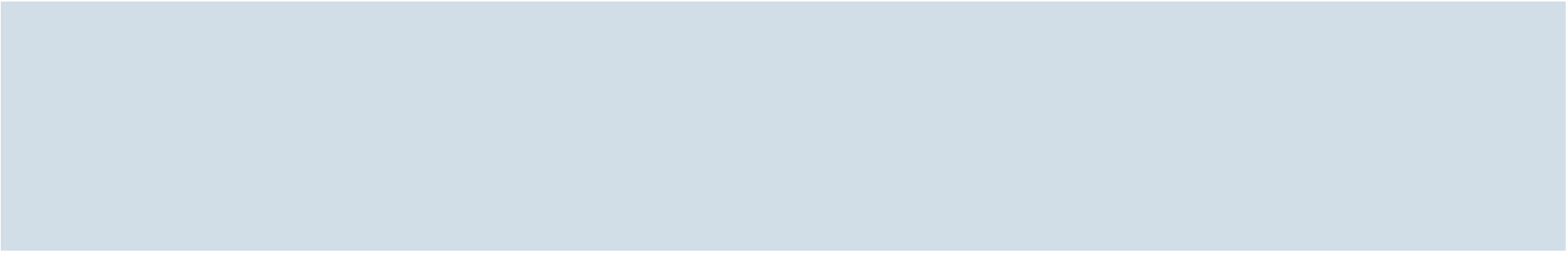


4CCS1DBS – Database Systems

Functional Dependencies and Normalisation for Relational Databases

- 
- Entity-Relationship Modelling
 - The Relational Data Model and Relational Database Constraints
 - Structured Query Language (SQL) - schema definition, constraints and queries
 - Relational Algebra
 - Functional Dependencies and Normalisation for Relational Databases

Purpose

- Develop measures for quality in database design
 - why is one grouping of attributes better than another?
- Two levels at which to discuss 'goodness' of design:
 - Logical or conceptual level
 - How users interpret the relation schemas and meaning of attributes?
 - Implementation or storage level
 - How are the tuples in a database stored and updated ?

Outline (1)

- Informal Design Guidelines for Relational Databases
 - Semantics of the Relation Attributes
 - Redundant Information in Tuples and Update Anomalies
 - Null Values in Tuples
 - Spurious Tuples
- Functional Dependencies (FDs)
 - Definition of FD
 - Inference Rules for FDs
 - Equivalence of Sets of FDs
 - Minimal Sets of FDs

Outline (2)

- Normal Forms Based on Primary Keys
 - Normalisation of Relations
 - Practical Use of Normal Forms
 - Definitions of Keys and Attributes Participating in Keys
 - First Normal Form
 - Second Normal Form
 - Third Normal Form
- General Normal Form Definitions (For Multiple Keys)
- BCNF (Boyce-Codd Normal Form)

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 (contd)

- 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)

Semantics of Relation Attributes

- **Semantics:** how to interpret the attribute values stored in a tuple of the relation, i.e. how the attribute values in a tuple relate to one another.
- Generally, the easier to explain the semantics of a relation, the better the relation schema design.
 - The *ease* with which the meaning of relation attributes can be explained, is an *informal* measure of relation design quality.

Semantics of Relation Attributes: Examples

EMPLOYEE F.K.

| | | | | |
|-------|------------|-------|---------|---------|
| Ename | <u>Ssn</u> | Bdate | Address | Dnumber |
|-------|------------|-------|---------|---------|

P.K.

DEPARTMENT F.K.

| | | |
|-------|----------------|----------|
| Dname | <u>Dnumber</u> | Dmgr_ssn |
|-------|----------------|----------|

P.K.

DEPT_LOCATIONS F.K.

| | |
|----------------|------------------|
| <u>Dnumber</u> | <u>Dlocation</u> |
|----------------|------------------|

P.K.

PROJECT F.K.

| | | | |
|-------|----------------|-----------|------|
| Pname | <u>Pnumber</u> | Plocation | Dnum |
|-------|----------------|-----------|------|

P.K.

WORKS_ON

| | | |
|------------|----------------|-------|
| <u>Ssn</u> | <u>Pnumber</u> | Hours |
|------------|----------------|-------|

P.K.

EMPLOYEE

| Ename | <u>Ssn</u> | 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 | <u>Dnumber</u> | Dmgr_ssn |
|----------------|----------------|-----------|
| Research | 5 | 333445555 |
| Administration | 4 | 987654321 |
| Headquarters | 1 | 888665555 |

DEPT_LOCATIONS

| <u>Dnumber</u> | <u>Dlocation</u> |
|----------------|------------------|
| 1 | Houston |
| 4 | Stafford |
| 5 | Bellaire |
| 5 | Sugarland |
| 5 | Houston |

WORKS_ON

| <u>Ssn</u> | <u>Pnumber</u> | 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 | <u>Pnumber</u> | 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 |

Semantics of Relation Attributes: Examples

Semantics of Relation Attributes: Examples

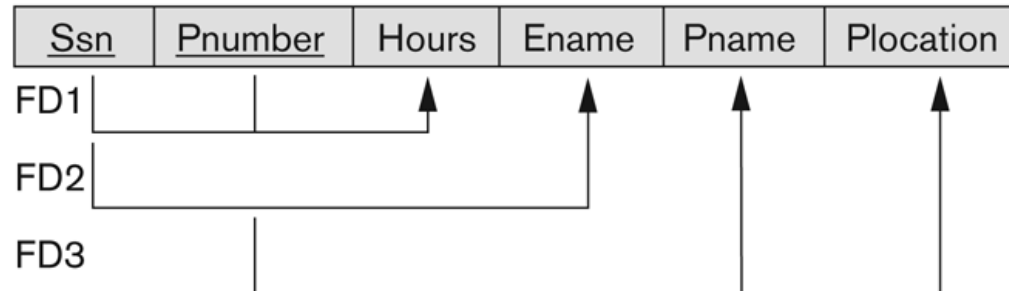
(a)

EMP_DEPT



(b)

EMP_PROJ



Semantics of Relation Attributes: Overall

- 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.
- Note: *Design a schema that can be explained easily relation by relation. The semantics of attributes should be easy to interpret.*

Redundant Information in Tuples and Update Anomalies

- When information is stored redundantly:
 - Wastes storage
 - Causes problems with update anomalies
 - Insertion anomalies
 - Deletion anomalies
 - Modification anomalies

Redundant Information in Tuples: example

EMPLOYEE

| Ename | <u>Ssn</u> | 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 | 291 Berry, 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 | <u>Dnumber</u> | Dmgr_ssn |
|----------------|----------------|-----------|
| Research | 5 | 333445555 |
| Administration | 4 | 987654321 |
| Headquarters | 1 | 888665555 |

EMP_DEPT

| Ename | <u>Ssn</u> | 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 |

Redundancy

Insertion Anomalies: example

- Consider the relation:

- EMP_PROJ (Emp#, Proj#, Ename, Pname, No_hours)

occurs when certain attributes cannot be inserted into the database without the presence of other attributes. For example this is the converse of delete anomaly - we can't add a new course unless we have at least one student enrolled on the course

- Insert Anomaly:

- Cannot insert a **new** project unless an employee is assigned to it.

- Conversely

- Cannot insert an employee unless a he/she is assigned to a project.

- Also: consistency problems when assigning projects to **every** employee

Deletion Anomalies: example

- Consider the relation:
 - EMP_PROJ (Emp#, Proj#, Ename, Pname, No_hours)
- Deletion Anomaly:
 - When a project is deleted, it will result in deleting all the employees who work on that project.
 - Alternatively, if an employee is the sole employee on a project, deleting that employee would result in deleting the corresponding project.

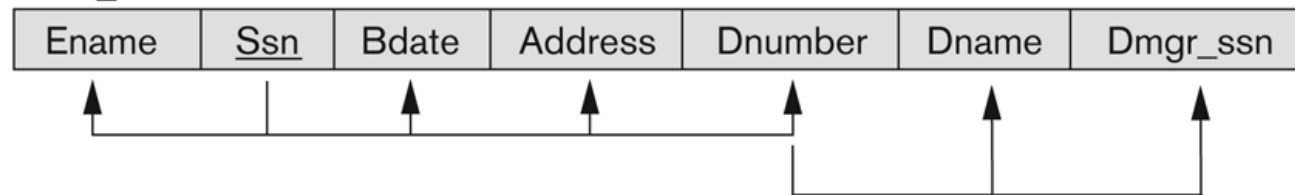
Update Anomalies: example

- Consider the relation:
 - EMP_PROJ (Emp#, Proj#, Ename, Pname, No_hours)
- Update Anomaly:
 - Changing the name of project number P1 (e.g from “Billing” to “Customer-Accounting”) may cause this update to be made for all 100 employees working on project P1.
 - If we fail to update some tuples, **inconsistencies** are introduced.

Two Relation Schemas Suffering from Update Anomalies

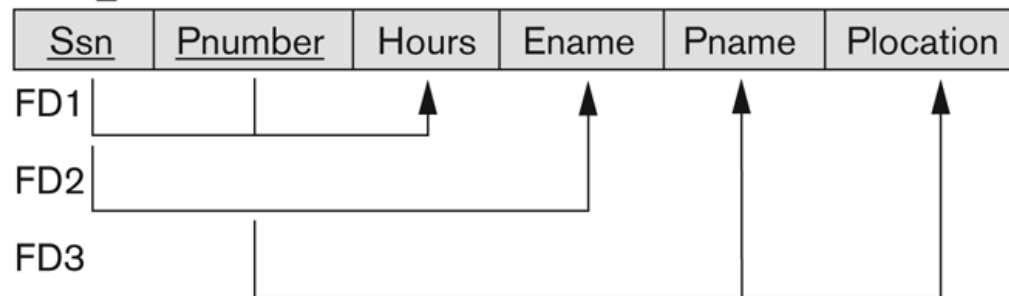
(a)

EMP_DEPT



(b)

EMP_PROJ



Base Relations EMP_DEPT and EMP_PROJ formed after a Natural Join with redundant information

Redundancy

EMP_DEPT

| Ename | <u>Ssn</u> | 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 |

Redundancy Redundancy

EMP_PROJ

| <u>Ssn</u> | <u>Pnumber</u> | Hours | Ename | Pname | Plocation |
|------------|----------------|-------|----------------------|-----------------|-----------|
| 123456789 | 1 | 32.5 | Smith, John B. | ProductX | Bellaire |
| 123456789 | 2 | 7.5 | Smith, John B. | ProductY | Sugarland |
| 666884444 | 3 | 40.0 | Narayan, Ramesh K. | ProductZ | Houston |
| 453453453 | 1 | 20.0 | English, Joyce A. | ProductX | Bellaire |
| 453453453 | 2 | 20.0 | English, Joyce A. | ProductY | Sugarland |
| 333445555 | 2 | 10.0 | Wong, Franklin T. | ProductY | Sugarland |
| 333445555 | 3 | 10.0 | Wong, Franklin T. | ProductZ | Houston |
| 333445555 | 10 | 10.0 | Wong, Franklin T. | Computerization | Stafford |
| 333445555 | 20 | 10.0 | Wong, Franklin T. | Reorganization | Houston |
| 999887777 | 30 | 30.0 | Zelaya, Alicia J. | Newbenefits | Stafford |
| 999887777 | 10 | 10.0 | Zelaya, Alicia J. | Computerization | Stafford |
| 987987987 | 10 | 35.0 | Jabbar, Ahmad V. | Computerization | Stafford |
| 987987987 | 30 | 5.0 | Jabbar, Ahmad V. | Newbenefits | Stafford |
| 987654321 | 30 | 20.0 | Wallace, Jennifer S. | Newbenefits | Stafford |
| 987654321 | 20 | 15.0 | Wallace, Jennifer S. | Reorganization | Houston |
| 888665555 | 20 | Null | Borg, James E. | Reorganization | Houston |

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.

Null Values in Tuples

- When grouping attributes in an enlarged relation:
 - End up with many nulls
 - Waste of storage
 - Problems in understanding meanings of attributes
 - Problems with aggregate functions (SUM etc)
- Reasons for nulls:
 - Attribute not applicable or invalid
 - Attribute value unknown (may exist)
 - Value known to exist, but unavailable

Null Values in Tuples (contd)

- 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)
- Example: if 10% of employees have individual offices, attribute should not be included in EMPLOYEE relation. Instead, use separate relation: EMP_OFFICES (ESSN, OFFICE-NUMBER).

(a)

EMP_LOCS

| <u>Ename</u> | <u>Plocation</u> |
|--------------|------------------|
|--------------|------------------|

P.K.

EMP_PROJ1

| <u>Ssn</u> | <u>Pnumber</u> | Hours | Pname | Plocation |
|------------|----------------|-------|-------|-----------|
|------------|----------------|-------|-------|-----------|

P.K.

(b)

EMP_LOCS

| Ename | Plocation |
|----------------------|------------|
| Smith, John B. | Bellaire |
| Smith, John B. | Surgarland |
| Narayan, Ramesh K. | Houston |
| English, Joyce A. | Bellaire |
| English, Joyce A. | Surgarland |
| Wong, Franklin T. | Surgarland |
| Wong, Franklin T. | Houston |
| Wong, Franklin T. | Stafford |
| Zelaya, Alicia J. | Stafford |
| Jabbar, Ahmad V. | Stafford |
| Wallace, Jennifer S. | Stafford |
| Wallace, Jennifer S. | Houston |
| Borg, James E. | Houston |

EMP_PROJ1

| Ssn | Pnumber | Hours | Pname | Plocation |
|-----------|---------|-------|-----------------|-----------|
| 123456789 | 1 | 32.5 | ProductX | Bellaire |
| 123456789 | 2 | 7.5 | ProductY | Sugarland |
| 666884444 | 3 | 40.0 | ProductZ | Houston |
| 453453453 | 1 | 20.0 | ProductX | Bellaire |
| 453453453 | 2 | 20.0 | ProductY | Sugarland |
| 333445555 | 2 | 10.0 | ProductY | Sugarland |
| 333445555 | 3 | 10.0 | ProductZ | Houston |
| 333445555 | 10 | 10.0 | Computerization | Stafford |
| 333445555 | 20 | 10.0 | Reorganization | Houston |
| 999887777 | 30 | 30.0 | Newbenefits | Stafford |
| 999887777 | 10 | 10.0 | Computerization | Stafford |
| 987987987 | 10 | 35.0 | Computerization | Stafford |
| 987987987 | 30 | 5.0 | Newbenefits | Stafford |
| 987654321 | 30 | 20.0 | Newbenefits | Stafford |
| 987654321 | 20 | 15.0 | Reorganization | Houston |
| 888665555 | 20 | NULL | Reorganization | Houston |

| EMP_PROJ | | | | | |
|-----------|---------|-------|----------------------|-----------------|-----------|
| Ssn | Pnumber | Hours | Ename | Pname | Plocation |
| 123456789 | 1 | 32.5 | Smith, John B. | ProductX | Bellaire |
| 123456789 | 2 | 7.5 | Smith, John B. | ProductY | Sugarland |
| 666884444 | 3 | 40.0 | Narayan, Ramesh K. | ProductZ | Houston |
| 453453453 | 1 | 20.0 | English, Joyce A. | ProductX | Bellaire |
| 453453453 | 2 | 20.0 | English, Joyce A. | ProductY | Sugarland |
| 333445555 | 2 | 10.0 | Wong, Franklin T. | ProductY | Sugarland |
| 333445555 | 3 | 10.0 | Wong, Franklin T. | ProductZ | Houston |
| 333445555 | 10 | 10.0 | Wong, Franklin T. | Computerization | Stafford |
| 333445555 | 20 | 10.0 | Wong, Franklin T. | Reorganization | Houston |
| 999887777 | 30 | 30.0 | Zelaya, Alicia J. | Newbenefits | Stafford |
| 999887777 | 10 | 10.0 | Zelaya, Alicia J. | Computerization | Stafford |
| 987987987 | 10 | 35.0 | Jabbar, Ahmad V. | Computerization | Stafford |
| 987987987 | 30 | 5.0 | Jabbar, Ahmad V. | Newbenefits | Stafford |
| 987654321 | 30 | 20.0 | Wallace, Jennifer S. | Newbenefits | Stafford |
| 987654321 | 20 | 15.0 | Wallace, Jennifer S. | Reorganization | Houston |
| 888665555 | 20 | Null | Borg, James E. | Reorganization | Houston |

How can EMP_PROJ be generated from EMP_PROJ1 and EMP_LOCS ?

Can we recover the information in EMP_PROJ from EMP_PROJ1 and EMP_LOCS ?

Why?

| | Ssn | Pnumber | Hours | Pname | Plocation | Ename |
|---|-----------|---------|-------|-----------------|-----------|--------------------|
| * | 123456789 | 1 | 32.5 | ProductX | Bellaire | Smith, John B. |
| | 123456789 | 1 | 32.5 | ProductX | Bellaire | English, Joyce A. |
| | 123456789 | 2 | 7.5 | ProductY | Sugarland | Smith, John B. |
| * | 123456789 | 2 | 7.5 | ProductY | Sugarland | English, Joyce A. |
| * | 123456789 | 2 | 7.5 | ProductY | Sugarland | Wong, Franklin T. |
| * | 666884444 | 3 | 40.0 | ProductZ | Houston | Narayan, Ramesh K. |
| | 666884444 | 3 | 40.0 | ProductZ | Houston | Wong, Franklin T. |
| * | 453453453 | 1 | 20.0 | ProductX | Bellaire | Smith, John B. |
| * | 453453453 | 1 | 20.0 | ProductX | Bellaire | English, Joyce A. |
| | 453453453 | 2 | 20.0 | ProductY | Sugarland | Smith, John B. |
| | 453453453 | 2 | 20.0 | ProductY | Sugarland | English, Joyce A. |
| * | 453453453 | 2 | 20.0 | ProductY | Sugarland | Wong, Franklin T. |
| * | 333445555 | 2 | 10.0 | ProductY | Sugarland | Smith, John B. |
| * | 333445555 | 2 | 10.0 | ProductY | Sugarland | English, Joyce A. |
| * | 333445555 | 2 | 10.0 | ProductY | Sugarland | Wong, Franklin T. |
| | 333445555 | 3 | 10.0 | ProductZ | Houston | Narayan, Ramesh K. |
| | 333445555 | 3 | 10.0 | ProductZ | Houston | Wong, Franklin T. |
| * | 333445555 | 10 | 10.0 | Computerization | Stafford | Wong, Franklin T. |
| | 333445555 | 20 | 10.0 | Reorganization | Houston | Narayan, Ramesh K. |
| | 333445555 | 20 | 10.0 | Reorganization | Houston | Wong, Franklin T. |

| | Ssn | Pnumber | Hours | Pname | Plocation | Ename |
|---|-----------|---------|-------|-----------------|-----------|--------------------|
| | 123456789 | 1 | 32.5 | ProductX | Bellaire | Smith, John B. |
| * | 123456789 | 1 | 32.5 | ProductX | Bellaire | English, Joyce A. |
| | 123456789 | 2 | 7.5 | ProductY | Sugarland | Smith, John B. |
| * | 123456789 | 2 | 7.5 | ProductY | Sugarland | English, Joyce A. |
| * | 123456789 | 2 | 7.5 | ProductY | Sugarland | Wong, Franklin T. |
| | 666884444 | 3 | 40.0 | ProductZ | Houston | Narayan, Ramesh K. |
| * | 666884444 | 3 | 40.0 | ProductZ | Houston | Wong, Franklin T. |
| * | 453453453 | 1 | 20.0 | ProductX | Bellaire | Smith, John B. |
| | 453453453 | 1 | 20.0 | ProductX | Bellaire | English, Joyce A. |
| * | 453453453 | 2 | 20.0 | ProductY | Sugarland | Smith, John B. |
| | 453453453 | 2 | 20.0 | ProductY | Sugarland | English, Joyce A. |
| * | 453453453 | 2 | 20.0 | ProductY | Sugarland | Wong, Franklin T. |
| * | 333445555 | 2 | 10.0 | ProductY | Sugarland | Smith, John B. |
| * | 333445555 | 2 | 10.0 | ProductY | Sugarland | English, Joyce A. |
| | 333445555 | 2 | 10.0 | ProductY | Sugarland | Wong, Franklin T. |
| * | 333445555 | 3 | 10.0 | ProductZ | Houston | Narayan, Ramesh K. |
| | 333445555 | 3 | 10.0 | ProductZ | Houston | Wong, Franklin T. |
| | 333445555 | 10 | 10.0 | Computerization | Stafford | Wong, Franklin T. |
| * | 333445555 | 20 | 10.0 | Reorganization | Houston | Narayan, Ramesh K. |
| | 333445555 | 20 | 10.0 | Reorganization | Houston | Wong, Franklin T. |

*
*

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 |

Spurious Tuples

- Bad designs for a relational database may result in erroneous results for certain JOIN operations.
- GUIDELINE 4:
 - Design relation schemas that can be joined with equality conditions on attributes that are either primary keys or foreign keys in a way that guarantees that no spurious tuples will be generated.
 - Avoid relations that contain matching attributes that are not (foreign key, primary key) combinations, as joining on such attributes will produce spurious tuples.

Summary of Design Guidelines

- Problems pointed out stemming from sub-optimal database design:
 - Anomalies that cause redundant work during insertion or modification, as well as accidental loss of information during deletion.
 - Waste of storage space due to nulls, difficulty in join or aggregation functions due to null values.
 - Generation of invalid or spurious data during joins or improperly related base relations.
- Next, discuss formal methods to determine quality of database design.

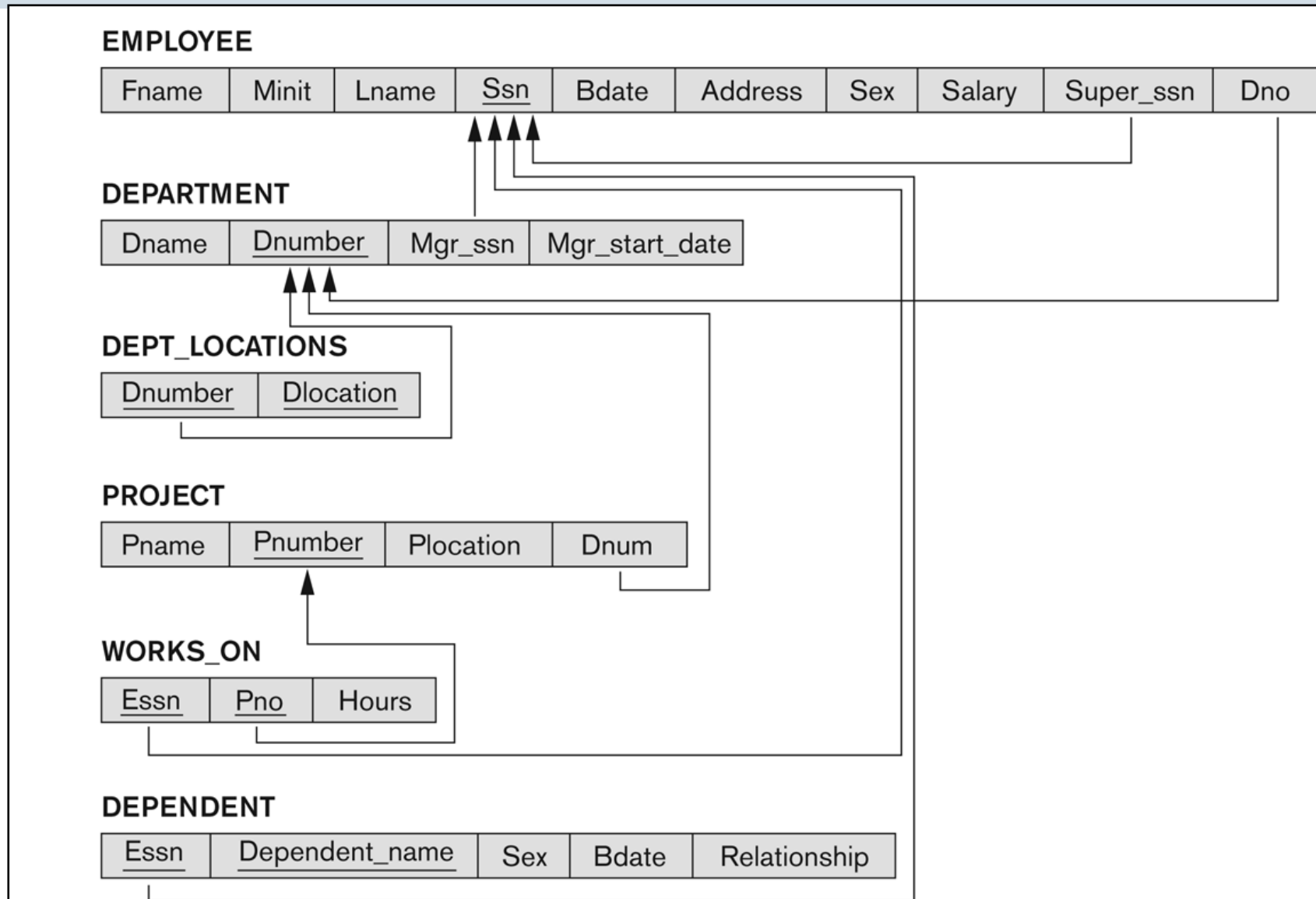
Functional Dependencies

- Functional dependencies (FDs)
 - Used to specify *formal measures* of how good a relational database design is
 - Keys are used to define **normal forms** for relations
 - **Constraints** 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 (contd)

- Functional dependency (FD) between attributes X and Y of relations R:
 - For any two tuples t1 and t2 in any relation instance r(R): If $t1[X]=t2[X]$, *then* $t1[Y]=t2[Y]$
 - $X \rightarrow Y$, if whenever two tuples have the same value for X, they *must have* the same value for 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 an arrow).
- FDs are derived from the real-world constraints on the attributes
- X: left-hand side, Y: right-hand side

Schema Showing Foreign Keys



Functional Dependency Example

EMP_PROJ

| | | | | | |
|------------|----------------|-------|-------|-------|-----------|
| <u>Ssn</u> | <u>Pnumber</u> | Hours | Ename | Pname | Plocation |
|------------|----------------|-------|-------|-------|-----------|

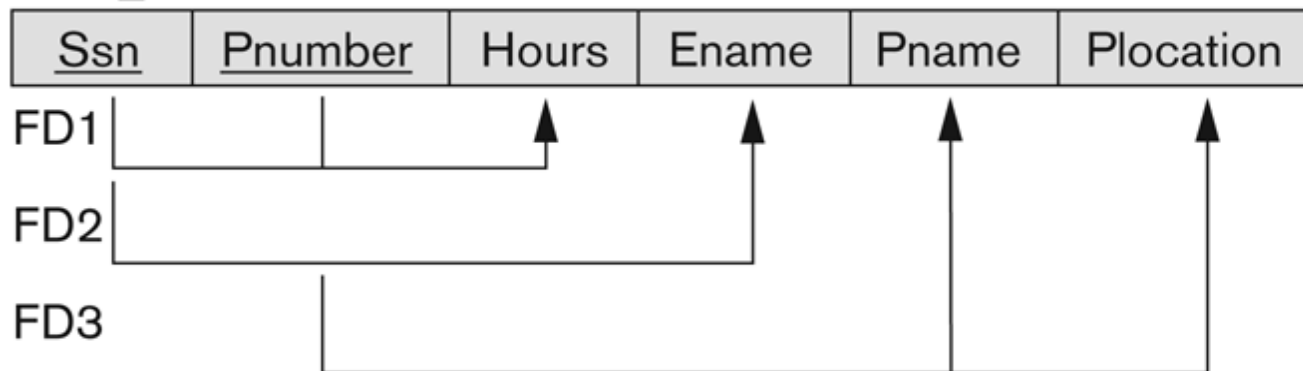
What are the functional dependencies in this relation?

Functional Dependency Example

EMP_PROJ

| <u>Ssn</u> | <u>Pnumber</u> | Hours | Ename | Pname | Plocation |
|------------|----------------|-------|-------|-------|-----------|
|------------|----------------|-------|-------|-------|-----------|

EMP_PROJ



Functional Dependency Example

EMP_DEPT

| | | | | | | |
|-------|------------|-------|---------|---------|-------|----------|
| Ename | <u>Ssn</u> | Bdate | Address | Dnumber | Dname | Dmgr_ssn |
|-------|------------|-------|---------|---------|-------|----------|

What are the functional dependencies in this relation?

Functional Dependency Example

EMP_DEPT

| | | | | | | |
|-------|------------|-------|---------|---------|-------|----------|
| Ename | <u>Ssn</u> | Bdate | Address | Dnumber | Dname | Dmgr_ssn |
|-------|------------|-------|---------|---------|-------|----------|

EMP_DEPT

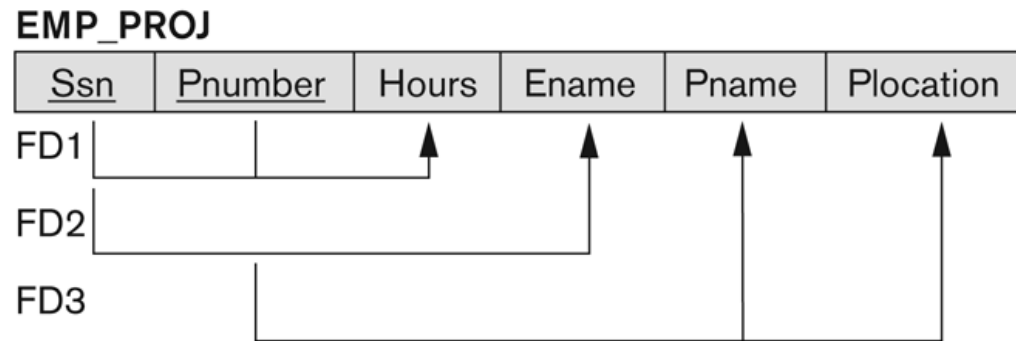


FDs in populated state

| | | | | | Redundancy | |
|----------------------|------------|------------|--------------------------|---------|----------------|-----------|
| EMP_DEPT | | | | | | |
| Ename | <u>Ssn</u> | 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 |

| | | | Redundancy | Redundancy | |
|-----------|---------|-------|----------------------|-----------------|-----------|
| EMP_PROJ | | | | | |
| Ssn | Pnumber | Hours | Ename | Pname | Plocation |
| 123456789 | 1 | 32.5 | Smith, John B. | ProductX | Bellaire |
| 123456789 | 2 | 7.5 | Smith, John B. | ProductY | Sugarland |
| 666884444 | 3 | 40.0 | Narayan, Ramesh K. | ProductZ | Houston |
| 453453453 | 1 | 20.0 | English, Joyce A. | ProductX | Bellaire |
| 453453453 | 2 | 20.0 | English, Joyce A. | ProductY | Sugarland |
| 333445555 | 2 | 10.0 | Wong, Franklin T. | ProductY | Sugarland |
| 333445555 | 3 | 10.0 | Wong, Franklin T. | ProductZ | Houston |
| 333445555 | 10 | 10.0 | Wong, Franklin T. | Computerization | Stafford |
| 333445555 | 20 | 10.0 | Wong, Franklin T. | Reorganization | Houston |
| 999887777 | 30 | 30.0 | Zelaya, Alicia J. | Newbenefits | Stafford |
| 999887777 | 10 | 10.0 | Zelaya, Alicia J. | Computerization | Stafford |
| 987987987 | 10 | 35.0 | Jabbar, Ahmad V. | Computerization | Stafford |
| 987987987 | 30 | 5.0 | Jabbar, Ahmad V. | Newbenefits | Stafford |
| 987654321 | 30 | 20.0 | Wallace, Jennifer S. | Newbenefits | Stafford |
| 987654321 | 20 | 15.0 | Wallace, Jennifer S. | Reorganization | Houston |
| 888665555 | 20 | Null | Borg, James E. | Reorganization | Houston |

Examples of FD Constraints



- 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 determine the hours per week that the employee works on the project
 - $\{SSN, PNUMBER\} \rightarrow HOURS$

Examples of FD Constraints

- 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
 - as we never have two distinct tuples with $t1[K]=t2[K]$

What is the Functional Dependency in this example?

TEACH

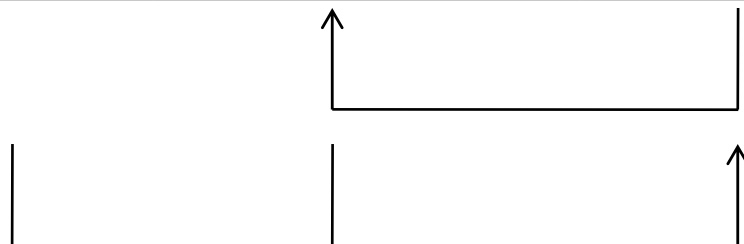
| Teacher | Course | Text |
|---------|-----------------|----------|
| Smith | Data Structures | Bartram |
| Smith | Data Management | Martin |
| Hall | Compilers | Hoffman |
| Brown | Data Structures | Horowitz |

FD's are a property of the meaning of data and hold at all times - certain FD's can be ruled out based on a given state of the database

Functional Dependency in TEACH Database

TEACH

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



A relation state of TEACH with a *possible* functional dependency $TEXT \rightarrow COURSE$. However, $TEACHER \rightarrow COURSE$ is ruled out.

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-of* 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 hold and all other rules that hold can be deduced from these

Inference Rules for FDs (contd)

- 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 $X \rightarrow YZ$
 - **Pseudotransitivity:**
 - 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)

Some Proofs

- **Decomposition:**

- If $X \rightarrow YZ$, then $X \rightarrow Y$

- **Proof:**

- $X \rightarrow YZ$ (given)
- using reflexive rule and knowing that YZ is-subset-of Y :
 $YZ \rightarrow Y$
- $X \rightarrow Y$

Some Proofs

- **Pseudotransitivity:**

- If $X \rightarrow Y$ and $WY \rightarrow Z$, then $WX \rightarrow Z$

- **Proof:**

- $X \rightarrow Y$ (given)
- $WY \rightarrow Z$ (given)
- $WX \rightarrow WY$ (using augmentation rule)
- $WX \rightarrow Z$ (using transitive rule)

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

Summary

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