Database Management System – 30
Database design – Functional Dependencies and Keys

Ajay James
Asst. Prof in CSE
Government Engineering College Thrissur

Outline

- Functional Dependency
- Normalization
- Keys
- Super key
- Primary Key
- Candidate Key
- Secondary Key

Functional Dependencies

- Constraints that are derived from the meaning and interrelationships of the data attributes
- Used to specify formal measures of the "goodness" of relational designs
- Constraint between two sets of attributes
- FD and keys are used to define normal forms for relations

Functional Dependency

- $R = \{A_1, A_2, \ldots, A_n\}$
- A set of attributes X functionally determines a set of attributes Y if the value of X determines a unique value for Y
- **Definition**, denoted by $X \rightarrow Y$
  - between two sets of attributes X and Y that are subsets of R specifies a constraint on the possible tuples
  - constraint :- for any two tuples t_1 and t_2 in r(R) that have $t_1[X] = t_2[X]$, they must also have $t_1[Y] = t_2[Y]$.

Examples of FD constraints

- Social security number determines employee name
  - SSN → ENAME
- Project number determines project name and location
  - PNUMBER →{PNAME, PLOCATION}
- Employee ssn and project number determines the hours per week that the employee works on the project
  - {SSN, PNUMBER} → HOURS

Functional Dependency contd...

- An FD is a property of the attributes in the schema R
- The constraint must hold on every relation instance r(R)
- If K is a key of R, then K functionally determines all attributes in R
  - (since we never have two distinct tuples with t1[K]=t2[K])

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

Example

TEACH

Teacher	Course	Text
Smith	Data Structures	Bartram
Smith	Data Management	Martin
Hall	Compilers	Hoffman
Brown	Data Structures	Horowitz

- TEXT → COURSE (FD)
- TEACHER → COURSE (no FD)
- TEXT → TEACHER (no FD)
- COURSE → TEXT (no FD)

Defining FDs from instances

- In order to define the FDs, we need to understand the **meaning of the attributes** involved and the **relationship** between them
- An FD is a property of the attributes in the schema R
- Given the instance (population) of a relation, all we can conclude is that an FD may exist between certain attributes
- What we can definitely conclude is –that **certain FDs do not exist** because there are tuples that show a violation of those dependencies

Example

- A relation R(A, B, C, D) with its extension.

A	B	C	D
a1	b1	c1	d1
a1	b2	c2	d2
a2	b2	c2	d3
a3	b3	c4	d3

- A → B, B → A, D → C (No FD)
- B → C; C → B;
- {A, B} → C;
- {A, B} → D;
- {C, D} → B

Normal Forms Based on Primary Keys

1. Normalization of Relations
2. Practical Use of Normal Forms
3. Definitions of Keys and Attributes Participating in Keys
4. First Normal Form
5. Second Normal Form
6. Third Normal Form

Normalization of Relations

- First proposed by Codd
- Takes a relation schema through a series of tests to certify whether it satisfies a certain **normal form**
- Top-down fashion (relational design by analysis)
- Initially, Codd proposed 3 normal forms - first, second, and third normal form.
- A stronger definition of 3NF—called Boyce-Codd normal form (BCNF)—was proposed later by Boyce and Codd.

Normalization of Relations contd...

- Process of decomposing unsatisfactory "bad" relations by breaking up their attributes into smaller relations
- Process of analyzing the given relation schemas based on their FDs and primary keys to achieve the desirable properties of
  - (1) minimizing redundancy and
  - (2) minimizing the insertion, deletion, and update anomalies

Normal Form definition

- Refers to the highest normal form condition that it meets
- Indicates the degree to which it has been normalized
- Condition using keys and FDs of a relation to certify whether a relation schema is in a particular normal form
- Additional properties may be needed to ensure a good relational design (lossless join, dependency preservation)

Practical Use of Normal Forms

- Normalization - the resulting designs are of high quality and meet the desirable properties
- Constraints on which they are based are hard to understand or to detect
- Need not normalize to the highest possible normal form
  - (usually up to 3NF and BCNF. 4NF rarely used in practice.)
- Denormalization:
  - The process of storing the join of higher normal form relations as a base relation—which is in a lower normal form

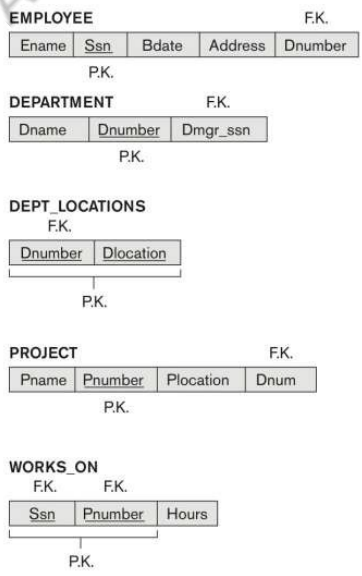
Definitions of Keys and Attributes Participating in Keys

- A **superkey** of a relation schema R = {A_1, A_2, ..., A_n} is a set of attributes **S** subset-of R with the property that no two tuples t_1 and t_2 in any legal relation state r of R will have $t_1[S] = t_2[S]$
- A **key K** is a superkey with the additional property that removal of any attribute from K will cause K not to be a superkey any more.
- {Ssn} is a key for EMPLOYEE, whereas {Ssn}, {Ssn, Ename}, {Ssn, Ename, Bdate}, and any set of attributes that includes Ssn are all superkeys.

Definitions of Keys and Attributes Participating in Keys contd...

- If a relation schema has more than one key, each is called a **candidate key**.
  - One of the candidate keys is arbitrarily designated to be the **primary key**, and the others are called **secondary keys**
- A **Prime attribute** must be a member of **some candidate key**
- A **Nonprime attribute** is not a prime attribute—that is, it is not a member of **any candidate key**.

Example



Example

- Student (Admisssion_no, Reg_no, Aadhar_no, Roll_no, Name, DOB, Address, Phone_no)
- (Admission_no, Name)
- (Reg_no,Phone_no)
- (Admission_no , Name,DOB)

Reference

- Elmasri R. and S. Navathe, Database Systems: Models, Languages, Design and Application Programming, Pearson Education 6th edition and 7th edition

Thank you