CSCI-620

Normalization

Roadmap

1. Introduction to normalization
2. Keys and functional dependencies
3. Normal forms and decomposition
4. Functional dependency discovery
5. Denormalization

Normalization

1	id	name	description	date	email	fn	ln	dob	st	city	zipcode	state
2	1	Ender's Game	The Sci-Fi book by Scott Card	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
3	2	American Gods	The Sci-Fi book by Gaiman	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
4	2	American Gods	The Sci-Fi book by Gaiman	2.0171E+11	u1@rit.edu	Sandra	Smith	19780905	134 West Henrietta Rd	Henrietta	14633	NY
5	3	I, Robot	The Sci-Fi book by Asimov	2.0171E+11	u1@rit.edu	Sandra	Smith	19780905	134 West Henrietta Rd	Henrietta	14633	NY
6	1	Ender's Game	The Sci-Fi book by Scott Card	2.01709E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
7	2	American Gods	The Sci-Fi book by Gaiman	2.01709E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
8	2	American Gods	The Sci-Fi book by Gaiman	2.01707E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
9	3	I, Robot	The Sci-Fi book by Asimov	2.01609E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
10	1	Ender's Game	The Sci-Fi book by Scott Card	2.01705E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
11	3	I, Robot	The Sci-Fi book by Asimov	2.01705E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
12	2	American Gods	The Sci-Fi book by Gaiman	2.01609E+11	u4@rit.edu	Sandra	Smith	19720225	156 Monroe Ave	Brighton	14512	NY
13	4	The Castle	The Sci-Fi book by Kafka	2.01709E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
14	4	The Castle	The Sci-Fi book by Kafka	2.01709E+11	u4@rit.edu	Sandra	Smith	19720225	156 Monroe Ave	Brighton	14512	NY
15	2	American Gods	The Sci-Fi book by Gaiman	2.01402E+11	u1@rit.edu	Sandra	Smith	19780905	189 Monroe Ave	Brighton	14512	NY
16	4	The Castle	The Sci-Fi book by Kafka	2.0151E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
17	5	I, Robot	The Sci-Fi book by Asimov	2.01609E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
18	6	The Graveyard Book	The Sci-Fi book by Gaiman	2.01707E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
19	1	Ender's Game	The Sci-Fi book by Scott Card	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
20	1	Ender's Game	The Sci-Fi book by Scott Card	2.01707E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY

Starting point

Reasons for Normalizing

- Designs *may* be more efficient
- Reduces the amount of redundant data stored (usually)
- Avoids *anomalies* when updating, inserting, or deleting data

1	id	name	description	date	email	fn	ln	dob	st	city	zipcode	state
2	1	Ender's Game	The Sci-Fi book by Scott Card	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
3	2	American Gods	The Sci-Fi book by Gaiman	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
4	2	American Gods	The Sci-Fi book by Gaiman	2.0171E+11	u1@rit.edu	Sandra	Smith	19780905	134 West Henrietta Rd	Henrietta	14633	NY
5	3	I, Robot	The Sci-Fi book by Asimov	2.0171E+11	u1@rit.edu	Sandra	Smith	19780905	134 West Henrietta Rd	Henrietta	14633	NY
6	1	Ender's Game	The Sci-Fi book by Scott Card	2.01709E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
7	2	American Gods	The Sci-Fi book by Gaiman	2.01709E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
8	2	American Gods	The Sci-Fi book by Gaiman	2.01707E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
9	3	I, Robot	The Sci-Fi book by Asimov	2.01609E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
10	1	Ender's Game	The Sci-Fi book by Scott Card	2.01705E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
11	3	I, Robot	The Sci-Fi book by Asimov	2.01705E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
12	2	American Gods	The Sci-Fi book by Gaiman	2.01609E+11	u4@rit.edu	Sandra	Smith	19720225	156 Monroe Ave	Brighton	14512	NY
13	4	The Castle	The Sci-Fi book by Kafka	2.01709E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
14	4	The Castle	The Sci-Fi book by Kafka	2.01709E+11	u4@rit.edu	Sandra	Smith	19720225	156 Monroe Ave	Brighton	14512	NY
15	2	American Gods	The Sci-Fi book by Gaiman	2.01402E+11	u1@rit.edu	Sandra	Smith	19780905	189 Monroe Ave	Brighton	14512	NY
16	4	The Castle	The Sci-Fi book by Kafka	2.0151E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
17	5	I, Robot	The Sci-Fi book by Asimov	2.01609E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
18	6	The Graveyard Book	The Sci-Fi book by Gaiman	2.01707E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
19	1	Ender's Game	The Sci-Fi book by Scott Card	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
20	1	Ender's Game	The Sci-Fi book by Scott Card	2.01707E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY

Anomalies

1	id	name	description	date	email	fn	ln	dob	st	city	zipcode	state
2	1	Ender's Game	The Sci-Fi book by Scott Card	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
3	2	American Gods	The Sci-Fi book by Gaiman	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
4	2	American Gods	The Sci-Fi book by Gaiman	2.0171E+11	u1@rit.edu	Sandra	Smith	19780905	134 West Henrietta Rd	Henrietta	14633	NY
5	3	I, Robot	The Sci-Fi book by Asimov	2.0171E+11	u1@rit.edu	Sandra	Smith	19780905	134 West Henrietta Rd	Henrietta	14633	NY
6	1	Ender's Game	The Sci-Fi book by Scott Card	2.01709E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
7	2	American Gods	The Sci-Fi book by Gaiman	2.01709E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
8	2	American Gods	The Sci-Fi book by Gaiman	2.01707E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
9	3	I, Robot	The Sci-Fi book by Asimov	2.01609E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
10	1	Ender's Game	The Sci-Fi book by Scott Card	2.01705E+11	u6@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
11	3	I, Robot	The Sci-Fi book by Asimov	2.01705E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
12	2	American Gods	The Sci-Fi book by Gaiman	2.01609E+11	u4@rit.edu	Sandra	Smith	19720225	156 Monroe Ave	Brighton	14512	NY
13	4	The Castle	The Sci-Fi book by Kafka	2.01709E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
14	4	The Castle	The Sci-Fi book by Kafka	2.01709E+11	u4@rit.edu	Sandra	Smith	19720225	156 Monroe Ave	Brighton	14512	NY
15	2	American Gods	The Sci-Fi book by Gaiman	2.01402E+11	u1@rit.edu	Sandra	Smith	19780905	189 Monroe Ave	Brighton	14512	NY
16	4	The Castle	The Sci-Fi book by Kafka	2.0151E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
17	5	I, Robot	The Sci-Fi book by Asimov	2.01609E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
18	6	The Graveyard Book	The Sci-Fi book by Gaiman	2.01707E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
19	1	Ender's Game	The Sci-Fi book by Scott Card	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
20	1	Ender's Game	The Sci-Fi book by Scott Card	2.01707E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY

Update Anomaly

1	id	name	description	date	email	fn	ln	dob	st	city	zipcode	state
2	1	Ender's Game	The Sci-Fi book by Scott Card	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
3	2	American Gods	The Sci-Fi book by Gaiman	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
4	2	American Gods	The Sci-Fi book by Gaiman	2.0171E+11	u1@rit.edu	Sandra	Smith	19780905	134 West Henrietta Rd	Henrietta	14633	NY
5	3	I, Robot	The Sci-Fi book by Asimov	2.0171E+11	u1@rit.edu	Sandra	Smith	19780905	134 West Henrietta Rd	Henrietta	14633	NY
6	1	Ender's Game	The Sci-Fi book by Scott Card	2.01709E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
7	2	American Gods	The Sci-Fi book by Gaiman	2.01709E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
8	2	American Gods	The Sci-Fi book by Gaiman	2.01707E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
9	3	I, Robot	The Sci-Fi book by Asimov	2.01609E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
10	1	Ender's Game	The Sci-Fi book by Scott Card	2.01705E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
11	3	I, Robot	The Sci-Fi book by Asimov	2.01705E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
12	2	American Gods	The Sci-Fi book by Gaiman	2.01609E+11	u4@rit.edu	Sandra	Smith	19720225	156 Monroe Ave	Brighton	14512	NY
13	4	The Castle	The Sci-Fi book by Kafka	2.01709E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
14	4	The Castle	The Sci-Fi book by Kafka	2.01709E+11	u4@rit.edu	Sandra	Smith	19720225	156 Monroe Ave	Brighton	14512	NY
15	2	American Gods	The Sci-Fi book by Gaiman	2.01402E+11	u1@rit.edu	Sandra	Smith	19780905	189 Monroe Ave	Brighton	14512	NY
16	4	The Castle	The Sci-Fi book by Kafka	2.0151E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
17	5	I, Robot	The Sci-Fi book by Asimov	2.01609E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
18	6	The Graveyard Book	The Sci-Fi book by Gaiman	2.01707E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
19	1	Ender's Game	The Sci-Fi book by Scott Card	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
20	1	Ender's Game	The Sci-Fi book by Scott Card	2.01707E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
	?	?	?	?	u8@rit.edu	John	Doe	19920302	249 Brown Street	Rochester	14623	NY

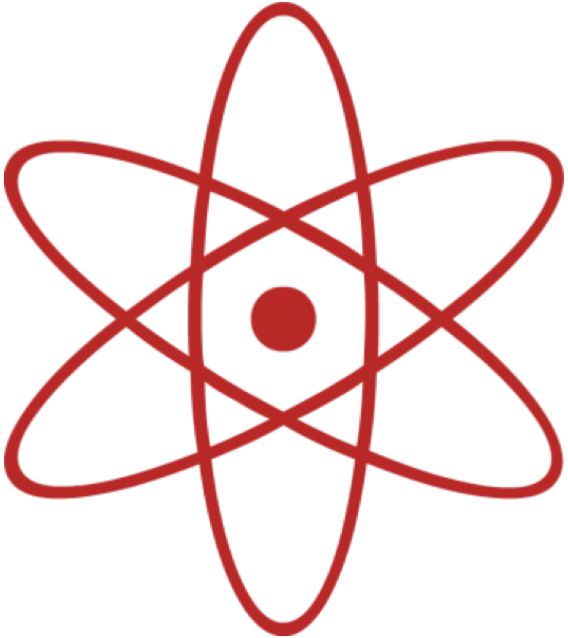
Insertion Anomalies

1	id	name	description	date	email	fn	ln	dob	st	city	zipcode	state
2	1	Ender's Game	The Sci-Fi book by Scott Card	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
3	2	American Gods	The Sci-Fi book by Gaiman	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
4	2	American Gods	The Sci-Fi book by Gaiman	2.0171E+11	u1@rit.edu	Sandra	Smith	19780905	134 West Henrietta Rd	Henrietta	14633	NY
5	3	I, Robot	The Sci-Fi book by Asimov	2.0171E+11	u1@rit.edu	Sandra	Smith	19780905	134 West Henrietta Rd	Henrietta	14633	NY
6	1	Ender's Game	The Sci-Fi book by Scott Card	2.01709E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
7	2	American Gods	The Sci-Fi book by Gaiman	2.01709E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
8	2	American Gods	The Sci-Fi book by Gaiman	2.01707E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
9	~~3~~	~~I, Robot~~	~~The Sci-Fi book by Asimov~~	~~2.01609E+11~~	~~u2@rit.edu~~	~~John~~	~~Johnson~~	~~19681231~~	~~156 Monroe Ave~~	~~Brighton~~	~~14512~~	~~NY~~
10	1	Ender's Game	The Sci-Fi book by Scott Card	2.01705E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
11	3	I, Robot	The Sci-Fi book by Asimov	2.01705E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
12	2	American Gods	The Sci-Fi book by Gaiman	2.01609E+11	u4@rit.edu	Sandra	Smith	19720225	156 Monroe Ave	Brighton	14512	NY
13	4	The Castle	The Sci-Fi book by Kafka	2.01709E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
14	4	The Castle	The Sci-Fi book by Kafka	2.01709E+11	u4@rit.edu	Sandra	Smith	19720225	156 Monroe Ave	Brighton	14512	NY
15	2	American Gods	The Sci-Fi book by Gaiman	2.01402E+11	u1@rit.edu	Sandra	Smith	19780905	189 Monroe Ave	Brighton	14512	NY
16	4	The Castle	The Sci-Fi book by Kafka	2.0151E+11	u4@rit.edu	Sandra	Smith	19720225	175 Ash St	Scranton	18519	PA
17	5	I, Robot	The Sci-Fi book by Asimov	2.01609E+11	u2@rit.edu	John	Johnson	19681231	156 Monroe Ave	Brighton	14512	NY
18	6	The Graveyard Book	The Sci-Fi book by Gaiman	2.01707E+11	u3@rit.edu	John	Richardson	19810415	156 Monroe Ave	Brighton	14512	NY
19	1	Ender's Game	The Sci-Fi book by Scott Card	2.01609E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY
20	1	Ender's Game	The Sci-Fi book by Scott Card	2.01707E+11	u1@rit.edu	Sandra	Smith	19780905	156 Monroe Ave	Brighton	14512	NY

Deletion Anomalies

Normalizing

1. Analyze first normal form (1NF)
2. Identify keys and functional dependencies (FDs)
3. Decompose to third normal form (3NF)
4. Analyze results



Values must be atomic

First normal form

Doctor		
ssn	lastName	patients
479-23-8734	Moore	235-14-785 … -48-0924
293-58-9309	Kent	821-13-2 …

No structure in attributes

One Lomb Memorial Drive,
Rochester, NY 14623-5603

Special case: addresses

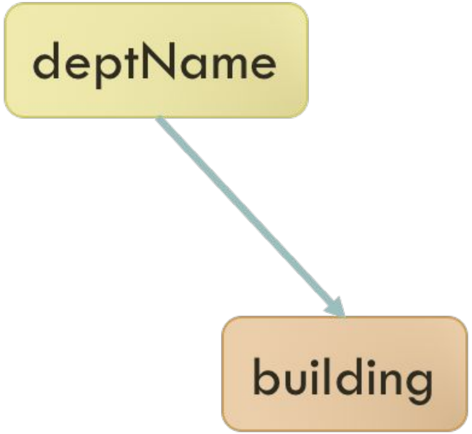
CSCI-620

Special case: course ID

Roadmap

1. Introduction to normalization
2. Keys and functional dependencies
3. Normal forms and decomposition
4. Functional dependency discovery
5. Denormalization

Keys



Functional dependencies

Identify keys and FDs

Types of keys

Super

Candidate

Primary

Patient		
ssn	firstName	lastName
235-14-7854	Sandra	Smith
192-48-0924	John	Moore
821-13-2108	Laura	Turner
874-72-0093	John	Moore

{firstName, lastName} ✗

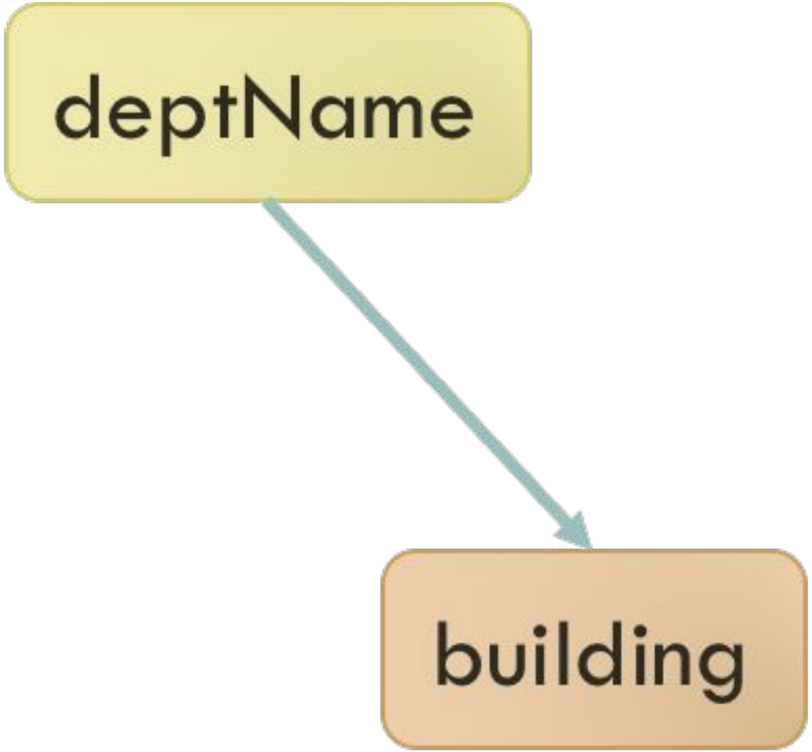
{firstName} ✗

{ssn, lastName} ✓

Superkeys (informally)

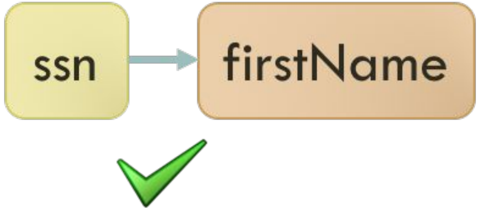
Superkeys (formally)

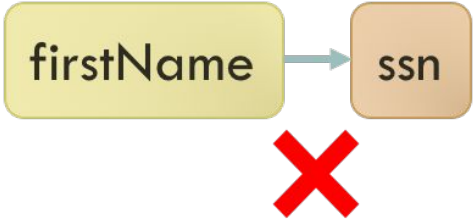
Let r(R) be a relation name with R attributes. K $\subseteq$ R is a superkey of r(R) if, for all tuples t_1 and t_2 in the instance of r, if $t_1 \neq t_2$ then $t_1[K] \neq t_2[K]$. That is, no two tuples in any instance of relation r may have the same value on attribute set K. Clearly, if no two tuples in r have the same value on K, then a K-value uniquely identifies a tuple in r.



Functional dependencies

Patient		
ssn	firstName	lastName
235-14-7854	Sandra	Smith
192-48-0924	John	Moore
821-13-2108	Laura	Turner
874-72-0093	John	Moore





FDs (informally)

FDs (formally)

Let $r(R)$ be a relation name with R attributes, and $\alpha \subseteq R$ and $\beta \subseteq R$. An instance of r satisfies the functional dependency $\alpha \rightarrow \beta$ if, for all tuples t_1 and t_2 in the instance such that $t_1[\alpha] = t_2[\alpha]$, it is also the case that $t_1[\beta] = t_2[\beta]$. We say that a functional dependency $\alpha \rightarrow \beta$ holds on $r(R)$ if every instance of $r(R)$ satisfies the functional dependency.

Defining FDs

Instructor

id	lastName	salary
44	Smith	$65K
47	Moore	$67K

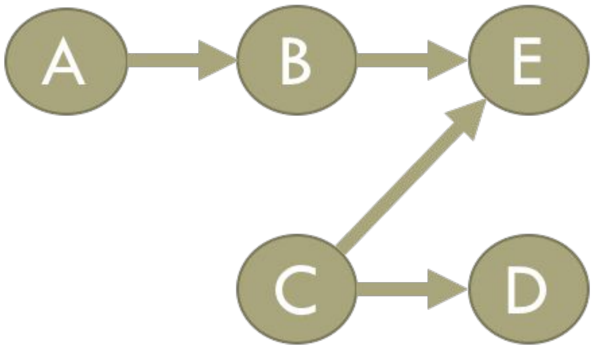
id	lastName	salary
49	Smith	$69K

lastName ✗→ salary

α (a set of attributes) is a superkey for relation $r(R)$ if $R \subseteq \alpha^+$.

Identifying superkeys

Starting with a set of attributes, can I reach all possible attributes?



Identifying superkeys

```
result := α
Do
        For each γ → δ in F
                If (γ ⊆ result)
                        result := result ∪ δ
While (result change);
Return result
```

Closure of α under F (α^+)

Having F, a simplified set of functional dependencies that works exactly the same as F.

Canonical cover

Attributes that we can remove without altering the functional dependencies.

Extraneous attributes

Extraneous attributes

Given $\alpha \rightarrow \beta$:

- $A \in \alpha$ is extraneous if $\beta \subseteq (\alpha - A)^+$ under F
  A is not necessary to infer β

or

- $A \in \beta$ is extraneous if $A \in \alpha^+$ under
  $F' = (F - \{\alpha \rightarrow \beta\}) \cup \{\alpha \rightarrow (\beta - A)\}$
  A can be inferred from other FDs in F using α

result := F

Do {

If $\alpha \rightarrow \beta$ in result has a extraneous attribute,

remove the attribute from $\alpha \rightarrow \beta$ in result

} While (result change)

Return result

Canonical cover

$F = \{A \rightarrow BC, B \rightarrow C, A \rightarrow B, AB \rightarrow C, A \rightarrow C\}$

$A \rightarrow BC$ can have B removed because of $A \rightarrow B$ (now $A \rightarrow C$)

$AB \rightarrow C$ can be removed because of $B \rightarrow C$ and $A \rightarrow C$

$A \rightarrow C$ can be removed because of $A \rightarrow B$ and $B \rightarrow C$

$F^+ = \{A \rightarrow B, B \rightarrow C\}$

Canonical cover

Core

1. All attributes that never appear in functional dependencies
2. All attributes that are only on the left-hand side of functional dependencies and never on the right

Exterior

3. All attributes that appear on the right-hand side of some functional dependency

Finding candidate keys

ABCD

$ABC \rightarrow D$

$D \rightarrow A$

Core	BC
Exterior	DA

ABCDEFG

$AB \rightarrow F$

$AD \rightarrow E$

$F \rightarrow G$

Core	ABCD
Exterior	EFG

Finding candidate keys

Test the closure of the core. If this is the complete set of attributes, the core is the only candidate key.

Otherwise, try combining the core with one attribute in the exterior and check the closure. If the closure is the complete set of attributes, this is another candidate key.

Otherwise, continue adding attributes and checking the closure until you find all candidate keys. Do *not* consider any attributes which are a superset of an existing candidate key.

Finding candidate keys

Finding candidate keys

ABCD

$ABC \rightarrow D$	**Core**	BC
$D \rightarrow A$	**Exterior**	AD

Closure of the core BC

Combination	**Closure**	
ABC	ABCD	✓
BCD	ABCD	✓

Roadmap

1. Introduction to normalization
2. Keys and functional dependencies
3. Normal forms and decomposition
4. Functional dependency discovery
5. Denormalization

$$\text{ssn} \longrightarrow \{\text{firstName, lastName, zipCode, city}\}$$

$$\{\text{zipCode, city}\} \longrightarrow \text{state}$$

Transitive dependencies

Testing 3NF

For each $\alpha \rightarrow \beta$:

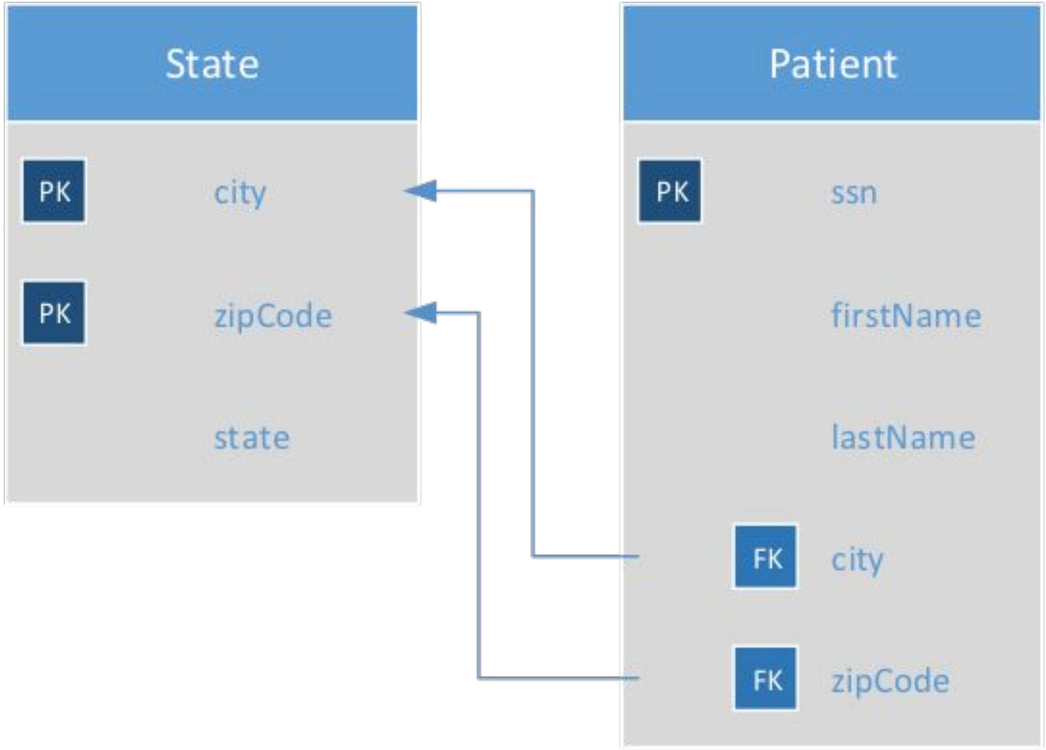
- Trivial ($\beta \subseteq \alpha$)

OR

- α is superkey

OR

- β - α is contained in some candidate key



3NF Decomposition

Let r be a relation and F a set of FDs, F_C is a canonical cover for F

For each $\alpha \rightarrow \beta$ in F_C

Create $r_i(\alpha\beta)$

If none r_i contains a candidate key for r

Create r_i(Any candidate key)

Remove all redundant $r_x(\alpha) \subseteq r_y(\beta)$

Return all relations

3NF Decomposition

Tournament Winners

Tournament	Year	Winner	Winner Date of Birth
Indiana Invitational	1998	Al Fredrickson	21 July 1975
Cleveland Open	1999	Bob Albertson	28 September 1968
Des Moines Masters	1999	Al Fredrickson	21 July 1975
Indiana Invitational	1999	Chip Masterson	14 March 1977

Tournament, Year → Winner, Winner Date of Birth

Winner → Winner Date of Birth

3NF Example

Canonical cover

Tournament, Year → Winner

~~Winner → Winner Date of Birth~~

The second FD is

Not trivial

The left-hand side is not a superkey

The right-hand side minus the left is not part of a candidate key

3NF Example

Tournament Winners

Tournament	Year	Winner
Indiana Invitational	1998	Al Fredrickson
Cleveland Open	1999	Bob Albertson
Des Moines Masters	1999	Al Fredrickson
Indiana Invitational	1999	Chip Masterson

Winner Dates of Birth

Winner	Date of Birth
Chip Masterson	14 March 1977
Al Fredrickson	21 July 1975
Bob Albertson	28 September 1968

3NF Example

TABLE_BOOK_DETAIL

Book ID	Genre ID	Genre Type	Price
1	1	Gardening	25.99
2	2	Sports	14.99
3	1	Gardening	10.00
4	3	Travel	12.99
5	2	Sports	17.99

Book ID → Genre ID, Genre Type, Price
Genre ID → Genre Type

3NF Example

Canonical cover

Book ID → Genre ID, Genre Type

~~Genre ID → Genre Type~~

The second FD is

Not trivial

The left-hand side is not a superkey

The right-hand side minus the left is not part of a candidate key

3NF Example

TABLE_BOOK

Book ID	Genre ID	Price
1	1	25.99
2	2	14.99
3	1	10.00
4	3	12.99
5	2	17.99

TABLE_GENRE

Genre ID	Genre Type
1	Gardening
2	Sports
3	Travel

3NF Example

ID	name	street	city	salary
⋮				
57766	Kim	Main	Perryridge	75000
98776	Kim	North	Hampton	67000
⋮				

employee

ID	name
⋮	
57766	Kim
98776	Kim
⋮	

name	street	city	salary
⋮			
Kim	Main	Perryridge	75000
Kim	North	Hampton	67000
⋮			

natural join

ID	name	street	city	salary
⋮				
57766	Kim	Main	Perryridge	75000
57766	Kim	North	Hampton	67000
98776	Kim	Main	Perryridge	75000
98776	Kim	North	Hampton	67000
⋮				

Lossy join

Decomposition of $R = (A, B, C)$

$R_1 = (A, B)$ $R_2 = (B, C)$

A	B	C
α	1	A
β	2	B

r

A	B
α	1
β	2

$\Pi_{A,B}(r)$

B	C
1	A
2	B

$\Pi_{B,C}(r)$

$\Pi_A (r) \bowtie \Pi_B (r)$

A	B	C
α	1	A
β	2	B

Lossless join

ER model and normalizing

- A carefully designed ER model means we don’t normally need normalization
- However, functional dependencies may indicate the design is not normalized
- Checking that our model is normalized can confirm the quality of ER model

Roadmap

1. Introduction to normalization
2. Keys and functional dependencies
3. Normal forms and decomposition
4. Functional dependency discovery
5. Denormalization

- All possible attribute combinations; for instance, r(a, b, c):
  - {a} → a
  - {a} → b
  - {a} → c
  - {b} → a
  - …
  - {a, b} → c
  - …
  - {a, b, c} → c

$$\mathcal{O}(n^2(\frac{m}{2})^2 2^m)$$

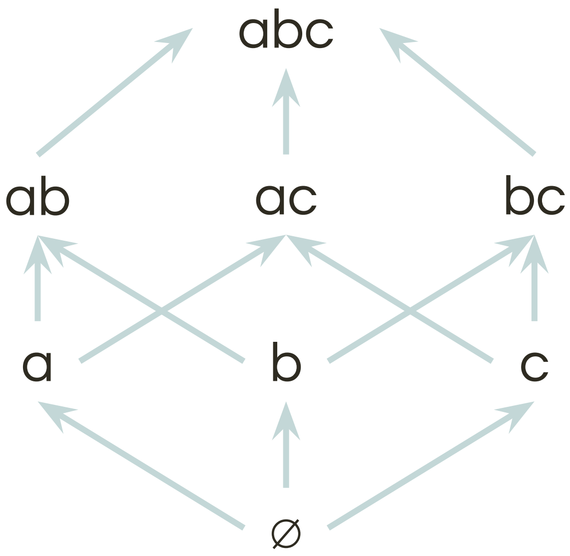
Naive approach

Two problems

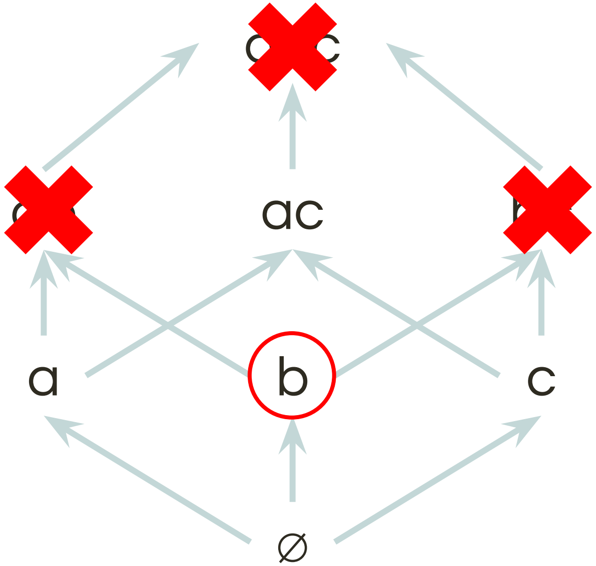
1. We check all dependencies even if we should already know some of them are true (e.g. $A \rightarrow B$ implies $AC \rightarrow B$)

2. Checking a single dependency is $O(n^2)$

r(a, b, c) has the following lattice:



Lattice traversal



For a given left-hand side (LHS), any dependency at lower levels of the lattice implies all dependencies which are higher are true.

Lattice traversal

- Π_α is a partition of the tuples of a relation r
- Tuples t1 and t2 belong to Π_α iff $\forall\ x \in \alpha$: t1[x] = t2[x]

Instructor		
	first	last
t1	Peter	Miller
t2	Thomas	Miller
t3	Peter	Moore
t4	Peter	Miller

- $\Pi_{\{first\}} = \{\{t1, t3, t4\}, \{t2\}\}$
- $\Pi_{\{last\}} = \{\{t1, t2, t4\}, \{t3\}\}$
- $\Pi_{\{first,\ last\}} = \Pi_{\{first\}} \cap \Pi_{\{last\}} = \{\{t1, t4\}, \{t2\}, \{t3\}\}$

Partitions

Refinement

- A partition Π refines another partition Π' if every equivalence class in Π is subset of some equivalence class in Π'
- For instance, $\Pi = \{\{t1, t2\}, \{t3\}\}$ refines $\Pi' = \{\{t1, t2, t3\}\}$ but Π' does not refine Π
- A functional dependency $\alpha \rightarrow a$ holds iff Π_α refines $\Pi_{\{a\}}$ (Proof omitted)

Algorithm

- Generate partitions for individual attributes ($\Pi_{\{a\}}$, $\Pi_{\{b\}}$, etc.)
- FDS := ∅
- For each level in the lattice starting from level zero (bottom-up) and each set of attributes α in each level:

(Inside loop)

- For each possible right-hand side (RHS) attribute b:
  - If FDS contains any functional dependency such that $\beta \rightarrow b$ where $\beta \subset \alpha$, then skip
  - Compute Π_α and if Π_α refines $\Pi_{\{b\}}$, then FDS := FDS $\cup$ $\{\alpha \rightarrow b\}$

Algorithm

Pruning

If FDS contains any functional dependency such that $\beta \rightarrow b$ where $\beta \subset \alpha$, then skip

Example:

$\alpha = \{A, B\}$ (current entry in lattice), $b = C$ (possible RHS)

If we found $A \rightarrow C$, then we can prune $A,B \rightarrow C$ because we are guaranteed that it holds.

Algorithm

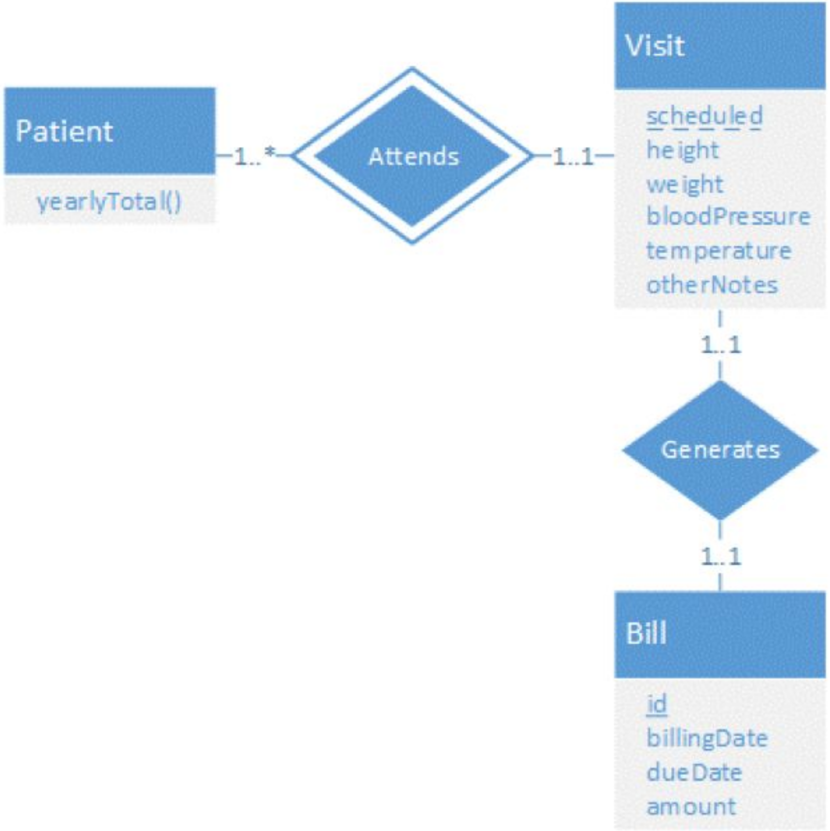
Stripped partitions

- A partition with equivalent classes of size one removed denoted by Π^S
- Example: $\Pi^S_{\{first, last\}} = \Pi^S_{\{first\}} \cap \Pi^S_{\{last\}} = \{\{t1, t4\}\}$
- Intuitive explanation: a singleton equivalence class (size one) can never break any functional dependency on the left-hand side (Proof omitted)

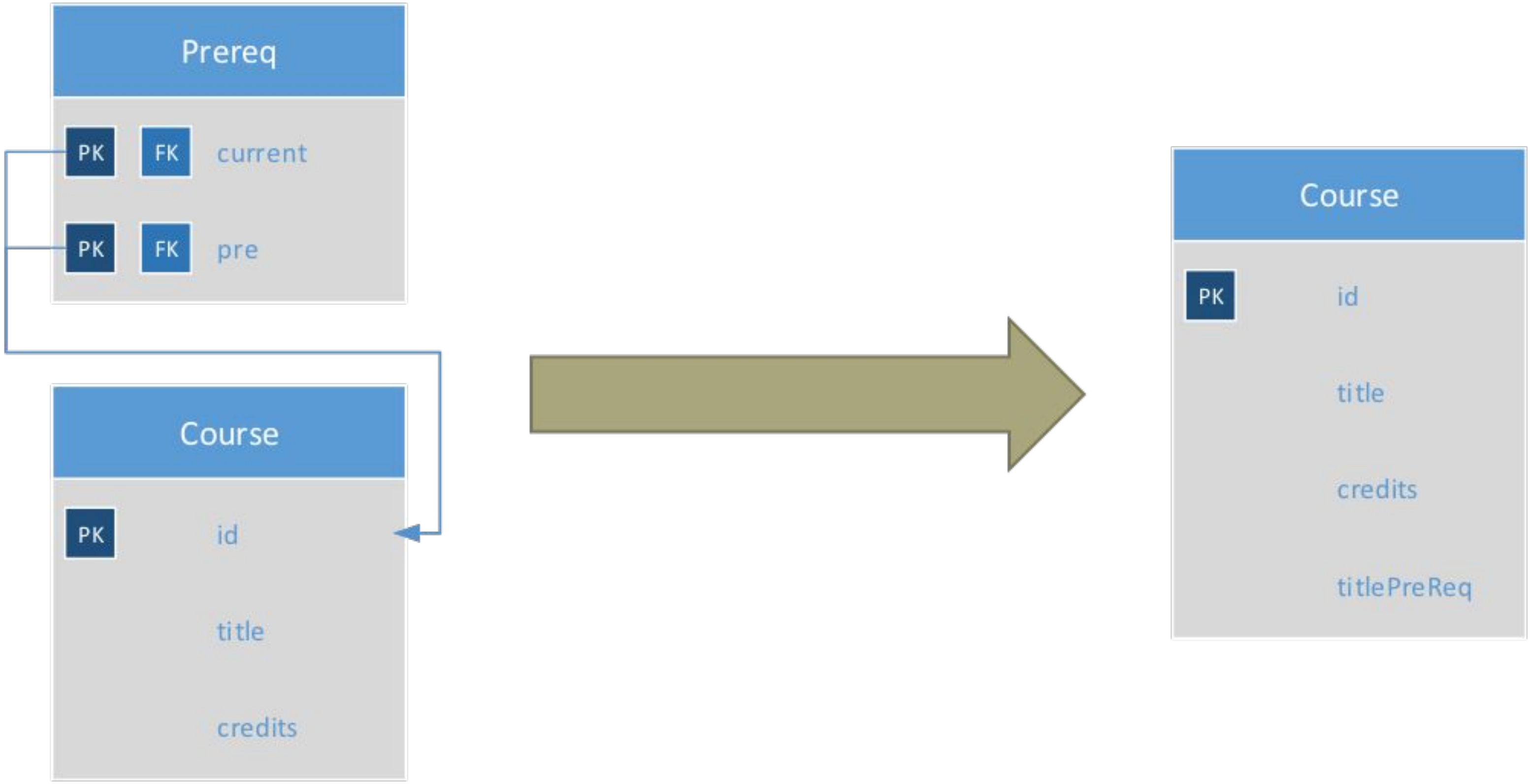
Roadmap

After normalization, we can update the logical model to introduce redundancy and (usually) improve performance

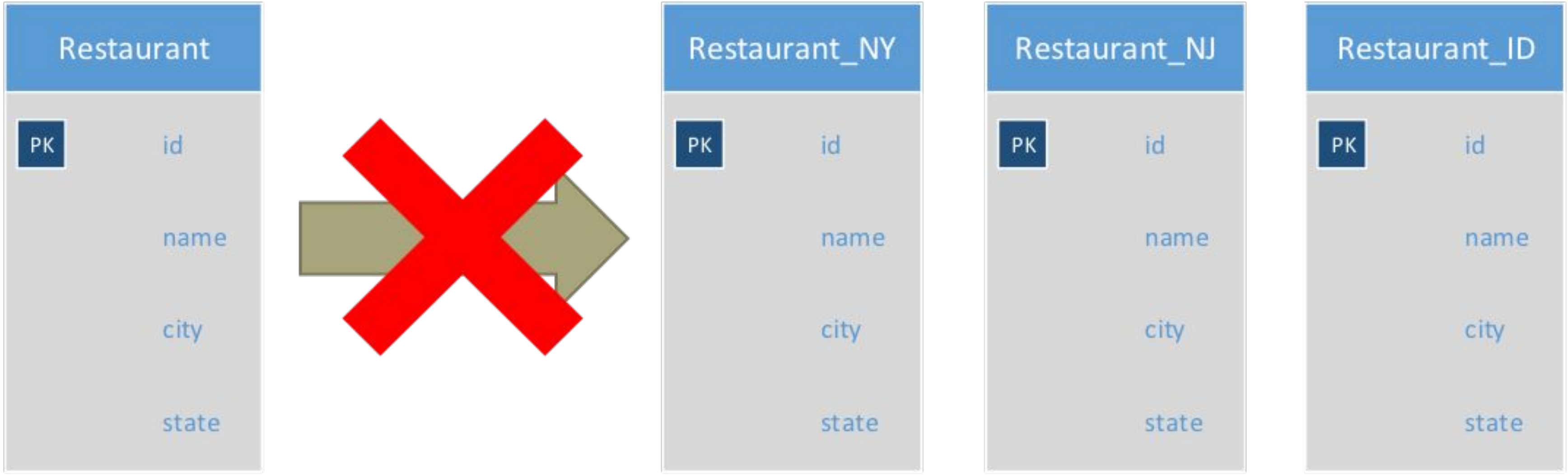
Denormalization



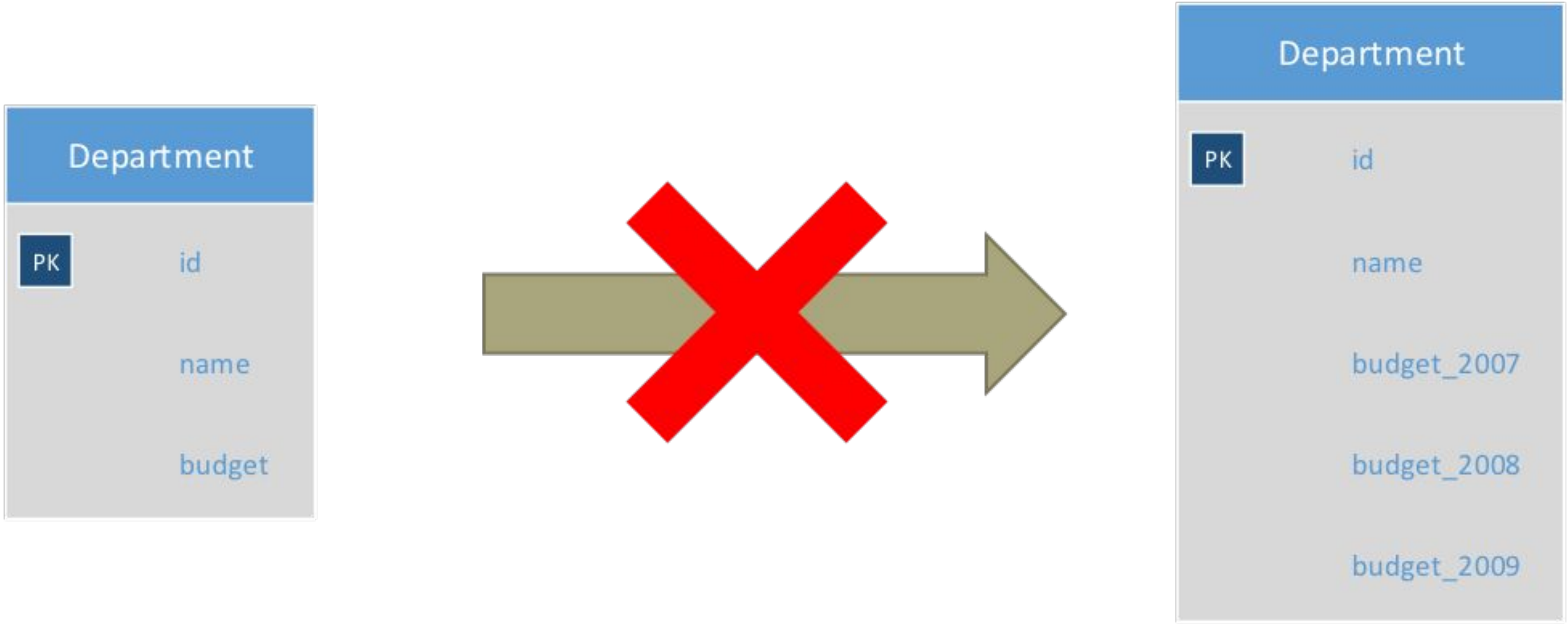
Derived attributes



Avoiding joins



Discrimination


Department
PK
id
name
budget
Department
PK
id
name
budget_2007
budget_2008
budget_2009

Crosstabs

Computed Columns

- Sometimes we have a value which is expensive to compute and used in many queries

- In this case, we can use a *computed column* and ask the database to store the result when the row is changed

```
CREATE MATERIALIZED VIEW salaries_by_dept AS
SELECT department, MIN(salary), MAX(salary)
FROM Doctor GROUP BY department;
```

CREATE MATERIALIZED VIEW

Denormalization

- Should be used **very** sparingly in most database applications
- Like indexing, denormalization should be done for the benefit of queries
- Use database features to help with denormalization (e.g. triggers)
- Careful with possible anomalies!