

CS3743 Introduction to Databases

Relational Data Model

Weining Zhang

Department of Computer Science
University of Texas at San Antonio

May 31, 2009

Outline

- 1 Schema of Relational DB
 - Schema and Instances
- 2 Constraints of Relational DB
 - Domain Constraint
 - Access-By-Content Constraint
 - Key Integrity Constraint
 - Entity Integrity Constraint
 - Referential Integrity Constraint

Schema and Instances

- A database has a schema (defining structure of data) and many instances (actual data in the database at an instance of time)
- Database schema is typically static and stable; instances may change all the time
- At any time, the data in the database is called a database instance
- To create a database, one must design and specify a schema, and input data

Relation Instance

- A relation is a table (or equivalently, a set of tuples)
 - Each column is also called an attribute
 - Each row is also called a tuple
 - The number of rows is called the cardinality

Example

Relation name → **Students**

SID	Name	Age	Major	GPA
102	J. Smith	20	CS	3.2
105	M. Day	18	Math	2.9
112	B. Lee	19	EE	2.7

Attribute name → SID, Name, Age, Major, GPA

↑ Tuple → (105, M. Day, 18, Math, 2.9)

↑ Column → SID

Relation Schema

- The schema of a relation is written as $R(A_1, A_2, \dots, A_k)$
 - R is the name of the relation; A_i or $R.A_i$ is the name of an attribute of R ; k is the degree of R
- Each attribute A_i has a domain $Dom(A_i)$, which is the set of valid values of A_i
- If t is a tuple of relation $R(A_1, A_2, \dots, A_k)$, $t[A_i]$ or $t.A_i$ is the component of t under column A_i

Example

Let $t = \langle 105, M. Day, 18, Math, 2.9 \rangle$ be a tuple in Students(SID, Name, Age, Major, GPA), then $t[Age] = 18$, $t[Name, GPA] = \langle M. Day, 2.9 \rangle$

State of a Relational Database

Example

Students

SID	Name	Age	Major	GPA
102	J. Smith	20	CS	3.2
105	M. Day	18	Math	2.9
112	B. Lee	19	EE	2.7

Takes

SID	Cno	Term	Grade
102	CS374	F2000	B
105	CS455	S2000	B
112	CS100	F1999	A
105	Math210	F1999	C

Courses

Cno	Name	Hour	Dept
CS374	Database	3	CS
CS455	Network	3	CS
CS100	Prog.Lang	4	CS
Math210	Calculus	3	Math

Departments

Name	Room	Chair
CS	SB220	Hansen
EE	EB318	Johnson
Math	AB119	Miller

Schema of A Relational Database

Concept

The schema of a relational database consists of

- The schema of a set of relations, written as $S = \{R_1, R_2, \dots, R_m\}$, where each R_j is the schema of a relation
- A set of *Integrity Constraints* (ICs) that must be satisfied by any valid database state

Constraints of A Relational DB

ICs

Relation instances of any relational database must satisfy the following ICs.

- Domain constraint
- Access-by-content constraint
- Key constraint
- Entity integrity constraint
- Referential integrity constraint

A RDBMS will help to enforce these constraints.

Domain Constraint

The domain of any attribute must contain only atomic values (cannot contain sets or composite values)

Any relation satisfying domain constraint is said to be in the First Normal Form

Example

How to handle data that do have set values, represented in following non-relational table?

Employees

EID	Name	Age	Dependents
1234	Bob	34	{Allen, Ann}
1357	Mary	23	{Kathy}
2468	Peter	54	{Mike, Sue, David}

Ways to Handle Set Values

Use Multiple Columns

Employees

EID	Name	Age	Dep1	Dep2	Dep3
1234	Bob	34	Allen	Ann	
1357	Mary	23	Kathy		
2468	Peter	54	Mike	Sue	David

What are the pros and cons?

Ways to Handle Set Values

Use Multiple Tuples

Employees

EID	Name	Age	Dependents
1234	Bob	34	Allen
1234	Bob	34	Ann
1357	Mary	23	Kathy
2468	Peter	54	Mike
2468	Peter	54	Sue
2468	Peter	54	David

What are the pros and cons?

Ways to Handle Set Values

Use Multiple Tables

Employees

EID	Name	Age
1234	Bob	34
1357	Mary	23
2468	Peter	54

Dependents

EID	Name
1234	Allen
1234	Ann
1357	Kathy
2468	Mike
2468	Sue
2468	David

What are the pros and cons?

Access-By-Content Constraint

Concept

Each tuple can only be identified by some or all of its attribute values (rather than by other means, such as its position in the table or some system assigned id)

Notice

Tuple ordering and other non-content information may be used by the DBMS to efficiently retrieve the tuple. But, such information cannot be used by a user to formulate queries

Superkey

Definition

A subset of attributes of a relation is a superkey if each tuple has a unique value under these attributes

Example

Consider Students(SID, Name, Age, Major, GPA). If all students have different SID, then SID is a superkey. Also, the set of all attributes, namely {SID, Name, Age, Major, GPA}, is a superkey

Notice

An instance relation cannot be used to identify a superkey, unless the relation has no other instance

Key Integrity Constraint

Concept

Every relation must have a primary key

So, what is a primary key? We need to define several concepts about keys.

The Idea

Identify a small part of tuples that is sufficient to distinct tuples from each other

Properties of Superkey

- Any relation must have at least one superkey
- If K is a superkey of a relation R and S is a set of attributes of R that contains K, then S is also a superkey of R

Example

Assume that a relation has no other instance than the following one.

A	B	C	D
a ₁	b ₂	c ₁	d ₂
a ₁	b ₂	c ₃	d ₂
a ₁	b ₁	c ₂	d ₁
a ₂	b ₃	c ₄	d ₁

Is A a superkey? What about {A, B, C}

Candidate Key

Concept

A candidate key is a superkey that does not contain any other superkey

Example

Consider relation Students(SID, Name, Age, Major, GPA). If SID is a superkey, then so is {SID, Major}. However, while SID is also a candidate key, {SID, Major} is not

Primary Key

Concept

The primary key of a relation is a candidate key chosen by a database designer

- Choose primary key is a task of database design
- The choice is based on the size of the candidate key, the intended use of the data, and other considerations

Candidate Key

Example

Assume that a relation has no other instance than the following one.

A	B	C	D
a ₁	b ₂	c ₁	d ₂
a ₁	b ₂	c ₃	d ₂
a ₁	b ₁	c ₂	d ₁
a ₂	b ₃	c ₄	d ₁

Is {A,B} a candidate key? Why? Find all candidate keys.

Question

What properties does a candidate key have?

Entity Integrity Constraint

Concept

None of the attribute of the primary key of a relation can contain a null value

What is a null value? Why can not a primary key contain a null value?

Referential Integrity Constraint

Concept

A foreign key must contain either a null value or a value in the referenced key (i.e., must not contain unmatched non-null value)

What is a foreign key?

Definition

A foreign key *FK* of relation R_1 is a set of attributes of R_1 that corresponds one-to-one to attributes of a candidate key *CK* of (not necessarily distinct) relation R_2 , such that, every pair of matching attributes have the same domain

Foreign Key

Example

Employees.DName is a foreign key referencing Departments.Name

Employees

EID	Name	Age	DName
1234	Bob	34	Sales
1357	Mary	23	Service
2468	Peter	54	null

Departments

Name	City	Manager
Sales	Huston	Bill
Payroll	Dallas	Steve
Service	Chicago	Tom

Example

In Employees(EID, Name, Age, Dept, ManagerID), ManagerID can be a foreign key referencing EID