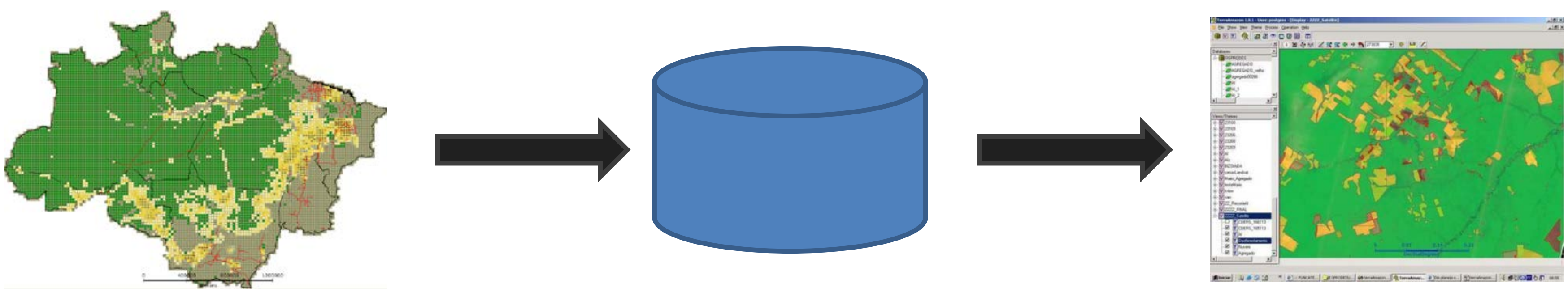
Spatial Databases: Lecture 3

Institute for Geoinformatics

Winter Semester 2014



Malumbo Chipofya: room 109

SPATIAL INTELLIGENCE LABORATORY

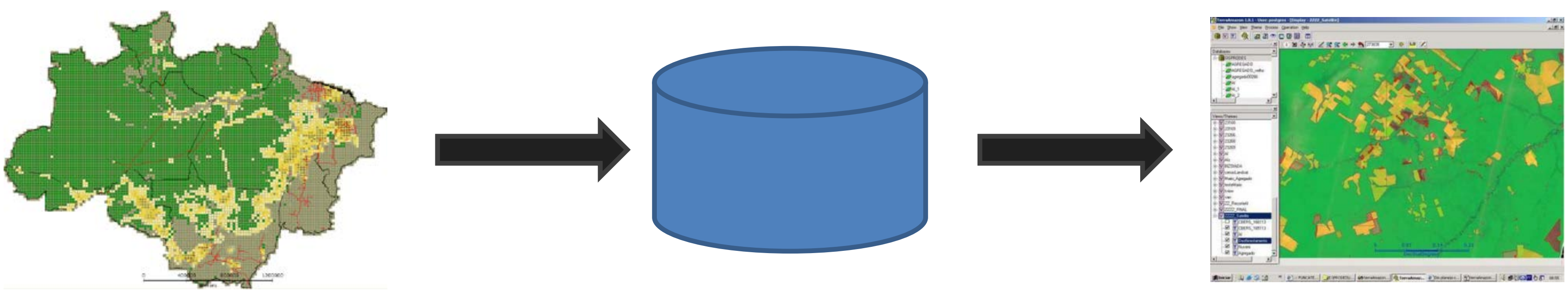
Topic Overview

1. Prelude: Data and problem solving in science and applications
2. **The Relational Database model**
3. Interacting with relational databases
4. Spatial Relational Database Management Systems
5. Applications: Terraview and Terralib: Prof. Dr. Gilberto Camara
6. A sample of Nosql Databases: brief introductions + example applications
    a. Array databases: SciDB
    b. Document databases: MongoDB
    c. Graph databases: Neo4J
7. Summary of all lectures given.

SPATIAL INTELLIGENCE LABORATORY

Recap

- What is a Relational Database?



SPATIAL INTELLIGENCE LABORATORY

What is a Relational Database

- Tables are formally known as “Relations”
- A row in table is called a “Tuple”
- A column in table is called an “Attribute”
  - Attributes have associated “Types”

- According to C.J. Date
  - Type are sets of things we can talk about
  - Relations are sets of things we can say about types

SPATIAL INTELLIGENCE LABORATORY

Relations, Attributes, and Tuples

- The set of all attributes of a relation must satisfy the **Uniqueness** property

At any given time no two tuples are equivalent

- This leads to the definition of keys

SPATIAL INTELLIGENCE LABORATORY

Special Attributes: Keys

- A **Candidate Key** for a relation, say ***R***, is a collection of attributes, say ***K*** satisfying (and being satisfied by, the following 2 properties:
  - **Uniqueness**: at **no** time does ***R*** have two distinct tuples whose projections to ***K*** have the same values
  - **Irreducibility**: no proper subset of ***K*** has the uniqueness property

SPATIAL INTELLIGENCE LABORATORY

Relational Operations

- **Restrict**
- **Project**
- **Join**

SPATIAL
INTELLIGENCE
LABORATORY

Database Design Principles

SPATIAL
INTELLIGENCE
LABORATORY

Entity/Relationship (E/R) Diagrams

- Consider the following description

An organisation is a collection of departments. Each department is named and has one or more employees. An employee is recorded with name (1^{st}, 2^{nd}, 3^{rd}), salary, and a list of dependents (and their relationship to the employee). The company executes a number of projects in which the employees participate. Every project has a project manager. A project will order parts of different kinds for maintaining its equipment. Parts can be made up of other subparts. The project keeps track of suppliers of each type of part including those who supply subparts. Every supplier is recorded with a name, a status, and the city of their location. In addition the project must keep track of the quantity of each part supplied by each supplier.

SPATIAL INTELLIGENCE LABORATORY

Entity/Relationship (E/R) Diagrams

- Task: design a database for the organisation described above.
- The design task involves identifying 4 kinds of things
  - Entities
  - Properties
  - Relationships
  - Subtypes

Entity/Relationship (E/R) Diagrams

Concept	Informal definition	Examples
ENTITY	A distinguishable object	Supplier, Part, Shipment Employee, Department Person Composition, Concerto Orchestra, Conductor Purchase order, Order line
PROPERTY	A piece of information that describes an entity	Supplier number Shipment quantity Employee department Person height Concerto type Purchase order date
RELATIONSHIP	An entity that serves to interconnect two or more other entities	Shipment (supplier-part) Assignment (employee-department) Recording (composition-orchestra-conductor)
SUBTYPE	Entity type *Y* is a subtype of entity type *X* if and only if every *Y* is necessarily an *X*	Employee is a subtype of Person Concerto is a subtype of Composition

Fig. 14.1 Some useful semantic concepts

SPATIAL INTELLIGENCE LABORATORY

Entity/Relationship (E/R) Diagrams

- Types of entities:
  - Regular entity (exists in its own right)
  - Weak entity (existence depends on other entities)

Entity/Relationship (E/R) Diagrams

- Types of properties:
  - Simple vs. composite
  - Key vs. non-key
  - Single- vs. multi- valued
  - Base vs. derived

SPATIAL
INTELLIGENCE
LABORATORY

Entity/Relationship (E/R) Diagrams

- Types of relationships:
  - 1-to-1
  - 1-to-many
  - Many-to-many

- Participation of an entity in a relationship
  - Total
  - Partial

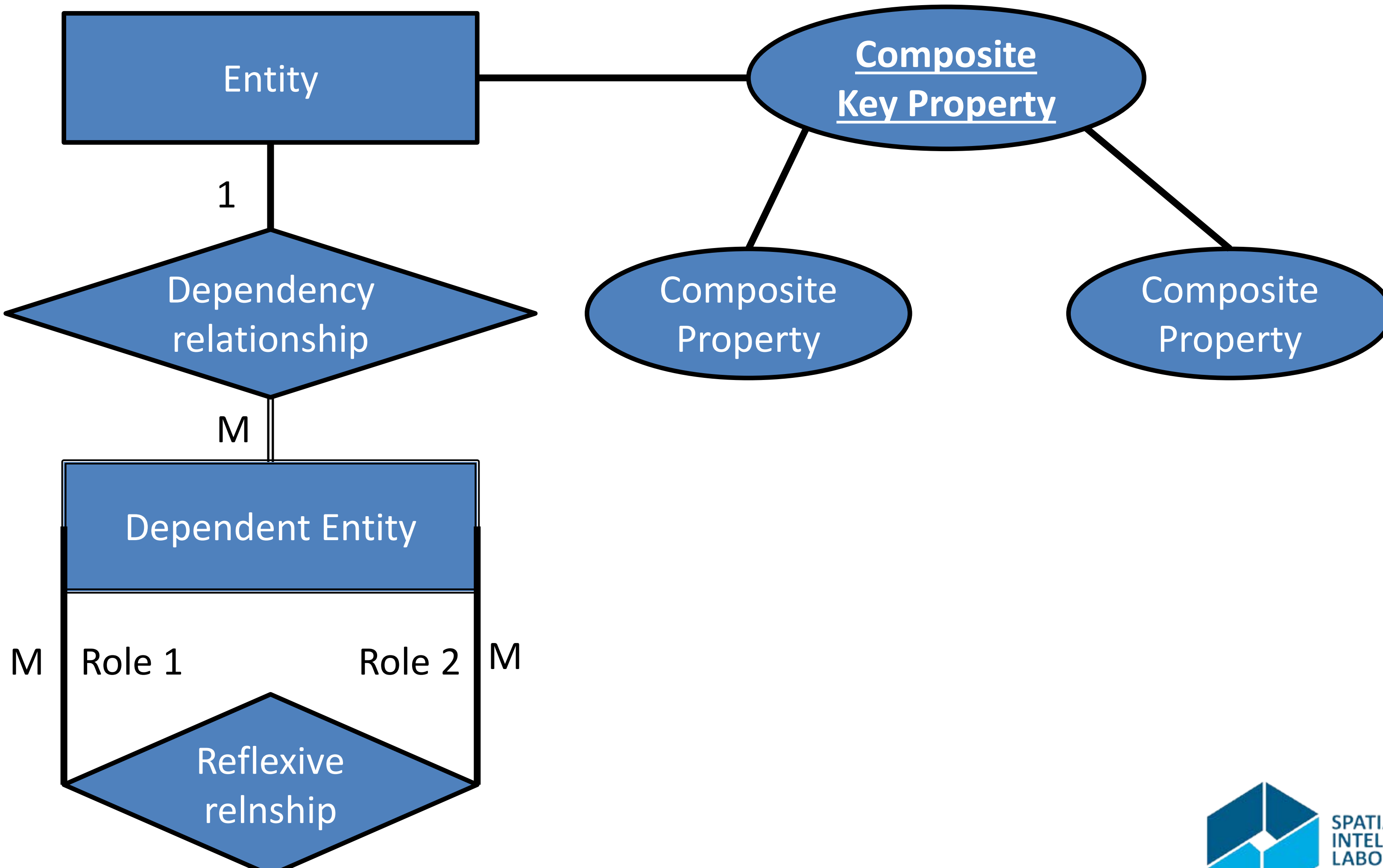
SPATIAL
INTELLIGENCE
LABORATORY

Entity/Relationship (E/R) Diagrams

- The four things identified form part of the E/R model
- An E/R diagram is a pictorial representation of an instance of the model

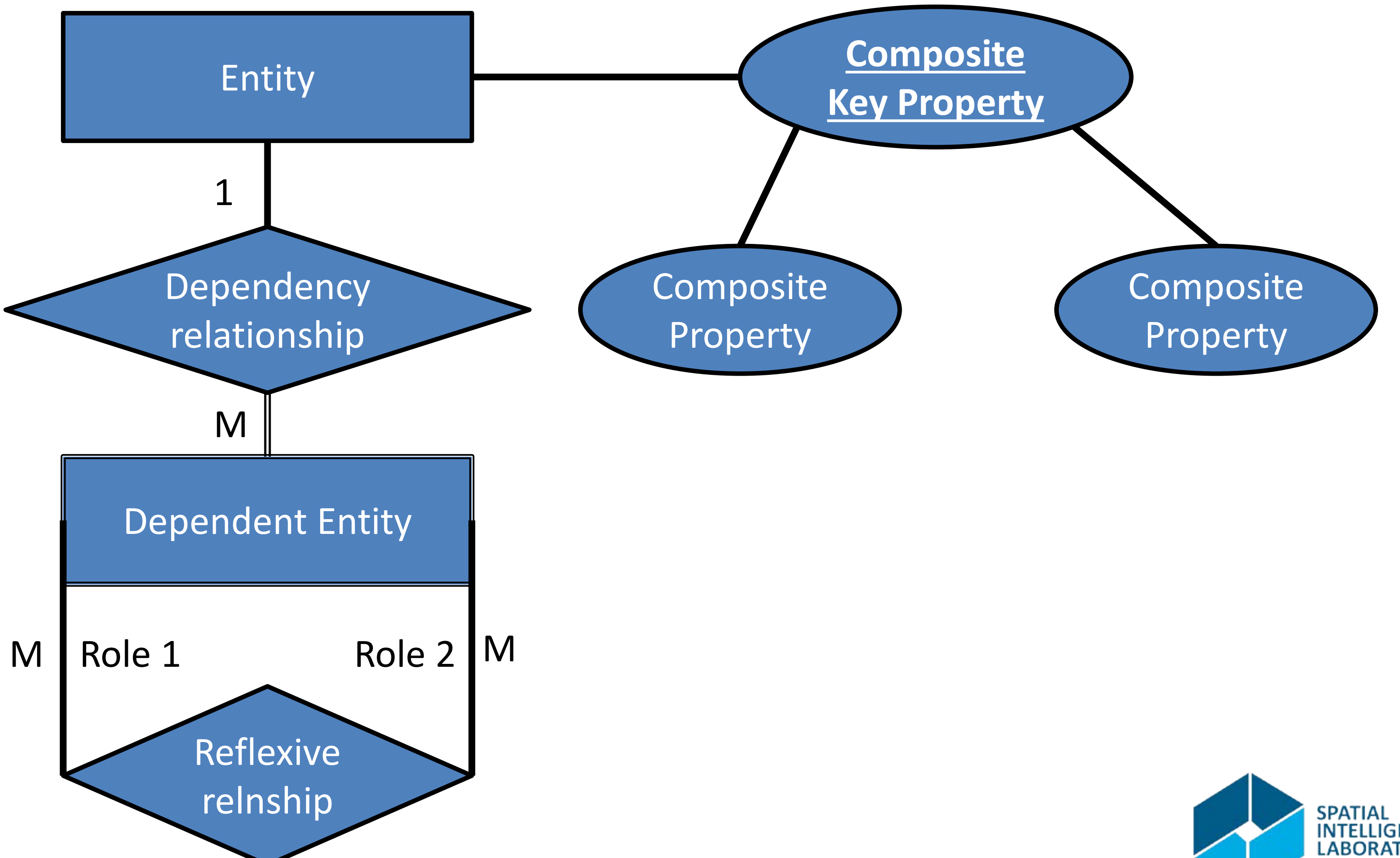
SPATIAL INTELLIGENCE LABORATORY

Entity/Relationship (E/R) Diagrams



SPATIAL INTELLIGENCE LABORATORY

Entity/Relationship (E/R) Diagrams



SPATIAL INTELLIGENCE LABORATORY

Exercise 4

- Draw an E/R diagram for the organisation described above.
- Map the E/R diagram to a relational database specification
  - i.e. decide on relations and their attributes and keys

SPATIAL
INTELLIGENCE
LABORATORY

Note on Constraints

- Type constraints (CHECK)
- Key constraints (UNIQUE, PRIMARY KEY)
- Referential constraint (FOREIGN KEY)
- Attribute constraints (NOT-NULL)

Exercise 4

- Draw an E/R diagram for the organisation described above.
- Map the E/R diagram to a relational database specification
  - i.e. decide on relations and their attributes and keys

Functional Dependencies

- Given two sets of attributes of a relation $\boldsymbol{R}$:

$$\boldsymbol{A} := \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \ldots\} \quad \boldsymbol{B} := \{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \ldots\}$$

- $\boldsymbol{A}$ is a functionally dependent on $\boldsymbol{B}$ written

$$\boldsymbol{B} \rightarrow \boldsymbol{A}$$

if and only if there is a function from the set of legal values of $\boldsymbol{B}$ to the set of legal values of $\boldsymbol{A}$ determined exactly by tuples of $\boldsymbol{R}$

SPATIAL INTELLIGENCE LABORATORY

Functional Dependencies

- Example: Consider the following relation

ID#	Name	M.St	#Chd	#Yrs	M.€	Date	#sticks	Wgt.	Hrs
1	Madelene Doe	M	0	2	30	1.06	55	9	6
2	Doe John	S	0	1	30	7.05	34	5	5
3	Julie Nuts	S	2	2	40	1.06	54	9	6
4	Mary Jane	M	3	4	50	3.11	61	12	8

- What are the FDs in this relation?

SPATIAL INTELLIGENCE LABORATORY

Functional Dependencies

- What are the FDs in this relation?

ID#	Name	M.St	#Chd	#Yrs	M.€	Date	#sticks	Wgt.	Hrs
1	Madelene Doe	M	0	2	30	1.06	55	9	6
2	Doe John	S	0	1	30	7.05	34	5	5
3	Julie Nuts	S	2	2	40	1.06	54	9	6
4	Mary Jane	M	3	4	50	3.11	61	12	8

- {ID#} → {Name, M.St, #Chd, #Yrs, M.€, #sticks, Wgt., Hrs}
- {ID#, Name} → {M.St, #Chd, #Yrs, M.€, #sticks, Wgt., Hrs}
- {Name, M.St, #Chd, #Yrs, M.€, #sticks, Wgt., Hrs} → {#Yrs, M.€, #sticks, Wgt., Hrs}

SPATIAL INTELLIGENCE LABORATORY

Functional Dependencies

- Trivial FD
  - {Name, M.St, #Chd, #Yrs, M.€} → {Name, M.€}
  - $LHS \supseteq RHS$

Functional Dependencies

- Closure of a set of FDs: Consider
  - {ID#, Name} → {M.St, #Chd, #Yrs, M.€}
  - {ID#} → {Name}
  - {Name} → {CountryOfOrigin}

- From the above dependency we can derive
  - {ID#, Name} → {M.St}
  - {ID#, M.St} → {Name, M.St}
  - {ID#} → {CountryOfOrigin}

SPATIAL INTELLIGENCE LABORATORY

Functional Dependencies

- Closure of a set of FDs: Consider
  - {ID#, Name} → {M.St, #Chd, #Yrs, M.€}
  - {ID#} → {Name}
  - {Name} → {CountryOfOrigin}

- The closure of a set of FDs, S, is the set of ***all*** FDs that can be derived from S.

Functional Dependencies

- Rules – we write 'A' for {A} and 'A,B,C' for {A, B, C}
  - Reflexivity:
    - $B \subseteq A$ implies $A \longrightarrow B$
  - Augmentation:
    - $A \longrightarrow B$ implies $A,C \longrightarrow B,C$
  - Transitivity:
    - $A \longrightarrow B$ and $B \longrightarrow C$ implies $A \longrightarrow C$
  - Self-determination:
    - $A \longrightarrow A$
  - Decomposition:
    - $A \longrightarrow B,C$ implies $A \longrightarrow B$ and $A \longrightarrow C$
  - Union:
    - $A \longrightarrow B$ and $A \longrightarrow C$ implies $A \longrightarrow B,C$
  - Composition:
    - $A \longrightarrow B$ and $C \longrightarrow D$ imples $A,C \longrightarrow B,C$

Functional Dependencies

- Irreducibility
  - A set of FDs, S, is irreducible if and only if it satisfies
    - RHS of every FD in S has only one attribute
    - LHS of every FD in S is irreducible in the sense that discarding any attribute changes the closure of S – left irreducibility
    - Discarding any FD in S changes the closure of S

SPATIAL INTELLIGENCE LABORATORY

Exercise 1

A manager in a tobacco manufacturing company has the following information for the Tobacco Rollers Department:

1. Employee ID of each Roller
2. Names of Each Roller
3. Tobacco sticks rolled for each day
4. Total weight of tobacco rolled per day
5. Number of hours worked each day

SPATIAL INTELLIGENCE LABORATORY

Exercise 1

1. Employee ID of each Roller
2. Names of Each Roller
3. Tobacco sticks rolled for each day
4. Total weight of tobacco rolled per day
5. Number of hours worked each day

Qn. 1: What are the possible attributes and their types?

Qn. 2: What relations can be formed?

1) a. Employee b. Productivity, 2) EmployeeProductivity

SPATIAL INTELLIGENCE LABORATORY

Exercise 2

The human resources department keeps track of the welfare of Tobacco Rollers and Tobacco Packers who belong to different departments:

1. Employee’s income
2. Monthly health reports
   a. Skin condition
   b. Lung condition
   c. Eye condition (left and right eye seperately)
3. Marital status
4. Number of children
5. Years of service

Qn. 1: Define the relations and specify the type for each attribute

Qn. 2: For each relation above plus those from exercise 1 (update the relations if necessary) specify

a. **some** candidate keys

b. a primary key

c. the foreign keys

SPATIAL INTELLIGENCE LABORATORY

Exercise 3

For the relations you have defined in exercises 1 and 2, which (how) of the operators can be used to answer the following ***queries***

Qn. 1: What are the names of the employees with bad left eye condition

Qn. 2: Which department has an employee earning less than 100 monies

Qn. 3: What are names of all tobacco rollers

SPATIAL INTELLIGENCE LABORATORY

Exercise 5

- Consider all relations you obtained from the previous exercises (1 – 3).

Qn. 1: Determine the join of all those relations.

Qn. 2: Determine as large a set of FDs as you can from this new relation (enough to answer Qns 3,4).

Qn. 3: Demonstrate the application of the 7 rules to derive new functional dependencies using your answer to Qn. 2: *Reflexivity, Augmentation, Transitivity, Self-determination, Decomposition, Union, Composition*.

E.g. ID# $\longrightarrow$ Name,M.St ***therefore*** ID# $\longrightarrow$ Name ***and*** ID# $\longrightarrow$ M.St

SPATIAL INTELLIGENCE LABORATORY

Exercise 5

Qn. 4: state two irreducible FDs among those you have listed.

Qn. 5: Is it feasible to compute the closure of every (i.e. any single but arbitrary) set of FDs?

Exercise 6

- Consider the following descriptions:
  - A polygon is given by a cyclic ordered list of coordinate pairs
  - A point is a coordinate pair
  - A line is pair of points
  - A linestring is an ordered list of points

Qn. 1: Design a set of relations using the E/R design process

Qn. 2: Convert your design into SQL data definition language statements – i.e. create a database with those relations

SPATIAL INTELLIGENCE LABORATORY

Exercise 6

Qn. 3: Would your databases be a suitable basis for storing spatial data for a GIS application?

- Why and/or why not?

References

- C.J. Date, *An Introduction to Database Systems, 8th Edition*. Pearson Education Inc., 2004.
- See www.geoinformatic.cc

Reading

- C.J. Date, *An Introduction to Database Systems, 8th Edition*. Pearson Education Inc., 2004.
  - Chapters 11, 12, and 14

SPATIAL INTELLIGENCE LABORATORY

That's all for today

Thank you!

Questions?

SPATIAL
INTELLIGENCE
LABORATORY