

# CS 245: Database System Principles

## Notes 6: Query Processing

Hector Garcia-Molina

CS 245

Notes 6

1

## Query Processing

Q → Query Plan

CS 245

Notes 6

2

## Query Processing

Q → Query Plan

Focus: Relational System

- Others?

CS 245

Notes 6

3

## Example

Select B,D

From R,S

Where  $R.A = "c" \wedge S.E = 2 \wedge R.C=S.C$

CS 245

Notes 6

4

R	A	B	C	S	C	D	E
	a	1	10		10	x	2
	b	1	20		20	y	2
	c	2	10		30	z	2
	d	2	35		40	x	1
	e	3	45		50	y	3

CS 245

Notes 6

5

R	A	B	C	S	C	D	E
	a	1	10		10	x	2
	b	1	20		20	y	2
	c	2	10		30	z	2
	d	2	35		40	x	1
	e	3	45		50	y	3

Answer 

B	D
2	x

CS 245

Notes 6

6

• How do we execute query?

One idea

- Do Cartesian product
- Select tuples
- Do projection

CS 245

Notes 6

7

RXS	R.A	R.B	R.C	S.C	S.D	S.E
	a	1	10	10	x	2
	a	1	10	20	y	2
	⋮					
	C	2	10	10	x	2
	⋮					

CS 245

Notes 6

8

RXS	R.A	R.B	R.C	S.C	S.D	S.E
	a	1	10	10	x	2
	a	1	10	20	y	2
	⋮					
Bingo! Got one...	C	2	10	10	x	2
	⋮					

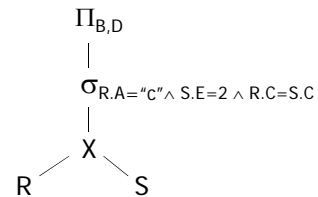
CS 245

Notes 6

9

Relational Algebra - can be used to describe plans...

Ex: Plan I



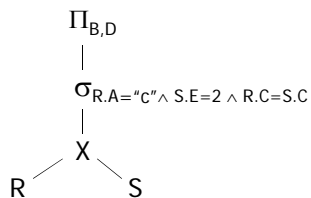
CS 245

Notes 6

10

Relational Algebra - can be used to describe plans...

Ex: Plan I



OR:  $\Pi_{B,D} [\sigma_{R.A='c' \wedge S.E=2 \wedge R.C=S.C} (RXS)]$

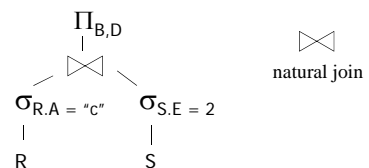
CS 245

Notes 6

11

Another idea:

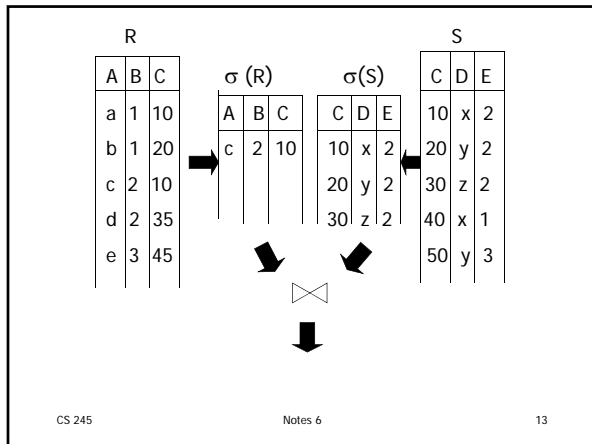
Plan II



CS 245

Notes 6

12



**Plan III**

Use R.A and S.C Indexes

- (1) Use R.A index to select R tuples with R.A = "c"
- (2) For each R.C value found, use S.C index to find matching tuples

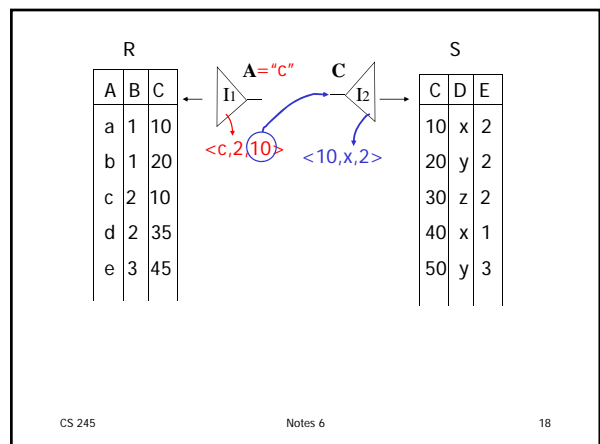
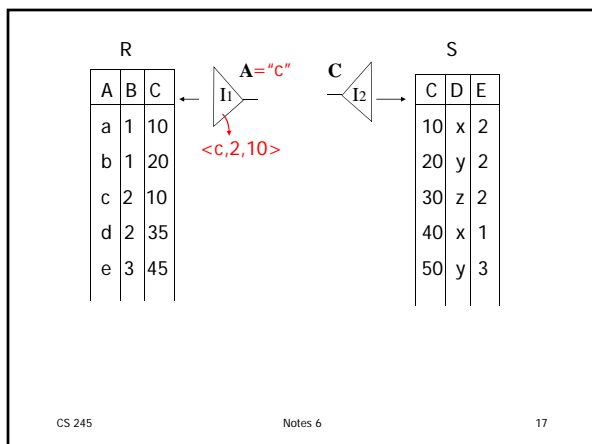
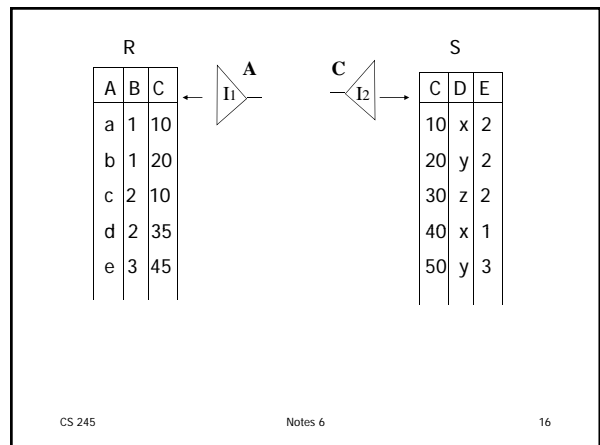
CS 245      Notes 6      14

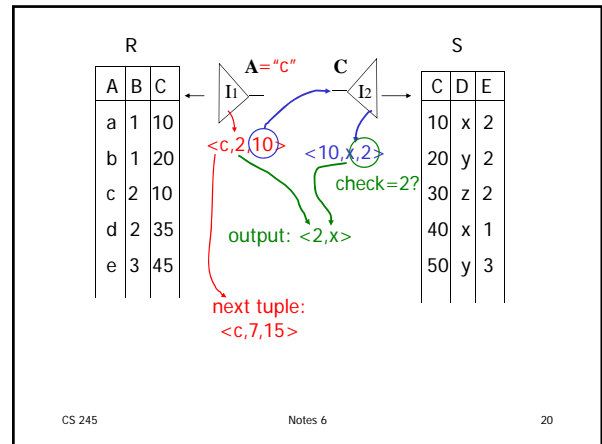
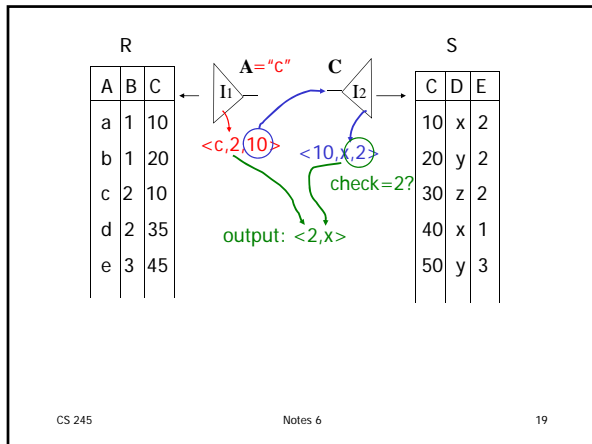
**Plan III**

Use R.A and S.C Indexes

- (1) Use R.A index to select R tuples with R.A = "c"
- (2) For each R.C value found, use S.C index to find matching tuples
- (3) Eliminate S tuples S.E  $\neq$  2
- (4) Join matching R,S tuples, project B,D attributes and place in result

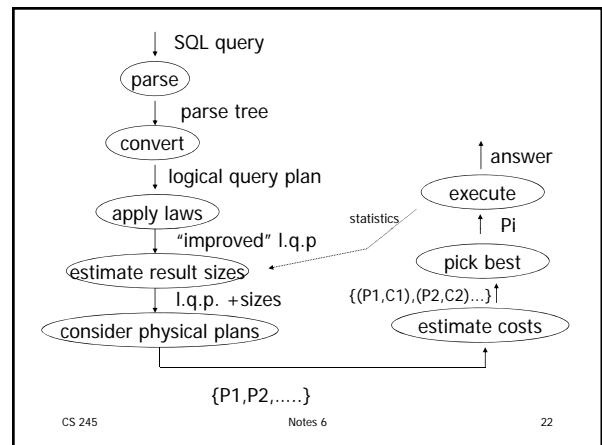
CS 245      Notes 6      15





## Overview of Query Optimization

CS 245 Notes 6 21



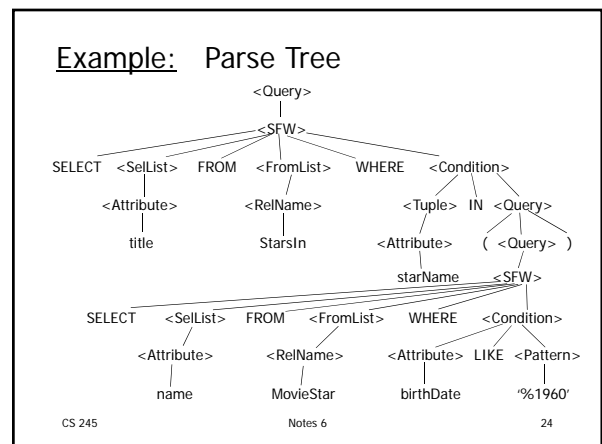
Example: SQL query

```

SELECT title
FROM StarsIn
WHERE starName IN (
  SELECT name
  FROM MovieStar
  WHERE birthdate LIKE '%1960'
);
  
```

(Find the movies with stars born in 1960)

CS 245 Notes 6 23



Example: Generating Relational Algebra

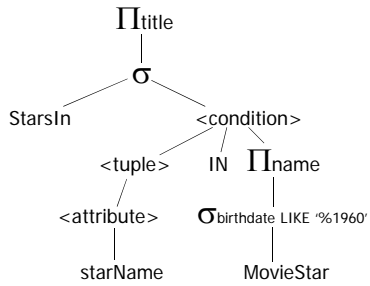


Fig. 7.15: An expression using a two-argument  $\sigma$ , midway between a parse tree and relational algebra

CS 245

Notes 6

25

Example: Logical Query Plan

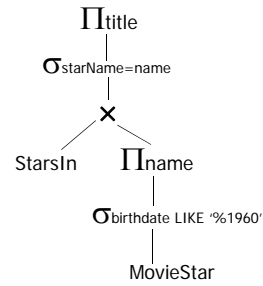


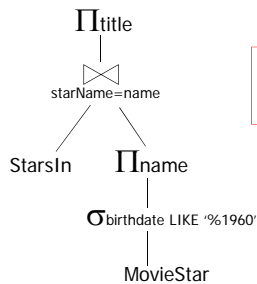
Fig. 7.18: Applying the rule for IN conditions

CS 245

Notes 6

26

Example: Improved Logical Query Plan



Question: Push project to StarsIn?

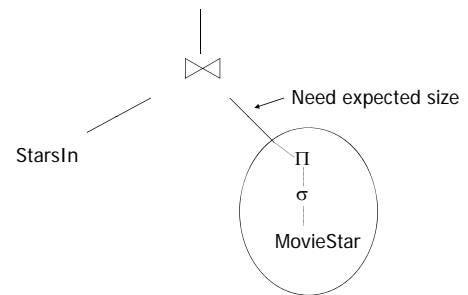
Fig. 7.20: An improvement on fig. 7.18.

CS 245

Notes 6

27

Example: Estimate Result Sizes

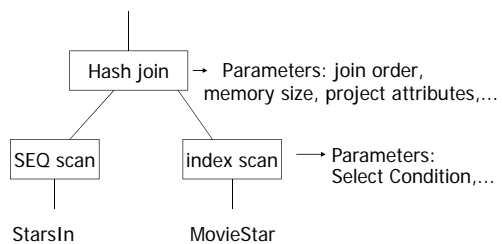


CS 245

Notes 6

28

Example: One Physical Plan

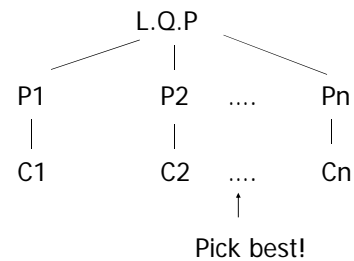


CS 245

Notes 6

29

Example: Estimate costs



CS 245

Notes 6

30

## Textbook outline

### Chapter 15

5 Algebra for queries [bags vs sets]  
[Ch 5] - Select, project, join, .... [project list  
a,a+b->x,...]  
- Duplicate elimination, grouping, sorting

### 15.1 Physical operators

[15.1] - Scan, sort, ...

### 15.2 - 15.6 Implementing operators + [15.2-15.6] estimating their cost

CS 245

Notes 6

31

## Chapter 16

16.1[16.1] Parsing

16.2[16.2] Algebraic laws

16.3[16.3] Parse tree -> logical query  
plan

16.4[16.4] Estimating result sizes

16.5-7[16.5-7] Cost based optimization

CS 245

Notes 6

32

## Reading textbook - Chapters 15, 16

### Optional:

- Sections 15.7, 15.8, 15.9 [15.7, 15.8]
- Sections 16.6, 16.7 [16.6, 16.7]

Optional: Duplicate elimination operator  
grouping, aggregation operators

CS 245

Notes 6

33

## Query Optimization - In class order

- Relational algebra level
- Detailed query plan level

CS 245

Notes 6

34

## Query Optimization - In class order

- Relational algebra level
- Detailed query plan level
  - Estimate Costs
    - without indexes
    - with indexes
  - Generate and compare plans

CS 245

Notes 6

35

## Relational algebra optimization

- Transformation rules  
(preserve equivalence)
- What are good transformations?

CS 245

Notes 6

36

Rules: Natural joins & cross products & union

$$R \bowtie S = S \bowtie R$$

$$(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T)$$

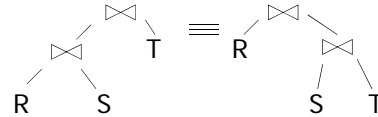
CS 245

Notes 6

37

Note:

- Carry attribute names in results, so order is not important
- Can also write as trees, e.g.:



CS 245

Notes 6

38

Rules: Natural joins & cross products & union

$$R \bowtie S = S \bowtie R$$

$$(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T)$$

$$R \times S = S \times R$$

$$(R \times S) \times T = R \times (S \times T)$$

$$R \cup S = S \cup R$$

$$R \cup (S \cup T) = (R \cup S) \cup T$$

CS 245

Notes 6

39

Rules: Selects

$$\sigma_{p_1 \wedge p_2}(R) =$$

$$\sigma_{p_1 \vee p_2}(R) =$$

CS 245

Notes 6

40

Rules: Selects

$$\sigma_{p_1 \wedge p_2}(R) = \sigma_{p_1} [ \sigma_{p_2}(R) ]$$

$$\sigma_{p_1 \vee p_2}(R) = [ \sigma_{p_1}(R) ] \cup [ \sigma_{p_2}(R) ]$$

CS 245

Notes 6

41

Bags vs. Sets

$$R = \{a, a, b, b, b, c\}$$

$$S = \{b, b, c, c, d\}$$

$$R \cup S = ?$$

CS 245

Notes 6

42

### Bags vs. Sets

R = {a,a,b,b,b,c}

S = {b,b,c,c,d}

RUS = ?

- Option 1 SUM  
RUS = {a,a,b,b,b,b,c,c,c,d}
- Option 2 MAX  
RUS = {a,a,b,b,b,c,c,d}

CS 245

Notes 6

43

### Option 2 (MAX) makes this rule work:

$$\sigma_{p_1 \vee p_2}(R) = \sigma_{p_1}(R) \cup \sigma_{p_2}(R)$$

Example: R={a,a,b,b,b,c}

P1 satisfied by a,b; P2 satisfied by b,c

CS 245

Notes 6

44

### Option 2 (MAX) makes this rule work:

$$\sigma_{p_1 \vee p_2}(R) = \sigma_{p_1}(R) \cup \sigma_{p_2}(R)$$

Example: R={a,a,b,b,b,c}

P1 satisfied by a,b; P2 satisfied by b,c

$$\sigma_{p_1 \vee p_2}(R) = \{a,a,b,b,b,c\}$$

$$\sigma_{p_1}(R) = \{a,a,b,b,b\}$$

$$\sigma_{p_2}(R) = \{b,b,b,c\}$$

$$\sigma_{p_1}(R) \cup \sigma_{p_2}(R) = \{a,a,b,b,b,c\}$$

CS 245

Notes 6

45

### "Sum" option makes more sense:

Senators (.....)

Rep (.....)

T1 =  $\pi_{yr,state}$  Senators; T2 =  $\pi_{yr,state}$  Reps

T1	Yr	State	T2	Yr	State
	97	CA		99	CA
	99	CA		99	CA
	98	AZ		98	CA

Union?

CS 245

Notes 6

46

### Executive Decision

- > Use "SUM" option for bag unions
- > Some rules cannot be used for bags

CS 245

Notes 6

47

### Rules: Project

Let: X = set of attributes

Y = set of attributes

XY = X U Y

$\pi_{xy}(R) =$

CS 245

Notes 6

48

Rules: Project

Let: X = set of attributes  
Y = set of attributes  
XY = X U Y

$$\pi_{xy}(R) = \pi_x[\pi_y(R)]$$

CS 245

Notes 6

49

Rules: Project

Let: X = set of attributes  
Y = set of attributes  
XY = X U Y

$$\pi_{xy}(R) = \pi_x[\pi_y(R)]$$

CS 245

Notes 6

50

Rules:  $\sigma + \bowtie$  combined

Let p = predicate with only R attribs  
q = predicate with only S attribs  
m = predicate with only R,S attribs

$$\sigma_p(R \bowtie S) =$$

$$\sigma_q(R \bowtie S) =$$

CS 245

Notes 6

51

Rules:  $\sigma + \bowtie$  combined

Let p = predicate with only R attribs  
q = predicate with only S attribs  
m = predicate with only R,S attribs

$$\sigma_p(R \bowtie S) = [\sigma_p(R)] \bowtie S$$

$$\sigma_q(R \bowtie S) = R \bowtie [\sigma_q(S)]$$

CS 245

Notes 6

52

Rules:  $\sigma + \bowtie$  combined (continued)

Some Rules can be Derived:

$$\sigma_{p \wedge q}(R \bowtie S) =$$

$$\sigma_{p \wedge q \wedge m}(R \bowtie S) =$$

$$\sigma_{p \vee q}(R \bowtie S) =$$

CS 245

Notes 6

53

Do one, others for homework:

$$\sigma_{p \wedge q}(R \bowtie S) = [\sigma_p(R)] \bowtie [\sigma_q(S)]$$

$$\sigma_{p \wedge q \wedge m}(R \bowtie S) = \sigma_m[(\sigma_p R) \bowtie (\sigma_q S)]$$

$$\sigma_{p \vee q}(R \bowtie S) = [(\sigma_p R) \bowtie S] \cup [R \bowtie (\sigma_q S)]$$

CS 245

Notes 6

54

--> Derivation for first one:

$$\sigma_{p \wedge q} (R \bowtie S) =$$

$$\sigma_p [\sigma_q (R \bowtie S)] =$$

$$\sigma_p [R \bowtie \sigma_q (S)] =$$

$$[\sigma_p (R)] \bowtie [\sigma_q (S)]$$

CS 245

Notes 6

55

Rules:  $\pi, \sigma$  combined

Let  $x$  = subset of R attributes

$z$  = attributes in predicate P  
(subset of R attributes)

$$\pi_x [\sigma_p (R)] =$$

CS 245

Notes 6

56

Rules:  $\pi, \sigma$  combined

Let  $x$  = subset of R attributes

$z$  = attributes in predicate P  
(subset of R attributes)

$$\pi_x [\sigma_p (R)] = \{ \sigma_p [ \pi_x (R) ] \}$$

CS 245

Notes 6

57

Rules:  $\pi, \sigma$  combined

Let  $x$  = subset of R attributes

$z$  = attributes in predicate P  
(subset of R attributes)

$$\pi_x [\sigma_p (R)] = \pi_x \{ \sigma_p [ \overset{\pi_{xz}}{\cancel{\pi_x}} (R) ] \}$$

CS 245

Notes 6

58

Rules:  $\pi, \bowtie$  combined

Let  $x$  = subset of R attributes

$y$  = subset of S attributes

$z$  = intersection of R,S attributes

$$\pi_{xy} (R \bowtie S) =$$

CS 245

Notes 6

59

Rules:  $\pi, \bowtie$  combined

Let  $x$  = subset of R attributes

$y$  = subset of S attributes

$z$  = intersection of R,S attributes

$$\pi_{xy} (R \bowtie S) =$$

$$\pi_{xy} \{ [ \pi_{xz} (R) ] \bowtie [ \pi_{yz} (S) ] \}$$

CS 245

Notes 6

60

$$\pi_{xy} \{ \sigma_P (R \bowtie S) \} =$$

CS 245

Notes 6

61

$$\pi_{xy} \{ \sigma_P (R \bowtie S) \} =$$

$$\pi_{xy} \{ \sigma_P [ \pi_{xz'} (R) \bowtie \pi_{yz'} (S) ] \}$$

$$z' = z \cup \{ \text{attributes used in } P \}$$

CS 245

Notes 6

62

Rules for  $\sigma$ ,  $\pi$  combined with  $X$

similar...

e.g.,  $\sigma_P (R \times S) = ?$

CS 245

Notes 6

63

Rules  $\sigma$ ,  $\cup$  combined:

$$\sigma_P (R \cup S) = \sigma_P (R) \cup \sigma_P (S)$$

$$\sigma_P (R - S) = \sigma_P (R) - S = \sigma_P (R) - \sigma_P (S)$$

CS 245

Notes 6

64

Which are "good" transformations?

- $\sigma_{P_1 \wedge P_2} (R) \rightarrow \sigma_{P_1} [ \sigma_{P_2} (R) ]$
- $\sigma_P (R \bowtie S) \rightarrow [ \sigma_P (R) ] \bowtie S$
- $R \bowtie S \rightarrow S \bowtie R$
- $\pi_x [ \sigma_P (R) ] \rightarrow \pi_x \{ \sigma_P [ \pi_{xz} (R) ] \}$

CS 245

Notes 6

65

Conventional wisdom:  
do projects early

Example:  $R(A,B,C,D,E)$   $x = \{E\}$

$P: (A=3) \wedge (B=\text{"cat"})$

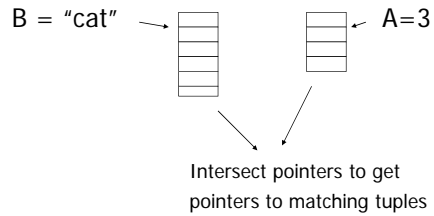
$$\pi_x \{ \sigma_P (R) \} \quad \text{vs.} \quad \pi_E \{ \sigma_P \{ \pi_{ABE} (R) \} \}$$

CS 245

Notes 6

66

**But** What if we have A, B indexes?



CS 245

Notes 6

67

Bottom line:

- No transformation is always good
- Usually good: early selections

CS 245

Notes 6

68

In textbook: more transformations

- Eliminate common sub-expressions
- Other operations: duplicate elimination

CS 245

Notes 6

69

Outline - Query Processing

- Relational algebra level
  - transformations
  - good transformations
- Detailed query plan level
  - estimate costs
  - generate and compare plans

CS 245

Notes 6

70

- Estimating cost of query plan

- (1) Estimating size of results
- (2) Estimating # of IOs

CS 245

Notes 6

71

Estimating result size

- Keep statistics for relation R
  - T(R) : # tuples in R
  - S(R) : # of bytes in each R tuple
  - B(R) : # of blocks to hold all R tuples
  - V(R, A) : # distinct values in R  
for attribute A

CS 245

Notes 6

72

Example

R	A	B	C	D
cat	1	10	a	
cat	1	20	b	
dog	1	30	a	
dog	1	40	c	
bat	1	50	d	

A: 20 byte string  
 B: 4 byte integer  
 C: 8 byte date  
 D: 5 byte string

Example

R	A	B	C	D
cat	1	10	a	
cat	1	20	b	
dog	1	30	a	
dog	1	40	c	
bat	1	50	d	

A: 20 byte string  
 B: 4 byte integer  
 C: 8 byte date  
 D: 5 byte string

$T(R) = 5$      $S(R) = 37$

$V(R,A) = 3$                        $V(R,C) = 5$

$V(R,B) = 1$                        $V(R,D) = 4$

Size estimates for  $W = R1 \times R2$

$T(W) =$

$S(W) =$

Size estimates for  $W = R1 \times R2$

$T(W) = T(R1) \times T(R2)$

$S(W) = S(R1) + S(R2)$

Size estimate for  $W = \sigma_{A=a}(R)$

$S(W) = S(R)$

$T(W) = ?$

Example

R	A	B	C	D
cat	1	10	a	
cat	1	20	b	
dog	1	30	a	
dog	1	40	c	
bat	1	50	d	

$V(R,A)=3$   
 $V(R,B)=1$   
 $V(R,C)=5$   
 $V(R,D)=4$

$W = \sigma_{z=val}(R)$      $T(W) =$

Example

R	A	B	C	D
cat	1	10	a	
cat	1	20	b	
dog	1	30	a	
dog	1	40	c	
bat	1	50	d	

$$V(R,A)=3$$

$$V(R,B)=1$$

$$V(R,C)=5$$

$$V(R,D)=4$$

$$W = \sigma_{Z=val}(R) \quad T(W) = \frac{T(R)}{V(R,Z)}$$

CS 245

Notes 6

79

Assumption:

Values in select expression  $Z = val$  are uniformly distributed over possible  $V(R,Z)$  values.

CS 245

Notes 6

80

Alternate Assumption:

Values in select expression  $Z = val$  are uniformly distributed over domain with  $DOM(R,Z)$  values.

CS 245

Notes 6

81

Example

R	A	B	C	D
cat	1	10	a	
cat	1	20	b	
dog	1	30	a	
dog	1	40	c	
bat	1	50	d	

Alternate assumption  
 $V(R,A)=3 \quad DOM(R,A)=10$   
 $V(R,B)=1 \quad DOM(R,B)=10$   
 $V(R,C)=5 \quad DOM(R,C)=10$   
 $V(R,D)=4 \quad DOM(R,D)=10$

$$W = \sigma_{Z=val}(R) \quad T(W) = ?$$

CS 245

Notes 6

82

$$C=val \Rightarrow T(W) = (1/10)1 + (1/10)1 + \dots$$

$$= (5/10) = 0.5$$

$$B=val \Rightarrow T(W) = (1/10)5 + 0 + 0 = 0.5$$

$$A=val \Rightarrow T(W) = (1/10)2 + (1/10)2 + (1/10)1$$

$$= 0.5$$

CS 245

Notes 6

83

Example

R	A	B	C	D
cat	1	10	a	
cat	1	20	b	
dog	1	30	a	
dog	1	40	c	
bat	1	50	d	

Alternate assumption  
 $V(R,A)=3 \quad DOM(R,A)=10$   
 $V(R,B)=1 \quad DOM(R,B)=10$   
 $V(R,C)=5 \quad DOM(R,C)=10$   
 $V(R,D)=4 \quad DOM(R,D)=10$

$$W = \sigma_{Z=val}(R) \quad T(W) = \frac{T(R)}{DOM(R,Z)}$$

CS 245

Notes 6

84

Selection cardinality

SC(R,A) = average # records that satisfy equality condition on R.A

$$SC(R,A) = \begin{cases} \frac{T(R)}{V(R,A)} \\ \frac{T(R)}{DOM(R,A)} \end{cases}$$

CS 245

Notes 6

85

What about  $W = \sigma_{z \geq val} (R)$  ?

$T(W) = ?$

CS 245

Notes 6

86

What about  $W = \sigma_{z \geq val} (R)$  ?

$T(W) = ?$

- Solution # 1:  
 $T(W) = T(R)/2$

CS 245

Notes 6

87

What about  $W = \sigma_{z \geq val} (R)$  ?

$T(W) = ?$

- Solution # 1:  
 $T(W) = T(R)/2$
- Solution # 2:  
 $T(W) = T(R)/3$

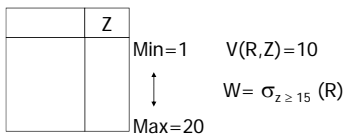
CS 245

Notes 6

88

- Solution # 3: Estimate values in range

Example R



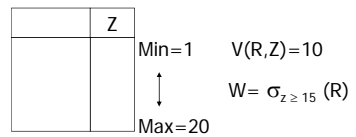
CS 245

Notes 6

89

- Solution # 3: Estimate values in range

Example R



$$f = \frac{20-15+1}{20-1+1} = \frac{6}{20} \quad (\text{fraction of range})$$

$$T(W) = f \times T(R)$$

CS 245

Notes 6

90

Equivalently:

$$f \times V(R,Z) = \text{fraction of distinct values}$$

$$T(W) = \frac{[f \times V(Z,R)] \times T(R)}{V(Z,R)} = f \times T(R)$$

CS 245

Notes 6

91

Size estimate for  $W = R1 \bowtie R2$

Let  $x$  = attributes of R1  
 $y$  = attributes of R2

CS 245

Notes 6

92

Size estimate for  $W = R1 \bowtie R2$

Let  $x$  = attributes of R1  
 $y$  = attributes of R2

Case 1

$$X \cap Y = \emptyset$$

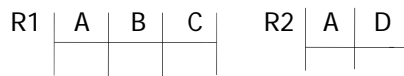
Same as  $R1 \times R2$

CS 245

Notes 6

93

Case 2  $W = R1 \bowtie R2$   $X \cap Y = A$



CS 245

Notes 6

94

Case 2

$W = R1 \bowtie R2$   $X \cap Y = A$



Assumption:

$V(R1,A) \leq V(R2,A) \Rightarrow$  Every A value in R1 is in R2  
 $V(R2,A) \leq V(R1,A) \Rightarrow$  Every A value in R2 is in R1

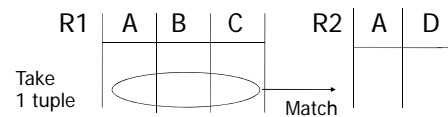
“containment of value sets” Sec. 7.4.4

CS 245

Notes 6

95

Computing  $T(W)$  when  $V(R1,A) \leq V(R2,A)$

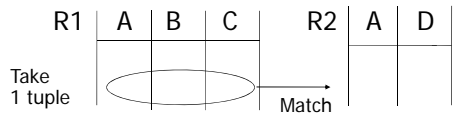


CS 245

Notes 6

96

Computing T(W) when  $V(R1,A) \leq V(R2,A)$



1 tuple matches with  $\frac{T(R2)}{V(R2,A)}$  tuples...

so  $T(W) = \frac{T(R2)}{V(R2,A)} \times T(R1)$

- $V(R1,A) \leq V(R2,A) \quad T(W) = \frac{T(R2) T(R1)}{V(R2,A)}$

- $V(R2,A) \leq V(R1,A) \quad T(W) = \frac{T(R2) T(R1)}{V(R1,A)}$

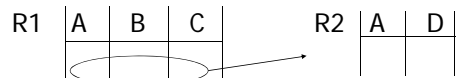
[A is common attribute]

In general  $W = R1 \bowtie R2$

$$T(W) = \frac{T(R2) T(R1)}{\max\{ V(R1,A), V(R2,A) \}}$$

Case 2 with alternate assumption

Values uniformly distributed over domain



This tuple matches  $T(R2)/\text{DOM}(R2,A)$  so

$$T(W) = \frac{T(R2) T(R1)}{\text{DOM}(R2, A)} = \frac{T(R2) T(R1)}{\text{DOM}(R1, A)}$$

Assume the same

In all cases:

$$S(W) = S(R1) + S(R2) - S(A)$$

size of attribute A

Using similar ideas,  
we can estimate sizes of:

$\Pi_{AB}(R)$  ..... Sec. 16.4.2 (same for either edition)

$\sigma_{A=a \wedge B=b}(R)$  .... Sec. 16.4.3

$R \bowtie S$  with common attribs. A,B,C  
Sec. 16.4.5

Union, intersection, diff, ...  
Sec. 16.4.7

Note: for complex expressions, need intermediate T,S,V results.

E.g.  $W = [\underbrace{\sigma_{A=a}(R1)}_U] \bowtie R2$

Treat as relation U

$T(U) = T(R1)/V(R1,A) \quad S(U) = S(R1)$

Also need  $V(U, *) !!$

CS 245

Notes 6

103

To estimate Vs

E.g.,  $U = \sigma_{A=a}(R1)$

Say R1 has attribs A,B,C,D

$V(U, A) =$

$V(U, B) =$

$V(U, C) =$

$V(U, D) =$

CS 245

Notes 6

104

Example

R1	A	B	C	D
cat	1	10	10	
cat	1	20	20	
dog	1	30	10	
dog	1	40	30	
bat	1	50	10	

$V(R1,A)=3$

$V(R1,B)=1$

$V(R1,C)=5$

$V(R1,D)=3$

$U = \sigma_{A=a}(R1)$

CS 245

Notes 6

105

Example

R1	A	B	C	D
cat	1	10	10	
cat	1	20	20	
dog	1	30	10	
dog	1	40	30	
bat	1	50	10	

$V(R1,A)=3$

$V(R1,B)=1$

$V(R1,C)=5$

$V(R1,D)=3$

$U = \sigma_{A=a}(R1)$

$V(U,A) = 1 \quad V(U,B) = 1 \quad V(U,C) = \frac{T(R1)}{V(R1,A)}$

$V(D,U) \dots$  somewhere in between

CS 245

Notes 6

106

Possible Guess  $U = \sigma_{A=a}(R)$

$V(U,A) = 1$

$V(U,B) = V(R,B)$

CS 245

Notes 6

107

For Joins  $U = R1(A,B) \bowtie R2(A,C)$

$V(U,A) = \min \{ V(R1, A), V(R2, A) \}$

$V(U,B) = V(R1, B)$

$V(U,C) = V(R2, C)$

[called "preservation of value sets" in section 7.4.4]

CS 245

Notes 6

108

Example:

$$Z = R1(A,B) \bowtie R2(B,C) \bowtie R3(C,D)$$

R1	T(R1) = 1000	V(R1,A)=50	V(R1,B)=100
R2	T(R2) = 2000	V(R2,B)=200	V(R2,C)=300
R3	T(R3) = 3000	V(R3,C)=90	V(R3,D)=500

CS 245

Notes 6

109

Partial Result:  $U = R1 \bowtie R2$

$$T(U) = \frac{1000 \times 2000}{200} \quad V(U,A) = 50$$

$$V(U,B) = 100$$

$$V(U,C) = 300$$

CS 245

Notes 6

110

$$Z = U \bowtie R3$$

$$T(Z) = \frac{1000 \times 2000 \times 3000}{200 \times 300} \quad V(Z,A) = 50$$

$$V(Z,B) = 100$$

$$V(Z,C) = 90$$

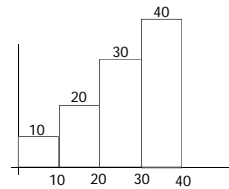
$$V(Z,D) = 500$$

CS 245

Notes 6

111

### A Note on Histograms



number of tuples  
in R with A value  
in given range

$$\sigma_{A=val}(R) = ?$$

CS 245

Notes 6

112

### Summary

- Estimating size of results is an "art"
- Don't forget:  
Statistics must be kept up to date...  
(cost?)

CS 245

Notes 6

113

### Outline

- Estimating cost of query plan
  - Estimating size of results ← done!
  - Estimating # of IOs ← next...
- Generate and compare plans

CS 245

Notes 6

114