Consider the following join which is given in your textbook (§19.5.2, p.720)

OP7: DEPARTMENT $\bowtie_{Mgr\_ssn=Ssn}$ EMPLOYEE

Calculate the **cost-based estimation** of executing this query using (a) DEPARTMENT as the outer loop and then (b) EMPLOYEE as the outer loop using textbook single-loop join algorithms, i.e. J2. (Note: The join formulas from the textbook (pp.718-19) will be provided during the final exam).

Assume that the Column, Table, and Index meta-data are the values in Figures 1-3.

Table_name	Column_name	Num_distinct
DEPARTMENT	Dnumber	50
DEPARTMENT	Mgr_ssn	50
EMPLOYEE	Ssn	10000
EMPLOYEE	Dno	50
EMPLOYEE	Salary	500

FIGURE 1: COLUMN INFORMATION

Table_name	Num_rows	Blocks
DEPARTMENT	50	5
EMPLOYEE	10000	2000

FIGURE 2: TABLE INFORMATION

Table_name	Column_index	Index_type	Unique	Blevel (i.e. x-value)
DEPARTMENT	Dnumber	PRIMARY	YES	1
DEPARTMENT	Mgr_ssn	SECONDARY	YES	1
EMPLOYEE	Ssn	PRIMARY	YES	1
EMPLOYEE	Dno	SECONDARY	NO	1
EMPLOYEE	Salary	SECONDARY	NO	1

FIGURE 3: INDEX INFORMATION

To begin, we must calculate the blocking factor of the join bfr_{DEP_EMP}. To do this, we need to first calculate the blocking factor of the individual tables bfr_{DEP} and.bfr_{EMP}.

The blocking factor of the DEPARTMENT table is the number of rows/records per block.

$bfr_{DEP} = 50 / 5 = 10$

The blocking factor of the EMPLOYEE table is the number of rows/records per block.

$bfr_{DEP} = 10000 / 2000 = 5$

The blocking factor of the joined table is the number of joined rows/records per block.

$bfr_{DEP\_EMP} = \lfloor 1/ (1/bfr_{DEP} + 1/bfr_{DEP}) \rfloor$

$bfr_{DEP\_EMP} = \lfloor 1/ (1/10 + 1/5) \rfloor$

$bfr_{DEP\_EMP} = \lfloor 1/ 3/10 \rfloor$

$bfr_{DEP\_EMP} = \lfloor 10/3 \rfloor$

$bfr_{DEP\_EMP} = 3$

Now we can calculate the costs of the two different J2 single-loop join operations…

Cost-based estimations

DEPARTMENT as the outer loop

If DEPARTMENT is the outer loop, then the index attribute of the inner loop is EMPLOYEE.Ssn, which is a primary key index. Therefore, we use join formula **J2c** to estimate the cost.

$Cost_{J2c} = b_{DEP} + (r_{DEP} * (x_{Ssn} + 1)) + \lceil((js_{OP6} * r_{DEP} * r_{EMP}) / bfr_{DEP\_EMP})\rceil$
$Cost_{J2c} = 5 + (50 * (1 + 1)) + \lceil((1/10000 * 50 * 10000) / 3)\rceil$
$Cost_{J2c} = 5 + 100 + \lceil(50 / 3)\rceil$
$Cost_{J2c} = 122$

EMPLOYEE as the outer loop

If EMPLOYEE is the outer loop, then the index attribute of the inner loop is DEPARTMENT.Mgr_ssn, which is a unique secondary index. Therefore, we use join formula **J2a** to estimate the cost.

The selection cardinality (s_{Mgr_ssn}) of DEPARTMENT.Mgr_ssn is 1, because there is exactly one department manager per department.

$Cost_{J2a} = b_{EMP} + (r_{EMP} * (x_{Mgr\_ssn} + 1 + s_{Mgr\_ssn})) + \lceil((js_{OP6} * r_{DEP} * r_{EMP}) / bfr_{DEP\_EMP})\rceil$
$Cost_{J2a} = 2000 + (10000 * (1 + 1 + 1)) + \lceil((1/10000 * 50 * 10000) / 3)\rceil$
$Cost_{J2a} = 2000 + 30000 + \lceil(50 / 3)\rceil$
$Cost_{J2a} = 32017$