CSC 430/530 : DATABASE MANAGEMENT SYSTEMS/ DATABASE THEORY

Lecture 5: Relations and Relational Algebra

149

Overview

- Relations
  - Domains of Attributes
    - Constraint: Domain Constraints
  - Keys
    - Constraint: Foreign Key Constraints
  - Relational Algebra
    - Operators

150

Example of a Relation

attributes
(or columns)

ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

tuples
(or rows)

151

Domain of an attribute

- The set of allowed values for each attribute is called the **domain of the attribute**
- Example of a **student relation** **(r)** and its attributes:
  - **CWID:** 8-digit number
  - **First Name, Last Name:** Alpha String
  - **DoB:** Date
  - **SSN:** 9-digit number
  - **Passport:** String (Letter followed by 7 digits) - nullable
  - **Program:** Alpha String

CWID	First Name	Last Name	DoB	SSN	Passport	Program
11122333	Jack	Smith	27-Mar-1987	733-34-272	L456781	Computer Science
22233444	Mary	Doe	13-May-1994	555-22-123	Null	Cyber Engineering

152

Domain of an attribute

- Attribute values are (normally) required to be **atomic**; that is, indivisible
- The special value ***null*** is a member of every domain. Indicated that the value is “unknown”
- IMPORTANT: The null value causes complications in the definition of many operations

CWID	First Name	Last Name	DoB	SSN	Passport	Program
11122333	Jack	Smith	27-Mar-1987	733-34-272	L456781	Computer Science
22233444	Mary	Doe	13-May-1994	555-22-12`	Null	Cyber Engineering

153

Relation Schema and Instance

- Let $A_1, A_2, \ldots, A_n$ are *attributes,* and $D_1, D_2, \ldots. D_n$ are the domains of the attributes
- We represent the set of attributes as a relation $R = (A_1, A_2, \ldots, A_n)$, using the *relation schema*

  Example:

  INSTRUCTOR = (ID, name, dept_name, salary)
- Formally, given sets $D_1, D_2, \ldots. D_n$ a **relation *R*** at a given instance of time is a subset of

  $D_1 \times D_2 \times \ldots \times D_n$

  Thus, a relation is a set of n-tuples of $(a_1, a_2, \ldots, a_n)$ where each $a_i \in D_i$

- The current values (**relation instance**) of a relation are specified by a table
- An element ***t*** of ***R*** is a *tuple*, represented by a *row* in a table

154

Relations are Unordered

- Order of tuples is irrelevant (tuples may be stored in an arbitrary order)
- Example: INSTRUCTOR relation with unordered tuples

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

155

Keys

- Let an attribute K $\subseteq$ R
- *K* is a **superkey** of the Relation *R* if values for *K* are sufficient to identify a unique tuple.
  - Example: {SSN}, {*CWID*} and {CWID,name} are superkeys of *a student at LaTech.*
- Superkey *K* is a **candidate key** if *K* is **minimal**
  - Example: {SSN}, {*CWID*} is a candidate key for *STUDENT*
- One of the candidate keys is selected to be the **primary key**.
  - which one?

156

Exercise:

CWID	First Name	Last Name	DoB	SSN	Passport	Program
11122333	Jack	Smith	27-Mar-1987	733-34-272	L456781	Computer Science
22233444	Mary	Doe	13-May-1994	555-22-122	Null	Cyber Engineering
33344555	Danny	Do	13-May-1996	123-45-276	Null	BioMED
44455666	Farah	Mendez	13-May-1994	346-09-987	Z6541113	Electrical
55566777	Kyle	Patrick	23-Aug-2007	876-56-423	A129331	Computer Science

- **Task:** Identify the super keys, candidate keys, and primary keys

157

Exercise:

CWID	First Name	Last Name	DoB	SSN	Passport	Program
11122333	Jack	Smith	27-Mar-1987	733-34-272	L456781	Computer Science
22233444	Mary	Doe	13-May-1994	555-22-122	Null	Cyber Engineering
33344555	Danny	Mo	13-May-1996	123-45-276	Null	BioMED
44455666	Farah	Mendez	13-May-1994	346-09-987	Z6541113	Electrical
55566777	Kyle	Patrick	23-Aug-2007	876-56-423	A129331	Computer Science

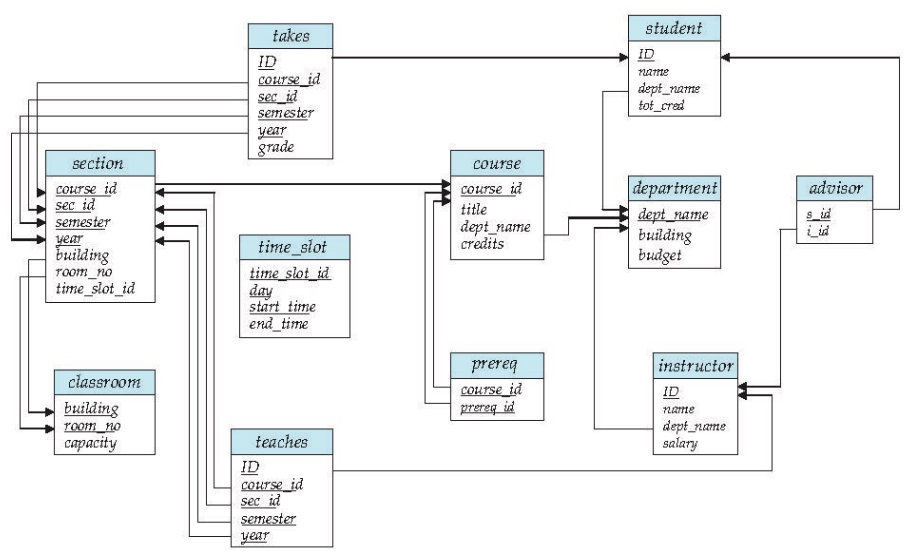
- Super keys: CWID,{CWID,DoB}, SSN, {First Name, Last Name}
- Candidate Keys: CWID, SSN
- Primary Key: your choice of Candidate Key

158

Keys

- **Foreign key** constraint: **Value in one relation must appear in another**
  - **Referencing** relation
  - **Referenced** relation
  - Example – *dept_name* in i*nstructor* is a foreign key from *instructor* referencing *department*

160



161

Points to remember: Equivalent Representations

- We know Relation is a set of tuples and not a list of tuples.
  - Order in which we present the tuples does not matter.

CoursesTaken(Student, Course, Grade)

- The attribute in a schema is also a set.
  - Schema is the same irrespective of order of attributes.

CoursesTaken (Student, Grade, Course)

  - We specify a “standard” order when we introduce schema.

162

Points to remember: Degree and Cardinality

CoursesTaken:

Student	Course	Grade
Amy, O'Brian	CSC-100	A
Bob, Castillo	CSC-100	F
Candice, De Mello	CSC-101	A
Darrel, West	CSC-101	B

- **Degree:** is the number of attributes in a schema
  - The degree for the above table is 3
- **Cardinality:** is the number of tuples in relation
  - The cardinality for the above table is 4

163

Questions

164

RELATIONAL ALGEBRA

165

Relational Query Languages

- Procedural vs .non-procedural, or declarative
- “Pure” languages:
  - Relational algebra
  - Tuple relational calculus
  - Domain relational calculus
- The above 3 pure languages are equivalent in computing power
- We will concentrate on **relational algebra**
  - consists of 6 basic operations

166

Queries in Relational Languages

- Each Query input is a table (or set of tables)
- Each Query output is a table.
- All data in the output table appears in one of the input tables
- Can we compute:
  - SUM
  - AVG
  - MAX
  - MIN

167

Select Operation – selection of rows (tuples)

A Unary Operator

Relation r

A	B	C	D
α	α	1	7
α	β	5	7
β	β	12	3
β	β	23	10

$\sigma_{((A=B)\land(D>5))}\ (r)$

A	B	C	D
α	α	1	7
β	β	23	10

168

Project Operation – selection of columns (Attributes)

A Unary Operator

Relation r:

A	B	C
α	10	1
α	20	1
β	30	1
β	40	2

$\Pi_{A,C}\ (r)$

A	C
α	1
α	1
β	1
β	2

=

A	C
α	1
β	1
β	2

169

Union of two relations

A Binary Operator

Relations *r, s:*

r

A	B
α	1
α	2
β	1

s

A	B
α	2
β	3

r ∪ s:

A	B
α	1
α	2
β	1
β	3

170

Set difference of two relations

A Binary Operator

Relations *r*, *s*:

r

A	B
α	1
α	2
β	1

s

A	B
α	2
β	3

r – *s:*

A	B
α	1
β	1

171

Set intersection of two relations

A Binary Operator

Relation *r, s*:

r

A	B
α	1
α	2
β	1

s

A	B
α	2
β	3

$r \cap s$

A	B
α	2

Note: $r \cap s = r - (r - s)$

172

joining two relations -- Cartesian-product

A Binary Operator

Relations *r, s*:

r

A	B
α	1
β	2

s

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

$r \times s$:

A	B	C	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b

173

Cartesian-product – naming issue

Relations *r*, *s*:

A	B
α	1
β	2

r

B	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

s

$r \times s$:

A	r.B	s.B	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b

174

Renaming a Table

A Unary Operator

- Allows us to refer to a relation, (say E) by more than one name.

$$\rho_x(E)$$

returns the expression *E* under the name *X*

Relations *r*

A	B
α	1
β	2

$r \times \rho_s(r)$

r.A	r.B	s.A	s.B
α	1	α	1
α	1	β	2
β	2	α	1
β	2	β	2

175

Composition of Operations

- Can build expressions using multiple operations
- Example: $\sigma_{A=C}(r \times s)$

A	B
α	1
β	2

r

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

s

176

Composition of Operations

- Can build expressions using multiple operations
- Example: $\sigma_{A=C}(\boxed{r \times s})$

A	B
α	1
β	2

r

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

s

- *r x s*

A	B	C	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b

177

Composition of Operations

- Can build expressions using multiple operations
- Example: $\sigma_{A=C}(r \times s)$

A	B
α	1
β	2

r

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

s

- *r x s*

A	B	C	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b

178

Composition of Operations

- Can build expressions using multiple operations
- Example: $\sigma_{A=C}(r \times s)$

A	B
α	1
β	2

r

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

s

- *r x s*

A	B	C	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b

- $\sigma_{A=C}(r \times s)$

A	B	C	D	E
α	1	α	10	a
β	2	β	10	a
β	2	β	20	b

179