

Database Management System – 31

Database design – Normal Forms (1NF, 2NF, 3NF)

Ajay James
Asst. Prof in CSE
Government Engineering College Thrissur

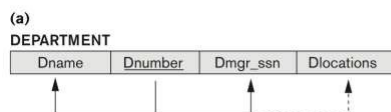
Outline

- First Normal Form
- Second Normal Form
 - Full Functional Dependency
- Third Normal Form
 - Transitive Dependency

First Normal Form

- Considered to be part of the definition of a relation
- Disallows
 - composite attributes
 - Multi-valued attributes
 - nested relations; attributes whose values for an individual tuple are non-atomic
- Most RDBMSs allow only those relations to be defined that are in First Normal Form

Normalization into 1NF



(b)

DEPARTMENT

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

(c)

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

Normalization into 1NF

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

Normalization of Nested relation

- EMP_PROJ(Ssn, Ename, {PROJS(Pnumber, Hours)})

(a)

EMP_PROJ		Projs	
Ssn	Ename	Pnumber	Hours

(b)

EMP_PROJ

Ssn	Ename	Pnumber	Hours
123456789	Smith, John B.	1	32.5
		2	7.5
666884444	Narayan, Ramesh K.	3	40.0
453453453	English, Joyce A.	1	20.0
		2	20.0
333445555	Wong, Franklin T.	2	10.0
		3	10.0
		10	10.0
		20	10.0
999887777	Zelaya, Alicia J.	30	30.0
		10	10.0
987987987	Jabbar, Ahmad V.	10	35.0
		30	5.0
987654321	Wallace, Jennifer S.	30	20.0
		20	15.0
888665555	Borg, James E.	20	NULL

(c)

EMP_PROJ1

Ssn	Ename
-----	-------

EMP_PROJ2

Ssn	Pnumber	Hours
-----	---------	-------

Nested relation example

- CANDIDATE (Ssn, Name, {JOB_HIST (Company, Highest_position, {SAL_HIST (Year, Max_sal)}}))
- CANDIDATE_1 (Ssn, Name)
- CANDIDATE_JOB_HIST (Ssn, Company, Highest_position)
- CANDIDATE_SAL_HIST (Ssn, Company, Year, Max-sal)

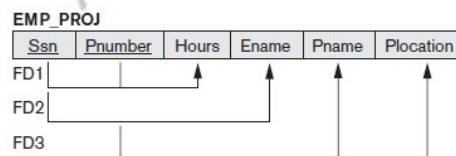
More than one Multivalued attribute example

- PERSON (Ss#, {Car_lic#}, {Phone#})
- PERSON_IN_1NF (Ss#, Car_lic#, Phone#)
- P1(Ss#, Car_lic#) and P2(Ss#, Phone#)

Second Normal Form

- 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 $X \rightarrow Y$ where removal of any attribute from X means the FD does not hold any more
- FULL FD - For any attribute $A \in X$, $(X - \{A\})$ does not functionally determine Y
- A functional dependency $X \rightarrow Y$ is a **partial dependency** if some attribute $A \in X$ can be removed from X and the dependency still holds
- For some $A \in X$, $(X - \{A\}) \rightarrow Y$

Partial dependency example

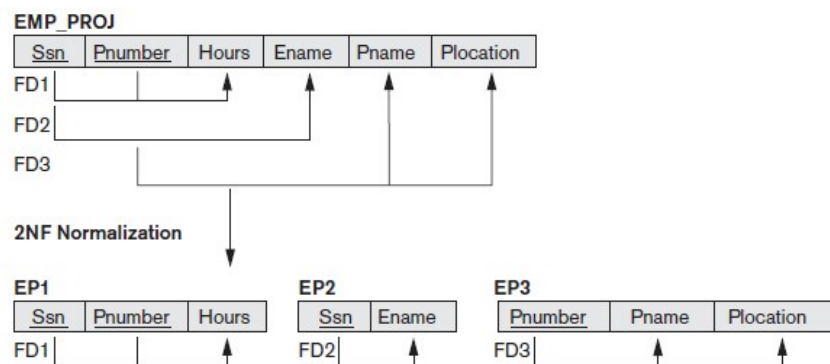


- $\{SSN, PNUMBER\} \rightarrow HOURS$ is a full FD since neither $SSN \rightarrow HOURS$ nor $PNUMBER \rightarrow HOURS$ hold
- $\{SSN, PNUMBER\} \rightarrow ENAME$ is not a full FD (it is called a partial dependency) since $SSN \rightarrow ENAME$ also holds

Second Normal form

- 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 or “second normalization”

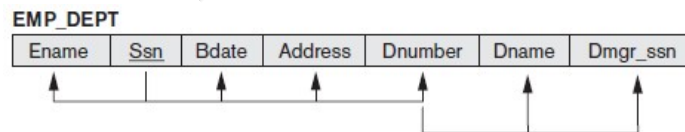
Normalization into 2NF



Third Normal Form

- Definition:
 - **Transitive functional dependency**: a FD $X \rightarrow Z$ that can be derived from two FDs $X \rightarrow Y$ and $Y \rightarrow Z$
 - Y is not a candidate key or not a subset of any key of R

Transitive dependency example

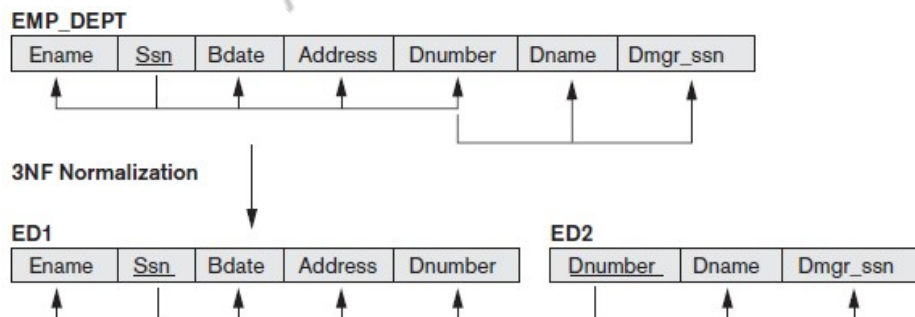


- Examples:
 - $SSN \rightarrow DMGRSSN$ is a **transitive FD**
 - Since $SSN \rightarrow DNUMBER$ and $DNUMBER \rightarrow DMGRSSN$ hold
 - $SSN \rightarrow ENAME$ is **non-transitive**
 - Since there is no set of attributes X where $SSN \rightarrow X$ and $X \rightarrow ENAME$

Third Normal Form

- 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 $X \rightarrow Y$ and $Y \rightarrow Z$, 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, $SSN \rightarrow Emp\# \rightarrow Salary$ and Emp# is a candidate key.

3NF normalization example



Summary

Table 14.1 Summary of Normal Forms Based on Primary Keys and Corresponding Normalization

Normal Form	Test	Remedy (Normalization)
First (1NF)	Relation should have no multivalued attributes or nested relations.	Form new relations for each multivalued attribute or nested relation.
Second (2NF)	For relations where primary key contains multiple attributes, no nonkey attribute should be functionally dependent on a part of the primary key.	Decompose and set up a new relation for each partial key with its dependent attribute(s). Make sure to keep a relation with the original primary key and any attributes that are fully functionally dependent on it.
Third (3NF)	Relation should not have a nonkey attribute functionally determined by another nonkey attribute (or by a set of nonkey attributes). That is, there should be no transitive dependency of a nonkey attribute on the primary key.	Decompose and set up a relation that includes the nonkey attribute(s) that functionally determine(s) other nonkey attribute(s).

Reference

- Elmasri R. and S. Navathe, Database Systems: Models, Languages, Design and Application Programming, Pearson Education 6th edition and 7th edition

Thank you

AJ-GEC THRISSUR