

CSE 562 Database Systems

Query Processing: Algebraic Optimization

Some slides are based or modified from originals by
Database Systems: The Complete Book,
Pearson Prentice Hall 2nd Edition
© 2008 Garcia-Molina, Ullman, and Widom

cse@buffalo

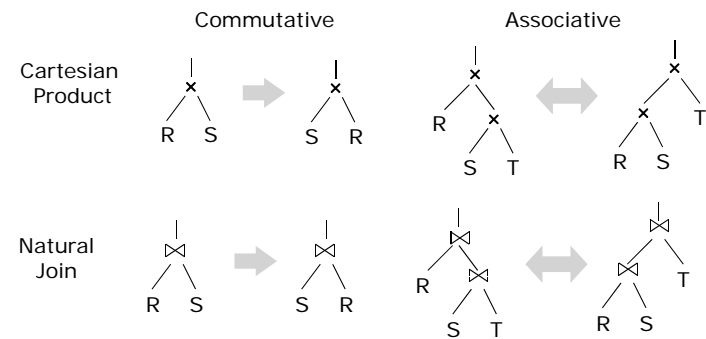
Outline – Query Optimization

- Overview
- Relational algebra level
 - Algebraic Transformations
- Detailed query plan level
 - Estimate Costs
 - Estimating size of results
 - Estimating # of IOs
 - Generate and compare plans

Relational Algebra Optimization

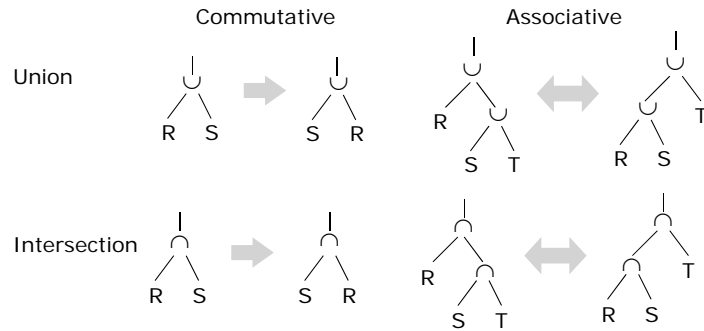
- Transformation rules
(preserve equivalence)
- What are good transformations?

Algebraic Rewritings: Commutative and Associative Laws



- **Question 1:** Do the above hold for both sets and bags?
- **Question 2:** Do commutative and associative laws hold for arbitrary Theta Joins?

Algebraic Rewritings: Commutative and Associative Laws

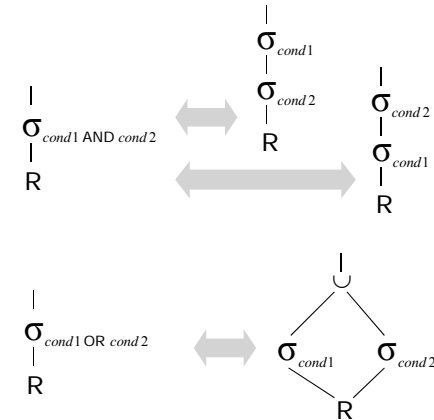


- **Question 1:** Do the above hold for both sets and bags?
- **Question 2:** Is difference commutative and associative?

UB CSE 562 Spring 2009

5

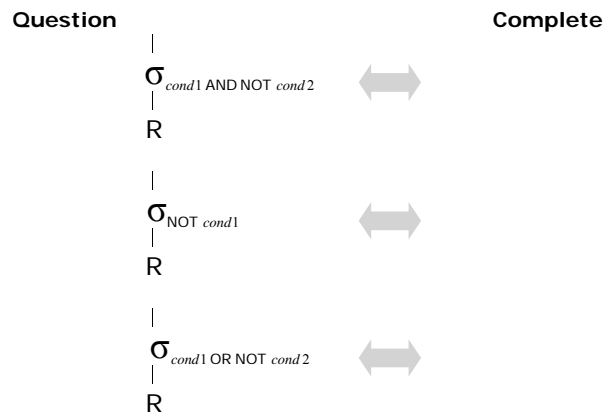
Algebraic Rewritings for Selection: Decomposition of Logical Connectives



UB CSE 562 Spring 2009

6

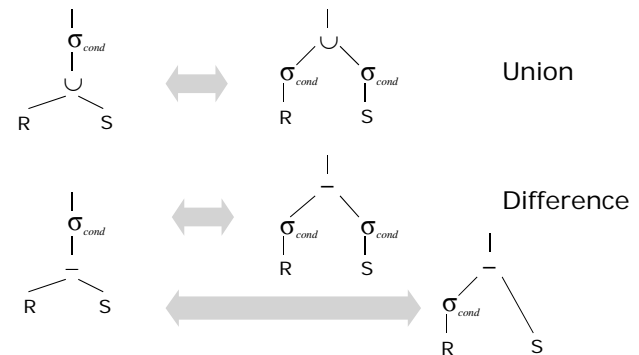
Algebraic Rewritings for Selection: Decomposition of Negation



UB CSE 562 Spring 2009

7

Pushing Selection Through Binary Operators: Union and Difference

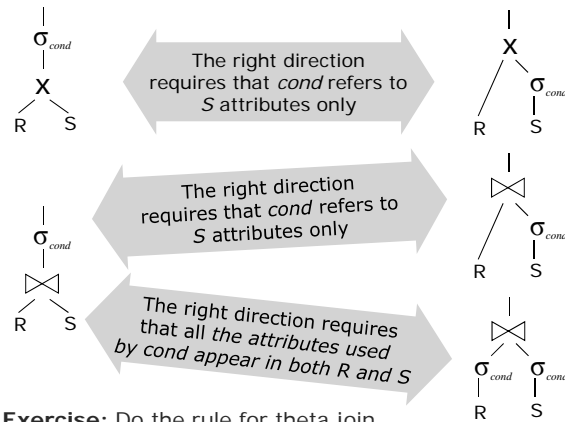


- **Exercise:** Do the rules for intersection

UB CSE 562 Spring 2009

8

Pushing Selection Through Cartesian Product and Join



- **Exercise:** Do the rule for theta join

UB CSE 562 Spring 2009

9

Rules: $\pi + \sigma$ combined

Let X = subset of R attributes
 Z = attributes in predicate P (subset of R attributes)

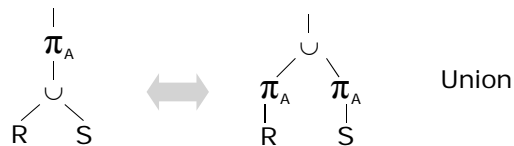
$$\pi_X[\sigma_P(R)] = \pi_X\{\sigma_P[\pi_{XZ}(R)]\}$$

UB CSE 562 Spring 2009

10

Pushing Simple Projections Through Binary Operators: Union

- A projection is simple if it only consists of an attribute list

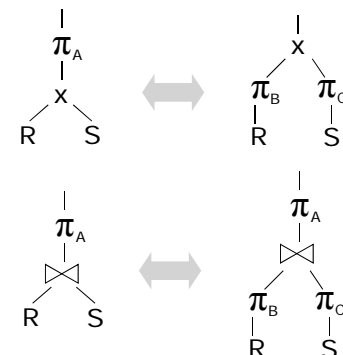


- **Question 1:** Does the above hold for both bags and sets?
- **Question 2:** Can projection be pushed below intersection and difference?
- Answer for both bags and sets

UB CSE 562 Spring 2009

11

Pushing Simple Projections Through Binary Operators: Join and Product



- B is the list of R attributes that appear in A
- Similar for C

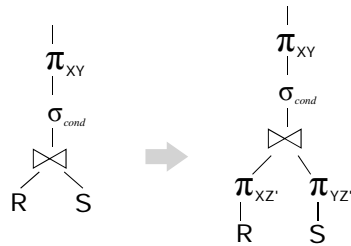
- **Question:** What is B and C ?

- **Exercise:** Write the rewriting rule that pushes projection below theta join

UB CSE 562 Spring 2009

12

Rules: $\pi + \sigma + \bowtie$ combined



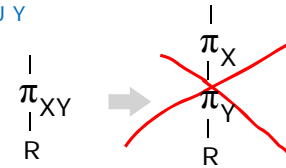
- $Z' = Z \cup \{\text{attributes used in } cond\}$

UB CSE 562 Spring 2009

13

Projection Decomposition

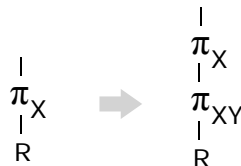
- Let $X =$ set of attributes
- $Y =$ set of attributes
- $XY = X \cup Y$



UB CSE 562 Spring 2009

14

Projection Decomposition



UB CSE 562 Spring 2009

15

Some Rewriting Rules Related to Aggregation: SUM

- $\sigma_{cond}[SUM_{GroupbyList; GroupedAttribute \rightarrow ResultAttribute}(R)]$
 \Leftrightarrow
 $SUM_{GroupbyList; GroupedAttribute \rightarrow ResultAttribute}[\sigma_{cond}(R)]$
 if $cond$ involves only the *GroupbyList*
- $SUM_{GL; GA \rightarrow RA}(R \cup S)$
 \Leftrightarrow
 $PLUS_{RA1, RA2: RA}[(SUM_{GL; GA \rightarrow RA1}(R)) \bowtie (SUM_{GL; GA \rightarrow RA2}(S))]$
- $SUM_{GL2; RA1 \rightarrow RA2}[SUM_{GL1; GA \rightarrow RA1}(R)]$
 \Leftrightarrow
 $SUM_{GL2; GA \rightarrow RA2}(R)$
- **Question:** does the above hold for both bags and sets?

UB CSE 562 Spring 2009

16

Derived Rules: σ + \bowtie combined

More Rules can be Derived:

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

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

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

- **p** only at **R**
- **q** only at **S**
- **m** at both **R** and **S**

UB CSE 562 Spring 2009

17

Derivation for first one

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

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

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

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

UB CSE 562 Spring 2009

18

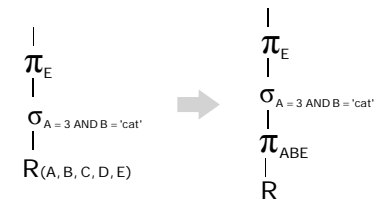
Which are "good" transformations?

- $\sigma_{p1 \wedge p2} (R) \rightarrow \sigma_{p1} [\sigma_{p2} (R)]$
- $\sigma_p (R \bowtie S) \rightarrow [\sigma_p (R)] \bowtie S$
- $R \bowtie S \rightarrow S \bowtie R$
- $\pi_x [\sigma_p (R)] \rightarrow \pi_x \{ \sigma_p [\pi_{xz} (R)] \}$

UB CSE 562 Spring 2009

19

Conventional Wisdom: Do Projects Early

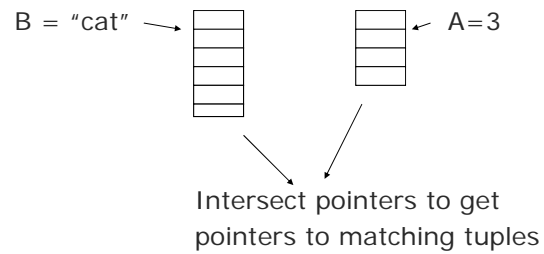


UB CSE 562 Spring 2009

20

But...

What if we have A, B indexes?



UB CSE 562 Spring 2009

21

More Transformations in Textbook

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

UB CSE 562 Spring 2009

22

Bottom line

- No transformation is always good at the logical query plan level
- Usually good:
 - early selections
 - elimination of Cartesian products
 - elimination of redundant sub-expressions
- Many transformations lead to “promising” plans
 - Commuting/rearranging joins
 - In practice too “combinatorially explosive” to be handled as rewriting of logical query plan

UB CSE 562 Spring 2009

23