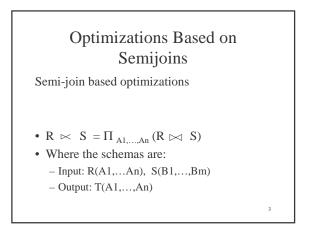
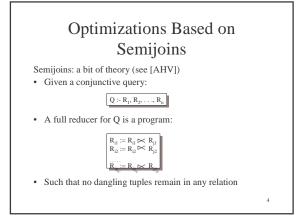
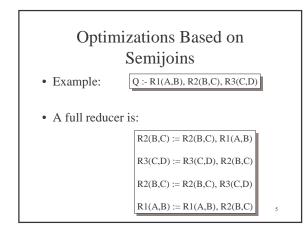


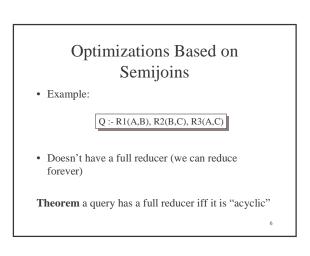


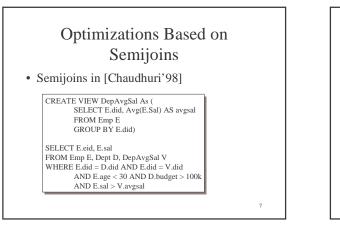
• Book, Chapter 15

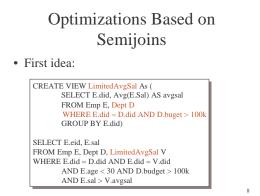


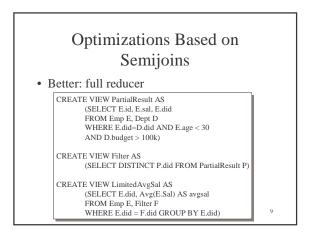


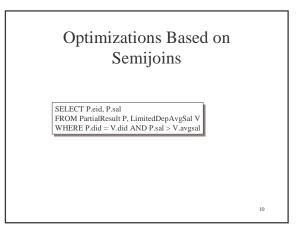








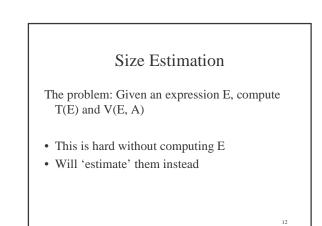




## Modern Query Optimizers

- Volcano
  - Rewrite rules
  - Extensible
- Starburst
  - Keeps query blocks
  - Interblock, intrablock optimizations

11



#### Size Estimation

Estimating the size of a projection

- Easy:  $T(\Pi_L(R)) = T(R)$
- This is because a projection doesn't eliminate duplicates

# Size Estimation

Estimating the size of a selection

- $S = \sigma_{A=c}(R)$ 
  - $\begin{array}{l} \ T(S) \mbox{ san be anything from 0 to } T(R) V(R,A) + 1 \\ \ Estimate: T(S) = T(R)/V(R,A) \end{array}$
  - When V(R,A) is not available, estimate T(S) = T(R)/10
- $S = \sigma_{A < c}(R)$ 
  - T(S) can be anything from 0 to T(R)
  - Estimate: T(S) = (c Low(R, A))/(High(R,A) Low(R,A))

14

16

18

– When Low, High unavailable, estimate T(S) = T(R)/3

#### Size Estimation

Estimating the size of a natural join, R  $|\times|_A S$ 

- When the set of A values are disjoint, then  $T(R \mid \! \times \! \mid_A S) = 0$
- When A is a key in S and a foreign key in R, then  $T(R |x|_A S) = T(R)$
- When A has a unique value, the same in R and S, then  $T(R |x|_A S) = T(R) T(S)$

15

13

## Size Estimation

Assumptions:

- <u>Containment of values</u>: if V(R,A) <= V(S,A), then the set of A values of R is included in the set of A values of S
  Note: this indeed holds when A is a foreign key in R, and a key in S
- <u>Preservation of values</u>: for any other attribute B,  $V(R |x|_A S, B) = V(R, B)$  (or V(S, B))

### Size Estimation

Assume V(R,A) <= V(S,A)

- Then each tuple t in R joins some tuple(s) in S
  - How many ?
  - On average T(S)/V(S,A)
  - $\ t \ will \ contribute \ T(S)/V(S,A) \ tuples \ in \ R \ |\times|_A S$
- Hence  $T(R | \times |_A S) = T(R) T(S) / V(S,A)$

In general:  $T(R |\times|_A S) = T(R) T(S) / max(V(R,A),V(S,A))$ 

17

### Size Estimation

Example:

- T(R) = 10000, T(S) = 20000
- V(R,A) = 100, V(S,A) = 200
- How large is  $R |\times|_A S$  ?

Answer:  $T(R |\times|_A S) = 10000 \ 20000/200 = 1M$ 



Joins on more than one attribute:

•  $T(R | \times |_{A,B} S) =$ 

 $T(R) \ T(S)/(max(V(R,A),V(S,A))*max(V(R,B),V(S,B)))$ 

19

#### Histograms

- Statistics on data maintained by the RDBMS
- Makes size estimation much more accurate (hence, cost estimations are more accurate)

20

		ame, salary				
Main	tain a his	togram on	salary:			
Salary:	020k	20k40k	40k60k	60k80k	80k100k	> 10

anks(rankName, salary)											
Estimate the size of Employee $ \times _{Salary}$ Ranks											
Employee	020k	20k40k	40k60k	60k80k	80k100k	> 100k	1				
	200	800	5000	12000	6500	500	1				
							-				
Ranks	020k	20k40k	40k60k	60k80k	80k100k	> 100k					
	8	20	40	80	100	2	1				

