Functional Dependencies

Roadmap

- We extracted relations/tables for our design.
- How good is our relational design?
- Can we reason about it?
- Is there any methodology that helps to obtain a good design or correcting possible problems with our current design?
  - E.g., redundancy

Facts

- Commonly we know certain facts about dependencies and/or constraints existing in our application scenario:
  - You cannot have two classes on the same room and at the same time/day in the campus!
- We will formalize such dependencies / constraints and utilize them to get a 'good' design.
  - How? By utilizing them to decompose relational schemas into "simpler" ones.

CSC43 N Koudas 3

Functional Dependencies (FD's)

$X \rightarrow A$ = assertion about a relation R that whenever two tuples agree on all the attributes of X, (the tuples have the same values in their respective components for each of these attributes) then they must also agree on attribute A.

Note: FD is a constraint, its ok if it does not hold on an schema instance

Examples??

CSC43 N Koudas 4

Example

Movies(title, year, length, film-type, studio-Name, star-Name)

title	*year*	*length*	*film-type*	*studio-Name*	*star-Name*
Star Wars	1977	124	color	Fox	Carrie Fisher
Star Wars	1977	124	color	Fox	Mark Hamill
Star Wars	2002	142	color	Fox	Natalie Portman
Wayne's	1992	104	color	Paramount	Dana Carvey
Wayne's	1992	104	color	Paramount	Mike Meyers

Reasonable FD's to assert:

1. *title*, *year* → *length*
2. *title*, *year* → *film-type*
3. *title*, *year* → *studio-Name*

What about the FD: *title*, *year* → *star-Name* !?!??

CSC43 N Koudas 5

Functional Dependencies (cont.)

- Shorthand: combine FD's with common left side by concatenating their right sides.
  Example:
  Given *title*, *year* → *length* and *title*, *year* → *film-type*
  we can have *title*, *year* → *length, film-type*
- Sometimes, several attributes jointly determine another attribute, although neither does by itself.
  Example:
  We know that *title*, *year* → *length* holds…
  It is easy to see that *title* → *length* does not hold
- Rules (combining/splitting rules):
  - A1A2 -> B1, A1A2 -> B2, A1A2 -> B3 <=> A1A2 -> B1 B2 B3

CSC43 N Koudas 6

Functional Depend. (cont')

- Assume A1A2..An -> B1B2..Bm
  - Trivial if the B's are a subset of the A's
  - Nontrivial if at least one of the B's is not among the A's.
  - Completely non trivial if none of the B's is also one of the A's.
- You can remove from the right side all attributes that appear on the left:
  - Trivial dependency rule

CSC43 N Koudas 7

Keys of Relations

Define key *K* for relation *R* w.r.t. FD's. Set *K* of attrs is key if:

1. *K* → all attributes of *R*. **(Uniqueness)**
2. For no proper subset of *K* is (1) true. **(Minimality)**
   If *K* at least satisfies (1), then *K* is a *superkey*.

AB -> CDE
A -> BCDE

A few words about notation…

- Pick one key; underline key attributes in the relation schema.
- *A, B, C* etc., represent single attributes;
  X, Y, Z etc., represent sets of attributes
- No set formers in FD's… use *ABC* instead of {*A*, *B*, *C*}

CSC43 N Koudas 8

Example

Movies(title, year, length, film-type, studio-Name, star-Name)

- {*title*, *year*, *star-Name*} → all attributes. Thus, it 's a superkey.
- *title* → *year* is false, so *title* is not a superkey.
- *star-Name* → *title* also false, so *star-Name* not a superkey.
- Thus, {*title*, *year*, *star-Name*} is a key.
- No other keys in this example.

CSC43 N Koudas 9

Who Determines Keys/FD's?

- We could assert a key *K*.
  - Then the only FD's asserted are that *K* → *A* for every attribute *A*.
  - No surprise: *K* is then the only key for those FD's, according to the formal definition of "key."
- Or, we could assert some FD's and *deduce* one or more keys by the formal definition.
  - E/R diagram implies FD's by key declarations and many-one relationship declarations.
- Rule of thumb: FD's either come from keys, many-1 relationships, or from known facts.
  - *E.g.*, "no two courses can meet in the same room at the same time" yields *rooms, times* → *courses*.

CSC43 N Koudas 10

Functional Dependencies and Many-One Relationships

- Consider $R(A_1,\ldots, A_n)$ and X is a key
  Then $X \rightarrow Y$ for any attributes Y in $A_1,\ldots, A_n$
  even if they overlap with X
- Suppose R is used to represent a many → one relationship:
  $E1$ entity set → $E2$ entity set
  where X key for $E1$, Y key for $E2$,
  Then, $X \rightarrow Y$ holds,
  And $Y \rightarrow X$ does not hold unless the relationship is one-one.
- What about many-many relationships?

CSC43 N Koudas 11

Inferring FD's

When we talk about improving relational designs, we often need to ask:

"does this FD hold in this relation?"

More formally, given FD's

$$X_1 \rightarrow A_1, X_2 \rightarrow A_2,\ldots, X_n \rightarrow A_n,$$

does FD $Y \rightarrow B$ necessarily hold in the same relation?

Try the following: Start by assuming two tuples agree in Y. Use given FD's to infer other attributes on which they must agree. If B is among them, then yes, else no.

CSC43 N Koudas 12

Example

Consider $R(A, B, C)$ and FD's $A \rightarrow B$ and $B \rightarrow C$

What about: $A \rightarrow C$?

Consider two tuples (a, b_1, c_1), (a, b_2, c_2)

Then, due to $A \rightarrow B$, we have (a, b, c_1), (a, b, c_2)

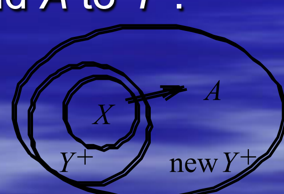
Since $B \rightarrow C$ we have (a, b, c), (a, b, c)

Thus, $A \rightarrow C$

Algorithm

The *closure* Y^+ of attribute set Y is the set of attributes functionally determined by Y

- Basis: $Y^+ := Y$.
- Induction: If $X \subseteq Y^+$, and $X \rightarrow A$ is a given FD, then add A to Y^+.

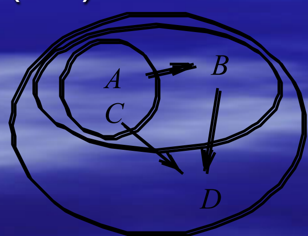


- Terminate: when Y^+ cannot be changed.

Example

Initial set of FDs: $A \rightarrow B$, $BC \rightarrow D$.

- $A^+ = AB$.
- $C^+=C$.
- $(AC)^+ = ABCD$.



Another Example

- Set of FD's in *ABCGHI*:

  $A \rightarrow B$
  $A \rightarrow C$
  $CG \rightarrow H$
  $CG \rightarrow I$
  $B \rightarrow H$
- Compute $(CG)^+$, $(BG)^+$, $(AG)^+$
  - $(CG)^+$ = CGHI
  - $(BG)^+$ = BGH
  - $(AG)^+$ = ABCGHI
- How can I use this algorithm to determine whether a set of attributes forms a key? What is the complexity?

Armstrong's axioms

- Armstrong's Axioms:
  - *Reflexivity:* If $Y \subseteq X$, then $X \rightarrow Y$ (trivial FD's)
  - *Augmentation:* If $X \rightarrow Y$, then $XZ \rightarrow YZ$ for any Z
  - *Transitivity:* If $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow Z$
- These are *sound* and *complete* inference rules for FDs!

Example:
Initial set of FDs: $A \rightarrow B$, $BC \rightarrow D$. Show that $AC \rightarrow D$
Solution:
$A \rightarrow B$, then $AC \rightarrow BC$ (Augmentation).
$AC \rightarrow BC$ and $BC \rightarrow D$, then $AC \rightarrow D$ (Transitivity)

Given Versus Implied FD's

Typically, we state a few FD's that are known to hold for a relation *R*.

- Other FD's may follow logically from the given FD's; these are *implied* FD's.
- We are free to choose any *basis* for the FD's of *R* – a set of FD's that imply all the FD's that hold for *R*.
- Minimal basis (no subset a basis)

Projecting FD's

Motivation: Suppose we have a relation *ABCD* with some FD's *F*. If we decide to decompose *ABCD* into *ABC* and *AD*, what are the FD's for *ABC*, *AD*?

Closure of a set of functional dependencies F.

- Example: $F = AB \rightarrow C, C \rightarrow D, D \rightarrow A$. It looks like just $AB \rightarrow C$ holds in *ABC*, but in fact $C \rightarrow A$ follows from *F* and applies to relation *ABC*.
- Get closure for each subset of ABC wrt F
- Problem is exponential in worst case.

CSC43 N Koudas 19

Algorithm

Assume relation *R*, set *F* of FD's and set $S \subseteq R$

- Compute X^+ for each $X \subseteq S$
  - But skip the cases $X = \emptyset$, X = all attributes.
  - Add $X \rightarrow A$ for each A in $X^+ - X$.
- If X^+ is all attributes, then there is no point in computing closure of supersets of X.
- Finally, project the FD's by selecting only those FD's that involve only attributes of *S*

CSC43 N Koudas 20

Example

$F = AB \to C, C \to D, D \to A$. What FD's follow from F?

- $A^+ = A$; $B^+ = B$ (nothing).
- $C^+ = ACD$ (add $C \to A$).
- $D^+ = AD$ (nothing new).
- $(AB)^+ = ABCD$ (add $AB \to D$; skip all supersets of AB).
- $(BC)^+ = ABCD$ (nothing new; skip all supersets of BC).
- $(BD)^+ = ABCD$ (add $BD \to C$; skip all supersets of BD).
- $(AC)^+ = ACD$; $(AD)^+ = AD$; $(CD)^+ = ACD$ (nothing new).
- $(ACD)^+ = ACD$ (nothing new).
- All other sets contain AB, BC, or BD, so skip.
- Thus, the only interesting FD's that follow from F are: $C \to A$, $AB \to D$, $BD \to C$.

CSC43 N Koudas 21

Another Example

In ABC with FD's $A \to B$, $B \to C$, project onto AC

1. $A^+ = ABC$; yields $A \to B$, $A \to C$.
2. $B^+ = BC$; yields $B \to C$.
3. $C^+ = C$ and $BC^+ = BC$; adds nothing.

- Resulting FD's: $A \to B$, $A \to C$, $B \to C$.
- Projection onto AC: $A \to C$.

CSC43 N Koudas 22

Design Anomalies and Normalization

Design Anomalies

Movies(title, year, length, film-type, studio-Name, star-Name)

title	*year*	*length*	*film-type*	*studio-Name*	*star-Name*
Star Wars	1977	124	color	???	Carrie Fisher
Star Wars	1977	???	color	Fox	Mark Hamill
Star Wars	1977	124	???	Fox	Harrison Ford
Wayne's	1992	104	color	Paramount	Mike Meyers

1. *title*, *year* → *length*
2. *title*, *year* → *film-type*
3. *title*, *year* → *studio-Name*

- ???'s are redundant, since we can figure them out from the FD's.
- Update anomalies: If Star Wars is actually 125 minutes long, will we change *length* in each of its tuples?
- Deletion anomalies: If we delete Mike Meyers, we lose the info for Wayne's World.

Eliminate Anomalies

- Decomposition
  - Intuitively break the schema into two or more pieces.
  - Premise
    - anomalies do not exist anymore
    - We can recover the original information by manipulating the pieces.
      - Example: R(A,B,C,D) = {(a1 b1 c1 d1),(a1 b1 c2 d2) (a1 b2 c1 d3)} decomposed to R1(A,B) and R2(C,D)
      - R1(A,B) = {(a1 b1),(a1,b2)} R2(C,D) = {(c1,d1),(c2,d2),(c1,d3)}
      - What is wrong?

CSC43 N Koudas 25

Decomposition of a Relation Scheme

- Suppose that relation R contains attributes $A_1 \ldots A_n$
  A *decomposition* of R consists of replacing R by two (or more) relations such that:
  - Each new relation scheme contains a subset of the attributes of R (and no attributes that do not appear in R), and
  - Every attribute of R appears as an attribute in at least one of the new relations.
- Intuitively, decomposing R means we will store instances of the relation schemes produced by the decomposition, instead of instances of R.

CSC43 N Koudas 26

Normal Forms

- Given relation R, the first question to ask is whether any refinement/decomposition is needed!
- If a relation is in a certain *normal form* (BCNF, 3NF etc.), it is known that certain kinds of problems are avoided/minimized. This can be used to help us decide whether decomposing the relation will help.
- Role of FD's in detecting redundancy:
  - Consider a relation R with 3 attributes, ABC.
    - No FD's hold: There is no redundancy here.
    - Given A → B: Several tuples could have the same A value, and if so, they'll all have the same B value!

CSC43 N Koudas 27

Boyce-Codd Normal Form (BCNF)

Formally, *R* is in BCNF if for every nontrivial FD for *R*, say *X* → *A*, then *X* is a superkey.

Advantages of BCNF:
1. Guarantees no redundancy due to FD's.
2. Guarantees no update anomalies
3. Guarantees no deletion anomalies

CSC43 N Koudas 28

Example

Movies(title, year, length, film-type, studio-Name, star-Name)

title	*year*	*length*	*film-type*	*studio-Name*	*star-Name*
Star Wars	1977	124	color	Fox	Carrie Fisher
Star Wars	1977	124	color	Fox	Mark Hamill
Star Wars	1977	124	color	Fox	Harrison Ford
Wayne's	1992	104	color	Paramount	Mike Meyers

Functional dependencies:

1. *title*, *year* → *length*
2. *title*, *year* → *film-type*
3. *title*, *year* → *studio-Name*

Each of the given FD's is a BCNF violation… Why?
Remember that the key is {*title*, *year*, *star-Name*}

CSC43 N Koudas 29

Decomposition to Reach BCNF

The setting: relation R, set F of FD's.

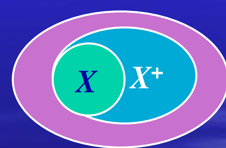
Suppose relation R has BCNF violation $X \rightarrow B$

- We need only look among FD's of F for a BCNF violation, not those that follow from F.
- Proof: If $Y \rightarrow A$ is a BCNF violation and follows from F, then the computation of Y^+ used at least one FD $X \rightarrow B$ from F.
  - X must be a subset of Y.
  - Thus, if Y is not a superkey, X cannot be a superkey either, and $X \rightarrow B$ is also a BCNF violation.

CSC43 N Koudas 30

Decomposition to Reach BCNF (cont.)

1. Compute X^+.
   – Cannot be all attributes – why?
2. Decompose R into X^+ and $(R–X^+) \cup X$.



3. Find the FD's for the decomposed relations. Project the FD's from F calculate all consequents of F that involve only attributes from X^+ or only from $(R\text{-}X^+) \cup X$.

Example

R = *Movies*(*title*, *year*, *length*, *film-type*, *studio-Name*, *star-Name*)
F =
1. *title*, *year* → *length*
2. *title*, *year* → *film-type*
3. *title*, *year* → *studio-Name*

Pick BCNF violation, e.g.,
title, *year* → *length*, *film-type*, *studio-Name*

- Find the closure
  (*title*, *year*)$^+$… = {*title*, *year* , *length*, *film-type*, *studio-Name*}
- Create decomposed relations:
  $Movies_1$(<u>*title*</u>, <u>*year*</u>, *length*, *film-type*, *studio-Name*)
  $Movies_2$(<u>*title*</u>, <u>*year*</u>, <u>*star-Name*</u>)

Another Example

R = Movies(title, year, studio-Name, president, presAddr)

F =

1. *title, year → studio-Name*
2. *studio-Name → president*
3. *president → presAddr*

Pick BCNF violation, e.g., *studio-Name → president*

- Find the closure *(studio-Name)*$^+$… = *{studio-Name, president, presAddr}*
- Create decomposed relations:
  Movies$_1$(<u>title</u>, <u>year</u>, studio-Name)
  Movies$_2$(<u>studio-Name</u>, president, presAddr)
- Projected FD's (skipping a lot of work that leads nowhere interesting):
  – For *Movies$_1$*: *title, year → studio-Name*
  – For *Movies$_2$*: *studio-Name → president* and *president → presAddr*

CSC43 N Koudas 33

Another Example (cont.)

(Continuing…)

- Decomposed relations:
  Movies$_1$(<u>title</u>, <u>year</u>, studio-Name)
  Movies$_2$(<u>studio-Name</u>, president, presAddr)
  Projected FD's:
  – For *Movies$_1$*: *title, year → studio-Name*
  – For *Movies$_2$*: *studio-Name → president* and *president → presAddr*

- BCNF violations?
  – For *Movies$_1$*, *{<u>title</u>, <u>year</u>}* is key and all left sides of FD's are superkeys.
  – For *Movies$_2$*, *{<u>studio-Name</u>}* is the key… but *president → presAddr* violates BCNF.

CSC43 N Koudas 34

Decomposing $Movies_2$…

- First set of decomposed relations:
  $Movies_1$(<u>title</u>, <u>year</u>, studio-Name)
  $Movies_2$(<u>studio-Name</u>, president, presAddr)
- Closure *president* $^+$… = {*president*, *presAddr*}
- Decompose $Movies_2$ into:
  $Movies_{2\text{-}1}$(<u>studio-Name</u>, president)
  $Movies_{2\text{-}2}$(<u>president</u>, presAddr)
- Resulting relations are all in BCNF:
  $Movies_1$(<u>title</u>, <u>year</u>, studio-Name)
  $Movies_{2\text{-}1}$(<u>studio-Name</u>, president)
  $Movies_{2\text{-}2}$(<u>president</u>, presAddr)

That's it for today…