Design guidelines for relational schema

Guideline 1:
- Semantic of the relation attributes
  - Do not mix attributes from distinct real world
- 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.

Guideline 1:
- Intuitively, 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.

example

EMP_DEPT

ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN

EMP_PROJ

SSN	PNUMBER	HOURS	ENAME	PNAME	PLOCATION

Guideline 1:
- A tuple in the EMP_DEPT relation schema represents a single employee but includes additional information- namely,
  - the name (DNAME) of the department for which the employee works and
  - the social security number (DMGRSSN) of the department manager.

Guideline 1:
- For the EMP_PROJ relation, each tuple relates an employee to a project but also includes
  - the employee name (ENAME),
  - project name (PNAME), and
  - project location (PLOCATION).

Guideline 1:

- Although there is nothing wrong logically with these two relations,
  - they are considered poor designs because
  - they violate Guideline 1 by mixing attributes from distinct real-world entities;
- They may be used as views, but they cause problems when used as base relations

Guideline 2:

- Reducing the redundant values in tuples
  - Reducing storage area
  - Avoiding update anomalies

Reducing storage area

- One goal of schema design is to minimize the storage space used by the base relations
- Grouping attributes into relation schemas has a significant effect on storage space.

Reducing storage area

- For example, compare the space used by the two base relations EMPLOYEE and DEPARTMENT with that for an EMP_DEPT base relation in, which is the result of applying the NATURAL JOIN operation to EMPLOYEE and DEPARTMENT.

EMPLOYEE

ENAME	SSN	BDATE	ADDRESS	DNUMBER
ith,John E.	123456789	1965-01-09	731 Fondren,Houston,TX	5
ng,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5
ava,Alicia J.	999887777	1968-07-19	3321 Castle,Spring,TX	4
llace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4
rayan,Ramesh K.	666884444	1962-09-15	975 Fire Oak,Humble,TX	5
glish,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5
bar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4
g,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1

DEPARTMENT

DNAME	DNUMBER	DMGRSSN
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

EMP DEPT						redundancy
NAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGR
John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5	Research	333443
Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5	Research	333443
Alicia J.	999887777	1968-07-19	3321 Castle,Spring,TX	4	Administration	987654
Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4	Administration	987654
Ramesh K.	666884444	1962-09-15	975 FireOak,Humble,TX	5	Research	333443
Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5	Research	333443
Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4	Administration	987654
James E.	888665555	1957-11-10	450 Stone,Houston,TX	1	Headquarters	888665

Reducing storage area

- In EMP_DEPT, the attribute values pertaining to a particular department (DNUMBER, DNAME, DMGRSSN) are repeated for *every employee who works for that department.*
- In contrast, each department's information appears only once in the DEPARTMENT relation

Reducing storage area

- Only the department number (DNUMBER) is repeated in the EMPLOYEE relation for each employee who works in that department.

Update Anomalies

- These can be classified into
  - insertion anomalies,
  - deletion anomalies, and
  - modification anomalies

Insertion Anomalies.

- Insertion anomalies can be differentiated into two types, illustrated by the following examples based on the EMP_DEPT relation:

EMP_DEPT						redundancy
ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
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

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).

Insertion Anomalies

- For example, to insert a new tuple for an employee who works in department number 5,
- we must enter the attribute values of department 5 correctly so that they are *consistent* with values for department 5 in other tuples in EMP_DEPT

Insertion 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 causes a problem because SSN is the primary key of EMP_DEPT, and each tuple is
- supposed to represent an employee entity- not a department entity

Deletion Anomalies

- If we delete from EMP_DEPT an employee tuple that happens to represent the last employee working for a particular department,
- The information concerning that department is lost from the database.

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

Modification Anomalies

- If we fail to update some tuples, the same department will be shown to have two different values for manager in different employee tuples,
- which would be wrong

Guideline 3

- Reducing the Null values in tuples
  - As far as possible, avoid placing attributes in a base relation whose values may frequently be null.
  - If nulls are unavoidable, make sure that they apply in exceptional cases only and do not apply to a majority of tuples in the relation.

Guideline 4

- Disallowing the generation of spurious tuples
- Design relation schemas so that they 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 are generated.

Guideline 4

- Avoid relations that contain matching attributes that are not (foreign key, primary key) combinations,
- because joining on such attributes may produce spurious tuples.
- This informal guideline is called the nonadditive (or lossless) join property, that guarantees that certain joins do not produce spurious tuples

nonadditive (or lossless) join property

- Design relations so that they can be JOINed with equality condition *"Equijoin"* on attributes that are either PKs or FKs

EMP_PROJ

| | | | | redundancy | redundancy |
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

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	Product X	Bellaire
123456789	2	7.5	Product Y	Sugarland
666884444	3	40.0	Product Z	Houston
453453453	1	20.0	Product X	Bellaire
453453453	2	20.0	Product Y	Sugarland
333445555	2	10.0	Product Y	Sugarland
333445555	3	10.0	Product Z	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

Guideline 4

- Suppose that we used EMP_PROJ1 and EMP_LaCS 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_LaCS.

Guideline 4

- If we attempt a NATURALJOIN operation on EMP_PROJ1 and EMP_LaCS,
- the result produces many more tuples than the original set of tuples in EMP_PROJ.
- Additional tuples that were not in EMP_PROJ are called spurious tuples because they represent spurious or *wrong* information that is not valid.
- The spurious tuples are marked by asterisks (*)

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.

Questions