Chapter 5

Spatial database technology

-Niraj K.C.

Table of Content

Introduction

In various fields there is a need to manage geometric, geographic, or spatial data, which means data related to space.

A spatial database system is a database system.

- It offers spatial data types (SDTs) in its data model and query language.
- It supports spatial data types in its implementation, providing at least spatial indexing and efficient algorithms for spatial join.

Value of SDBMS

- Traditional (non-spatial) database management systems provide:
  - Persistence/Determination across failures
  - Allows concurrent access to data
  - Scalability to search queries on very large datasets which do not fit inside main memories of computers
  - Efficient for non-spatial queries, but not for spatial queries
- Non-spatial queries:
  - List the names of all bookstore with more than ten thousand titles.
  - List the names of ten customers, in terms of sales, in the year 2001
  - Use an index to narrow down the search
- Spatial Queries:
  - List the names of all bookstores with ten miles of Minneapolis
  - List all customers who live in Tennessee and its adjoining states
  - List all the customers who reside within fifty miles of the company headquarter

Value of SDBMS – Spatial Data Examples

- Examples of non-spatial data
  - Names, phone numbers, email addresses of people
- Examples of Spatial data
  - Census Data
  - NASA satellites imagery - terabytes of data per day
  - Weather and Climate Data
  - Rivers, Farms, ecological impact
  - Medical Imaging
- Assignment 1:

Identify spatial and non-spatial data items in

- A phone book
- A Product catalog

Value of SDBMS – Users, Application Domains

- Many important application domains have spatial data and queries. Some Examples follow:
  - **Army Field Commander**: Has there been any significant enemy troop movement since last night?
  - **Insurance Risk Manager**: Which homes are most likely to be affected in the next great flood on the Mississippi?
  - **Medical Doctor**: Based on this patient's MRI, have we treated somebody with a similar condition ?
  - **Molecular Biologist**:Is the topology of the amino acid biosynthesis gene in the genome found in any other sequence feature map in the database ?
  - **Astronomer**:Find all blue galaxies within 2 arcmin of quasars.

- Assignment 2 :

List two ways you have used spatial data. Which software did you use to manipulate spatial data?

What is a SDBMS ?

- A SDBMS is a software module that
  - can work with an underlying DBMS
  - supports spatial data models, spatial abstract data types (ADTs) and a query language from which these ADTs are callable
  - supports spatial indexing, efficient algorithms for processing spatial operations, and domain specific rules for query optimization
- Example: Oracle Spatial data cartridge, ESRI SDE
  - can work with Oracle 8i DBMS
  - Has spatial data types (e.g. polygon), operations (e.g. overlap) callable from SQL3 query language
  - Has spatial indices, e.g. R-trees
- IBM: Spatial Option
- Informix: Spatial Datablade

SDBMS Example

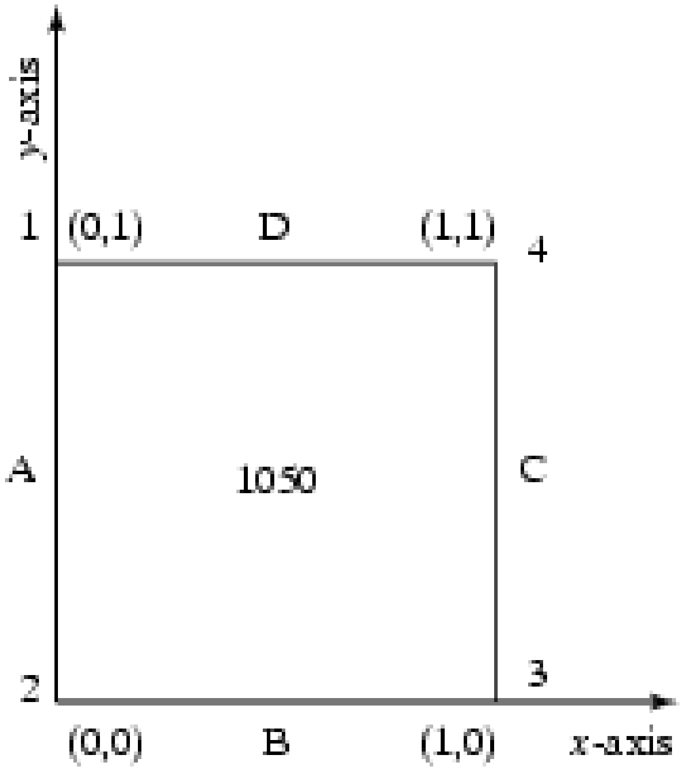
- Consider a spatial dataset with:
  - County boundary (dashed white line)
  - Census block - name, area, population, boundary (dark line)
  - Water bodies (dark polygons)
  - Satellite Imagery (gray scale pixels)

- Storage in a SDBMS table:

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

Modeling Spatial Data in Traditional DBMS

- A row in the table census_blocks (Figure below)
- Question: Is Polyline datatype supported in DBMS?



Census_blocks

Name	Area	Population	Boundary
1050	1	1839	Polyline((0,0),(0,1),(1,1),(1,0))

Spatial Data Types and Traditional Databases

- Traditional relational DBMS
  - Support simple data types, e.g. number, strings, date
  - Modeling Spatial data types is tedious
- Example: Fig above shows modeling of polygon using numbers
  - Three new tables: polygon, edge, points

    Note: Polygon is a polyline where last point and first point are same
  - A simple unit square represented as 16 rows across 3 tables
  - Simple spatial operators, e.g. area(), require joining tables
  - Tedious and computationally inefficient

- Assignment 3

Question. Name post-relational database management systems which facilitate modeling of spatial data types, e.g. polygon. How?

Mapping “census_table” into a Relational Database

Census_blocks

Name	Area	Population	boundary-ID
340	1	1839	1050

Polygon

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

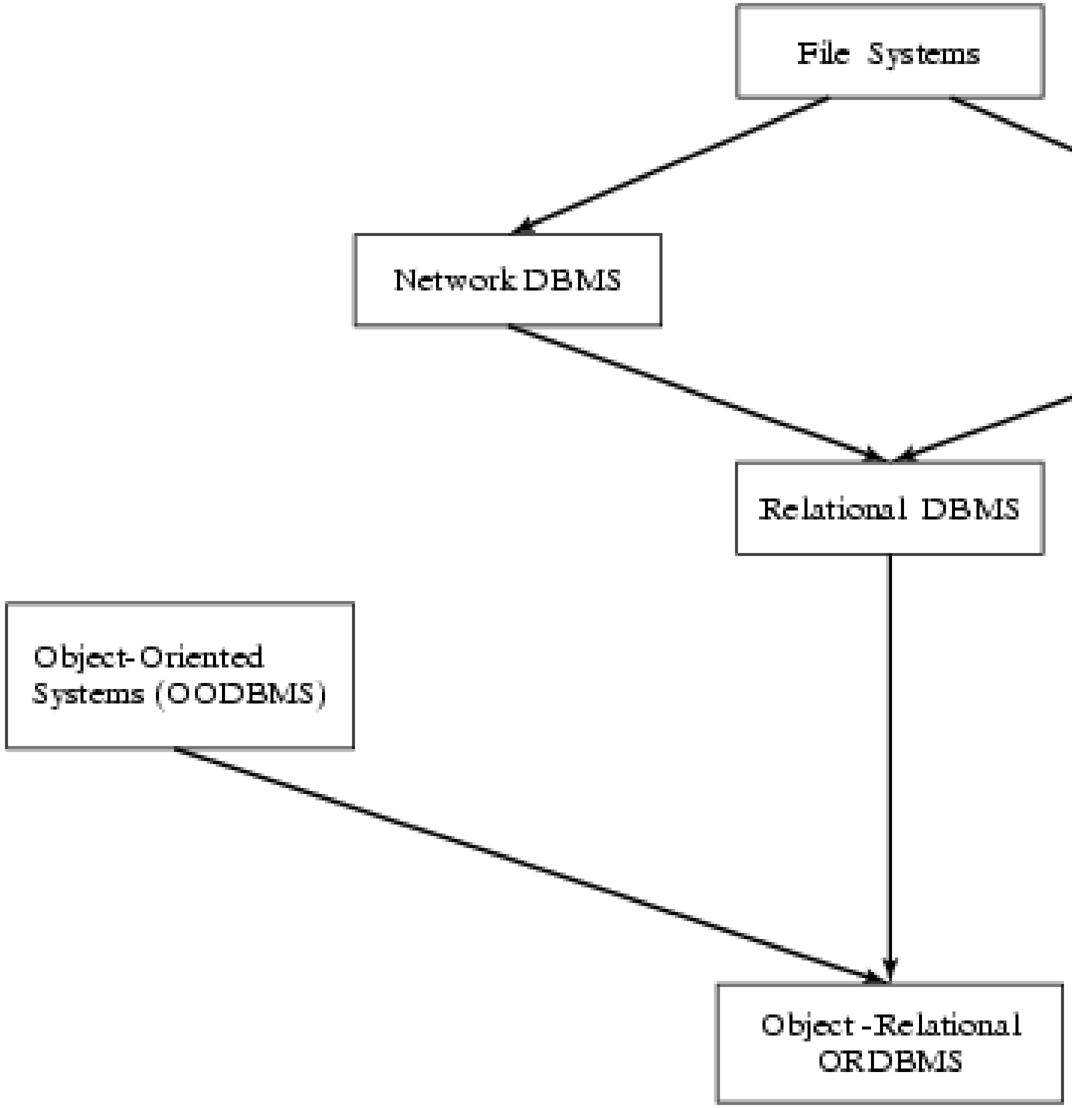
Edge

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

Evolution of DBMS technology



Spatial Data Types and Post-relational Databases

- Post-relational DBMS
  - Support user defined abstract data types
  - Spatial data types (e.g. polygon) can be added
- Choice of post-relational DBMS
  - Object oriented (OO) DBMS
  - Object relational (OR) DBMS
- A spatial database is a collection of spatial data types, operators, indices, processing strategies, etc. and can work with many post-relational DBMS as well as programming languages like Java, Visual Basic etc.

An Example : OO-Model

- Class Highway

  tuple (highway_name: string, highway_type: string,
  sections: list(Section))
- Class Section

  tuple (section_name: string,
  number_lanes: integer,
  city_start: City,
  city_end: City,
  geometry: Line)
- Class City

  Tuple (city_name: string,
  population: integer,
  geometry: Region)

An Example : OO-Model

- Method

  Method length in class Line: real (Computes the length of a line.)

- Query: Length of Interstate 99 (I99)

```
Sum (Select s. geometry->length()
          from h in All_highways, s in h.sections
          where h.highway_name = 'I99')
```

How is a SDBMS different from a GIS？

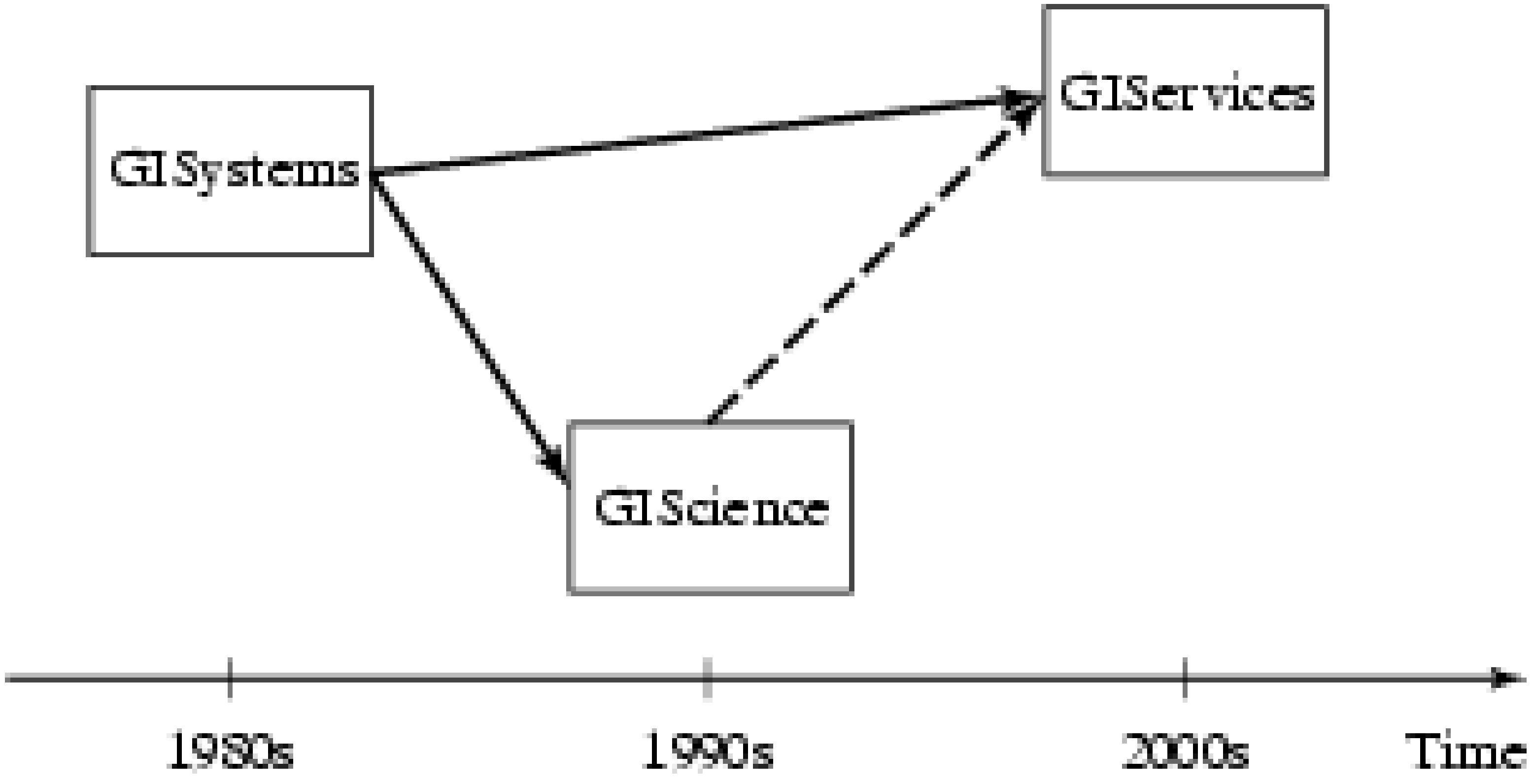
- GIS is a software to visualize and analyze spatial data using spatial analysis functions such as
  - **Search** Thematic search, search by region, (re-)classification
  - **Location analysis** Buffer, corridor, overlay
  - **Terrain analysis** Slope/aspect, catchment, drainage network
  - **Flow analysis** Connectivity, shortest path
  - **Distribution** Change detection, proximity, nearest neighbor
  - **Spatial analysis/Statistics** Pattern, centrality, autocorrelation, indices of similarity, topology: hole description
  - **Measurements** Distance, perimeter, shape, adjacency, direction
- GIS uses SDBMS
  - to store, search, query, share large spatial data sets

How is a SDBMS different from a GIS ?

- SDBMS focuses on
  - Efficient storage, querying, sharing of large spatial datasets
  - Provides simpler set based query operations
  - Example operations: search by region, overlay, nearest neighbor, distance, adjacency, perimeter etc.
  - Uses spatial indices and query optimization to speedup queries over large spatial datasets.
- SDBMS may be used by applications other than GIS
  - Astronomy, Genomics, Multimedia information systems, ...

Evolution of acronym "GIS"

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



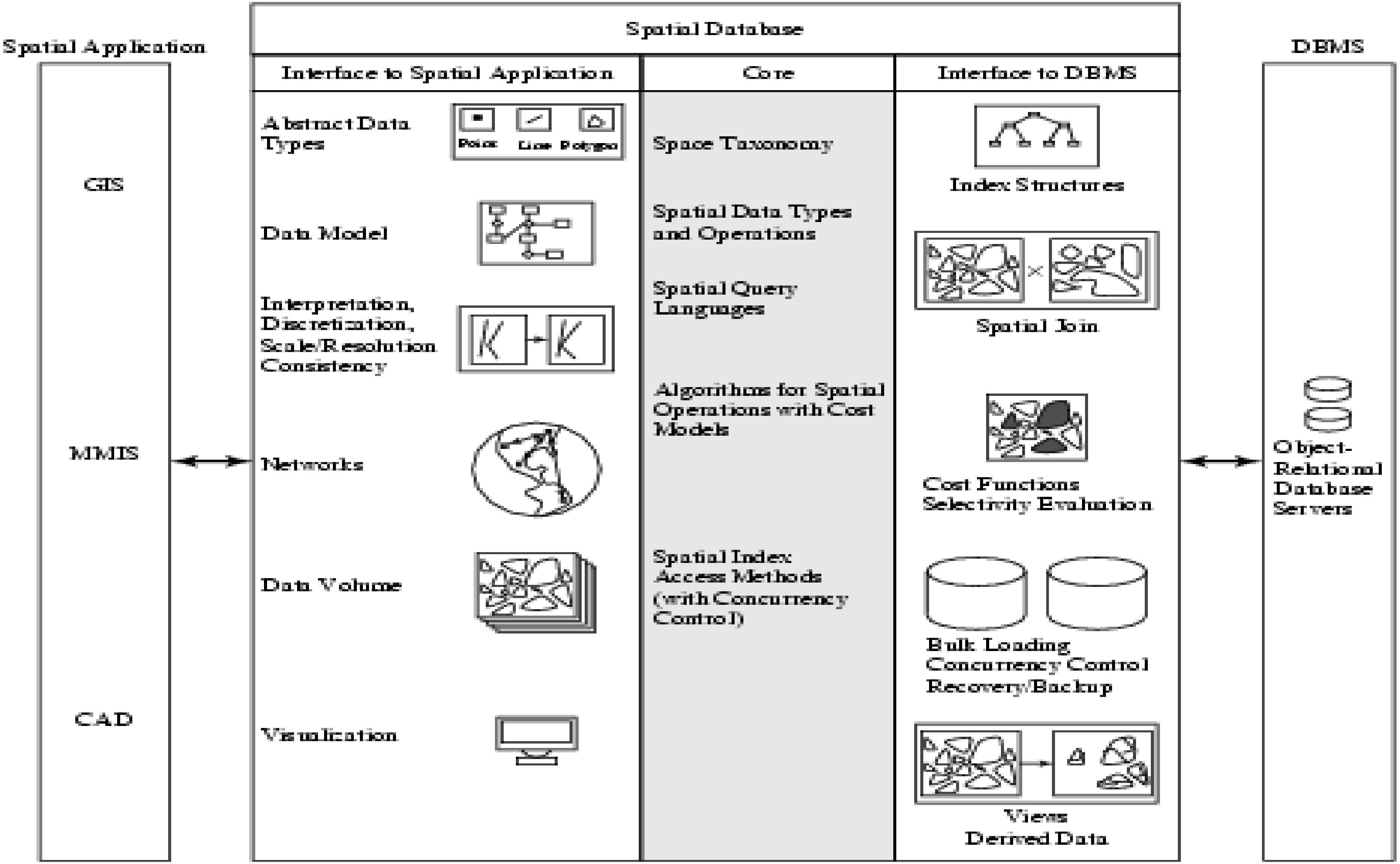
Three meanings of the acronym GIS

- Geographic Information Services
  - Web-sites and service centers for casual users, e.g. travelers
  - Example: Service (e.g. AAA, mapquest) for route planning
- Geographic Information Systems
  - Software for professional users, e.g. cartographers
  - Example: ESRI Arc/View software
- Geographic Information Science
  - Concepts, frameworks, theories to formalize use and development of geographic information systems and services
  - Example: design spatial data types and operations for querying

Components of a SDBMS

- Recall: a SDBMS is a software module that
  - can work with an underlying DBMS
  - supports spatial data models, spatial ADTs and a query language from which these ADTs are callable
  - supports spatial indexing, algorithms for processing spatial operations, and domain specific rules for query optimization
- Components include
  - spatial data model, query language, query processing, file organization and indices, query optimization, etc.

Three Layer Architecture



Spatial Taxonomy, Data Models

➢Spatial Taxonomy:

- multitude of descriptions available to organize space.
- Topology models homeomorphic relationships, e.g. overlap, adjacent
- Euclidean space models distance and direction in a plane
- Graphs models connectivity, Shortest-Path

➢Spatial data models

- rules to identify identifiable objects and properties of space
- Object model help manage identifiable things, e.g. mountains, cities, land-parcels etc.
- Field model help manage continuous and amorphous phenomenon, e.g. temperature, satellite imagery, snowfall etc.

Spatial Query Language

- Spatial query language
  - Spatial data types, e.g. point, linestring, polygon, …
  - Spatial operations, e.g. overlap, distance, nearest neighbor, …
  - Callable from a query language (e.g. SQL3) of underlying DBMS

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

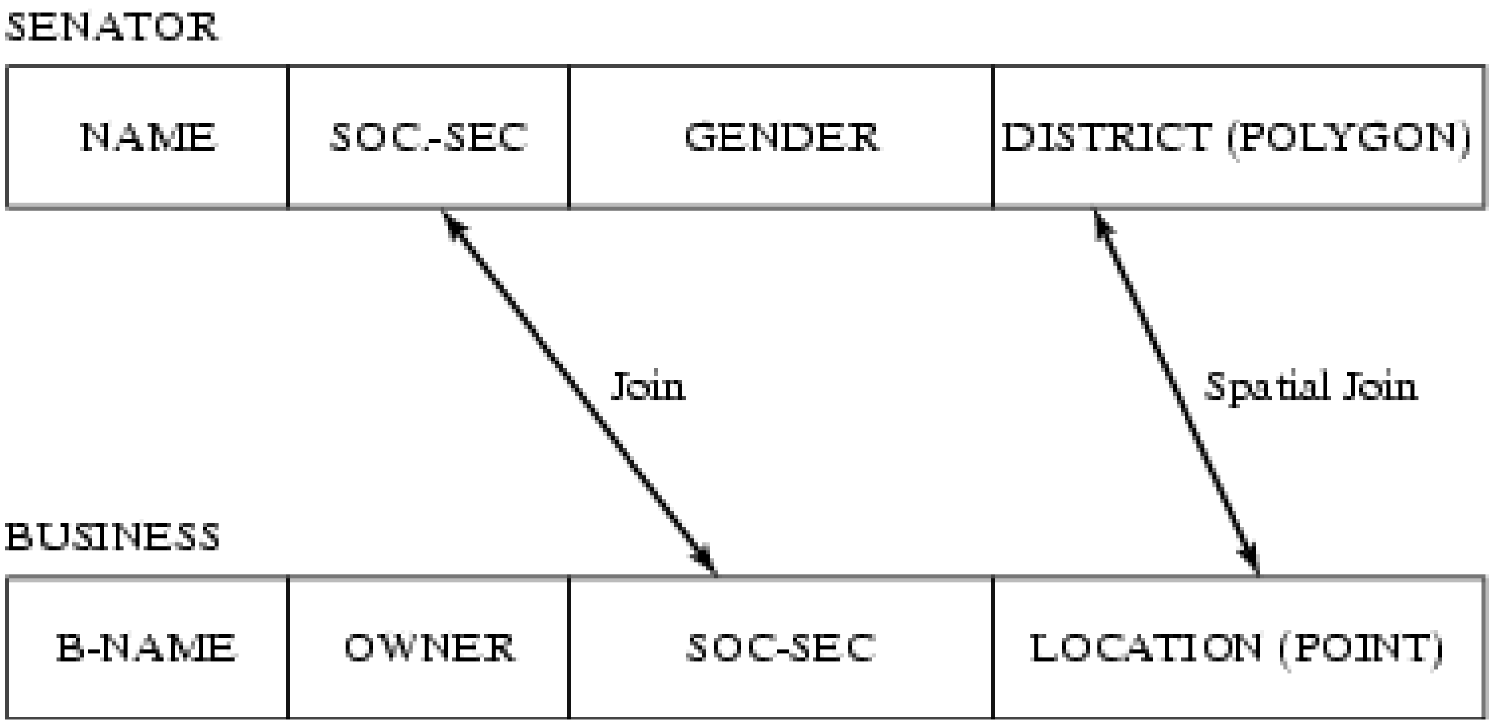
- Standards
  - SQL3 (a.k.a. SQL 1999) is a standard for query languages
  - OGIS is a standard for spatial data types and operators
  - Both standards enjoy wide support in industry

Multi-scan Query Example

- Find all senators who serve a district of are greater than 300 square miles and who own a business within the district
- Spatial join example

  SELECT S.name FROM Senator S, Business B

  WHERE S.district.Area() > 300 AND Within(B.location, S.district)

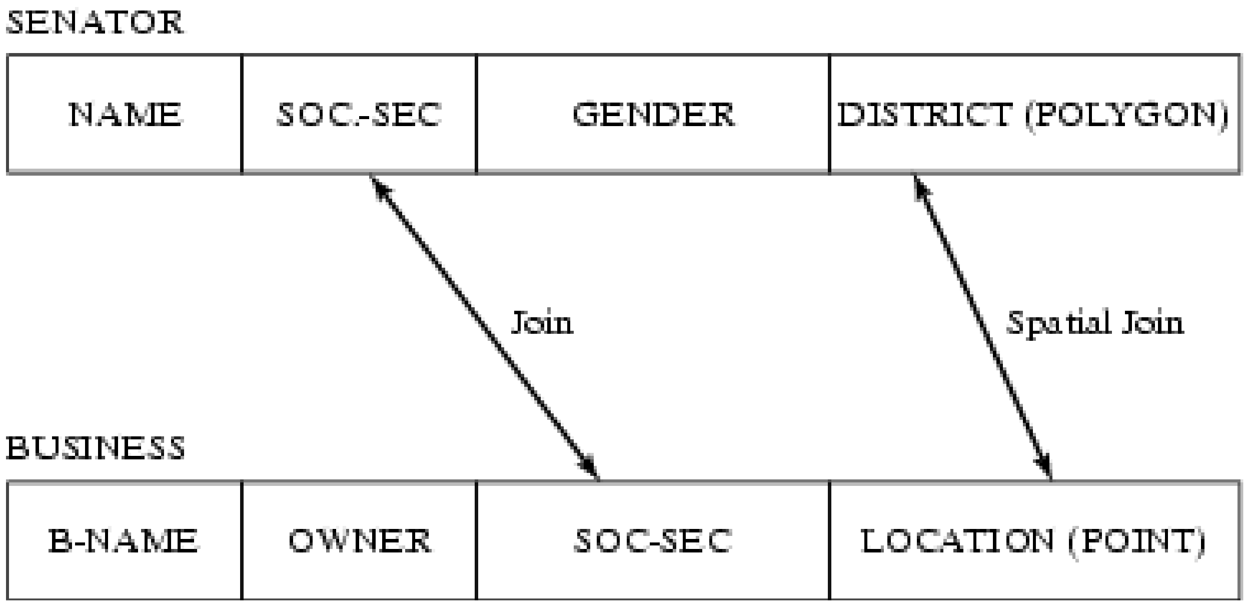


Multi-scan Query Example

- Find the name of all female senators who own a business
- Non-Spatial Join example

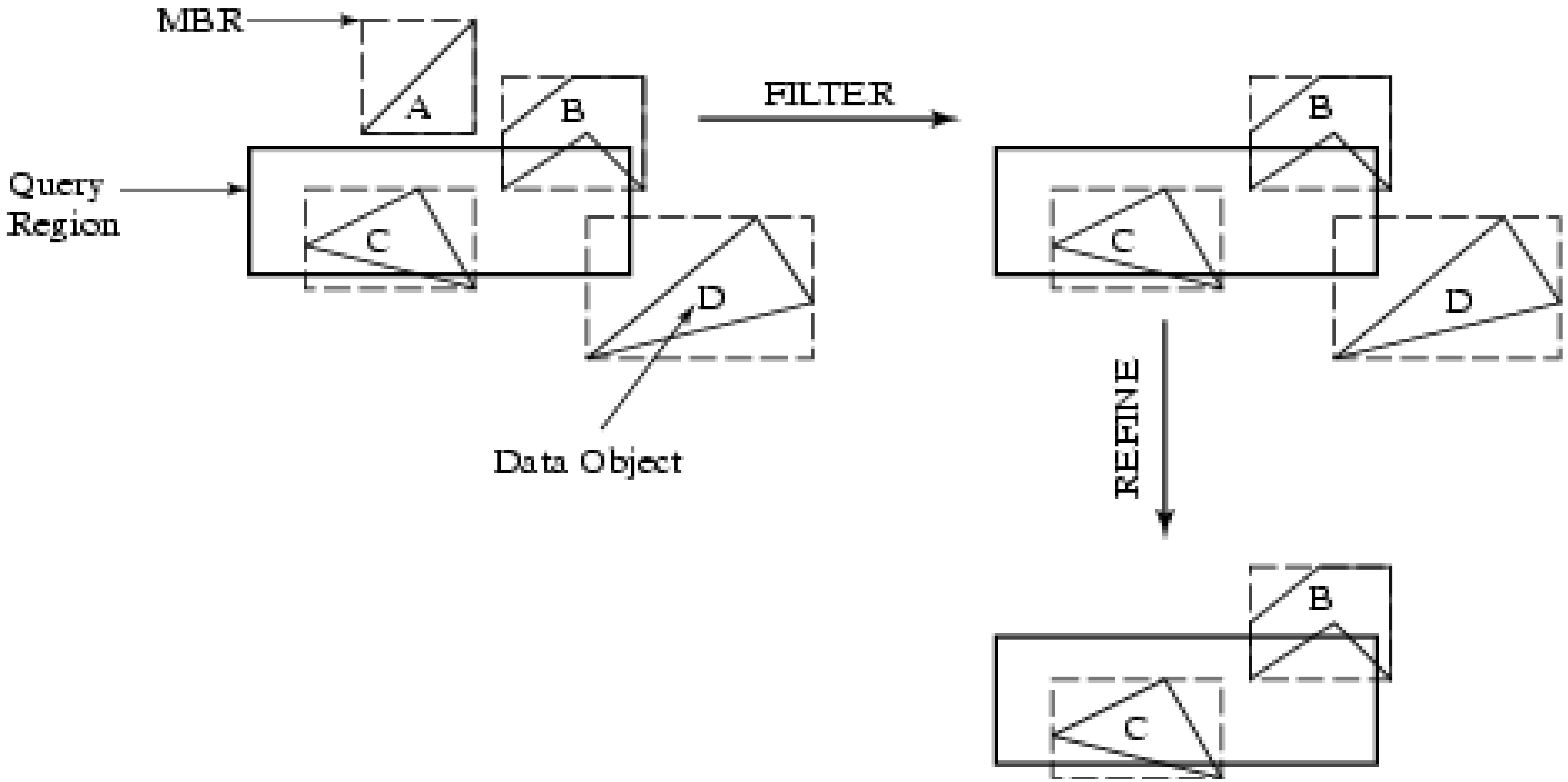
  SELECT S.name FROM Senator S, Business B

  WHERE S.soc-sec = B.soc-sec AND S.gender = 'Female'



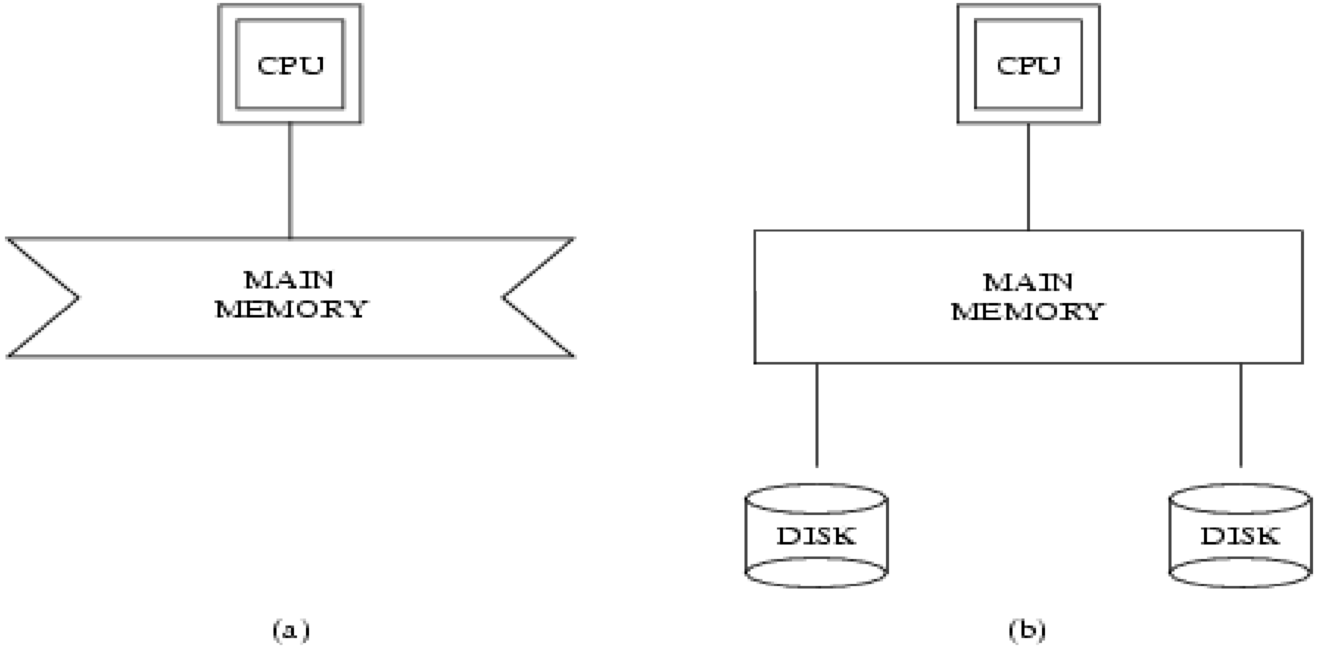
Query Processing

- Efficient algorithms to answer spatial queries
- Common Strategy - filter and refine
  - Filter Step: Query Region overlaps with MBRs of B,C and D
  - Refine Step: Query Region overlaps with B and C



File Organization and Indices

- A difference between GIS and SDBMS assumptions
  - GIS algorithms: dataset is loaded in main memory (Fig. (a))
  - SDBMS: dataset is on secondary storage e.g disk (Fig. (b))
  - SDBMS uses space filling curves and spatial indices
    - to efficiently search disk resident large spatial datasets
- Memory access: 10ns, disk access 10,000,000ns (2000)



Organizing spatial data with space filling curves

- Issue:
  - Sorting is not naturally defined on spatial data
  - Many efficient search methods are based on sorting datasets
- Space filling curves
  - Impose an ordering on the locations in a multi-dimensional space
  - Examples: row-order (Fig. (a), z-order (Fig . (b))
  - Allow use of traditional efficient search methods on spatial data

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

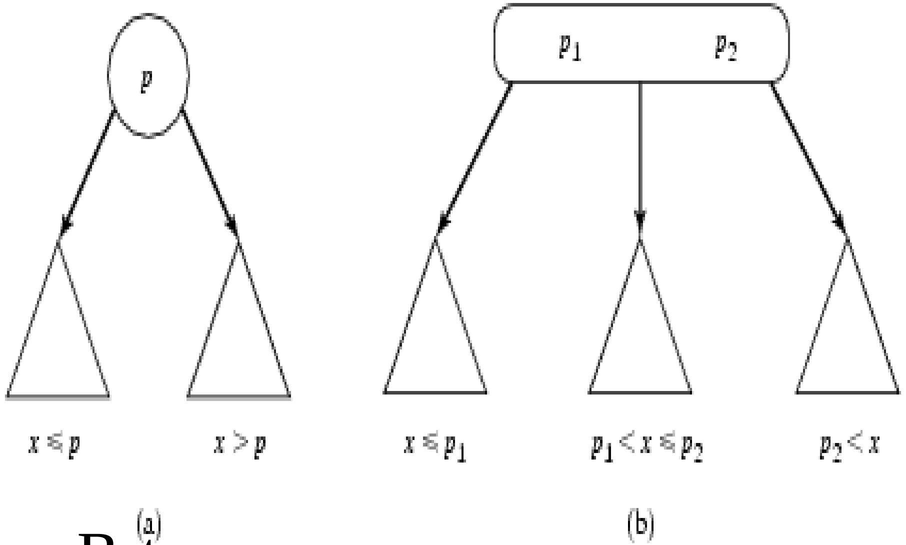
(a)

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

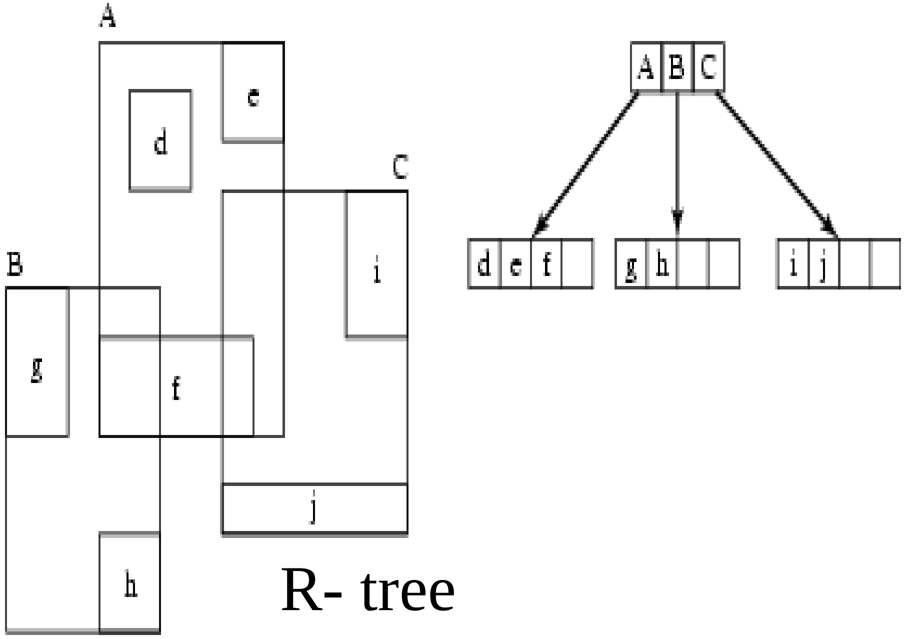
(b)

Spatial Indexing: Search Data-Structures

- Choice for spatial indexing:
  - B-tree is a hierarchical collection of ranges of linear keys, e.g. numbers
  - B-tree index is used for efficient search of traditional data
  - B-tree can be used with space filling curve on spatial data
  - R-tree provides better search performance yet!
  - R-tree is a hierarchical collection of rectangles



B-tree



R- tree

Query Optimization

- Query Optimization
  - A spatial operation can be processed using different strategies
  - Computation cost of each strategy depends on many parameters
  - Query optimization is the process of
    - ordering operations in a query and
    - selecting efficient strategy for each operation
    - based on the details of a given dataset
- Example Query:

  SELECT S.name FROM Senator S, Business B

  WHERE S.soc-sec = B.soc-sec AND S.gender = ‘Female’
- Optimization decision examples
  - Process (S.gender = ‘Female’) before (S.soc-sec = B.soc-sec)
  - Do not use index for processing (S.gender = ‘Female’)

Data Mining

- Analysis of spatial data is of many types
  - Deductive Querying, e.g. searching, sorting, overlays
  - Inductive Mining, e.g. statistics, correlation, clustering, classification, …
- Data mining is a systematic and semi-automated search for interesting non-trivial patterns (Significant patterns) in large spatial databases

- Example applications include
  - Infer land-use classification from satellite imagery
  - Identify cancer clusters and geographic factors with high correlation
  - Identify crime hotspots to assign police patrols and social workers

Summary

- SDBMS is valuable to many important applications
- SDBMS is a software module
  - works with an underlying DBMS
  - provides spatial ADTs callable from a query language
  - provides methods for efficient processing of spatial queries
- Components of SDBMS include
  - spatial data model, spatial data types and operators,
  - spatial query language, processing and optimization
  - spatial data mining
- SDBMS is used to store, query and share spatial data for GIS as well as other applications

Assignment 5

1. What do you understand by Data Mining? Describe the applications of Data Mining in SDBMS in detail with suitable exampl
2. Describe the query optimization process in SDBMS in detail with suitable example.
3. What do you understand by Spatial Indexing? List out and describe different types of Spatial indexing in detail with sui diagram.
4. What is space filling curves? Describe organizing spatial data with space filling curves in detail with suitable example.
5. What is Spatial taxonomy? Describe spatial data model and its type in detail with suitable example.
6. What is query processing? Describe the common strategy of Query processing in detail.
7. What do you understand by spatial query? Describe the three layer architecture of SDBMS in detail with suitable diagram.
8. Discuss the differences between Spatial and non-spatial data.
9. Geographic applications are a common source of spatial data . List at least four other important sources of spatial data.
10. What are the advantages of storing spatial data in a DBMS as opposed to file system.
11. How can object relational databases be used to implement an SDBMS?
12. Compare and Contrast

a. GIS Vs SDBMs
b. OODBMS vs RDBMS
c. GI Systems VS GI services
d. Data Model VS Query Language
e. Query Processing VS File organization and indices
f. Querying and Data Mining
g. Main Memory Vs DISK