Query Optimization

November 2020

1 Exercise 2

1.1 Task 1

Task 1.1
To be shown:

$$\sigma_a(R_1 \bowtie_b R_2) \equiv \sigma_a(R_1) \bowtie_b R_2 \qquad \mid a = p_1,\ b = p_2,\ F(a) \subseteq A(R_1)$$

$$\begin{aligned}
&\textit{Let } c \in \sigma_a(R_1 \bowtie_b R_2) \\
\iff &\exists c \in (R_1 \bowtie_b R_2) \mid a(c) \\
\iff &\exists c_1 \in R_1, c_2 \in R_2 \mid c = c_1 \circ c_2 \wedge b(c) \wedge a(c) \\
&\textit{Because } F(a) \subseteq A(R_1) \\
\iff &\exists c_1 \in R_1, c_2 \in R_2 \mid c = c_1 \circ c_2 \wedge b(c) \wedge a(c_1) \\
\iff &\exists c_1 \in \sigma_a(R_1), c_2 \in R_2 \mid c = c_1 \circ c_2 \wedge b(c) \\
\iff &\exists c \mid \sigma_a(R_1) \bowtie_b R_2
\end{aligned}$$

Task 1.2
The equivalence does not hold true for outer joins.

If R_1:

$p1$	$p2$

And if R_2:

p_2
1
2

Then:

$\sigma_a(R_1 \bowtie_b R_2) = \emptyset$

And:

$\sigma_a(R_1) \bowtie_b R_2$:

p_2	p_1
1	-
2	-

1.2 Task 2

Task 2.1

$R_1.x = key$	$1 \div \lvert R_1 \rvert$
$R_1.x \neq key$	$1 \div distinct(R_1)$

Task 2.2

	$R_2.y = key$	$R_2.y \neq key$
$R_1.x = key$	$1 \div max(\lvert R_1 \rvert, \lvert R_2 \rvert)$	$1 \div \lvert R_2 \rvert$
$R_1.x \neq key$	$1 \div \lvert R_1 \rvert$	$1 \div max(distinct(R_1), distinct(R_2))$

1.3 Task 3

Task 3.1: NL

Given, average access time for relation $x = aat_x$, pg_x is amount of seconds needed to transfer x pages, performing the Nested Loops Join of R and S takes:

$$\begin{aligned} & \lvert R \rvert * (aat_R + pg_1 + aat_S + pg_{100000}) \\ = & \ 50000 * (10^{-2} + 10^{-4} + 10^{-2} + \frac{100000}{10000}) \\ = & \ 501005 \end{aligned}$$

seconds.
Task 3.2: BNL

$$\begin{aligned} & \lvert R \rvert * (aat_R + pg_{100} + aat_S + pg_{100000}) \\ = & \ 10 * (10^{-2} + 10^{-2} + 10^{-2} + \frac{100000}{10000}) \\ = & \ 100.3 \end{aligned}$$

seconds.