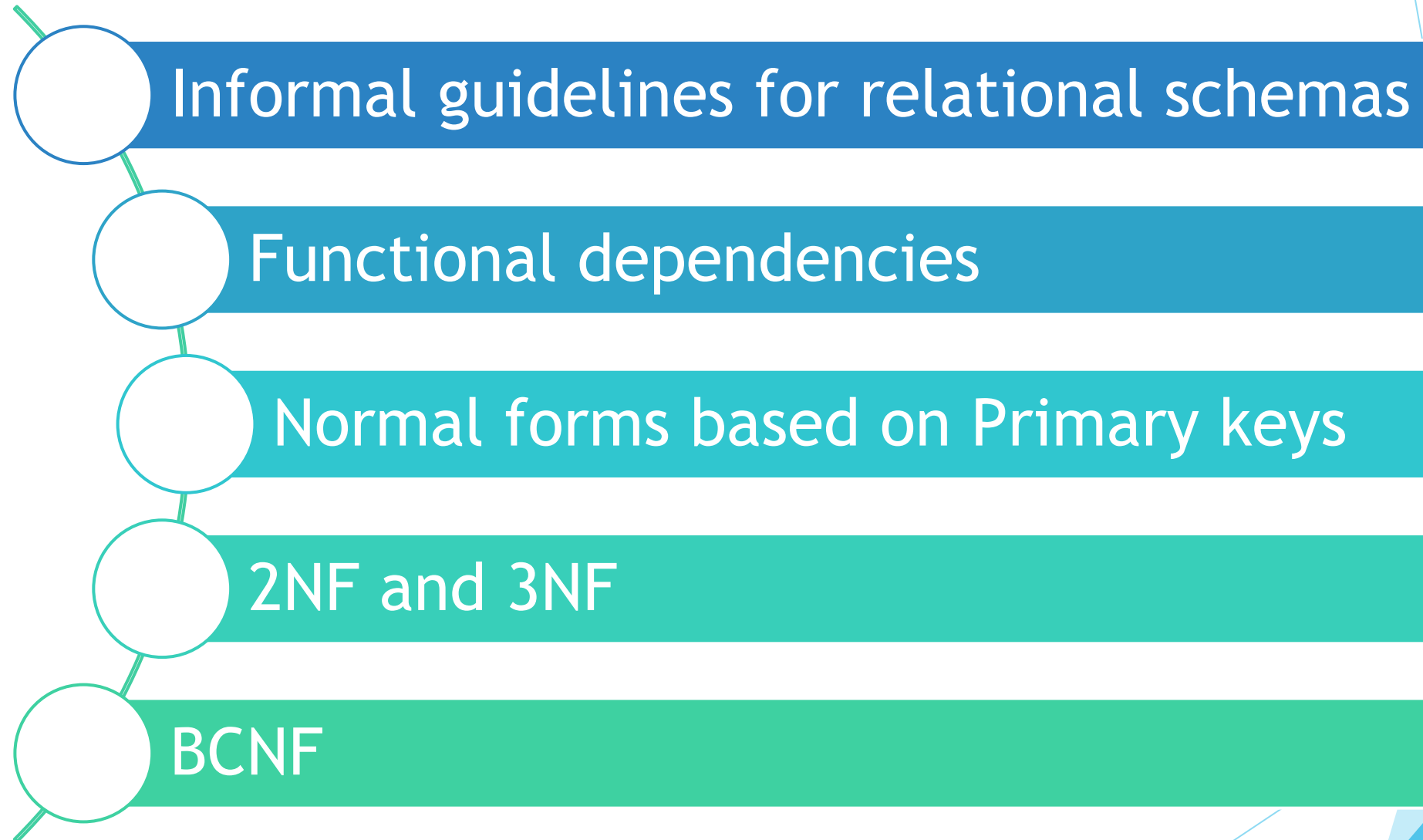


The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the left and right sides of the frame, creating a modern, layered effect. The central area is a plain, light grayish-white, providing a clean backdrop for the text.

Chapter 14

Normalization

Content

- 
- Informal guidelines for relational schemas
 - Functional dependencies
 - Normal forms based on Primary keys
 - 2NF and 3NF
 - BCNF

Informal Design Guideline for Relation Schemas

- ▶ We will discuss four informal guidelines that may be used as measures to determine the quality of relation schema design:
 - ▶ Making sure that the semantics of the attributes is clear in the schema
 - ▶ Reducing the redundant information in tuples
 - ▶ Reducing the NULL values in tuples
 - ▶ Disallowing the possibility of generating spurious tuples

Informal Design Guideline for Relation Schemas

- ▶ *Imparting Clear Semantics to Attributes in Relations*
- ▶ semantics of a relation:
 - ▶ **refers to its meaning** resulting from the interpretation of attribute values in a tuple.
 - ▶ i.e., what a relation exactly means and stands for

Informal Design Guideline for Relation Schemas

➤ Imparting Clear Semantics to Attributes in Relations

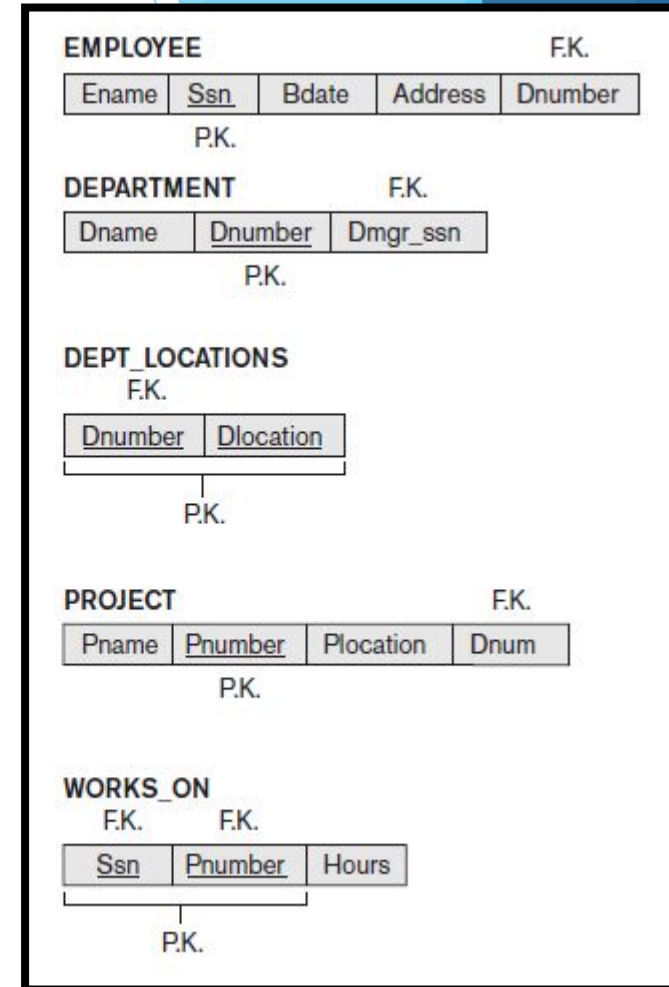
➤ The meaning of the EMPLOYEE relation schema is simple:

- Each tuple represents an employee, with values for:
 - the employee's name (Ename),
 - Social Security number (Ssn),
 - birth date (Bdate),
 - address (Address),
 - number of the department that the employee works for (Dnumber).

➤ **Dnumber attribute:** A FK that represents a relationship between EMPLOYEE and DEPARTMENT.

➤ The semantics of the DEPARTMENT and PROJECT schemas are also straightforward:

- Each DEPARTMENT tuple represents a department entity,
- Each PROJECT tuple represents a project entity.
- The attribute **Dmgr_ssn** of DEPARTMENT relates a department to the employee who is its manager,
- whereas **Dnum** of PROJECT relates a project to its controlling department; both are foreign key attributes.



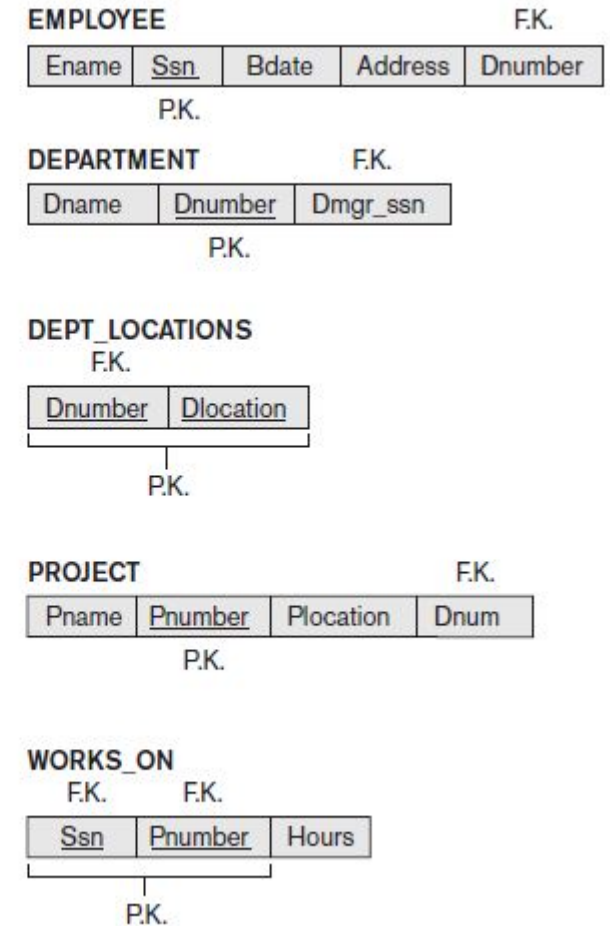
Informal Design Guideline for Relation Schemas

Imparting Clear Semantics to Attributes in Relations

- Each tuple in DEPT_LOCATIONS gives
 - a department number (Dnumber)
 - and one of the locations of the department (Dlocation).
- Each tuple in WORKS_ON gives
 - an employee Social Security number (Ssn),
 - the project number of one of the projects that the employee works on (Pnumber),
 - and the number of hours per week that the employee works on that project (Hours).

Figure 14.1

A simplified COMPANY relational database schema.



Informal Design Guideline for Relation Schemas

- ▶ Imparting Clear Semantics to Attributes in Relations
- ▶ Guideline 1.
- ▶ **Design** a relation schema so that it is **easy to explain its meaning**.
- ▶ **Do not combine attributes from multiple entity types and relationship types into a single relation.**
- ▶ If a relation schema corresponds to one **entity type or one relationship type**, it is straightforward to explain its meaning.
- ▶ Otherwise, if the relation corresponds to a **mixture of multiple entities and relationships**, semantic ambiguities will result, and the relation cannot be easily explained.

Informal Design Guideline for Relation Schemas

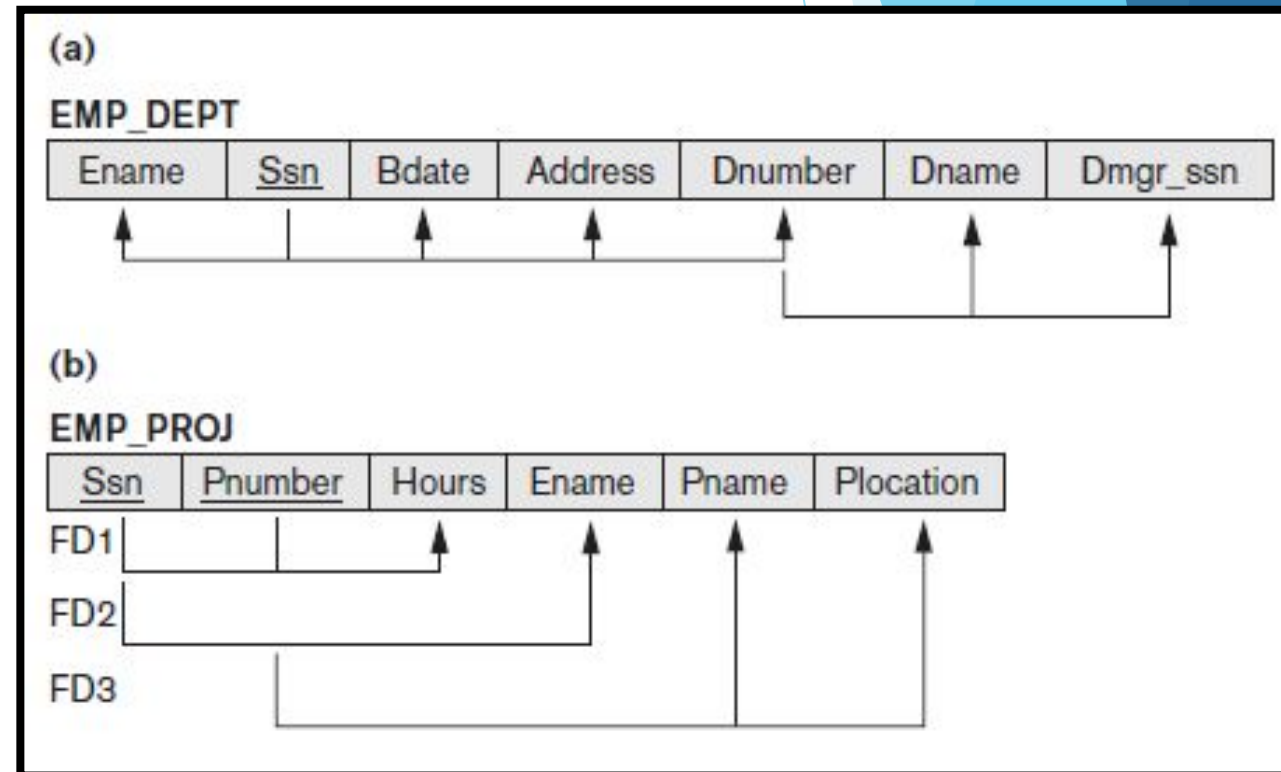
- ▶ Imparting Clear Semantics to Attributes in Relations

- ▶ Guideline 1.

- ▶ **Examples of Violating Guideline 1**

- ▶ The relation schemas in Figures (a) and (b) also have clear semantics.

- ▶ A tuple in the EMP_DEPT relation schema in Figure (a) represents a single employee but includes, along with the Dnumber (the identifier for the department he/she works for), additional information—namely, the name (Dname) of the department for which the employee works and the Social Security number (Dmgr_ssn) of the department manager.



Informal Design Guideline for Relation Schemas

▶ Imparting Clear Semantics to Attributes in Relations

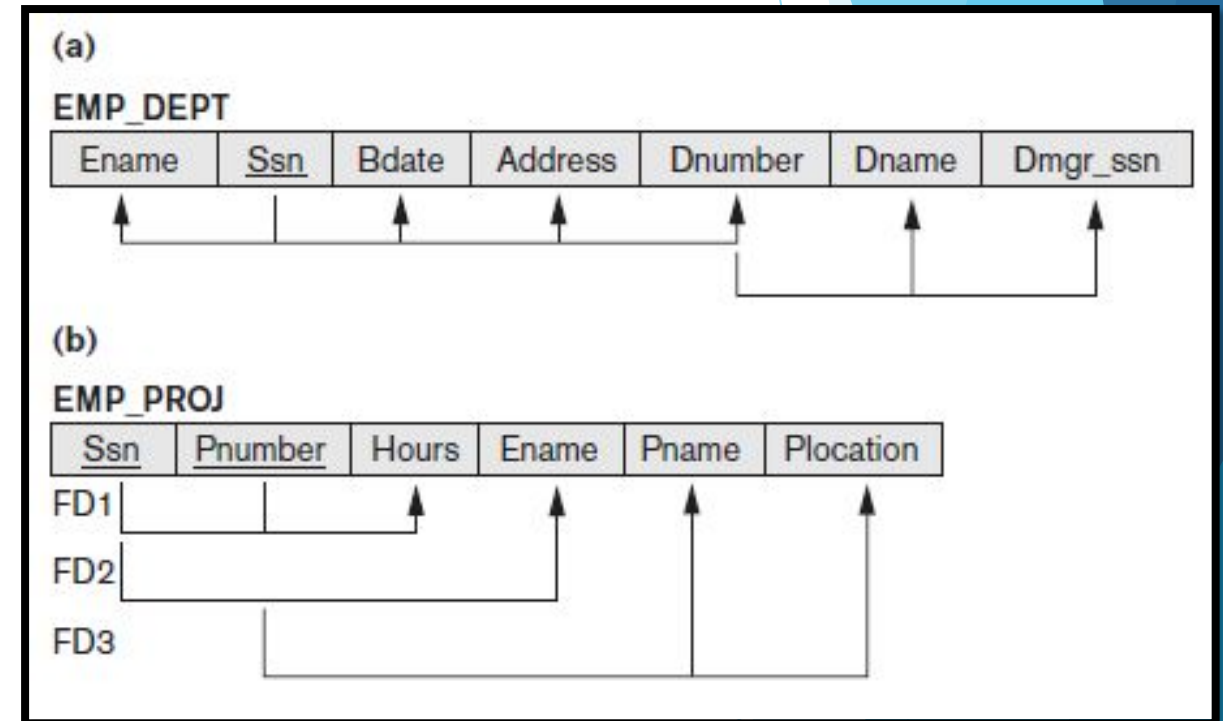
▶ Guideline 1.

▶ **Examples of Violating Guideline 1**

▶ For the EMP_PROJ relation in Figure 14.3(b), each tuple relates an employee to a project but also includes the employee name (Ename), project name (Pname), and project location (Plocation).

▶ Although there is nothing wrong logically with these two relations, they violate Guideline 1 by mixing attributes from distinct real-world entities:

- ▶ **EMP_DEPT** mixes attributes of employees and departments, and **EMP_PROJ** mixes attributes of employees and projects and the WORKS_ON relationship.



Informal Design Guideline for Relation Schemas

- ▶ Redundant Information in Tuples and Update Anomalies
- ▶ One goal of schema design is to **minimize the storage space used by the base relations**.
- ▶ For example,
 - ▶ **compare the space used by the two base relations EMPLOYEE and DEPARTMENT** with that for an **EMP_DEPT** base relation, (which is the result of applying the **NATURAL JOIN** operation to EMPLOYEE and DEPARTMENT).
- ▶ In EMP_DEPT, the attribute **values** pertaining to a particular department (Dnumber, Dname, Dmgr_ssn) are **repeated** for every employee who works for that department.

Informal Design Guideline for Relation Schemas

- ▶ Redundant Information in Tuples and Update Anomalies
- ▶ In contrast, each department's information appears only once in the DEPARTMENT relation.
- ▶ Only the department number (**Dnumber**) is repeated in the EMPLOYEE relation for each employee who works in that department as a foreign key.
- ▶ Similar comments apply to the EMP_PROJ relation (see Figure 14.4), which augments the WORKS_ON relation with additional attributes from EMPLOYEE and PROJECT.

Informal Design Guideline for Relation Schemas

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	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	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

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

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

Informal Design Guideline for Relation Schemas

- ▶ Redundant Information in Tuples and Update Anomalies
- ▶ Storing natural joins of base relations leads to an additional problem referred to as update anomalies.
- ▶ These can be classified into:
 - ▶ insertion anomalies,
 - ▶ deletion anomalies,
 - ▶ modification anomalies.
- ▶ **Insertion Anomalies**. Insertion anomalies can be differentiated into two types, illustrated by the following examples based on the **EMP_DEPT** relation.

Informal Design Guideline for Relation Schemas

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

Redundancy

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	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	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

- ▶ Redundant Information in Tuples and Update Anomalies
- ▶ **Insertion Anomalies**
- ▶ To insert a new employee tuple into EMP_DEPT, **we must include either the attribute values for the department that the employee works for, or NULLs (if the employee does not work for a department as yet).**
 - ▶ For example, to insert a new tuple for an employee who works in department number 5, we must enter all the attribute values of department 5 correctly so that they are consistent with the corresponding values for department 5 in other tuples in EMP DEPT.
 - ▶ In Figure 2, we do not have to worry about this consistency problem because we enter only the department number in the employee tuple; all other attribute values of department 5 are recorded only once in the database, as a single tuple in the DEPARTMENT relation.

Informal Design Guideline for Relation Schemas

EMP_DEPT					Redundancy	
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 in Tuples and Update Anomalies
- ▶ It is difficult to insert a new department that has no employees as yet in the EMP_DEPT relation.
 - ▶ The only way to do this is to place NULL values in the attributes for employee.
 - ▶ This violates the entity integrity for EMP_DEPT because its primary key SSN cannot be null.
 - ▶ This problem does not occur in the design of Figure 14.2 because a department is entered in the DEPARTMENT relation whether any employees work for it or not, and whenever an employee is assigned to that department, a corresponding tuple is inserted in EMPLOYEE.

Informal Design Guideline for Relation Schemas

EMP_DEPT					Redundancy	
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

- ▶ Redundant Information in Tuples and Update Anomalies
- ▶ **Deletion Anomalies.**
- ▶ If we delete from EMP_DEPT,
 - ▶ and that employee is the last employee working in a particular department,
 - ▶ then the information related to that department will be lost as well.
- ▶ This problem does not occur in the 14.2 designed database because DEPARTMENT tuples are stored separately.

Informal Design Guideline for Relation Schemas

EMP_DEPT					Redundancy	
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

- ▶ Redundant Information in Tuples and Update Anomalies
- ▶ **Modification Anomalies.**
 - ▶ In EMP_DEPT, if we change the value of one of the attributes of a particular department
 - ▶ —say, the manager of department 5
 - ▶ —we must update the tuples of all employees who work in that department;
 - ▶ otherwise, the database will become inconsistent.

Informal Design Guideline for Relation Schemas

- ▶ Redundant Information in Tuples and Update Anomalies
- ▶ Guideline 2.
- ▶ Design the base relation schemas so that no insertion, deletion, or modification anomalies are present in the relations.
- ▶ These guidelines may sometimes have to be violated in order to improve the performance of certain queries.
 - ▶ For example,
 - ▶ If EMP DEPT is stored as a materialized view in addition to the base relations of EMPLOYEE and DEPARTMENT, the anomalies in EMP DEPT must be noted and accounted for
 - ▶ for example, by using triggers or stored procedures that would make **automatic updates**.
- ▶ This way, whenever the base relation is updated, we do not end up with inconsistencies.

Informal Design Guideline for Relation Schemas

▶ NULL Values in Tuples

▶ Problem 1:

- ▶ If most of the attributes do not apply to all tuples in the relation, then we have many NULLs in the tuples.
- ▶ This can **waste space at the storage level.**

▶ Problem 2:

- ▶ What to do with NULLs when aggregate operations such as COUNT or SUM are applied.

▶ Problem 3:

- ▶ SELECT and JOIN operations involve comparisons; with NULL the results may become unpredictable.

Informal Design Guideline for Relation Schemas

- ▶ NULL Values in Tuples
- ▶ Moreover, **NULLs can have multiple interpretations**, such as the following:
 - ▶ The attribute **does not apply** to this tuple. For example, Visa_status may not apply to U.S. students.
 - ▶ The attribute **value for this tuple is unknown**. For example, the Date_of_birth may be unknown for an employee.
 - ▶ The value is **known but absent**; that is, it has not been recorded yet. For example, the Home_Phone_Number for an employee may exist, but may not be available and recorded yet.

Informal Design Guideline for Relation Schemas

- ▶ *NULL Values in Tuples*

- ▶ *Guideline 3.*

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

- ▶ If NULLs are **unavoidable**,

- ▶ apply in exceptional cases only
 - ▶ and do not apply to most tuples in the relation.

- ▶ For example,

- ▶ if only 15% of employees have individual offices,
 - ▶ Then there is no need to include an attribute **Office_number** in the EMPLOYEE relation;
 - ▶ rather, a relation **EMP_OFFICES (Essn, Office_number)** can be created to include tuples for only the employees with individual offices.

Informal Design Guideline for Relation Schemas

Generation of Spurious Tuples

(a)

EMP_LOCS

<u>Ename</u>	<u>Plocation</u>

P.K.

EMP_PROJ1

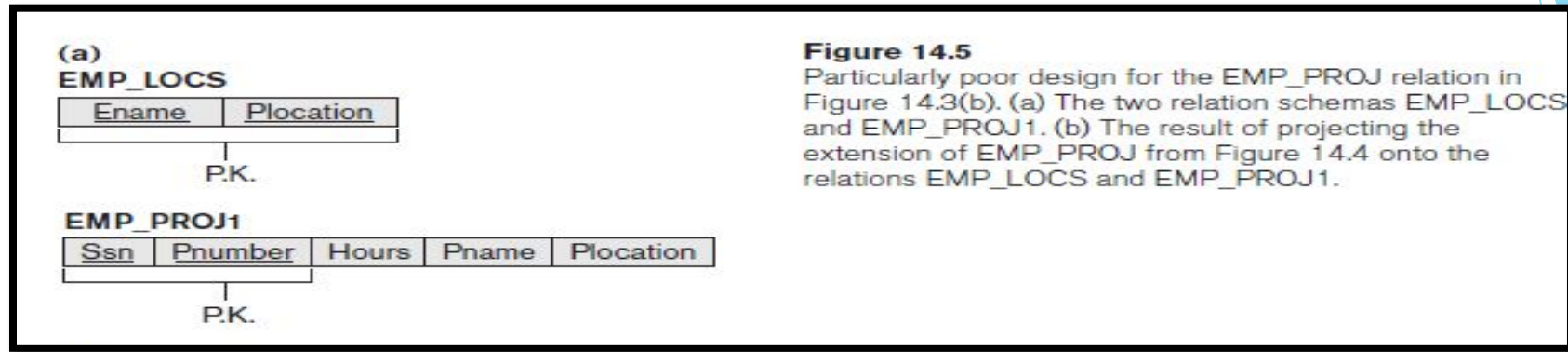
<u>Ssn</u>	<u>Pnumber</u>	Hours	Pname	Plocation

P.K.

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

Informal Design Guideline for Relation Schemas



- ▶ Generation of Spurious Tuples
- ▶ *Spurious tuples: tuples that actually doesn't exists in a table.*
- ▶ Consider the two relation schemas EMP_LOCS and EMP_PROJ1.
- ▶ A tuple in EMP_LOCS means that the employee whose name is Ename works on at least one project located at Plocation.
- ▶ A tuple in EMP_PROJ1 refers to the fact that the employee whose Social Security number is Ssn works the given Hours per week on the project whose name, number, and location are Pname, Pnumber, and Plocation.

Informal Design Guideline for Relation Schemas

- ▶ Generation of Spurious Tuples
- ▶ Figure 14.5(b) shows relation states of EMP_LOCS and EMP_PROJ1 corresponding to the EMP_PROJ relation in Figure 14.4.

(b)

EMP_LOCS

Ename	Plocation
Smith, John B.	Bellaire
Smith, John B.	Sugarland
Narayan, Ramesh K.	Houston
English, Joyce A.	Bellaire
English, Joyce A.	Sugarland
Wong, Franklin T.	Sugarland
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

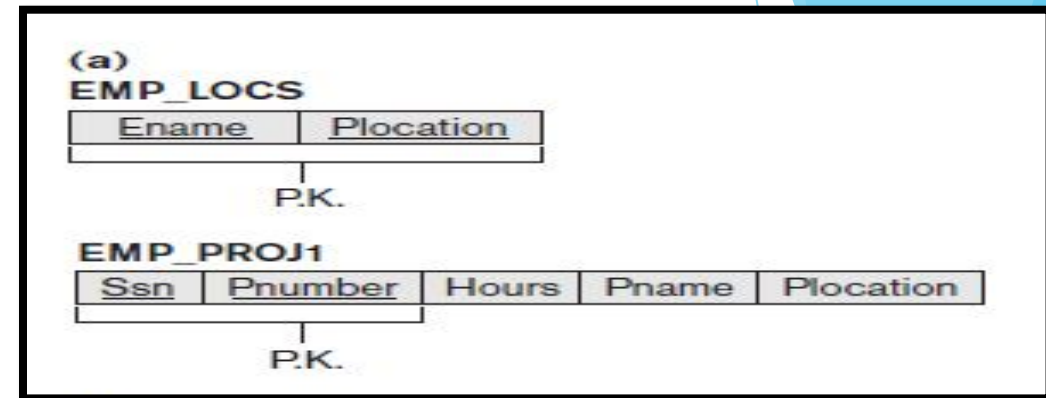
EMP_PROJ1

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

Informal Design Guideline for Relation Schemas

► Generation of Spurious Tuples

- Suppose that we used EMP_PROJ1 and EMP_LOCS as the base relations instead of EMP_PROJ.
- This produces a particularly **bad schema design** because we cannot recover the information that was originally in EMP_PROJ from EMP_PROJ1 and EMP_LOCS.
- If we attempt a NATURAL JOIN operation on EMP_PROJ1 and EMP_LOCS, the result produces many more tuples than the original set of tuples in EMP_PROJ.



			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	

Informal Design Guideline for Relation Schemas

► Generation of Spurious Tuples

- In Figure 14.6, the result of applying the join to only the tuples for employee with Ssn = “123456789” is shown (to reduce the size of the resulting relation).
- Additional tuples that were not in EMP_PROJ are called spurious tuples because they represent spurious information that is not valid.
- The spurious tuples are marked by asterisks (*) in Figure 14.6.

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.

Figure 14.6

Result of applying NATURAL JOIN to the tuples in EMP_PROJ1 and EMP_LOCS of Figure 14.5 just for employee with Ssn = “123456789”. Generated spurious tuples are marked by asterisks.

Informal Design Guideline for Relation Schemas

► Generation of Spurious Tuples

- Decomposing EMP_PROJ into EMP_LOCS and EMP_PROJ1 is undesirable because when we JOIN them back using NATURAL JOIN, we do not get the correct original information.
- This is because in this case **Plocation** happens to be the attribute that relates **EMP_LOCS** and **EMP_PROJ1**, and Plocation is neither a primary key nor a foreign key in either EMP_LOCS or EMP_PROJ1.

► Guideline 4.

- Design relation schemas so that they can be joined with equality conditions on attributes that are appropriately related (primary key, foreign key) pairs in a way that guarantees that no spurious tuples are generated.
- Join should be done on whole PK projected as FK in another table, not a portion of that.
- Avoid relations that contain matching attributes that are not (foreign key, primary key) combinations because joining on such attributes may produce spurious tuples.

Informal Design Guideline for Relation Schemas

- ▶ Summary and Discussion of Design Guidelines
- ▶ The problems are:
- ▶ ■ Anomalies:
 - ▶ cause redundant work to be done during insertion and modification of a relation,
 - ▶ cause accidental loss of information during a deletion from a relation
- ▶ ■ Waste of storage space due to NULLs and the difficulty of performing selections, aggregation operations, and joins due to NULL values
- ▶ ■ Generation of invalid and spurious data during joins on base relations with matched attributes that may not represent a proper (foreign key, primary key) relationship
- ▶ Strategy for achieving a **good design**:
 - ▶ decompose a badly designed relation appropriately to achieve higher normal forms.