Chapter 10

Functional Dependencies and Normalization for Relational Databases
Chapter Outline

1 Informal Design Guidelines for Relational Databases
   1.1 Semantics of the Relation Attributes
   1.2 Redundant Information in Tuples and Update Anomalies
   1.3 Null Values in Tuples
   1.4 Spurious Tuples

2 Functional Dependencies (FDs)
   2.1 Definition of FD
   2.2 Inference Rules for FDs
   2.3 Equivalence of Sets of FDs
   2.4 Minimal Sets of FDs
Chapter Outline

- 3 Normal Forms Based on Primary Keys
  - 3.1 Normalization of Relations
  - 3.2 Practical Use of Normal Forms
  - 3.3 Definitions of Keys and Attributes Participating in Keys
  - 3.4 First Normal Form
  - 3.5 Second Normal Form
  - 3.6 Third Normal Form

- 4 General Normal Form Definitions (For Multiple Keys)

- 5 BCNF (Boyce-Codd Normal Form)
1 Informal Design Guidelines for Relational Databases (1)

- 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 (2)

- We first discuss informal guidelines for good relational design
- Then we discuss formal concepts of functional dependencies and normal forms
  - 1NF (First Normal Form)
  - 2NF (Second Normal Form)
  - 3NF (Third Normal Form)
  - BCNF (Boyce-Codd Normal Form)
- Additional types of dependencies, further normal forms, relational design algorithms by synthesis are discussed in Chapter 11
1.1 Semantics of the Relation Attributes

- **GUIDELINE 1:** Informally, each tuple in a relation should represent one entity or relationship instance. (Applies to individual relations and their attributes).
  - 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.

- **Bottom Line:** *Design a schema that can be explained easily relation by relation. The semantics of attributes should be easy to interpret.*
Figure 10.1 A simplified COMPANY relational database schema
1.2 Redundant Information in Tuples and Update Anomalies

- Information is stored redundantly
  - Wastes storage
  - Causes problems with update anomalies
    - Insertion anomalies
    - Deletion anomalies
    - Modification anomalies
EXAMPLE OF AN UPDATE ANOMALY

- Consider the relation:
  - EMP_PROJ(Emp#, Proj#, Ename, Pname, No_hours)

- Update Anomaly:
  - Changing the name of project number P1 from “Billing” to “Customer-Accounting” may cause this update to be made for all 100 employees working on project P1.
EXAMPLE OF AN INSERT ANOMALY

- Consider the relation:
  - EMP_PROJ(Emp#, Proj#, Ename, Pname, No_hours)
- Insert Anomaly:
  - Cannot insert a project unless an employee is assigned to it.
- Conversely
  - Cannot insert an employee unless he/she is assigned to a project.
EXAMPLE OF AN DELETE ANOMALY

- Consider the relation:
  - EMP_PROJ(Emp#, Proj#, Ename, Pname, No_hours)

- Delete Anomaly:
  - When a project is deleted, it will result in deleting all the employees who work on that project.
  - Alternately, if an employee is the sole employee on a project, deleting that employee would result in deleting the corresponding project.
Figure 10.3 Two relation schemas suffering from update anomalies.

(a) EMP_DEPT

<table>
<thead>
<tr>
<th>Ename</th>
<th>Ssn</th>
<th>Bdate</th>
<th>Address</th>
<th>Dnumber</th>
<th>Dname</th>
<th>Dmgr_ssn</th>
</tr>
</thead>
</table>

(b) EMP_PROJ

<table>
<thead>
<tr>
<th>Ssn</th>
<th>Pnumber</th>
<th>Hours</th>
<th>Ename</th>
<th>Pname</th>
<th>Plocation</th>
</tr>
</thead>
</table>

FD1

FD2

FD3
Figure 10.4 Example States for EMP_DEPT and EMP_PROJ

Example states for EMP_DEPT and EMP_PROJ resulting from applying NATURAL JOIN to the relations in Figure 10.2. These may be stored as base relations for performance reasons.

### EMP_DEPT

<table>
<thead>
<tr>
<th>Ename</th>
<th>Sn</th>
<th>Bdate</th>
<th>Address</th>
<th>Dnumber</th>
<th>Dname</th>
<th>Dmgr_sn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith, John B.</td>
<td>123456789</td>
<td>1965-01-09</td>
<td>731 Fondren, Houston, TX</td>
<td>5</td>
<td>Research</td>
<td>333445555</td>
</tr>
<tr>
<td>Wong, Franklin T.</td>
<td>333445555</td>
<td>1955-12-08</td>
<td>638 Yoss, Houston, TX</td>
<td>5</td>
<td>Research</td>
<td>333445555</td>
</tr>
<tr>
<td>Zelaya, Alicia J.</td>
<td>999887777</td>
<td>1968-07-19</td>
<td>3321 Castle, Spring, TX</td>
<td>4</td>
<td>Administration</td>
<td>987654321</td>
</tr>
<tr>
<td>Wallace, Jennifer S.</td>
<td>987654321</td>
<td>1941-06-20</td>
<td>291 Berry, Bellaire, TX</td>
<td>4</td>
<td>Administration</td>
<td>987654321</td>
</tr>
<tr>
<td>Narayan, Ramesh K.</td>
<td>666884444</td>
<td>1962-09-15</td>
<td>975 FireOak, Humble, TX</td>
<td>5</td>
<td>Research</td>
<td>333445555</td>
</tr>
<tr>
<td>English, Joyce A.</td>
<td>453453453</td>
<td>1972-07-31</td>
<td>5631 Rice, Houston, TX</td>
<td>5</td>
<td>Research</td>
<td>333445555</td>
</tr>
<tr>
<td>Jabbar, Ahmad V.</td>
<td>987987987</td>
<td>1969-03-20</td>
<td>980 Dallas, Houston, TX</td>
<td>4</td>
<td>Administration</td>
<td>987654321</td>
</tr>
<tr>
<td>Borg, James E.</td>
<td>988665555</td>
<td>1997-11-10</td>
<td>450 Stone, Houston, TX</td>
<td>1</td>
<td>Headquarters</td>
<td>988665555</td>
</tr>
</tbody>
</table>

### EMP_PROJ

<table>
<thead>
<tr>
<th>Sn</th>
<th>Pnumber</th>
<th>Hours</th>
<th>Ename</th>
<th>Pname</th>
<th>Plocation</th>
</tr>
</thead>
<tbody>
<tr>
<td>123456789</td>
<td>1</td>
<td>32.5</td>
<td>Smith, John B.</td>
<td>ProductX</td>
<td>Bellaire</td>
</tr>
<tr>
<td>123456789</td>
<td>2</td>
<td>7.5</td>
<td>Smith, John B.</td>
<td>ProductY</td>
<td>Sugarland</td>
</tr>
<tr>
<td>666884444</td>
<td>3</td>
<td>40.0</td>
<td>Narayan, Ramesh K.</td>
<td>ProductZ</td>
<td>Houston</td>
</tr>
<tr>
<td>453453453</td>
<td>1</td>
<td>20.0</td>
<td>English, Joyce A.</td>
<td>ProductX</td>
<td>Bellaire</td>
</tr>
<tr>
<td>453453453</td>
<td>2</td>
<td>20.0</td>
<td>English, Joyce A.</td>
<td>ProductY</td>
<td>Sugarland</td>
</tr>
<tr>
<td>333445555</td>
<td>2</td>
<td>10.0</td>
<td>Wong, Franklin T.</td>
<td>ProductY</td>
<td>Sugarland</td>
</tr>
<tr>
<td>333445555</td>
<td>3</td>
<td>10.0</td>
<td>Wong, Franklin T.</td>
<td>ProductZ</td>
<td>Houston</td>
</tr>
<tr>
<td>333445555</td>
<td>10</td>
<td>10.0</td>
<td>Wong, Franklin T.</td>
<td>Computerization</td>
<td>Stafford</td>
</tr>
<tr>
<td>333445555</td>
<td>20</td>
<td>10.0</td>
<td>Wong, Franklin T.</td>
<td>Reorganization</td>
<td>Houston</td>
</tr>
<tr>
<td>999887777</td>
<td>30</td>
<td>30.0</td>
<td>Zelaya, Alicia J.</td>
<td>Newbenefits</td>
<td>Stafford</td>
</tr>
<tr>
<td>999887777</td>
<td>10</td>
<td>10.0</td>
<td>Zelaya, Alicia J.</td>
<td>Computerization</td>
<td>Stafford</td>
</tr>
<tr>
<td>987987987</td>
<td>10</td>
<td>35.0</td>
<td>Jabbar, Ahmad V.</td>
<td>Computerization</td>
<td>Stafford</td>
</tr>
<tr>
<td>987987987</td>
<td>30</td>
<td>5.0</td>
<td>Jabbar, Ahmad V.</td>
<td>Newbenefits</td>
<td>Stafford</td>
</tr>
<tr>
<td>987654321</td>
<td>30</td>
<td>20.0</td>
<td>Wallace, Jennifer S.</td>
<td>Newbenefits</td>
<td>Stafford</td>
</tr>
<tr>
<td>987654321</td>
<td>20</td>
<td>15.0</td>
<td>Wallace, Jennifer S.</td>
<td>Reorganization</td>
<td>Houston</td>
</tr>
<tr>
<td>888665555</td>
<td>20</td>
<td>Null</td>
<td>Borg, James E.</td>
<td>Reorganization</td>
<td>Houston</td>
</tr>
</tbody>
</table>
Guideline to Redundant Information in Tuples and Update Anomalies

- GUIDELINE 2:
  - Design a schema that does not suffer from the insertion, deletion and update anomalies.
  - If there are any anomalies present, then note them so that applications can be made to take them into account.
1.3 Null Values in Tuples

- **GUIDELINE 3:**
  - Relations should be designed such that their tuples will have as few NULL values as possible
  - Attributes that are NULL frequently could be placed in separate relations (with the primary key)

- **Reasons for nulls:**
  - Attribute not applicable or invalid
  - Attribute value unknown (may exist)
  - Value known to exist, but unavailable
1.4 Spurious Tuples

- Bad designs for a relational database may result in erroneous results for certain JOIN operations.
- The "lossless join" property is used to guarantee meaningful results for join operations.

**GUIDELINE 4:**
- The relations should be designed to satisfy the lossless join condition.
- No spurious tuples should be generated by doing a natural-join of any relations.
Spurious Tuples (2)

- There are two important properties of decompositions:
  a) Non-additive or losslessness of the corresponding join
  b) Preservation of the functional dependencies.

- Note that:
  - Property (a) is extremely important and *cannot* be sacrificed.
  - Property (b) is less stringent and may be sacrificed. (See Chapter 11).
2.1 Functional Dependencies (1)

- Functional dependencies (FDs)
  - Are used to specify *formal measures* of the "goodness" of relational designs
  - And keys are used to define *normal forms* for relations
  - Are *constraints* that are derived from the *meaning* and *interrelationships* of the data attributes

- A set of attributes $X$ *functionally determines* a set of attributes $Y$ if the value of $X$ determines a unique value for $Y$
Functional Dependencies (2)

- $X \rightarrow Y$ holds if whenever two tuples have the same value for $X$, they **must have** the same value for $Y$
  - For any two tuples $t_1$ and $t_2$ in any relation instance $r(R)$: If $t_1[X] = t_2[X]$, then $t_1[Y] = t_2[Y]$

- $X \rightarrow Y$ in $R$ specifies a *constraint* on all relation instances $r(R)$

- Written as $X \rightarrow Y$; can be displayed graphically on a relation schema as in Figures. (denoted by the arrow: $\rightarrow$).

- FDs are derived from the real-world constraints on the attributes
Examples of FD constraints (1)

- Social security number determines employee name
  - SSN \(\rightarrow\) ENAME

- Project number determines project name and location
  - PNUMBER \(\rightarrow\) \{PNAME, PLOCATION\}

- Employee ssn and project number determines the hours per week that the employee works on the project
  - \{SSN, PNUMBER\} \(\rightarrow\) HOURS
Examples of FD constraints (2)

- 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 \( t1[K]=t2[K] \))
2.2 Inference Rules for FDs (1)

- Given a set of FDs $F$, we can infer additional FDs that hold whenever the FDs in $F$ hold.

- Armstrong's inference rules:
  - IR1. (Reflexive) If $Y \subset X$, then $X \rightarrow Y$.
  - IR2. (Augmentation) If $X \rightarrow Y$, then $XZ \rightarrow YZ$.
    - (Notation: $XZ$ stands for $X \cup Z$)
  - IR3. (Transitive) If $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow Z$.

- IR1, IR2, IR3 form a sound and complete set of inference rules.
  - These are rules that hold and all other rules that hold can be deduced from these.
Some additional inference rules that are useful:

- **Decomposition**: If $X \rightarrow YZ$, then $X \rightarrow Y$ and $X \rightarrow Z$
- **Union**: If $X \rightarrow Y$ and $X \rightarrow Z$, then $X \rightarrow YZ$
- **Psuedotransitivity**: If $X \rightarrow Y$ and $WY \rightarrow Z$, then $WX \rightarrow Z$

The last three inference rules, as well as any other inference rules, can be deduced from IR1, IR2, and IR3 (completeness property)
Inference Rules for FDs (3)

- **Closure** of a set $F$ of FDs is the set $F^+$ of all FDs that can be inferred from $F$

- **Closure** of a set of attributes $X$ with respect to $F$ is the set $X^+$ of all attributes that are functionally determined by $X$

- $X^+$ can be calculated by repeatedly applying IR1, IR2, IR3 using the FDs in $F$
2.3 Equivalence of Sets of FDs

Two sets of FDs $F$ and $G$ are equivalent if:
- Every FD in $F$ can be inferred from $G$, and
- Every FD in $G$ can be inferred from $F$
- Hence, $F$ and $G$ are equivalent if $F^+ = G^+$

Definition (Covers):
- $F$ covers $G$ if every FD in $G$ can be inferred from $F$
  - (i.e., if $G^+$ subset-of $F^+$)
- $F$ and $G$ are equivalent if $F$ covers $G$ and $G$ covers $F$
- There is an algorithm for checking equivalence of sets of FDs
2.4 Minimal Sets of FDs (1)

- A set of FDs is **minimal** if it satisfies the following conditions:
  1. Every dependency in F has a single attribute for its RHS.
  2. We cannot remove any dependency from F and have a set of dependencies that is equivalent to F.
  3. We cannot replace any dependency $X \rightarrow A$ in F with a dependency $Y \rightarrow A$, where $Y$ proper-subset-of $X$ ( $Y$ subset-of $X$) and still have a set of dependencies that is equivalent to F.
Minimal Sets of FDs (2)

- Every set of FDs has an equivalent minimal set
- There can be several equivalent minimal sets
- There is no simple algorithm for computing a minimal set of FDs that is equivalent to a set F of FDs
- To synthesize a set of relations, we assume that we start with a set of dependencies that is a minimal set
  - E.g., see algorithms 11.2 and 11.4
3 Normal Forms Based on Primary Keys

- 3.1 Normalization of Relations
- 3.2 Practical Use of Normal Forms
- 3.3 Definitions of Keys and Attributes Participating in Keys
- 3.4 First Normal Form
- 3.5 Second Normal Form
- 3.6 Third Normal Form
3.1 Normalization of Relations (1)

- **Normalization:**
  - The process of decomposing unsatisfactory "bad" relations by breaking up their attributes into smaller relations

- **Normal form:**
  - Condition using keys and FDs of a relation to certify whether a relation schema is in a particular normal form
Normalization of Relations (2)

- 2NF, 3NF, BCNF
  - based on keys and FDs of a relation schema
- 4NF
  - based on keys, multi-valued dependencies : MVDs; 5NF
  - based on keys, join dependencies : JDs (Chapter 11)
- Additional properties may be needed to ensure a good relational design (lossless join, dependency preservation; Chapter 11)
3.2 Practical Use of Normal Forms

- **Normalization** is carried out in practice so that the resulting designs are of high quality and meet the desirable properties.

- The practical utility of these normal forms becomes questionable when the constraints on which they are based are hard to understand or to detect.

- The database designers *need not* normalize to the highest possible normal form.
  - (usually up to 3NF, BCNF or 4NF)

- **Denormalization:**
  - The process of storing the join of higher normal form relations as a base relation—which is in a lower normal form.
3.3 Definitions of Keys and Attributes Participating in Keys (1)

- A **superkey** of a relation schema \( R = \{A_1, A_2, \ldots, A_n\} \) 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.
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.
3.2 First Normal Form

- **Disallows**
  - composite attributes
  - multivalued attributes
  - **nested relations**; attributes whose values for an *individual tuple* are non-atomic

- Considered to be part of the definition of relation
Figure 10.8 Normalization into 1NF

(a) A relation schema that is not in 1NF. (b) Example state of relation DEPARTMENT. (c) 1NF version of the same relation with redundancy.

### Figure 10.8

<table>
<thead>
<tr>
<th>Dname</th>
<th>Dnumber</th>
<th>Dmgr_ssn</th>
<th>Dlocations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>5</td>
<td>333445555</td>
<td>{Bellaire, Sugarland, Houston}</td>
</tr>
<tr>
<td>Administration</td>
<td>4</td>
<td>987654321</td>
<td>{Stafford}</td>
</tr>
<tr>
<td>Headquarters</td>
<td>1</td>
<td>888665555</td>
<td>{Houston}</td>
</tr>
</tbody>
</table>

### Figure 10.8

<table>
<thead>
<tr>
<th>Dname</th>
<th>Dnumber</th>
<th>Dmgr_ssn</th>
<th>Dlocation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>5</td>
<td>333445555</td>
<td>Bellaire</td>
</tr>
<tr>
<td>Research</td>
<td>5</td>
<td>333445555</td>
<td>Sugarland</td>
</tr>
<tr>
<td>Research</td>
<td>5</td>
<td>333445555</td>
<td>Houston</td>
</tr>
<tr>
<td>Administration</td>
<td>4</td>
<td>987654321</td>
<td>Stafford</td>
</tr>
<tr>
<td>Headquarters</td>
<td>1</td>
<td>888665555</td>
<td>Houston</td>
</tr>
</tbody>
</table>
Figure 10.9 Normalization nested relations into 1NF

(a) Schema of the EMP_PROJ relation with a nested relation attribute PROJS.
(b) Example extension of the EMP_PROJ relation showing nested relations within each tuple.
(c) Decomposition of EMP_PROJ into relations EMP_PROJ1 and EMP_PROJ2 by propagating the primary key.

<table>
<thead>
<tr>
<th>EMP_PROJ</th>
<th>Projs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ssn</td>
<td>Ename</td>
</tr>
<tr>
<td>123456789</td>
<td>Smith, John B.</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>666884444</td>
<td>Narayan, Ramesh K.</td>
</tr>
<tr>
<td>453453453</td>
<td>English, Joyce A.</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>333445555</td>
<td>Wong, Franklin T.</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
<tr>
<td>999987777</td>
<td>Zelaya, Alicia J.</td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>997987987</td>
<td>Jabbar, Ahmad V.</td>
</tr>
<tr>
<td>30</td>
<td></td>
</tr>
<tr>
<td>987654321</td>
<td>Wallace, Jennifer S.</td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
<tr>
<td>888665555</td>
<td>Borg, James E.</td>
</tr>
</tbody>
</table>
3.3 Second Normal Form (1)

- Uses the concepts of **FDs, primary key**
- **Definitions**
  - **Prime attribute**: An attribute that is member of the primary key K
  - **Full functional dependency**: a FD \( Y \rightarrow Z \) where removal of any attribute from \( Y \) means the FD does not hold any more
- **Examples**:
  - \( \{\text{SSN, PNUMBER}\} \rightarrow \text{HOURS} \) is a full FD since neither \( \text{SSN} \rightarrow \text{HOURS} \) nor \( \text{PNUMBER} \rightarrow \text{HOURS} \) hold
  - \( \{\text{SSN, PNUMBER}\} \rightarrow \text{ENAME} \) is not a full FD (it is called a partial dependency) since \( \text{SSN} \rightarrow \text{ENAME} \) also holds
Second Normal Form (2)

- A relation schema R is in **second normal form (2NF)** if every non-prime attribute A in R is fully functionally dependent on the primary key.

- R can be decomposed into 2NF relations via the process of 2NF normalization.
Figure 10.10 Normalizing into 2NF and 3NF

(a) Normalizing EMP_PROJ into 2NF relations. (b) Normalizing EMP_DEPT into 3NF relations.
Figure 10.11 Normalization into 2NF and 3NF

(a) LOTS

<table>
<thead>
<tr>
<th>Property_id#</th>
<th>County_name</th>
<th>Lot#</th>
<th>Area</th>
<th>Price</th>
<th>Tax_rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) LOTS1

<table>
<thead>
<tr>
<th>Property_id#</th>
<th>County_name</th>
<th>Lot#</th>
<th>Area</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) LOTS1A

<table>
<thead>
<tr>
<th>Property_id#</th>
<th>County_name</th>
<th>Lot#</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(d) LOTS1B

<table>
<thead>
<tr>
<th>Area</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD4</td>
<td></td>
</tr>
</tbody>
</table>

Figure 10.11
Normalization into 2NF and 3NF. (a) The LOTS relation with its functional dependencies FD1 through FD4. (b) Decomposing into the 2NF relations LOTS1 and LOTS2. (c) Decomposing LOTS1 into the 3NF relations LOTS1A and LOTS1B. (d) Summary of the progressive normalization of LOTS.
3.4 Third Normal Form (1)

- **Definition:**
  - **Transitive functional dependency:** a FD $X \rightarrow Z$ that can be derived from two FDs $X \rightarrow Y$ and $Y \rightarrow Z$

- **Examples:**
  - $SSN \rightarrow DMGRSSN$ is a transitive FD
    - Since $SSN \rightarrow DNUMBER$ and $DNUMBER \rightarrow DMGRSSN$ hold
  - $SSN \rightarrow ENAME$ is non-transitive
    - Since there is no set of attributes $X$ where $SSN \rightarrow X$ and $X \rightarrow ENAME$
Third Normal Form (2)

- A relation schema \( R \) is in **third normal form (3NF)** if it is in 2NF and no non-prime attribute \( A \) in \( R \) is transitively dependent on the primary key.
- \( R \) can be decomposed into 3NF relations via the process of 3NF normalization.
- **NOTE:**
  - In \( X \to Y \) and \( Y \to Z \), with \( X \) as the primary key, we consider this a problem only if \( Y \) is not a candidate key.
  - When \( Y \) is a candidate key, there is no problem with the transitive dependency.
  - E.g., Consider EMP (SSN, Emp#, Salary).
    - Here, SSN \( \to \) Emp\# \( \to \) Salary and Emp\# is a candidate key.
Normal Forms Defined Informally

- **1st normal form**
  - All attributes depend on the key

- **2nd normal form**
  - All attributes depend on the whole key

- **3rd normal form**
  - All attributes depend on nothing but the key
4 General Normal Form Definitions (For Multiple Keys) (1)

- The above definitions consider the primary key only
- The following more general definitions take into account relations with multiple candidate keys
- A relation schema \( R \) is in **second normal form (2NF)** if every non-prime attribute \( A \) in \( R \) is fully functionally dependent on every key of \( R \)
Definition:

- **Superkey** of relation schema R - a set of attributes S of R that contains a key of R

A relation schema R is in **third normal form (3NF)** if whenever a FD X -> A holds in R, then either:

  - (a) X is a superkey of R, or
  - (b) A is a prime attribute of R

**NOTE:** Boyce-Codd normal form disallows condition (b) above
5 BCNF (Boyce-Codd Normal Form)

- A relation schema R is in **Boyce-Codd Normal Form (BCNF)** if whenever an FD $X \rightarrow A$ holds in R, then $X$ is a superkey of R.

- Each normal form is strictly stronger than the previous one.
  - Every 2NF relation is in 1NF.
  - Every 3NF relation is in 2NF.
  - Every BCNF relation is in 3NF.

- There exist relations that are in 3NF but not in BCNF.
- The goal is to have each relation in BCNF (or 3NF).
**Figure 10.12 Boyce-Codd normal form**

(a) LOTS1A

<table>
<thead>
<tr>
<th>Property_id#</th>
<th>County_name</th>
<th>Lot#</th>
<th>Area</th>
</tr>
</thead>
</table>

FD1

FD2

FD5

BCNF Normalization

LOTS1AX

<table>
<thead>
<tr>
<th>Property_id#</th>
<th>Area</th>
<th>Lot#</th>
</tr>
</thead>
</table>

LOTS1AY

<table>
<thead>
<tr>
<th>Area</th>
<th>County_name</th>
</tr>
</thead>
</table>

(b) $R$

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
</table>

FD1

FD2

**Figure 10.12**

Boyce-Codd normal form. (a) BCNF normalization of LOTS1A with the functional dependency FD2 being lost in the decomposition. (b) A schematic relation with FDs; it is in 3NF, but not in BCNF.
A relation TEACH that is in 3NF but not in BCNF.

<table>
<thead>
<tr>
<th>Student</th>
<th>Course</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narayan</td>
<td>Database</td>
<td>Mark</td>
</tr>
<tr>
<td>Smith</td>
<td>Database</td>
<td>Navathe</td>
</tr>
<tr>
<td>Smith</td>
<td>Operating Systems</td>
<td>Ammar</td>
</tr>
<tr>
<td>Smith</td>
<td>Theory</td>
<td>Schulman</td>
</tr>
<tr>
<td>Wallace</td>
<td>Database</td>
<td>Mark</td>
</tr>
<tr>
<td>Wallace</td>
<td>Operating Systems</td>
<td>Ahamad</td>
</tr>
<tr>
<td>Wong</td>
<td>Database</td>
<td>Omiecinski</td>
</tr>
<tr>
<td>Zelaya</td>
<td>Database</td>
<td>Navathe</td>
</tr>
<tr>
<td>Narayan</td>
<td>Operating Systems</td>
<td>Ammar</td>
</tr>
</tbody>
</table>
Achieving the BCNF by Decomposition (1)

Two FDs exist in the relation TEACH:

- \( fd_1: \{ \text{student, course} \} \rightarrow \text{instructor} \)
- \( fd_2: \text{instructor} \rightarrow \text{course} \)

\( \{\text{student, course}\} \) is a candidate key for this relation and that the dependencies shown follow the pattern in Figure 10.12 (b).

- So this relation is in 3NF but not in BCNF

A relation **NOT** in BCNF should be decomposed so as to meet this property, while possibly forgoing the preservation of all functional dependencies in the decomposed relations.

- (See Algorithm 11.3)
Achieving the BCNF by Decomposition (2)

- Three possible decompositions for relation TEACH
  - \{student, instructor\} and \{student, course\}
  - \{course, instructor\} and \{course, student\}
  - \{instructor, course\} and \{instructor, student\}

- All three decompositions will lose fd1.
  - We have to settle for sacrificing the functional dependency preservation. But we cannot sacrifice the non-additivity property after decomposition.

- Out of the above three, only the 3rd decomposition will not generate spurious tuples after join. (and hence has the non-additivity property).

- A test to determine whether a binary decomposition (decomposition into two relations) is non-additive (lossless) is discussed in section 11.1.4 under Property LJ1. Verify that the third decomposition above meets the property.
Chapter Outline

- Informal Design Guidelines for Relational Databases
- Functional Dependencies (FDs)
  - Definition, Inference Rules, Equivalence of Sets of FDs, Minimal Sets of FDs
- Normal Forms Based on Primary Keys
- General Normal Form Definitions (For Multiple Keys)
- BCNF (Boyce-Codd Normal Form)