Assignment #3

1. **[Parallel Data Models] (30)**
   a. (15) Assume a program P running on a single-processor system takes time T to complete. 20% of P can only be executed sequentially on a single processor, and the rest is "embarrassingly parallel" in that it can be easily divided into smaller tasks executing concurrently across multiple processors. What are the best time costs to execute P using 2, 4, 8 machines (expressed by T). What are the speed-up respectively? What are the optimal speed-up given infinitely amount of machines?
   b. (15) Describe and compare the pros and cons of the three architecture for parallel systems.

2. **[ACID vs BASE] (40)** This set of questions are related to data consistency
   a. (15) What is CAP Theory? Consider an example cluster that contains three servers S1, S2 and S3, each located in a different geographical areas. Assuming data is partitioned across the three servers, explain CAP theory for the example cluster.
   b. (15) Consider the relation Accounts(acctNo, customerName, balance) with acctNo as primary key. Consider the following two SQL statements that conduct a request "transfer $200 from account A1 to B1".
      i. Add $200 to account B1:

         UPDATE Accounts SET balance=balance+200 WHERE acctNo = B1

      ii. Subtract $200 from account A1:

         UPDATE Accounts SET balance=balance - 200 WHERE acctNo=A1

Use this example and necessary scenarios to show when Atomicity, Consistency, Isolation and Durability can be violated.

c. (10) What is "BASE"? give an example of "BASE" data consistency model and compare it to ACID.

3. **[Quorum Consensus] (15)**

   (15) We introduced Quorum Consensus method to ensure consistent data can be fetched in read/write operations. Describe Quorum Consensus and explain why it works.

4. **[Relational DB - Query Processing]** (15) This question tests the understanding of basic relational database search operators. Consider a join $\bowtie_{R.A=S.B}$. We ignore the cost of output the result, and measure the cost with the number of I/Os. Given the information about relations to be joined below:
   Relation S contains 20,000 tuples and has 10 tuples per block. Relation R contains 100,000 tuples and has 10 tuples per block. Attribute B is the primary key of S. In total, 52 blocks are available in memory. Assume neither relation has any index.

   - (15) Describe a block nested join algorithm. Give the cost of joining R and S with a block nested loops join.