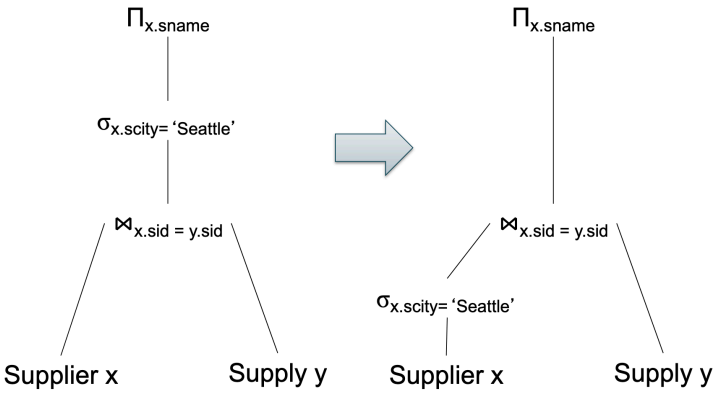
Query Optimization

Friday, May 10, 2019 1:31 PM

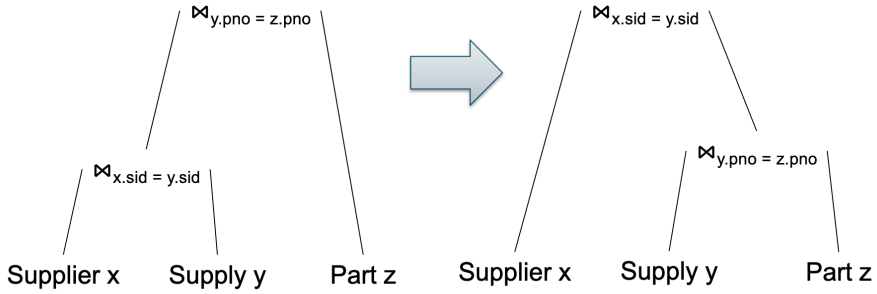
- Main idea: replace a query plan with another one that is equivalent but cheaper.

- Push selections down
  - Do selection before grouping or joining.

e.g. $\sigma_{c_1 and c_2}(R \bowtie S) = \sigma_{c_1}(\sigma_{c_2}(R \bowtie S)) = \sigma_{c_1}(R) \bowtie \sigma_{c_2}(S)$



- Join Reorder: need to know the size of table.
  - First join similiar size tables.
  - Cardinality problem



better when supplier x and supply y are both small

(R ⋈ S) ⋈ T= R ⋈ (S ⋈ T)

better when supply y and part z are both small

⇒ Always do the smaller join first.

Why so many joins: schema normalize

- Size & Cost Estimation
  - B(R) number of blocks (pages) for relation of R.
  - T(R) # of tuples in relation R

  } T(R) ≥ B(R)
  - V(R,A) # of instinct values of attribute A If A is key, V(R,A) = T(R)
- Size estimation: estimate the size of a logical subplan

Cost estimation: estimate the cost of a physical subplan.

- Estimate size of a query plan |P|
  - Worst case sizes:

    $|\sigma_c(R)| \leq |R|$

    $|R \bowtie S| \leq |R * S|$
  - f * worst case f: selectivity factor
  - Estimating size of a selection R(A,B) c = v

- Estimating size of a selection $R(A,B)$, $S(C,D)$
  - Assumption 1. Uniform distribution of values

$$|\sigma_{A=v}(R)| \approx \frac{|T(R)|}{V(R,A)} \qquad f = \frac{1}{V(R,A)}$$

  - Assumption 2. Independence of attributes

$$|\sigma_{A=v \text{ and } B=w}(R)| = \frac{|T(R)|}{(V(R,A) \cdot V(R,B))}$$

  - ✡ $f = f_{A=v} \cdot f_{B=w}$
- Estimating Size of a join
  - Assumption 3: Inclusion

if $V(R,B) \leq V(S,C)$ then $\Pi_B(R) \subseteq \Pi_C(S)$

$$|R \bowtie_{B=C} S| \approx \frac{|R| \cdot |S|}{V(S,C)}$$

  - In general $|R \bowtie_{B=C} S| \approx \frac{|R| * |S|}{\max(V(R,B), V(S,C))}$

- I/O Cost of selection
  - Sequential scan for relation costs $B(R)$
  - Index-based selection
    -
    - Clustered $B(R) \cdot \frac{1}{V(R,A)}$
    - Unclustered $T(R) \cdot \frac{1}{V(R,A)}$

e.g. $B(R) = 2000$, $T(R) = 100{,}000$, $V(R, A) = 20$.