Lecture 23:
Query Execution

Wednesday, March 8, 2006

1

Outline

- Query optimization: algebraic laws 16.2

2

Example

Product(pname, maker), **Company**(cname, city)

```
Select Product.pname
From  Product, Company
Where Product.maker=Company.cname
   and  Company.city = "Seattle"
```

- How do we execute this query ?

3

Example

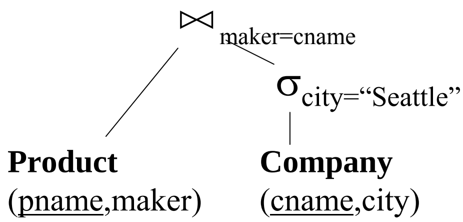
Product(pname, maker), **Company**(cname, city)

Assume:

Clustered index: **Product**.pname, **Company**.cname
Unclustered index: **Product**.maker, **Company**.city

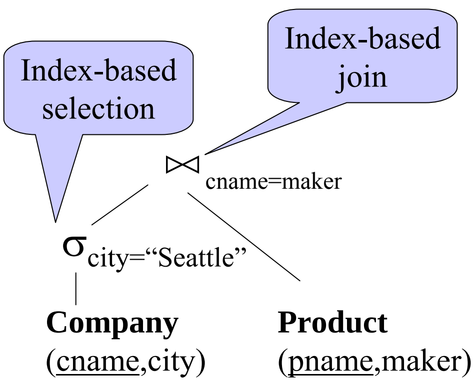
4

Logical Plan:



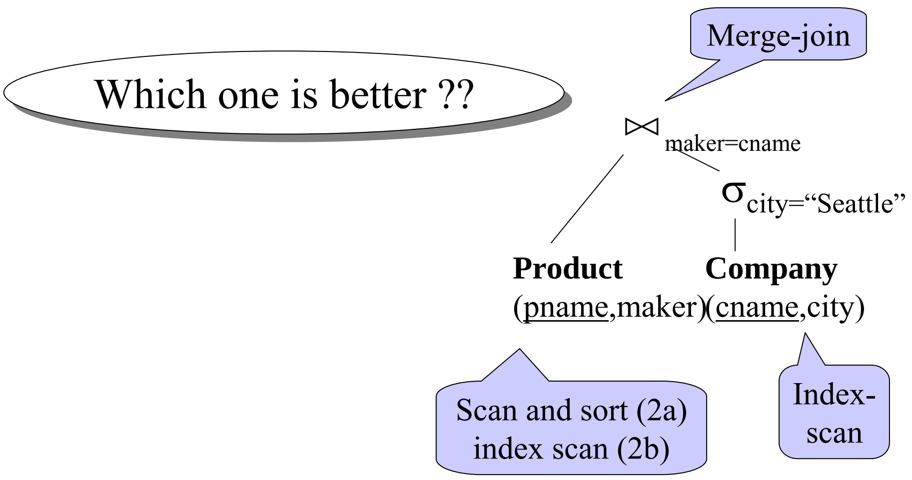
5

Physical plan 1:



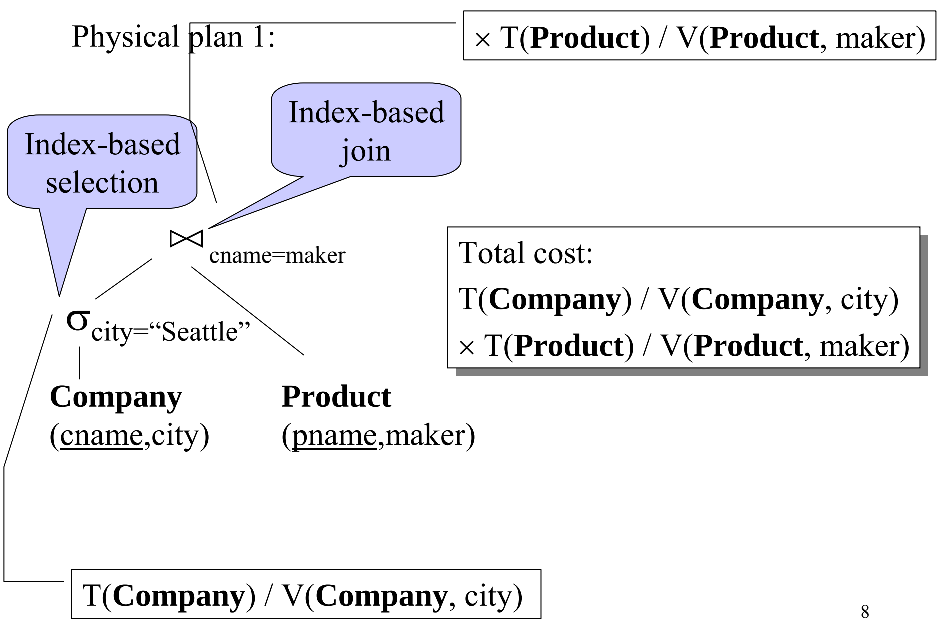
6

Physical plans 2a and 2b:

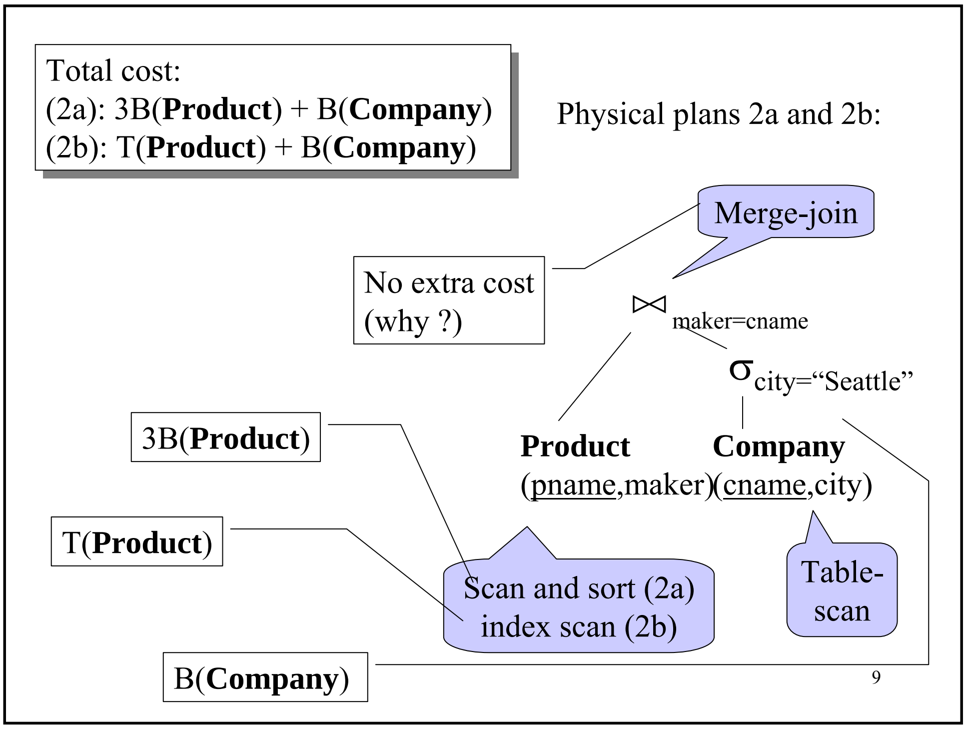


7

Physical plan 1:



8



Plan 1: T(**Company**)/V(**Company**,city) × T(**Product**)/V(**Product**,maker)
Plan 2a: B(**Company**) + 3B(**Product**)
Plan 2b: B(**Company**) + T(**Product**)

Which one is better ??

It depends on the data !!

10

Example

T(**Company**) = 5,000 B(**Company**) = 500 M = 100
T(**Product**) = 100,000 B(**Product**) = 1,000

We may assume V(**Product**, maker) ≈ T(**Company**) (why ?)

- Case 1: V(**Company**, city) ≈ T(**Company**)

V(**Company**,city) = 2,000

- Case 2: V(**Company**, city) << T(**Company**)

V(**Company**,city) = 20

11

Which Plan is Best ?

Plan 1: T(**Company**)/V(**Company**,city) × T(**Product**)/V(**Product**,maker)
Plan 2a: B(**Company**) + 3B(**Product**)
Plan 2b: B(**Company**) + T(**Product**)

Case 1:

Case 2:

12

Lessons

- Need to consider several physical plan
  - even for one, simple logical plan
- No magic "best" plan: depends on the data
- In order to make the right choice
  - need to have ***statistics*** over the data
  - the B's, the T's, the V's

13

Query Optimzation

- Have a SQL query Q
- Create a plan P

HOW ??

- Find equivalent plans P = P' = P'' = …
- Choose the "cheapest".

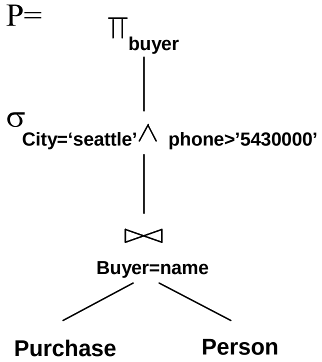
14

Logical Query Plan

```
SELECT  P.buyer
FROM Purchase P, Person Q
WHERE P.buyer=Q.name AND
      P.city='seattle' AND
      Q.phone > '5430000'
```

Purchasse(buyer, city)
Person(name, phone)

P=



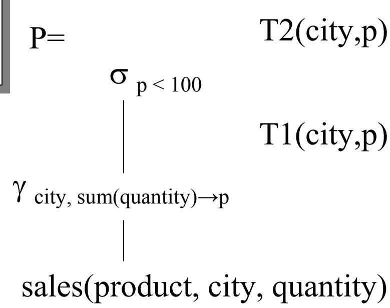
In class:
find a "better" plan P'

Logical Query Plan

Q=

```
SELECT city, sum(quantity)
FROM sales
GROUP BY city
HAVING sum(quantity) < 100
```

P=



In class:
find a "better" plan P'

The three components of an optimizer

We need three things in an optimizer:

- Algebraic laws
- An optimization algorithm
- A cost estimator

17

Algebraic Laws

- Commutative and Associative Laws

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

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

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

- Distributive Laws

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

18

Algebraic Laws

- Laws involving selection:

  $\sigma_{C \text{ AND } C'}(R) = \sigma_C(\sigma_{C'}(R)) = \sigma_C(R) \cap \sigma_{C'}(R)$

  $\sigma_{C \text{ OR } C'}(R) = \sigma_C(R) \cup \sigma_{C'}(R)$

  $\sigma_C (R \mid\times\mid S) = \sigma_C (R) \mid\times\mid S$
- When C involves only attributes of R

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

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

  $\sigma_C (R \mid\times\mid S) = \sigma_C (R) \mid\times\mid S$

19

Algebraic Laws

- Example: R(A, B, C, D), S(E, F, G)

  $\sigma_{F=3} (R \mid\times\mid_{D=E} S) =$?

  $\sigma_{A=5 \text{ AND } G=9} (R \mid\times\mid_{D=E} S) =$?

20

Algebraic Laws

- Laws involving projections

  $\Pi_M(R \mid\times\mid S) = \Pi_M(\Pi_P(R) \mid\times\mid \Pi_Q(S))$

  $\Pi_M(\Pi_N(R)) = \Pi_{M,N}(R)$

- Example R(A,B,C,D), S(E, F, G)

  $\Pi_{A,B,G}(R \mid\times\mid_{D=E} S) = \Pi_{?}(\Pi_?(R) \mid\times\mid_{D=E} \Pi_?(S))$

21

Algebraic Laws

- Laws involving grouping and aggregation:

  $\delta(\gamma_{A,\ agg(B)}(R)) = \gamma_{A,\ agg(B)}(R)$

  $\gamma_{A,\ agg(B)}(\delta(R)) = \gamma_{A,\ agg(B)}(R)$ if agg is "duplicate insensitive"

- Which of the following are "duplicate insensitive" ? sum, count, avg, min, max

$\gamma_{A,\ agg(D)}(R(A,B) \mid\times\mid_{B=C} S(C,D)) =$
$\quad \gamma_{A,\ agg(D)}(R(A,B) \mid\times\mid_{B=C} (\gamma_{C,\ agg(D)}S(C,D)))$

22