CSC 370

Activity Worksheet: BCNF Decomposition

Dr. Sean Chester

Fall 2022

Notes

This worksheet should provide extra practice questions for decomposing a relation into Boyce-Codd Normal Form (BCNF). In these questions, you are given a relation and set of functional dependencies and you should show the process of decomposition into a set of sub-relations that are all in BCNF. The first question is already answered as an example.

Questions

1. Political Contributions

   PoliticalContributions(donor_id, donor_name, candidate_name, party, year, riding, amount)

   donor_id → donor_name
   candidate_name → party riding
   party riding → candidate_name
   donor_id candidate_name year → amount

   Solution:

   First find any BCNF violations:

   Considering the first FD and check if the left-hand side is a superkey. Let A denote the set of all six attributes.

   $\{donor\_id\}^+ = \{donor\_id, donor\_name\} \neq A$ // Therefore, the first FD is a BCNF violation.

   Partition A into A_1 and A_2, based on $\{donor\_id\}^+$:
   $A_1 = \{donor\_id, donor\_name\}$
   $A_2 = \{candidate\_name, party, riding, year, amount, donor\_id\}$

   Recurse on $R_1(donor\_id, donor\_name)$ with FD's: donor_id → donor_name.

   Here, donor_id is a key so R_1 is already in BCNF and we return from this step of the recursion.

   So we will return $\{R_1\}$ with union of relations from recursing now on...

   $R_2(candidate\_name, riding, party, year, amount, donor\_id)$ with FD's:
   candidate_name → party riding
   party riding → candidate_name
   donor_id candidate_name year → amount

   We check if candidate_name is a superkey:
   $\{candidate\_name\}^+ = \{candidate\_name, party, riding\} \neq A_2$,
   so we decompose based on $\{candidate\_name\}^+$:
   $A_{21} = \{candidate\_name, party, riding\}$
   $A_{22} = \{candidate\_name, donor\_id, year, amount\}$

   Next we recurse on $R_{21}(candidate\_name, party, riding)$ with FD's:
   candidate_name → party riding
   party riding → candidate_name

   However, we see that both $\{candidate\_name\}$ and $\{party, riding\}$ are keys; so, R_{21} is in BCNF. We simply return $\{R_{21}\}$ from the recursion.

   So, we will return the union of $\{R_{21}\}$ with the result of recursing on...

$R_{22}(candidate\_name, donor\_id, year, amount)$ with FD's:
donor_id candidate_name year $\rightarrow$ amount

Here, we see that $\{donor\_id, candidate\_name, year\}$ is a key; so, R_{21} is also already in BCNF. We simply return $\{R_{22}\}$ from the recursion.

Thus, our final solution is:
$\{R_1\} \cup (\{R_{21}\} \cup \{R_{22}\})$
$= \{R_1\} \cup \{R_{21}, R_{22}\}$
$= \{R_1, R_{21}, R_{22}\}$, i.e.,:

Donor(donor_id, donor_name)
Candidate(candidate_name, party, riding)
AnnualContribution(donor_id, candidate_name, year, amount)

2. Songs

   Songs(artist_id, artist_name, song_name, album_name, run_length)
   artist_id → artist_name
   song_name artist_id → run_length

   Solution:

3. R(A, B, C, D, E)

   $AB \rightarrow CDE$

   Solution:

4. R(A, B, C, D, E)

   $AC \rightarrow BD$

   Solution:

5. R(A,B,C,D,E) [Difficult]
   $A \rightarrow BC$
   $BC \rightarrow D$
   $C \rightarrow B$

   This question was on the Fall 2021 take-home final exam.

   Solution:

Solutions

Question 1

First find any BCNF violations:

Considering the first FD and check if the left-hand side is a superkey. Let A denote the set of all six attributes.

$\{donor\_id\}^+ = \{donor\_id, donor\_name\} \neq A$ // Therefore, the first FD is a BCNF violation.

Partition A into A_1 and A_2, based on $\{donor\_id\}^+$:
$A_1 = \{donor\_id, donor\_name\}$
$A_2 = \{candidate\_name, party, riding, year, amount, donor\_id\}$

Recurse on $R_1(donor\_id, donor\_name)$ with FD's: donor_id → donor_name.

Here, donor_id is a key so R_1 is already in BCNF and we return from this step of the recursion.

So we will return $\{R_1\}$ with union of relations from recursing now on...

$R_2(candidate\_name, riding, party, year, amount, donor\_id)$ with FD's:
candidate_name → party riding
party riding → candidate_name
donor_id candidate_name year → amount

We check if candidate_name is a superkey: $\{candidate\_name\}^+ = \{candidate\_name, party, riding\} \neq A_2$,
so we decompose based on $\{candidate\_name\}^+$:
$A_{21} = \{candidate\_name, party, riding\}$
$A_{22} = \{candidate\_name, donor\_id, year, amount\}$

Next we recurse on $R_{21}(candidate\_name, party, riding)$ with FD's:
candidate_name → party riding
party riding → candidate_name

However, we see that both $\{candidate\_name\}$ and $\{party, riding\}$ are keys; so, R_{21} is in BCNF. We simply return $\{R_{21}\}$ from the recursion.

So, we will return the union of $\{R_{21}\}$ with the result of recursing on...

$R_{22}(candidate\_name, donor\_id, year, amount)$ with FD's:
donor_id candidate_name year → amount

Here, we see that $\{donor\_id, candidate\_name, year\}$ is a key; so, R_{21} is also already in BCNF. We simply return $\{R_{22}\}$ from the recursion.

Thus, our final solution is:
$\{R_1\} \cup (\{R_{21}\} \cup \{R_{22}\})$

$= \{R_1\} \cup \{R_{21}, R_{22}\}$
$= \{R_1, R_{21}, R_{22}\}$, i.e.,:

Donor(donor_id, donor_name)
Candidate(candidate_name, party, riding)
AnnualContribution(donor_id, candidate_name, year, amount)

Question 2

$\{artist\_id\}^+ = \{artist\_id, artist\_name\}$ // BCNF Violation
$\{artist\_id, song\_name\}^+ = \{artist\_id, song\_name, artist\_name, run\_length\}$ // BCNF Violation

Solution A - Starting with Violation: artist_id → artist_name

$R_1 = \{artist\_id\}^+ = \{artist\_id, artist\_name\}$
$R_2 = \{artist\_id, artist\_name, song\_name, album\_name, run\_length\} \setminus \{artist\_id\}^+ \cup \{artist\_id\}$

Recursing on R_1 with FD's: $\{artist\_id \rightarrow artist\_name\}$ will simply return $\{R_1\}$
Recursing on R_2 with FD's: $\{song\_nameartist\_id \rightarrow run\_length\}$:

$\{artist\_id, song\_name\}^+ = \{artist\_id, song\_name, run\_length\}$ // BCNF Violation. It does not determine album_name.

$R_{21} = \{artist\_id, song\_name\}^+ = \{artist\_id, song\_name, run\_length\}$
$R_{22} = \{artist\_id, song\_name, album\_name, run\_length\} \setminus \{artist\_id, song\_name\}^+$
$\cup \{artist\_id, song\_name\}$

Recursing on R_{21} with FD's: $\{song\_name\ artist\_id \rightarrow run\_length\}$ will simply return $\{R_{21}\}$.
Recursing on R_{22} with FD's: $\emptyset$ will simply return $\{R_{22}\}$.

Thus, our solution is:

Artist(artist_id, artist_name)
Song(artist_id, song_name, run_length)
Album(artist_id, song_name, album_name)

Solution B - Starting with Violation: song_name artist_id → run_length

$R_1 = \{artist\_id, song\_name\}^+ = \{artist\_id, song\_name, artist\_name, run\_length\}$
$R_2 = \{artist\_id, artist\_name, song\_name, album\_name, run\_length\} \setminus \{artist\_id, song\_name\}^+$
$\cup \{artist\_id, song\_name\}$

Recursing on R_2 with FD's: $\emptyset$ will simply return $\{R_2\}$
Recursing on R_1 with FD's: $\{artist\_id \rightarrow artist\_name, song\_name\ artist\_id \rightarrow run\_length\}$:

$\{artist\_id\}^+ = \{artist\_id, artist\_name\}$ // BCNF Violation
$\{artist\_id, song\_name\}^+ = \{artist\_id, song\_name, artist\_name, run\_length\}$ // A Key!

$R_{11} = \{artist\_id\}^+ = \{artist\_id, artist\_name\}$
$R_{12} = \{artist\_id, artist\_name, song\_name, run\_length\} \setminus \{artist\_id\}^+ \cup \{artist\_id\}$

Recursing on R_{11} with FD's: $\{artist\_id \rightarrow artist\_name\}$ will simply return $\{R_{11}\}$
Recursing on R12 with FD's: $\{song\_name\ artist\_id \rightarrow run\_length\}$ will simply return $\{R_{12}\}$

Thus, our final solution is:

Artist(artist_id, artist_name)

Song(artist_id, song_name, run_length)
Album(artist_id, song_name, album_name)

Observe that we ended up with the same result, irrespective of which order we processed the violations. This is not, in general, true.

Question 3

$\{AB\}^+ = \{A, B, C, D, E\}$ // Thus $\{A, B\}$ is a (super)key

This relation has no BCNF violations; it is already in BCNF.

Question 4

$\{AC\}^+ = \{A, B, C, D\}$ // This is a BCNF violation, because we cannot functionally determine E from A and C.

$R_1 = \{A, C\}^+ = \{A, B, C, D\}$
$R_2 = \{A, B, C, D, E\} \setminus \{A, C\}^+ \cup \{A, C\} = \{A, C, E\}$

Recursing on R_1 with FD's: $AC \rightarrow BD$ will simply return $\{R_1\}$
Recursing on R_2 with FD's: $\emptyset$ will simply return $\{R_2\}$

Thus, our final solution is:

R_1(A, B, C, D)
R_2(A, C, E)

Question 5

$\{A\}^+ = \{A, B, C, D\}$ // BCNF violation; does not determine E.
$\{B, C\}^+ = \{B, C, D\}$ // BCNF violation; does not determine A nor E.
$\{C\}^+ = \{B, C, D\}$ // BCNF violation; does not determine A nor E.

Solution A - Starting with Violation: A → BC

$R_1 = \{A\}^+ = \{A, B, C, D\}$
$R_2 = \{A, B, C, D, E\} \setminus \{A\}^+ \cup \{A\} = \{A, E\}$

Recursing on R_2 with FD's: $\emptyset$ will simply return $\{R_2\}$
Recursing on R_1 with FD's: $\{A \to BC, BC \to D, C \to B\}$:

$\{A\}^+ = \{A, B, C, D\}$ // A Key!
$\{B, C\}^+ = \{B, C, D\}$ // BCNF violation; does not determine A.
$\{C\}^+ = \{B, C, D\}$ // BCNF violation; does not determine A.

Solution A' - Starting with Violation: BC → D

$R_{11} = \{B, C\}^+ = \{B, C, D\}$
$R_{12} = \{A, B, C, D\} \setminus \{B, C\}^+ \cup \{B, C\} = \{A, B, C\}$

Recursing on R_{12} with FD's: $\{A \to BC, C \to B\}$:
$\{A\}^+ = \{A, B, C\}$ // A Key!
$\{C\}^+ = \{B, C\}$ // BCNF violation; does not determine A.

$R_{121} = \{C\}^+ = \{B, C\}$
$R_{122} = \{A, B, C\} \setminus \{C\}^+ \cup \{C\} = \{A, C\}$

Recursing on R_{121} with FD's: $\{C \to B\}$ will return the set $\{R_{121}\}$
Recursing on R_{122} with FD's: $\{A \to C\}$ will return the set $\{R_{122}\}$

Recursing now on R_{11} with FD's: $\{BC \to D, C \to B\}$:

$\{B, C\}^+ = \{B, C, D\}$ // A Key!
$\{C\}^+ = \{B, C, D\}$ // A Key!

There are no BCNF violations; so, we simply return R_{11}.

So, solution A' looks like:

$R_{11}(B, C, D)$
$R_{121}(B, C)$
$R_{122}(A, C)$
$R_2(A, E)$

However, observe that $R_{121} \subset R_{11}$; so, R_{121} is redundant. Thus our correct final solution A' is:

$R_{11}(B, C, D)$

$R_{122}(A, C)$
$R_2(A, E)$

Note that this solution has lost FD A $\rightarrow$ B due to excessive decomposition; BCNF does not have the dependency preservation property.

Solution A" - Starting with Violation: C $\rightarrow$ B

$R_{21} = \{C\}^+ = \{B, C, D\}$
$R_{22} = \{A, B, C, D\} \setminus \{C\}^+ \cup \{C\} = \{A, C\}$

At this point, you can probably observe that we will end up with the same solution as Solution A', though it will take several steps.

Solution B - Starting with Violation: BC $\rightarrow$ D

$R_1 = \{B, C\}^+ = \{B, C, D\}$
$R_2 = \{A, B, C, D, E\} \setminus \{B, C\}^+ \cup \{B, C\} = \{A, B, C, E\}$

Recursing on R_1 with FD's: $\{C \rightarrow B, BC \rightarrow D\}$ will return the set $\{R_1\}$ since both $\{C\}$ and $\{BC\}$ are superkeys.
Recursing on R_2 with FD's: $\{C \rightarrow B, A \rightarrow BC\}$:

$\{C\}^+ = \{B, C\}$ // a BCNF violation; C cannot functionally determine A nor E.
$\{A\}^+ = \{A, B, C\}$ // a BCNF violation; A cannot functionally determine E.

$R_{21} = \{C\} = \{B, C\}$
$R_{22} = \{A, B, C, E\} \setminus \{C\}^+ \cup \{C\} = \{A, C, E\}$

Recursing on R_{21} with FD's: $\{C \rightarrow B\}$ will return the set $\{R_{21}\}$ since $\{C\}$ is a superkey.
Recursing on R_{22} with FD's: $\{A \rightarrow C\}$:

$\{A\}^+ = \{A, C\}$ // a BCNF violation; A cannot functionally determine E.

So we will decompose into and eventually return:

$R_{221} = \{A\}^+ = \{A, C\}$
$R_{222} = \{A, C, E\} \setminus \{A\}^+ \cup \{A\} = \{A, E\}$

Thus, our solution so far is:

$R_1(B, C, D)$
$R_{21}(B, C)$
$R_{221}(A, C)$
$R_{222}(A, E)$

But clearly R_{21} is redundant as it is a subset of R_1; so, our final solution is:

$R_1(B, C, D)$
$R_{221}(A, C)$

$R_{222}(A, E)$

Observe that we have still lost FD A → B.

Solution C - Starting with Violation: C → B

$R_1 = \{C\}^+ = \{B, C, D\}$
$R_2 = \{A, B, C, D, E\} \setminus \{C\}^+ \cup \{C\} = \{A, C, E\}$

Already at this point, you can probably recognise that we will end up with the same solution as Solution B. The details are not provided.

Summary

There is one unique decomposition for this problem; a previously posted version of this solution incorrectly stated a second solution, but this was because of a missing FD in one of the projections on Solution B.

$\{(B, C, D), (A, C), (A, E)\}$
$\{(B, C, D), (A, C, E)\}$