Distributed Databases

CS347

Lecture 14

May 30, 2001

Topics for the Day

- Query processing in distributed databases
  - Localization
  - Distributed query operators
  - Cost-based optimization

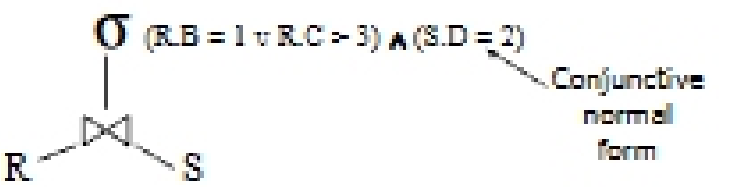
Query Processing Steps

- Decomposition
  - Given SQL query, generate one or more algebraic query trees
- Localization
  - Rewrite query trees, replacing relations by fragments
- Optimization
  - Given cost model + one or more localized query trees
  - Produce minimum cost query execution plan

Decomposition

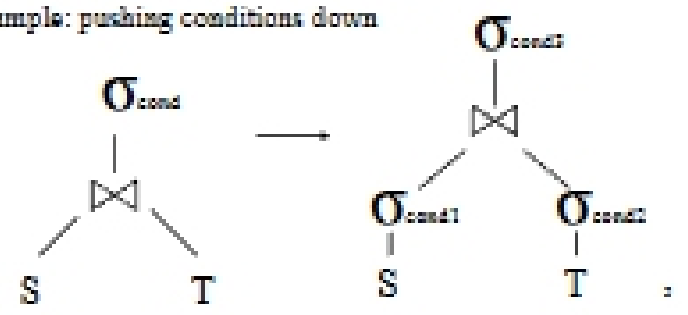
- Same as in a centralized DBMS
- Normalization (usually into relational algebra)

Select A,C
From R Natural Join S
Where (R.B = 1 and S.D = 2) or (R.C > 3 and S.D = 2)



Decomposition

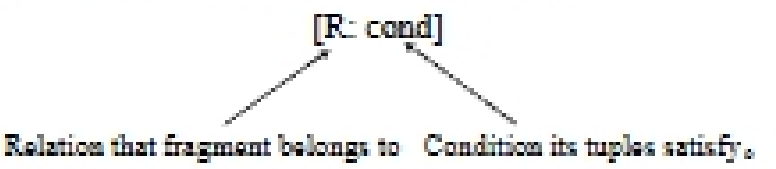
- Redundancy elimination
  - (S.A = 1) ∧ (S.A > 5) → False
  - (S.A < 10) ∧ (S.A < 5) → S.A < 5
- Algebraic Rewriting
  - Example: pushing conditions down



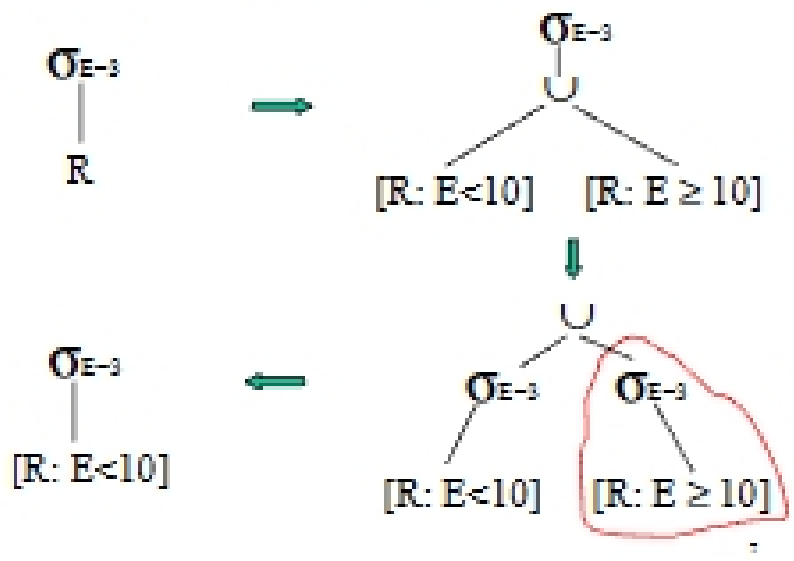
Localization Steps

1. Start with query tree
2. Replace relations by fragments
3. Push ∪ up & π,σ down (CS245 rules)
4. Simplify – eliminating unnecessary operations

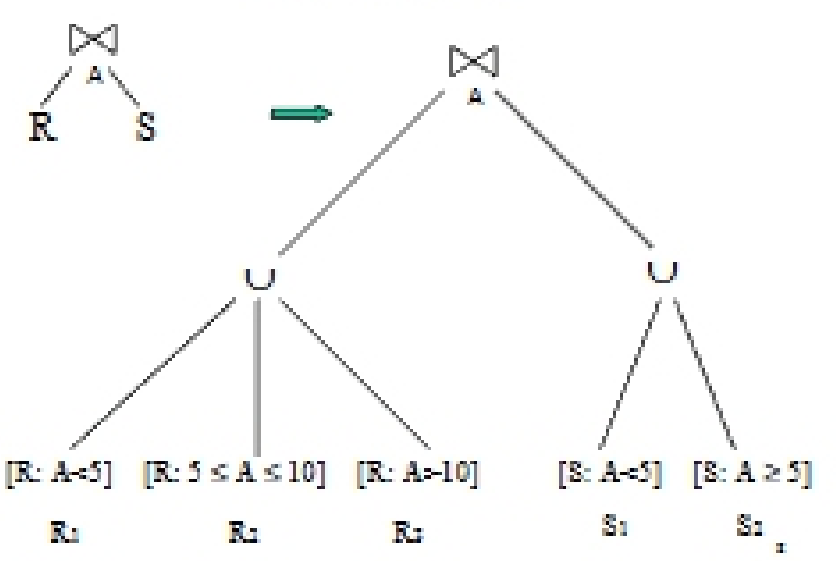
Note: To denote fragments in query trees

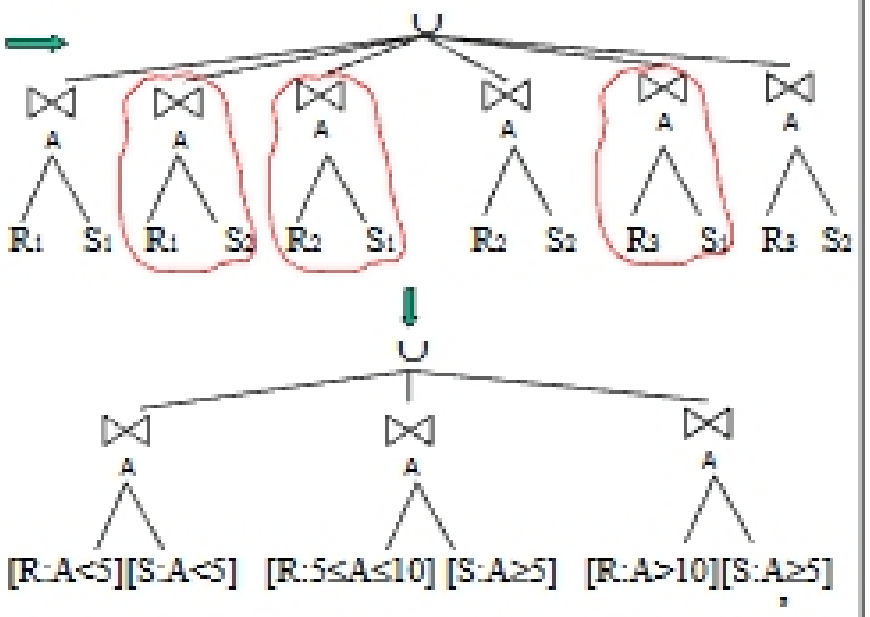


Example 1



Example 2





Rules for Horiz. Fragmentation

- $\sigma_{C1}[R: C_2] \Rightarrow [R: C_1 \wedge C_2]$
- $[R: False] \Rightarrow \emptyset$
- $[R: C_1] \underset{A}{\bowtie} [S: C_2] \Rightarrow [R \underset{A}{\bowtie} S: C_1 \wedge C_2 \wedge R.A = S.A]$

- In Example 1:

  $\sigma_{E=3}[R_2: E \geq 10] \Rightarrow [R_2: E=3 \wedge E \geq 10]$

  $\Rightarrow [R_2: False] \Rightarrow \emptyset$
- In Example 2:

  $[R: A<5] \,?\, [S: A \geq 5]$

  $\Rightarrow [R \,?\, S: R.A < 5 \wedge S.A \geq 5 \wedge R.A = S.A]$

  $\Rightarrow [R \,?\, S: False] \Rightarrow \emptyset$

Example 3 – Derived Fragmentation

S's fragmentation is derived from that of R.