4. Normal Forms

Stéphane Bressan

January 27, 2015

This lecture is based on material by Professor Ling Tok Wang.

CS 4221: Database Design

The Relational Model

Ling Tok Wang
National University of Singapore

1

CS4221: The Relational Model

https://www.comp.nus.edu.sg/~lingtw/cs4221/rm.pdf

Content

Readings

- Kent, William, "A Simple Guide to Five Normal Forms in Relational Database Theory", Communications of the ACM 26 (2), Feb. 1983, pp. 120-125.
- Zaniolo, Carlo, "A New Normal Form for the Design of Relational Database Schemata." ACM Transactions on Database Systems 7(3), September 1982.
- Maier, David, "The Theory of Relational Databases", `http://web.cecs.pdx.edu/~maier/TheoryBook/TRD.html`, (1983).

Second Normal Form

TAKE						
STUDENT#	DEPARTMENT	FACULTY	COURSE#	SNAME	CDESC	MARK
95001	CS	SoC	CS1101	Tan CK	Programming	75
95023	CEG	Eng	CS1101	Lee SL	Programming	58
95023	CEG	Eng	CS2103	Tan CK	D.S. and Alg.	64
. . .						

Example

TAKE = {STUDENT#, DEPARTMENT, FACULTY, COURSE#, SNAME, CDESC, MARK}.

$\Sigma = \{\{STUDENT\#\} \rightarrow \{SNAME, DEPARTMENT\},$
$\{DEPARTMENT\} \rightarrow \{FACULTY\},$
$\{COURSE\#\} \rightarrow \{CDESC\},$
$\{STUDENT\#, COURSE\#\} \rightarrow \{MARK\}\}$

The candidate key is {STUDENT#, COURSE#}

Problem

SNAME, DEPARTMENT, FACULTY, CDESC and MARK do not fully depend on the key! This creates anomalies.

Reminder

We say that Y is fully dependent on X if and only if there exists a non-trivial functional dependency $X \rightarrow Y$ such that no proper subset X' of X is such such that $X' \rightarrow Y \in \Sigma^+$.

SNAME is not fully dependent on the key $\{\text{STUDENT\#}, \text{COURSE\#}\}$.

$$\{\text{STUDENT\#}\} \rightarrow \{\text{SNAME}\}$$

First Idea

Let us make sure that every attribute fully depends on the primary key.

But prime attributes do not fully depend on the key (if there is only one)!

Example

STUDENT# is not fully dependent on the key {STUDENT#, COURSE#}.

$$\{\text{STUDENT\#}, \text{COURSE\#}\} \to \{\text{STUDENT\#}\}$$

is trivial.

First Idea (refined)

Let us make sure that every non-prime attribute fully depends on the primary key.

But there could be more than one candidate key!

Example

$R = \{A, B, C, D, E\}$.
$\Sigma = \{\{A, B\} \to \{C, D\}, \{C, D\} \to \{A, B\}, \{C\} \to \{E\}\}$
The candidate keys are $\{A, B\}$ and $\{C, D\}$.
E fully depends on $\{A, B\}$.
Yet E does not fully depend on $\{C, D\}$. This can create anomalies.

First Idea (further refined)

Let us make sure that every non-prime attribute fully depends on each candidate key.

"A nonkey field must provide a fact about the key, the whole key [...]", W. Kent in "A Simple Guide to Five Normal Forms in Relational Database Theory", Communication of the ACM, Volume 26, Number 2 (1983).

Definition

A relation R with a set of functional dependencies Σ is in Second Normal Form, or 2NF for short, if and and only if every non-prime attribute is fully dependent on each candidate key.

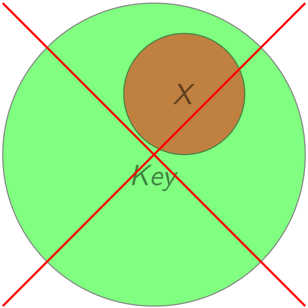
Theorem

A relation R with a set of functional dependencies Σ is in Second Normal Form if and only if for every functional dependency $X \rightarrow \{A\} \in \Sigma^+$:

- *$X \rightarrow \{A\}$ is trivial or*
- *X is not a proper subset of a candidate key or*
- *A is a prime attribute.*

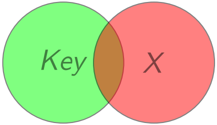
It is sufficient to look at Σ.

This situation where X is a proper subset of a candidate key is forbidden:

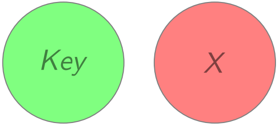


X cannot be proper subset of a candidate key.
A must be fully dependent on each candidate key.

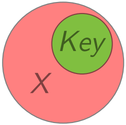
For all candidate keys, we must have one of the following:



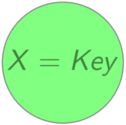
X instersect with the candidate key.



X and the candidate key are disjoint.



X is a superset of the candidate key (X is a superkey).



X is a candidate key (X is a superkey).

STUDENT			
STUDENT#	DEPARTMENT	FACULTY	SNAME
95001	CS	SoC	Tan CK
95023	CEG	Eng	Lee SL
...			

COURSE	
COURSE#	CDESC
CS1101	Programming
CS2103	D.S. and Alg.
...	

TAKE		
STUDENT#	COURSE#	MARK
95001	CS1101	75
95023	CS1101	58
95023	CS2103	64
...		

Verify that all relations are in 2NF.
What are the (projected) functional dependencies? What are the candidate keys? Is it prone to anomalies?

Example

A supplier with supplier number (S#) and name (SNAME) supplies a part with part number (P#) and name (PNAME) with a price (PRICE).

$$\mathsf{SP} = \{\mathsf{S\#}, \mathsf{SNAME}, \mathsf{P\#}, \mathsf{PNAME}, \mathsf{PRICE}\}$$

$$\Sigma = \{\{\mathsf{S\#}\} \rightarrow \{\mathsf{SNAME}\},$$

$$\{\mathsf{P\#}\} \rightarrow \{\mathsf{PNAME}\},$$

$$\{\mathsf{S\#}, \mathsf{P\#}\} \rightarrow \{\mathsf{PRICE}\}\}$$

Question

Is SP with Σ in 2NF?

The only candidate key is $\{S\#, P\#\}$. How to prove it?

Compute all the attribute set closures or observe that PRICE cannot be prime as it appears in the right-hand-side of a functional dependency and does not appear in the left-hand-side of a functional dependency.

Question

Is SP with Σ in 2NF?

One way: SNAME is not fully dependent on the candidate key.

$$\{S\#\} \rightarrow \{SNAME\}$$

There is redudant information about SNAME and about PNAME in SP.

Or another: $\{S\#\} \rightarrow \{SNAME\}$ is neither trivial, nor is SNAME a prime attribute, and $\{S\#\}$ is a proper subset of a candidate key ($\{S\#\} \subset \{S\#, P\#\}$ and $\{S\#\} \neq \{S\#, P\#\}$).

Answer

SP with Σ in not in 2NF.

Third Normal Form

STUDENT			
STUDENT#	DEPARTMENT	FACULTY	SNAME
95001	CS	SoC	Tan CK
95011	CS	SoC	Wee LK
95023	CEG	Eng	Lee SL
...			

Example

$R = \{\text{STUDENT\#}, \text{FACULTY}, \text{COURSE\#}, \text{SNAME}\}$.
$\Sigma = \{\{\text{STUDENT\#}\} \to \{\text{SNAME}, \text{DEPARTMENT}\}$,
$\{\text{DEPARTMENT}\} \to \{\text{FACULTY}\}\}$
The candidate key is $\{\text{STUDENT\#}\}$.
The relation is in 2NF.

Problem

FACULTY is transitively dependent on the key! This creates anomalies.

Reminder

We say that a non-prime attribute A is transitively dependent on a candidate key if and only if there exists a set of attributes S such that S is not a superkey and $S \rightarrow \{A\}$ holds and is a non-trivial functional dependency.

FACULTY is transitively dependent on the key.
$\{\text{STUDENT\#}\} \rightarrow \{\text{DEPARTMENT}\} \rightarrow \{\text{FACULTY}\}$

Second Idea

Let us make sure that every non-prime attribute is not transitively dependent on any candidate key.

Second Idea

"A nonkey field must provide a fact about the key, the whole key, and nothing but the key. [So help me Codd.]",
W. Kent in "A Simple Guide to Five Normal Forms in Relational Database Theory", Communication of the ACM, Volume 26, Number 2 (1983).

Definition

A relation R with a set of functional dependencies Σ is in (Codd) Third Normal Form, or 3NF for short, if and and only if it is in Second Normal Form and no non-prime attribute is transitively dependent on some candidate key.

The test requires to check every non-prime attribute with every candidate key.

Theorem

If a non-prime attribute is not transitively dependent on a given candidate key, then it fully depends on a candidate key.

Theorem

A relation R with a set of functional dependencies Σ is in Third Normal Form if and and only if it is in 2NF and every non-prime attribute is not transitively dependent on any given candidate key.

The test now only requires to check every non-prime attribute with one candidate key.

The test requires to check 2NF.

Theorem

If a non-prime attribute is not transitively dependent on any given candidate key, then it fully depends on a candidate key.

Theorem

A relation R with a set of functional dependencies Σ is in Third Normal Form if and and only if every non-prime attribute is not transitively dependent on any given candidate key.

The test now does not requires to check 2NF.

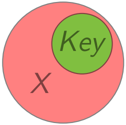
Theorem

A relation R with a set of functional dependencies Σ is in Third Normal Form, or 3NF for short, if and only if for every functional dependency $X \rightarrow \{A\} \in \Sigma^+$:

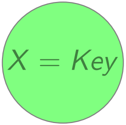
- *$X \rightarrow \{A\}$ is trivial or*
- *A is a prime attribute or*
- *X is a superkey.*

It is sufficient to look at Σ.

For some candidate key, we must have one of the following:



X is a superset of the candidate key (X is a superkey).



X is the candidate key (X is a superkey).

STUDENT		
STUDENT#	DEPARTMENT	SNAME
95001	CS	Tan CK
95011	CS	Wee LK
95023	CEG	Lee SL
...		

DEPARTMENT	
DEPARTMENT	FACULTY
CS	SoC
CS	SoC
CEG	Eng
...	

Verify that all relations are in 3NF.
What are the (projected) functional dependencies? What are the candidate keys? Is it prone to anomalies?

Example

A supplier with supplier number (S#) and name (SNAME) supplies a part with part number (P#) and name (PNAME) with a price (PRICE).

$$\text{SP} = \{\text{S\#}, \text{SNAME}, \text{P\#}, \text{PNAME}, \text{PRICE}\}$$

$$\Sigma = \{\{\text{S\#}\} \rightarrow \{\text{SNAME}\},$$

$$\{\text{P\#}\} \rightarrow \{\text{PNAME}\},$$

$$\{\text{S\#}, \text{P\#}\} \rightarrow \{\text{PRICE}\}\}$$

Question

Is SP with Σ in 3NF?

The only candidate key is $\{S\#, P\#\}$.

One way: SNAME is transitively dependent on the candidate key.

$$\{S\#, P\#\} \rightarrow \{S\#\} \quad \{S\#\} \rightarrow \{SNAME\}$$

Or another: $\{S\#\} \rightarrow \{SNAME\}$ is neither trivial, nor is SNAME a prime attribute, nor is $\{S\#\}$ a superkey.

Answer

SP with Σ in not in 3NF.

Elementary Key Normal Form

STAFF		
DEPARTMENT	HEAD	PROFESSOR
Tan Kian Lee	CS	Lee Mong Li
Zhu Chengbo	Math	Frank Stefan
Tan Kian Lee	CS	Frank Stefan
Zhu Chengbo	Math	Bao Weizhu
...		

Example

$STAFF = \{\text{DEPARTMENT}, \text{HEAD}, \text{PROFESSOR}\}$.
$\Sigma = \{\{\text{DEPARTMENT}\} \to \{\text{HEAD}\}, \{\text{HEAD}\} \to \{\text{DEPARTMENT}\}\}$
The candidate keys are $\{\text{PROFESSOR}, \text{HEAD}\}$ and $\{\text{PROFESSOR}, \text{DEPARTMENT}\}$.
The relation is in 3NF. Every attribute is prime!

Problem

$\{\text{HEAD}\} \to \{\text{DEPARTMENT}\}$ and $\{\text{DEPARTMENT}\} \to \{\text{HEAD}\}$ cannot be enforced in most SQL dialects. This leads to anomalies.

Definition

An elementary functional dependency is a full dependency.

Definition

A candidate key K is an elementary candidate key if and only if there exists an attribute A such that $K \to \{A\}$ is an elementary functional dependency.

Definition

An elementary prime attribute is an attribute of some elementary candidate key.

The two candidate keys, $\{\text{PROFESSOR}, \text{HEAD}\}$ and $\{\text{PROFESSOR}, \text{DEPARTMENT}\}$, are not elementary: there is no attribute that fully depends on them.

Third Idea

Let us make sure that some candidate keys and all transitively dependent prime attributes are elementary.

Definition

A relation R with a set of functional dependencies Σ is in Elementary Key Normal Form, or EKNF for short, if and only if for every functional dependency $X \rightarrow \{A\} \in \Sigma^+$:

- $X \rightarrow \{A\}$ is not elementary or
- A is an elementary prime attribute or
- X is an elementary candidate key.

It is sufficient to look at Σ.

Theorem

A relation R with a set of functional dependencies Σ is in Elementary Key Normal Form, or EKNF for short, if and only if for every functional dependency $X \to \{A\} \in \Sigma^+$:

- *$X \to \{A\}$ is trivial or*
- *A is an elementary prime attribute or*
- *X is a superkey.*

It is sufficient to look at Σ.

Example

$STAFF = \{\text{DEPARTMENT}, \text{HEAD}, \text{PROFESSOR}\}$.
$\Sigma = \{\{\text{DEPARTMENT}\} \rightarrow \{\text{HEAD}\}, \{\text{HEAD}\} \rightarrow \{\text{DEPARTMENT}\}\}$
The candidate keys are $\{\text{PROFESSOR}, \text{HEAD}\}$ and $\{\text{PROFESSOR}, \text{DEPARTMENT}\}$.
No candidate key is elementary.

The relation is not in EKNF.

STAFF	
DEPARTMENT	PROFESSOR
CS	Lee Mong Li
Math	Frank Stefan
CS	Frank Stefan
Math	Bao Weizhu
...	

MANAGEMENT	
DEPARTMENT	HEAD
Tan Kian Lee	CS
Zhu Chengbo	Math
...	

Verify that all relations are in EKNF.
What are the (projected) functional dependencies? What are the candidate keys? Is it prone to anomalies?

Boyce-Codd Normal Form

STAFF		
DEPARTMENT	HEAD	PROFESSOR
Tan Kian Lee	CS	Lee Mong Li
Zhu Chengbo	Math	Frank Stefan
Tan Kian Lee	CS	Frank Stefan
Zhu Chengbo	Math	Bao Weizhu
...		

Example

$STAFF = \{\text{DEPARTMENT}, \text{HEAD}, \text{PROFESSOR}\}$.
$\Sigma = \{\{\text{DEPARTMENT}\} \rightarrow \{\text{HEAD}\}, \{\text{HEAD}\} \rightarrow \{\text{DEPARTMENT}\}\}$
The candidate keys are $\{\text{PROFESSOR}, \text{HEAD}\}$ and $\{\text{PROFESSOR}, \text{DEPARTMENT}\}$.
The relation is in 3NF. Every attribute is prime!

Problem

$\{\text{HEAD}\} \rightarrow \{\text{DEPARTMENT}\}$ and $\{\text{DEPARTMENT}\} \rightarrow \{\text{HEAD}\}$ cannot be enforced in most SQL dialects. This leads to anomalies.

Fourth Idea

Why do we focus on prime and elementary attributes?
If something non-trivially depends on something else, then it should be on

“a key, a whole key, and nothing but a key”

Definition

A relation R with a set of functional dependencies Σ is in Boyce-Codd Normal Form, or BCNF for short, if and only if for every attribute set $S \subset R$, if any attribute of R not in S is functionally dependent on S, then all attributes in R are functionally dependent on S.

Theorem

A relation R with a set of functional dependencies Σ is in BCNF if and only if no attribute is transitively dependent on any key. [David Maier]

Theorem

A relation R with a set of functional dependencies Σ is in BCNF if and only if for every functional dependency $X \to \{A\} \in \Sigma^+$:

- *$X \to \{A\}$ is trivial or*
- *X is a superkey.*

It is sufficient to look at Σ.

STAFF	
DEPARTMENT	PROFESSOR
CS	Lee Mong Li
Math	Frank Stefan
CS	Frank Stefan
Math	Bao Weizhu
...	

MANAGEMENT	
DEPARTMENT	HEAD
Tan Kian Lee	CS
Zhu Chengbo	Math
...	

Verify that all relations are in BCNF.
What are the (projected) functional dependencies? What are the candidate keys? Is it prone to anomalies?

The prototypical example of a relation in 3NF (and EKNF) and not in BCNF is:

$$R(A, B, C)$$

with

$$\{A, B\} \rightarrow \{C\}$$

and

$$\{C\} \rightarrow \{B\}$$

The candidate keys are $\{A, B\}$ and $\{A, C\}$.
The only elementary candidate keys is $\{A, B\}$
(Why isn't $\{A, C\}$ elementary?).
$\{C\} \rightarrow \{B\}$ is non trivial, $\{C\}$ is not a candidate key but B is an elementary prime attribute.

DIRECTORY		
PROFESSOR	UNIVERSITY	TELEPHONE
Ling Tok wang	NUS	(65) 6516-2734
Lee Mong Li	NUS	(65) 6516 2905
Gillian Dobbie	U. Auckland	(64 9) 373-7599 83949
Lee Mong Li	U. Auckland	(64 9) 373-7599 83949
...		

Example

$R = \{\text{PROFESSOR}, \text{UNIVERSITY}, \text{TELEPHONE}\}$.
$\Sigma = \{\{\text{PROFESSOR}, \text{UNIVERSITY}\} \rightarrow \{\text{TELEPHONE}\}, \{\text{TELEPHONE}\} \rightarrow \{\text{UNIVERSITY}\}\}$
The candidate keys are $\{\text{PROFESSOR}, \text{UNIVERSITY}\}$ and $\{\text{PROFESSOR}, \text{TELEPHONE}\}$.
The relation is in EKNF but not in BCNF.

Problem

$\{\text{TELEPHONE}\} \rightarrow \{\text{UNIVERSITY}\}$ cannot be enforced in most SQL dialects).

DIRECTORY		
PROFESSOR	UNIVERSITY	TELEPHONE
Ling Tok wang	NUS	(65) 6516-2734
Lee Mong Li	NUS	(65) 6516 2905
Gillian Dobbie	U. Auckland	(64 9) 373-7599 83949
Lee Mong Li	U. Auckland	(64 9) 373-7599 83949
. . .		

SUBDIRECTORY	
UNIVERSITY	TELEPHONE
NUS	(65) 6516-2734
NUS	(65) 6516 2905
U. Auckland	(64 9) 373-7599 83949
. . .	

Verify that all relations are in EKNF but still not in BCNF.
What are the (projected) functional dependencies? What are the candidate keys? Is it prone to anomalies?

Example

$DIRECTORY = \{\text{PROFESSOR}, \text{UNIVERSITY}, \text{TELEPHONE}\}$.
$\Sigma_1 = \{\{\text{PROFESSOR}, \text{UNIVERSITY}\} \to \{\text{TELEPHONE}\}, \{\text{TELEPHONE}\} \to \{\text{UNIVERSITY}\}\}$
The candidate keys are $\{\text{PROFESSOR}, \text{UNIVERSITY}\}$ and $\{\text{PROFESSOR}, \text{TELEPHONE}\}$.
$SUBDIRECTORY = \{\text{UNIVERSITY}, \text{TELEPHONE}\}$.
$\Sigma_2 = \{\{\text{TELEPHONE}\} \to \{\text{UNIVERSITY}\}\}$
The candidate keys are $\{\text{PROFESSOR}, \text{UNIVERSITY}\}$ and $\{\text{PROFESSOR}, \text{TELEPHONE}\}$.

In addition, one needs a referential constraint from $\{\text{DIRECTORY.UNIVERSITY}, \text{DIRECTORY.TELEPHONE}\}$ to $\{\text{SUBDIRECTORY.UNIVERSITY}, \text{SUBDIRECTORY.TELEPHONE}\}$

One possible way is to declare a primary key which is a superkey.

```
CREATE TABLE DIRECTORY
UNIVERSITY ... ,
TELEPHONE ... UNIQUE NOT NULL,
PRIMARY KEY (UNIVERSITY, TELEPHONE))
```

And to declare the foreign key.

```
DIRECTORY(UNIVERSITY, TELEPHONE) REFERENCES
              SUBDIRECTORY(UNIVERSITY, TELEPHONE)
```

DIRECTORY	
PROFESSOR	TELEPHONE
Ling Tok wang	(65) 6516-2734
Lee Mong Li	(65) 6516 2905
Gillian Dobbie	(64 9) 373-7599 83949
Lee Mong Li	(64 9) 373-7599 83949
...	

SUBDIRECTORY	
UNIVERSITY	TELEPHONE
NUS	(65) 6516-2734
NUS	(65) 6516 2905
U. Auckland	(64 9) 373-7599 83949
...	

Verify that all relations are in BCNF.
What are the (projected) functional dependencies? What are the candidate keys? Is it prone to anomalies?

Example

DIRECTORY = {PROFESSOR, TELEPHONE}.
$\Sigma_1 = \emptyset$
The candidate key is {PROFESSOR, TELEPHONE}.
SUBDIRECTORY = {UNIVERSITY, TELEPHONE}.
$\Sigma_2 = \{\{\text{TELEPHONE}\} \rightarrow \{\text{UNIVERSITY}\}\}$
The candidate key is {TELEPHONE}.

But we have lost one functional dependency.
$\{\text{PROFESSOR}, \text{UNIVERSITY}\} \rightarrow \{\text{TELEPHONE}\}$

In some cases, there is no BCNF lossless dependency preserving decomposition.

Theorem

$$1NF \subset 2NF \subset 3NF \subset EKNF \subset BCNF$$

Theorem

$$1NF \neq 2NF \neq 3NF \neq EKNF \neq BCNF$$

We have not discussed inter-relational dependencies.
See `https://www.comp.nus.edu.sg/~lingtw/ltk.pdf` or read T.-W. Ling, F.W. Tompa, and T. Kameda, "An Improved Third Normal Form for Relational Databases", ACM Transactions on Database Systems, 6(2), June 1981, 329-346.