

Shifting Gears:

From ***Introductory Concepts*** to ***Design Steps***

Or,

“Now that we’ve learned a bit about relational databases, their history, and conceptual principles, let’s look into how one goes about designing a database for an organization.”

Design Steps:

Collect all the information you can about the organization's data (i.e. "requirements")

- What do they call their entities and attributes?
- Data types and lengths
- How and when do they use each piece of data?
- Where does it come from? Where does it go to?
- Who uses it for what purpose?
- How many occurrences/instances of each entity do they deal with?

How?

1. Conduct interviews
2. Review documentation
3. Review current systems and processes
 - User data entry screens
 - Paper forms
 - Paper reports
 - Computer reports
 - Process flows

Design Steps:

Once data requirements are collected and documented, organize the data:

- Identify all Functional Dependencies
- Put the data into 3rd Normal Form
- Create a logical data model

Functional Dependency

Definition: (Ullman text, page 68)

“If the values of one or more attributes (A1, A2, A3, etc.) of a relation functionally determine the value of another attribute (B) of that relation, then we can say that (B) is functionally dependent on (A1, A2, A3, etc.)”

In other words,

If I know the value of an attribute (or set of attributes),
I can determine the value of another attribute.

Functional Dependency

Examples: (from the text)

The “Movies1” relation

<i>title</i>	<i>year</i>	<i>length</i>	<i>genre</i>	<i>studioName</i>	<i>starName</i>
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne's World	1992	95	comedy	Paramount	Mike Meyers

Functional Dependency

Follows Codd's relational rules, but it is bad design.

Why?

Let's look at functional dependencies.

Functional Dependency

<i>title</i>	<i>year</i>	<i>length</i>	<i>genre</i>	<i>studioName</i>	<i>starName</i>
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne's World	1992	95	comedy	Paramount	Mike Meyers

FD = title + year \rightarrow length + genre + studioName

If 2 tuples have the same values in title + year,
those 2 tuples will have the same values in length, genre
and studioName

TRUE

Functional Dependency

<i>title</i>	<i>year</i>	<i>length</i>	<i>genre</i>	<i>studioName</i>	<i>starName</i>
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne's World	1992	95	comedy	Paramount	Mike Meyers

FD = title + year \rightarrow starName

If 2 tuples have the same values in title + year,
those 2 tuples will have the same values in starName

FALSE

Functional Dependency

A formal definition of a Key

1. “A set of attributes (A1, A2, A3,etc.) is a KEY for a relation if that set of attributes functionally determines all the other attributes of the relation.

And

2. No subset of those attributes (A1, A2, A3,etc.) functionally determines all the other attributes of the relation.”

(FD doesn't differentiate Primary Key versus Candidate Key)

Functional Dependency

Apply some FD tests to this relation.

<i>title</i>	<i>year</i>	<i>length</i>	<i>genre</i>	<i>studioName</i>	<i>starName</i>
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne's World	1992	95	comedy	Paramount	Mike Meyers

title

title+year

title+starName

title+year+starName

- *The Manchurian Candidate* (1962 film), starring Frank Sinatra
- *The Manchurian Candidate* (2004 film), starring Denzel Washington
- *Blade Runner* (1982 film) starring Harrison Ford
- *Blade Runner* (2017 film) starring Harrison Ford

Functional Dependency

Let's practice

A relation of people in the U.S.

name, SSN, address, city, state, ZIP, area code, 7-digit
phone

what FD's are TRUE?

Functional Dependency

More practice

Project Code	Project Title	Project Manager	Project Budget	Employee No.	Employee Name	Department No.	Department Name	Hourly Rate
PC010	Pensions System	M Phillips	24500	S10001	A Smith	L004	IT	22.00
PC010	Pensions System	M Phillips	24500	S10030	L Jones	L023	Pensions	18.50
PC010	Pensions System	M Phillips	24500	S21010	P Lewis	L004	IT	21.00
PC045	Salaries System	H Martin	17400	S10010	B Jones	L004	IT	21.75
PC045	Salaries System	H Martin	17400	S10001	A Smith	L004	IT	18.00
PC045	Salaries System	H Martin	17400	S31002	T Gilbert	L028	Database	25.50
PC045	Salaries System	H Martin	17400	S13210	W Richards	L008	Salary	17.00
PC064	HR System	K Lewis	12250	S31002	T Gilbert	L028	Database	23.25
PC064	HR System	K Lewis	12250	S21010	P Lewis	L004	IT	17.50
PC064	HR System	K Lewis	12250	S10034	B James	L009	HR	16.50

Functional Dependency

Some FD rules

If I know a true FD for a relation, I can deduce other true FD's for that relation.

Functional Dependency

Transitive Rule

In a relation $R(A, B, C)$

if $A \rightarrow B$, and $B \rightarrow C$, then $A \rightarrow C$

Example:

Project Code	Project Title	Project Manager	Project Budget	Employee No.	Employee Name	Department No.	Department Name	Hourly Rate
PC010	Pensions System	M Phillips	24500	S10001	A Smith	L004	IT	22.00
PC010	Pensions System	M Phillips	24500	S10030	L Jones	L023	Pensions	18.50
PC010	Pensions System	M Phillips	24500	S21010	P Lewis	L004	IT	21.00
PC045	Salaries System	H Martin	17400	S10010	B Jones	L004	IT	21.75
PC045	Salaries System	H Martin	17400	S10001	A Smith	L004	IT	18.00
PC045	Salaries System	H Martin	17400	S31002	T Gilbert	L028	Database	25.50
PC045	Salaries System	H Martin	17400	S13210	W Richards	L008	Salary	17.00
PC064	HR System	K Lewis	12250	S31002	T Gilbert	L028	Database	23.25
PC064	HR System	K Lewis	12250	S21010	P Lewis	L004	IT	17.50
PC064	HR System	K Lewis	12250	S10034	B James	L009	HR	16.50

EmployeeNo \rightarrow EmployeeName

EmployeeName \rightarrow DepartmentNo, therefore

EmployeeNo \rightarrow DepartmentNo

Functional Dependency

Augmentation Rule

In a relation R(A, B, C)

if $A \rightarrow C$, then $A+B \rightarrow C+B$

Example:

Project Code	Project Title	Project Manager	Project Budget	Employee No.	Employee Name	Department No.	Department Name	Hourly Rate
PC010	Pensions System	M Phillips	24500	S10001	A Smith	L004	IT	22.00
PC010	Pensions System	M Phillips	24500	S10030	L Jones	L023	Pensions	18.50
PC010	Pensions System	M Phillips	24500	S21010	P Lewis	L004	IT	21.00
PC045	Salaries System	H Martin	17400	S10010	B Jones	L004	IT	21.75
PC045	Salaries System	H Martin	17400	S10001	A Smith	L004	IT	18.00
PC045	Salaries System	H Martin	17400	S31002	T Gilbert	L028	Database	25.50
PC045	Salaries System	H Martin	17400	S13210	W Richards	L008	Salary	17.00
PC064	HR System	K Lewis	12250	S31002	T Gilbert	L028	Database	23.25
PC064	HR System	K Lewis	12250	S21010	P Lewis	L004	IT	17.50
PC064	HR System	K Lewis	12250	S10034	B James	L009	HR	16.50

EmployeeNo + DepartmentNo \rightarrow EmployeeName + DepartmentNo

Functional Dependency

Combining + Splitting Rule

In a relation $R(A, B, C, D)$

if $A+B \rightarrow C$, and $A+B \rightarrow D$, then $A+B \rightarrow C, D$

<i>title</i>	<i>year</i>	<i>length</i>	<i>genre</i>	<i>studioName</i>	<i>starName</i>
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne's World	1992	95	comedy	Paramount	Mike Meyers

$\text{title}+\text{year} \rightarrow \text{length}$

$\text{title}+\text{year} \rightarrow \text{genre}$

$\text{title}+\text{year} \rightarrow \text{studioName}$

$\text{title}+\text{year} \rightarrow \text{length, genre, studioName}$ (combine the right side)

Functional Dependency

Combining + Splitting Rule

<i>title</i>	<i>year</i>	<i>length</i>	<i>genre</i>	<i>studioName</i>	<i>starName</i>
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne's World	1992	95	comedy	Paramount	Mike Meyers

But these are false: (can't split the left side)

title → length

year → length

Functional Dependency

Trivial FDs – are assumed to be true.

the right side is a subset of the left side

<i>title</i>	<i>year</i>	<i>length</i>	<i>genre</i>	<i>studioName</i>	<i>starName</i>
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne's World	1992	95	comedy	Paramount	Mike Meyers

title+year \rightarrow title

title \rightarrow title

Functional Dependency

Closure

For a relation, we seek ALL TRUE functional dependency rules

A relation: $R(A, B, C, G, H, I)$

FD's =
 $A \twoheadrightarrow B$
 $A \twoheadrightarrow C$
 $C+G \twoheadrightarrow H$
 $C+G \twoheadrightarrow I$
 $B \twoheadrightarrow H$

Functional Dependency

Closure

1. $A \rightarrow H$ (transitive)
2. $C+G \rightarrow H, I$ (combining)
3. $A+G \rightarrow I$ (augmentation)
 - $A \rightarrow C$
 - $A+G \rightarrow C+G$
 - $C+G \rightarrow I$

FD's =

$A \rightarrow B$
$A \rightarrow C$
$C+G \rightarrow H$
$C+G \rightarrow I$
$B \rightarrow H$

So, the closure (full set of FD's) is all of these.

Functional Dependency

Why do we care about all this?

- If I am designing a database for an organization, I must consider all their data. I must organize that data into relations.
- I need to understand Functional Dependencies so I can determine keys for every relation.
- I need keys so that I can normalize the data.
- I must normalize the data to design the database properly.