

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

Null Values

Used to represent missing information

At least two kinds

dne — does not exist

unk — value exists but unknown

PILOT	FLIGHT	DATE	TIME
Cushing	615	5Oct	unk
dne	704	6Oct	5:50a
Cook	dne	dne	dne

Unit 2, Part 4

David Maier

1

Principles of Databases

“Does Not Exist” Null

Like adding an additional value to
domain of an attribute

So it enlarges the set of possible
relation instances

Unit 2, Part 4

David Maier

2

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

Issue: dne = dne?

Maybe no: LD → P

PILOT	FLIGHT	DATE	TIME
Cook	dne	dne	dne
Cadiz	dne	dne	dne

Maybe yes: FIRST MI LAST → AGE

FIRST	MI	LAST	AGE
Alex	dne	Cook	37
Alex	dne	Cook	39

Unit 2, Part 4

David Maier

3

Principles of Databases

Unknown Null

More like variables

Can view an instance with **unk** nulls as a set of possible fully-defined instances

PILOT	FLIGHT	DATE	TIME
Cushing	615	50ct	unk
Cook	704	60ct	5:50a
Cushing	unk	60ct	unk

Unit 2, Part 4

David Maier

4

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

Some Possible Instances

PILOT	FLIGHT	DATE	TIME
Cushing	615	50ct	5:00p
Cook	704	60ct	5:50a
Cushing	872	60ct	11:00p

PILOT	FLIGHT	DATE	TIME
Cushing	615	50ct	6:00p
Cook	704	60ct	5:50a
Cushing	615	60ct	6:00p

Unit 2, Part 4

David Maier

5

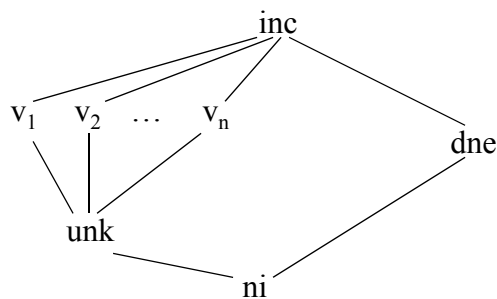
Principles of Databases

Also, “No-Information” and “Inconsistent”

$x \sqsubseteq y$: y more informative than x

$ni \sqsubseteq unk \sqsubseteq v_i \sqsubseteq inc$

$ni \sqsubseteq dne \sqsubseteq inc$



Unit 2, Part 4

David Maier

6

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

Extend Order to Tuples

Use \sqsubseteq on each component

```
<Fred, unk, 6>  $\sqsubseteq$  <Fred, 10.25, 6>  
<Fred, unk, 6>  $\sqsubseteq$  <Fred, unk, 6>  
<ni, dne, 6>  $\sqsubseteq$  <unk, inc, 6>  
<ni, dne, 6>  $\sqsubseteq$  <inc, inc, inc>
```

Unit 2, Part 4

David Maier

7

Principles of Databases

Extending to Relations

Concentrate on unk null

```
ord1(Person Price Amt)  
Fred   unk   6  
Fred   10.25 unk  
Fritz  unk   unk
```

```
ord2(Person Price Amt)  
Fred   11.50 6  
unk    10.25 7
```

```
ord3(Person Price Amt)  
Fred   11.50 6  
Fred   10.25 7  
Fritz  12.25 7
```

Unit 2, Part 4

David Maier

8

What Does a Tuple Mean

For partial relations r, q , what should $r \sqsubseteq q$ require?

For tuple t in r

- exactly one s in q with $t \sqsubseteq s$?
- at least one s in q with $t \sqsubseteq s$?

For tuples t_1, t_2 in r , can there be a single s in q with $t_1 \sqsubseteq s$ and $t_2 \sqsubseteq s$?

Can there be s in q where $t \not\sqsubseteq s$ for all t in r ? (closed world assumption)

Book's Definition

"Fill in the blanks"

$r \sqsubseteq q$ means there is a total, onto map θ from r to q where for every t in r

$$t \sqsubseteq \theta(t)$$

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

Examples

ord1(<u>Person</u> Price Amt)	ord3(<u>Person</u> Price Amt)
Fred unk 6	→ Fred 11.50 6
Fred 10.25 unk	→ Fred 10.25 7
Fritz unk unk	→ Fritz 12.25 7

θ

ord1(<u>Person</u> Price Amt)	ord4(<u>Person</u> Price Amt)
Fred unk 6	→ Fred 10.25 6
Fred 10.25 unk	→ Fritz 12.25 unk
Fritz unk unk	→ Fritz 12.25 unk

θ

Unit 2, Part 4

David Maier

11

Principles of Databases

But Consider

ord2(<u>Person</u> Price Amt)	ord3(<u>Person</u> Price Amt)
Fred 11.50 6	→ Fred 11.50 6
unk 10.25 7	→ Fred 10.25 7
?	→ Fritz 12.25 7

θ

Unit 2, Part 4

David Maier

12

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

Possible Worlds

View a partial relation as specifying a set of possible worlds

$$\text{POSS}(r) = \{q \mid r \sqsubseteq q \text{ and } q \text{ is fully defined}\}$$

Can include constraints C

$$\text{POSS}(r, C) = \{q \mid q \text{ in } \text{POSS}(r) \text{ and } q \text{ in } \text{SAT}(C)\}$$

Unit 2, Part 4

David Maier

13

Principles of Databases

What Happens if We Have Constraints?

Recall

PILOT	FLIGHT	DATE	TIME
Cushing	615	50ct	unk
Cook	704	60ct	5:50a
Cushing	unk	60ct	unk

Unit 2, Part 4

David Maier

14

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

Not All Satisfy the FDs

$$F = \{L \rightarrow T, PDT \rightarrow L, LD \rightarrow P\}$$

PILOT	FLIGHT	DATE	TIME
Cushing	615	50ct	
Cook	704	60ct	5:50a
Cushing		60ct	

PILOT	FLIGHT	DATE	TIME
Cushing	615	50ct	
Cook	704	60ct	5:50a
Cushing		60ct	

Unit 2, Part 4

David Maier

15

Principles of Databases

Use FD-Chase on Instances with Unknowns

r

PILOT	FLIGHT	DATE	TIME
Cushing	704	50ct	unk
Cook	704	60ct	5:50a
Cushing	unk	60ct	unk

r'

PILOT	FLIGHT	DATE	TIME
Cushing	704	50ct	5:50a
Cook	704	60ct	5:50a
Cushing	unk	60ct	unk

$$\text{POSS}(r, F) = \text{POSS}(r', F)$$

Unit 2, Part 4

David Maier

16

Querying Relations with Unknowns

Let r be a relation with unknowns, and
 Q a query on full relations

Would like Q' on partial relations
 $Q'(r)$ represents $\{Q(q) \mid q \text{ in } POSS(r)\}$

When r is fully defined, would like
 $Q'(r)$ and $Q(r)$ to agree. (Faithful)

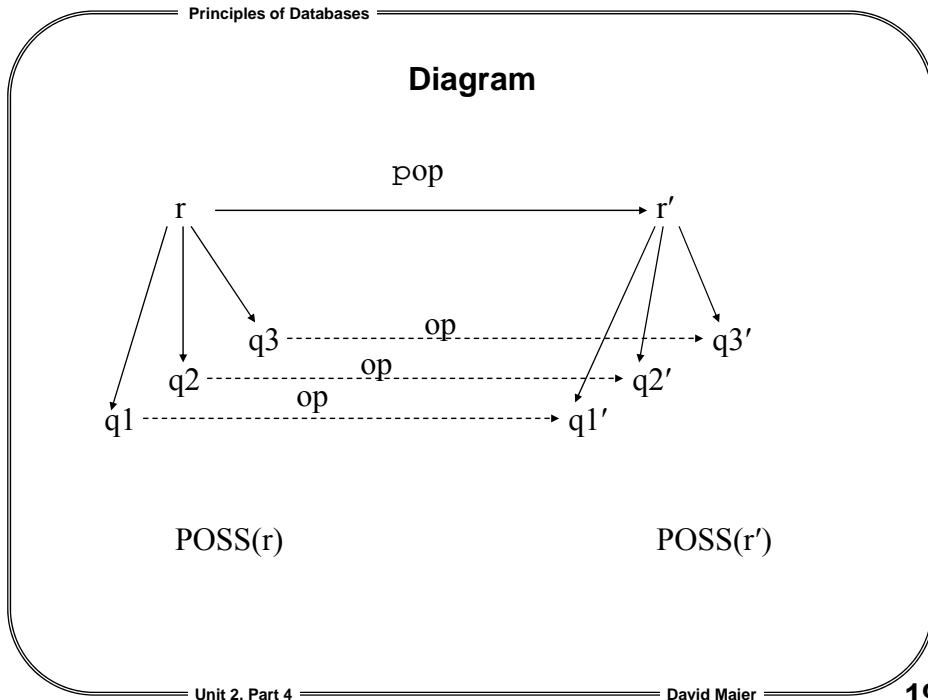
Relational Algebra

One approach: Multi-relation

Let ρ be partial version of operator op
 $\rho\pi$ and π

$POSS(\rho op(r)) = \{op(q) \mid q \text{ in } POSS(r)\}$

CS 589 Principles of Databases, Unit 2, Part 4



19

Principles of Databases

Sometimes Exists

r				
	PILOT	FLIGHT	DATE	TIME
	Cushing	615	50ct	unk
	Cook	704	60ct	5:50a
	Cushing	unk	60ct	unk

$p\pi_{DT}(r)$	
DATE	TIME
50ct	unk
60ct	5:50a
60ct	unk

Unit 2, Part 4 David Maier

20

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

Not Always

$\sigma_{L=872}(r)$

r1

PILOT	FLIGHT	DATE	TIME
Cushing	615	5Oct	5:00p
Cook	704	6Oct	5:50a
Cushing	872	6Oct	11:00p

ans(r1)

PILOT	FLIGHT	DATE	TIME
Cushing	872	6Oct	11:00p

Unit 2, Part 4

David Maier

21

Principles of Databases

Not Always

$\sigma_{L=872}(r)$

r2

PILOT	FLIGHT	DATE	TIME
Cushing	615	5Oct	5:00p
Cook	704	6Oct	5:50a
Cushing	615	6Oct	5:00p

ans(r2)

PILOT	FLIGHT	DATE	TIME
\emptyset			

Unit 2, Part 4

David Maier

22

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

“For sure semantics”

Truth preserving: common part of all possible results

$$\bigcap \{q \mid q \text{ in } \text{POSS}(\text{pop}(r))\} = \bigcap \{\text{op}(q') \mid q' \text{ in } \text{POSS}(r)\}$$

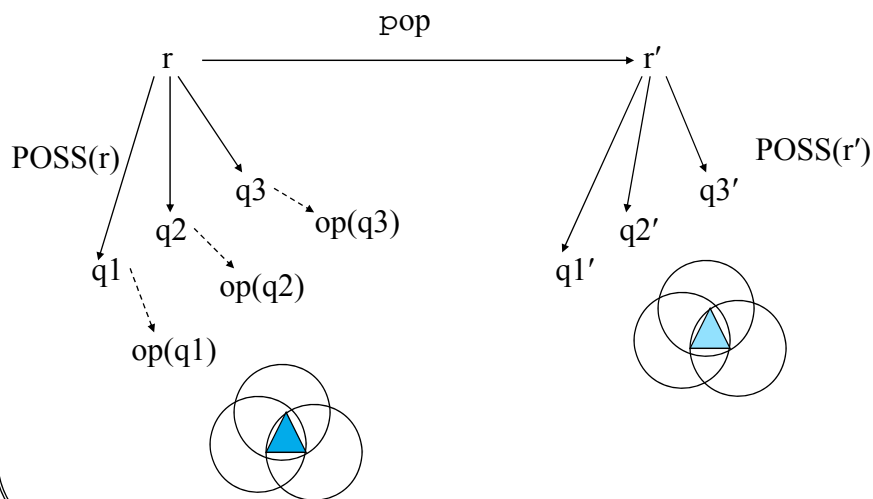
Unit 2, Part 4

David Maier

23

Principles of Databases

Diagram



Unit 2, Part 4

David Maier

24

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

Projection

$\rho\pi_Y(r) = \{t[Y] \mid t \text{ in } r\}$
ord1(Person Price Amt)
Fred unk 6
Fred 10.25 unk
Fritz unk unk

$\rho\pi_{\text{Person Price}}(r) =$
ans(Person Price)
Fred unk
Fred 10.25
Fritz unk

Unit 2, Part 4

David Maier

25

Principles of Databases

Projection 2

POSS(r) includes

(Person Price Amt)
Fred 11.50 6
Fred 10.25 7
Fritz 12.25 7

POSS($\rho\pi_{\text{Person Price}}(r)$) includes

(Person Price)
Fred 11.50
Fred 10.25
Fritz 12.25

Common part will be <Fred 10.25>

Unit 2, Part 4

David Maier

26

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

Union

$$r1 \cup r2 = \{t \mid t \text{ in } r1 \text{ or } t \text{ in } r2\}$$

$$\text{ord1} \cup \text{ord2}$$

ans(Person Price Amt)

Fred	unk	6
Fred	10.25	unk
Fritz	unk	unk

Fred	11.50	6
unk	10.25	7

Unit 2, Part 4

David Maier

27

Principles of Databases

What if We Let Intersection Be the Result?

Might only want tuples that are in every possible answer

$$\text{pop}(r) = \{t \mid t \text{ in } \text{op}(q) \text{ for every } q \text{ in } \text{POSS}(r)\}$$

For example, r

PILOT	FLIGHT	DATE	TIME
Cushing	615	50ct	unk
Cook	704	60ct	5:50a
Cushing	unk	60ct	unk

$\rho\pi_{PL}(r)$

PILOT	FLIGHT
Cushing	615
Cook	704

Unit 2, Part 4

David Maier

28

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

But Note

r

PILOT	FLIGHT	DATE	TIME
Cushing	615	50ct	unk
Cook	704	60ct	5:50a
Cushing	unk	60ct	unk

$$\rho\sigma_{P=Cushing}(r) = \emptyset$$

$$\rho\pi_{LD}(\rho\sigma_{P=Cushing}(r)) =$$

FLIGHT	DATE
615	50ct

Unit 2, Part 4

David Maier

29

Principles of Databases

“Could Be” Semantics

Possibility Preserving: combination of all possible results

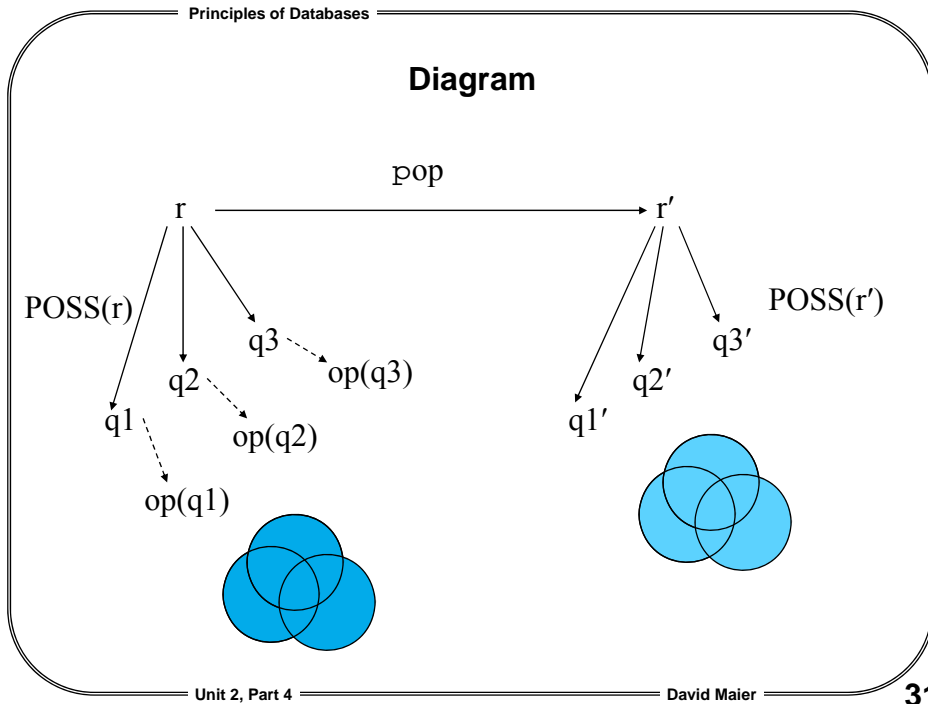
$$\cup\{q \mid q \text{ in } POSS(\rho op(r))\} = \cup\{op(q') \mid q' \text{ in } POSS(r)\}$$

Unit 2, Part 4

David Maier

30

CS 589 Principles of Databases, Unit 2, Part 4



Principles of Databases

Selection Conditions

$t.B = 6$

<Fred 6>	true
<Fred 4>	false
<Fred unk>	maybe

3-valued logic

or	T	F	M
T	T	T	T
F	T	F	M
M	T	M	M

Unit 2, Part 4 David Maier **32**

CS 589 Principles of Databases, Unit 2, Part 4

Principles of Databases

Example

t1 <Fred unk>

t2 <Fritz unk>

t.A = Fritz **or** t.B = 6

t1 F **or** M = M

t2 T **or** M = T

Unit 2, Part 4

David Maier

33

Principles of Databases

But

t1 <Fred unk>

t.B < 6 **or** t.B ≥ 6

t1 M **or** M = M

But it would be T for any value for unk

Unit 2, Part 4

David Maier

34