תכנון מסדי נתונים:

- תהי r רלציה, תהי R הסכמה של r, ותהיינה $X,Y \square R$. אז r מקיימת את **התלות הפונקציונלית** $X \rightarrow Y$ (סימון: $r \square X \rightarrow Y$) אם לכל $t_1, t_2 \square r$ כך ש-$t_1[X]=t_2[X]$ מתקיים $t_1[Y]=t_2[Y]$

- אם $r \square X \rightarrow Y$ לכל מופע "חוקי" r של הסכמה R, נסמן $R \square X \rightarrow Y$ או פשוט $X \rightarrow Y$

F קבוצת תלויות פונקציונליות

- $r \square F$ אם לכל $X \rightarrow Y \square F$ מתקיים $r \square X \rightarrow Y$
- $F \square X \rightarrow Y$ אם לכל רלציה r, $r \square F \Rightarrow r \square X \rightarrow Y$
- X **מפתח-על** (superkey) של R אם $F \square X \rightarrow R$
- X **מפתח קביל** (candidate key) של R אם הוא מפתח-על, ולא קיים $Y \square X$ כך ש-Y מפתח-על של R
- **המפתח הראשי:** אחד מהמפתחות הקבילים
- **הסגור** של F הוא $F^+ \equiv \{X \rightarrow Y \mid F \square X \rightarrow Y\}$
- $X \rightarrow Y$ נמצאת ב-F^+ אםם כל רלציה שמקיימת את F מקיימת בהכרח גם את $X \rightarrow Y$

אקסיומות ארמסטרונג: כללי הסק ל-$F+$

A1 **רפלקסיביות** (reflexivity): אם $Y \square X \square R$ אז $X \rightarrow Y$
A2 **הכללה** (augmentation): אם $X \rightarrow Y$ ו-$Z \square R$ אז $XZ \rightarrow YZ$
A3 **טרנזיטיביות** (transitivity): אם $X \rightarrow Y$ ו-$Y \rightarrow Z$ אז $X \rightarrow Z$
A4 **איחוד** (union): אם $X \rightarrow Y$ ו-$X \rightarrow Z$ אז $X \rightarrow YZ$
A5 **פרוק** (decomposition): אם $X \rightarrow Y$ ו-$Z \square Y$ אז $X \rightarrow Z$
A6 **טרנזיטיביות-למחצה** (pseudo-transitivity): אם $X \rightarrow Y$ ו-$WY \rightarrow Z$ אז $WX \rightarrow Z$

הסגור (closure) של X בהינתן קב' תלויות פונקציונליות F הוא $X+F \equiv \{A \mid F \square X \rightarrow A\}$

אלגוריתם לחישוב $X+F$

```
Result := X;
    repeat
        for each Y→Z □ F
            if Y □ Result
                then Result := Result U Z
    until no more changes to Result.
```

כיסוי מינימלי של קבוצת תלויות: הגדרה

- קב' תלויות פונקציונליות F **מינימלית**
  אם לכל $X \rightarrow Y \in F$ מתקיים
  1. $|Y| = 1$ (תכונה יחידה באגפי ימין של כל התלויות)
  2. $(F - \{X \rightarrow Y\})^+ \neq F^+$ (אף תלות לא מיותרת)
  3. לכל $A \in X$ מתקיים $(F - \{X \rightarrow Y\} \cup \{X{-}A \rightarrow Y\})^+ \neq F^+$ (אף תכונה באגף שמאל של תלות לא מיותרת)

- **כיסוי מינימלי** של F: קב' מינימלית F_C כך ש-$F_C^+ = F^+$ (תמיד קיים בגלל פרוק/איחוד; לא בהכרח יחיד)

כיסוי מינימלי של קב' תלויות: אלגוריתם

1. פצל את התלויות ב-F לתכונות יחידות באגפי ימין

2. כל עוד יש שינויים ב-F בצע

   i. אם קיימת תלות $X \rightarrow B \in F$ כך ש-$B \in X^+_{F'}$
   (כאשר $F' = F - \{X \rightarrow B\}$), החלף את F ב-F'

   ii. אם קיימת תלות $X \rightarrow B \in F$
   ותכונה $A \in X$ כך ש-$B \in (X{-}A)^+_F$,
   החלף את F ב-$F - \{X \rightarrow B\} \cup \{X{-}A \rightarrow B\}$

שימור מידע: הגדרה

- **פרוק** (decomposition) של סכמה רלציונית R:
  קב' סכמות $\sigma = \{R_1, \ldots, R_n\}$ כך ש-$R = R_1 \cup \ldots \cup R_n$
- פרוק **משמר מידע** (lossless-join decomposition) של R בהינתן קבוצת תלויות פונקציונליות F:
  פרוק $\sigma = \{R_1, \ldots, R_n\}$ כך שלכל רלציה r מעל R,
  אם $r \models F$ אז $r = \pi_{R1} r \bowtie \ldots \bowtie \pi_{Rn} r$
- בפרוק שאינו משמר מידע עלול לקרות $r \subset \pi R1 r \bowtie \ldots \bowtie \pi Rn r$
  באי-שימור מידע לא מאבדים רשומות — להפך!
- בפרוק משמר מידע, אחת מתת-הסכמות חייבת להכיל מפתח של הסכמה

משפט: בהנתן קבוצת אטריביוטים U, בהנתן סכמה R[U], ובהנתן קבוצת תלויות פונקצונליות F, בהנתן פירוק חוקי של R ל-2 סכמות, R1[U1], R2[U2], הפירוק הוא משמר מידע אמ"מ קבוצת האטריביוטים $U1 \cap U2$ מפתח על של U1 או U2 או של שניהם.

שימור תלויות: הגדרה

- **ההיטל** של קב' תלויות פונקציונליות F על תת-סכמה
$R_i \square R$ הוא $\pi_{R_i} F = \{X\rightarrow Y \mid X\rightarrow Y \square F^+ \square X \square Y \square R_i\}$ כלומר ההיטל של F על V היא קבוצת התלויות מתוך F+ הרלוונטיות ל- V.
- פרוק $\sigma=\{R_1,\ldots,R_n\}$ **משמר תלויות** בהינתן F אם לכל $r_1,\ldots,r_n$ מעל תת-הסכמות $R_1,\ldots,R_n$, בהתאמה, אם $r_i \square \pi_{R_i} F$ לכל $i=1,\ldots,n$ אז $r_1 \square \ldots \square r_n \square F$. כלומר, פירוק הוא משמר תלויות אם האיחוד של כל ההיטלים של F מעל קבוצות האטריביוטים מוכיח את F.

- **משפט:** פרוק $\sigma=\{R_1,\ldots,R_n\}$ משמר תלויות בהינתן F
אםם $(\pi_{R1} F \square \ldots \square \pi_{Rn} F)^+ = F^+$
- **טענה:** אם $X \square R_i$ אז $X^+_{\pi Ri\, F} = X^+_F \cap R_i$

שימור תלויות: אלגוריתם לבדיקה

- בהינתן פרוק $\sigma=\{R_1,\ldots,R_n\}$ ותלויות פונקציונליות F, לכל תלות $X\rightarrow Y \square F$ חשב את $X^+{}_{\pi R1\, F \,\square \ldots \square\, \pi Rn\, F}$ ע"י:

$Z := X;$
repeat
for $i = 1,\ldots,n$ **do**
$Z := Z \cup (((Z \cap Ri)+F) \cap Ri)$
until no more changes to Z.

$X\rightarrow Y$ נשמרת אםם בסיום $Y \square Z$.
σ משמר תלויות אםם כל תלות $X\rightarrow Y \square F$ נשמרת.

BCNF: הגדרה

- סכמה רלציונית R היא ב-**Boyce-Codd Normal Form** בהינתן קב' תלויות פונקציונליות F אם לכל $X\rightarrow Y \square F^+$ לא טריוויאלית, X מפתח-על של R

אלגוריתם פרוק ל-BCNF

- כל עוד קיימת סכמה R ותלות לא טריוויאלית $X\rightarrow Y \square \pi_R F$ כך ש-X איננו מפתח-על של R, פרק את R לשתי תת-סכמות: $X \square Y$ ו-$R-(Y-X)$
הפרוק משמר מידע
אינו מבטיח שימור תלויות (גם אם קיים פרוק משמר תלויות ל-BCNF!)
אי-דטרמיניסטי—יכול למצוא פרוקים שונים

3NF (Third Normal Form): הגדרה

- סכמה רלציונית R היא ב-**3NF** בהינתן קב' תלויות פונקציונליות F אם לכל $X\rightarrow Y \square F^+$ לא טריוויאלית, X מפתח-על של R **או שכל תכונות Y נמצאות במפתחות קבילים של R**
- כל סכמה ב-BCNF היא גם ב-3NF, לא בהכרח להפך
- 3NF מאפשרת כפילויות ש-BCNF מונעת
- תמיד קיים פרוק משמר מידע **ותלויות** ל-3NF

אלגוריתם פרוק ל-3NF

1. מצא כיסוי מינימלי לקבוצת התלויות הנתונה
2. לכל קבוצת תלויות $X\rightarrow A_1, \ldots, X\rightarrow A_n$ עם אותו אגף שמאל X, אחד את התלויות ל-$X\rightarrow A_1\ldots A_n$
3. אם קיימת תלות $X\rightarrow A_1\ldots A_n$ המכילה את כל תכונות R אז R כבר ב-3NF — **עצור**
4. לכל תלות $X\rightarrow A_1\ldots A_n$ צור תת-סכמה $XA_1\ldots A_n$
5. אם אף תת-סכמה אינה מכילה מפתח קביל של R, הוסף תת-סכמה שהיא מפתח קביל כלשהו של R

- עוצר ומוצא פרוק משמר מידע **ותלויות**
- כל תת-סכמה R_i ב-3NF בהינתן $\pi_{R_i} F$

D קבוצת תלויות (פונקציונליות+רב-ערכיות)

- $r \models D$ אם לכל $d \in D$ מתקיים $r \models d$
- $D \models d$ (כאשר d תלות כלשהי) אם לכל רלציה r, $r \models D \;\Rightarrow\; r \models d$
- **הסגור** של D הוא $D^+ \equiv \{d \mid D \models d\}$ (תלות פונקציונלית/רב-ערכית d נמצאת ב-D^+ אמם כל רלציה שמקיימת את D מקיימת בהכרח גם את d)

כללי הסק לתלויות פונק' + רב-ערכיות

A1–A3 עדיין בתוקף **(רק לתלויות פונקציונליות!)**

A4 **השלמה** (complementation):
אם $X \twoheadrightarrow Y$ אז $X \twoheadrightarrow R{-}X{-}Y$

A5 **הכללה רב-ערכית** (multi-valued aug.):
אם $X \twoheadrightarrow Y$ ו-$V \subseteq W$ אז $WX \twoheadrightarrow VY$

A6 **טרנזיטיביות רב-ערכית** (multi-valued trans.):
אם $X \twoheadrightarrow Y$ ו-$Y \twoheadrightarrow Z$ אז $X \twoheadrightarrow Z{-}Y$

A7 **הכפלה** (doubling): אם $X\rightarrow Y$ אז $X \twoheadrightarrow Y$

A8 **מיזוג** (merging): אם $X \twoheadrightarrow Y$, $Z \subseteq Y$,
וקיימת W כך ש-$W\cap Y=\emptyset$ ו-$W\rightarrow Z$, אז $X\rightarrow Z$

$D \vdash d$: אפשר להוכיח את d מ-D ע"י A1–A8

- **איחוד רב-ערכי** (multi-valued union):
  אם $X \twoheadrightarrow Y$ ו-$X \twoheadrightarrow Z$ אז $X \twoheadrightarrow YZ$
  (ראינו הוכחה בשקף הקודם)
- **חיתוך רב-ערכי** (multi-valued intersection):
  אם $X \twoheadrightarrow Y$ ו-$X \twoheadrightarrow Z$ אז $X \twoheadrightarrow Y\cap Z$
- **הפרש רב-ערכי** (multi-valued difference):
  אם $X \twoheadrightarrow Y$ ו-$X \twoheadrightarrow Z$ אז $X \twoheadrightarrow Y{-}Z$

4NF (Fourth Normal Form)

- סכמה רלציונית R היא ב-**4NF** בהינתן קב' תלויות פונקציונליות ורב-ערכיות D אם לכל תלות $X \rightarrow Y \in D^+$ או $X \twoheadrightarrow Y \in D^+$ לא טריוויאלית, X מפתח-על של R

פרוק ל-4NF

- בדומה לפרוק ל-BCNF, תמיד קיים פרוק משמר מידע ל-4NF; לא תמיד קיים פרוק משמר תלויות
- **אלגוריתם** לפרוק משמר מידע ל-4NF:
  כל עוד קיימת סכמה R
  ותלות לא טריוויאלית $X \twoheadrightarrow Y \in \pi_R D$
  כך ש-$X \cap Y = \emptyset$ ו-X איננו מפתח-על של R,
  פרק את R לשתי תת-סכמות: $X \cup Y$ ו-$R-Y$