Week 10 Tutoring

CSE 180

Final

Tuesday, 12/12 8:00-11:00 am (3 hours)

- The Final is cumulative, covering the entire quarter, with somewhat greater emphasis on later material.
- Bring:
  - Red Scantron
  - Cheat sheet

Design Theory

- A set of design principles that allows one to decide what constitutes a “good” or “bad” database schema design
- A set of algorithms for modifying a “bad” design to a “better” one

How to make a good schema design?

- Remove redundancies

Functional Dependencies

- The information that rank **determines** salary_scale is a type of integrity constraint known as a functional dependency (FD).
- Functional dependencies can help us detect anomalies that may exist in a given schema
- $X \rightarrow Y$ = "X determines Y"
- A relation instance r of R satisfies the FD $X \rightarrow Y$ if for every pair of tuples t and t in r:

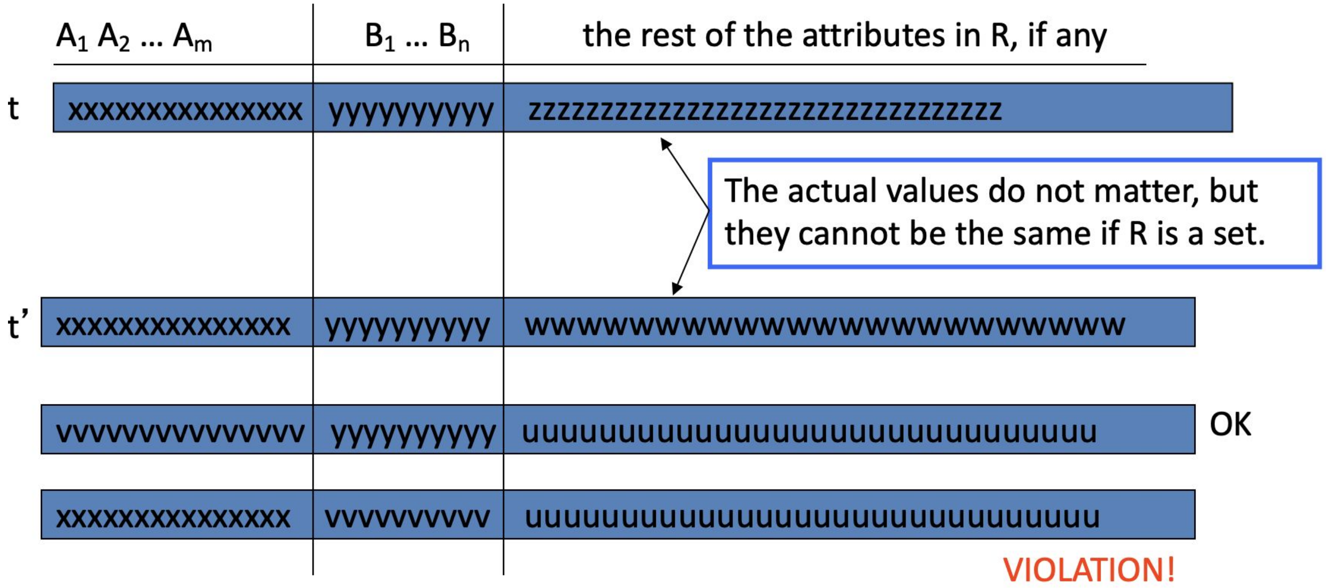
If $\pi_X(t) = \pi_X(t')$ holds, then $\pi_Y(t) = \pi_Y(t')$ also holds.

- That is, if the two tuples t and t' agree on the the values of all the attributes in X, then the two tuples t and t' must also agree on the values of all the attributes in Y.

FD example

This is a violation because $A_1 \ldots A_m \rightarrow B_1 \ldots B_m$ therefore, the rest of the attributes in R should match as well.

In this case zzzz… does not match wwww…



Implications

We say that a set $\mathcal{F}$ of FDs implies an FD F if for every instance r that satisfies $\mathcal{F}$, it must also be true that r satisfies F.

Notation: $\mathcal{F} \vDash F$

Armstrong’s Axioms

Let X, Y, and Z denote sets of attributes over a relation schema R.

- Reflexivity: If $Y \subseteq X$, then $X \rightarrow Y$.
    - FDs in this category are called trivial FDs.
- Augmentation: If $X \rightarrow Y$, then $XZ \rightarrow YZ$ for any set Z of attributes.
- Transitivity: If $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow Z$.

Derived from Armstrong’s axioms

- Union: If $X \rightarrow Y$ and $X \rightarrow Z$, then $X \rightarrow YZ$.
- Decomposition: If $X \rightarrow YZ$, then $X \rightarrow Y$ and $X \rightarrow Z$
- Pseudo-Transitivity: If $X \rightarrow Y$ and $WY \rightarrow Z$, then $XW \rightarrow Z$.

Rules cont.

We use the notation $\mathcal{F} \vdash F$ to mean that F can be derived from F using Armstrong's axioms.

VS. $\mathcal{F} \vDash F$: We say that a set $\mathcal{F}$ of FDs implies an FD F if for every instance r that satisfies $\mathcal{F}$, it must also be true that r satisfies F.

Completeness and Soundness of Armstrong's Axioms

Completeness: If a set $\mathcal{F}$ of FDs implies F, then F can be derived from $\mathcal{F}$ using Armstrong's axioms.

$$\text{If } \mathcal{F} \vDash F, \text{ then } \mathcal{F} \vdash F.$$

Soundness: If F can be derived from a set of FDs $\mathcal{F}$ using Armstrong's axioms, then $\mathcal{F}$ implies F.

$$\text{If } \mathcal{F} \vdash F, \text{ then } \mathcal{F} \vDash F.$$

With Completeness and Soundness, we know that

$$\mathcal{F} \vdash F \text{ if and only if } \mathcal{F} \vDash F$$

Closure of a Set of FD $\mathcal{F}$

Let $\mathcal{F}^+$ denote the set of all FDs implied by a given set $\mathcal{F}$ of FDs.

- Also called the closure of $\mathcal{F}$

To decide whether $\mathcal{F}$ implies F, first compute $\mathcal{F}^+$, then see whether F is a member of $\mathcal{F}^+$.

Attribute Closure Algorithm

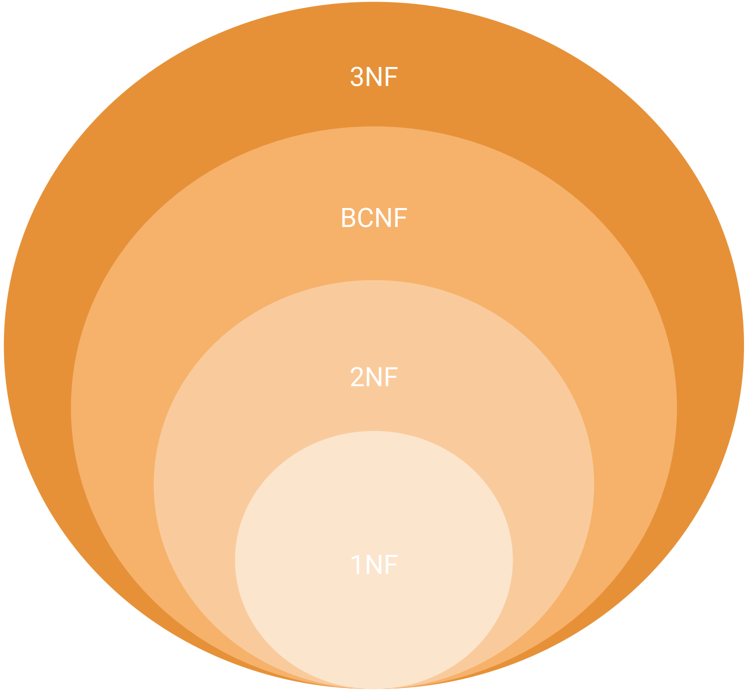
Let X be a set of attributes, and $\mathcal{F}$ be a set of FD s.

The **Attribute Closer X^+ with respect to $\mathcal{F}$** is the set of all attributes A such that $X \rightarrow A$ is derivable from $\mathcal{F}$

Should be clear that: If $A \in$ "Closure" (that is, if $A \in X^+$) , then $X \rightarrow A$.
More strongly: $\mathcal{F} \vdash X \rightarrow A$ if and only if $A \in X^+$.

Normal Forms

- First Normal Form (1NF)
  - Every attribute is atomic
- Second Normal Form (2NF) -- Not important
- **Boyce-Codd Normal Form (BCNF)**
- **Third Normal Form (3NF)**



BCNF and 3NF

R is in _NF if for every FD X → A in F, at least one of following is true:

- X → A is a trivial FD (i.e., A ∈ X), or
- X is a superkey for R, or
- A is a Prime Attribute. That is, A is part of some key of R

1NF

BCNF

3NF

Trivial

Trivial FD

- $A \rightarrow A$, $B \rightarrow B$, $C \rightarrow C$, $AB \rightarrow A$, $AB \rightarrow B$, $BC \rightarrow B$, $BC \rightarrow C$, $AC \rightarrow A$, $AC \rightarrow C$, $ABC \rightarrow A$, $ABC \rightarrow B$, $ABC \rightarrow C$, $ABC \rightarrow AB$, $ABC \rightarrow AC$, $ABC \rightarrow BC$, $ABC \rightarrow ABC$

FD Example 2 using Attribute Closure

- $\mathcal{F}$ = { AB $\rightarrow$ E, B $\rightarrow$ AC, BE $\rightarrow$ C }
- Question: Does BC $\rightarrow$ E?
- Compute $(BC)^+$

We'll write $(BC)^+$
Instead of $\{B,C\}^+$

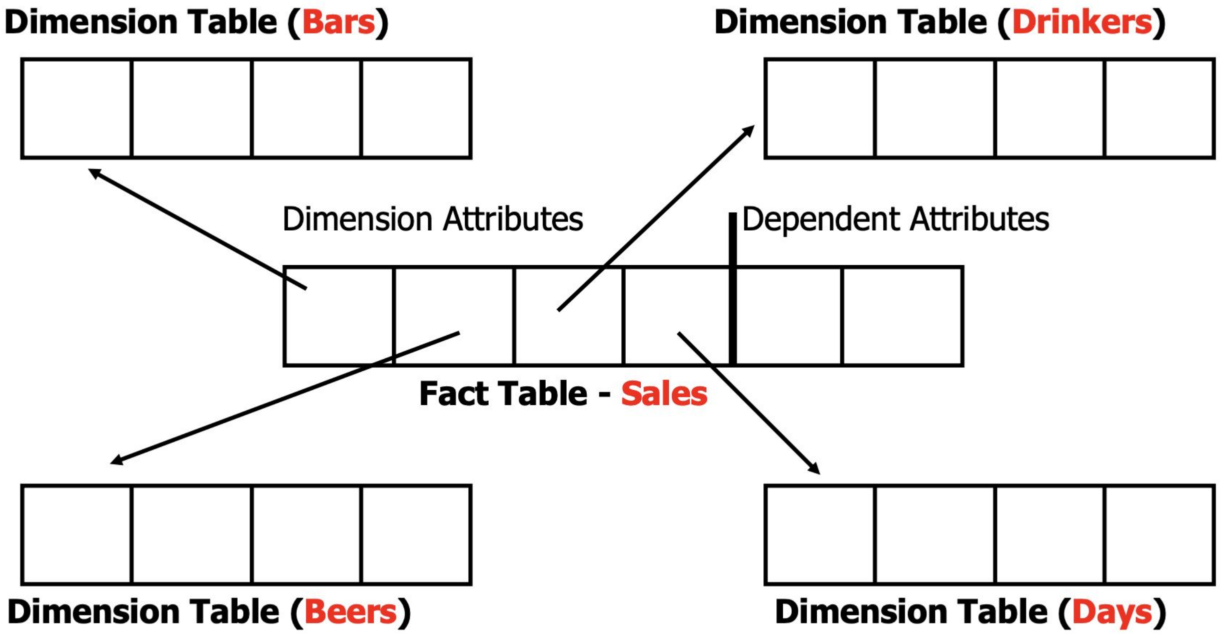
- Closure = { B, C }
- Closure = { A, B, C } (due to B $\rightarrow$ AC)
- Closure = { A, B, C, E } (due to AB $\rightarrow$ E)
- Closure = { A, B, C, E } (due to BE $\rightarrow$ C)
  - No change, so stop.
- Therefore $(BC)^+ = \{A,B,C,E\}$
- Since $E \in (BC)^+$, answer YES.

Superkeys, Keys and Prime Attributes

For $X \rightarrow A$...

- If $X^+ = attr(R)$, then X is a **superkey**
- If there is not proper subset of X that is a super key, then X is a **key**
  - I.e. if A and AD are superkeys, only A is a key
- A is a **prime attribute** if it is part of some key

Star Schemas



Dimension and Dependent Attributes

Dimension Attribute: The key of a dimension table

Dependent Attribute: a fact value determined by the dimension attributes of the tuple

Online Analytical Processing Operations

$ of Anheuser-Busch by drinker/bar

	Jim	Bob	Mary
Joe's Bar	45	33	30
Nut-House	50	36	42
Blue Chalk	38	31	40

Roll-up by Bar

$ of A-B / drinker

Jim	Bob	Mary
133	100	112

Drill-down by Beer

$ of A-B Beers / drinker

	Jim	Bob	Mary
Bud	40	29	40
M'lob	45	31	37
Bud Light	48	40	35

- Pivoting: Changing the dimensions used in a cross-tab
- Slicing: Creating a cross-tab for fixed values only
- Rollup: Moving from finer-granularity data to a coarser granularity
- Drill down: Moving from coarser granularity data to finer granularity data

Joins

Outer Join: preserves dangling tuples by padding them with NULL

FULL Outer Join: pad both sides; default

R

A	B
1	2
4	5

S

B	C
2	3
6	7

SELECT * FROM R FULL OUTER JOIN S ON R.B = S.B;

A	B	B	C
1	2	2	3
4	5	N	N
N	N	6	7

Right Outer Join

Pad dangling tuples of S only

R

A	B
1	2
4	5

S

B	C
2	3
6	7

SELECT * FROM R RIGHT OUTER JOIN S ON R.B = S.B;

A	B	B	C
1	2	2	3
N	N	6	7

Left Outer Join

Pad dangling tuples of R only

R

A	B
1	2
4	5

S

B	C
2	3
6	7

SELECT * FROM R LEFT OUTER JOIN S ON R.B = S.B;

A	B	B	C
1	2	2	3
4	5	N	N

Setting NULL values to 0

How do you change NULL value to 0?

- COALESCE(x, 0) has value x if x isn't NULL, and value 0 if x is NULL.
- Using LEFT OUTER JOIN
- Using UNION