Functional Dependencies

- Need a special type of constraint to help us with normalization
- $X \rightarrow Y$ is an assertion about a relation R that whenever two tuples of R agree on all the attributes in set X, they must also agree on all attributes in set Y.
- E.g., suppose X = {AB}, Y = {C}

R

A	B	C
x1	y1	c2
x1	y1	c2
x2	y2	c3
x2	y2	c3

Splitting Right Sides of FDs

- $X \rightarrow A_1A_2\ldots A_n$ holds for R exactly when each of $X \rightarrow A_1$, $X \rightarrow A_2, \ldots,$ $X \rightarrow A_n$ hold for R.
- Example: $A \rightarrow BC$ is equivalent to $A \rightarrow B$ and $A \rightarrow C$.
- Combining: if $A \rightarrow F$ and $A \rightarrow G$, then $A \rightarrow FG$

- There is no splitting rule for the left side
  - $ABC \rightarrow DEF$ is NOT the same as $AB \rightarrow DEF$ and $C \rightarrow DEF$!
- We'll generally express FDs with singleton right sides.

Trivial FDs

- Not all functional dependencies are useful
  - A → A always holds
  - ABC → A also always holds (right side is subset of left side)
- FD with an attribute on both sides
  - ABC → AD becomes ABC → D
  - Or, in singleton form, delete trivial FDs

    ABC → A and ABC → D becomes just ABC → D

FDs are a generalization of keys

- Functional dependency: $X \rightarrow Y$
- Superkey: $X \rightarrow R$
- A superkey must include all the attributes of the relation on the RHS.
- An FD can involve just a subset of them
  - Example:

Houses (street, city, value, owner, tax)

- street,city → value, owner, tax *(both FD and key)*
- city,value → tax *(FD only)*

Identifying functional dependencies

- FDs are domain knowledge
  - Intrinsic features of the data you're dealing with
  - Something you know (or assume) about the data
- Database engine cannot identify FDs for you
  - Designer must specify them as part of schema
  - DBMS can only enforce FDs when told to
- DBMS cannot "optimize" FDs either
  - It has only a finite sample of the data
  - An FD constrains the entire domain

Armstrong's Axioms

X, Y, Z are sets of attributes

1. **Reflexivity:** If $Y \subseteq X$, then $X \to Y$
2. **Augmentation:** If $X \to Y$, then $XZ \to YZ$ for any Z
3. **Transitivity:** If $X \to Y$ and $Y \to Z$, then $X \to Z$
4. **Union:** If $X \to Y$ and $X \to Z$, then $X \to YZ$
5. **Decomposition:** If $X \to YZ$, then $X \to Y$ and $X \to Z$

Inferring FDs

- Given a set of FDs, we can often infer further FDs.
- This will come in handy when we apply FDs to the problem of database design.

Dependency Inference

- Suppose we are given FDs
  $X_1 \rightarrow A_1$,
  $X_2 \rightarrow A_2$,
  …,
  $X_n \rightarrow A_n$.
- Does the FD $Y \rightarrow B$ also hold in any relation that satisfies the given FDs?
- Example: If $A \rightarrow B$ and $B \rightarrow C$ hold, surely $A \rightarrow C$ holds, even if we don't say so.
  $A \rightarrow C$ is ***entailed (implied)*** by $\{A \rightarrow B\ ,\ B \rightarrow C\ \}$

Transitive Property

The transitive property holds for FDs

- Consider the FDs: $A \rightarrow B$ and $B \rightarrow C$; then $A \rightarrow C$ holds
- Consider the FDs: $AD \rightarrow B$ and $B \rightarrow CD$; then $AD \rightarrow CD$ holds or just $AD \rightarrow C$ (because of trivial FDs)

Method 1: Prove it from first principles

- To test if $Y \rightarrow B$, start by assuming two tuples agree on all attributes of Y.

$\leftarrow Y \rightarrow$

t1: aaaaa bb. . . b

t2: aaaaa ?? . . . ?

Example

ClientID	Income	OtherProd	Rate	Country	City	State
225	High	A	2.1%	USA	San Francisco	MD
420	High	A	2.1%	USA	San Francisco	CA
333	High	B	3.0%	USA	San Francisco	CA
576	High	B	3.0%	USA	San Francisco	CA
128	Low	C	4.5%	UK	Reading	Berkshire
193	Low	C	4.5%	UK	London	London
550	Low	B	3.5%	UK	London	London

F1: [Income, OtherProd] → [Rate]

F2: [Country, City] → [State]

How to prove it in the general case?

Closure Test for FDs

- Given attribute set Y and FD set F
  - Denote Y_F^+ or Y^+ the closure of Y relative to F

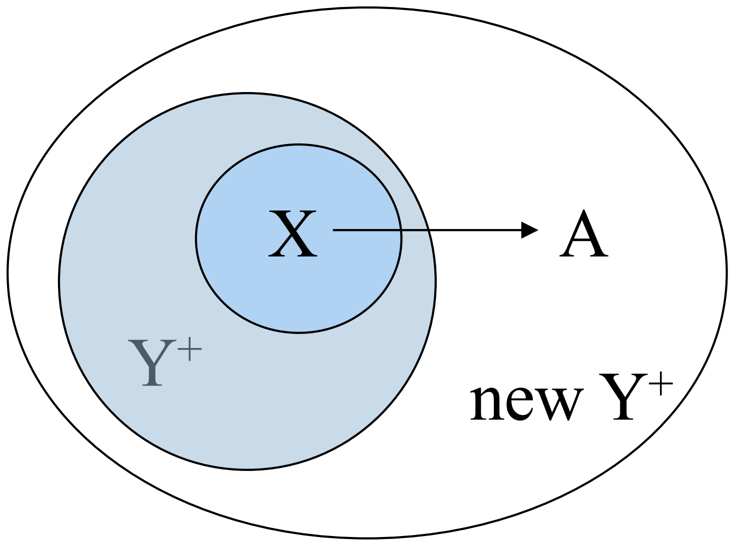
    Y_F^+ = set of all FDs given or implied by Y
- Computing the closure of Y
  - Start: $Y_F^+ = Y$, $F' = F$
  - While there exists an $f \in F'$ s.t. $LHS(f) \subseteq Y_F^+$:

    $Y_F^+ = Y_F^+ \cup RHS(f)$

    $F' = F' - f$
  - At end: $Y \rightarrow B$ for all $B \in Y_F^+$

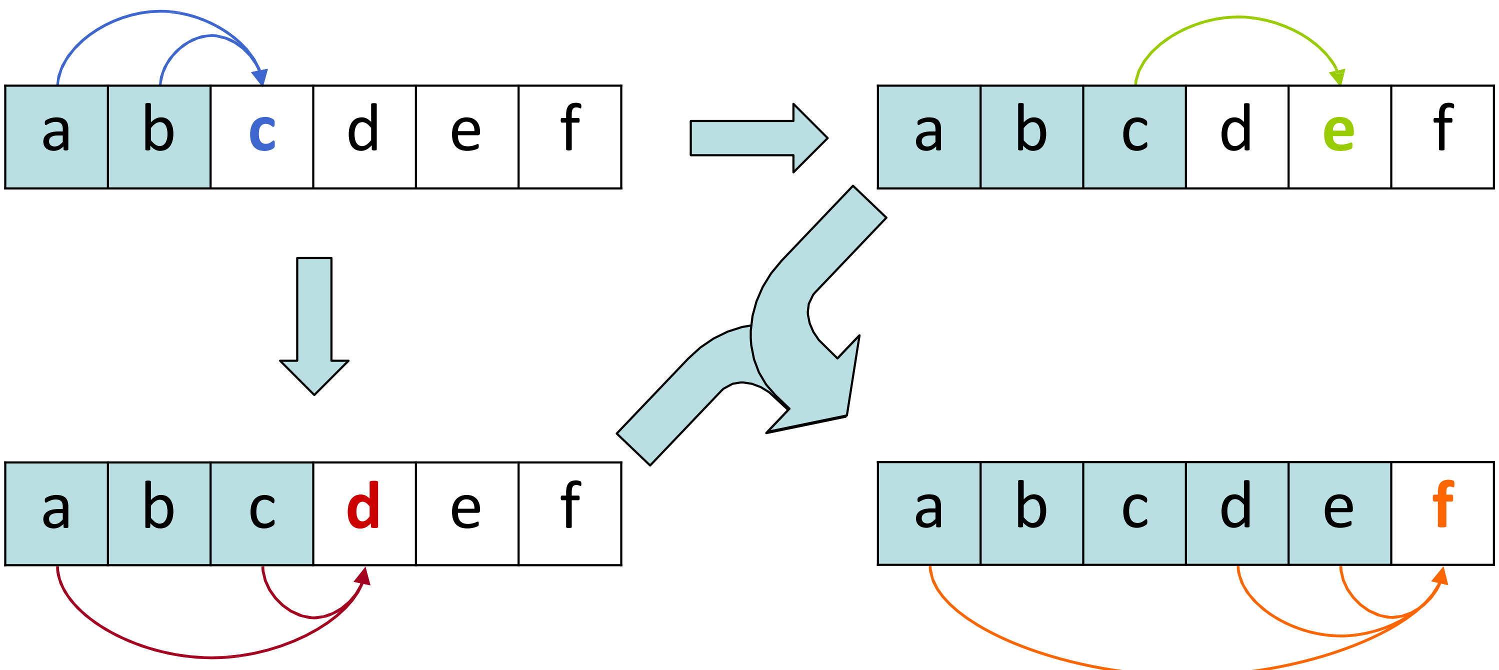
Acknowledgements: M. Papagelis

Computing the closure Y^+ of a set of attributes Y
Given FDs F:



Example: Closure Test

- Consider R(a,b,c,d,e,f)
  with FDs ab → c, ac → d, c → e, ade → f
- Find Y^+ if Y = ab or find $\{a,b\}^+$



$\{a,b\}^+=\{a,b,c,d,e,f\}$ or *ab → cdef*

ab is a candidate key!

Your Turn: Closure Test

F:

$AB \rightarrow C$
$A \rightarrow D$
$D \rightarrow E$
$AC \rightarrow B$

$\boldsymbol{X}$	$\boldsymbol{X_F^+}$
A	$\{A, D, E\}$
AB	$\{A, B, C, D, E\}$
AC	$\{A, C, B, D, E\}$
B	$\{B\}$
D	$\{D, E\}$

Is $AB \rightarrow E$ entailed by ***F***? *Yes*

Is $D \rightarrow C$ entailed by ***F***? *No*

Result: X_F^+ allows us to determine all FDs of the form $X \rightarrow Y$ entailed by ***F***