CS 5300 module5

Name Student ID

Problem 1(25 Points)

1) Define the term query processing.

Query processing is defined as the activities involved in parsing, validation, translation, optimization, and execution of a query.

2) Define the term query optimization.

Query optimization is the activity of choosing an efficient execution strategy for processing a query

3) Query optimization can be done in two fashions: static or dynamic. Explain each and enumerate advantages and compare and contrast them against each other.

In dynamic approach query decomposition and optimization is carried out every time the query is run.

In static approach query is parsed, validated, and optimized once.

In dynamic approach, query optimization will be done up-to-date and changes in database statistics will be enforced and hence we may end up with a better execution time, however, the overhead of repeating the query processing phases is the penalty to pay.

So if databases are static in nature, static approach would be a better choice.

4) Assume the following SQL query and related relations:

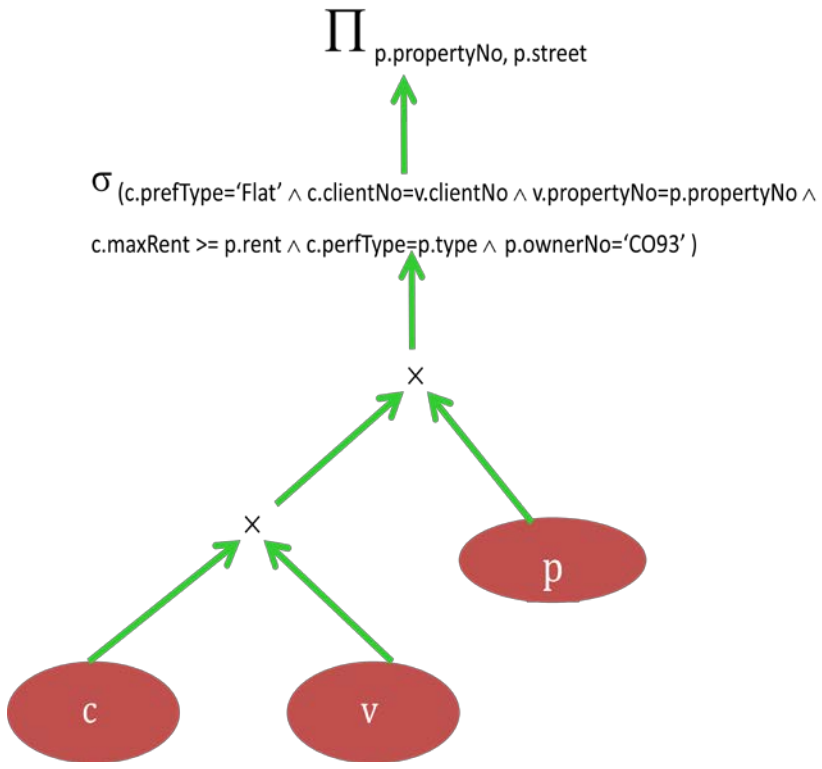
PropertyForRent(propertyNo, street, city, postcode, type, room, rent, ownerNo)
Client(clientNo, fName, lName, telNo, prefType, maxRent)
Viewing(clientNo, propertyNo, viewDate, comment)

```
SELECT    p.propertyNo, p.street
FROM      Client c, Viewing v, PropertyFor Rent p
WHERE     c.prefType = 'Flat' AND c.clientNo = v.clientNo AND
          v.propertyNo = p.propertyNo AND c.maxRent >= p.rent AND
          c.prefType = p.type AND p.ownerNo = 'CO93';
```

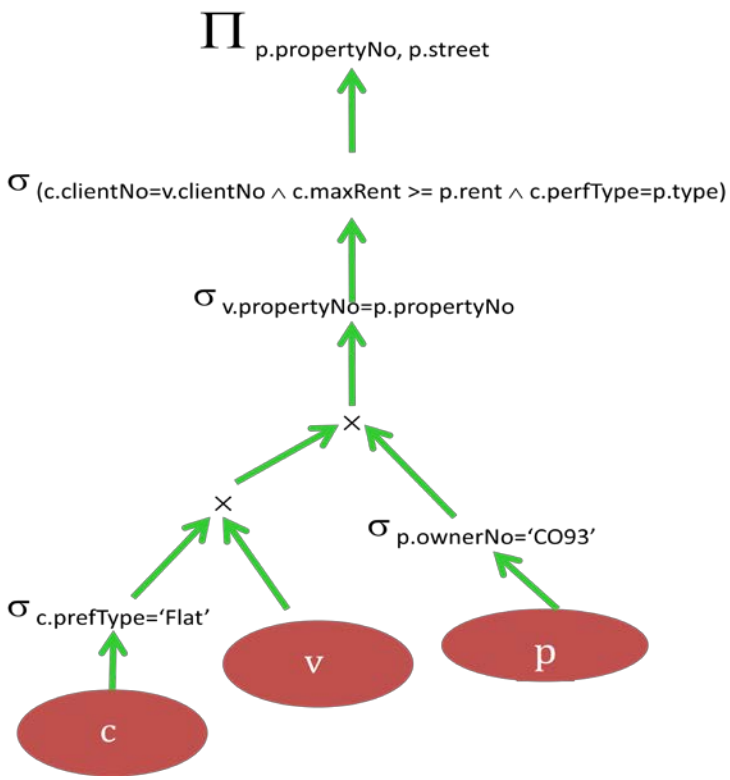
For the purposes of this example, assume that there are fewer properties owned by owner CO93 than prospective renters who have specified a preferred property type of FLAT.

Go through different stages of query processing to translate and optimize the aforementioned query (make sure to justify your actions, show relational algebra expression, and the query tree after each optimization step).

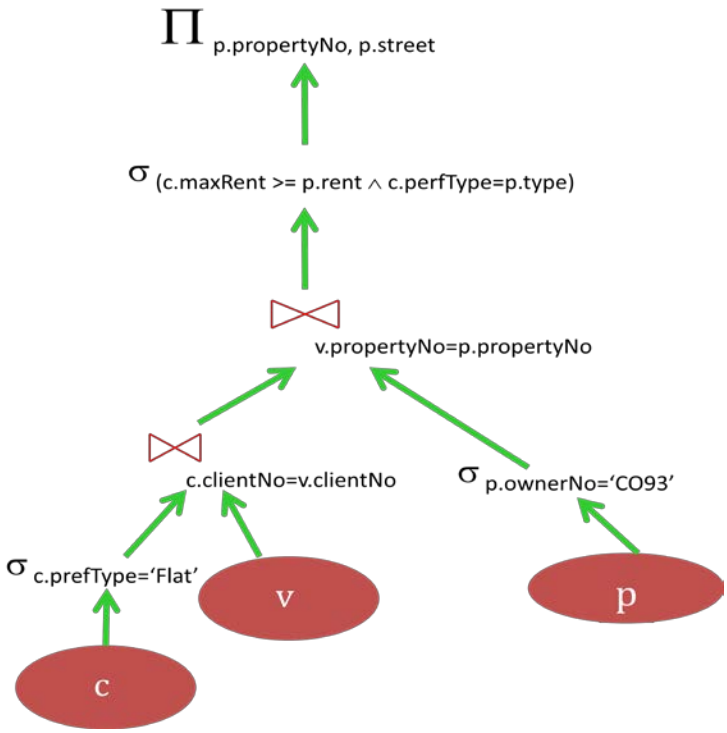
$$\Pi_{p.propertyNo,\ p.street}(\sigma_{c.prefType='Flat' \wedge c.clientNo=v.clientNo \wedge v.propertyNo=p.propertyNo \wedge c.maxRent >= p.rent \wedge c.prefType=p.type \wedge p.ownerNo='CO93'}((c \times v) \times p))$$



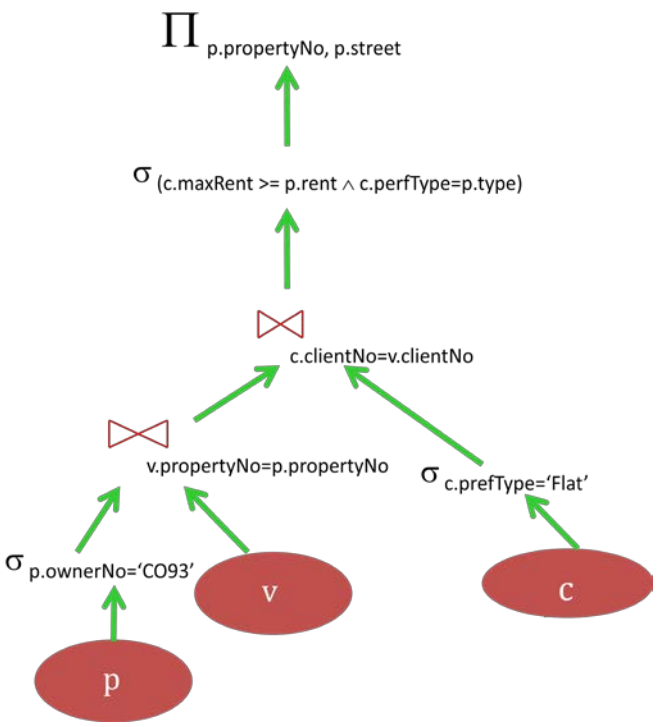
Breaking selection and pushing it down the tree



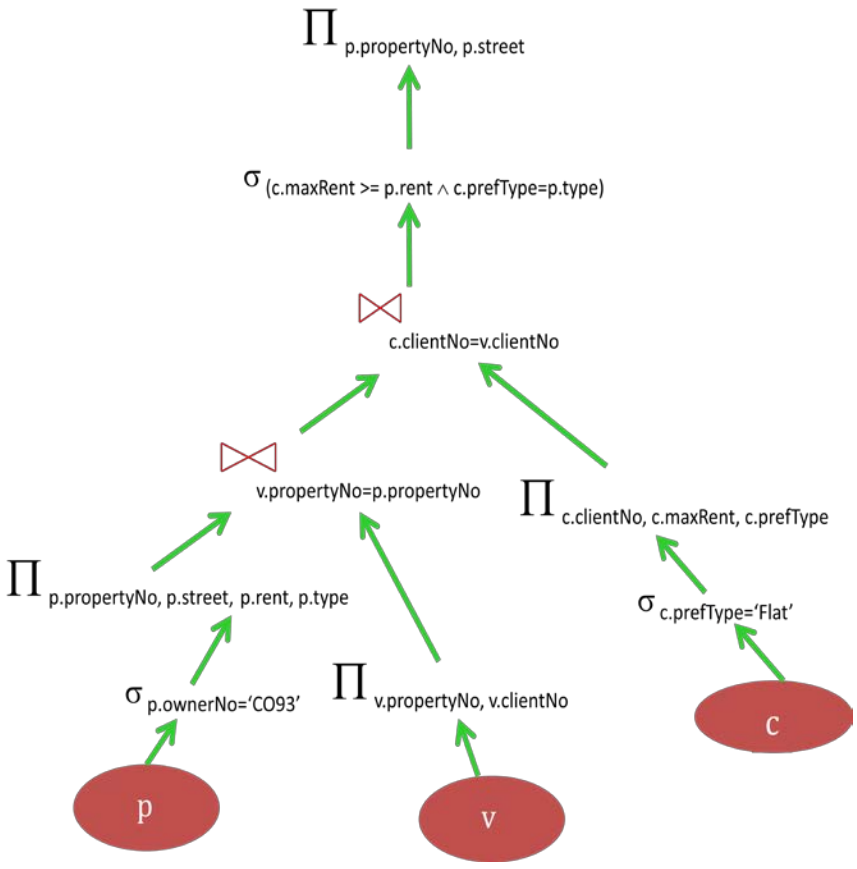
Cartesian products to Join



Reorder join operations to perform more restrictive selection first



Pushing projections down



Problem #2 (10 Points)
Consider the following two relations:

S (S_{id}: integer, S_{name}: string, rating: integer, age: real)
R (S_{id}: integer, bid: integer, day: dates, r_{name}: string)

Further assume the following query:

```
SELECT   S.Sname
FROM     R, S
WHERE    R.Sid = S.Sid
    AND R.bid = 100  AND S.rating > 5
    AND R.day = '8/9/09'
```

a) What is the relational algebra expression of this query?

$$\Pi_{sname}((\sigma_{bid=100 \wedge rating=5 \wedge date='8/9/09'})(S \times R))$$

b) What is an optimized query plan for the aforementioned SQL query assuming that the underlying platform allows pipeline execution of operations (justify your choice of the plan)?