Assignment: Advanced Database Management Systems (Query Processing)

Part 1: Relational Algebra

Task: Solve the following problems using relational algebra expressions.

1. Write the relational algebra query for the following SQL statement:

```
SELECT sname, city
FROM Supplier
WHERE rate > 5;
```

2. Given the relations below:
   - **Suppliers**: (sid, sname, rate, city)
   - **Products**: (pno, pname, weight)
   - **Supplies**: (sid, pno, qty)

   Write relational algebra queries to:

   - Find all supplier names who supply a product with weight > 20.
   - Find the distinct product names supplied by suppliers from 'New York'.

Part 2: Query Processing

1. **Parsing and Translating**:
   Translate the following query into its internal relational algebra form:

```
SELECT DISTINCT sname
FROM Supplier S, Supplies SP
WHERE S.sid = SP.sid AND SP.pno = 'P2';
```

2. **Optimization**:
   - Transform the above relational algebra expression into an optimized form using equivalence rules.
   - Provide a step-by-step explanation of how optimization reduces query cost.

Part 3: Query Execution and Cost Estimation

1. **Cost Estimation**:
   - For a relation R with:
     - NR=1000NR = 1000NR=1000 tuples
     - BR=200BR = 200BR=200 blocks
     - FR=5FR = 5FR=5 (blocking factor)
     - SC(A,R)=10SC(A, R) = 10SC(A,R)=10
   - Estimate the cost of:

        - Linear search
        - Binary search
    - Compare the results.
2. **Join Algorithms**:
    - Explain the working of any two join algorithms (e.g., Nested-Loop Join, Sort-Merge Join).
    - Choose the most efficient algorithm to join relations R and S, assuming RRR has 500 tuples and SSS has 200 tuples. Justify your choice.

Part 4: Case Study

1. Consider the execution plans for the following query:

```
SELECT *
FROM Students S, Enrolled E, Courses C
WHERE S.name = E.name AND E.course = C.course;
```

    - Compare the two execution plans provided in the slides.
    - Suggest improvements to minimize disk access or computation time.