Exercises

[Note: The symbol for the natural join is written as $\times_{A=B}$]

Q1. If you have a file occupying 1500 blocks and 5 buffer blocks are available in main memory.

a) How many runs will you produce in sort phase using external sort-merge algorithm?
b) How many passes are necessary to completely sort the files during the merge phase?

Q2. Consider a relation R(A,B) containing 1000000 records with bfr =10. Assume A is a candidate key and that exists a secondary index of height 4 on A. Estimate the cost of the query $\sigma_{A=100}(R)$ for methods

i) Linear search ii) Using the index

Q3. Consider two relations Depositor and Customer. Depositor constitutes of 100 blocks holding 5000 records. Customer constitutes of 400 blocks holding 10,000 records. Evaluate the cost of the join operation DEPOSITOR $\times_{A=B}$ CUSTOMER using only the number of blocks need to perform the operation.

a) Nested loop join
b) Single loop join assuming Customer has primary index on B and height is 4. Assume main memory can hold only one block of Depositor in buffer.
c) Sort-merge join assuming two relations are already sorted.

Assume that the blocking factor for the joined relation is 20.

Q4. Show how to derive the equivalence by sequence of transformations using the heuristic rules.

$\sigma_{\theta 1 \text{and } \theta 2}(E1 \times_{\theta 3} E2) = \sigma_{\theta 1}(E1 \times_{\theta 3} (\sigma_{\theta 2}(E2)))$, where θ2 involves only attributes from E2.

Q5. Given the relation below

Customer (<u>CustSK</u>, Name, Phone, Email)
Stove (<u>SNo</u>, Type, Version, DateofManufacture)
Registration (<u>CustSK, SNo</u>, Date)
StoreRepair (<u>RepairInvoice</u>, SNo, Date, Description, Cost, CustSK)

The following are the details of each relation:

i) Customer has 500 records with 10 disk blocks.
ii) Stove has 20 records with 5 disk blocks
iii) Registration has 100 records with 10 disk blocks
iv) StoreRepair has 50 records with 5 disk blocks
v) There is an ordering index on CustSk of Customer, levels is 1
vi) There is secondary index on SNo of Registration, levels is 4. There are 10 distinct values for SNo
vii) Blocking factor of the joined relation is 8.

Operations are:

a) $\sigma_{\text{CustSK} > 34890}$ (Customer)
b) Stove $\times_{\text{SNo} = \text{SNo}}$ Registration

Translate the following queries into Relational Algebra and also calculate the cost estimates:

Q6. Given the relation below

Book (ISBN, Title, Publisher, PubYear)
Author (AName, Bdate)
Publisher (PName, Address)
Wrote (ISBN, AName)

a) Select W.AName
From Book B, Wrote W
Where B.ISBN = W.ISBN and Title = 'Programming In C';

b) Select P.Address
From Publisher P, Book B
Where P.PName = B.Publisher and Title = 'Database Fundamentals' or Title = 'Data Management';

The details for the relations are as follows:

i) Book has 5000 records with 100 disk blocks.
ii) Author has 100 records with 20 disk blocks
iii) Publisher has 30 records with 10 disk blocks
iv) Wrote has 200 records with 30 disk blocks
v) Secondary index on ISBN of Wrote, level is 4, selection cardinality is 3, first level index is 2
vi) Clustering index on Title, level is 4, no of distinct values is 10
vii) Primary index on PName, level is 1, selection cardinality = 10
viii) Blocking factor of the joined relation is 5.