Warm up (20 points)
1. (5 points) Write your name and student id on your blue book.
2. (5 points) Provide the definition for the disk seek operation.
3. (5 points) True or False: One may construct a clustered B+-tree on the salary attribute of the Employee table and a non-clustered hash index on the SS# of the same table. Explain your answer.
4. (5 points) True or False: The values in the internal nodes of a B+-tree must appear in at least one of its leaf nodes. Explain your answer.

Normal Forms (10 points)
5. (10 points) State the difference between 2\textsuperscript{nd} and 3\textsuperscript{rd} normal forms. Provide an example for illustration purposes.

Physical Data Design (10 points)
6. (10 points) Describe the indexed heap organization of records in a disk page. Provide a drawing to show its record organization.

Relational Algebra (20 points)
Recall the yacht club reservation database consisting of the following three tables:
\begin{itemize}
  \item Sailors(sid, sname, rating)
  \item Boats(bid, bname, bcolor)
  \item Reserve(sid, bid, date)
\end{itemize}
7. (10 points) Write the algebraic expression to retrieve those sailors who have reserved boats named Tenacious and Resolute.

8. (10 points) Write the algebraic expression to delete the reservations for a boat named Tenacious by those sailors whose rating is lower than 5.

SQL (20 points)
Assume the same three tables for the yacht club reservation database.
9. (10 points) Retrieve those sailors who have reserved Tenacious and not Resolute. It is acceptable to retrieve either sid or the whole row of the qualifying sailors.

10. (10 points) For each boat with more than 5 reservations, retrieve the boat name and the average rating of the sailors who have reserved it.

Query Processing and Cost Estimation (20 points)
Assume the yacht club has 100,000 members, t(Sailors)=100,000. 10 Sailor records fit per disk page, P(Sailors)=10,000. The yacht club assigned integer ratings of 1 to 10 to each sailor, v(rating, Sailors)=10.
11. (10 points) Assuming a non-clustered, secondary B+-tree index on the rating attribute of the Sailors table, how does a system process a range selection predicate that retrieves those sailors with a rating greater than or equal to 7, rating ≥ 7. State your assumptions (if any) clearly.

12. (10 points) Assume a 2 level deep clustered B+-tree on the rating attribute of sailors. It consists of 20,000 leaf page, lp(I)=20,000. What is the estimated number of disk I/O operations to retrieve those sailors whose rating equals 2. State your assumptions (if any) clearly.