

COL 362 & COL 632

Normal Forms and Relational Algebra

18 Jan 2023

First normal form

- 1NF (First normal form)
 - A relation is in 1NF iff every tuple contains an atomic value for each attribute
 - Follows directly from definition of relation
 - Relation contains a key
- A relation that is **only** in 1NF (and not 2NF & above), has
 - Trouble with inserting new items – no actor unless we have a movie, no student without registering for at least one course, ...
 - Trouble with delete – if we remove a movie, we may lose the actor as well!
 - Trouble with update – any change must be done to all instances in the relation instance

Second normal form (1/2)

1NF + No non-prime attribute in the table is functionally dependent on a proper subset of any candidate key

Name DOB → Address

(Name DOB) MTitle Year → Address

Name	DOB	Address	MTitle	Year	Language
Priyanka Chopra	1992	Mumbai	Don	2006	Hindi
Priyanka Chopra	1992	Mumbai	Don II	2011	Hindi
Tom Cruise	1962	LA	MI-IV	2011	English
Anthony Hopkins	1937	LA	Thor: Ragnarok	2017	English
Bill Nighy	1949	LA	Valkyrie	2008	English

What are we missing here?

Name	DOB	Address
Priyanka Chopra	1992	Mumbai
Anthony Hopkins	1937	LA
Bill Nighy	1949	LA
Tom Cruise	1962	LA

MTitle	Year	Language
Don	2006	Hindi
Don II	2011	Hindi
MI-IV	2011	English
Valkyrie	2008	English
Thor: Ragnarok	2017	English

Second normal form (2/2)

Bridge tables

VR

Name	DOB	Address	MTitle	Year	Language
Priyanka Chopra	1992	Mumbai	Don	2006	Hindi
Priyanka Chopra	1992	Mumbai	Don II	2011	Hindi
Anthony Hopkins	1937	LA	MI-IV	2011	English
Anthony Hopkins	1937	LA	Valkyrie	2017	English
Bill Nighy	1949	LA	Valkyrie	2008	English

No non-prime attribute in the table is functionally dependent on a proper subset of any candidate key

R₁

ID	Name	DOB	Address
1	Priyanka Chopra	1992	Mumbai
2	Anthony Hopkins	1937	LA
3	Bill Nighy	1949	LA
4	Tom Cruise	1962	LA

MID

AID	MID
1	1
1	2
2	3
3	4
4	5
4	3

R₂

ID	MTitle	Year	Language
1	Don	2006	Hindi
2	Don II	2011	Hindi
3	MI-IV	2011	English
4	Valkyrie	2008	English
5	Thor: Ragnarok	2017	English

Third normal form

non-trivial
For any $FD \in F^+$

$$Y \subseteq X$$

$$\left. \begin{array}{l} \alpha\beta \rightarrow \gamma \\ \alpha\beta \rightarrow \delta \\ \alpha\beta \rightarrow \gamma\delta \end{array} \right\}$$

• 2NF +

• For a non-trivial FD $X \rightarrow Y$, X is a superkey or Y is prime

Address \rightarrow Country

Name, DOB \rightarrow Address

Name, DOB \rightarrow Country

↓ ↓ ↓

<u>Name</u>	<u>DOB</u>	Address	Country
Priyanka Chopra	1992	Mumbai	India
Anthony Hopkins	1937	LA	USA
Bill Nighy	1949	LA	USA

↓ ↓ ↓ ↓

<u>Name</u>	<u>DOB</u>	AID	ID	Address	Country
Priyanka Chopra	1992	1	1	Mumbai	India
Anthony Hopkins	1937	2	2	LA	USA
Bill Nighy	1949	2			

Boyce-Codd normal form

BCNF

For all **non-trivial** FDs $X \rightarrow Y$ in F^+

X is a superkey

Addresses the following additional scenarios:

- Multiple candidate keys with intersecting elements ✓
- All attributes are part of some key ✓

Title	Theatre	City
-------	---------	------

Keys: Title, City

Theatre, Title

FDs:

Theatre \rightarrow City

Title, City \rightarrow Theatre

Theatre, Title \rightarrow City

Lossless decomposition

- Algorithm:
 - If $X \rightarrow Y$ is a BCNF violation, then form two relations:
 - with attributes from $X \cup Y$
 - with attributes from $X \cup (all-X-Y)$

Relational algebra

What is an “Algebra”

- Mathematical system consisting of:
 - *Operands* --- variables or values from which new values can be constructed.
 - *Operators* --- symbols denoting procedures that construct new values from given values.

What is relational algebra?

rel vars

- An algebra whose operands are relations or variables that represent relations.
- Operators are designed to do the most common things that we need to do with relations in a database.
- Defines basic operations on relation instances
 - composition of operations to form queries
- The result is an algebra that can be used as a query language for relations.
 - Basis for SQL
- Useful to represent execution plans
 - what are the operations needed to execute a query
 - what is the order of execution of these operations

$$\left(\frac{(a+b) \times x}{y} \right) \div c$$

Basic operations

- Selection σ (choose subset of rows)
- Projection Π (choose subset of columns)
- Cross product \times
- Union \cup
- Difference $-$
- Rename ρ
- Join \bowtie

Same operations as those on any sets
Apply operations on tuples with same schema

Rename

$$\rho_{R(A_1, A_2, \dots)}(S)$$
$$\rho_{Stars(Name, Age, City)}(Actors)$$

$$\rho_{Stars(SName, Sage, Scity)}$$

Actors

Name	Age	Addr
Priyanka Chopra	38	Mumbai
Anthony Hopkins	81	LA
Bill Nighy	69	LA
Abhishek Bachchan	45	Mumbai



Stars

Name	Age	City
Priyanka Chopra	38	Mumbai
Anthony Hopkins	81	LA
Bill Nighy	69	LA
Abhishek Bachchan	45	Mumbai

Selection (1/2)

$$R1 = \sigma_C(R2)$$

C is a condition on attributes of R2

Selected
Set

Actors

Name	Age	Addr
Priyanka Chopra	38	Mumbai
Anthony Hopkins	81	LA
Bill Nighy	69	LA
Abhishek Bachchan	45	Mumbai

Return all actors living in Mumbai

$\sigma_{Addr='Mumbai'}(Actors)$

Name	Age	Addr
Priyanka Chopra	38	Mumbai
Abhishek Bachchan	45	Mumbai

}

Selection (2/2)

$$R1 = \sigma_C(R2)$$

Actors

Name	Age	Addr
Priyanka Chopra	38	Mumbai
Anthony Hopkins	81	LA
Bill Nighy	69	LA
Abhishek Bachchan	45	Mumbai

Return all actors whose age is more than 42.

$\sigma_{Age > 42}(Actors)$

Name	Age	Addr
Anthony Hopkins	81	LA
Bill Nighy	69	LA
Abhishek Bachchan	45	Mumbai

Return all actors whose age is more than 42 and who live in Mumbai

$\sigma_{Age > 42 \text{ and } Addr = 'Mumbai'}(Actors)$

Name	Age	Addr
Abhishek Bachchan	45	Mumbai

Projection (1/2)

$$R1 = \Pi_L(R2)$$

Actors

Name	Age	Addr
Priyanka Chopra	38	Mumbai
Anthony Hopkins	81	LA
Bill Nighy	69	LA
Abhishek Bachchan	45	Mumbai

Return the name and age of all actors

$$\Pi_{Name, Age}(Actors)$$

Name	Age
Priyanka Chopra	38
Anthony Hopkins	81
Bill Nighy	69
Abhishek Bachchan	45



Projection (2/2)

$$R1 = \Pi_L(R2)$$

Actors

Name	Age	Addr
Priyanka Chopra	38	Mumbai
Anthony Hopkins	81	LA
Bill Nighy	69	LA
Abhishek Bachchan	45	Mumbai

Return the addresses of the actors

$\Pi_{Addr}(Actors)$

Addr
Mumbai
LA

Addr
Mumbai
LA
LA
Mumbai

Cross product

$$A \times B = \{ \underline{(x, y)} \mid x \in A, y \in B \}$$

$(\alpha \beta)$ $R_1.Name, Movies.Name$

$$R3 = R1 \times R2$$

Actors

Name	Age	Addr
Priyanka Chopra	38	Mumbai
Anthony Hopkins	81	LA
Bill Nighy	69	LA
Abhishek Bachchan	45	Mumbai

Movies

Name	Year	Title
Priyanka Chopra	2011	Don-II
Anthony Hopkins	2011	MI-IV
Bill Nighy	2009	Valkyrie
Abhishek Bachchan	2010	Raavan

Actor.name	Age	Addr	Movies.Name	Year	Title
Priyanka Chopra	38	Mumbai	Priyanka Chopra	2011	Don-II
...15 more rows...					