Effects of Join Order

Plan 1:

\[
\text{sel}_{C \times B} := \frac{|C \times B|}{|C| \times |B|} = \frac{100,000}{1,000 \times 1,000} = 0.1
\]

Plan 2:

\[
\text{sel}_{A \times B} := \frac{|A \times C|}{|A| \times |C|} = \frac{100}{1,000 \times 1,000} = 0.0001.
\]

Plan 1: Top-level join has to process 1,000 + 100,000 tuples.

Plan 2: Top-level join has to process 100 + 1,000 tuples.
Effects of a Scan Access Path

\[ \prod_{\text{title}} \]
\[ \times_{A.\text{id}=B.\text{dz}} \]
\[ \prod_{\text{title}, \text{id}} \]
\[ \sigma_{A.\text{name}='Hugo'} \]

scan on A

\[ \prod_{\text{title}} \]
\[ \times_{A.\text{id}=B.\text{dz}} \]
\[ \prod_{\text{title}, \text{id}} \]
\[ \sigma_{A.\text{name}='Hugo'} \]
Effects of an Index Access Path

\[ \Pi_{\text{title}} \]
\[ \Pi_{\text{title}, \text{id}} \]
\[ \sigma_{A.\text{name}='Hugo'} \]

\[ \prod_{\text{title}} \]
\[ \prod_{\text{title}, \text{id}} \]
\[ \sigma_{A.\text{name}='Hugo'} \]
Which Index Access Path Exactly?

clustered index on A.name

\[ \Pi_{\text{title}} \quad \left\{ A.\text{id}=B.\text{dz} \right\} \]
\[ \Pi_{\text{title, id}} \quad B \]
\[ \sigma_{A.\text{name}=\text{'Hugo'}} \]

unclustered index on A.name

\[ \Pi_{\text{title}} \quad \left\{ A.\text{id}=B.\text{dz} \right\} \]
\[ \Pi_{\text{title, id}} \quad B \]
\[ \sigma_{A.\text{name}=\text{'Hugo'}} \]

covering index on A.name, A.title, A.id

\[ \Pi_{\text{title}} \quad \left\{ A.\text{id}=B.\text{dz} \right\} \]
\[ \Pi_{\text{title, id}} \quad B \]
\[ \Pi_{\text{title, id}} \quad \sigma_{A.\text{name}=\text{'Hugo'}} \]

\[ \Rightarrow \text{index-only mode} \]
Estimating Index Access Costs

Clustered index on A.name

\[ \Pi_{\text{title, id}} \]

\[ \sigma_{A.\text{name}=\text{‘Hugo’}} \]

expected costs:
one random I/O to fetch leaf,
one random I/O in store,
then ISAM

Unclustered index on A.name

\[ \Pi_{\text{title, id}} \]

\[ \sigma_{A.\text{name}=\text{‘Hugo’}} \]

expected costs:
as clustered index,
plus random I/O to store depending on selectivity

Covering index on A.name, A.title, A.id

\[ \Pi_{\text{title, id}} \]

\[ \sigma_{A.\text{name}=\text{‘Hugo’}} \]

expected costs:
one random I/O to fetch leaf,
then ISAM no need to go to store
Estimating Scan vs Index Access Costs

**Scan on A**

\[ \prod_{\text{title}, \text{id}} \sigma_{A.\text{name}='Hugo'} \]

**Unclustered index on A.name**

\[ \prod_{\text{title}, \text{id}} \sigma_{A.\text{name}='Hugo'} \]

**Expected costs:**
- One random I/O to go to start of A, then sequential read

**Expected costs:**
- As clustered index, plus random I/O to store depending on selectivity

**Graph:**
- Costs vs Selectivity
- Results: Few vs Many

**Legend:**
- Unclustered index on A.name
- Scan on A
Copyrights

© iStock.com:

Rastan, moenez