Database Management Systems (DBMS)

Lec 16: Relational database design (Contd.)

Ramesh K. Jallu

IIIT Raichur

Date: 30/03/21

Recap

- General definitions of 2NF and 3NF
- Properties of normalization
  - Nonadditive join property
  - The dependency preservation
- Boyce-Codd normal form
  - Whenever a nontrivial functional dependency $\boldsymbol{X} \rightarrow \boldsymbol{A}$ holds in $\boldsymbol{R}$, then $\boldsymbol{X}$ is a superkey in $\boldsymbol{R}$

Today's plan

- Multivalued dependency
- The fourth normal form (4NF)
- The join dependency
- The fifth normal form (5NF)

1NF: Illustration with an example (recap)

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocation
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocation 1	Dlocaton 2	Dlocation 3
Research	5	333445555	Bellaire	Sugarland	Houston
Administration	4	987654321	Stafford	**NULL**	**NULL**
Headquarters	1	888665555	Houston	**NULL**	**NULL**

Multivalued dependency (MVD)

- Some relations have constraints that cannot be specified as functional dependencies and hence are not in violation of BCNF

EMP

Ename	Pname	Dname
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

Multivalued dependency (Contd.)

- The fourth normal form is based on MVD
- Multivalued dependencies are a consequence of 1NF
- If more than one multivalued attribute is present, the second way of normalizing the relation introduces a multivalued dependency
- MVD occurs when two attributes in a table are independent of each other but, both depend on a third attribute

Multivalued dependency (Contd.)

- Whenever two independent 1:N relationships between A, B, and A, C are mixed in the same relation, MVD occurs
- This MVD is denoted as $A \twoheadrightarrow B$ and $A \twoheadrightarrow$C (or $A \twoheadrightarrow B|C$)
- $A \twoheadrightarrow B$ is read as A multidetermines B
- In order to occur an MVD, a relation should have at least three attributes
- An MVD $A \twoheadrightarrow B$ in a relation R is said to be ***trivial*** if (i) B is a subset of A, or (ii) $A \cup B = R$; Otherwise, we call it ***non-trivial*** MVD

Multivalued dependency (Contd.)

- If we have a nontrivial MVD in a relation, we may have to repeat values redundantly in the tuples
- The values 'X' and 'Y' of Pname are repeated with each value of Dname
- Relations containing nontrivial MVDs tend to be ***all-key relations***
- The EMP schema is in BCNF because no functional dependencies hold in EMP
- Therefore, MVDs problematic and we need a NF stronger than BCNF

EMP

Ename	Pname	Dname
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

The fourth normalform (4NF)

- A relation schema ***R*** is in 4NF if,
  i. it satisfies BCNF, and
  ii. it doesnot have any non-trivial MVDs

- A relation that is not in 4NF due to a nontrivial MVD must be decomposed to convert it into a set of relations in 4NF

  - The decomposition removes the redundancy caused by the MVD

- During the decompostion each non-trivial MVD is represented by a separate relation where it becomes a trivial MVD

The fourth normal form (4NF): Formal definition

- A relation schema R is in 4NF with respect to a set of dependencies ***F*** (that includes functional dependencies and multivalued dependencies) if, for every nontrivial MVD ***X*** →→ ***Y*** in F^+, ***X*** is a superkey for ***R***
- An all-key relation is always in BCNF since it has no FDs
- The EMP relation is an all-key relation, which has no FDs but has the MVD Ename →→ Pname | Dname, is not in 4NF

Example

EMP

Ename	Pname	Dname
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

EMP_PROJECTS

Ename	Pname
Smith	X
Smith	Y

EMP_DEPENDENTS

Ename	Dname
Smith	John
Smith	Anna

Join dependencies

- So far, we have discussed the FDs which are problematic and tried to eliminate them by a process of repeated binary decomposition
  - This normalization process achieves 1NF, 2NF, 3NF, and BCNF
  - And the decomposition obeys the NJB property
- We also have seen the problematic MVDs and eliminating them by repeated binary decomposition

Join dependencies (Contd.)

- So far, we have seen problematic FDs and MVDs, and tried to eliminate by a process of repeated binary decomposition
- In some cases, there may be no NJ decomposition of ***R*** into *two* relation schemas, but there may be a NJ decomposition into more than two relation schemas
- Also, there may be no FDs in ***R*** that violates any normal form up to BCNF, and there may be no nontrivial MVD present in ***R*** either that violates 4NF

Join dependency (JD)

- Let ***R*** be a relation and ***r*** is a relation state. A join dependency specifies a constraint on the relations state ***r*** that every legal state ***r*** of ***R*** should have a nonadditive join decomposition into $R_1, R_2, \ldots, R_n$
  - $*(\pi_{R1}(r), \pi_{R2}(r), \ldots, \pi_{Rn}(r)) = r$

- Observe that MVD is a special case of JD where n = 2
  - JD(R_1, R_2) implies an MVD $(R_1 \cap R_2) \rightarrow\rightarrow (R_1 - R_2)$ (or, by symmetry, $(R_1 \cap R_2) \rightarrow\rightarrow (R_2 - R_1)$

- A join dependency JD($R_1, R_2, \ldots, R_n$) is a ***trivial*** JD if one of the relation schemas R_i in JD($R_1, R_2, \ldots, R_n$) is equal to ***R***

The fifth normal form (5NF)

- The 5NF is based on another dependency called ***multiway decomposition***
- A relation schema ***R*** is in 5NF, w.r.t a set ***F*** of FD, MVD, and JDs, if for every nontrivial join dependency JD($R_1, R_2, \ldots, R_n$) in F^+, every R_i is a superkey of ***R***
- The 5NF is also known as ***project-join normal form*** (PJNF)

Example

SUPPLY

Sname	Part_name	Proj_name
Smith	Bolt	ProjX
Smith	Nut	ProjY
Adamsky	Bolt	ProjY
Walton	Nut	ProjZ
Adamsky	Nail	ProjX
Adamsky	Bolt	ProjX
Smith	Bolt	ProjY

Whenever a supplier ***s*** supplies part ***p***, *and* a project ***j*** uses part ***p***, *and* the supplier ***s*** supplies at least one part to project ***j***, *then* supplier ***s*** will also be supplying part ***p*** to project ***j***

R_1

Sname	Part_name
Smith	Bolt
Smith	Nut
Adamsky	Bolt
Walton	Nut
Adamsky	Nail

R_2

Sname	Proj_name
Smith	ProjX
Smith	ProjY
Adamsky	ProjY
Walton	ProjZ
Adamsky	ProjX

R_3

Part_name	Proj_name
Bolt	ProjX
Nut	ProjY
Bolt	ProjY
Nut	ProjZ
Nail	ProjX

Thank you!