COSC 3380
Design of Database Systems

Functional Dependencies and Normalization for Relational Databases

March 25, 2024

Informal Design Guidelines for Relational Databases

- What is relational database design?
  - The grouping of attributes to form "good" relation schemas

- Two levels of relation schemas
  - The logical "user view" level
  - The storage "base relation" level

- Design is concerned mainly with base relations

- What are the criteria for "good" base relations?

Informal Design Guidelines for Relational Databases

- Approaches to database design:
  - Bottom-up (*synthesis*)
    - Basic relationships among individual attributes to start with
    - Collecting large number of binary relationships among pairs of attributes
  - Top-down (*analysis*)
    - Group attributes into relations as they exist together naturally, in real-life situations
    - Analyze relations to figure out the details

Implicit goals of design activity

- Information preservation
  - Maintain all concepts – attribute types, entity types, relationship types
  - Everything from the conceptual design
- Minimum redundancy
  - Minimize redundant storage of same information
  - Reducing multiple updates to ensure consistency

Informal Design Guidelines for Relational Schemas

- Measures of quality
  - Making sure attribute semantics are clear
  - Reducing redundant information in tuples
  - Reducing NULL values in tuples
  - Disallowing possibility of generating spurious tuples

Semantics of the Relational Attributes must be clear

- **GUIDELINE 1**: Informally, each tuple in a relation should represent one entity or relationship instance.
  - Applies to individual relations and their attributes
  - Design relation schema so that it is easy to explain its real-world meaning
  - Attributes of different entities (EMPLOYEEs, DEPARTMENTs, PROJECTs) should not be mixed in the same relation
  - Only foreign keys should be used to refer to other entities
  - Entity and relationship attributes should be kept apart as much as possible.
  - Do not combine attributes from multiple entity types and relationship types into a single relation
- Bottom Line: *Design a schema that can be explained easily relation by relation. The semantics of attributes should be easy to interpret.*

COMPANY relational database schema

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date

DEPT_LOCATIONS

Dnumber	Dlocation

PROJECT

Pname	Pnumber	Plocation	Dnum

WORKS_ON

Essn	Pno	Hours

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship

A simplified COMPANY relational database schema

EMPLOYEE   F.K.

Ename	Ssn	Bdate	Address	Dnumber

P.K.

DEPARTMENT   F.K.

Dname	Dnumber	Dmgr_ssn

P.K.

DEPT_LOCATIONS

F.K.

Dnumber	Dlocation

P.K.

PROJECT   F.K.

Pname	Pnumber	Plocation	Dnum

P.K.

WORKS_ON

F.K.   F.K.

Ssn	Pnumber	Hours

P.K.

A simplified COMPANY state

EMPLOYEE

Ename	Ssn	Bdate	Address	Dnumber
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4
Wallace, Jennifer S.	987654321	1941-06-20	291Berry, Bellaire, TX	4
Narayan, Ramesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1

DEPARTMENT

Dname	Dnumber	Dmgr_ssn
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

WORKS_ON

Ssn	Pnumber	Hours
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	Null

PROJECT

Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

Mixing attributes from multiple entities

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn

EMP_PROJ

Ssn	Pnumber	Hours	Ename	Pname	Plocation

May be 'views', and not base relations!

Guideline 1

- Design relation schema so that it is easy to explain its real-world meaning
- Do not combine attributes from multiple entity types and relationship types into a single relation

Guideline 2

- Design base relation schemas so that no update anomalies are present in the relations
- If any anomalies are present:
  - Note them clearly
  - Make sure that application programs which update the database will operate correctly
  - Use triggers or stored procs for automatic updates

Redundant Information and Update Anomalies

- Grouping attributes into relation schemas
  - Significant effect on storage space
- Storing natural joins of base relations leads to update anomalies
- Types of update anomalies:
  - Insertion
  - Deletion
  - Modification

Redundant Information and Update Anomalies

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Redundant Information and Update Anomalies

- To insert new employees, all attributes related to the department have to be correctly entered in EMP_DEPT
- Difficult to add a new department with no employees – violation of entity integrity as SSN is a primary key
- If you delete an employee who happens to be the last employee in a department – that department is lost from the database
- Updating the manager of department in EMP_DEPT
  - Update all employees who work in that department

Guideline 2

- Design base relation schemas so that no update anomalies are present in the relations
- If any anomalies are present:
  - Note them clearly
  - Make sure that the programs that update the database will operate correctly
  - Use triggers or stored procs for automatic updates

Guideline 3

- Avoid placing attributes in a base relation whose values may frequently be NULL

- If NULLs are unavoidable:
  - Make sure they apply in exceptional cases only, not to a majority of tuples

NULL values in tuples

- When many attributes grouped together into a 'fat' relation
  - Can end up with many NULLs
- Problems with NULLs
  - Wasted storage space
  - Joins with NULL values
  - Understanding the meaning of attributes

NULL values in tuples

- If only a small number of employees have individual offices
  - Attribute *Office_number* in EMPLOYEE relation – not justified
  - New relation EMP_OFFICES (Essn, Office_number)
- Other examples?

Guideline 3

- Avoid placing attributes in a base relation whose values may frequently be NULL

- If NULLs are unavoidable:

  - Make sure that they apply in exceptional cases only, not to a majority of tuples