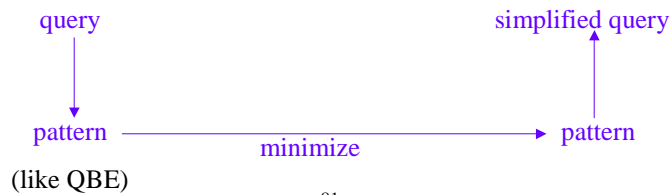


## Exact Minimization of # of Joins

Possible for large set of queries:

- Basic, unnested SQL queries
- One-stage QBE queries
- Relational algebra with  $\pi, \sigma, \bowtie$
- Relational calculus with  $\exists, =, \wedge$

- Basic idea:



81

## Minimization for a simple case

- Relational algebra with:
  - one input relation
  - $\pi, \bowtie$
  - $\sigma_{\text{cond}}$  where cond is  $A = \text{const}$

↓  
rspj query

- First step: write query in QBE style, using a pattern  $\rightarrow$  tableau
  - e.g. tableau for  $\pi_{AB}(R) \bowtie \pi_{BC}(R)$ , where R: ABC, is:

A	B	C	
a	b	c <sub>1</sub>	pattern
a <sub>1</sub>	b	c	
a	b	c	answer

82

## Tableaux

- Tableau over R:  $\langle s, t \rangle$  where
  - $s$  is the set of rows over R, with variables or constants
  - $t$  is the “answer” row

$t$  is over a subset of  $\text{att}(R)$

$t(A)$  is a variable occurring in the A column of S, or a constant
  
- Terminology
  - variables in  $t$ : distinguished
    - denoted  $a, b, c$
  - other variables: nondistinguished
    - denoted  $a_1 a_2 \dots b_1 b_2 \dots c_1 c_2$

83

## Example

- Directors who are also actors
  
- QBE:

movie	title	dir	actor
		-d	
			-d
answer	dir		
I.	-d		

- Tableau:

	title	dir	actor	
S	$t_1$	d	$a_1$	
	$t_2$	$d_1$	d	
T		d		answer row

84

## Example

R: ABC

q:  $\pi_{AC}(\pi_{AB}(R) \bowtie \pi_{BC}(\sigma_{A=5}(\pi_{AB}(R)) \bowtie \pi_{AC}(R)))$

Relational calculus:

$\exists b_2[\exists c_1(R(ab_2c_1)) \wedge \exists a_1(\exists c_2(R(a_1b_2c_2)) \wedge a_1=5 \wedge \exists b_1(R(a_1b_1c)))]$

Prenex form (move  $\exists$  to left):

$\exists a_1b_1b_2c_1c_2[R(ab_2c_1) \wedge R(a_1b_2c_2) \wedge R(a_1b_1c) \wedge a_1=5]$

85

## Example, continued

Tableau:

A	B	C	
a	b <sub>2</sub>	c <sub>1</sub>	Note: # rows in tableau = 1 + # joins
5	b <sub>2</sub>	c <sub>2</sub>	
5	b <sub>1</sub>	c	
a		c	

Note: like QBE query

R	A	B	C
	-a	-b <sub>2</sub>	
	5	-b <sub>2</sub>	
	5		-c
answer	A	C	
I.	-a	-c	

86

## Minimizing Tableau

- $\langle s, t \rangle$ : tableau
- Definition
  - Mapping  $f$  on variables is a **homomorphism on  $\langle s, t \rangle$**  iff:
    - $h(t) = t$
    - $h(c) = c$  if  $c$  is constant
    - $h(s) \subseteq s$
- Fact
  - A row  $r \in S$  is redundant iff for some homomorphism
    - $f$  of  $\langle s, t \rangle, r \notin f(s)$
  - Note:  $\langle f(s), t \rangle$  **equivalent to  $\langle s, t \rangle$**

87

## Example

- Tableau  $\langle s, t \rangle$

A	B	C	
a	$b_1$	$c_1$	
$a_1$	b	$c_1$	
a	$b_2$	$c_2$	S
$a_2$	$b_2$	c	
$a_2$	$b_1$	c	
a	b	c	t

- Let  $f: c_2 \rightarrow c_1$ 
  - $b_2 \rightarrow b_1$
  - $a_1 \rightarrow a_1$
  - $a_2 \rightarrow a_2$

88

## Example, continued

- $f(\langle s, t \rangle)$ :

A	B	C
a	$b_1$	$c_1$
$a_2$	$b_1$	c
$a_1$	b	$c_1$
a	b	c

No redundant rows: MINIMAL

- Fact: all minimal equivalent tableaux are isomorphic!

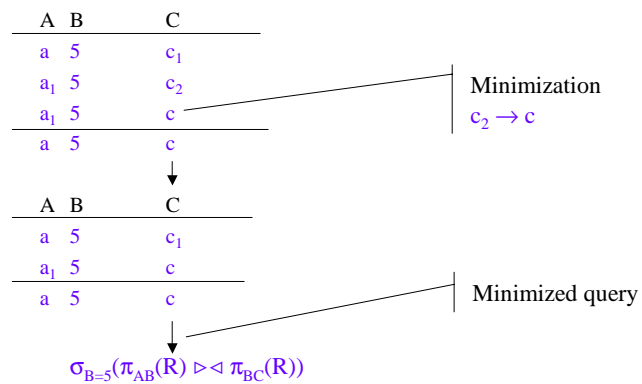
89

## Example

R: ABC

q:  $\pi_{AB}(\sigma_{B=5}(R)) \triangleright \triangleleft \pi_{BC}(\pi_{AB}(R)) \triangleright \triangleleft \pi_{AC}(\sigma_{B=5}(R))$

Tableau:



90

## Functional Dependencies

- Dependencies: statements about properties of valid data
  - e.g.: “Every student is a person”
    - inclusion dependency
  - “Each employee works in more than one department”
    - $NAME \rightarrow DEPARTMENT$
    - functional dependency
- Use of dependencies:
  - check data integrity
  - query optimization
  - schema design  $\rightarrow$  “normal forms”

91

## Functional Dependencies

- Functional dependency over R:
  - expression  $x \rightarrow y$  where  $x, y \subseteq \text{att}(R)$
- A relation r over R **satisfies**  $x \rightarrow y$ 
  - iff whenever two tuples in r agree on x, they also agree on y

e.g.

SCHEDULE	THEATER	TITLE
	odeon	tango
	forum	delicatessen

Satisfies  $THEATER \rightarrow TITLE$

SCHEDULE	THEATER	TITLE
	odeon	tango
	forum	delicatessen
	forum	l'amant

Violates  $THEATER \rightarrow TITLE$ , satisfies  $TITLE \rightarrow THEATER$

92

## Using FDs in Query Optimization

- Example: R: ABC with  $B \rightarrow C$

- query  $\pi_{AB}(R) \triangleright \triangleleft \pi_{BC}(R)$

- Fact: if R satisfies  $B \rightarrow C$  then

- $\pi_{AB}(R) \triangleright \triangleleft \pi_{BC}(R) = R$

- Why: tableau of query is

A	B	C
a	b	$c_1$
$a_1$	b	c
a	b	c

- Note: if  $\langle a, b, c \rangle \in R$  and  $\langle a_1, b, c \rangle \in R$  then  $c_1 = c$  since R satisfies  $B \rightarrow C$

- So, tableau is equivalent to  $\frac{A \ B \ C}{a \ b \ c} \equiv R$

93

- In general: can simplify tableau  $\langle s, t \rangle$  over R if R satisfies a set F of FDs.

- Algorithm: **The Chase**

- Input: tableau  $\langle s, t \rangle$ , set F of FDs

- Output: tableau  $\text{CHASE}_F \langle s, t \rangle$  on all relations satisfying F

- Note: assume without loss of generality that FDs in F are of the form  $X \rightarrow A$  where A is one attribute

94

## The Chase

- Repeat until no change
  - For each  $X \rightarrow A$  in  $F$  do
    - For each  $t_1, t_2$  in  $s$  such that  $t_1(x) = t_2(x), t_1(A) \neq t_2(A)$  do
      - if  $t_1(A), t_2(A)$  are nondistinguished then replace one by the other in  $s$
      - if  $t_1(A)$  distinguished,  $t_2(A)$  nondistinguished then replace  $t_2(A)$  by  $t_1(A)$  in  $s$
      - if  $t_1(A)$  is constant,  $t_2(A)$  is variable then replace  $t_2(A)$  by  $t_1(A)$  in  $s$
      - if  $t_1(A)$  is constant,  $t_2(A)$  is constant then STOP and output  $\emptyset$

95

## Optimization of RSPJ Queries with FDs

- $q$  over  $R$ , set of FDs  $F$  over  $R$ 
  - build tableau  $\langle s, t \rangle$  of  $q$
  - compute  $\text{CHASE}_F \langle s, t \rangle$
  - minimize  $\text{CHASE}_F \langle s, t \rangle$
  - construct rspj query from minimal tableau

96

## Example

– R: ABC    F = {B → A}

–  $q = \pi_{BC}(\sigma_{A=5}(R)) \triangleright \triangleleft \pi_{AB}(R)$

<s, t>:

A	B	C
5	b	c
a	b	$c_1$
a	b	c

CHASE<s, t>:

A	B	C
5	b	c
5	b	$c_1$
5	b	c

MIN:

A	B	C
5	b	c
5	b	c

RSPJ:  $\sigma_{A=5}(R)$

97

## Example

– R: ABC    F = {B → A},  $q = \pi_{BC}(\sigma_{A=5}(R)) \triangleright \triangleleft \pi_{AB}(\sigma_{A=6}(R))$

<s, t>:

A	B	C
5	b	c
6	b	$c_1$
6	b	c

CHASE<s, t>:  $\emptyset$

QUERY:  $\emptyset$

98

## Example

– R: ABC    F = {A → B}, q =  $\pi_{AB}(R) \bowtie \pi_A(\sigma_{B=5}(R)) \bowtie \pi_{AB}(\pi_{AC}(R)) \bowtie \pi_{BC}(R)$

<s, t>:

A	B	C
a	b	c <sub>1</sub>
a	b <sub>1</sub>	c <sub>2</sub>
a <sub>1</sub>	b	c <sub>2</sub>
a	5	c <sub>3</sub>
a	b	

CHASE<s, t>:

A	B	C
a	5	c <sub>1</sub>
a	5	c <sub>2</sub>
a <sub>1</sub>	5	c <sub>2</sub>
a	5	c <sub>3</sub>
a	5	

MIN:	A    B    C	RSPJ: $\pi_{AB}(\sigma_{B=5}(R))$
	a    5    c <sub>1</sub>	
	a    5	