Chapter 4
The Relational Model

1

COMP211

Objectives

- **Terminologies of relational model.**
- **How tables are used to represent data.**
- **Properties of database relations.**
- **Candidate key, Primary key and Foreign Key.**
- **Meaning of entity integrity and referential integrity.**
- **Why do NoSQL systems emerge in the 2000s?**

2

COMP211

Relational Model Concepts

- The relational model was first proposed by Dr. T.F. Codd of IBM in 1970 in the following paper:

  "A Relational Model for Large Shared Data Banks," Communications of the ACM, June 1970.

- The above paper caused a major revolution in the field of Database management and earned Ted Codd the ACM Turing Award.

3

COMP211

Relational Model Terminologies

- Data is organized into tables (relations), with columns (attributes) and rows (tuples)
  - Only applies to logical structure of the database, not the physical structure.

- Attribute is a named column of a relation.

  Attribute types:
  - Simple and composite attributes
  - Single-valued and multi-valued attributes
 - Null attributes
  - Derived attributes

4

COMP211

Relational Model Terminologies

- Domain defines the **data type and the set of allowable values for one or more attributes.**
  - E.g. Dates have various formats such as yyyy-mm-dd, or dd mm,yyyy etc.
- Tuple is a row / record of a relation.
- Degree is the number of attributes in a relation.
- Cardinality is the number of tuples in a relation.

5

COMP211

Instances of Branch and Staff Relations (Fig. 4.1)
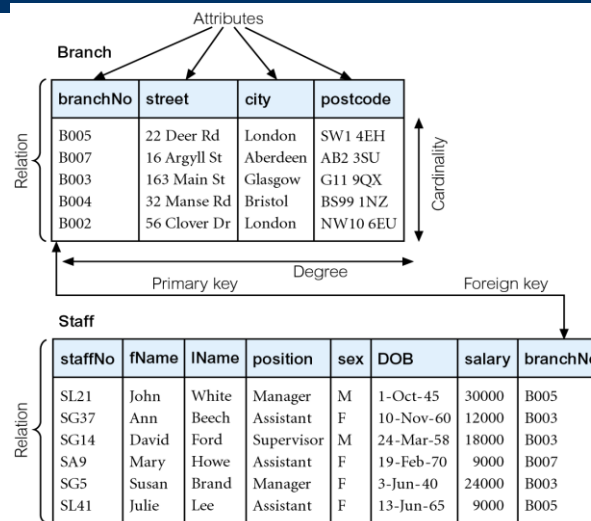
6

COMP211

Examples of Attribute Domains (Fig. 4.2)

Attribute	Domain Name	Meaning	Domain Definition
branchNo	BranchNumbers	The set of all possible branch numbers	character: size 4, range B001–B999
street	StreetNames	The set of all street names in Britain	character: size 25
city	CityNames	The set of all city names in Britain	character: size 15
postcode	Postcodes	The set of all postcodes in Britain	character: size 8
sex	Sex	The sex of a person	character: size 1, value M or F
DOB	DatesOfBirth	Possible values of staff birth dates	date, range from 1-Jan-20, format dd-mmm-yy
salary	Salaries	Possible values of staff salaries	monetary: 7 digits, range 6000.00–40000.00

Domain defines the **data type** and **the set of allowable values** for one or more attributes.

E.g. Dates have various formats such as yyyy-mm-dd, or dd mm,yyyy etc.

7

DEFINITION SUMMARY

Informal Terms	Formal Terms
Table	Relation
Column	Attribute
Row	Tuple
Values in a column	Domain
Table Definition	Schema of a Relation

8

Properties of Relations

- Relation name is distinct from all other relation names in relational schema.
- Each cell of relation contains exactly one atomic (single) value.
- Each attribute within a relation has a distinct name.
- Values of an attribute are all from the same domain.
- Each record is distinct; there are no duplicate records.
- Order of attributes has no significance.
- Order of records has no significance, theoretically.

9   COMP211

Candidate Key (CK)

- Candidate Key (CK)
    - An attribute, or smallest set of attributes, that uniquely identifies a record within a relation.
    - There may be several candidate keys for a relation
    - A key with more than one attribute is a **composite** key.

10   COMP211

Examples on Candidate Key

Consider the <u>Branch relation</u> in Fig. 4.1

- Given a value of city, we can determine several branch offices (e.g. London has 2 branch offices)
- Hence, city cannot be a candidate key.
- Given a branch number value, branchNo, we can determine at most one record.
- So, branchNo is a candidate key.

11   COMP211

Examples on Candidate Key

Consider the Viewing relation in Fig. 4.3

clientNo	propertyNo	viewDate	Comment
CR56	PA14	24-May-08	Too small
CR76	PG4	20-Apr-08	Too remote
CR56	PG4	26-May-08	
CR62	PA14	14-May-08	No dining room
CR56	PG36	28-Apr-08	

- Given a clientNo, there may be several viewings for different properties.
- Given a propertyNo, there may be several clients who viewed this property.
- Therefore, clientNo by itself or propertyNo by itself cannot be selected as a candidate key.
- However, combination of clientNo and propertyNo identifies at most one record for this instance.
- So, together they form the (composite) candidate key.
- **How about a client may view a property more than once?**

12   COMP211

Examples on Candidate Key

Consider the Viewing relation in Fig. 4.3

clientNo	propertyNo	viewDate	Comment
CR56	PA14	24-May-08	Too small
CR76	PG4	20-Apr-08	Too remote
CR56	PG4	26-May-08	
CR62	PA14	14-May-08	No dining room
CR56	PG36	28-Apr-08	

- **If a client may view a property more than once:**
  - **Given that your design already has (clientNo, propertyNo) as the PK, the operation of adding CR56 viewing PA14 will FAIL!!!!**
  - **In this case, ask if updating viewDate to hold the latest viewDate is feasible to solve the issue.**
  - **If NOT accepted, that means your design has to be changed to use (clientNo, propertyNo, viewDate) as the composite primary key.**

13

COMP211

Primary Key (PK)

- Primary Key (PK)
  - Candidate key selected to identify records uniquely within relation.
  - In the worst case, the entire set of attributes could serve as the primary key.
  - Every relation must have a primary key.

14

COMP211

Choices for Primary Key

In choosing the column(s) to be used as primary key, consider these properties:

- Values of primary key shall be unique, i.e. no duplicate value.

- Primary key shall always have a value. In other words, it shall not contain NULL.

- Primary key shall be simple and familiar.

- The value of primary key should not change.
  Primary key is used to reference other tables. If you change its value, you have to change all its references; otherwise, the references will be lost.

15   COMP211

Choices for Primary Key (con't)

- Primary key is usually a single column (e.g., customerID or productCode). But some circumstances, it could be made up of several columns. You should use as few columns as possible.

- Primary key often uses integer (or number) type. It could also be other types, such as texts. However, it is best to use numeric column as primary key for efficiency.

16   COMP211

Surrogate Key

- A surrogate key is an artificially generated key.
- Useful when your records essentially have no natural key.
- Most often, implemented as integers in an automatically incrementing field.
- The main advantage of the surrogate key is that they're easy to guarantee as unique.
- The main disadvantage is that they don't have any meaning.

17   COMP211

Surrogate key - disadvantage

- Primary key could take an arbitrary number. Most RDBMSs support so-called *auto-increment* (or AutoNumber type) for integer primary key, where (current maximum value + 1) is assigned to the new record. *However, in some circumstances, it is not desirable where a composite primary key is better to avoid inconsistency of data.*
  - E.g. We have a CustomerProducts table that relates customers to products with an "orderLimit" column to indicate the amount of product they are allowed to order.
    CustomerProducts (custID, productID, orderLimit)
    If you use an autonumber as the primary key, you might run into the following situation:
  - CustomerProductID CustomerID ProductID OrderLimit

CustomerProductID	CustomerID	ProductID	OrderLimit
1	1	100	25
2	1	100	30

  - What is the problem here? The primary issue is data integrity.

18   COMP211

Alternate Keys

- Alternate Keys
  - Candidate keys that are not selected to be primary key.

19 COMP211

Primary Key Example in MS Access

Primary key
(table in "design
view" in
MsAccess)

20 COMP211

Composite Primary Key Example in MS Access

Composite primary key

21 COMP211

Foreign Key (FK)

- A foreign key (FK) is a "copy" of a primary key that has been exported from one table and added as a new column in another table to represent the relationship between them.
- A foreign key is a copy of the whole of its parent primary key
  - if the primary key is composite then so is the foreign key
- A foreign key is needed to obtain the details from the parent table which are not duplicated in the child table because **data duplication might result in data inconsistency**.
- The foreign key is the minimal amount of data redundancy that is needed.

22 COMP211

Foreign Key (FK)

- For example, the inclusion of branchNo in both the Branch and Staff relations. (see Fig. 4.1)
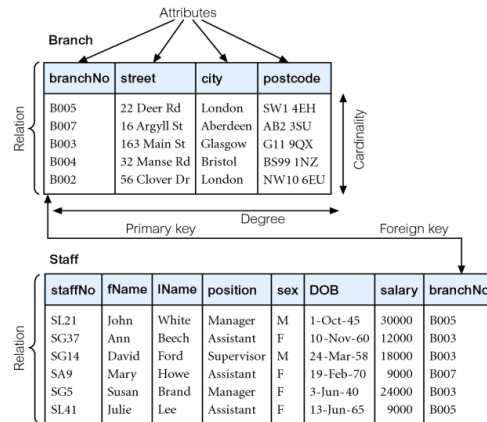
23

COMP211

Foreign Key Example

Textbook

Foreign Key

Title	ISBN	Edition	Publisher ID
Database Systems	0-201-34287-1	2	P091
Database Processing	0-02-366881-4	5	P473
Analysis and Design	0-202-36995-4	null	P091

Primary Key

Publisher

Primary Key

Publisher ID	Name
P091	Longman
P473	University Press

24

COMP211

12

Foreign Key Example in MS Access

Primary key of Lecturer table →

lecturer
- **Name**
- Department
- Address
- Telephone
- Title

1

∞

course
- **Code**
- Name
- Lecturer
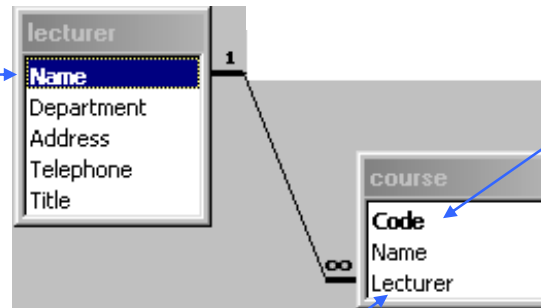
Primary key of Course table

Foreign key

Note: The name of the FK attribute does not need to match the PK – but the domain should be the same.

25

COMP211

Representing Relational Database Schemas

- A common convention for representing a relation schema is to give the name of the relation following by the attribute names in parentheses.
- Normally, the primary key is underlined.

    Branch (branchNo, street, city, postcode)

- The *conceptual model*, or *conceptual schema*, is the set of all such schemas for the database.
- The relational schema for part of the DreamHome case study is as follows:

Branch	(branchNo, street, city, postcode)
Staff	(staffNo, fName, lName, position, sex, DOB, salary, branchNo)
PropertyForRent	(propertyNo, street, city, postcode, type, rooms, rent, ownerNo, staffNo, branchNo)
Client	(clientNo, fName, lName, telNo, prefType, maxRent, eMail)
PrivateOwner	(ownerNo, fName, lName, address, telNo, eMail, password)
Viewing	(clientNo, propertyNo, viewDate, comment)
Registration	(clientNo, branchNo, staffNo, dateJoined)

26

COMP211

An instance of all such schemas for DreamHome case study (Fig. 4.3)

27 COMP211

Integrity Constraints

- Integrity constraints ensure that data is accurate.
- Domain constraints
  - form restrictions on the data type and set of values allowed for the attributes of relations.
- Two important integrity rules:
  - Entity integrity
  - Referential integrity

28 COMP211

Entity integrity

- Entity Integrity: In a base relation, no attribute of a primary key can be null.
  - Null
    - Represents value for an attribute that is currently unknown or not applicable for record.
    - Deals with incomplete or exceptional data.
    - Represents the absence of a value

29

COMP211

Referential integrity

- Referential Integrity
  - If foreign key exists in a relation, either foreign key value must match a candidate key value of some record in its home relation or foreign key value must be wholly null

30

COMP211

Referential integrity

- Staff oversees PropertyForRent

Staff

PK → (staffNo)

Relation

staffNo	fName	lName	position	sex	DOB	salary	branchNo
SL21	John	White	Manager	M	1-Oct-45	30000	B005
SG37	Ann	Beech	Assistant	F	10-Nov-60	12000	B003
SG14	David	Ford	Supervisor	M	24-Mar-58	18000	B003
SA9	Mary	Howe	Assistant	F	19-Feb-70	9000	B007
SG5	Susan	Brand	Manager	F	3-Jun-40	24000	B003
SL41	Julie	Lee	Assistant	F	13-Jun-65	9000	B005

FK

PropertyForRent

PK →

propertyNo	street	city	postcode	type	rooms	rent	ownerNo	staffNo	branchNo
PA14	16 Holhead	Aberdeen	AB7 5SU	House	6	650	CO46	SA9	B007
PL94	6 Argyll St	London	NW2	Flat	4	400	CO87	SL41	B005
PG4	6 Lawrence St	Glasgow	G11 9QX	Flat	3	350	CO40		B003
PG36	2 Manor Rd	Glasgow	G32 4QX	Flat	3	375	CO93	SG37	B003
PG21	18 Dale Rd	Glasgow	G12	House	5	600	CO87	SG37	B003
PG16	5 Novar Dr	Glasgow	G12 9AX	Flat	4	450	CO93	SG14	B003

31

Referential Integrity Example

Foreign Key

Course

Name	Level	Credits	Lecturer ID
Database Systems	Undergrad	3	38421
Data Networks	Undergrad	6	null
Speech Processing	Postgrad	null	34561

Note:

Foreign Key attribute name not necessarily the same as primary key

Primary Key

Violates referential integrity

Lecturer

Staff ID	StaffName
91027	Goscinski
38421	Nguyen

32

Why do NoSQL systems emerge in the 2000s?

- NoSQL systems relax the rigidity of storing data in tables by allowing a diverse set of data types.
- They allow for faster initial application development.
- However, NoSQL systems lack traditional systems' support for strong data consistency, instead relying on a weaker concept of eventual consistency.

33   COMP211

Summary

We have covered the following:
- **Relational model concepts**
  - Relations
  - Attributes
  - Relationship
- **Integrity constraints**
  - Entity integrity
  - Referential integrity
- **Terms:**
  - **Domain**
  - **Candidate key, Primary key and Foreign Key**

34   COMP211