Join Algorithms

Comparing Join Algorithms

Options:

Transformations: R1 ⋈ R2, R2 ⋈ R1

- Join algorithms:
  - Iteration (nested loops join)
  - Merge join
  - Join with index
  - Hash join

Factors that affect performance

(1) Tuples of relation stored physically together?

(2) Relations sorted by join attribute?

(3) Indexes exist?

Running Example

Example: R1 ⋈ R2 over common attribute C

T(R1) = 10,000

T(R2) = 5,000

S(R1) = S(R2) = 1/10 block

Memory available = 101 blocks

→ Metric: # of IOs (ignoring writing of result)

Iteration Join (Nested Loops Join)

```
for each r ∈ R1 do
    for each s ∈ R2 do
        if r.C = s.C then output r,s pair
```