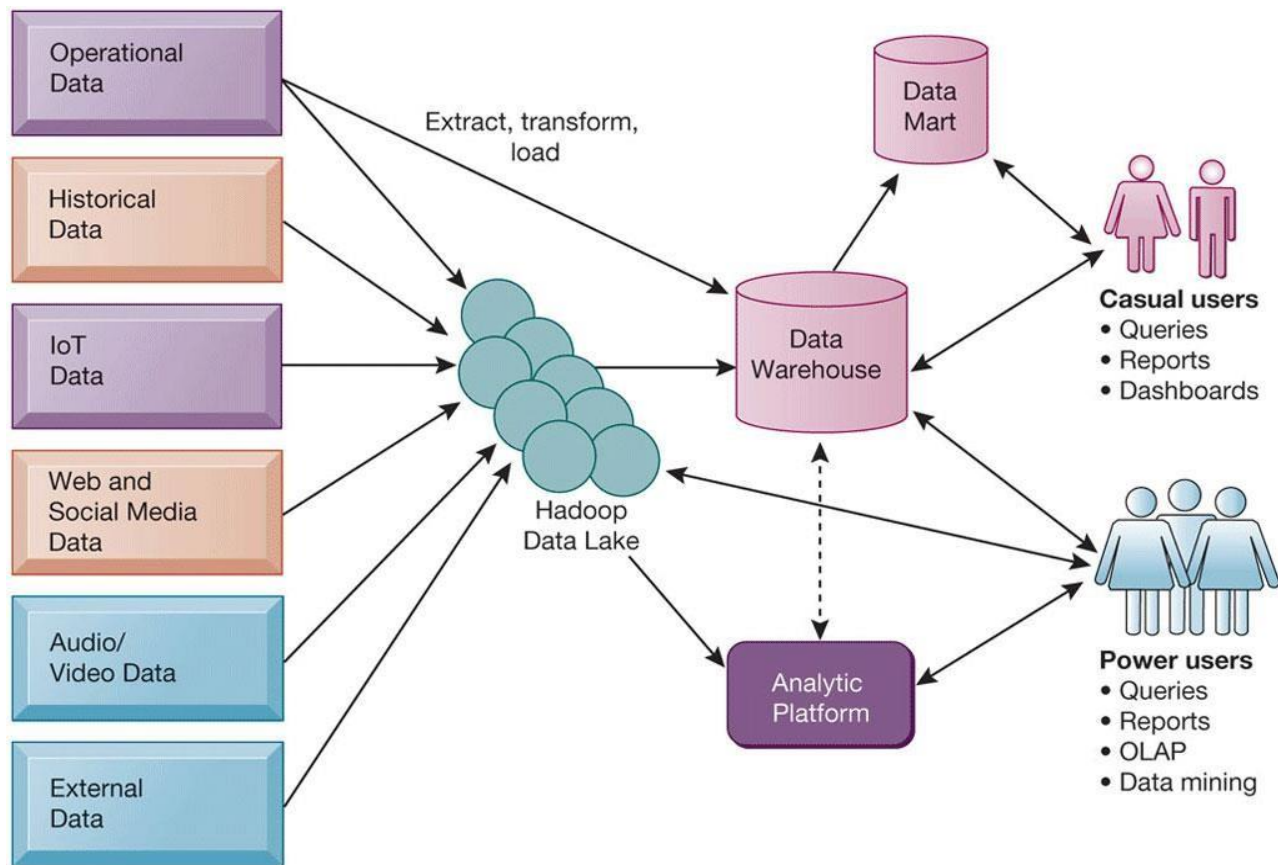
- Plataformas de alta velocidade que utilizam ferramentas otimizadas para grande quantidade de dados para ambos relacionais e não relacionais

Sessão interativa: Tecnologia: Kraft Heinz encontra uma nova receita para a análise de seus dados

- **Discussão em classe**

- Identificar o problema neste estudo de caso. Em que medida foi um problema tecnológico? Estavam envolvidos fatores administrativos e organizacionais?
- Como a tecnologia da informação estava afetando o desempenho comercial na Kraft Heinz?
- Como a nova tecnologia forneceu uma solução para o problema? Qual foi a eficácia da solução?
- Identificar os fatores gerenciais, organizacionais e tecnológicos que tiveram que ser abordados na seleção e implementação da nova solução de armazenamento de dados da Kraft Heinz.

Figura 6.13: Infraestrutura de Inteligência de Negócios Contemporânea



Ferramentas analíticas: Relacionamentos, Padrões, Tendências

- **Ferramentas para consolidar, analisar e fornecer acesso a grandes quantidades de dados para ajudar os usuários a tomar melhores decisões comerciais**
 - Análise de dados multidimensionais (OLAP)
 - Mineração de dados
 - Mineração de textos
 - Mineração de web

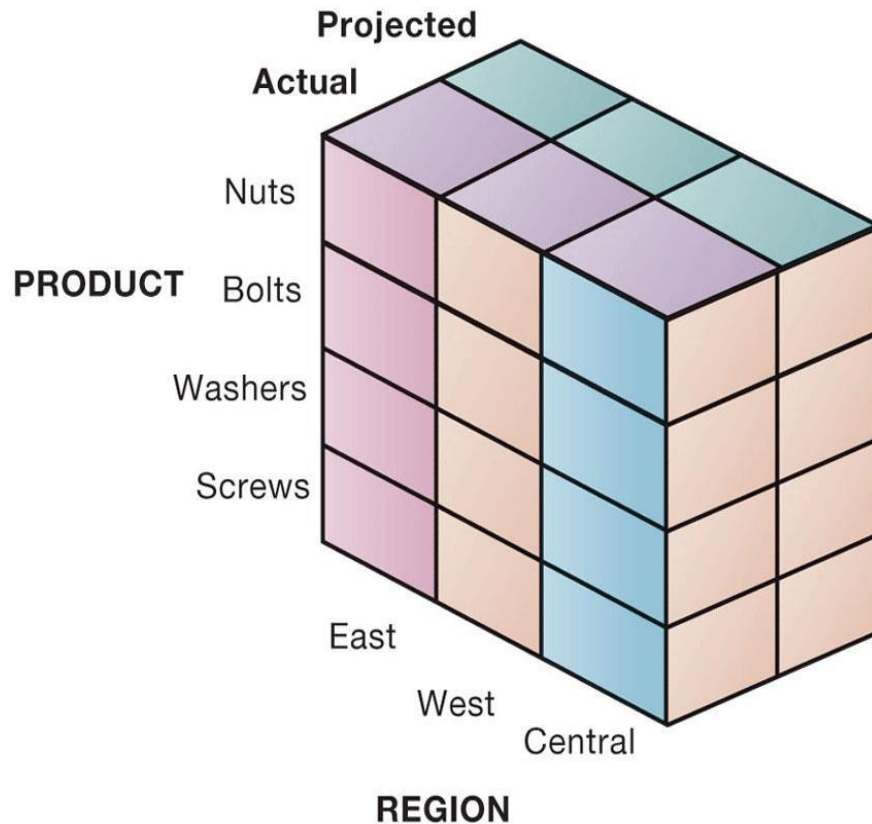
Processamento Analítico Online (OLAP)

- **Suporta análise de dados multidimensionais**

- Visualização de dados usando múltiplas dimensões
- Cada aspecto da informação (produto, preço, custo, região, período de tempo) tem uma dimensão diferente
 - Exemplo: Quantas lavadoras foram vendidas no Leste em junho, em comparação com outras regiões?

- **OLAP permite respostas rápidas e online a consultas ad hoc**

Figura 6.14: Modelo de Dados Multidimensionais



Mineração de Dados

- **Encontra padrões ocultos, relacionamentos em conjuntos de dados**
 - Exemplo: padrões de compra do cliente
- **Infere regras para prever o comportamento futuro**
- **Tipos de informações que podem ser obtidas através da mineração de dados:**
 - Associações
 - Sequências
 - Classificação
 - Agrupamento
 - Previsão

Mineração de Texto e Mineração da Web

- **Mineração de Texto**

- Extrai elementos-chave de grandes conjuntos de dados não estruturados
- Software de análise de sentimentos

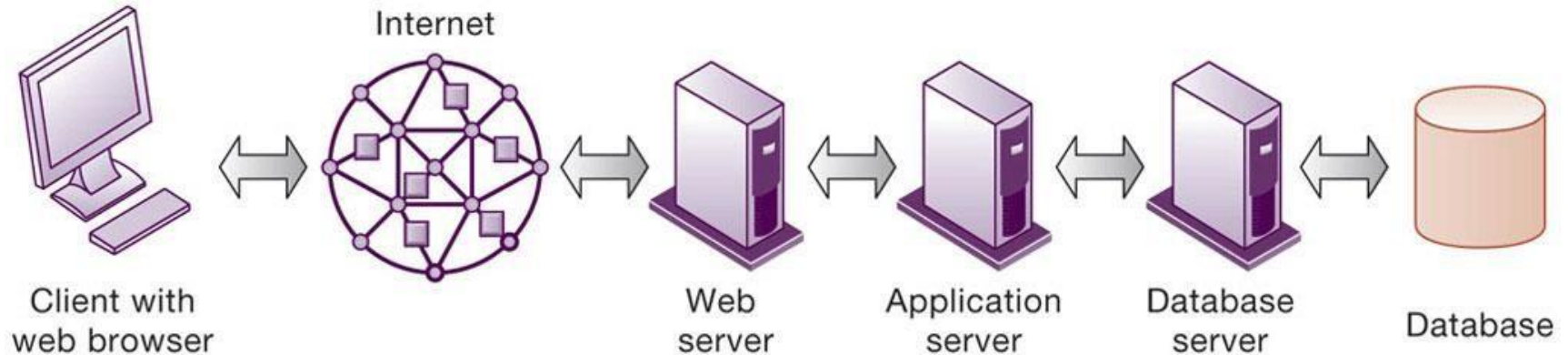
- **Mineração da Web**

- Descoberta e análise de padrões úteis e informações da web
- Mineração de conteúdo da web
- Mineração de estrutura da web
- Mineração de uso da Web

Bancos de dados e a Web

- **Muitas empresas usam a web para disponibilizar alguns bancos de dados internos para clientes ou parceiros**
- **A configuração típica inclui:**
 - Servidor web
 - Servidor de aplicativos/middleware/scripts CGI
 - Servidor de banco de dados (hospedando DBMS)
- **Vantagens de usar a web para acesso ao banco de dados:**
 - Facilidade de uso do software do navegador
 - A interface da Web requer poucas ou nenhuma alteração no banco de dados
 - Barato para adicionar interface web ao sistema

Figura 6.15: Vinculando Bancos de Dados Internos à Web



Estabelecimento de uma Política de Informação

- **Regras, procedimentos, funções da empresa para compartilhar, gerenciar, padronizar dados**
- **Administração de dados**
 - Estabelece políticas e procedimentos para gerenciar dados
- **Gestão de dados**
 - Lida com políticas e processos para gerenciar disponibilidade, usabilidade, integridade e segurança de dados, especialmente em relação a regulamentações governamentais
- **Administração de banco de dados**

Assegurando a qualidade dos dados

- **Mais de 25 por cento dos dados críticos na empresa de Banco de Dados Fortune 1000 são imprecisas ou incompletas**
- **Antes da criação de um novo banco de dados, uma empresa precisa:**
 - Identificar e corrigir dados errôneos
 - Estabelecer melhores rotinas para a edição de dados uma vez que o banco de dados esteja em operação
- **Auditoria de qualidade dos dados**
- **Limpeza de dados**