Question 1. Consider the two relations Works(eid, did, pct_time) and Dept(did, budget, floor). Translate the following SQL query to the relational algebra and heuristically optimize this expression by applying the algebraic rewriting laws:

```sql
SELECT D.did, COUNT(W.eid)
FROM Dept D, Works W,
WHERE
    (D.floor = 1 OR D.budget < 15000)
    AND D.did IN
        ( SELECT D2.did FROM Dept D2
            WHERE 2000 >= (SELECT SUM(W2.pct_time)
                FROM Works W2
                WHERE W2.did = D2.did)
        )
GROUP BY D.did
```

Use the algorithm studied in the course for the translation. Give sub-results and motivate your answer.

Question 2. Prove that a conjunctive query \( Q_1 \) is contained in a conjunctive query \( Q_2 \) if, and only if, there exists a homomorphism from \( Q_2 \) into \( Q_1 \).

Question 3. Consider the following relational algebra expression over the relation \( R(A, B, C) \):

\[
\pi_{R_2.A,R_3.B,R_1.C} \\
\sigma_{R_1.A=R_5.A} \sigma_{R_2.A=R_4.A} \sigma_{R_1.B=R_2.B} \sigma_{R_4.B=R_5.B} \sigma_{R_2.C=R_3.C} \sigma_{R_5.C=R_1.C} \\
\rho_{R_1(R)} \times \rho_{R_2(R)} \times \rho_{R_3(R)} \times \rho_{R_4(R)} \times \rho_{R_5(R)}
\]

Optimize this expression by removing redundant joins. Give sub-results and motivate your answer.
**Question 4.** Consider the clustered relations \( R(A, B, C) \) and \( S(B, C) \). Relation \( R \) has a clustered B-tree index on \( A \). Relation \( S \) has a hashing index on \((B, C)\) together. The attributes \((B, C)\) form a key for \( S \). An \( R \)-record comprises 50 bytes while an \( S \)-record comprises 40 bytes. Blocks are 4000 bytes large, and there are 5 main memory buffers available. The statistics show that \( R \) has 36000 tuples, and that \( S \) has 10000 tuples. They also show that the values of \( R.A \) lie within the range \([0, 100]\), uniformly distributed.

The query compiler has already obtained the following logical query plan:

\[
\sigma_{\Sigma_{\text{sum}(C)}>1500} \gamma_{B, \Sigma_{\text{sum}(C)}}(\pi_{B,C}(\sigma_{A\geq 90}(R)) \cap S)
\]

Construct a sufficiently optimal physical query plan. Use disk I/Os as your optimization metric. Motivate your answer, and describe any assumptions that you make. It suffices to make only locally-optimal decisions (in other words: you may use the greedy algorithm, and your solution need not be globally optimal.)

**Question 5.** Assume that we control concurrency by means of a validation-based scheduler. Explain what this scheduler does upon the following sequence of actions. Here, \( R_i(X) \) is used to mean “transaction \( T_i \) starts, and its read set is the list of database elements \( X \); \( V_i \) means “\( T_i \) attempts to validate”; and \( W_i(X) \) means “\( T_i \) finishes, and its write set was \( X \).”

1. \( R_1(A, B); R_2(B, C); R_3(C); V_1; V_2; V_3; W_1(C); W_2(B); W_3(A); \)
2. \( R_1(A, B); R_2(B, C); V_1; R_3(C, D); V_3; W_1(A); V_2; W_2(A); W_3(D); \)