M.Tech Geoinformatics

INTRODUCTION TO SPATIAL DATABASE

Prakash. K
Guest Faculty
Department of Geography
Bharathidasan University, Tiruchirappalli.

Introduction

- A SDBMS is a software module that can work with an underlying database management system, for example,
  - An Relational Database Management System, or
  - An Object-Oriented Database Management System
- SDBMS supports multiple spatial data models
- A data model is a conceptual description of the database
- SDBMS supports
  - spatial indexing
  - domain-specific rules for query optimization, and
  - efficient algorithms for spatial operations

...

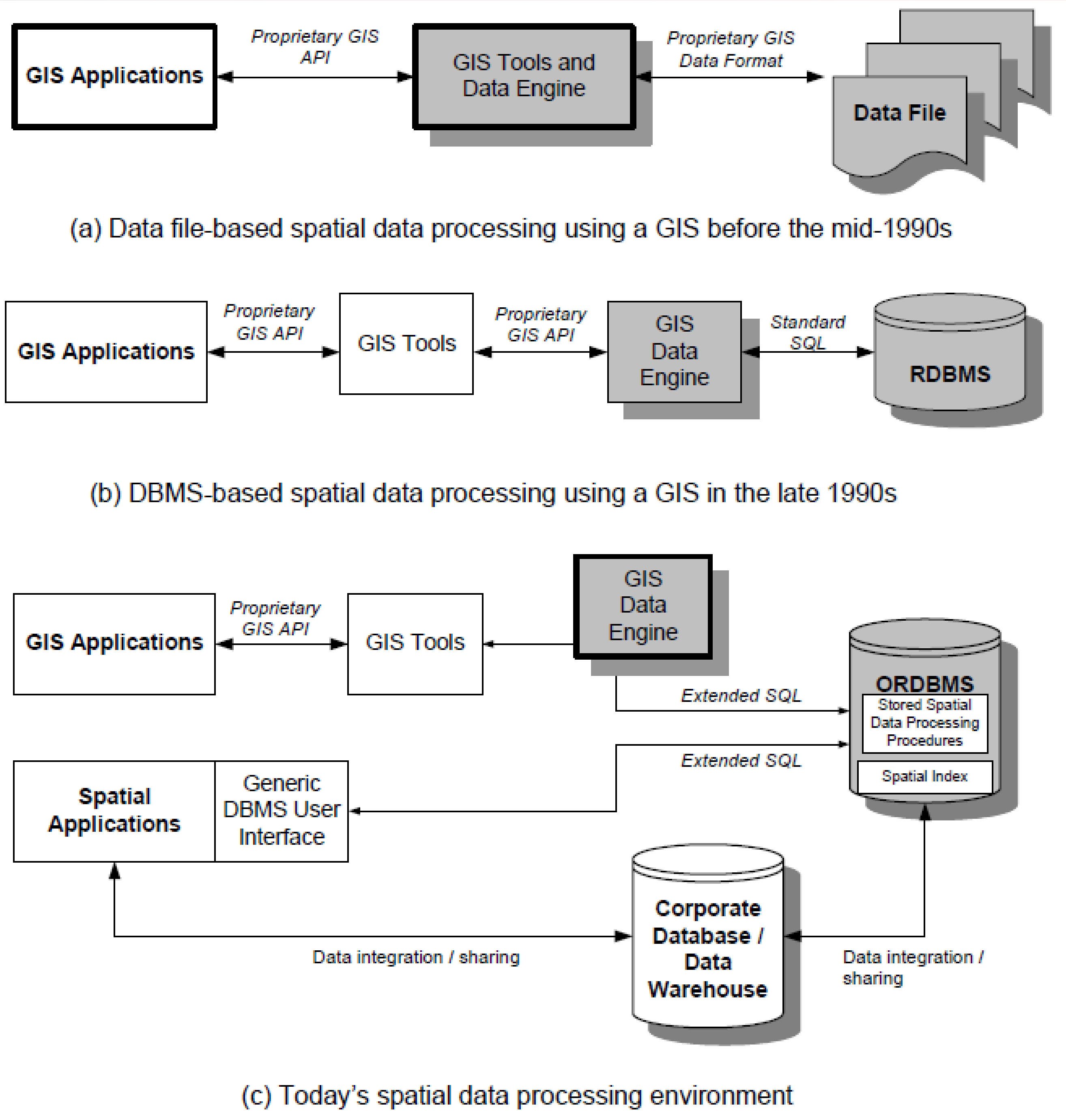
- The physical layout of the database, which describes how data are organised and stored in the database, is called the **database schema** or simply the **schema**
- A **database engine**, which is also commonly referred to as a **database server,** like,
  - Microsoft Jet Database Engine and SQL Server,
  - Oracle Database Server, and
  - IBM DB2 Universal Server

  are some of the computer programs that manipulate the data in a database

...

- Güting (1994) defined spatial database systems as a class of database systems that have the following three characteristics:
  - A spatial database system is a database system
  - It offers spatial data types (SDT) 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 joins

SDBMS Evolution



Characteristics of Spatial Database Systems

1. Spatial Data Types

- Several database software vendors have made use of this capability to define spatial data types (SDT) that enable their products for managing and processing spatial data.

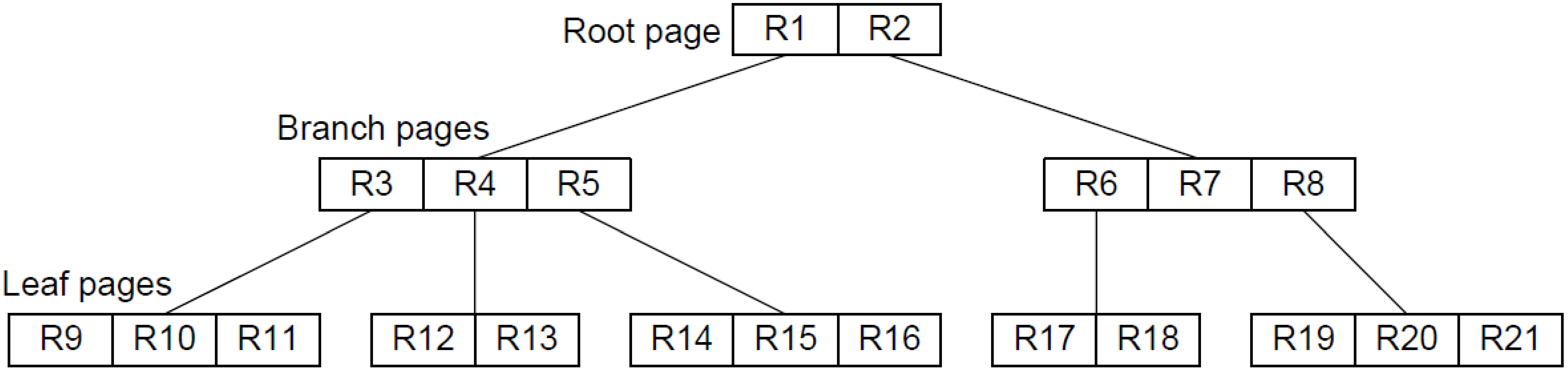
Oracle
- Points
- point clusters
- Line strings
- Polygons
- Arc strings
- Arc polygons
- Compound polygons
- Compound line string
- Circle
- Optimised rectangles

IBM DB2
- Point
- Multi-point
- Line string
- Multi-line strings
- Polygons
- Multi-polygons
- Ellipses

ESRI
- Point
- Multi-point
- Polyline
- Polygon
- Malti-patch
- Turns
- Dimension
- Annotation

2. Spatial Data Indexing and Access Method

- A fundamental concept of spatial indexing is the use of approximation whereby the spatial access process gradually narrows its search area until the required database objects are found.



(a) The R-tree indexing hierarchy

3. Spatial Data Integrity and Constraints

General Database constraints:

- Domain constraints: which specify the types of data values such as numeric, character or string, Boolean, date and time, and user-defined.
- Key and relationship constraints: which govern the use of entities as primary, secondary and foreign keys in data tables.
- Semantic integrity constraints: which are written rules stating what is allowed and not allowed in both data structure and data management.

…

- Cockcroft (1997) extended this classification to encompass the special requirements of spatial data. The spatial database constraints are,
- **Topological integrity constraints:** which are concerned with the geometric properties of spatial relationships (i.e. adjacency, containment and connectivity) between spatial features.
- **Semantic integrity constraints:** which are database rules governing the spatial behaviours of objects in the database (for example, no land parcels can be located in areas shown as water bodies).
- **User-defined constraints:** which are business rules similar to those identified in non-spatial data modelling (for example, no wood harvesting is allowed in a 200-metre buffer zone along the shore of lakes).

4. Long Transaction Management

- They involve not merely the movement of data in and out of the database, but also the recording of each and every step of the transaction process so the database can restored back to a consistent state if required.
- Transactions in spatial data processing, for example updating the road fabric as part of pavement maintenance program, commonly take several days or even several weeks to complete, while many users may need to access the data at the same time.
- Oracle, for example, uses the techniques of workspace management to handle long transactions. Workspace management refers to the ability of the database to hold different versions of the same record or row in one or more database workspaces. Users of the database can then change these versions independently.

SPATIAL DATA STRUCTURE AND DATABASE MODELS

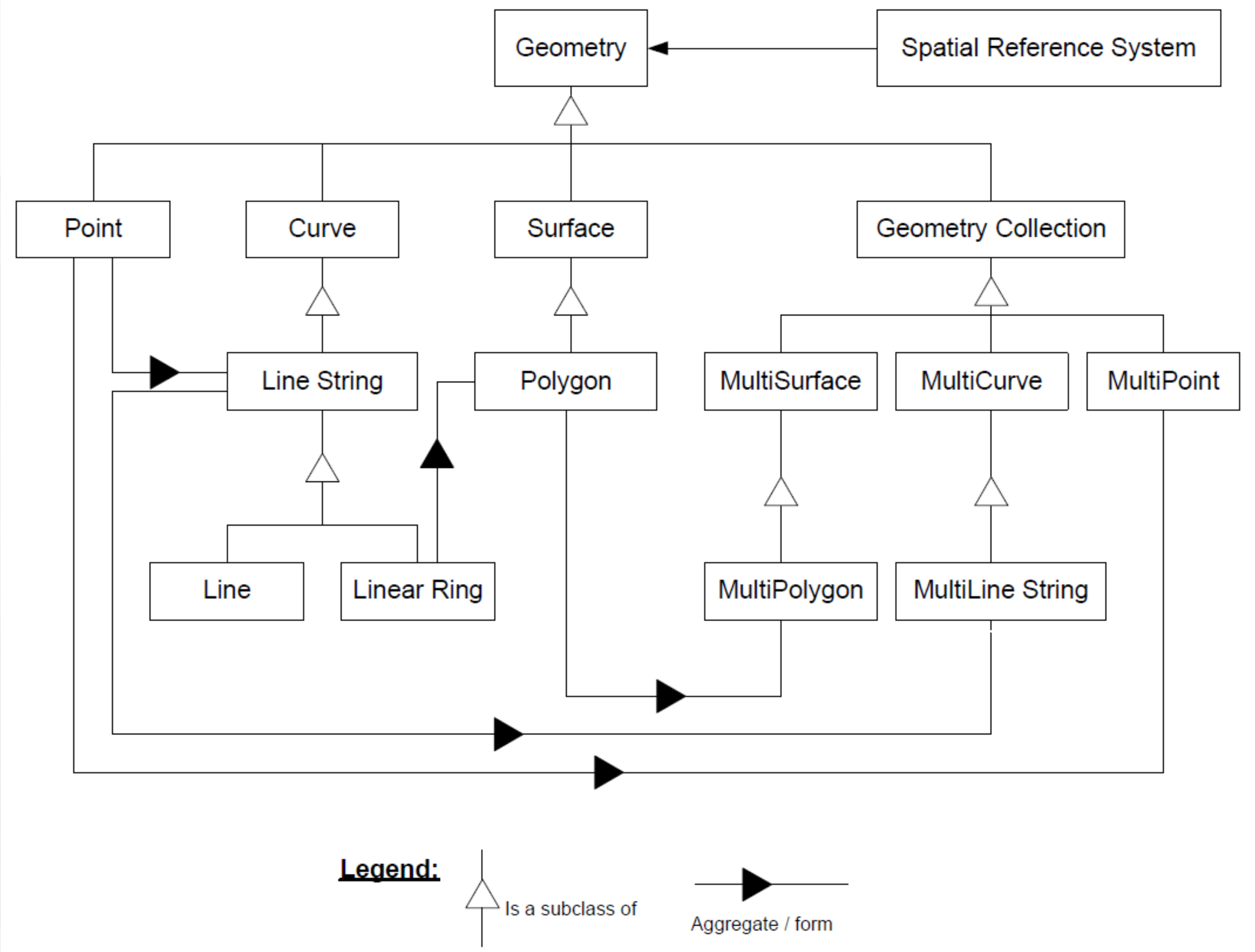
How to store spatial data?

A data structure is a specialized format for organizing, processing, retrieving and storing data

• Geometry • Coordinates • Relationship •

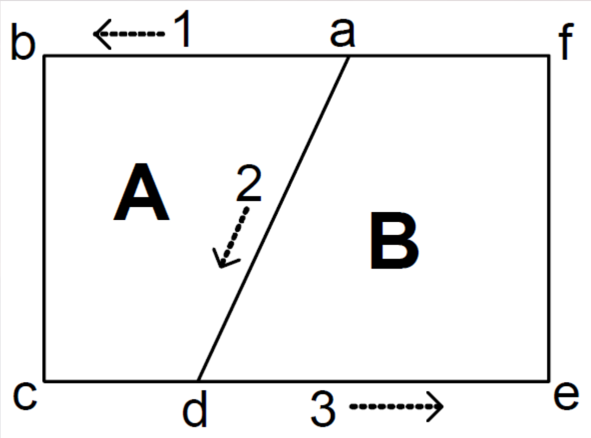
Geometrical data structure:

- OGC
- Simple geometries – Point, Curve, Surface, Geometry Collection



Topological data structure:

- Spatial relationships – adjacency, connectivity and containment
- Coverage files



Polygon A = (403600, 275700), (403000, 275700), (403000, 275000), (403300, 275000), (403600, 275700)

Polygon B = (403600, 275700), (403300, 275000), (404000, 275700), (404000, 275700)

- Polygon File
- Coordinate File
- Arc File
- Network Topology File
- Node File
- Polygon Topology File

3. Non-topological data structure:

- Limited cartographic rendering, Incompatible RDBMS, Absence of topology, and Metadata transferability
- Shapefile
- Separate external files of related properties for each layers

4. Geo-relational data structure

- Independent layer that represent a feature type
- Collection of layers that represent an incident area
- Separate external files of related properties for each layers

5. Geodatabase structure

- Object-oriented and object-relational database systems
- Indexing, transaction management, extensive database operations and spatial data integrity
- Personal Geodatabase

SPATIAL OPERATORS / RELATIONSHIPS

...

Spatial operators explain the relationships between objects

- Egenhofer (1994) classified them as *unary* and *binary* operators
- Clementini and Di Felice (1997) classified them into *topological, projective* and *metric* operators
- OGC (1999) has developed a comprehensive classification of spatial operators

OGC spatial operators…

Basic Operators	Topological Operators	Spatial Analysis Operators
Spatial Reference: Returns the reference system of the geometry Envelope: Returns the minimum bounding rectangle of the geometry Export: Converts the geometry into a different representation IsEmpty: Tests if the geometry is the empty set or not IsSimple: Returns TRUE if the geometry is simple Boundary: Returns the boundary of the geometry	Equal: Tests if the geometries are spatially equal Disjoint: Tests if the geometries are disjoint Intersect: Tests if the geometries intersect Touch: Tests if the geometries touch each other Cross: Tests if the geometries cross each other Within: Tests if a geometry is within another geometry Contain: Tests if a given geometry contains another geometry Overlap: Tests if a given geometry overlaps another given geometry Relate: Returns TRUE if the spatial relationship specified by the 9-Intersection matrix holds	Distance: Returns the shortest distance between any two points of two given geometries Buffer: Returns a geometry that represents all points whose distance from the given geometry is less than or equal to a specified distance ConvexHull: Returns the convex hull of a given geometry Intersection: Returns the intersection of two geometries Union: Returns the union of two geometries Difference: Returns the difference of two geometries SymDifference: Returns the symmetric difference of two geometries

Query Language

• Basic structure: DDL, DML •

• Relational Algebra

• Select & Project Operation •

• Set Operation • Join Operation •

• Formal Queries •

1. Relational Algebra

- Operand and operations
- RA has only one operand, relation/table
- Six operations, Select, Project, Union, Difference, Intersection, and Cross-product

<u>Select:</u>

- Select a set of data
- The select operation retrieves a subset of rows of the relational table, and the project operation extracts a subset of the columns.
  - σ cont = NAM(Country)

Table

Column

…

Project:

- Prepares for the display of data
- Subsets of columns for all rows in a relation are extracted by applying the project operation
- ΠName (σCont=NAM(Country))

Table

Shows names of countries

Column

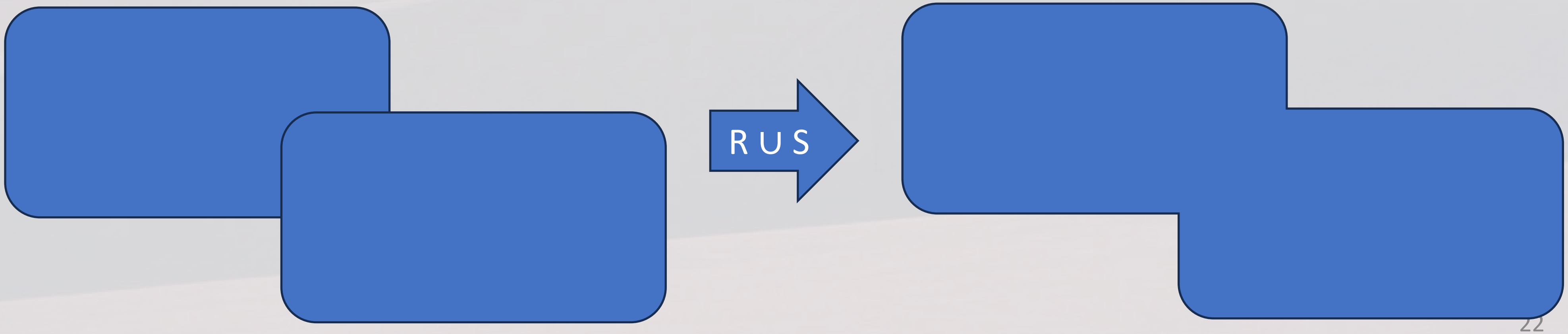
Set Operations

Relations/tables must be union-compatible

- They have the same number of columns, share the same domain, and if the columns appear in the same order from left to right.

- Union, Difference, and Intersection

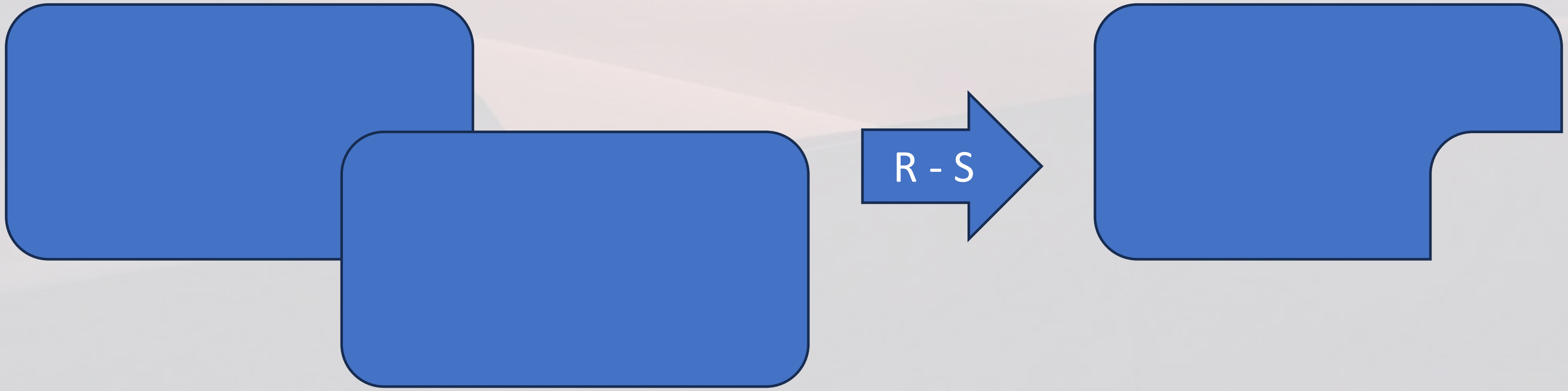
- <u>Union:</u> If R and S are relations/tables, then R ∪ S returns all tuples which are either in R or S.
- R ∪ S



…

Difference:

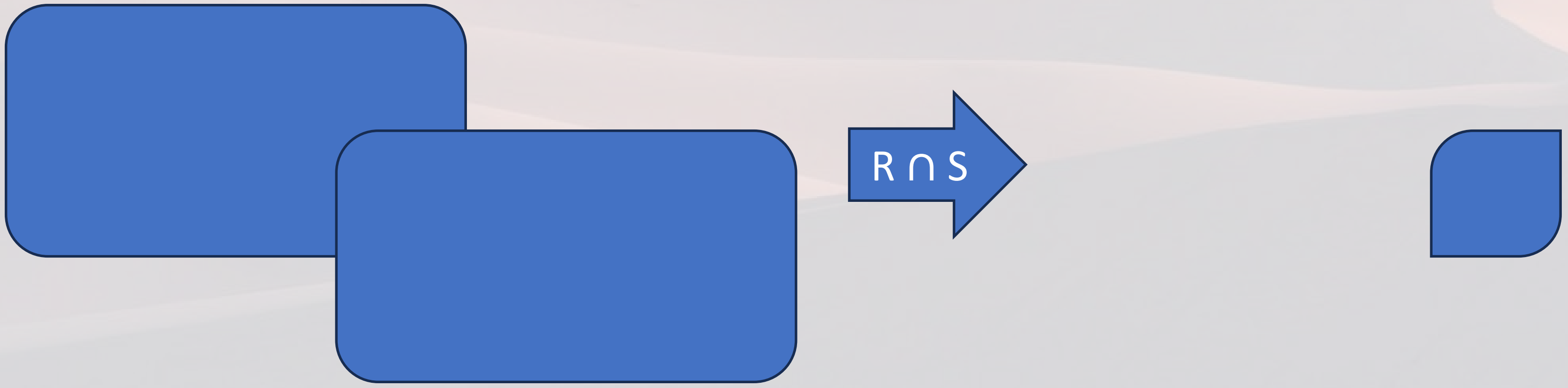
- Returns all tuples in R that are not in S.
- The difference operation can be used, for example, to list all countries in North America that have no river (listed in the River table) originating in them.
- R – S.



...

Intersection:

- For two union-compatible relations R and S, the intersection operation returns all tuples which occur both in R and S.
- R ∩ S



Join Operation

- To query data across multiple related tables
- Join is a cross-product of two relations/tables
- $R \bowtie_c S = \sigma_c (R \times S)$
- $U = R \bowtie S = \sigma(R.Name = \text{'Mexico'}) \land (R.Pop > S.Pop)(R \times S)$

2. Formal Queries

Data Definition Language (DDL):

- CREATE DATABASE testDB;

- CREATE TABLE Persons (

```
CREATE TABLE Persons (
    PersonID int,
    Name varchar(50),
    Age int,
    Salary int,
    Address varchar(255),
    City varchar(50)
);
```

...

Data Manipulation Language (DML):

- INSERT INTO Persons (PersonID, Name, Age, Salary, Address, City)
  VALUES (‘GEO1801’, ‘Harish Kumar’, 32, 15000, ‘21, VS Nagar’, ‘Tiruchirappalli’);

- SELECT * FROM Persons   SELECT City FROM Persons
- SELECT * FROM Persons
  WHERE City=‘Tiruchirappalli’;

- UPDATE Persons
  SET Address = ‘215, VS Nagar’, City= ‘Coimbatore'
  WHERE PersonID = GEO1801;

- DELETE FROM Persons WHERE Name =‘Harish Kumar’;

2.1 Aggregate Functions

- An aggregate function is a function that performs a calculation on a set of values, and returns a single value.
- The most commonly used SQL aggregate functions are:

  MIN() - returns the smallest value within the selected column

  MAX() - returns the largest value within the selected column

  COUNT() - returns the number of rows in a set

  SUM() - returns the total sum of a numerical column

  AVG() - returns the average value of a numerical column

- Aggregate functions ignore null values (except for COUNT()).

...

MIN

- SELECT MIN(Salary)
  FROM Persons;

MAX

- SELECT MAX(Salary)
  FROM Persons
  WHERE City='Tiruchirappalli';

COUNT

- SELECT COUNT(*)
  FROM Persons;

SELECT COUNT(PersonID)
FROM Persons WHERE City='Tiruchirappalli';

…

SUM

- SELECT SUM(Salary)
  FROM Persons;

  SELECT SUM(Salary)
  FROM Persons WHERE City='Tiruchirappalli';

AVG

- SELECT AVG(Salary)
  FROM Persons;

  SELECT AVG(Salary)
  FROM Persons WHERE City='Coimbatore';

Exercise:

- Refer syntax for all the SQL Statements

Views

- A view is a virtual table
- It contains rows and columns, just like a real table.
- The fields in a view are fields from one or more real tables in the database.

```
CREATE VIEW [Human Resource] AS
SELECT Name, Age, Salary
FROM Persons
WHERE City = ‘Coimbatore';
```

Triggers / Stored Procedure

- A stored procedure is a prepared SQL code that you can save, so the code can be reused over and over again, similar to a function.
- So if you have an SQL query that you write over and over again, save it as a stored procedure, and then just call it to execute it.
- You can also pass parameters to a stored procedure, so that the stored procedure can act based on the parameter value(s) that is passed.

- ```
  CREATE PROCEDURE SelectAllPersons
  AS
  SELECT * FROM Persons
  GO;
  ```

Execute the stored procedure as follows:

```
EXEC SelectAllCustomers;
```

- ```
  CREATE PROCEDURE SelectAllPersons @City varchar(50)
  AS
  SELECT * FROM Persons WHERE City = @City
  GO;
  ```

Execute the stored procedure as follows:

```
EXEC SelectAllCustomers @City = ‘Tiruchirappalli’;
```

Metadata

...

- Data or information about data
  - what is in the data
  - where the data originated
  - who produced them
  - when they were produced and modified
  - why they were produced, and
  - how the data can be obtained

…

- Spatial metadata are a special type of metadata that are associated with a spatial database, a spatial data set, or a particular class or instance of spatial features.
- "Author", "Title", "Keywords", "Subject" etc.. are examples of metadata keywords
- Standardized by metadata schemas or standards
- For spatial metadata schema, the most prominent examples include the three standards,
  - Content Standard for Digital Geospatial Metadata (CSDGM, developed by the FGDC),
  - ISO 19115 – Metadata, and
  - GEO Profile of Z39.50

Significance of Metadata

- Uniformity of Data Collection
- Data Management
- Data Understanding
- Data Use
- Data Sharing
- Data Archiving and Warehousing

Principles of Metadata Standards

Modularity: which is a key organising principle for the user environment of metadata characterised by a high degree of diversity in terms of sources, contents, and approaches to resource description.

Namespaces: which are defined as formal collections of terms managed according to a policy or algorithm that provides the mechanism to ensure global uniformity in the vocabulary used by a metadata standard.

Extensibility: which allows for profiles to be developed so that particular needs of a given application can be accommodated by a standard without unduly compromising the functionality of the base metadata schema.

Granularity: which allows a metadata designer to choose a level of detail appropriate to a given application.

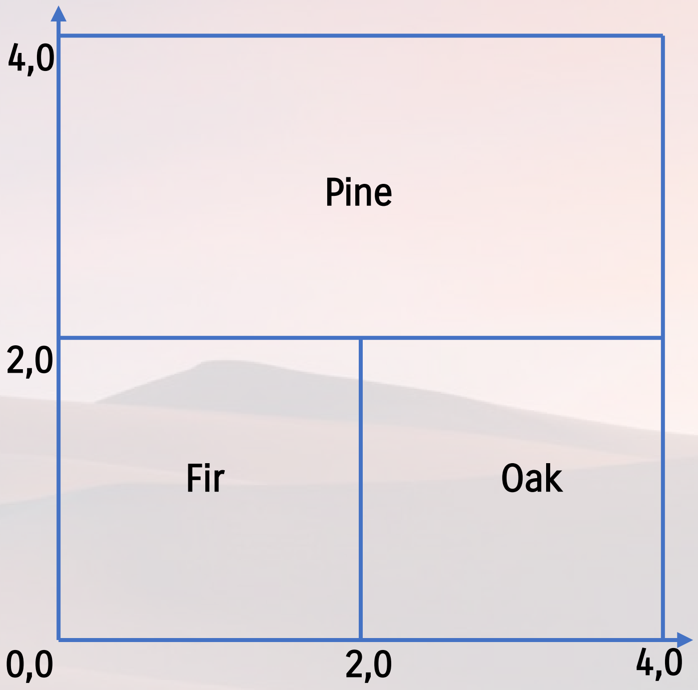
Multilingualism: which is the ability of a metadata standard to accommodate documentation that accommodates the linguistic and cultural diversities of its users.

Spatial Data Models

…

Models of spatial information are grouped into two categories: **Field and Object**

- **Field:** These models are used to model spatial data that is *continuous in nature*, e.g. terrain elevation, air quality index, temperature data, and soil variation characteristics.
  - Equivalent to raster data
- **Object:** These models have been used to model *discrete data* for applications such as transportation networks, land parcels, buildings, and other objects that possess both spatial and non-spatial attributes.
  - Equivalent to vector data



Object-based model

AreaID	Dominant Tree Species	Area/Boundary
FS1	Pine	[(0,0),(4,2),(4,4),(0,4)]
FS2	Fir	[(0,0),(2,0),(2,2),(0,2)]
FS3	Oak	[(2,0),(4,0),(4,2),(2,2)]

Field-based model

f(x, y) =

- *"Pine"* $2 \leq x \leq 4;\ 2 < y \leq 4$
- *"Fir* $0 \leq x \leq 2;\ 0 \leq y \leq 2$
- *"Oak"* $2 < x \leq 4;\ 0 \leq y \leq 2$

Thank you!