

Bases de Dados Espaciais

Livro: Spatial Databases: A Tour
Shashi Shekhar – Sanjay Chawla

Biblioteca do DCC – G70.212

Utilidade dos SDBMS (Spatial Database Manag. Systems)

- Os sistemas tradicionais (não espaciais) de gestão de bases de dados suportam:
 - Persistência dos dados a falhas
 - Permitem o acesso concorrente a dados
 - Escalabilidade na resolução de consultas em grandes conjuntos de dados que não cabem na memória principal dos computadores
 - Eficiência para consultas não-espaciais, mas não para consultas espaciais.
- Consultas não-espaciais:
 - Listar o nomes de todas as livrarias com mais de 10000 títulos.
 - Listar os nomes dos 10 maiores clientes, em termos de vendas, no ano de 2006
- Consultas espaciais:
 - Listar os nomes de todas as livrarias a menos de 20 km do Porto
 - Listar os nomes dos clientes que vivem no Porto e nos seus concelhos adjacentes.

Utilidade dos SDBMS – Exemplos de Dados Espaciais

- Exemplos de dados não-espaciais
 - Nomes, números de telefone, emails de pessoas
- Exemplos de dados Espaciais
 - Dados de Recenseamentos
 - Imagens de satélite da NASA - terabytes de dados por dia
 - Meteorologia e dados climáticos
 - Rios, terrenos, impacto ecológico
 - Imagens médicas, radiografias, TACs, etc.
- Exercício: Identificar dados espaciais e não espaciais em
 - Lista telefónica
 - Livro de receitas culinárias

O que é um SDBMS ?

- Um SDBMS é um módulo de software que
 - Pode trabalhar com uma camada suportada por um DBMS
 - Suporta modelos de dados espaciais, tipos abstractos de dados espaciais (ADTs) e uma linguagem de consulta da qual estes ADTs possam ser chamados
 - Suporta indexação espacial, algoritmos eficientes para processar operadores espaciais, e regras especializadas por domínios para optimização de consultas.
- Exemplos: Oracle Spatial data cartridge, ESRI SDE
 - Pode trabalhar com o Oracle 8i DBMS
 - Tem tipos de dados espaciais (e.g. polygon), operações (e.g. overlap) que podem ser chamadas da linguagem SQL3
 - Tem índices espaciais, e.g. R-trees

Exemplo de SDBMS

- Considerem um conjunto de dados espaciais com:
 - Limite de concelho (linha tracejada branca)
 - Dados de recenseamento de bairros (census)- name, area, population, boundary (linha a preto)
 - Zonas de água (polígonos pretos)
 - Imagens de Satélite (pixels a cinzento)

Armazenamento numa tabela de um SDBMS:

```
create table census_blocks (  
  name string,  
  area float,  
  population number,  
  boundary polygon );
```



Fig 1.2

Modelação de Dados Espaciais em DBMS tradicionais

- Um tuplo na tabela census_blocks (Figura 1.3)
- Pergunta: Será o tipo de dados **Polyline** suportado em DBMS?

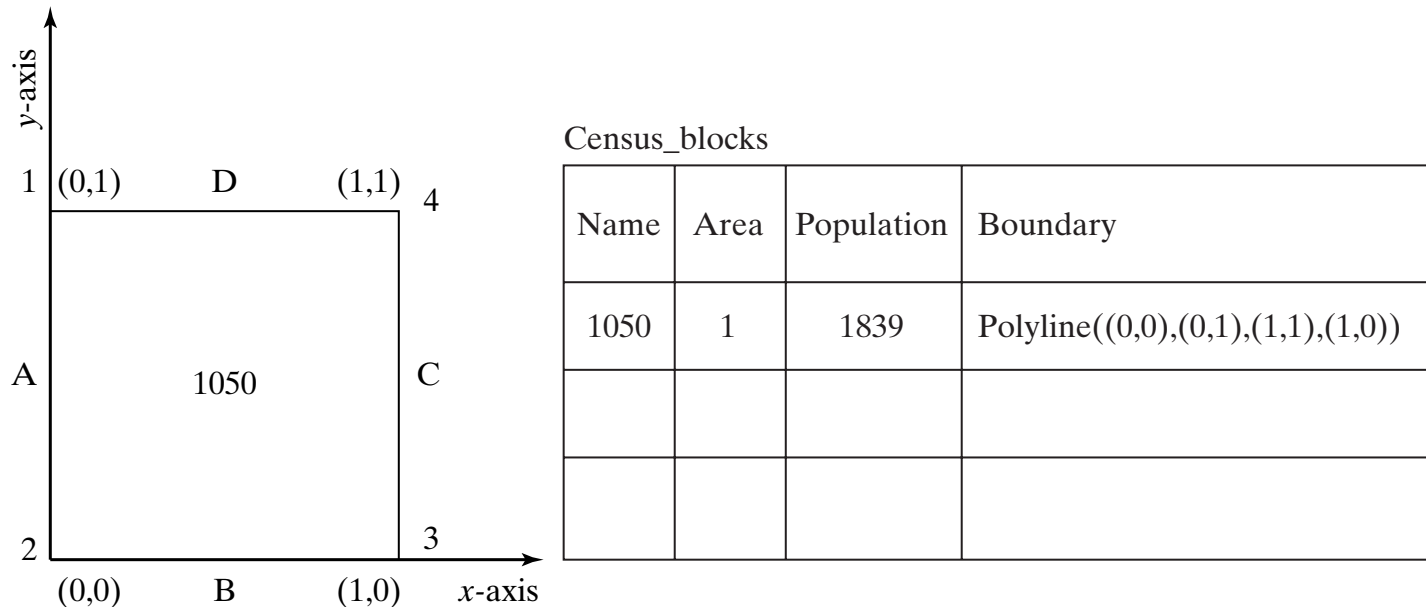


Figura 1.3

Tipos de Dados Espaciais e Bases de Dados Tradicionais

- As bases de dados relacionais tradicionais
 - Suportam tipos de dados simples, e.g. number, string, date
 - A modelação de tipos espaciais é morosa e complexa
- Exemplo: A Figura 1.4 mostra a modelação de um polígono com números
 - Três novas tabelas: polygon, edge, points
 - Nota: Um Polygon é uma polyline onde o primeiro e o último ponto são o mesmo
 - Um único e simples quadrado é representado por 16 tuplos em 3 tabelas
 - Uma operação espacial simples, e.g. area(), requer a junção de tabelas
 - É moroso e computacionalmente ineficiente

Mapeamento da “census_table” numa BD relacional

Census_blocks

Name	Area	Population	boundary-ID
340	1	1839	1050

Polygon

boundary-ID	edge-name
1050	A
1050	B
1050	C
1050	D

Edge

edge-name	endpoint
A	1
A	2
B	2
B	3
C	3
C	4
D	4
D	1

Point

endpoint	x-coor	y-coor
1	0	1
2	0	0
3	1	0
4	1	1

Fig 1.4

Evolução na tecnologia dos DBMS

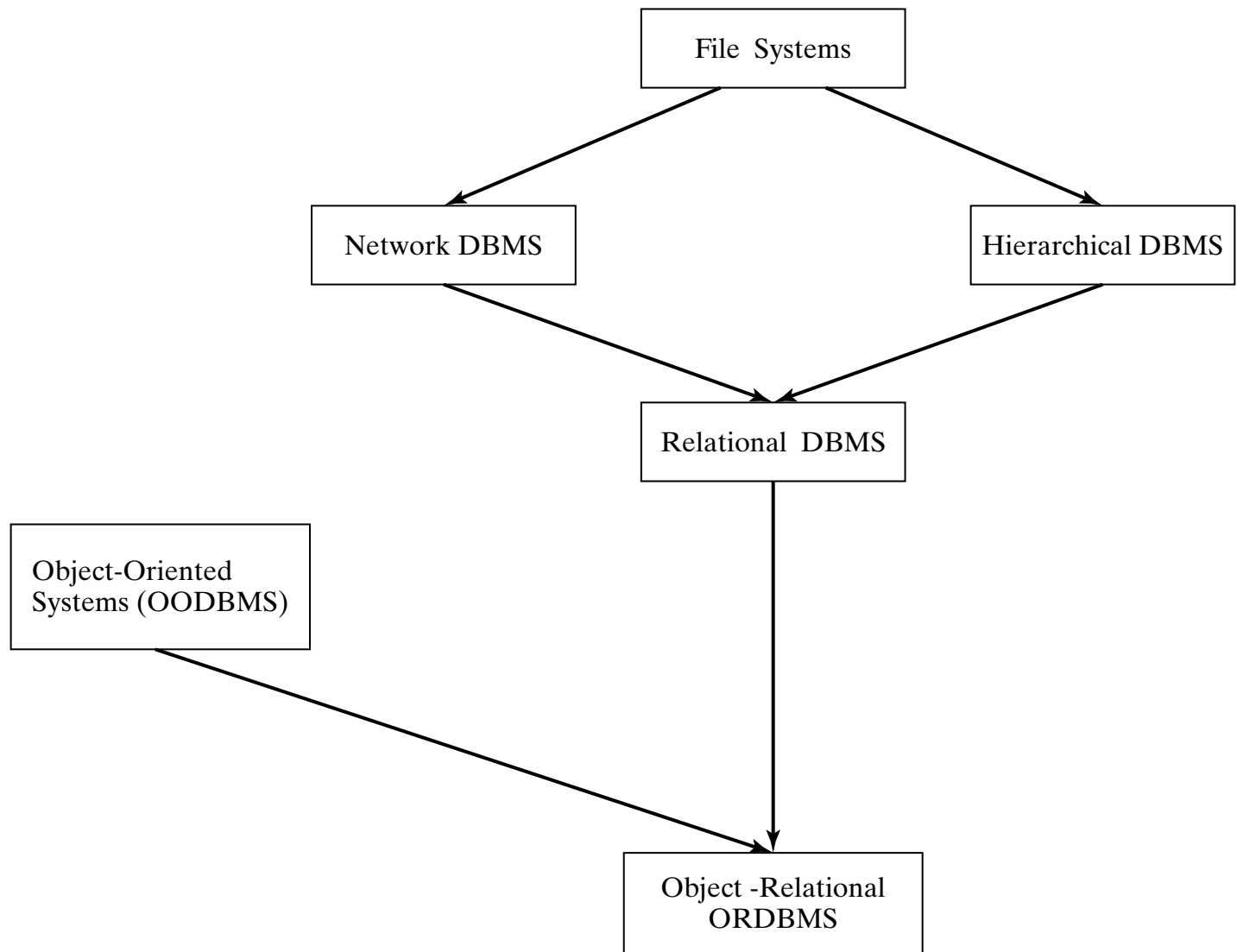


Fig 1.5

Tipos de dados espaciais e BD pós-relacionais

- DBMS pós-relacionais
 - Suportam tipos abstractos de dados definidos pelo utilizador
 - Tipos de dados espaciais (e.g. polygon) podem ser adicionados

- Alternativas em DBMS pós-relacionais
 - Object oriented (OO) DBMS
 - Object relational (OR) DBMS

- Uma base de dados espacial é um conjunto de tipos de dados espaciais, operadores, índices, estratégias de processamento, etc. e podem trabalhar com vários DBMS pós-relacionais, assim como com linguagens de programação como o C, Java, Visual Basic etc.

De que forma é um SDBMS diferente de um GIS ?

- Um GIS é um software para visualizar e analisar dados espaciais usando funções de análise espacial como
 - ❏ **Procura** Procura temática, procura por região, (re-)classificação
 - ❏ **Análise de localização** “Buffer”, “corridor”, “overlay”
 - ❏ **Análise de terreno** Declive/aspecto, drenagem
 - ❏ **Análise de fluxo** Conectividade, caminho mais curto
 - ❏ **Distribuição** Detecção de mudança, proximidade, “nearest neighbor”
 - ❏ **Análise espacial/Estatística** Padrão, centralidade, autocorrelação, índices de semelhança, topologia
 - ❏ **Medidas** Distância, perímetro, forma, adjacência, direcção
- Os GIS usam SDBMS
 - ❏ Para guardar, procurar, consultar, partilhar grandes conjuntos de dados espaciais

Quão diferente é um SDBMS de um GIS ?

- Os SDBMS focam-se em
 - ❑ Armazenamento, consulta e partilha eficiente de grandes conjuntos de dados espaciais
 - ❑ Fornece um conjunto de operações espaciais mais simples
 - ❑ Exemplos de operações: procura por região, "overlay", "nearest neighbor", distância, adjacência, perímetro, etc.
 - ❑ Uso de índices espaciais e optimizadores de consultas para acelerar as consultas sobre grandes conjuntos de dados espaciais.
- Os SDBMS podem ser utilizados por outras aplicações que não GIS
 - ❑ Astronomia, Genómica, Sistemas de Informação Multimédia, ...
- Será necessário um GIS ou um SDBM para a seguinte questão:
 - ❑ Quantos países vizinhos tem Portugal?
 - ❑ Que país tem o maior número de vizinhos?

Evolução do acrónimo “GIS”

- Geographic Information Systems (1980s)
- Geographic Information Science (1990s)
- Geographic Information Services (2000s)

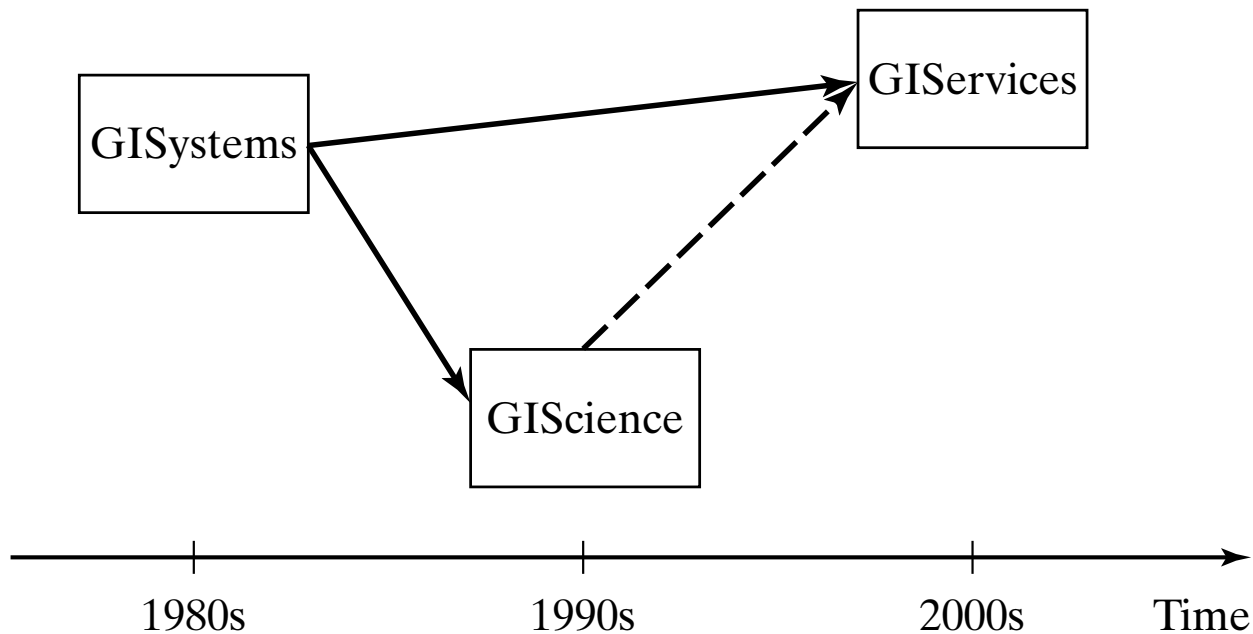


Fig 1.1

Três significados do acrónimo GIS

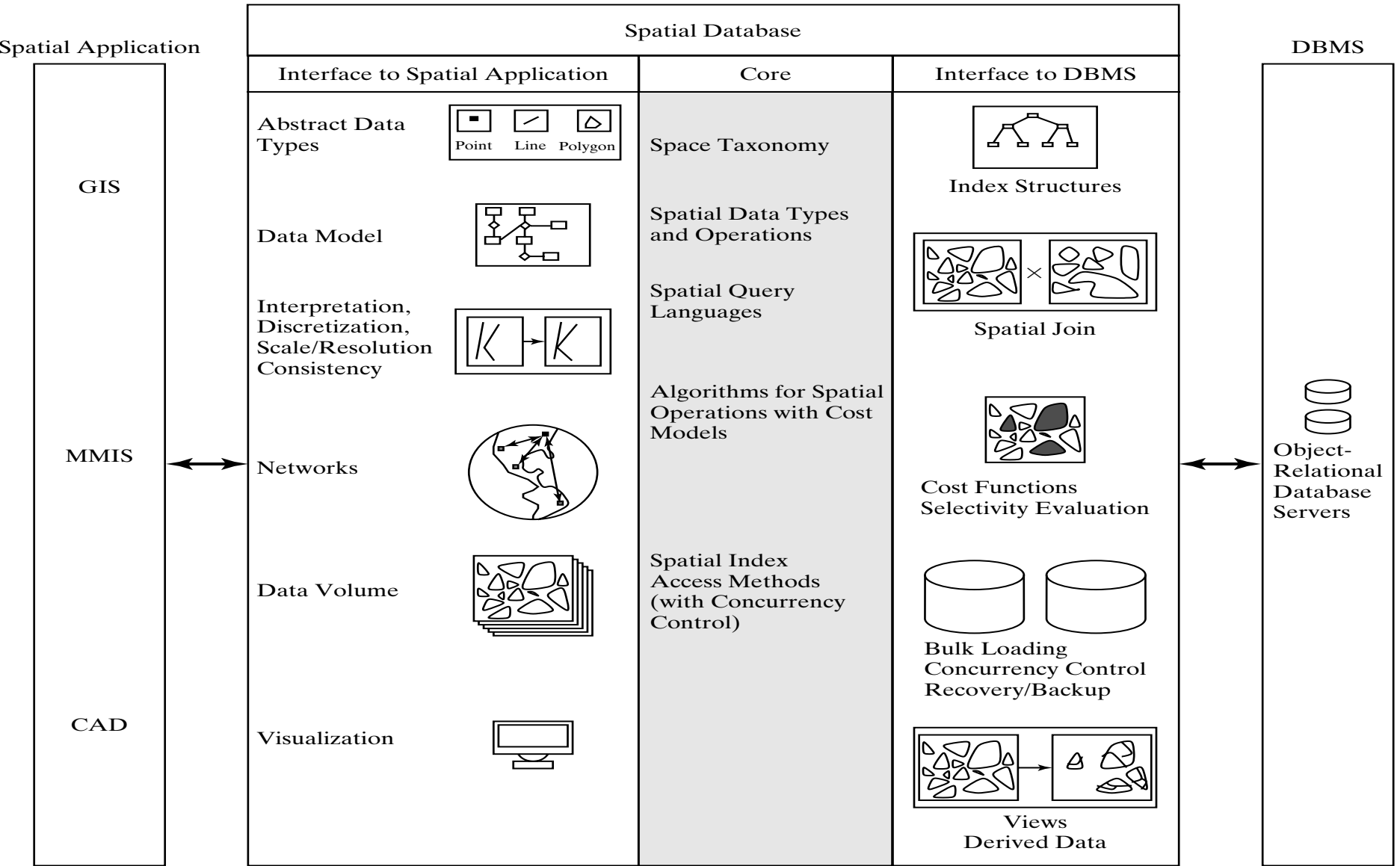
- Geographic Information Services
 - Web-sites e centro de serviços para utilizadores casuais, e.g. viajantes
 - Exemplo: Os serviços (e.g. AAA, mapquest) para planeamento de rotas
- Geographic Information Systems
 - Software para utilizadores profissionais, e.g. cartógrafos
 - Exemplo: software ESRI Arc/View
- Geographic Information Science
 - Conceitos e teorias para formalizar o uso e desenvolvimento de sistemas e serviços de informação geográfica
 - Exemplo: desenho de tipos de dados espaciais e operações de consulta

Componentes de um SDBMS

- Relembrando: um SDBMS é um módulo de software que
 - ❏ Pode trabalhar com uma camada de suporte dada por um DBMS
 - ❏ Suporta modelos de dados espaciais, ADTs espaciais e uma linguagem de consulta da qual estes ADTs podem ser chamados
 - ❏ Suporta indexação espacial, algoritmos para o processamento de operadores espaciais, e regras especializadas por domínios para otimização de consultas
- As componentes incluem
 - ❏ Um modelo de dados espaciais, uma linguagem de consulta, processamento de consultas, organização de ficheiros e índices, otimização de consultas, etc.
 - ❏ A Figure 1.6 mostra estas componentes

Arquitectura de 3 camadas

Fig 1.6



1.6.1 Taxonomia, Modelos de Dados

● Taxonomia Espacial:

- ❏ multitude de descrições disponíveis para organizar o espaço.
- ❏ A topologia modela relacionamentos homeomorficos, e.g. overlap (sobreposição)
- ❏ O espaço euclideano modela a distância e direcção no plano
- ❏ O grafos modelam conectividade, caminho mais curto

● Modelos de Dados Espaciais

- ❏ Regras para identificar objectos e propriedades do espaço
- ❏ Modelo de Objecto é orientado para a gestão de coisas identificáveis, e.g. montanhas, cidades, terrenos, etc.
- ❏ Modelo de Campo é orientado para a gestão de fenómenos contínuos e amorfos, e.g. zonas pantanosas, imagens de satélite, queda de neve, etc.

1.6.2 Linguagem de Consulta Espacial

- Linguagem de consulta espacial
 - Tipos de dados espaciais, e.g. point, linestring, polygon, ...
 - Operações espaciais, e.g. overlap, distance, nearest neighbor, ...
 - São chamáveis de uma linguagem de consulta (e.g. SQL3) do sistema de suporte (DBMS)

```
SELECT S.name  
FROM   Senator S  
WHERE  S.district.Area() > 300
```

- Standards
 - SQL3 (a.k.a. SQL 1999) é um standard para linguagem de consulta
 - OGIS é um standard para tipos de dados espaciais e operações
 - Ambos os standards gozam de ampla difusão

Exemplos de Consultas com várias passagens (multiscan)

- Exemplo de uma junção espacial

```
SELECT S.name      FROM Senator S, Business B
      WHERE S.district.Area() > 300 AND Within(B.location, S.district)
```

- Exemplo de uma junção não-espacial

```
SELECT S.name      FROM Senator S, Business B
      WHERE S.soc-sec = B.soc-sec AND S.gender = 'Female'
```

SENATOR

NAME	SOC.-SEC	GENDER	DISTRICT (POLYGON)
------	----------	--------	--------------------

Join

Spatial Join

BUSINESS

B-NAME	OWNER	SOC-SEC	LOCATION (POINT)
--------	-------	---------	------------------

Fig 1.7

1.6.3 Processamento de Consultas

- Algoritmos eficientes para calcular consultas espaciais
- Uma estratégia comum é – “filter and refine”
 - Passo de “filter”: Região “overlaps” c/ MBRs de B,C e D
 - Passo de “refine”: Região “overlaps” c/ B e C

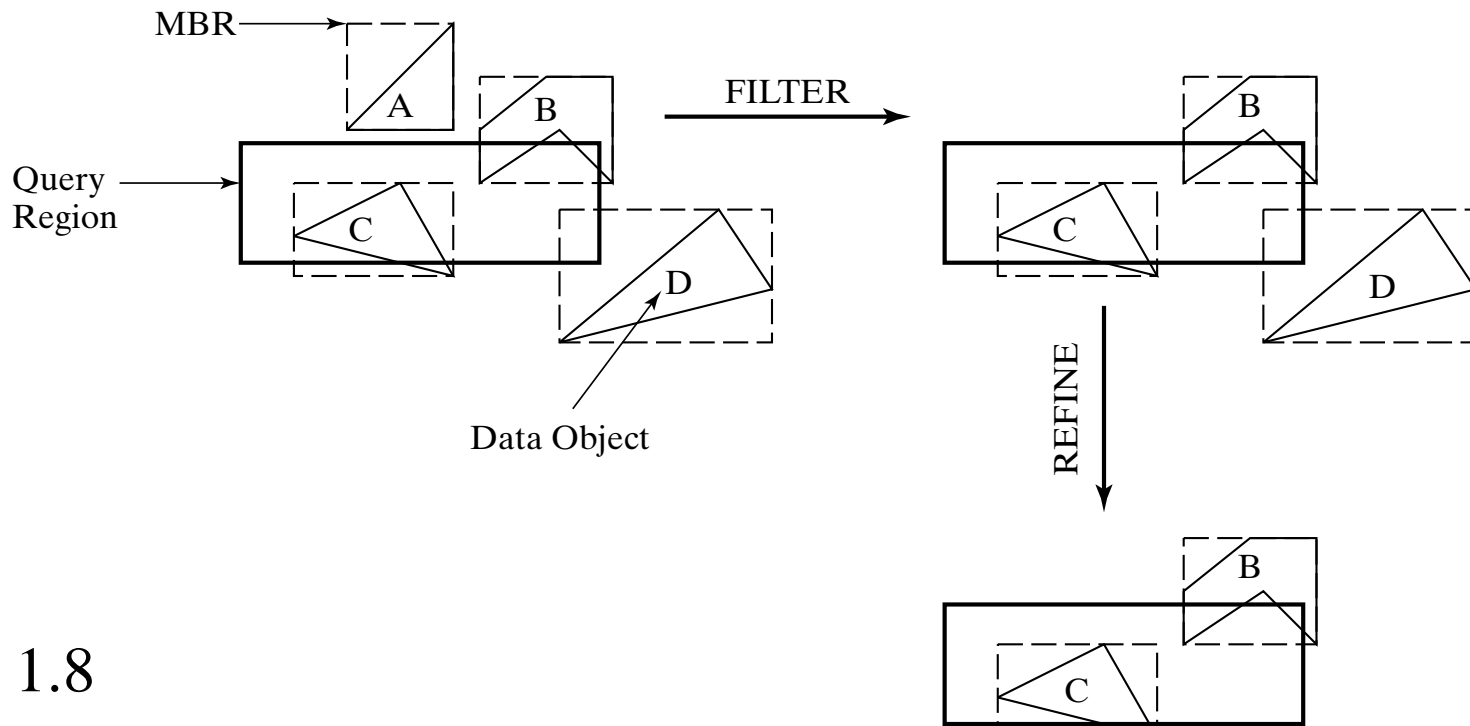


Fig 1.8

1.6.4 Organização de Ficheiros e Índices

- Pressupostos diferentes entre SDBMS e GIS
 - Algoritmos GIS: o “dataset” é carregado na memória principal (Fig. 1.10(a))
 - SDBMS: o dataset reside em memória secundária e.g disco (Fig. 1.10(b))
 - SDBMS usam curvas de preenchimento de espaço e índices espaciais
 - para procurarem de forma eficiente em grandes conjuntos de dados

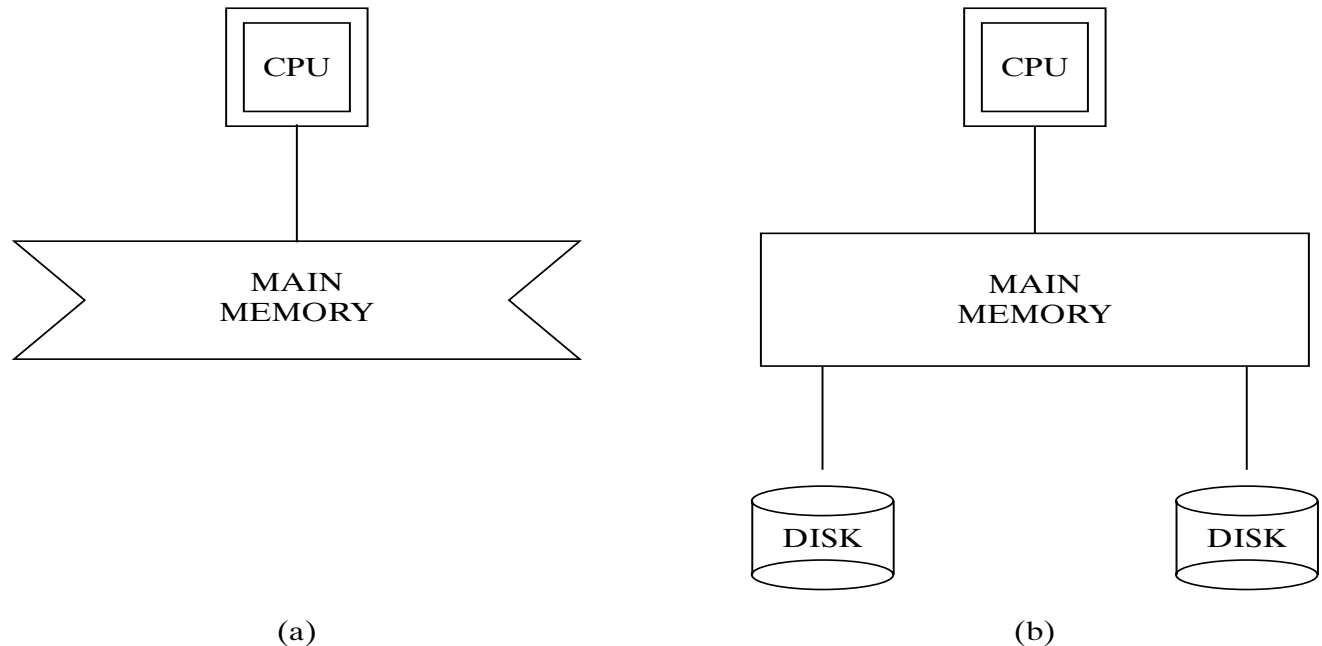


Fig 1.10

Organização de dados espaciais com curvas de preenchimento do espaço

•Problema:

- A ordenação não é definida naturalmente sobre dados espaciais
- Muito dos métodos de procura eficiente são baseados em ordenação

•“Space filling curves”

- Forçar uma ordem nas localizações num espaço multi-dimensional
- Exemplos: row-order (Fig. 1.11(a)), z-order (Fig 1.11(b))
- Assim permite-se o uso dos algoritmos tradicionais baseados em ordenação

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

(a)

7	8	14	16
5	6	13	15
2	4	10	12
1	3	9	11

(b)

Fig 1.11

Indexação Espacial

• Alternativas para indexação espacial:

- B-tree é um conjunto hierárquico de intervalos de chaves lineares, e.g. n^os
- Índice B-tree é usado para procura eficiente sobre dados tradicionais
- B-tree pode ser usada com “space filling curves” sobre dados espaciais
- R-tree permite uma performance superior com dados multi-dimensionais.
- R-tree é um conjunto hierárquico de rectângulos.

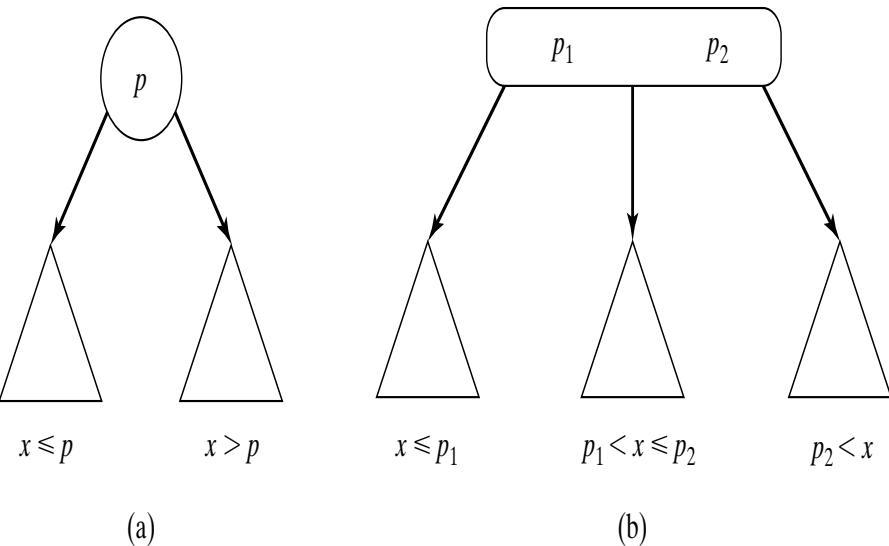


Fig 1.12: B-tree

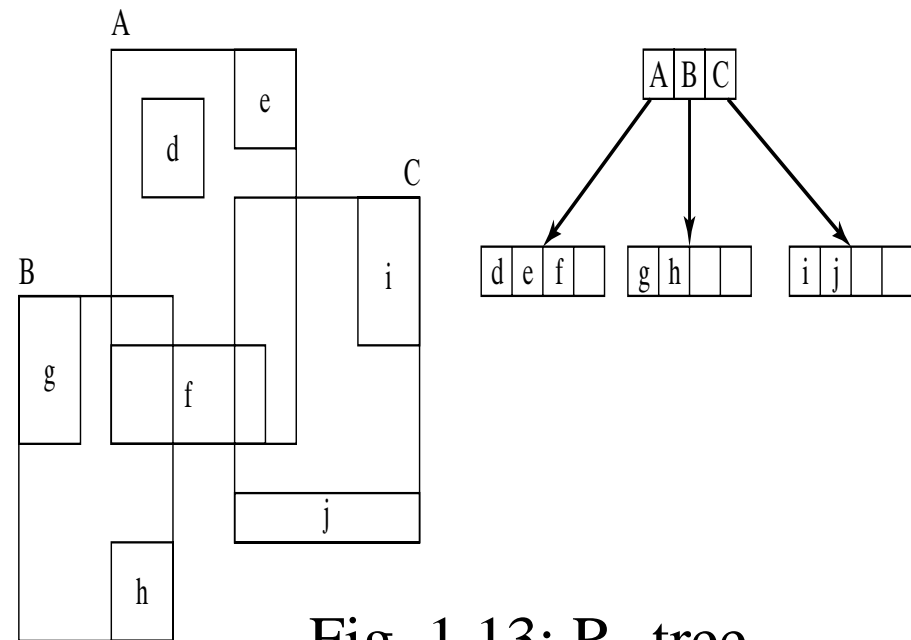


Fig. 1.13: R- tree

1.6.5 *Optimização de Consultas*

- **Optimização de Consultas**

- Uma operação espacial pode ser executada usando diferentes estratégias
- O custo computacional de cada estratégia depende de vários parâmetros
- A otimização de consultas é o processo de
 - Ordenar as operações numa consulta e
 - Seleccionar a melhor estratégia para cada operação baseado nos detalhes de cada dataset

- **Consulta exemplo:**

```
SELECT S.name      FROM Senator S, Business B
WHERE S.soc-sec = B.soc-sec AND S.gender = 'Female'
```

- **Exemplos de decisões de optimização**

- Processar (S.gender = 'Female') antes de (S.soc-sec = B.soc-sec)
- Não usar indexação para processar (S.gender = 'Female')

1.6.6 Data Mining

- Temos vários tipos de análise da dados espaciais
 - “Deductive Querying”, e.g. procura, ordenação, sobreposição
 - “Inductive Mining”, e.g. estatísticas, correlação, “clustering”, classificação, ...
- “Data mining” é a procura sistemática e semi-automática de padrões não triviais e interessantes em grandes bases de dados espaciais
- Aplicações-exemplo incluem
 - Inferir a ocupação de solo através de imagens de satélite
 - Identificar zonas de alta ocorrência de cancro e factores geográficos altamente correlacionados
 - Identificar zonas com altos índices criminais para escalonar patrulhas policiais e assistentes sociais