COSC265 — Relational Database Systems

Neville Churcher

Department of Computer Science & Software Engineering
University of Canterbury

2021

Multi-Valued Dependencies

MVDs arise because of the way attributes are structured into relations. Unlike FDs, they are not essentially properties of the information represented by the relations.

Definition (MVD)

If $X, Y \subseteq R$ then $X \twoheadrightarrow Y$ (X multi-determines Y) if for any given value of X there is a set of associated values for Y which is *independent* of the values of $R - X - Y$

Definition (Alternative MVD Definition)

If two tuples agree on their X values, we can exchange their Y values to obtain two tuples that must also occur, *whatever the values of the other attributes*

- ☆ Every FD is also a MVD
- ☆ This is the special case where RHS has only one value for given LHS value
- ☆ The reverse is not true in general — an MVD is not usually also an FD
- ☆ e.g. $Employee \twoheadrightarrow Project$ but $Employee \not\rightarrow Project$

In More Detail...

Definition (Yet Another Version)

$X \twoheadrightarrow Y$ holds in R if $\forall r(R)$ when $\exists t, s \in r, t[X] = s[X]$ then $\exists u, v \in r$ where:

- ☆ $u[X] = v[X] = t[X] = s[X]$
- ☆ $u[Y] = t[Y]$ and $u[R - X - Y] = s[R - X - Y]$
- ☆ $v[Y] = s[Y]$ and $v[R - X - Y] = t[R - X - Y]$

- ☆ *MVDs* are *tuple-generating* dependencies
- ☆ *FDs* are *equality-generating* dependencies.

Example (*Course*, *Lecturer*, *Text*)

A course may be taught by any of the specified lecturers, and uses all of the specified texts. The same texts are used no matter who teaches the course. A given lecturer or text may be associated with many courses.

$R = \{C, L, T\}$ $\mathcal{F} = \varnothing$

Course	**Lecturer**	**Text**
Physics	Jones	Mechanics
Physics	Jones	Optics
Physics	Smith	Mechanics
Physics	Smith	Optics
Maths	Jones	Mechanics
Maths	Jones	Algebra
Maths	Jones	Calculus

- ☆ $C \twoheadrightarrow L$, $C \twoheadrightarrow T$ hold
- ☆ $C \twoheadrightarrow (R - C - L)$, $C \twoheadrightarrow (R - C - T)$
- ☆ Some redundancy evident
- ☆ BCNF — "all-key" relation

CLT (Continued)

☆ Problem arises because lecturer and text are *independent*

☆ Can (intuitively?) solve by decomposition:

Course	**Lecturer**
Physics	Jones
Physics	Smith
Maths	Jones

Course	**Text**
Physics	Mechanics
Physics	Optics
Maths	Mechanics
Maths	Algebra
Maths	Calculus

☆ MVDs always occur in pairs, as in this example.

☆ $Course \twoheadrightarrow Lecturer$, $Course \twoheadrightarrow Text$ may be written as $Course \twoheadrightarrow Lecturer | Text$

MVDs and Inference

Given $\mathcal{D} = \{\mathcal{F}, \mathcal{M}\}$, the set of FDs and MVDs, we can compute closures, derive dependencies etc.

Complementation Axiom $X \twoheadrightarrow Y \models X \twoheadrightarrow (R - X - Y)$

Augmentation Axiom If $X \twoheadrightarrow Y$ holds and $V \subseteq W$ then $WX \twoheadrightarrow VY$

Transitivity Axiom $\{X \twoheadrightarrow Y, Y \twoheadrightarrow Z\} \models X \twoheadrightarrow (Z - Y)$

Union Rule $\{X \twoheadrightarrow Y, X \twoheadrightarrow Z\} \models X \twoheadrightarrow YZ$

Pseudotransitivity Rule $\{X \twoheadrightarrow Y, WY \twoheadrightarrow Z\} \models XW \twoheadrightarrow (Z - WY)$

Decomposition Rule If $X \twoheadrightarrow Y$ and $X \twoheadrightarrow Z$ hold then $X \twoheadrightarrow (Y \cap Z), X \twoheadrightarrow (Y - Z)$, and $X \twoheadrightarrow (Z - Y)$ also hold.

Mixing FDs and MVDs

Replication Axiom $\{X \to Y\} \models X \twoheadrightarrow Y$

Coalescence Axiom If $X \twoheadrightarrow Y$ and $W \to Z$ hold, where $Z \subseteq Y$, and $W \cap Y = \varnothing$, then $X \to Z$ also holds.

Mixed Pseudotransitivity Rule $\{X \twoheadrightarrow Y, XY \to Z\} \models X \to (Z - Y)$

Lossless-Joins Revisited

Remember that for FDs we had a lossless join test for decompositions into two schemes

- ✰ Replication Axiom $\{X \rightarrow Y\} \models X \twoheadrightarrow Y$ allows extension of previous test.
- ✰ A decomposition $\rho_{\mathcal{D}}(R) = (R_1, R_2)$ is lossless if:

  $R_1 \cap R_2 \twoheadrightarrow R_1 - R_2$ or $R_1 \cap R_2 \twoheadrightarrow R_2 - R_1$
- ✰ In other words: $R = ABC$ can be non-loss decomposed into $R_1 = AB$ and $R_2 = AC$ iff $A \twoheadrightarrow B|C$ holds in R

Example (CLT: Lossless-Join Test)

$R = \{Course, Lecturer, Text\}$,
$R_1 = \{Course, Lecturer\}, R_2 = \{Course, Text\}$

$$R_1 \cap R_2 = Course, R_1 - R_2 = Lecturer$$

$$Course \twoheadrightarrow Lecturer \in \mathcal{D}$$

$\therefore \rho_{\mathcal{D}}(R)$ non-loss

4th Normal Form

This looks familiar — remember BCNF?

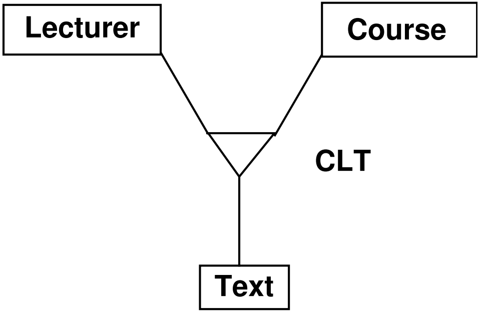
- ☆ A relation scheme is 4NF if all determinants(FD or MVD) are superkeys
- ☆ A relation scheme is 4NF if it is BCNF and all MVDs are also FDs
- ☆ If $X \twoheadrightarrow Y$ then $X \rightarrow R$
- ☆ Assuming *non-trivial* MVDs i.e. $XY \neq R$ and $X \not\supseteq Y$
- ☆ The only dependencies (MVD or FD) are from a candidate key to some other attribute(s)
- ☆ All 4NF schemes are also BCNF

Example (CLT: 4NF Test)

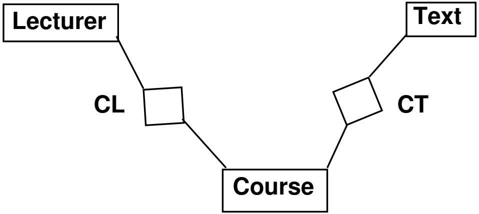
$\{Course, Lecturer, Text\}$ is not in 4NF since *Course* is not a key. However, both $\{Course, Lecturer\}$ and $\{Course, Text\}$ are 4NF (contain only trivial MVDs)

In Conceptual Model Terminology

Genuine Ternary Relationship?



or Pair of Binary Relationships?



Points to Ponder. . .

- Complementation for MVDs

$$X \twoheadrightarrow Y \models X \twoheadrightarrow (R - X - Y)$$

  has no counterpart for FDs.
- Reflexivity for FDs *appears* to have no direct counterpart for MVDs. However, combining it with replication

$$X \rightarrow Y \models X \twoheadrightarrow Y$$

  leads to

$$X \twoheadrightarrow Y \quad \text{if} \quad Y \subseteq X$$

- Transitivity is more restrictive for MVDs.

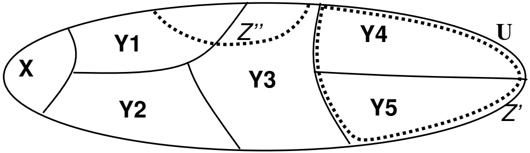
$$\{X \twoheadrightarrow Y, Y \twoheadrightarrow Z\} \models X \twoheadrightarrow (Z - Y)$$

  but

$$\{X \twoheadrightarrow Y, Y \twoheadrightarrow Z\} \not\models X \twoheadrightarrow Z$$

Dependency Basis

- ☆ The decomposition rule for MVDs is stronger than corresponding FD rule
- ☆ For FDs, $X \rightarrow Y, A \subseteq Y \models X \rightarrow A$
- ☆ For MVDs, we can only conclude $X \twoheadrightarrow A$ from $X \twoheadrightarrow Y$ if we can find some Z such that $X \twoheadrightarrow Z$ and $Z \cap Y = A$ or $Y - Z = A$ or $Z - Y = A$ holds
- ☆ The MVD decomposition rule and union rule lead to the definition of $dep_{\mathcal{D}}(X)$ — the dependency basis for X.
- ☆ $dep_{\mathcal{D}}(X)$ is a partition of $\mathcal{U}$ and represents a statement about the sets $Y : X \twoheadrightarrow Y$



- ☆ $X \twoheadrightarrow Z$ iff Z is the union of one or more of the Y_i
- ☆ $X \twoheadrightarrow Z'$ since $Z' = Y_4 \cup Y_5$ but $X \not\twoheadrightarrow Z''$

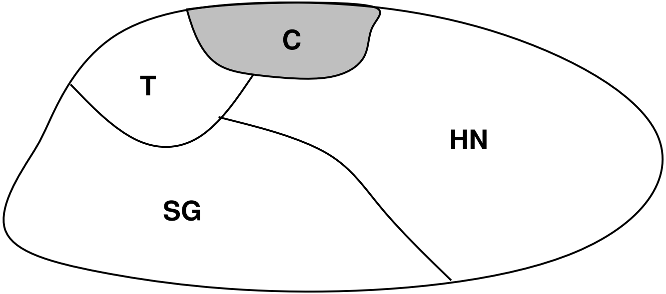
Dependency Basis Example

- ☆ Start with set of MVDs, $\mathcal{M}$, containing MVDs corresponding to a minimal cover of FDs, plus any other MVDs.
- ☆ For example from handout, $R = CTHNSG$,

$$\mathcal{M} = \left\{ \begin{array}{lll} C \twoheadrightarrow HN & C \twoheadrightarrow T & HN \twoheadrightarrow C \\ HT \twoheadrightarrow N & CS \twoheadrightarrow G & HS \twoheadrightarrow N \end{array} \right\}$$

- ☆ $C \twoheadrightarrow HN \models C \twoheadrightarrow TSG$ (complementation)
- ☆ $C \rightarrow T \models C \twoheadrightarrow T$ (replication)
- ☆ Decomposition Rule (with $X \equiv C, Y \equiv TSG, Z \equiv T$)
  $\{X \twoheadrightarrow Y, X \twoheadrightarrow Z\} \models \{X \twoheadrightarrow (Y \cap Z), X \twoheadrightarrow (Y - Z), X \twoheadrightarrow (Z - Y)\}$
- ☆ $\{C \twoheadrightarrow TSG, C \twoheadrightarrow T \models C \twoheadrightarrow (TSG - T)\}$. So $C \twoheadrightarrow SG$
- ☆ $dep_{\mathcal{M}}(C) = \{T, HN, SG\}$

Using $dep_{\mathcal{M}}(C)$



- ☆ To test $\mathcal{M} \models C \twoheadrightarrow Y$ check whether $Y - C$ is the union of members of $dep_{\mathcal{M}}(C)$
- ☆ $C \twoheadrightarrow HNSG$ since $HNSG = HN \cup SG$
- ☆ $C \not\twoheadrightarrow TH$
- ☆ $C \twoheadrightarrow CSGT$

Decomposition into 4NF

Every determinant is a candidate key

$$\mathcal{M} = \left\{ \begin{array}{lll} C \twoheadrightarrow HN & C \twoheadrightarrow T & HN \twoheadrightarrow C \\ HT \twoheadrightarrow N & CS \twoheadrightarrow G & HS \twoheadrightarrow N \end{array} \right\}$$

- ☆ $\rho^{(0)} = \{CTHNSG\}$
- ☆ $\rho^{(0)}$ not 4NF since $C \twoheadrightarrow HN$ but $C \not\to CTHNSG$
  i.e. C is a determinant but not a candidate key (only SH is (DIY))
- ☆ Replace by CHN and $(CTHNSG - HN) = CTSG$
- ☆ $\rho^{(1)} = \{CHN, CTSG\}$
- ☆ CHN is in 4NF with key HN
- ☆ The MVD $C \twoheadrightarrow HN$ is OK as LHS + RHS includes all attributes
- ☆ $CTSG$ is not in 4NF. Key is CS but $C \to T \models C \twoheadrightarrow T$ so have determinant that isn't a candidate key.
- ☆ Replace by CT and $(CTSG - T) = CSG$
- ☆ $\rho^{(1)} = \{CHN, CT, CSG\}$ has all schemes in 4NF

Sanity Check

Each of the decompositions is non-loss

- ☆ $\rho^{(0)} \Rightarrow \rho^{(1)} : \{CTHNSG\} \Rightarrow \{CHN, CTSG\}$
  $CHN \cap CTSG = C$
  $CHN - CTSG = HN$
  $C \twoheadrightarrow HN \in \mathcal{M}$,
  $\therefore$ non-loss
- ☆ $\rho^{(1)} \Rightarrow \rho^{(2)} : \{CTSG\} \Rightarrow \{CT, CSG\}$
  $CT \cap CSG = C$
  $CT - CSG = T$
  $C \twoheadrightarrow T \in \mathcal{M}$,
  $\therefore$ non-loss

Join Dependencies

- ☆ $\bowtie (R_1, R_2, \ldots, R_k)$ satisfied iff $m_\rho(r) = r$ where $\rho = (R_1, R_2, \ldots, R_k)$
- ☆ Thus a JD states that a relation is the join of its projections

Definition (5NF)

Scheme R is in 5NF (also known as projection/join normal form) if every JD $\bowtie (R_1, R_2, \ldots, R_k)$ is trivial (i.e. $R = R_i$ for some $1 \leq i \leq k$)

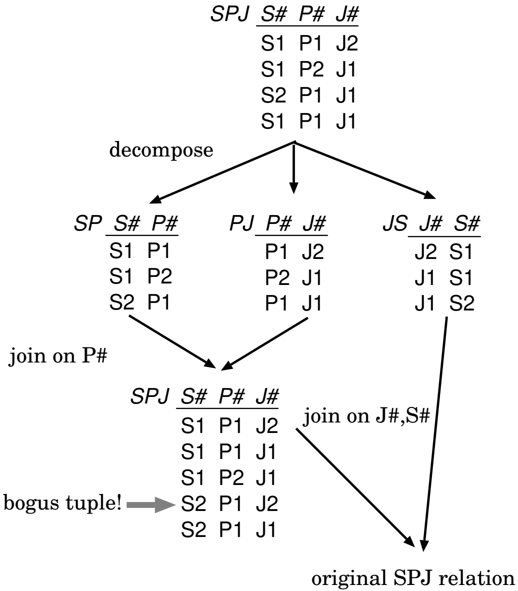
or

every R_i is a superkey for R.

☞ Finding JDs is hard and so is the implication problem for dependency sets including JDs.

JD Example

- Supplier, Part, proJect example
- $\bowtie(S\#P\#, P\#J\#, J\#S\#)$ holds
- *SPJ* is the join of all three of its binary projections *but not of any two*



Interpreting JDs

☆ If the ternary relation $S\#P\#J\#$ is decomposed then $\bowtie(S\#P\#, P\#J\#, J\#S\#)$ represents a constraint.

☆ e.g.

if supplier S1 supplies part P4
and P4 is used on project J2
and S1 supplies J2
then S1 supplies P4 to J2

☆ This is the kind of constraint required to distinguish between relationships *'supplies part to project'* and *'could supply part to project'*