

Query processing cost formulae

Legend

Symbol	Description
NKeys(Col)	The number of distinct values of column Col
High(Col)	The highest value of column Col
Low(Col)	The lowest value of column Col
NTuples(R)	The number of tuples of relation R
NPages(R)	The number of pages of relation R
NPages(I)	The number of pages of index I
Height(I)	The height of index I
$\prod RF_i$	The product of all reduction factors
$\prod NTuples(R_i)$	The product of the numbers of tuples of all relations taking part in a join
num_passes(R)	The number of passes for sorting relation R
PF	Projection factor (portion of all columns)

1. Reduction factor (Selectivity)

a. Col = value

$$RF = 1/NKeys(Col)$$

b. Col > value

$$RF = (High(Col) - value) / (High(Col) - Low(Col))$$

c. Col < value

$$RF = (val - Low(Col)) / (High(Col) - Low(Col))$$

d. Col_A = Col_B (for joins)

$$RF = 1/ (Max (NKeys(Col_A), NKeys(Col_B)))$$

e. In no information about NKeys or interval, use a “magic number” 1/10

$$RF = 1/10$$

2. Result size calculations

a. Single table

$$Result_size = NTuples(R) * \prod RF_i$$

b. Joins

$$Result_size = \prod NTuples(R_i) * \prod RF_i$$

3. Indexing Cost

a. B+-tree index

- i. Just a single tuple (selection over a primary key)

$$\text{Cost} = \text{Height(I)} + 1$$

- ii. Clustered index (multiple tuples)

$$\text{Cost} = (\text{NPages(I)} + \text{NPages(R)}) * \prod \text{RF}_i$$

- iii. Unclustered (multiple tuples)

$$\text{Cost} = (\text{NPages(I)} + \text{NTuples(R)}) * \prod \text{RF}_i$$

b. Hash Index

- i. Just a single tuple (selection over a primary key)

$$\text{Cost} = 1.2 + 1 = 2.2$$

- ii. Clustered index (multiple tuples)

$$\text{Cost} = (\text{NPages (R)}) * \prod \text{RF}_i * 2.2$$

- iii. Unclustered index (multiple tuples)

$$\text{Cost} = (\text{NTuples(R)}) * \prod \text{RF}_i * 2.2$$

4. Sequential Scan (i.e. Heap Scan) Cost

$$\text{Cost} = \text{NPages(R)}$$

5. Joins (between relations R and S, R = outer, S = inner) Cost

a. NLJ

i. Tuple-oriented NLJ

$$\text{Cost} = \text{NPages}(\text{R}) + \text{NTuples}(\text{R}) * \text{NPages}(\text{S})$$

ii. Page-oriented NLJ

$$\text{Cost} = \text{NPages}(\text{R}) + \text{NPages}(\text{R}) * \text{NPages}(\text{S})$$

iii. Block-oriented NJL (for block_size B)

$$\text{Cost} = \text{NPages}(\text{R}) + \text{ceil}(\text{NPages}(\text{R}) / (B - 2)) * \text{NPages}(\text{S})$$

b. Hash Join

$$\text{Cost} = 3 * (\text{NPages}(\text{R}) + \text{NPages}(\text{S}))$$

c. Sort-Merge Join

$$\begin{aligned} \text{Cost}_{\text{SMJ}} &= \text{NPages}(\text{R}) + \text{NPages}(\text{S}) + \\ &2 * \text{NPages}(\text{R}) * \text{num_passes}(\text{R}) + \\ &2 * \text{NPages}(\text{S}) * \text{num_passes}(\text{S}) \end{aligned}$$

6. Projections Cost

a. Sort-based

$$\begin{aligned} \text{Cost} &= \text{ReadTable} + \text{WriteProjectedPages} + \text{SortProjectedPages} + \\ &\text{ReadProjectedPages} = \text{NPages}(\text{R}) + \text{NPages}(\text{R}) * \text{PF} + 2 * \text{num_passes} * \\ &\text{NPages}(\text{R}) * \text{PF} + \text{NPages}(\text{R}) * \text{PF} \end{aligned}$$

b. Hash-Based

$$\begin{aligned} \text{Cost} &= \text{ReadTable} + \text{WriteProjectedPages} + \text{ReadProjectedPages} \\ &= \text{NPages}(\text{R}) + \text{NPages}(\text{R}) * \text{PF} + \text{NPages}(\text{R}) * \text{PF} \end{aligned}$$