CSCE-608 Database Systems

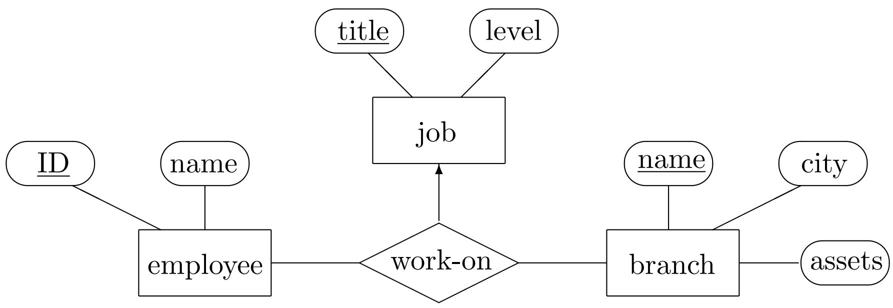
Spring 2025

Instructor: Dr. Jianer Chen
Office: PETR 428
Phone: 845-4259
Email: `chen@cse.tamu.edu`
Office Hours: MW 1:30 pm–3:00 pm

TA: Shahin John John Stella
Office: PETR 445
Phone: (979) 575-5523
Email: `ivin-98@tamu.edu`
Office Hours: TR 11:00 am–12:00 pm

Assignment #1 Solution

1. Convert the following E/R diagram into a relational database schema.



Answer. There are three entity sets, i.e., "employee," "branch," and "job." These entity sets result in three relations, given below, where underlined attributes mark the keys of the relations:

Employee($\underline{\text{ID}}$, name)
Branch($\underline{\text{name}}$, city, assets)
Job($\underline{\text{title}}$, level)

The relationship "work-on" gives a relation:

Work-On($\underline{\text{ID}}$, $\underline{\text{branch.name}}$, assets).

Note that since "work-on" is a many-to-one relationship from "employee" and "branch" to "job," each pair of entities from "employee" and "branch" uniquely determines an entity in "job." As a result, the key "title" in the relation "Job" is not part of the key of "Work-On."

2. List all functional dependencies for the relations constructed in Question 1.

Answer. Below are the functional dependencies for the relations:

- relation "Employee": {ID → name}
- relation "Branch: {name → city, assets} (or {name → city, name → assets})
- relation "Job": {title → level}
- relation "Work-On": {ID, branch.name → title}.

3. Consider a relation about people in the United States, including their name, Social Security number, street address, city, state, ZIP code, area code, and phone number (7 digits). What FDs would you expect to hold? What are the keys for the relation?

Answer. The relation is given as

USpeople(name, SSNo, st-addr, city, state, ZIP-code, area-code, phone)

The **functional dependencies** can be:

- SSNo → name, st-addr, city, state, ZIP-code, area-code, phone (Assuming each person has a single address and a single phone number).
- area-code → city, state (Assuming a single area code belongs to a single city and thus a state)
- ZIP-code → city, state (Assuming a single ZIP code belongs to a single city and thus a state)
- st-addr, city, state → ZIP-code
- name, st-addr, city, state → SSNo (Assuming no two people of the same name live in the same place).

There can be other functional dependencies.

The **keys** are: {SSNo} and {name, st-addr, city, state}.

4. Consider a relation with schema $R(A, B, C, D, E)$ and FDs $AB \to C$, $C \to D$, $D \to B$, $D \to E$..

a) Indicate all the BCNF violations. Do not forget to consider FDs that are not in the given set, but follow from them. Only give violations that have a single attribute on the right side.

Answer. We compute the closure for each subset of attributes. First, we observe

$$\textbf{keys:} \qquad (AB)^+ = (AC)^+ = (AD)^+ = ABCDE.$$

Thus, AB, AC, and AD are superkeys (actually, keys). Any attribute subsets that contain either AB, AC, or AD (there are totally 11 such subsets) are superkeys.

superkeys:
$ABC, ABD, ABE, ACD, ACE, ADE, ABCD, ABCE, ABDE, ACDE, ABCDE.$

Now consider the other subsets of attributes.

non-superkeys:
$A^+ = A, \quad B^+ = B, \quad C^+ = BCDE, \quad D^+ = BDE, \quad E^+ = E,$
$(AE)^+ = AE, \quad (BC)^+ = BCDE, \quad (BD)^+ = BDE, \quad (BE)^+ = BE,$
$(CD)^+ = (CE)^+ = (BCD)^+ = (BCE)^+ = (CDE)^+ = (BCDE)^+ = BCDE,$
$(DE)^+ = (BDE)^+ = BDE.$

None of the above subsets is a superkey. Thus, for each such a non-superkey X, if $X \neq X^+$, then for each attribute A in $X^+ \setminus X$, $X \to A$ is a BCNF violation. For example, from $C^+ = BCDE$, we have BCNF violations $C \to B$, $C \to D$, $C \to E$, and from $D^+ = BDE$, we have BCNF violations $D \to B$, $D \to E$. There are totally 16 BCNF violations. However, all other 11 BCNF violations are "implied" by the above 5 BCNF violations.

b) Decompose the relation into a collection of relations in BCNF.

Answer. We first use the BCNF violation $C \to D$ to decompose the relation $R(A, B, C, D, E)$ and derive FDs for the resulting relations. We get

- $R_1 = C^+ = BCDE$. The FDs in R_1 can be derived using the closure information given in a), which are $\{C \to D, D \to B, D \to E\}$.
- $R_2 = AC$. Since the relation R_2 has only two attributes, by the discussion given in the class, R_2 is in BCNF. The relation R_2 has no nontrivial FDs.

The attribute C is the only key in the relation R_1. Thus, $D \to B$ is a BCNF violation in R_1. We decompose R_2 using $D \to B$, which gives

- $R_3 = D^+ = DBE$. The FDs in R_3 are $\{D \to B, D \to E\}$.
- $R_4 = CD$. This 2-attribute relation is in BCNF. The FD is $\{C \to D\}$.

The only key in R_3 is D. Thus, R_3 is also in BCNF. Thus, the result is

(1) $R_2 = AC$ with no nontrivial FDs,
(2) $R_3 = DBE$ with the FDs $\{D \to B, D \to E\}$, and
(3) $R_4 = CD$ with the FD $\{C \to D\}$.

5. For the relation schema and FDs given in Question 4:
a) Indicate all the 3NF violations.

Answer. By the discussion in Question 4, AB, AC, and AD are all keys of the relation R. Thus, A, B, C, and D are all primes. Therefore, a 3NF violator of R must be of the form $X \to E$, where X is a non-superkey. Thus, the 3NF violations in R are

3NF violations:

$$C \to E, \quad D \to E, \quad BC \to E, \quad BD \to E, \quad CD \to E, \quad BCD \to E,$$

Note that the last four are implied by the first two.

b) Decompose the relation into a collection of relations in 3NF.

Answer. Look at the FDs given for R: $\{AB \to C, \quad C \to D, \quad D \to B, \quad D \to E\}$. The FD basis is already minimized. To verify this, we need to check: (1) no dependency is implied by the other dependencies (for example, we cannot derive $C \to D$ using the other three dependencies); (2) we cannot eliminate any attribute from the left side of a dependency (only $AB \to C$ needs to be considered for this).

Therefore, using the algorithm given in the class (and in the textbook), we can decompose the relation R into four relations in the 3NF:

(1) $R_1 = ABC$ with FDs $\{AB \to C, C \to B\}$,
(2) $R_2 = CD$ with the FD $\{C \to D\}$,
(3) $R_3 = BD$ with the FD $\{D \to B\}$, and
(4) $R_4 = DE$ with the FD $\{D \to E\}$.

By the discussion we had in class, all these four relations are in 3NF. Note that the relation R_1 is not in BCNF ($C \to B$ is a BCNF violation) but is in 3NF.