The Relational Data Model
Data Models

Types
- Relational data model
- Object data model
- Hierarchical data model (Trees)
- Network data model (Graphs)
- Semistructured Data Model

Structure - Operations - Constraints
Relational vs Semistructured

Figure 2.1: An example relation

<table>
<thead>
<tr>
<th>title</th>
<th>year</th>
<th>length</th>
<th>genre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gone With the Wind</td>
<td>1939</td>
<td>231</td>
<td>drama</td>
</tr>
<tr>
<td>Star Wars</td>
<td>1977</td>
<td>124</td>
<td>sciFi</td>
</tr>
<tr>
<td>Wayne’s World</td>
<td>1992</td>
<td>95</td>
<td>comedy</td>
</tr>
</tbody>
</table>

Figure 2.2: Movie data as XML

```xml
<Movies>
  <Movie title="Gone With the Wind">
    <Year>1939</Year>
    <Length>231</Length>
    <Genre>drama</Genre>
  </Movie>
  <Movie title="Star Wars">
    <Year>1977</Year>
    <Length>124</Length>
    <Genre>sciFi</Genre>
  </Movie>
  <Movie title="Wayne’s World">
    <Year>1992</Year>
    <Length>95</Length>
    <Genre>comedy</Genre>
  </Movie>
</Movies>
```

Relational Database Model

Introduced by E.F. Codd (1970)
http://www.acm.org/classics/nov95/

Based on relational algebra and logic developed by

Schröder (1880s)
Charles Peirce (1890s)
Russell and Whitehead (1900s)
Codd’s Twelve Rules

1. Information represented at the logical level in tables.
2. Data is determined by table, primary key, and column.
3. Missing information is modeled as null values.
4. Metadata is part of the database.
5. Single language for all tasks in DBMS.
6. Views and tables must change simultaneously.
7. Single operations for retrieve, insert, delete, update.
8. Operations independent of physical storage and access.
9. Database modifiable without affecting applications.
10. Constraints are part of database.
11. DML independent of physical layer (distributed, etc.)
12. Row-processing obeys same rules as set-processing.
Relations

Extensional versus intensional

Extensional Representation:

- table of values
  - rows = records
  - columns = attributes

Note:
- rows in tables are ordered,
- instances of relations are not
Domains

Set of atomic values for an attribute

atomic = indivisible
  (e.g. CSC 355 = CSC + 355 is divisible)

Examples

  age: integer
  sex: {male, female}

Physical Level: data type + format
Relation Schema

\( R(A_1, A_2, \ldots, A_n) \)  

- **R**: Name of Relation  
- **A_1, A_2, \ldots, A_n**: Attributes  
- **n**: degree (arity) of R

**Example**

Movie(movieID, title, genre, length, rating)

Domains ?
Relation Schema with Domains

R(A₁:D₁, A₂:D₁, ..., Aₙ:Dₙ) ← Aᵢ has domain Dᵢ

Example

Movie(movieID: integer, title:string, genre: Genres, length: Lengths, rating: Ratings)

dom(genre) = Genres = {Musical, Horror, ...}
dom(length) = Lengths = {x: x is valid time}
dom(rating) = Ratings = {NR, G, PG, PG-13, R, NC-17}

or

Movie(movieID: integer, title:string, genre: string, length: integer, rating: string)
Relational Schemas Example

CUSTOMER(Customer_ID, Customer_Name, City, State, Postal_Code)

ORDER(Order_ID, Order_Date, Customer_ID)

ORDERLINE(Order_ID, Product_ID, Ordered_Quantity)

PRODUCT(Product_ID, Product_Description, Product_Finish, Standard_Price)
Instances

Given $R(A_1, A_2, \ldots, A_n)$, $A_i$ has domain $D_i$

Instance of schema $R$ is a table with data from domains

Example: Student(LastName, FirstName, SID, SSN, Career, Program, City, Started)
Records

<510, “Monkey Business”, comedy, null, null>

are possible records (or tuples) in MOVIE(movieID, title, genre, length, rating).

null: value unknown, or attribute does not apply

values atomic: no multiple values (first normal form)
  (e.g. several genres)
  indivisible (name = first name + last name)
What about multiple values?

<table>
<thead>
<tr>
<th>LastName</th>
<th>FirstName</th>
<th>SID</th>
<th>SSN</th>
<th>Career</th>
<th>Program</th>
<th>City</th>
<th>Started</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snowdon</td>
<td>Jonathan</td>
<td>8871</td>
<td>123123123</td>
<td>GRD</td>
<td>INFO-SYS</td>
<td>Springfield</td>
<td>2009</td>
</tr>
<tr>
<td>Winter</td>
<td>Abigail</td>
<td>11035</td>
<td>111111111</td>
<td>GRD</td>
<td>PHD</td>
<td>Chicago</td>
<td>2009</td>
</tr>
<tr>
<td>Patel</td>
<td>Deepa</td>
<td>14662</td>
<td>GRD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degroff</td>
<td>Jarvis</td>
<td>14998</td>
<td>113311331</td>
<td>GRD</td>
<td>COMP-GAM</td>
<td>Evanston</td>
<td>2012</td>
</tr>
<tr>
<td>Starck</td>
<td>Jason</td>
<td>19992</td>
<td>789789789</td>
<td>UGRD</td>
<td>INFO-SYS</td>
<td>Springfield</td>
<td>2009</td>
</tr>
<tr>
<td>Johnson</td>
<td>Peter</td>
<td>32105</td>
<td>123456789</td>
<td>UGRD</td>
<td>COMP-SCI</td>
<td>Evanston</td>
<td>2013</td>
</tr>
<tr>
<td>Pollard</td>
<td>Joya</td>
<td>39077</td>
<td>GRD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kubik</td>
<td>Dwayne</td>
<td>57923</td>
<td>975797979</td>
<td>UGRD</td>
<td>COMP-SCI</td>
<td>Springfield</td>
<td>2013</td>
</tr>
<tr>
<td>Skelly</td>
<td>Trinity</td>
<td>58992</td>
<td>555222555</td>
<td>GRD</td>
<td>PHD</td>
<td>Springfield</td>
<td>2012</td>
</tr>
<tr>
<td>Krol</td>
<td>Angelo</td>
<td>60573</td>
<td>UGRD</td>
<td></td>
<td>COMP-SCI</td>
<td>Springfield</td>
<td>2011</td>
</tr>
<tr>
<td>Patel</td>
<td>Prakash</td>
<td>75234</td>
<td>UGRD</td>
<td></td>
<td>COMP-SCI</td>
<td>Chicago</td>
<td>2011</td>
</tr>
<tr>
<td>Brennigan</td>
<td>Marcus</td>
<td>90421</td>
<td>987654321</td>
<td>UGRD</td>
<td>COMP-GAM</td>
<td>Evanston</td>
<td>2010</td>
</tr>
<tr>
<td>Snowdon</td>
<td>Jennifer</td>
<td>93321</td>
<td>321321321</td>
<td>GRD</td>
<td>COMP-SCI</td>
<td>Springfield</td>
<td>2012</td>
</tr>
</tbody>
</table>

Multiple programs? Telephone numbers? Movie Genres?
Constraints

- Domain constraints
- Key (uniqueness) constraints
- Entity integrity constraints
- Referential integrity constraints

- Data dependencies (functional dependencies, etc.)
Domain Constraints

Restriction on values of attributes (domain).

Specified as **data-type**: integer, char, etc., or user-defined type

Operations on data-types: +, *, <, =, ...

Not null constraint for an attribute
Keys

Key: smallest set of attributes that uniquely identify a record in the relational schema

composite: more than one attribute

Examples:

- \text{MOVIE}(\text{MovieID, Title, Year, Length, Rating})

  \{\text{MovieID, Title}\} \text{ is not a key!}

- \text{MEMBER}(\text{StudentID, Groupname, Joined})
Key Examples

PRODUCT(ProductID, Name, Description, PricePerUnit, UnitSize)

ACTIVITY(StudentID, Activity, Fee)

COURSE(CourseID, Title, Enrolment)

ENROLED(StudentID, CourseID, Qt, Year)
Candidate Keys

If a relation has more than one key, these keys are called candidate keys.

Examples

- EMPLOYEE(EmpID, FirstName, LastName, Salary, Gender)
- DePaul students: peoplesoftID and SSN
- COURSE(Department, Number, Name, Instructor)
- CAR(OwnerName, Vehicle#, Engine#, Color)

One candidate key is declared the primary key of the relation (underlined in schema)
Relational Databases and Schemas

Relational Database **Schema**: Collection of Relations

Relational Database **State**: Collection of Instances

ACTIVITIES = \{STUDENT, ACTIVITY\}

<table>
<thead>
<tr>
<th>Student</th>
<th>SID</th>
<th>FName</th>
<th>LName</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>101</td>
<td>Mark</td>
<td>Spencer</td>
</tr>
<tr>
<td></td>
<td>971</td>
<td>Charles</td>
<td>Loeffler</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Activity</th>
<th>StudentID</th>
<th>Activity</th>
<th>Fee</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>971</td>
<td>Piano</td>
<td>$20</td>
</tr>
<tr>
<td></td>
<td>971</td>
<td>Swimming</td>
<td>$10</td>
</tr>
</tbody>
</table>

**Note**: different names (SID, StudentID) for the same concepts
A set of attributes in one relation ($R_1$) referring to a unique tuple in a second relation ($R_2$) through $R_2$’s primary key.

<table>
<thead>
<tr>
<th>Student</th>
<th>SID</th>
<th>FName</th>
<th>LName</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>Mark</td>
<td>Spencer</td>
<td></td>
</tr>
<tr>
<td>353</td>
<td>Gil</td>
<td>Ryle</td>
<td></td>
</tr>
<tr>
<td>971</td>
<td>Charles</td>
<td>Loeffler</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Activity</th>
<th>StudentID</th>
<th>Activity</th>
<th>Fee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piano</td>
<td>971</td>
<td></td>
<td>$20</td>
</tr>
<tr>
<td>Reading</td>
<td>353</td>
<td></td>
<td>$5</td>
</tr>
<tr>
<td>Swimming</td>
<td>971</td>
<td></td>
<td>$10</td>
</tr>
</tbody>
</table>
Foreign Key 2

Examples

REGISTRATION = \{ STUDENT, ENROLMENT, COURSE \}
COMPANY = \{ EMPLOYEE, WORKS_ON, PROJECT \}
SUPPLY = \{ SUPPLIES, SUPPLIER, PART, COMPANY \}

Note: R1 = R2 is possible

Example

EMPLOYEE (with supervisor)
MOVIE (remakes)
Referential Integrity

declaration of foreign keys in a database schema

\[
\text{STUDENT} (\text{SID, FName, LName}) \\
\text{ACTIVITIES} (\text{StudentID references STUDENT, Activity, Fee})
\]

or visually, by an arrow from foreign key to primary key

\[
\text{STUDENT} (\text{SID, FName, LName}) \\
\text{ACTIVITIES} (\text{StudentID, Activity, Fee})
\]
Integrity Constraints

- Domain Constraints: declaration of domains
- Not Null Constraints: attribute values cannot be null
- Key Constraints: candidate keys (uniqueness)
- Entity Integrity Constraint: primary key is not null
- Referential Integrity Constraint: declaring foreign keys

A valid state is a database state fulfilling all integrity constraints

Integrity Constraints defined by DDL

Semantic constraints (transitions) later