Functional Dependencies

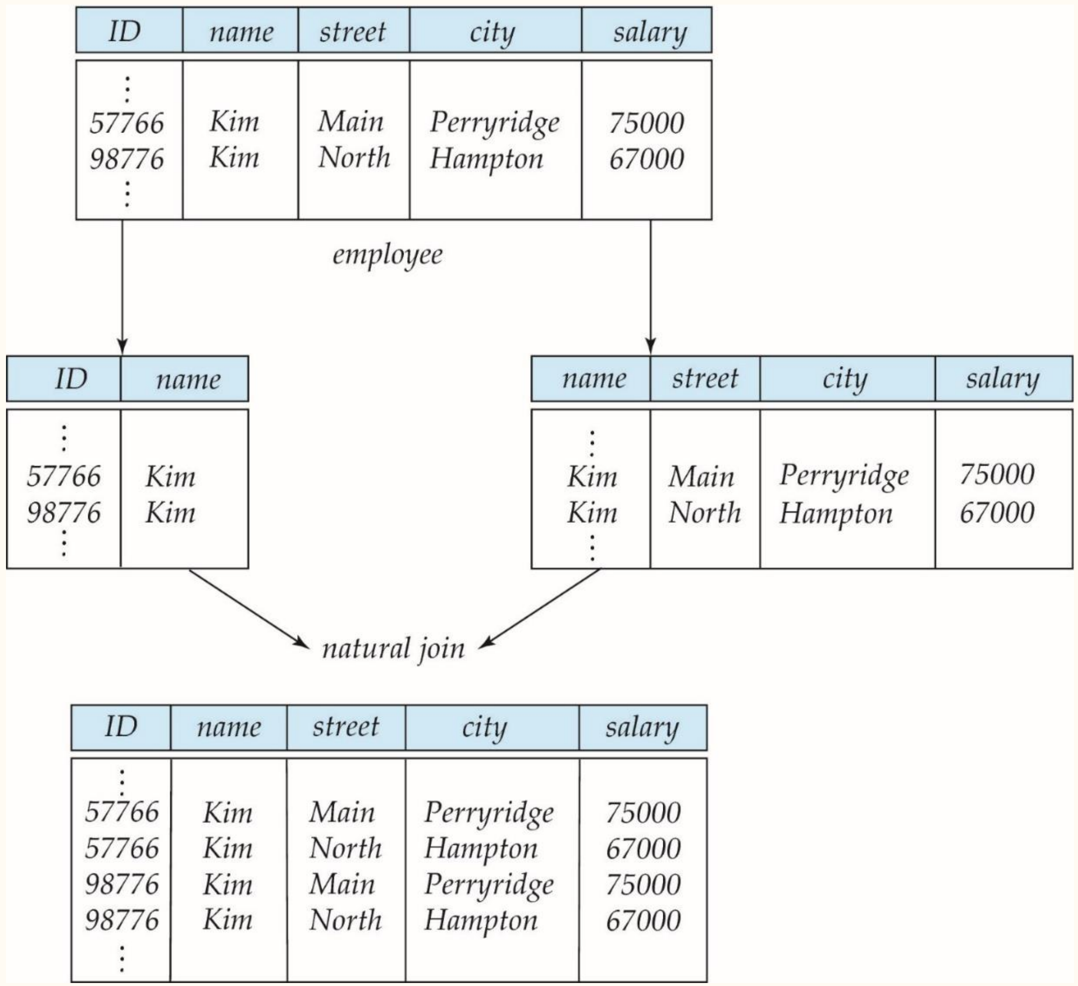
Why RDBS Design Again?

Suppose we have a table *inst_dept which contain information for both instructor* and *department*.

Result is possible repetition of information

ID	*name*	*salary*	*dept_name*	*building*	*budget*
22222	Einstein	95000	Physics	Watson	70000
12121	Wu	90000	Finance	Painter	120000
32343	El Said	60000	History	Painter	50000
45565	Katz	75000	Comp. Sci.	Taylor	100000
98345	Kim	80000	Elec. Eng.	Taylor	85000
76766	Crick	72000	Biology	Watson	90000
10101	Srinivasan	65000	Comp. Sci.	Taylor	100000
58583	Califieri	62000	History	Painter	50000
83821	Brandt	92000	Comp. Sci.	Taylor	100000
15151	Mozart	40000	Music	Packard	80000
33456	Gold	87000	Physics	Watson	70000
76543	Singh	80000	Finance	Painter	120000

Relational Design by Decomposition



Lossless-Join Decomposition

A Simple example of Lossy-Join (Non-Lossless Join):

Decomposition of $R = (A, B)$ *into* $R_1 = (A)$ *and* $R_2 = (B)$

A	B
α	1
α	2
β	1

r

A
α
β

$\Pi_A(r)$

B
1
2

$\Pi_{B(r)}$

A	B
α	1
α	2
β	1
β	2

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

Lossless-Join decomposition:

For all possible relation instance r on schema R

$$r = \Pi_{R1}(r) \bowtie \Pi_{R2}(r)$$

Goal - Devise a Theory for what is Good

We want to do two things:

1. Decide whether a particular relation R is in “good” form.
2. If a relation R is not in “good” form, decompose it into a set of relations {R1, R2, ..., Rn} such that
   - each relation is in good form
   - the decomposition is a lossless-join decomposition

Our theory/properties are defined based on functional dependencies.

Attribute Values can be Related

ID	*name*	*salary*	*dept_name*	*building*	*budget*
22222	Einstein	95000	Physics	Watson	70000
12121	Wu	90000	Finance	Painter	120000
32343	El Said	60000	History	Painter	50000
22222	Einstein	95000	Biology	Watson	90000
98345	Kim	80000	Elec. Eng.	Taylor	85000
76766	Crick	72000	Biology	Watson	90000
10101	Srinivasan	65000	Comp. Sci.	Taylor	100000
58583	Califieri	62000	History	Painter	50000
83821	Brandt	92000	Comp. Sci.	Taylor	100000
15151	Mozart	40000	Music	Packard	80000
33456	Gold	87000	Physics	Watson	70000
76543	Singh	80000	Finance	Painter	120000

Functional Dependencies

A functional dependency describes a **relation** between attributes

Whenever any two tuples t_1 and t_2 of r agree on one attribute α, they also agree on another attribute β.

$$t_1[\alpha] = t_2[\alpha] \Rightarrow t_1[\beta] = t_2[\beta]$$

This relation is denoted $\alpha \rightarrow \beta$.

Functional Dependencies

$ID \rightarrow Name, Depart\_name \rightarrow Building$

ID	*name*	*salary*	*dept_name*	*building*	*budget*
22222	Einstein	95000	Physics	Watson	70000
12121	Wu	90000	Finance	Painter	120000
32343	El Said	60000	History	Painter	50000
22222	Einstein	95000	Biology	Watson	90000
98345	Kim	80000	Elec. Eng.	Taylor	85000
76766	Crick	72000	Biology	Watson	90000
10101	Srinivasan	65000	Comp. Sci.	Taylor	100000
58583	Califieri	62000	History	Painter	50000
83821	Brandt	92000	Comp. Sci.	Taylor	100000
15151	Mozart	40000	Music	Packard	80000
33456	Gold	87000	Physics	Watson	70000
76543	Singh	80000	Finance	Painter	120000

Describes the **semantics** or **meaning of the attributes**

Functional Dependencies

The functional dependency

$X \rightarrow Y$ is true (holds)

if and only if

$$t_1[X] = t_2[Y] \Rightarrow t_1[X] = t_2[Y]$$

in relation R

ID	Name	Code	Grade
100	J	3550	A
200	X	3550	B
100	J	4540	B
100	J	4550	A

- Example: R = {ID, Name, Code, Grade}
  - ID → Name (OK)
  - ID → Grade (not OK), ID → Code (not OK)
  - ID, Name → Grade (not OK), ID, Code → Grade (OK)
  - ID, Name → Name (trivial)

Functional Dependencies: Test

Let's see if you understand (Test1)

F: $X \rightarrow Y$

X	Y
a	b
a	?

Functional Dependencies: Test

Let's see if you understand (Test2)

F: $X \rightarrow Y$

X	Y
a	b
?	b

c $\rightarrow$ b okay?

Functional Dependencies: Test

Let's see if you understand (Test3)

F: $X \rightarrow Y$

X	Y	
a	b	
c	b	?

Functional Dependencies: Test

Let's see if you understand (Test3)

F: $X \to Y$

X	Y
a	b
c	b ? possible

What does $X \to Y$ say about $Y \to X$?

Functional Dependencies: Test

Let's see if you understand (Test4,5)

$X, Y \rightarrow X$

$X \rightarrow X$

Note: Functional dependencies like these are trivial

Functional Dependencies: Test

Let's see if you understand (Test6)

Consider R (A , B) with the following instance r.

1	4
1	5
3	7

On this instance, $A \rightarrow B$ does NOT hold, but $B \rightarrow A$ does hold.

FD: relation between two sets

A functional dependency is a relation between two **sets** of attributes.

I.e., the value for a set of attributes determines the value for another set of attributes.

A functional dependency describes relation between two sets of attributes from are relation.

Examples:

$XY \rightarrow WZ$

$XW \rightarrow Z$

$Z \rightarrow XQ$

Functional Dependencies

A functional dependency is a **constraint** between two sets of attributes for all its **relation instances**.

A constraint means a constraint across all it's relation instances (extensions), that it must hold for all relation instances.

F is a set of FD specified on relation R must hold on all relation instances.

Constraint on all Relations

Example: *course* $\rightarrow$ *course_code in* Students

STUDENTS				
id	course	course_code	major	prof
1	Database	353	Comp Sci	Smith
2	Chem101	427	Chemistry	Turner
3	Database	353	Comp Sci	Clark

…

STUDENTS				
id	course	course_code	major	prof
1	Database	353	Comp Sci	Yu
4	Agile Dev	821	Comp Sci	Turner
		…		
5	Compiler	237	Comp Sci	Clark

Legal Extensions of R

Relation extensions $r(R)$ that satisfy the functional dependency constraints are called **legal relation states** (or **legal extensions**) of R.

Let $course \rightarrow course\_code$ *be the only FD for* Students

STUDENTS				
id	course	course_code	major	prof
1	Database	353	Comp Sci	Smith
2	Chem101	427	Chemistry	Turner
3	Database	353	Comp Sci	Clark

Legal

STUDENTS				
id	course	course_code	major	prof
1	Database	353	Comp Sci	Yu
4	Agile Dev	821	Comp Sci	Turner
5	Compiler	237	Comp Sci	Clark

Also legal

Violations in FD

But in practice, it's possible to have FD violations.

STUDENTS				
id	course	course_code	major	prof
1	Database	353	Comp Sci	Smith
2	Chem101	427	Chemistry	Turner
3	Database	353	Comp Sci	Clark

⇩

STUDENTS				
id	course	course_code	major	prof
1	Database	353	Data Science	Smith
2	Chem101	427	Chemistry	Turner
3	Database	353	Comp Sci	Clark

where is it? Isn't it a constraint? How should it be enforced? how come I seen it yet?

Notation and Terminology

Let $\underline{X \rightarrow Y}$ be a functional dependency on <u>relation R</u>

We say that

- $X \rightarrow Y$ *holds* on R

We say that

- X *functionally determines* Y
- Y is *functionally dependent* on X

We say that

- X is *determinant* of the dependency
- Y is *dependent* of the dependency

OR

- X is *left-hand side* of the dependency
- Y is *right-hand side* of the dependency

Functional Dependencies

A *WORKS_ON* relation

- *Ssn* = social security number
- *Pnumber* = project number

Question:
What might be the FDs of *WORKS_ON*?

WORKS_ON

Ssn	Pnumber	Hours
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	Null

Functional Dependencies

A *EMPLOYEE* relation

- *SSn* = social security number
- *Bdate* = birthday
- *Dnumer* = department number

Question: What might be the FDs of *EMPLOYEE?*

EMPLOYEE

Ename	<u>Ssn</u>	Bdate	Address	Dnumber
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4
Wallace, Jennifer S.	987654321	1941-06-20	291Berry, Bellaire, TX	4
Narayan, Ramesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1

Functional Dependencies

Question: What about Students? which columns might be dependent functionally?

STUDENTS				
ID		Major	Prof	Grade
1	237-4539	Comp Sci	Smith	A
2	427-7390	Chemistry	Turner	B
3	237-4539	Comp Sci	Clark	B
4	388-5183	Physics	James	A
5	371-6259	Decision Sci	Cook	C
6	823-7293	Mathematics	Lamb	B
7	823-7293	Mathematics	Bond	UN
8	237- 4539	Comp Sci	Cross	UN
9	839-0827	English	Broes	C

Functional Dependencies

Example: R = {ID, Name, Code, Grade}

r(R) Instance A

ID	Name	Code	Grade
100	J	3550	A
200	X	3550	B
100	J	4540	B
100	J	4550	A

- ID → Name (OK),
- ID → Grade (not OK),
- ID → Code (not OK),
- ID, Name → Grade (not OK),
- ID, Code → Grade (OK).

r(R) Instance B

ID	Name	Code	Grade
100	J	3550	A
200	X	3550	B
100	J	4540	A
100	J	4550	A

- ID → Name (OK)
- ID → Grade (OK),
- ID → Code (not OK)
- ID, Name → Grade (not OK),
- ID, Code → Grade (OK).

Functional Dependencies

Example: R = {ID, Name, Code, Grade}

r(R) Instance A

ID	Name	Code	Grade
100	J	3550	A
200	X	3550	B
100	J	4540	B
100	J	4550	A

- ID → Name (OK),
- ID → Grade (not OK),
- ID → Code (not OK),
- ID, Name → Grade (not OK),
- ID, Code → Grade (OK).

r(R) Instance B

ID	Name	Code	Grade
100	J	3550	A
200	X	3550	B
100	J	4540	A
100	J	4550	A

- ID → Name (OK)
- ID → Grade (OK),
- ID → Code (not OK)
- ID, Name → Grade (not OK),
- ID, Code → Grade (OK).

Important: You can't infer FD's from a relations instances

Functional Dependencies

Functional dependencies exist to:

- specify the semantics between attributes
  - semantics of a relation should be kept across its extensions, yes?
- specify constraints on a relational schema
- this semantics is not captured by ER
  - hasn't mentioned anything about this kind of relationship about attributes

Designing FDs

FD cannot be inferred automatically from a given relation extension r.

So given a relation, where do its FDs come from? Where do we find it?

Deciding the FDs of a table is part of a design decision.

- Defined explicitly by someone who knows the semantics of the attributes of R.

Designing FDs

Assume we need to define the FDs of this relation

STUDENTS					
ID	Course	Phone	Major	Prof	Grade

What can we know about the columns?

Could each ID have a unique phone number and major?

Which Columns are Related?

STUDENTS					
ID	Course	Phone	Major	Prof	Grade

Every ID has a unique phone number and major?

- We can say $\{ID\} \rightarrow \{Phone, Major\}$

Other relations between columns:

- Every course has a unique professor $\{Course\} \rightarrow \{Prof\}$
- Every ID and course has a unique grade $\{ID, Course\} \rightarrow \{Grade\}$

Whenever the semantics of two sets of attributes in R indicate that a functional dependency should hold, we specify the dependency as a constraint.

Final Notations

We may denote the attributes sets with/without curly brackets

- With curly brackets, attributes are comma separated
- {X,Y} = XY

The order of the attribute sets doesn’t matter

- ZY = YZ
- {Z,Y} = {Y, Z}

Test

(Q1) What is a functional dependency?

(Q2) What could decide the functional dependencies that hold among the attributes of a relation schema?

(Q3) Why can we not infer a functional dependency automatically from a particular relation state?

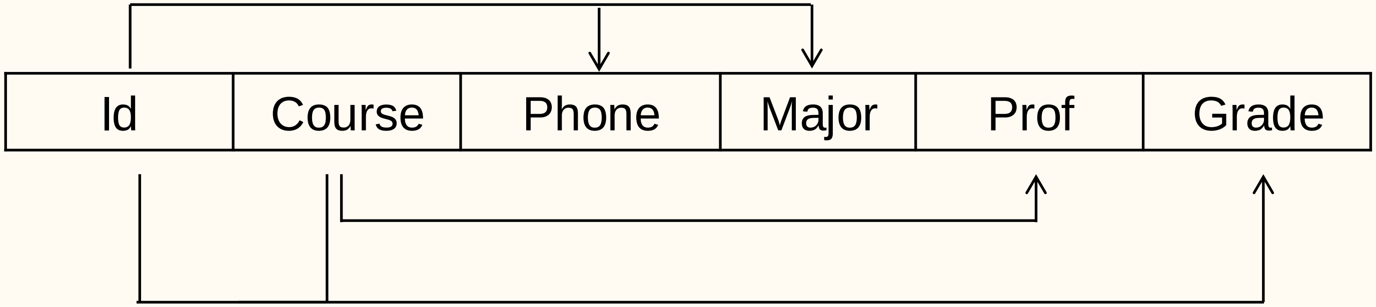
(Q4) Are there always functional dependencies in any relation?

Dependency Diagram

Each *horizontal line* represents a FD

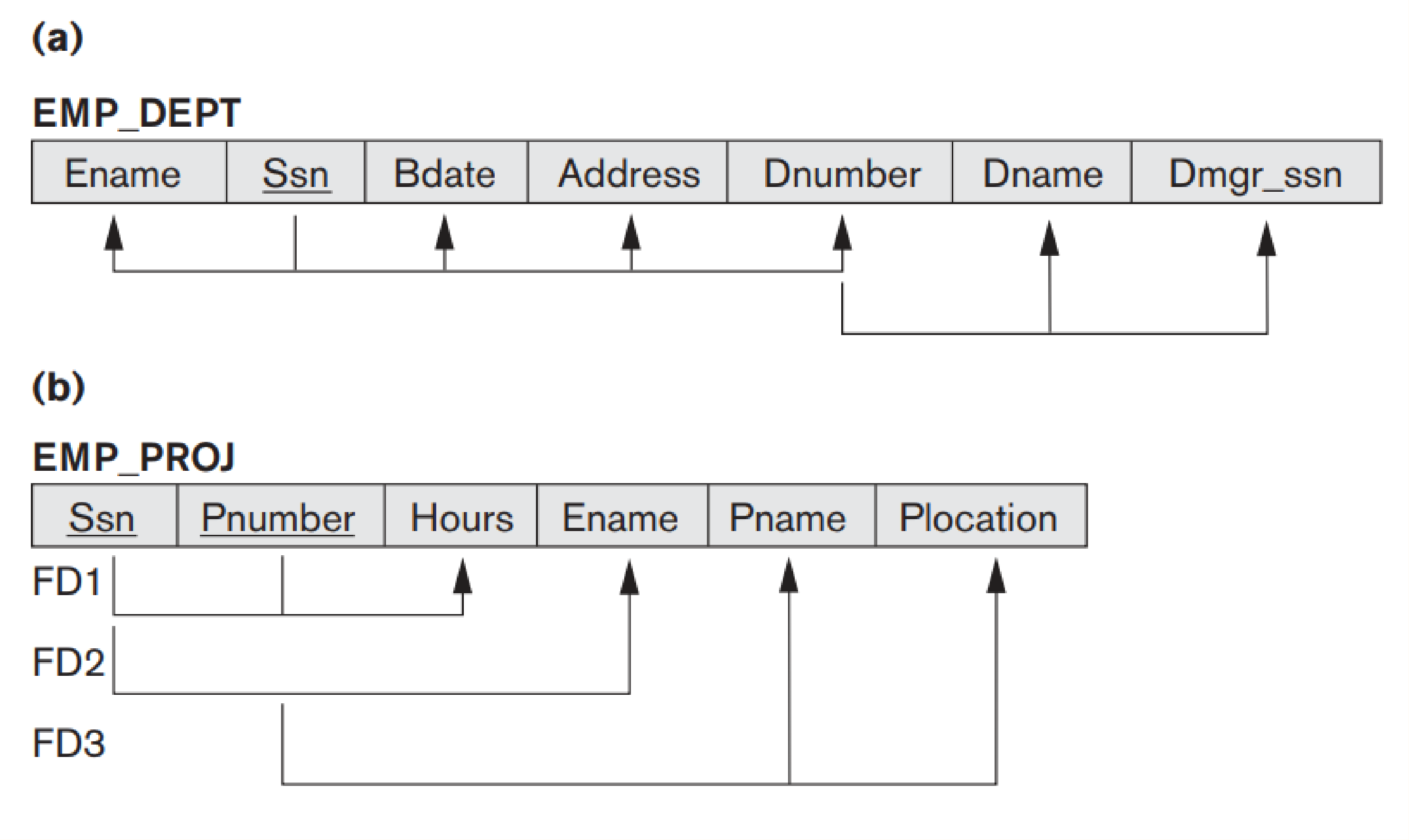
- Left-hand side attr. connected by vertical lines to the line,
- Right-hand side attr. connected by vertical lines <u>with arrows</u>
  - Arrow pointing toward the attributes

Dependency diagram from previous example.



Dependency Diagram (Cont.)

Some more examples of dependency diagrams.



Infering other FD's

Infering other FDs

$A \to B$ and $B \to C$, what do we know about $A \to C$?

Given $A \rightarrow B$ and $B \rightarrow C$ on relation R,

We know $A \rightarrow C$ holds on R, given A determines B, and B determines C.

There may be additionally functional dependencies that also hold on R!

Infering Other FDs

It's true that given a set F of functional dependencies, there are other functional dependencies that are logically implied by F.

$F \models X \rightarrow Y$

Denotes that set of FDs F infers $X \rightarrow Y$ if all relation instances satisfying F also satisfies $X \rightarrow Y$.

Example:
$F = A \rightarrow B, B \rightarrow C,$
$F \models A \rightarrow C$

Usually, the schema designer will only specify the functional dependencies that are semantically obvious.

Inference Rules

These are the inference rules for functional dependencies

- Rule 1 (reflexivity)
  - if $\beta \subseteq \alpha$, then $\alpha \to \beta$
- Rule 2 (augmentation)
  - if $\alpha \to \beta$, then $\gamma\,\alpha \to \gamma\,\beta$
- Rule 3 (transitivity)
  - if $\alpha \to \beta$, and $\beta \to \gamma$, then $\alpha \to \gamma$
- Where α, β, γ are all (nonempty) sets of attributes

The above are the primary rules/axioms from **Armstrong's Axioms**

Practice

$R = (A, B, C, G, H, I)$
$F = \{ A \rightarrow B, A \rightarrow C, CG \rightarrow H, CG \rightarrow I, B \rightarrow H \}$

These FDs can be inferred/ deduced.

$A \rightarrow H$

$AG \rightarrow I$

$CG \rightarrow HI$

Practice

$R = (A, B, C, G, H, I)$
$F = \{ A \to B, A \to C, CG \to H, CG \to I, B \to H \}$

These FDs can be inferred/ deduced.

$A \to H$

$AG \to I$

$CG \to HI$

But are they part of F?

(Solutions)

$R = (A, B, C, G, H, I)$
$F = \{ A \to B, A \to C, CG \to H, CG \to I, B \to H \}$

$A \to H$

by transitivity from $A \to B$ and $B \to H$

$AG \to I$

by augmenting $A \to C$ to get $AG \to CG$
then transitivity with given $CG \to I$

$CG \to HI$

by augmenting $CG \to I$ to infer $CG \to CGI$,
then augmenting $CG \to H$ to infer $CGI \to HI$,
followed up by a transitivity

Note: We denote by *F* the set of functional dependencies that are specified on relation schema *R*.

Functional dependencies logically implied other functional dependencies.

The set of all functional dependencies logically implied by F is the closure of F.

F+ denotes the **closure** of F

In general, given F, we can find F+ (the closure of F) by repeatedly applying Armstrong’s Axioms.

F+ is a superset of F.

Armstrong’s Axioms:

Armstrong’s Axioms are proven to be sound and complete

- Sound = generates only functional dependencies that hold
- Complete = generates all functional dependencies that hold

Armstrong's Axioms (Cont.)

Additional Rules we inferred from Armstrong's axioms.

- Rule 4 (**additivity**):
  - If $\alpha \rightarrow \beta$ holds and $\alpha \rightarrow \gamma$ holds, then $\alpha \rightarrow \beta\gamma$ holds
- Rule 5 (**projectivity**):
  - If $\alpha \rightarrow \beta\gamma$ holds, then $\alpha \rightarrow \beta$ holds and $\alpha \rightarrow \gamma$ holds
- Rule 6 (**pseudo-transitivity**):
  - If $\alpha \rightarrow \beta$ holds and $\gamma\beta \rightarrow \delta$ holds, then $\alpha\gamma \rightarrow \delta$ holds

Other names:
Additivity aka Union
Projectivity aka Decomposition

Proving Secondary Rules

Let’s try prove rule 5: projectivity

$\{X \rightarrow Y\,Z\} \models X \rightarrow Y$

Cheat Sheet

F1 (Reflexivity) If $X \supseteq Y$ then $X \rightarrow Y$.

F2 (Augmentation) $\{X \rightarrow Y\} \models XZ \rightarrow YZ$.

F3 (Transitivity) $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$.

(Solution)

Let's try prove rule 5: <u>projectivity</u>

$\{X \rightarrow Y\,Z\} \models X \rightarrow Y$

Step 1. $X \rightarrow Y\,Z$ (Given)

Step 2. $YZ \rightarrow Y$ (Reflexivity)

Step 3. $X \rightarrow Y$ (Transitivity of 1 and 2)

Cheat Sheet

F1 (Reflexivity) If $X \supseteq Y$ then $X \rightarrow Y$.

F2 (Augmentation) $\{X \rightarrow Y\} \models XZ \rightarrow Y\,Z$.

F3 (Transitivity) $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$.

Proving Secondary Rules

Let's prove rule 6: Pseudo-transitivity

$\{X \rightarrow Y\,,\, Y Z \rightarrow W\} \models XZ \rightarrow W$

Cheat Sheet

F1 (Reflexivity) If $X \supseteq Y$ then $X \rightarrow Y$.

F2 (Augmentation) $\{X \rightarrow Y\} \models XZ \rightarrow YZ$.

F3 (Transitivity) $\{X \rightarrow Y\,,\, Y \rightarrow Z\} \models X \rightarrow Z$.

(Solution)

Let's prove rule 6: Pseudo-transitivity

$\{X \rightarrow Y, YZ \rightarrow W\} \models XZ \rightarrow W$

Step 1. $X \rightarrow Y$ (Given)

Step 2. $XZ \rightarrow YZ$ (Augmentation of 1)

Step 3. $YZ \rightarrow W$ (Given)

Step 4. $XZ \rightarrow W$ (Transitivity, from 2 and 3)

Cheat Sheet
F1 (Reflexivity) If $X \supseteq Y$ then $X \rightarrow Y$.
F2 (Augmentation) $\{X \rightarrow Y\} \models XZ \rightarrow YZ$.
F3 (Transitivity) $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$.

Proving Secondary Rules

Let's prove rule 4: Additivity

$\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow Y Z$

Cheat Sheet

F1 (Reflexivity) If $X \supseteq Y$ then $X \rightarrow Y$.

F2 (Augmentation) $\{X \rightarrow Y\} \models XZ \rightarrow YZ$.

F3 (Transitivity) $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$.

(Solution)

Let's prove rule 4: <u>Additivity</u>

$\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow Y Z$

Step 1. $X \rightarrow Y$ (Given)

Step 2 . $XX \rightarrow XY$ (Augmentation of 1); that is, $X \rightarrow XY$

Step 3. $X \rightarrow Z$ (Given)

Step 4. $X Y \rightarrow Y Z$ (Augmentation of 2)

Step 5. $X \rightarrow Y Z$ (Transitivity, from 2 and 4)

Cheat Sheet

F1 (Reflexivity) If $X \supseteq Y$ then $X \rightarrow Y$.

F2 (Augmentation) $\{X \rightarrow Y\} \models XZ \rightarrow YZ$.

F3 (Transitivity) $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$.

Practice FD Inference

Cheat Sheet

F1 (Reflexivity) If $X \supseteq Y$ then $X \rightarrow Y$.

F2 (Augmentation) $\{X \rightarrow Y\} \models XZ \rightarrow YZ$.

F3 (Transitivity) $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$.

F4 (Additivity) $\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow YZ$.

F5 (Projectivity) $\{X \rightarrow YZ\} \models X \rightarrow Y$.

F6 (Pseudo-transitivity) $\{X \rightarrow Y, YZ \rightarrow W\} \models XZ \rightarrow W$.

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

Prove $A \rightarrow D$:

(Solution)

Cheat Sheet

F1 (Reflexivity) If $X \supseteq Y$ then $X \rightarrow Y$.

F2 (Augmentation) $\{X \rightarrow Y\} \models XZ \rightarrow YZ$.

F3 (Transitivity) $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$.

F4 (Additivity) $\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow YZ$.

F5 (Projectivity) $\{X \rightarrow YZ\} \models X \rightarrow Y$.

F6 (Pseudo-transitivity) $\{X \rightarrow Y, YZ \rightarrow W\} \models XZ \rightarrow W$.

Given F = {A → B, A → C, BC → D}
Prove A → D:

Step 1. A → B (Given)

Step 2. A → C (Given)

Step 3. A → BC (Additivity, from 1 and 2)

Step 4. BC → D (Given)

Step 5. A → D (Transitivity, from 3 and 4)

Closure of F

Definition. the set of all dependencies that can be inferred from F is called the **closure** of F.

F+ denotes the closure of F.

F+ includes dependencies in F.

Note: We typically reserve F to denote the set of functional dependencies that are specified on relation schema R.

Key Points on Closures

F denotes the set of FD's of a relation

F+ is the **closure** of F.

F+ is the *smallest set* of FD's that

- contains F, and
- is closed under Armstrong's axioms.

How do we check if a functional dependency can be inferred from FD's F (is a member of F+)?

The Procedure for Computing F+

To compute the closure of a set of functional dependencies F:

$F^+ = F$
repeat
  for each functional dependency f in F^+
    apply reflexivity and augmentation rules on f
    add the resulting functional dependencies to F^+
  for each pair of functional dependencies f_1 and f_2 in F^+
    if f_1 and f_2 can be combined using transitivity
      then add the resulting functional dependency to F^+
until F^+ does not change any further

The Procedure for Computing F+

F = { X → Y, Y → Z}

F+ = {XY → X, XY → Y, XY → Z, XZ → X, XZ → Y,
XZ → Z, XYZ → X, XYZ → Y, XYZ → Z, XY → XY,
XY → YZ, XY → XZ, …}

Checking Membership by F+

Given $F = \{ X \to Y, Y \to Z\}$

Question: Can $X \to Z$ be inferred or derived from the FDs in F?

How to do it? Check $X \to Z$ by computing F+?

Checking Membership by F+

Given F = { $X \rightarrow Y$, $Y \rightarrow Z$}

Question: Can $X \rightarrow Z$ be inferred or derived from the FDs in F?

How to do it? Check $X \rightarrow Z$ by computing F+?

F+ = {$XY \rightarrow X$, $XY \rightarrow Y$, $XY \rightarrow Z$, $XZ \rightarrow X$, $XZ \rightarrow Y$,
$XZ \rightarrow Z$, $XYZ \rightarrow X$, $XYZ \rightarrow Y$, $XYZ \rightarrow Z$, $XY \rightarrow XY$,
$XY \rightarrow YZ$, $XY \rightarrow XZ$, … , $X \rightarrow Z$, … }

Oh yes… $X \rightarrow Z$ is in the closure of F.

Checking Membership by F+

Given F = { $X \rightarrow Y$, $Y \rightarrow Z$}

Question: Can $X \rightarrow Z$ be inferred or derived from the FDs in F?

How to do it? Check $X \rightarrow Z$ by computing F+?

F+ = {$XY \rightarrow X$, $XY \rightarrow Y$, $XY \rightarrow Z$, $XZ \rightarrow X$, $XZ \rightarrow Y$,
$XZ \rightarrow Z$, $XYZ \rightarrow X$, $XYZ \rightarrow Y$, $XYZ \rightarrow Z$, $XY \rightarrow XY$,
$XY \rightarrow YZ$, $XY \rightarrow XZ$, … , $X \rightarrow Z$, … }

Oh yes… $X \rightarrow Z$ is in the closure of F.

Problem: In real life, it is impossible to specify all possible functional dependencies for a given situation. The size of F+ is always exponential size w.r.t |F|.

Closure of Attributes

Given F = { $X \rightarrow Y, Y \rightarrow Z$}

Question: How else to check if $X \rightarrow Z$ without computing F+ ?

Definition: Given a set of attributes α, define the ***closure*** of α **under** *F* (denoted by α^+) as the set of attributes that are functionally determined by α under *F*.

Realistically:
Narrow our attention to X, which is smaller than F.
Compute X+ instead of F+
Then check if Z is covered by X+

X+ is the largest set of attributes functionally determined by X.

Closure of Attribute Sets

Pseudocode to the closure of α under F

result := a;
while (changes to *result*) **do**
    for each $\beta \rightarrow \gamma$ **in** *F* **do**
      begin
        if $\beta \subseteq$ *result* **then** *result* := *result* $\cup\ \gamma$
      end

When no additional changes to result is possible, the final value of variable result is $\alpha+$

Example of Attribute Set Closure

$R = (A, B, C, G, H, I)$

$F = \{A \rightarrow B, A \rightarrow C, CG \rightarrow H,$
$CG \rightarrow I, B \rightarrow H\}$

Compute the closure of AG (AG+)

Cheat Sheet:

result := a;
while (changes to *result*) **do**
  for each $\beta \rightarrow \gamma$ **in** *F* **do**
    begin
      if $\beta \subseteq$ *result*
        then
          result := *result* $\cup\ \gamma$
      end

Algorithm to Compute X+

An **algorithm** for you to follow step by step

```
X := X;
change := true;
while change do
        begin
        change := false;
        for each FD W → Z in F do
                begin
                if (W ⊆ X+) and (Z ⊈ X+) then do
                        begin
                        X+ := X+ ∪ Z;
                        change := true;
                        end
                end
        end
```

Exercise

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

Practice: **Compute A+**

Cheat Sheet:

```
X+ := X;
change := true;
while change do
        begin
        change := false;
        for each FD W → Z in F do
          begin
          if (W ⊆ X+) and (Z ⊈ X+)
          then do
                 begin
                 X+ := X+ ∪ Z;
                 change := true;
                 end
          end
        end
```

(Solution)

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

Task: **Compute {A}+**

1st scan of F:

X+ := {A}
X+ := {A, B}
X+ := {A, B, C}

2nd scan of F:

X+ := {A, B, C, D }

3rd scan of F: no change, therefore, the algorithm terminates.

{A}+ := {A, B, C, D }

Cheat Sheet:

```
X+ := X;
change := true;
while change do
        begin
        change := false;
        for each FD W → Z in F do
          begin
          if (W ⊆ X+) and (Z ⊄ X+)
          then do
                  begin
                  X+ := X+ ∪ Z;
                  change := true;
                  end
          end
        end
```

FDs and Keys

Recall Exp. of Attribute Set Closure

R = (A, B, C, G, H, I)

F = {$A \rightarrow B$, $A \rightarrow C$, $CG \rightarrow H$, $CG \rightarrow I$, $B \rightarrow H$}

We know (AG)+ = ABCGHI

Observation: could AG a candidate key?

Is AG a super key?

Does $AG \rightarrow R$? == Is $(AG)^+ \supseteq R$

Is any subset of AG a super key?

Does $A \rightarrow R$? == Is $(A)^+ \supseteq R$

Does $G \rightarrow R$? == Is $(G)^+ \supseteq R$

Functional Dependencies (Cont.)

K is a super key for relation schema R if and only if $K \rightarrow R$

K is a candidate key for R if and only if

- $K \rightarrow R$, and
- for no $\alpha \subset K$, $\alpha \rightarrow R$

Example

Consider schema *STUDENT(zid, name, address)*

Where $zid \rightarrow name, address$

zid functionally determines all tuples in the table *STUDENT*

Notes:

Although key of a relation will always functionally determine every attributes in the relation.

The left-hand side of a dependency does not imply uniqueness.

Notes: functionally dependencies are a generalization of a concept of a key

Functional Dependencies (Cont.)

Functional dependencies allow us to express constraints that cannot be expressed using superkeys.

Consider the schema:

inst_dept (<u>*ID, dept_name,*</u> *name, salary, building, budget*) .

We can also express functional dependencies to hold:

$$dept\_name \rightarrow building$$

$$ID \rightarrow building$$

Procedurally Determine Keys

Motivation: use FDs to procedurally determine the keys of a relation.

Procedurally Determine Keys

How to compute a candidate key of a relation R based on the FD's belonging to R

Algorithm:

- *Step 1 : Assign a super-key of R in F to X.*
- *Step 2 : Iteratively remove attributes from X while retaining the property X+ = R till no reduction on X is possible.*
- *The remaining X is a key.*

Let's try an example

Practice

Step 1 : Assign a super-key of R in F to X.
Step 2 : Iteratively remove attributes from X while retaining the property $X^+ = R$ till no reduction on X is possible.
The remaining X is a key.

Given:
R = {A, B, C, D}
F = { A $\rightarrow$ B, BC $\rightarrow$ D, A $\rightarrow$ C }

(Solution)

Given:
$R = \{A, B, C, D\}$
$F = \{ A \rightarrow B, BC \rightarrow D, A \rightarrow C \}$

Let $X = \{A, B, C\}$
($\{A, B, C, D\}$ is also a super key)

A cannot be removed
because $\{BC\}+ = \{B, C, D\} \neq R$

B can be removed
because $\{AC\}+ = \{A, B, C, D\} = R$

We remove B from X
and update X to be $\{ A, C\}$

C can be further removed
because $\{A\}+ = \{A, B, C, D\}$

We remove C from X
and update X to be $\{ A\}$

Step 1 : Assign a super-key of R in F to X.
Step 2 : Iteratively remove attributes from X while retaining the property $X^+ = R$ till no reduction on X is possible.
The remaining X is a key.

Compute all Candidate Keys

Given a relational schema R and a set of functional dependencies F on R, find all the possible ways we can identify a row.

Note: we know how to compute one candidate key already.

Compute all Candidate Keys

The algorithm to compute all the candidate keys is as follows:

```
T := ∅
Main:
    X := S
    remove := true
    While remove do
        For each attribute A ∈ X
        Compute {X-A}+ with respect to F
        If {X-A}+ contains all attributes of R then
            X := X – {A}
        Else
            remove := false
    T :=T ∪ X
```

Repeat until no available S can be found.

Finally, T contains all the candidate keys.

Compute all Candidate Keys

Given relation R(A, B, C, D, E)

with set of FDs $\{A \rightarrow B, BC \rightarrow A, D \rightarrow E\}$

Find all the candidate keys for relation R

Compute all Candidate Keys

```
T := ∅
Main:
        X := S
        remove := true
        While remove do
                For each attribute A ∈ X
                Compute {X-A}+ with respect to F
                If {X-A}+ contains all attributes of R then
                        X := X – {A}
                Else
                        remove := false
        T :=T ∪ X
```

Relation R(A, B, C, D, E)
with set of FDs $\{A \rightarrow B, BC \rightarrow A, D \rightarrow E\}$

Compute all Candidate Keys

```
T := ∅
Main:
    X := S
    remove := true
    While remove do
        For each attribute A ∈ X
        Compute {X-A}+ with respect to F
        If {X-A}+ contains all attributes of R then
            X := X – {A}
        Else
            remove := false
    T :=T ∪ X
```

Relation R(A, B, C, D, E)
with set of FDs $\{A \rightarrow B,\ BC \rightarrow A,\ D \rightarrow E\}$

Compute all Candidate Keys

```
T := ∅
Main:
        X := S
        remove := true
        While remove do
                For each attribute A ∈ X
                Compute {X-A}+ with respect to F
                If {X-A}+ contains all attributes of R then
                        X := X – {A}
                Else
                        remove := false
        T :=T ∪ X
```

Relation R(A, B, C, D, E)
with set of FDs $\{A \rightarrow B, BC \rightarrow A, D \rightarrow E\}$

(Solution)

Step 1:
Let X := {A, B, C, D}

Step 2:
Try to remove A
{B, C, D}+ = {A, B, C, D, E}
Thus X := {B, C, D}

Steps 3,4,5:
Attempts to remove B, C, D separately

{C, D}+ = {C, D, E}

{B, D}+ = {B, D, E}

{B, C}+ = {A, B, C}

None can be removed
So {<u>B, C, D</u>} is a candidate key and add to T

(Solution)

Step 6:
Find another super key
Let X := {A, C, D}

Step 7,8,9:
Attempts to remove A, C, D separately

{C, D}+ = {C, D, E}

{A, D}+ = {A, B, D, E}

{A, C}+ = {A, B, C}

None cannot be removed
So, {A, C, D} is another candidate key and add to T

(Solution)

Step 10:
Cannot find any other super keys,

Conclusion: candidate keys are *{B, C, D}* and {A, C, D}

Lecture Learning Outcomes

Take aways

- Functional Dependencies
- Armstrong’s axioms
- Given a FD, check if the FD can be derived from a given set of FD
- How to compute one candidate key
- How to compute all candidate keys

Next Lecture

- Normal Forms

Acknowledgement: Several slides in this lecture were inspired by the textbook slides made by Siberschatz, Korth and Sudarshan from Database System Concepts. 6th Ed.