Homework 3 – Part 2 (Query Optimization)

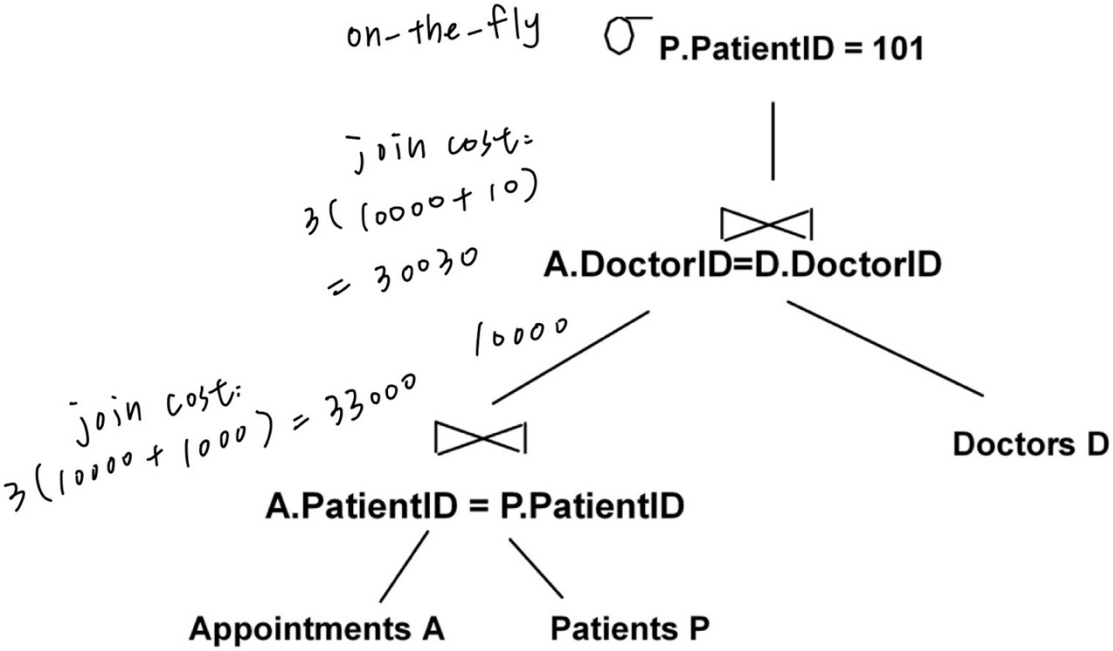
Q1.a

A.PatientID = P.PatientID # of rows = **1,000,000**

A.DoctorID=D.DoctorID # of rows = **1,000,000**

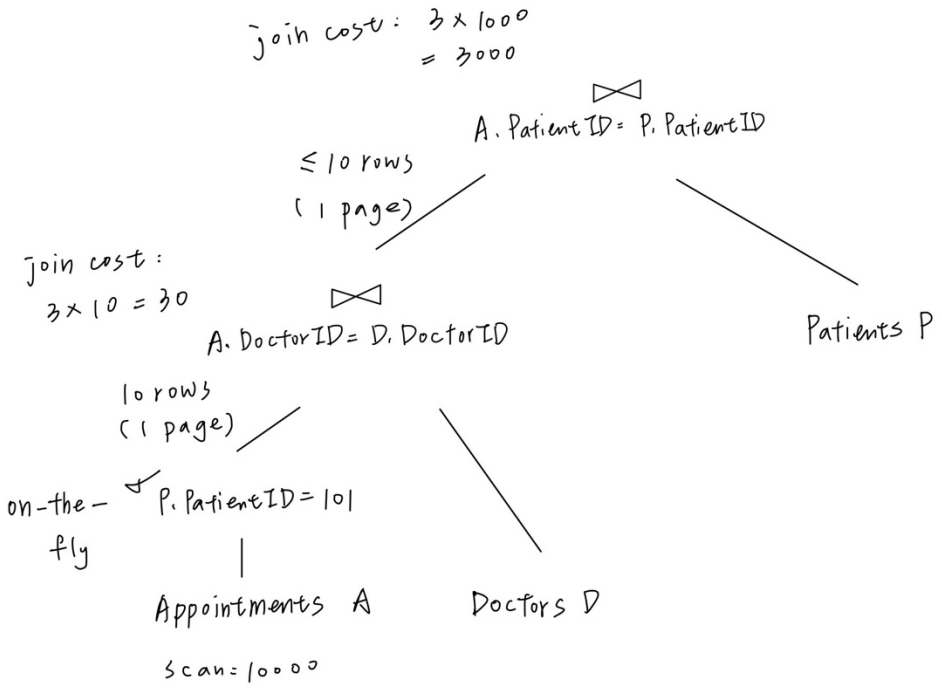
P.PatientID = 101 # of rows = **10**

Q1.b



Total cost: 33000 + 1000 + 30030 = **64030**

Q1.c



Total cost: 10000 + 30 + 3000 = **13030**

For the bottom hash join, it becomes 3N thanks to on-the-fly (it only has one pages, so it can sit in memory) so that it just needs to take the inner relations (D) into consideration.

For the upper hash join, it is the same with the bottom one.

Q2.a

Independence: # of rows = 3282 * (1/3) * (1/50) ≈ **22**

Dependence: # of rows = 3282 * (1/3) * (3/50) ≈ **66**

Q2.b

# of rows = 180 / (44 – 25 + 1) = **9**

Q3.a

False.

e.g.

$A = \{1, 4\}$

$B = \{1, 3\}$

$C = \{1, 2, 3\}$

$C - (A \cap B) = \{2, 3\}$

$(C - A) \cap B = \{3\}$

Accordingly, this equation is false.

Q3.b

False.

e.g.

Assume

R relation

Y	Z
2	3
5	6

S relation

C	Y
4	1
6	2

$\pi(z)\ (R \bowtie S)$

Z
3

$(\pi(z)\ (R)) \bowtie S$

Z	Y	Z
3	4	1
3	6	2
6	4	1
6	6	2

Accordingly, this equation is false.