CSE 344 Introduction to Data Management

Section 8: NoSQL and Parallel DBMS
Firebase Demo
Architectures for Parallel DBMS

- Shared memory

- Shared disk

- Shared nothing
Shared Nothing

- Clusters of machines
- Each machine has its own memory and disk
- Easy to maintain and scale
- Difficult to administer and tune
Horizontal Data Partitioning

- Block Partition
  - Partition tuples arbitrarily s.t. \( \text{size}(R_1) \approx \ldots \approx \text{size}(R_p) \)

- Hash partitioned on attribute A
  - Tuple \( t \) goes to chunk \( i \), where \( i = h(t.A) \mod P + 1 \)

- Range partitioned on attribute A
Example Parallel Query Plan
Example Parallel Query Plan

Order(oid, item, date), Line(item, ...)}
SELECT a, count(b) as countb
FROM R
WHERE a > 0
GROUP BY a
SELECT a, count(b) as countb
FROM R
WHERE a > 0
GROUP BY a
SELECT a, count(b) as countb
FROM R
WHERE a > 0
GROUP BY a
\[
\gamma_{a, \text{count}(b) \rightarrow b}
\]

\[
\sigma_{a > 0}
\]

\[
n \frac{1}{3} \text{ of } R
\]

\[
\text{Machine 1}
\]

\[
\gamma_{a, \text{count}(b) \rightarrow b}
\]

\[
\sigma_{a > 0}
\]

\[
n \frac{1}{3} \text{ of } R
\]

\[
\text{Machine 2}
\]

\[
\gamma_{a, \text{count}(b) \rightarrow b}
\]

\[
\sigma_{a > 0}
\]

\[
n \frac{1}{3} \text{ of } R
\]

\[
\text{Machine 3}
\]

SELECT a, count(b) as countb
FROM R
WHERE a > 0
GROUP BY a
SELECT a, count(b) as countb
FROM R WHERE a > 0 GROUP BY a

⅓ of R

Machine 1

Hash on a

γ_a, count(b) -> b

σ_{a>0}

scan

⅓ of R

Machine 2

Hash on a

γ_a, count(b) -> b

σ_{a>0}

scan

⅓ of R

Machine 3

Hash on a

γ_a, count(b) -> b

σ_{a>0}

scan
SELECT a, count(b) as countb
FROM R
WHERE a > 0
GROUP BY a
SELECT a, count(b) as countb
FROM R
WHERE a > 0
GROUP BY a

⅓ of R

Machine 1

⅓ of R

Machine 2

⅓ of R

Machine 3