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: tables are ordered, relations are not
Domains

Set of atomic values for an attribute.
atomic = indivisible
(e.g. IT 240 = IT + 240 is divisible)

Examples
age: \( \{ x : x \text{ is a valid age} \} \)
sex: \( \{ \text{male, female} \} \)

Physical Level: data type + format

Relation Schema

\[ R(A_1, A_2, \ldots, A_n) \] relational schema

\begin{align*}
R & : \text{Name of Relation} \\
A_1, A_2, \ldots, A_n & : \text{Attributes} \quad A_i \text{ has domain } D_i \\
N & : \text{degree (arity) of } R
\end{align*}

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

\begin{align*}
dom(movieID) = \text{MovieIDs} & = \{ x : x \text{ is a number} \} \\
dom(title) = \text{Titles} & = \{ x : x \text{ is a word} \} \\
dom(genre) = \text{Genres} & = \{ \text{Musical, Horror, } \ldots \} \\
dom(length) = \text{Lengths} & = \{ x : x \text{ is valid time} \} \\
dom(rating) = \text{Ratings} & = \{ \text{NR, G, PG, PG-13, R, NC-17} \}
\end{align*}

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, Product_Line_ID)

More Examples: page 32
Instances

R(A₁, A₂, ..., Aₙ) relational schema,
Aᵢ has domain Dᵢ.

Instance of schema R is a table with data from domains

Example:

Rep(RepNum, LastName, FirstName, Street, City, State, Zip, Commission, Rate)

<table>
<thead>
<tr>
<th>RepNum</th>
<th>LastName</th>
<th>FirstName</th>
<th>Street</th>
<th>City</th>
<th>State</th>
<th>Zip</th>
<th>Commission</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>J. Brown</td>
<td>John</td>
<td>123 Doe</td>
<td>New York</td>
<td>CA</td>
<td>12345</td>
<td>100</td>
<td>12.50</td>
</tr>
<tr>
<td>60</td>
<td>S. Black</td>
<td>Sarah</td>
<td>567 Doe</td>
<td>San Francisco</td>
<td>CA</td>
<td>12345</td>
<td>150</td>
<td>14.00</td>
</tr>
</tbody>
</table>

Figure 2-1 page 30

Schema and Relation Examples

STUDENT(Name, SSN, HomePhone, Address, OfficePhone, Age, GPA)

MOVIE

VIDEO

RENTAL

Records

<101, "Thirty-Nine Steps", mystery, 101, R>
<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?

Telephone numbers?

Functional Dependencies I

ORDER_DETAIL(Product, Quantity, PricePerUnit, Subtotal)

Example

<Soymilk, 3, $3.19, $9.57> in ORDER_DETAIL

(Quantity, PricePerUnit) → Subtotal

Quantity and PricePerUnit determine Subtotal

Functional Dependencies II

A → B (A determines B)

if setting the values of attribute(s) of A
determines the values of B

Example

SID → Name in STUDENT(SID, Name)
Functional Dependencies III

- EMPLOYEE(EmpID, FirstName, LastName, Salary, Sex)
- STUDENT(Name, SID, HomePhone, Address, OfficePhone, Age, GPA)
- PRODUCT(ProductID, Name, Description, PricePerUnit, UnitSize)
- CAR(OwnerName, Vehicle#, Engine#, Color)
- CUSTOMER(accountID, lastName, firstName, street, city, state, zipcode)

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 determines all attributes in a relation.
composite: more than one attribute

Example
STUDENT(Name, SID, HomePhone, Address, OfficePhone, Age, GPA)
[SID]: key
(SID, Name, Address): not key

Key Examples

MOVIE(Title, Year, Director, Length, Rating)
PRODUCT(ProductID, Name, Description, PricePerUnit, UnitSize)
ACTIVITY(StudentID, Activity, Fee)
COURSE(CourseID, Title, Enrolment)
ENROLLED(StudentID, CourseID, CourseTitle, Grade)

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>
<th>Activity</th>
<th>StudentID</th>
<th>Activity</th>
<th>Fee</th>
</tr>
</thead>
<tbody>
<tr>
<td>353</td>
<td>Mark Spence</td>
<td>101</td>
<td>Mark</td>
<td>Spencer</td>
<td>971</td>
<td>Piano</td>
<td>$20</td>
</tr>
<tr>
<td>971</td>
<td>Charles Loeffer</td>
<td>353</td>
<td>Gil</td>
<td>Ryle</td>
<td>971</td>
<td>Reading</td>
<td>$5</td>
</tr>
<tr>
<td>971</td>
<td>Charles Loeffer</td>
<td>971</td>
<td>Charles</td>
<td>Loeffer</td>
<td>971</td>
<td>Swimming</td>
<td>$10</td>
</tr>
</tbody>
</table>

*Note: different names (SID, StudentID) for the same concepts*

Foreign Key 1

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.

Terminology

- $R_1$: referencing relation
- $R_2$: referenced relation

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

\[
\text{STUDENT}(	ext{SID}, \text{FName}, \text{LName}) \\
\text{ACTIVITIES}(	ext{StudentID} \text{ references } \text{STUDENT}, \\
\text{Activity, Fee})
\]
or visually, by an arrow from foreign key to primary key

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

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