Query Processing

1. What are the steps in Query Processing? Explain each step.

2. What is the goal of Query optimization? Why is it important?

3. What types of information is stored in the system catalogs? Why?

4. Consider the following schema:

   Sailors (sid, sname, rating, age)

   Boats (bid, bname, color)

   Reserves (sid, bid, day)

   Consider the following query:

   SELECT S.sname

   FROM Sailors S, Reserves R, Boats B

   WHERE S.sid = R.sid AND R.bid = B.bid AND B.color = 'red'

   a. Create the relational algebra expression for the query above.
   b. Draw the graphical representation of the relational algebra expression

5. Consider the join $R \bowtie_{R.a=S.b} S$, given the following information about the relations to be joined. The cost metric is the number of page I/Os unless otherwise noted, and the cost of writing the result is ignored.

   Relation R contains 10,000 tuples and has 10 tuples per page.

   Relation S contains 2,000 tuples and also has 10 tuples per page.

   Attribute *b* of relation S is the primary key for S.

   Both relations are stored as simple heap files.

   Neither relation has any indexes built on it.

   52 buffer pages are available.

   a. What is the cost of joining R and S using simple nested loops join?

   b. What is the cost of joining R and S using a page-oriented nested loops join?

   c. What is the cost of joining R and S using a block-nested loops join?

   d. Assuming that there exists a B+ tree index (with height 3) on *b* column of relation S, what is the cost of performing an index nested loops join? Explain your answer.