# CSC343 - A3

Juliano Garcia de Oliveira, UTORID: olive207

December 5th, 2018

## Database Design - Functional Dependencies

**Consider** $R(A, B, C, D, E, F, G, H)$**, with these FDs,** $S = \{A \to E \ \ , AD \to BE \ \ , AC \to E \ \ , E \to B \ \ , BG \to F \ \ , BE \to D \ \ , BDH \to E \ \ , F \to A \ \ , D \to H \ \ , CD \to A\}$

## a) Find the (candidate) keys of this schema:

Since G only appears in the LHS, it must be part of the candidate keys. So, let's consider all subsets of attributes which contain G. But from those, C also only appears in the LHS, so it must also be part of the candidate keys: 2:

$(CG)^+ = CG$ (not a key)

3:

$(ACG)^+ = ABCDEFGH$ (it's a key!)

$(BCG)^+ = ABCDEFGH$ (it's a key!)

$(DCG)^+ = ABCDEFGH$ (it's a key!)

$(ECG)^+ = ABCDEFGH$ (it's a key!)

$(FCG)^+ = ABCDEFGH$ (it's a key!)

$(HCG)^+ = CGH$ (not a key)

Every other combination which is candidate to be a key would need to CG in it. But if the other attributes are all combinations of A, B, D, E, F, then they're just an extension of the keys with three attributes (therefore not minimal). The only case left would be subsets with H and CG. But to have H, C and G and a cardinality over 3, it means that H would have to be paired with an atribute from {A, B, D, E, F}, therefore it would not be minimal also.

So the candidate keys are: $\{ACG, BCG, DCG, ECG, FCG\}$

$S = \{A \to E \ \ , AD \to BE \ \ , AC \to E \ \ , E \to B \ \ , BG \to F \ \ , BE \to D \ \ , BDH \to E \ \ , F \to A \ \ , D \to H \ \ , CD \to A\}$

## b) Using 3NF

I. **Keys =** $\{ACG, BCG, DCG, ECG, FCG\}$

II. **Calculate the Minimum Basis ($S$):**

II.a: Expand all FD's RHS if possible:
$S = \{A \to E \ \ , AD \to B \ \ , AD \to E \ \ , AC \to E \ \ , E \to B \ \ , BG \to F \ \ , BE \to D \ \ , BDH \to E \ \ , F \to A \ \ , D \to H \ \ , CD \to A\}$

II.b: Replace all FD's which can be inferred with another one with less LHS attributes, if the LHS has cardinality greater than or equal 2:

a) $A \to E$: Skip

b) $AD \to B$: Replace $AD \to B$ with $A \to B$

Calculating Closures: $(A)^+ = AEB$ , stop.

$S = \{A \to E \ \ , A \to B \ \ , AD \to E \ \ , AC \to E \ \ , E \to B \ \ , BG \to F \ \ , BE \to D \ \ , BDH \to E \ \ , F \to A \ \ , D \to H \ \ , CD \to A\}$

c) $AD \to E$: Replace $AD \to E$ with $A \to E$

Calculating Closures: $(A)^+ = AEB$ , stop.

$S = \{A \to E \ \ , A \to B \ \ , AC \to E \ \ , E \to B \ \ , BG \to F \ \ , BE \to D \ \ , BDH \to E \ \ , F \to A \ \ , D \to H \ \ , CD \to A\}$

d) $AC \to E$: Replace $AC \to E$ with $A \to E$

Calculating Closures: $(A)^+ = AEB$ , stop.

$S = \{A \to E \ \ , A \to B \ \ , E \to B \ \ , BG \to F \ \ , BE \to D \ \ , BDH \to E \ \ , F \to A \ \ , D \to H \ \ , CD \to A\}$

e) $E \to B$: Skip

f) $BG \to F$: Stays

Calculating Closures: $(B)^+ = B$, $(G)^+ = G$ , stop.

g) $BE \to D$: Replace $BE \to D$ with $E \to D$

Calculating Closures: $(B)^+ = B$, $(E)^+ = EBD$ , stop.

$S = \{A \to E \ \ , A \to B \ \ , E \to B \ \ , BG \to F \ \ , E \to D \ \ , BDH \to E \ \ , F \to A \ \ , D \to H \ \ , CD \to A\}$

h) $BDH \to E$: Replace $BDH \to E$ with $BD \to E$

Calculating Closures: $(B)^+ = B$, $(D)^+ = DH$ , $(H)^+ = H$, $(BD)^+ = BDHEB$ stop.

$S = \{A \to E \ \ , A \to B \ \ , E \to B \ \ , BG \to F \ \ , E \to D \ \ , BD \to E \ \ , F \to A \ \ , D \to H \ \ , CD \to A\}$

i) $F \to A$: Skip

j) $D \to H$: Skip

k) $CD \to A$: Stays

Calculating Closures: $(C)^+ = C$, $(D)^+ = DH$, stop.

The current set of FDs is:

$S = \{A \to E \ \ , A \to B \ \ , E \to B \ \ , BG \to F \ \ , E \to D \ \ , BD \to E \ \ , F \to A \ \ , D \to H \ \ , CD \to A\}$

`II.c: For each functional dependency K in S, remove K from S if it can be inferred from S - {K}:`

a) $A \to E$: Stays

$(A)^+_{S-\{(a)\}} = AB$

b) $A \to B$: Remove from S

$(A)^+_{S-\{(b)\}} = AEB$, stop.

c) $E \to B$: Stays

$(E)^+_{S-\{(b),(c)\}} = EDH$

d) $BG \to F$: Stays

$(BG)^+_{S-\{(b),(d)\}} = BG$

e) $E \to D$: Stays

$(E)^+_{S-\{(b),(e)\}} = EB$

f) $BD \to E$: Stays

$(BD)^+_{S-\{(b),(f)\}} = BDH$

g) $F \to A$: Stays

$(F)^+_{S-\{(b),(g)\}} = F$

h) $D \to H$: Stays

$(D)^+_{S-\{(b),(h)\}} = D$

i) $CD \to A$: Stays

$(CD)^+_{S-\{(b),(i)\}} = CDH$

The calculated Minimum Basis is: $M = \{A \to E \ \ , BG \to F \ \ , E \to BD \ \ , BD \to E \ \ , F \to A \ \ , D \to H \ \ , CD \to A\}$

III. **For each FD $X \to Y$ in $M$ define a new relation with schema $X \cup Y$**

$R_1 = \{A, E\}$

$R_2 = \{B, G, F\}$

$R_3 = \{B, D, E\}$

$R_4 = \{F, A\}$

$R_5 = \{D, H\}$

$R_6 = \{C, D, A\}$

IV. **If no relation is a superkey/key for the attributes, add a new which is some key.**

In this case no relation is a superkey, so I'll add this one:

$R_7 = \{A, C, G\}$

**Answer:** So our final decomposition is:

$R_1 = \{A, E\}, \ \ F_1 = \{A \to E\}$

$R_2 = \{B, G, F\}, \ \ F_2 = \{BG \to F\}$

$R_3 = \{B, D, E\}, \ \ F_3 = \{E \to BD, \ \ BD \to E\}$

$R_4 = \{F, A\}, \ \ F_4 = \{F \to A\}$

$R_5 = \{D, H\}, \ \ F_5 = \{D \to H\}$

$R_6 = \{C, D, A\},\ \ F_6 = \{CD \to A\}$

$R_7 = \{C, G, A\},\ \ F_7 = \{\}$

## c) Find a BCNF decomposition of the schema

A BCNF decomposition is a decomposition which doesn't violates the BCNF. I will use the relations I got from the 3NF in **b)**, since the question asks for a BCNF decomposition of the schema, not to follow the BCNF decomposition algorithm. I'll check if every relation is in BCNF, if not, apply the algorithm to that specific relation.

- $R_1 = \{A, E\},\ \ F_1 = \{A \to E\}$

$(A)+ = AE$, A is superkey

Violates BCNF? No

- $R_2 = \{B, G, F\},\ \ F_2 = \{BG \to F\}$

$(BG)+ = BGF$, BG is superkey

Violates BCNF? No

- $R_3 = \{B, D, E\},\ \ F_3 = \{E \to BD,\ \ BD \to E\}$

$(E)+ = EBD$, E is a superkey

$(BD)+ = EBD$, BD is a superkey

Violates BCNF? No

- $R_4 = \{F, A\},\ \ F_4 = \{F \to A\}$

$(F)+ = FA$, F is a superkey

Violates BCNF? No

- $R_5 = \{D, H\},\ \ F_5 = \{D \to H\}$

$(D)+ = DH$, D is a superkey

Violates BCNF? No

- $R_6 = \{C, D, A\},\ \ F_6 = \{CD \to A\}$

$(CD)+ = ACD$, CD is a superkey

Violates BCNF? No

- $R_7 = \{C, G, A\},\ \ F_7 = \{\}$

Violates BCNF? No

Since in this specific decomposition every relation follows the BCNF, it means this is a valid BCNF decomposition. Also, this is dependency preserving, since all FD's from the minimum basis are preserved (the 'lost' ones can be inferred from those). Here is the final BCNF decomposition of the schema:

$R_1 = \{A, E\},\ \ F_1 = \{A \to E\}$

$R_2 = \{B, G, F\},\ \ F_2 = \{BG \to F\}$

$R_3 = \{B, D, E\},\ \ F_3 = \{E \to BD,\ \ BD \to E\}$

$R_4 = \{F, A\},\ \ F_4 = \{F \to A\}$

$R_5 = \{D, H\}, \;\; F_5 = \{D \to H\}$

$R_6 = \{C, D, A\}, \;\; F_6 = \{CD \to A\}$

$R_7 = \{C, G, A\}, \;\; F_7 = \{\}$

3. Let $R(A_1, A_2, ..., A_n)$ be a relation with n attributes. Assume that the only keys are $\{A_1, A_2, A_3\}$ and $\{A_1, A_3, A_4\}$. Express as a function of n the total number of **superkeys** in R.

Since the second key is $\{A_1, A_3, A_4\}$ (has $A_4$), I'll assume that $n \geq 4$.

Let $K$ be the cardinality of a superkey, and let's look for the first few cases. - Considering the key $\{A_1, A_2, A_3\}$ (i.e. superkeys that contains this one) : With $K = 4$ we have $\binom{n-3}{1}$ superkeys (each of the $n-3$ attributes remaining can form a superkey). With $K = 5$ we have $\binom{n-3}{2}$ superkeys (choose groups of 2 from the remaining attributes, order doesn't matter). … - Considering the key $\{A_1, A_3, A_4\}$ : The argument works the same as above, we do a combination of the remaining attributes.

**OBS:** Some of the superkeys we get by the combinations of the remaining attributes union the first key $\{A_1, A_2, A_3\}$ are also part of some of the superkeys generated by the same procedure but with key $\{A_1, A_3, A_4\}$. That is, fix $A_4$ as a member of a superkey which also have $\{A_1, A_2, A_3\}$. If we consider $K = 4$, we have superkey $\{A_1, A_2, A_3, A_4\}$. However this same key is part of the keys we could generate from the **SECOND** key if we choose $A_2$. So we need to subtract the number of superkeys which overlap when we do the final calculation.

We know that the biggest value for $K$ is $n$, and given we already fixed 3 values, we can do the procedure above (varying $K$) for $n-3$ times in total. The number of superkeys who overlap is, w.l.o.g., if we are constructing a superkey "from" the first key $\{A_1, A_2, A_3\}$, we fix $A_4$ and vary the remaining attributes. This is valid because the same number of superkeys built from this procedure will appear when we fix $A_2$if we were to build form the second key, and they will actually be the same. This number is, given a cardinality of the superkey, $\binom{n-4}{K-4}$. That is, given that we fixed the first 4 attributes of the superkey (now we have $n-4$ remaining attributes to choose from), we choose the remaining number of times to complete a set with cardinality $K$.

So we can sum up the amount of superkeys we can build from both original keys (the number is the same for both), and subtract the overlapping keys (that we would be counting twice). This is the formula, rearranging $K$ to a new index, and without taking into account the original keys.

$$(1) : \sum_{i=1}^{n-3} 2 \cdot \binom{n-3}{i} - \binom{n-4}{i-1}$$

However, let's **take into account** the original keys. That is, $\{A_1, A_2, A_3\}$ and $\{A_1, A_3, A_4\}$ count as superkeys too. So we just need to add 2 to the formula (1). However, let's expand 2 to be $2 \cdot \binom{n-3}{0}$. And let's add this term to (1) then expand it:

$$2 \cdot \binom{n-3}{0} + \sum_{i=1}^{n-3} 2 \cdot \binom{n-3}{i} - \binom{n-4}{i-1} =$$

$$2 \cdot \binom{n-3}{0} + 2 \cdot \binom{n-3}{1} - \binom{n-4}{0} + ... + 2 \cdot \binom{n-3}{n-3} - \binom{n-4}{n-4}$$

We can organize this last equation like this:

$2 \cdot \binom{n-3}{0} +$

$2 \cdot \binom{n-3}{1} - \binom{n-4}{0} +$

$2 \cdot \binom{n-3}{2} - \binom{n-4}{1} +$

$2 \cdot \binom{n-3}{3} - \binom{n-4}{2} +$

...+

$2 \cdot \binom{n-3}{n-3} - \binom{n-4}{n-4}$

Which can be rearranged in (summing up the 'columns'):

$$2 \cdot \sum_{i=0}^{n-3} \binom{n-3}{i} - \sum_{i=0}^{n-4} \binom{n-4}{i}$$

Using the Pascal Triangle sum property, we can reduce this to:

$$2^{n-3} - 2^{n-4} = 2^{n-4}(2^2 - 1) = 3 \cdot 2^{n-4}$$

**Answer:** So $\forall n \geq 4$ the total number of superkeys in $R$ is given by the following equation:

$$f(n) = 3 \cdot 2^{n-4}$$

4. Prove or disprove that:

a) If $A \to B$ then $B \to C$:

False. Consider this counterexample:

| A | B | C |
|---|---|---|
| x | y | a |
| x | y | b |

b) If $AB \to C$ then $A \to C$ and $B \to C$:

False. Consider this counterexample:

| A | B | C |
|---|---|---|
| x | y | a |
| x | z | b |