Design Theory of Relational DBs.

- What makes a good database?
  - Normalization
  - Avoid "anomalies"

Functional Dependencies (FD)

- A FD is a constraint between two set of attributes of a relation.
- Given R, a set of attributes X in R is said to functionally determine another set of attributes Y in R ($X \rightarrow Y$) iff 2 tuples have the same values of attributes X then they must have the same values for attributes Y.

We write them as:

$$A_1 \ldots A_n \longrightarrow B_1 \ldots B_m$$

Attributes written as list.

Example:

title	Year	Lenght	Genre	StudioName	StarName
Star Wars	1977	124	Sci Fi	Fox	Carrie Fisher
Star Wars	1977	124	Sci Fi	Fox	Mark Hamill
Star Wars	1977	124	Sci Fi	Fox	Harrison Ford
The Godfather	1972	175	Drama	Paramount	Robert Duvall
The Godfather	1972	175	Drama	Paramount	Marlon Brandon
Apocalypse Now	1979	153	War	Zoetrope	Marlon Brandon

I claim by design that

title, year $\longrightarrow$ lenght, genre, studioName

title, year $\not\longrightarrow$ starName

Superkey (SK)

A set of attributes $\{A_1 \ldots A_n\}$ is a superkey of R iff $A_1 \ldots A_n \rightarrow R$

Candidate key (key)

A candidate key is a superkey that is minimal:

There is no proper subset of C of $\{A_1 \ldots A_n\}$ s.t $C \rightarrow R$

One candidate key becomes the Primary Key!

<u>2</u>

For our example:

- All attributes of R are <u>always</u> a SK.
- title, year, starname is a candidate Key

  title, year, starName $\rightarrow$ R.

- Any superset of a candidate Key is a SK.

Reasoning about FD's.

- Given a relation R two sets of FDs A & B are equivalent if.

  The set of instances of R that satisfy A is exactly the same that satisfy B

- A follows from B if every instance of R that satisfies B also satisfies A.

- A & B are equivalent iff.

  A follows B and B follows A

<u>3</u>

Armstrong's Axioms (3.2 page 81)

Given relation R with subsets of attributes
$X, Y, Z \subseteq R$

Reflexivity (Trivial)
$Y \subseteq X$ then $X \rightarrow Y$

Augmentation:
If $X \rightarrow Y$ then $XZ \rightarrow YZ$ for any Z

Transitivity:
If $X \rightarrow Y, Y \rightarrow Z$ then $X \rightarrow Z$

Additional Rules.
They can be derived from axioms.

Union:
If $X \rightarrow Y, X \rightarrow Z$ then $X \rightarrow YZ$

Decomposition
If $X \rightarrow YZ$ then $X \rightarrow Y$ and $X \rightarrow Z$

Ex :

Derive Union from axioms.

$X \rightarrow Y$, $X \rightarrow Z$

$XZ \rightarrow YZ$ Augmentation

$\left.\begin{array}{l} XX \rightarrow XZ \\ X \rightarrow XZ \end{array}\right\}$ Augmentation.

$\left.\begin{array}{l} X \rightarrow XZ \rightarrow YZ \\ X \rightarrow YZ \end{array}\right\}$ Transitivity

Derive decomposition

$X \rightarrow YZ$

$\left.\begin{array}{l} YZ \rightarrow Y \\ YZ \rightarrow Z \end{array}\right\}$ Reflexivity.

$\left.\begin{array}{l} X \rightarrow Y \\ X \rightarrow Z \end{array}\right\}$ Transitivity.

Closure of attributes (3.2.4)

Given a relation R and a set f of FDs, what other FDs can be computed from a set of FDs f?

The closure of a set of attributes $A_1 \ldots A_n$ denoted $\{A_1 \ldots A_n\}^+$ is the set of attr. that can be derived from $A_1 \ldots A_n$ using f.

Hard to do and error prone via axioms!!

Alg:

1) Rewrite FDs in f in canonical form
2) $X \leftarrow A_1 \ldots A_n$
3) for each $B_1 \ldots B_m \rightarrow C$ in FDs
   if $B_1 \ldots B_m \subseteq X$ and $C \notin X$
   add C to X
4) Repeat (3) until X does not change

$$X \text{ is } \{A_1 \ldots A_n\}^+$$

Ex:

R(A BC DE)

$$f = \begin{cases} AB \to C \\ BC \to AD \\ D \to E \\ CE \to B \end{cases}$$

①
AB → C	1
BC → A	2
BC → D	3
D → E	4
CE → B	5

Compute $\{AB\}^+$

X ← AB

First pass:

X ← ABC   fd 1

X ← ABCD   fd 3

X ← <u>ABCDE</u>   fd 4

all attributes hence X will not change any more

$\{AB\}^+ = \{ABCDE\}$

AB is a SK of R. Is it a candidate key?

Closure of attr. can help us find CKs of a relation: Compute $\{A\}^+$ and $\{B\}^+$

<u>7</u>

Closure of set of FDs

Given a set f of FDs, its closure f^+ is the set of all FDs derived from f.

Ex:

$$f = \left\{ \begin{array}{l} A \to B \\ B \to C \end{array} \right\} \qquad A \to C \in f^+$$

Two sets A & B of FDs are equivalent iff. $A^+ = B^+$

$$\left\{ \begin{array}{l} A \to B \\ B \to C \end{array} \right\}^+ = \left\{ \begin{array}{l} A \to B \\ B \to C \\ A \to C \end{array} \right\}^+$$

We can easily test if $X \to Y \in f^+$

$$X \to Y \in f^+ \text{ iff } Y \subseteq X^+ \text{ using } f.$$

Ex:

$$A \to C \in \left\{ \begin{array}{l} A \to B \\ B \to C \end{array} \right\}$$

$$\{A\}^+ = \{ABC\} \Rightarrow A \to ABC$$
$$\Rightarrow \begin{array}{l} A \to A \\ A \to B \\ A \to C \end{array}$$

Basis of a relation

Given a relation R and FDs f we say
that any set g s.t. $f^+ = g^+$ is a
basis of R.

Minimal Basis of FDs. (3.2.7)

Any relation R has many equivalent
set of FDs (many basis-es).
To avoid an explosion of FDs we
usually use a minimal basis

A **minimal basis** B of a relation R is
a basis of R s.t.

1) All FDs in B are in canonical form

2) If for any FD we remove one or more
attr. from the left hand side the
result is no longer a basis,

3) If any FD is removed from B, the
result is no longer a basis

Ex: Is A a minimal basis of R with fds B?

$$A = \begin{cases} A \to B \\ B \to A \\ B \to C \\ C \to B \end{cases}$$

$$B = \begin{cases} A \to B & ① \\ A \to C & ② \\ B \to A & ③ \\ B \to C & ④ \\ C \to A & ⑤ \\ C \to B & ⑥ \\ AB \to C & ⑦ \\ AC \to B & ⑧ \\ BC \to A & ⑨ \end{cases}$$

Given FDs in A, can we generate FDs in B?

①, ③, ④, ⑥ already in A.

Can we generate ②, ⑤, ⑦, ⑧, ⑨?

$\{A\}^+ = \{A B C\}$.

⟹ ② can be generated, and also ⑦, ⑧ by using augmentation. (or compute $\{AB\}^+$, $\{AC\}^+$)

⑤? $\{C\}^+ = \{C B A\}$ ⟹ yes ⑤ and

⑨ (augmentation) can be generated.

So from A we can generate B.

Hence A is a basis of R

B is also a basis of R with fds A

$$B \subset A$$

$$\underset{\text{subset}}{\longleftarrow}$$

Is A minimal?

- Can we drop $A \to B$?

  $A \to B$ be generated from $\left\{ \begin{array}{l} B \to A \\ C \to B \\ B \to C \end{array} \right\}$

  $\{A\}^+ = \{A\}$ so no, $A \to B$ cannot be removed.

- Can we drop $B \to A$?

  $B \to A$ be generated from $\left\{ \begin{array}{l} A \to B \\ C \to B \\ B \to C \end{array} \right\}$

  $\{B\}^+ = \{B \ C\}$, so no, it can not be removed.

- Repeat for $C \to B$ and $B \to C$.

Yes, A is minimal

(11)

Another Example.

Given
$$AC \rightarrow D$$
$$AD \rightarrow C$$
$$A \rightarrow CD$$
$$C \rightarrow D$$

Compute its minimal basis

1) Write in canonical form

① $AC \rightarrow D$ ③ $A \rightarrow C$ ⑤ $C \rightarrow D$

② $AD \rightarrow C$ ④ $A \rightarrow D$

2) Remove redundant attr. for LHS

Test A in ①

Can we generate A from C?

$\{C\}^+ = CD$ NO

Test C:

Can we generate C from A:

$\{A\}^+ = AC...$ yes

Drop C from ① $\Rightarrow$ $A \rightarrow D$. (Replace ①)

_____

②

Can we generate A from D? NO

Can we generate D from A? Yes.

$\Rightarrow$ $A \rightarrow C$ (Replace ②)

12

Now we have.

① A → D   ③ A → C   ⑤ C → D

② A → C   ④ A → D

③ Remove redundant FDs.

   ① and ② are obviously redundant.

      ⇒ Remove.

   New FDs:

         ③ A → C   ⑤ C → D
         ④ A → D

   can ③ be generated from ④,⑤?

         No.   Keep.

   Can ④ be generated from ③⑤?

      Yes. Remove.

   Can ⑤ be generated from ③?

         No.   Keep.

Minimal   Basis

         ③   A → C

         ⑤   C → D

13

Projection of FDs   (3.2.8)

Given R and set F of FDs
the projection of F on $R_1 = \Pi_L R$
is the set of FDs that follows from F
that involve only attributes in R_1.

Algorithm:

T ← ∅
for each subset X ∈ L compute $\{X\}^+$
  for every attribute A in $\{X\}^+$
    add X → A to T
      iff A ∈ L and
           A ∉ X (non trivial)

Ex: R(ABCD) $F = \begin{cases} A \rightarrow B \\ B \rightarrow C \\ C \rightarrow D \end{cases}$

Compute FDs of $\Pi_{ACD} R$

                    closure
1) A C D | ②A̶ B̶ C̶ D̶   3) Remove
    A C | A̶ C̶ B̶ D   RHS att not
    A   D | A̶ D̶ B̶ C   in L.
    A | A̶ B̶ C D   4) Remove
      C D | C̶ D   trivial FDs.
      C | C̶ D
        D | D̶

Result.: $\begin{cases} AC \rightarrow D \quad A \rightarrow C \\ AD \rightarrow C \quad A \rightarrow D \\ \qquad\qquad\quad C \rightarrow D \end{cases}$ ▨

Is it a minimal basis? <u>No</u>

$\begin{cases} A \rightarrow D \\ AC \rightarrow D \\ AD \rightarrow C \end{cases}$ can be generated from $\begin{cases} A \rightarrow C \\ C \rightarrow D \end{cases}$

                    Prove it!!

Design of Relational DBs (3.3)

title	Year	Length	Genre	Studio Name	Star Name
Star Wars	977	124	Sci Fi	Fox	Carry Fisher
Star Wars	977	124	Sci Fi	Fox	Mark Hamill
Star Wars	977	124	Sci Fi	Fox	Harrison Ford
The Godfather	972	175	Drama	Paramount	Robert Duvall
The Godfather	972	175	Drama	Paramount	Marlon Brandon
Apocalyse Now	1953	153	War	Zoetrope	Marlon Brandon

FDs: title, year → length, studio Name

Anomalies

- Redundancy: Unnecessary repeated info
- Update anomalies: If we change one tuple we might have to change another: Ex: change Lenght of a movie.
- Deletion anomalies: If we delete a tuple we might lose other info: Ex: Remove M. Brandon from Ap. Now.

Decomposing Relations

To deal with anomalies we decompose relations.

Given $R(A_1 .. A_n)$ a decomposition into $S(B_1 .. B_m)$ and $T(C_1 .. C_k)$ s.t.

1) $\{A_1 ... A_n\} = \{B_1 .. B_m\} \cup \{C_1 ... C_k\}$ and

2) $S = \Pi_{B_1 ... B_m} R$ and
$T = \Pi_{C_1 ... C_k} R$

We can decompose Movies into
S(title, year, length, genre, studio Name)
T(title, year, star Name)

title	Year	length	Genre	Studio Name	
Star Wars	1977	124	Sci Fi	Fox	=S
The Godfather	1972	175	Drama	Paramount	
Apocalyte Now	1953	153	War	Zoetrope	

Call this Movies 2

title	year	starName
Star Wars	1977	Carry Fisher
Star Wars	1977	Mark Hamill
Star Wars	1977	Harrison Ford
The Godfather	1972	Robert Duvall
The Godfather	1972	Marlon Brandon
Apocalyte Now	1953	Marlon Brandon

$= T.$

Call this Movies 3

Good decompositions.

Given a relation R we want to decompose
it into two relations S and T s.t.

1) $R = S \bowtie T$ lossless join

2) The projection of FDs F_R of R
   into S (F_S) and T (F_T)
   satisfies:
$$(F_S \cup F_T)^+ = F_R^+$$

Dependency preserving.

Boyce Codd Normal Form (BCNF)

A relation R is in BCNF iff
for every FD

$$A_1 \ldots A_n \rightarrow B_1 \ldots B_m$$

$A_1 \ldots A_n$ is a Superkey.
or
$\{B_1 \ldots B_m\} \subseteq \{A_1 \ldots A_n\}$ (it is trivial).

Ex:

Movies is not BCNF

title, year $\rightarrow$ lenght, studioName

is a not a SK of Movies

If a relation R is not BCNF then
decompose into relations $R_1 \ldots R_n$
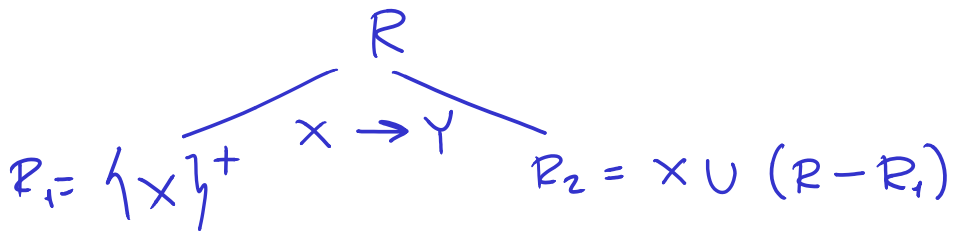s.t $R_1 \bowtie R_2 \ldots \bowtie R_n = R$

$\Rightarrow$ Loss-less join decomposition.

Algorithm to decompose into BCNF relations

Given R and set F of FDs:

R is <u><u>not</u></u> BCNF.

1) Choose one FD $X \rightarrow Y$ not in BCNF

2). Decompose:

$$R$$

$$X \rightarrow Y$$

$R_1 = \{X\}^+$ $R_2 = X \cup (R - R_1)$

3) Compute FDs for R_1 and R_2
   (projection of FDs of R into R_1, R_2)

4) If R_1 or R_2 are not BCNF recursively decompose.

Guaranteed to be lossless join. but not FD preserving.

Any non BCNF relation has a
  BCNF loss-less join decomposition
  but not a BCNF FD preserving
  decomposition.

Ex: $R(TCH)$ $H \to C$
                                    $TC \to H.$

      $H \to C$ not BCNF.
               $TCH.$

               $H \to C$

  $\{H\}^+ =$ $H \cup \{TCH - HC\}$

  $HC \leftarrow$ any 2 att rel $\nearrow$ HT

FDs   is BCNF   FDs

~~HC~~	~~HC~~
H	~~HC~~
~~C~~	~~C~~

$H \to C$

~~HT~~	~~HTC~~
~~H~~	~~HC~~
~~T~~	~~T~~

$\emptyset$

Decomposition $R_1 = HC$ $FD_1 = \{H \to C\}$
                     $R = HT$ $FD_2 = \emptyset$

Not FD preserving lost $TC \to H.$

3rd Normal Form (3NF)

If we cannot decompose a relation into BCNF relations that are FD preserving we are happy if they can be decompose into 3NF relations.

Any relation R with a set of FDs F has a 3NF decomposition that is loss-less join and FD preserving.

A relation R with FDs F is in 3NF if for every FD

$$A_1 .. A_n \rightarrow B_1 ... B_m$$

- it is a : SK
  or for all
- $C \in \{B_1 ... B_m\}$ is either

  $C \in \{A_1 .. A_n\}$

  or C is part of __some candidate key__.

Ex:

R(ABCDE)

$F = \begin{cases} AB \to C \\ C \to B \\ A \to D \end{cases}$

Is it 3NF?

AB → C   AB not a SK.

is C part of a CK?

Need to compute candidate keys of R

• Heuristic:

AE never on right hand side of FD

⟹ always part of a key.

Use closure of attr. to compute SKs.

all combination of attr.   closure.

$\Downarrow$

all combination	closure	
AE B C D	AEBCD	
AE B C	AE BC D	
AE B D	AE BD C	all SK.
AE B	AEB C D	
AE C D	AE CD B	
AE C	AE CBD	
AE D	AE D	
AE	AED.	

minimal ⟹ Candidate Keys. $\begin{cases} AEB \\ AEC \end{cases}$

23

Back to testing if R is 3NF.

$AB \rightarrow C$ C is part of CK. ✓

$C \rightarrow B$ C is not SK. but
           B is part of CK.

$A \rightarrow D$ A is not SK.
           D is not part of CK

$\Rightarrow$ R is not 3NF.

Decomposition of a Relation into 3NF
relations that is loss-less join and
FD preserving: (Synthesis alg 3.5.2)

Given R with set of FDs F

1) Find G, a minimal basis of F

2) For each FD $A_1 \ldots A_n \rightarrow B$ in G.
   add a relation with schema
   $A_1 \ldots A_n B$ with FD
   $A_1 \ldots A_n \rightarrow B$

3) If **none** of the added relations in
   step 2 is a **SK** of R (i.e. does
   not contain a candidate key of R)
   then add another relation whose
   schema is a key of R.

4) **Simplify:** if the attributes of a
   subrelation are a subset of another
   subrelation $\Rightarrow$ merge them.
   i.e. delete subset relation and
   add its FDs to other relation.

EX:

R(ABCDE)

$$F = \begin{cases} AB \to C \\ C \to B \\ A \to D \end{cases}$$ F is minimal basis

$R_1 = ABC$ with $AB \to C$
$R_2 = CB$ with $C \to B$
$R_3 = AD$ with $A \to D$

None of them   contains a SK.
(see previous exercise, you can
verify it by computing $\{R_i\}^+$)
We know its keys are AEB and AEC
We need only one. So add

$R_4 = AEB$ with no FDs.
But R_2 is a subset of R_1 :
     So we merge them!

$\Rightarrow$ Decomposition! $\longrightarrow$ $R_1 = ABC$ $\begin{cases} AB \to C \\ C \to B \end{cases}$

3NF

BCNF $\begin{cases} R_3 = AD \end{cases} \begin{cases} A \to D \end{cases}$

$R_4 = AEB$