Data Modeling × One Big Table   1

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*

Forget about data modeling in RDBMS
i.e., conceptual (ER), logical (relational) & physical (SQL) levels

Data Modeling × One Big Table   2

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*

Redundancy: repeated directors, genres!

+Simple

- Not efficient! Why?

Data Modeling × One Big Table   3

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*America*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*

UPDATE
An edit to a director info needs an update on all his movies.
Otherwise, there would be inconsistencies!

An update to a genre's title, …

Data Modeling × One Big Table   4

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*America*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
					Roman	*Polanski*	*Aug. 18, 1933*	*France*

INSERT
Adding a new director should be with a movie.
Otherwise, there would be empty spaces!

Adding a new genre, …

Data Modeling × One Big Table

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*				
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*

DELETE
Removing a director must be done for all his movies.
Otherwise, there would be inconsistencies!

Removing a genre, …

Data Modeling × Anomaly 6

Anomaly | Inconsistency
Something that deviates from our expectations

To avoid anomaly in RDBMS
Data Integrity | Integrity Constraints

SQL × DML × Data Integrity   W09B-50

Data Integrity | Integrity Constraints MUST always be assured by DBMS.
ACID Properties (Atomicity, Consistency, Isolation, Durability)

INSERT, UPDATE, DELETE will fail and their effect will be rolled backed if they violate (conflict with) any integrity constraints!

SQL × DML × Data Integrity

I) Domain Integrity
II) Entity Integrity
III) Referential Integrity
IV) User-defined Integrity

Data Modeling × Normalization

To avoid anomaly in RDBMS
   <u>Table decomposition</u> to minimize redundancy
   and improve data integrity.

Data Modeling × Normalization

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*

Data Modeling × Normalization 9

Movie

Id	Title	Language	ReleaseDate
1	2001: A Space Odyssey	1	1968
2	The Shining	1	1980
3	A Clockwork Orange	1	1971
4	The Birds	1	1963
5	Psycho	1	1960

Director

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth
1	Stanley	Kubrick	Jul. 26, 1928	USA
2	Alfred	Hitchcock	Aug. 13, 1899	England

Language

Id	Title
1	English

Genre

Id	Title
1	Sci-fi
2	Drama
3	Crime
4	Mystery
5	Thriller
6	Adventure
7	Horror

MovieGenre

MovieId	GenreId
1	1
1	6
2	2
2	7
3	3
3	2
3	1
4	2
4	7
5	7
5	4
5	5

MovieDirector

MovieId	DirectorId
1	1
2	1
3	1
4	2
5	2

MovieRunningTime

MovieId	RunningTime	Scope
1	142	Globe
2	144	US
2	119	EU
3	136	Globe
3	119	Globe
3	109	Globe

Data Modeling × Normalization

Given a big table of all information, the process of decomposing it into tables in order to avoid redundancy and improve data integrity.

Machine-based!
Algorithm-based!

No conceptual level design. No E/R!
No semantics!

Data Modeling × Normalization

Given a big table of <u>all information</u>, the process of decomposing it into tables in order to avoid redundancy and improve data integrity.

Machine-based!
Algorithm-based!

No conceptual level design. No E/R!
<mark>At conceptual or logical levels, we do not have actual data!</mark>

Data Modeling × Normalization

I) Functional Dependencies
II) Normal Forms

I) Functional Dependencies
II) Normal Forms

Functional Dependencies

A functional dependency occurs when the value of one (set of) attribute(s) determines the value of a second (set of) attribute(s)

Functional Dependencies

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*

Title → Title

(Title, ReleaseDate) → FirstName

(Title, ReleaseDate) → (FirstName, LastName)

Genre ↛ Title

Functional Dependencies

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*

Title → Title
(Title, ReleaseDate) → FirstName
(Title, ReleaseDate) → (FirstName, LastName)
Genre ↛ Title

Determinant

Functional Dependencies

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*

Title → Title
(Title, ReleaseDate) → FirstName
(Title, ReleaseDate) → (FirstName, LastName)
Genre ↛ Title

Composite Determinant

Functional Dependencies

Functional dependencies may be based on equations, e.g., in derived attributes:

TotalPrice = Quantity × UnitPrice
(Quantity, UnitPrice) → TotalPrice

But,

Not Equation
Not Function

Functional Dependencies

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	English	1968	142	Sci-fi, Adventure	Stanley	Kubrick	Jul. 26, 1928	USA
The Shining	English	1980	144 (US), 119 (EU)	Drama, Horror,	Stanley	Kubrick	Jul. 26, 1928	USA
A Clockwork Orange	English	1971	136	Crime, Drama, Sci-Fi	Stanley	Kubrick	Jul. 26, 1928	USA
The Birds	English	1963	119	Drama, Horror	Alfred	Hitchcock	Aug. 13, 1899	England
Psycho	English	1960	109	Horror, Mystery, Thriller	Alfred	Hitchcock	Aug. 13, 1899	England

Functional dependencies are based on the existing data:

Title → ReleaseDate

Title → Genre

Title → FirstName, LastName

PlaceOfBirth → FirstName

Functional Dependencies

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1998*	*104*	*Horror, Mystery, Thriller*	*Gus*	*Van Sant*	*July 24, 1952*	*USA*

Functional dependencies are based on the existing data:

Title ↛ ReleaseDate

Title → Genre

Title ↛ FirstName, LastName

Functional Dependencies

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	English	1968	142	Sci-fi, Adventure	Stanley	Kubrick	Jul. 26, 1928	USA
The Shining	English	1980	144 (US), 119 (EU)	Drama, Horror,	Stanley	Kubrick	Jul. 26, 1928	USA
A Clockwork Orange	English	1971	136	Crime, Drama, Sci-Fi	Stanley	Kubrick	Jul. 26, 1928	USA
The Birds	English	1963	119	Drama, Horror	Alfred	Hitchcock	Aug. 13, 1899	England
Psycho	English	1960	109	Horror, Mystery, Thriller	Alfred	Hitchcock	Aug. 13, 1899	England
Psycho	English	1998	104	Horror, Mystery, Thriller	Gus	Van Sant	July 24, 1952	USA

Determinant might not be unique:

Title → Genre

FirstName → LastName

Functional Dependencies

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1998*	*104*	*Horror, Mystery, Thriller*	*Gus*	*Van Sant*	*July 24, 1952*	*USA*

But if a determinant is unique, then ...

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1998*	*104*	*Horror, Mystery, Thriller*	*Gus*	*Van Sant*	*July 24, 1952*	*USA*

But if a determinant is unique, then it is determinant of ==ALL== other attributes.

RunningTime → Title, Language, …, PlaceOfBirth
ReleaseDate → Title, Language, …, PlaceOfBirth

Functional Dependencies

Table				
A	B	C	D	E
1	1	1	1	1
1	1	2	1	1
2	1	1	1	1
2	2	1	2	1
2	2	2	3	2

BC   ?→ D

B   ?→ A

D   ?→ BE

AB   ?→ C

Functional Dependencies

Table				
A	B	C	D	E
1	1	1	1	1
1	1	2	1	1
2	1	1	1	1
2	2	1	2	1
2	2	2	3	2

BC → D

B ↛ A   Given B=1, two values for A, 1 or 2

D → BE

AB ↛ C   Given (A=1,B=1), two values for C, 1 or 2

Functional Dependencies × Rules

IF	THEN	Not a complete list
A → BC	A → B A → C AB → C AC → B	
A → B B → C	A → C	Transitivity
AB → C	A ↛ C B ↛ C A ↛ BC	Be Careful!
A → B A → C	B ↛ C C ↛ A	Be Careful!

Functional Dependencies × Trivial

A functional dependency is trivial if it is satisfied by every tables

A → A
AB → A
AB → B

Generally, X → Y, where Y ⊆ X.

Trivial FD does not make a significant statement about real world constraints and we only interested in <u>non-trivial</u> FD's.

Functional Dependencies × Super Key

Super Key is a set of attributes that functionally determines <u>ALL</u> the attributes in a table.

Super Key is a set of attributes that identify an entity (row) uniquely.

The trivial Super Key is a set of all attributes of a table.

Functional Dependencies × Super Key

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1998*	*104*	*Horror, Mystery, Thriller*	*Gus*	*Van Sant*	*July 24, 1952*	*USA*

Super Key:
A) {Title, Language, ..., DateOfBirth, PlaceOfBirth}
B) {Title}
C) {Title, Language, Genre}
D) {Title, ReleaseDate}

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	English	1968	142	Sci-fi, Adventure	Stanley	Kubrick	Jul. 26, 1928	USA
The Shining	English	1980	144 (US), 119 (EU)	Drama, Horror,	Stanley	Kubrick	Jul. 26, 1928	USA
A Clockwork Orange	English	1971	136	Crime, Drama, Sci-Fi	Stanley	Kubrick	Jul. 26, 1928	USA
The Birds	English	1963	119	Drama, Horror	Alfred	Hitchcock	Aug. 13, 1899	England
Psycho	English	1960	109	Horror, Mystery, Thriller	Alfred	Hitchcock	Aug. 13, 1899	England
Psycho	English	1998	104	Horror, Mystery, Thriller	Gus	Van Sant	July 24, 1952	USA

Super Key:

A) {Title, Language, ..., DateOfBirth, PlaceOfBirth} ✔

B) {Title} ✖

C) {Title, Language, Genre} ✖

D) {Title, ReleaseDate} ✔

Candidate Key is a minimal Super Key.

Super Key is minimal if it is not possible to remove an attribute from it. Otherwise, it is not Super Key anymore.

Functional Dependencies × Candidate Key 32

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1998*	*104*	*Horror, Mystery, Thriller*	*Gus*	*Van Sant*	*July 24, 1952*	*USA*

Candidate Key (Minimal Super Key):
A) {Title, Language, ..., DateOfBirth, PlaceOfBirth}
B) {Title, FirstName}
C) {RunningTime}
D) {ReleaseDate}

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1998*	*104*	*Horror, Mystery, Thriller*	*Gus*	*Van Sant*	*July 24, 1952*	*USA*

Candidate Key (Minimal Super Key):

A) {Title, Language, ..., DateOfBirth, PlaceOfBirth} ✘

B) {Title, FirstName} ✔

C) {RunningTime} ✔

D) {ReleaseDate} ✔

Functional Dependencies × Primary Key

Primary Key is a Candidate Key selected out of multiple Candidate Keys.

A) (Title, FirstName)
B) (RunningTime)
C) (ReleaseDate)

Best Practice: The best candidate for Primary Key:
I) Less #attributes AND
II) The attributes are mandatory (must have value)

Functional Dependencies × Keys

Consider the table T(A, B, C, D) and the following functional dependencies: B → D, ACD → B, BC → A

Super Keys?
Candidate Keys?
Primary Key?

Functional Dependencies × Keys

Consider the table T(A, B, C, D) and the following functional dependencies: B → D, ACD → B, BC → A

Super Keys?   SK1={A,B,C,D}, SK2={A,C,D}, SK3={B,C}
Candidate Keys?   CK1={A,C,D}, CK2={B,C}
Primary Key?   PK={A,C,D} | PK={B,C}

I) Functional Dependencies
II) Normal Forms

Normalization × Normal Forms

Normalization is done through decomposing tables based on series of normal forms.

There are 11 normal forms:

UNF (1970)	1NF (1971)	2NF (1971)	3NF (1971)	EKNF (1982)	BCNF (1974)	4NF (1977)	ETNF (2012)	5NF (1979)	DKNF (1981)	6NF (2003)

But we only consider UNF, 1NF, 2NF, 3NF and BCNF.
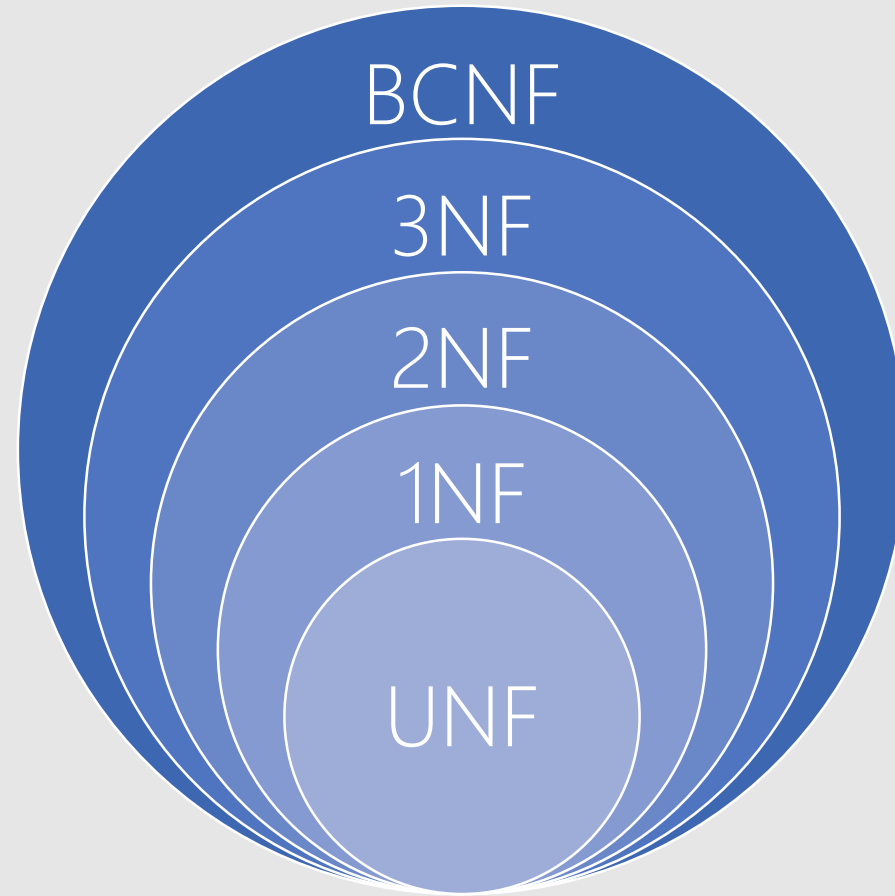
Normalization × UNF   40

Unnormalized Form is the initial date where <u>there is no duplicate tuple (row) in a table</u>.

<mark>Any table MUST have a Primary Key (Entity Integrity).</mark>

This normal form is usually taken for granted.

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1998*	*104*	*Horror, Mystery, Thriller*	*Gus*	*Van Sant*	*July 24, 1952*	*USA*

UNF ✔

BCNF
3NF
2NF
1NF
UNF

Normalization × 1NF

1st Normal Form requires that the domain of each attribute contains only atomic (indivisible) values & UNF.

No composite attribute
No multivalued attribute

Normalization × 1NF

Movie								
Title	Language	ReleaseDate	RunningTime	Genre	FirstName	LastName	DateOfBirth	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*142*	*Sci-fi, Adventure*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Shining	*English*	*1980*	*144 (US), 119 (EU)*	*Drama, Horror,*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
A Clockwork Orange	*English*	*1971*	*136*	*Crime, Drama, Sci-Fi*	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*
The Birds	*English*	*1963*	*119*	*Drama, Horror*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1960*	*109*	*Horror, Mystery, Thriller*	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*
Psycho	*English*	*1998*	*104*	*Horror, Mystery, Thriller*	*Gus*	*Van Sant*	*July 24, 1952*	*USA*

1NF ✖

BCNF
3NF
2NF
1NF
UNF

Movie										
Title	Genre1	Genre2	Genre3	RunningTime1	Scope1	RunningTime2	Scope2	MonthOfBirth	DayOfBirth	YearOfBirth	...
2001: A Space Odyssey	*Sci-fi*	*Adventure*		*142*	*Globe*			*Jul.*	*26*	*1928*	
The Shining	*Drama*	*Horror*		*144*	*US*	*119*	*EU*	*Jul.*	*26*	*1928*	
A Clockwork Orange	*Crime*	*Drama*	*Sci-Fi*	*136*	*Globe*			*Jul.*	*26*	*1928*	
The Birds	*Drama*	*Horror*		*119*	*Globe*			*Aug.*	*13*	*1899*	
Psycho	*Horror*	*Mystery*	*Thriller*	*109*	*Globe*			*Aug.*	*13*	*1899*	
Psycho	*Horror*	*Mystery*	*Thriller*	*104*	*Globe*			*July*	*24*	*1952*	

1NF ✔
What's the problem with this way of normalization?
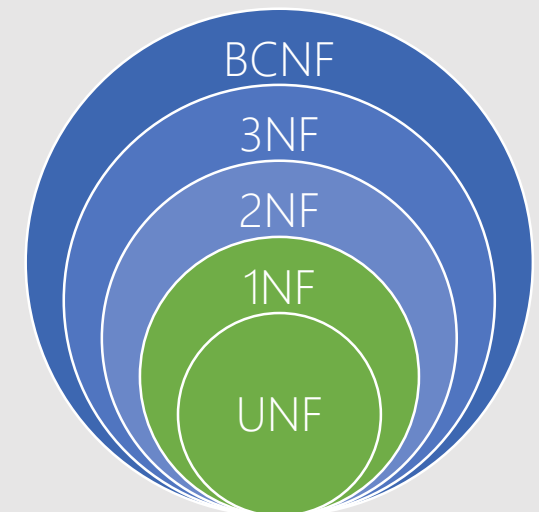
Normalization × 1NF

Movie

Id	Title	Language	ReleaseDate	MonthOfBirth	DayOfBirth	YearOfBirth
1	2001: A Space Odyssey	English	1968	Jul.	26	1928
2	The Shining	English	1980	Jul.	26	1928
3	A Clockwork Orange	English	1971	Jul.	26	1928
4	The Birds	English	1963	Aug.	13	1899
5	Psycho	English	1960	Aug.	13	1899
6	Psycho	English	1998	104	24	1952

MovieGenre

MovieId	GenreId
1	1
1	6
2	2
2	7
3	3
3	2
3	1
4	2
4	7
5	7
5	4
5	5

Genre

Id	Title
1	Sci-fi
2	Drama
3	Crime
4	Mystery
5	Thriller
6	Adventure
7	Horror

1NF ✔

MovieRunningTime

MovieId	RunningTime	Scope
1	142	Globe
2	144	US
2	119	EU
3	136	Globe
3	119	Globe
3	109	Globe

BCNF
3NF
2NF
1NF
UNF
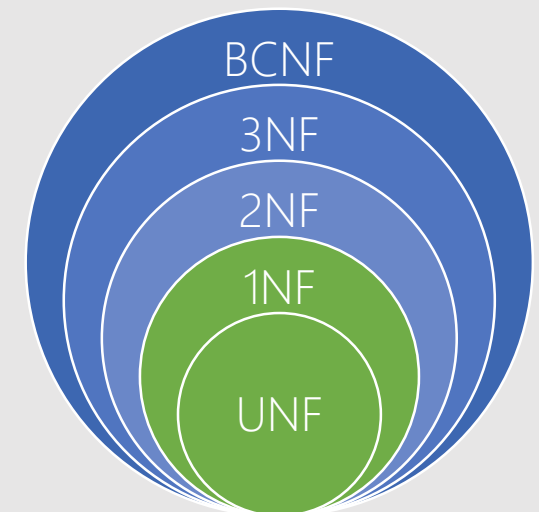
Normalization × 2NF

2nd Normal Form requires that a table:

I) Be in 1NF

II) Does NOT have any non-key attribute that is dependent on any proper subset of <mark>any</mark> candidate key of the table.

A non-key attribute of a table is an attribute
that is not a part of any candidate key of the relation.
🙄
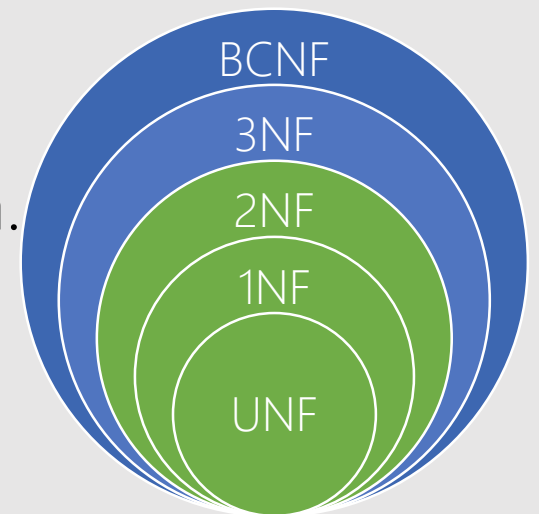
Normalization × 2NF

E.g., T(A, B, C, D, E, F)
CK1 = {A, B}   i.e. AB→CDEF
CK2 = {C, D, E}   i.e. CDE→ABF

F is a non-key attribute since F∉CK1 and F∉CK2
If T is in 2NF, there must be none of these functional dependencies
A→F | B→F
C→F | D→F | E→F | CD→F | DE→F | CE→F
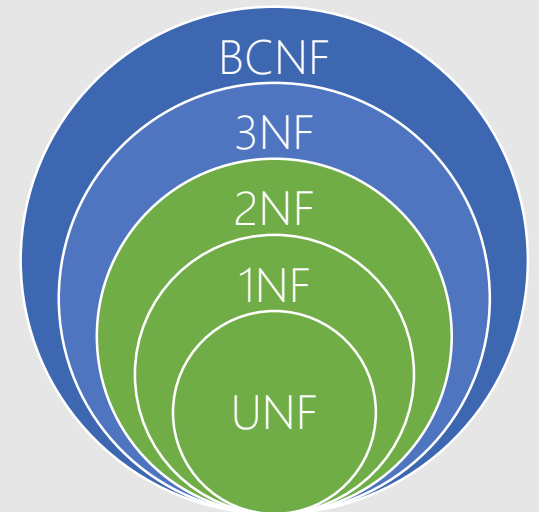
Movie					
Title	Language	ReleaseDate	FirstName	LastName	PlaceOfBirth
2001: A Space Odyssey	*English*	*1968*	*Stanley*	*Kubrick*	*USA*
The Shining	*English*	*1980*	*Stanley*	*Kubrick*	*USA*
A Clockwork Orange	*English*	*1971*	*Stanley*	*Kubrick*	*USA*
The Birds	*English*	*1963*	*Alfred*	*Hitchcock*	*England*
Psycho	*English*	*1960*	*Alfred*	*Hitchcock*	*England*
Psycho	*English*	*1998*	*Gus*	*Van Sant*	*USA*

Candidate Keys:
CK1={Title, FirstName}
CK2={RunningTime}
CK3={ReleaseDate}

non-keys:
Language
LastName
PlaceOfBirth

2NF violations:
Title→Language
FirstName→LastName

BCNF
3NF
2NF
1NF
UNF
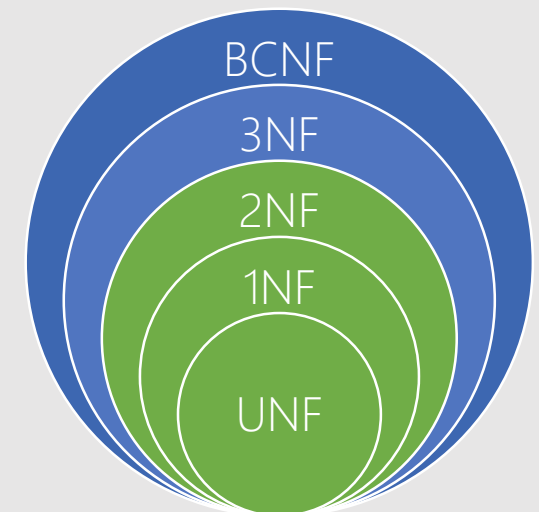
ALGORITHM: Normalize a table in 2NF
Move out data into new tables for functional dependencies
that violate 2NF
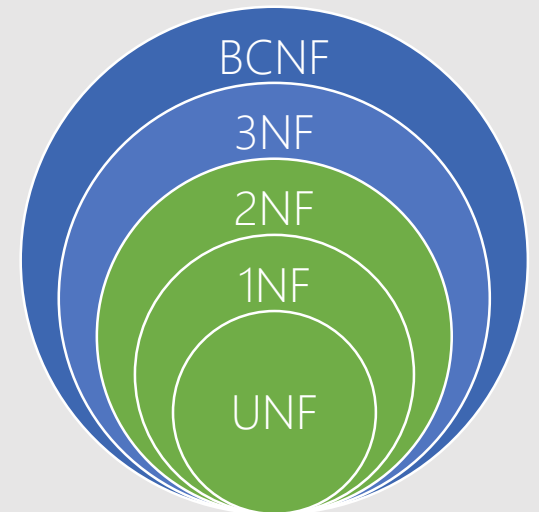
Movie

Id	Title	ReleaseDate	DirectorId
1	2001: A Space Odyssey	1968	1
2	The Shining	1980	1
3	A Clockwork Orange	1971	1
4	The Birds	1963	2
5	Psycho	1960	2
6	Psycho	1998	3

MovieLanguage

MovieId	Title
1	English
2	English
3	English
4	English
5	English
6	English

Director

Id	FirstName	LastName	PlaceOfBirth
1	Stanley	Kubrick	USA
2	Alfred	Hitchcock	England
3	Gus	Van Sant	USA

FirstName→LastName

2NF ✔

Title→Language

BCNF
3NF
2NF
1NF
UNF
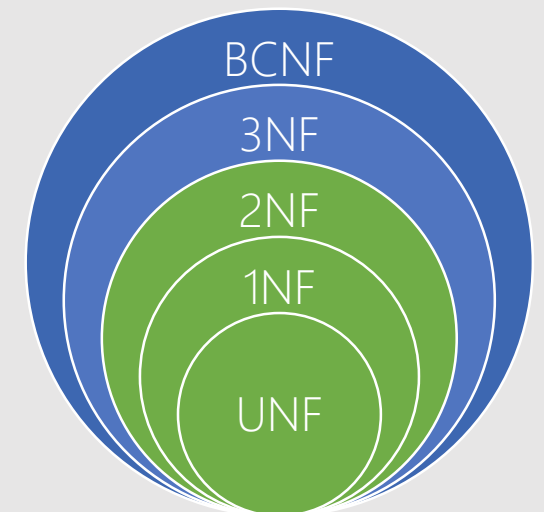
Movie				
Id	Title	LanguageId	ReleaseDate	DirectorId
1	2001: A Space Odyssey	1	1968	1
2	The Shining	1	1980	1
3	A Clockwork Orange	1	1971	1
4	The Birds	1	1963	2
5	Psycho	1	1960	2
6	Psycho	1	1998	3

Language	
Id	Title
1	English

Director			
Id	FirstName	LastName	PlaceOfBirth
1	Stanley	Kubrick	USA
2	Alfred	Hitchcock	England
3	Gus	Van Sant	USA

2NF ✘
Why?

BCNF
3NF
2NF
1NF
UNF
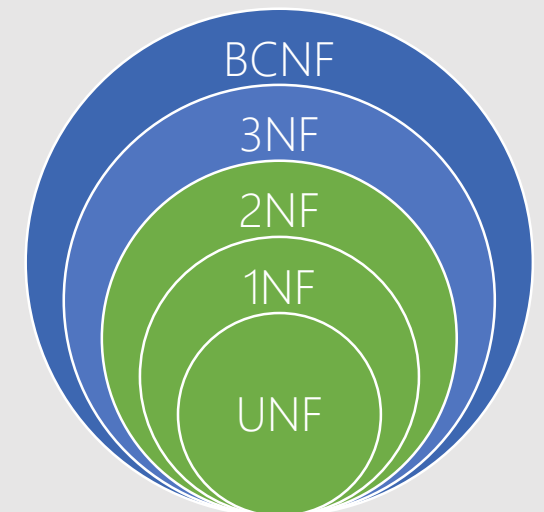
Normalization × 3NF

3rd Normal Form requires that a table:

I)   Be in 2NF

II) All the attributes in a table are determined only by the candidate keys and not by any non-key attributes.

Every non-key attribute of the table is <mark>non-transitively</mark> dependent only on all candidate keys.

No functional dependencies between non-keys.
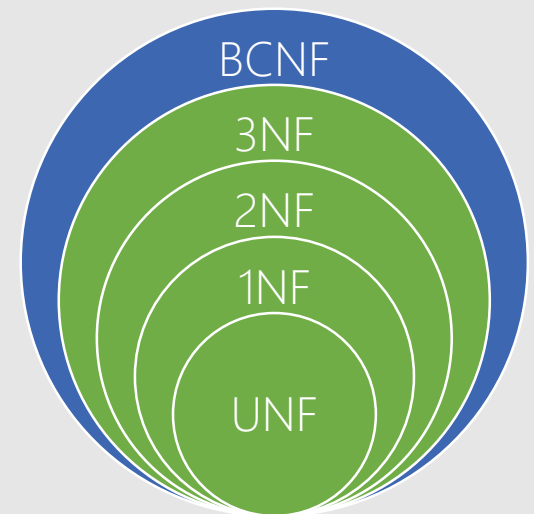
E.g., T(A, B, C, D, E, F)
CK1 = {A, B},   i.e., AB→CDEF
CK2 = {C, D},   i.e., CD→ABEF

E and F are non-key attributes since E,F∉CK1 and E,F∉CK2

If T is in 3NF, there must be <mark>NO</mark> E → F | F → E such that

AB → E & E → F then AB→F (transitivity)
CD → E & E → F then CD → F (transitivity)
AB → F & F → E then AB→ E (transitivity)
CD → F & F → E then CD → E (transitivity)
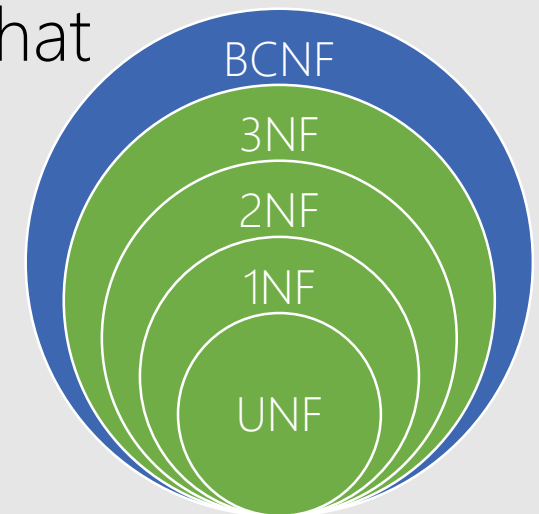
BCNF
3NF
2NF
1NF
UNF

Normalization × 3NF

E.g. Invoice(<u>OrderId</u>, CustomerId, ProductId, Quantity, Price)

PK={OrderId} i.e., OrderId → CustomerId, ProductId, Quantity, Price

Also, there are functional dependencies:

ProductId → Price

Invoice is NOT in 3NF since …
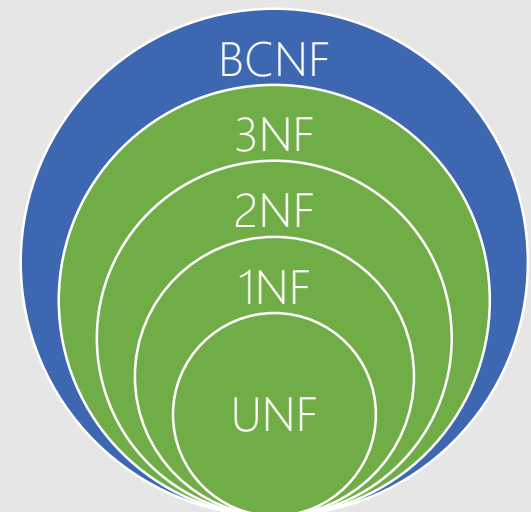
Normalization × 3NF

E.g. Invoice(<u>OrderId</u>, CustomerId, ProductId, Quantity, Price)

PK={OrderId} i.e., OrderId → CustomerId, ProductId, Quantity, Price

Also, there are functional dependencies:

ProductId → Price

ProductId as a non-key is a determinant for Price as another non-key

OrderId → ProductId & ProductId → Price THEN OrderId → Price

BCNF

3NF

2NF

1NF

UNF
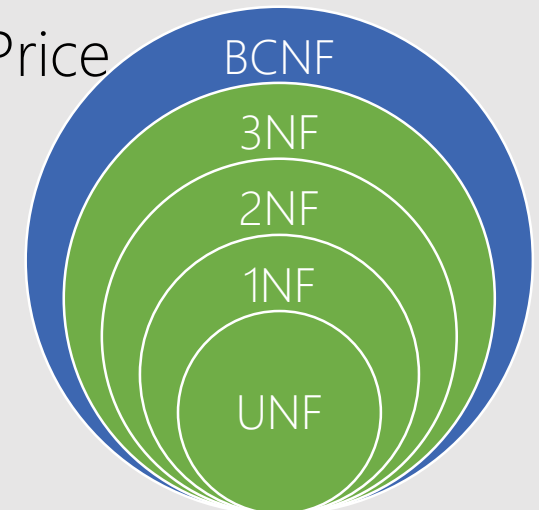
ALGORITHM: Normalize a table in 3NF
Move out data into new tables for functional dependencies that violate 3NF
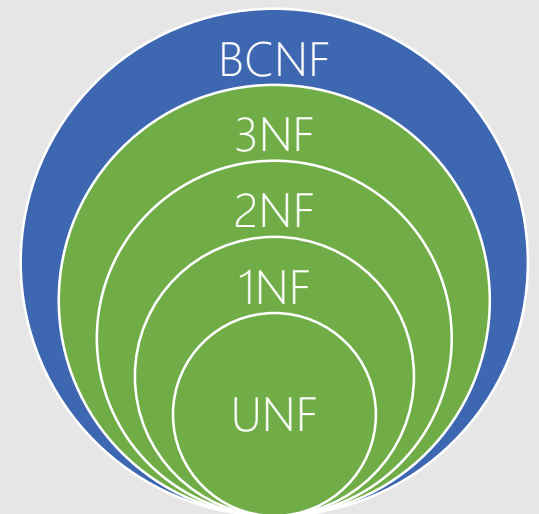
Normalization × 3NF

E.g. Invoice(<u>OrderId</u>, CustomerId, ProductId, Quantity, Price)

PK={OrderId} i.e., OrderId → CustomerId, ProductId, Quantity, Price

Also, there are functional dependencies:

ProductId → Price

Invoice(<u>OrderId</u>, CustomerId, ProductId, Quantity)

ProductPrice(<u>ProductId</u>, Price)
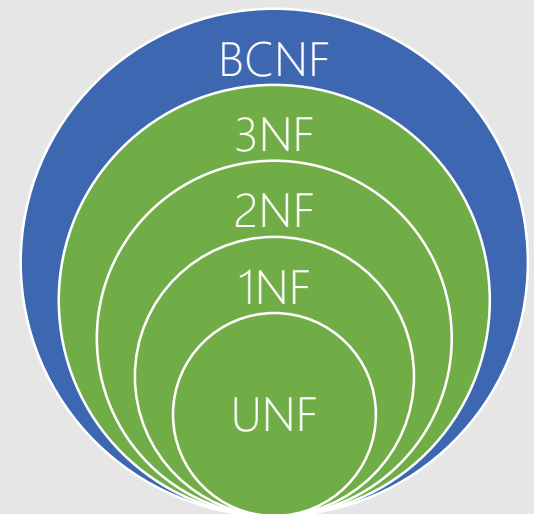
Normalization × 1,2,3NF

1NF: Requiring existence of <mark>"the key"</mark> in the table
2NF: Requiring that non-key attributes be dependent on <mark>"the whole key"</mark>
3NF: Requiring that non-key attributes be dependent on <mark>"nothing but the key"</mark>

○ Both 2NF and 3NF are concerned equally with <mark>ALL</mark> Candidate Keys of a table and not just any one key

○ If there is no non-key, i.e., all attributes are part of at least a candidate key, then table is already in 2NF and 3NF.

○ In 3NF, a non-key attribute is able to be a determinant of a key attribute!
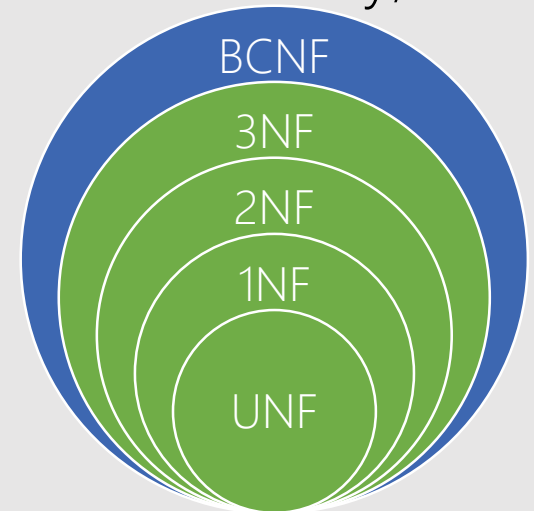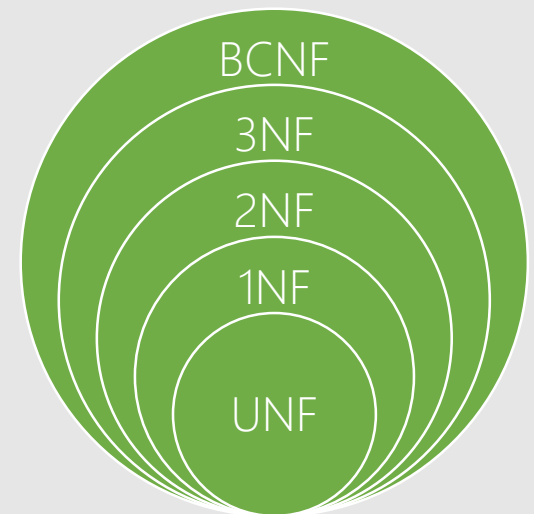
BCNF
3NF
2NF
1NF
UNF

Normalization × BCNF

Boyce-Codd NF | BCNF | 3.5NF, requires that at least one of the following conditions hold for all functional dependencies like X → Y of a table:

I) X → Y is trivial, i.e., X⊆Y
II) X is a Super Key

The different between BCNF and 3NF is …
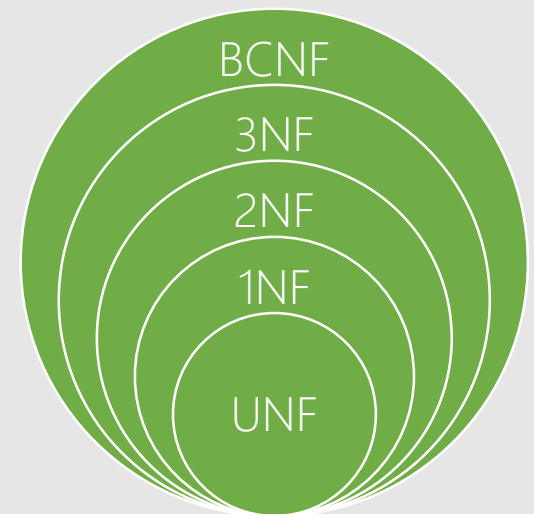
Normalization × BCNF

Boyce-Codd NF | BCNF | 3.5NF, requires that at least one of the following conditions hold for all functional dependencies like X → Y of a table:

I) X → Y is trivial, i.e., X⊆Y
II) X is a Super Key

The different between BCNF and 3NF is:
In 3NF if Y is a Candidate Key, X is could be a non-key
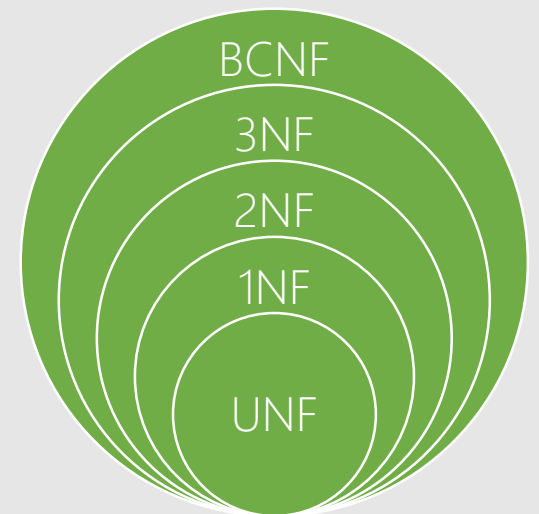In BCNF X must be a Super Key

Normalization × BCNF

E.g. T(A, B, C) with functional dependencies as AB → C and C → B

2NF? If not, decompose T to comply with 2NF.

3NF? If not, decompose T to comply with 3NF.

BCNF? If not, decompose T to comply with BCNF.

Normalization × BCNF

E.g. T(A, B, C) with functional dependencies as AB ➔ C and C ➔ B

Since C ➔ B and AB ➔ C, therefore AC ➔ B. Now, A, B, C are all key attributes and there is no non-key attributes. So, T is in 3NF (2NF).
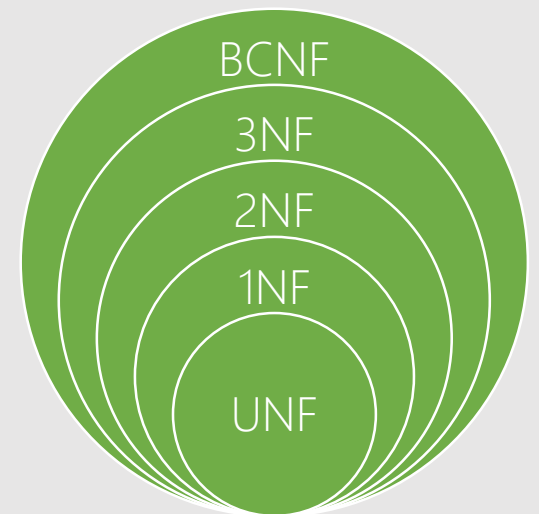
Super Keys: {A, B, C}, {A, B}, {A, C}

However, C ➔ B is violating BCNF as C is not a Super Key!

To make it BCNF, move C ➔ B to a new table, i.e.,

T(<u>A</u>, <u>B</u>)

T'(<u>C</u>, B)

BCNF

3NF

2NF

1NF

UNF

Normalization × BCNF

Interview				
ClientNo	InterviewDate	InterviewTime	StaffNo	RoomNo
76	13-May-02	10.30	05	G101
56	13-May-02	12.00	05	G101
74	13-May-02	12.00	37	G102
56	01-July-02	10.30	05	G102
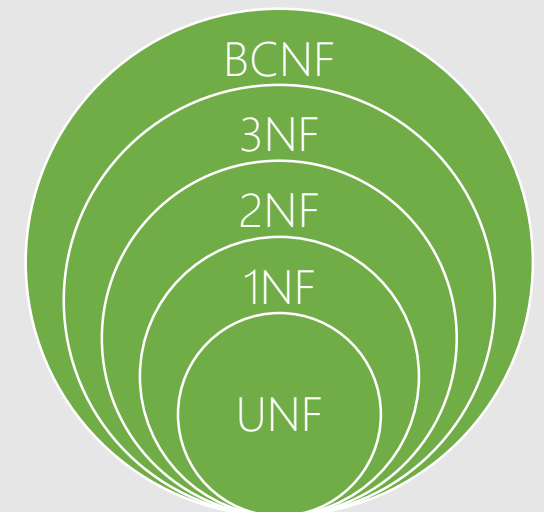
BCNF
3NF
2NF
1NF
UNF

Normalization × BCNF

Interview				
ClientNo	InterviewDate	InterviewTime	StaffNo	RoomNo
76	13-May-02	10.30	05	G101
56	13-May-02	12.00	05	G101
74	13-May-02	12.00	37	G102
56	01-July-02	10.30	05	G102

CK1: ClientNo, InterviewDate → InterviewTime, StaffNo, RoomNo

CK2: StaffNo, InterviewDate, InterviewTime → ClientNo, RoomNo

CK3: RoomNo, InterviewDate, InterviewTime → ClientNo, StaffNo
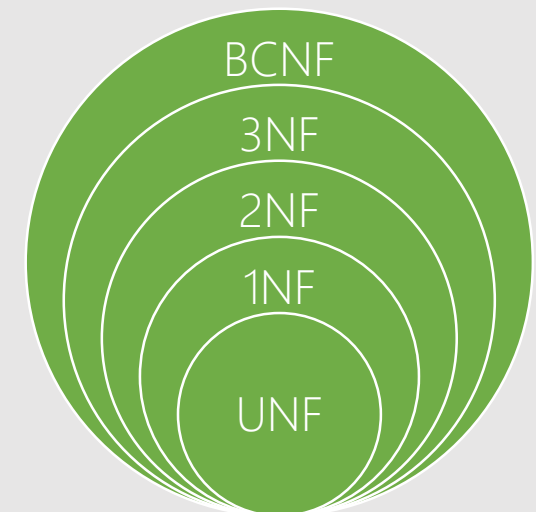
StaffNo, InterviewDate → RoomNo

In UNF, since has a super key

In 1NF, since has no multivalued or composite attribute

In 2NF, since there is no non-key attribute

In 3NF, since there is no non-key attribute

Not in BCNF, since {StaffNo, InterviewDate} is not Super Key

Normalization × BCNF

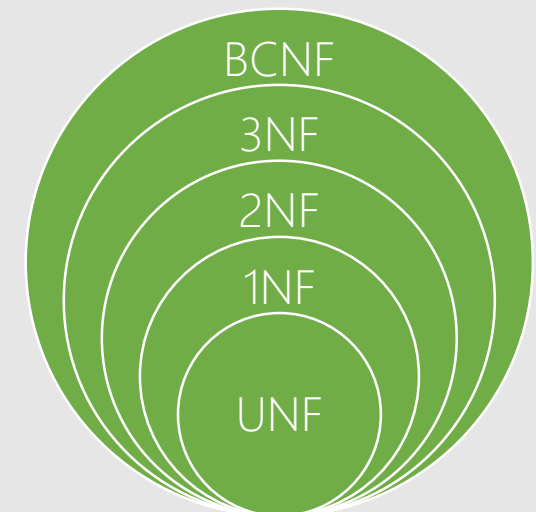| Interview			
ClientNo	InterviewDate	InterviewTime	StaffNo
76	13-May-02	10.30	05
56	13-May-02	12.00	05
74	13-May-02	12.00	37
56	01-July-02	10.30	05

CK1: ClientNo, InterviewDate → InterviewTime, StaffNo, RoomNo

CK2: StaffNo, InterviewDate, InterviewTime → ClientNo, RoomNo

CK3: RoomNo, InterviewDate, InterviewTime → ClientNo, StaffNo

InterviewDate	StaffNo	RoomNo
13-May-02	05	G101
13-May-02	05	G101
13-May-02	37	G102
01-July-02	05	G102

StaffNo, InterviewDate → RoomNo

Denormalization