Problem 1 (Adapted from Problem 5a Fall 2016)

Given the following query with relations `R(A,B)` and `S(C)`:

```
SELECT R.A,  MAX(R.B) AS MaxR_B
FROM R, S
WHERE R.A = S.C AND R.B > 10
GROUP BY R.A
```

Suppose that `R` and `S` are partitioned across 3 different machines using block partitioning, and no indexes are available on any of the machines. Draw the relational algebra plan that you would use to execute the query above. Use only shuffle for joins. You do not need to indicate how joins are executed locally on each machine. (17 points)

(Start the diagram with 3 nodes in the bottom.)

Steps:

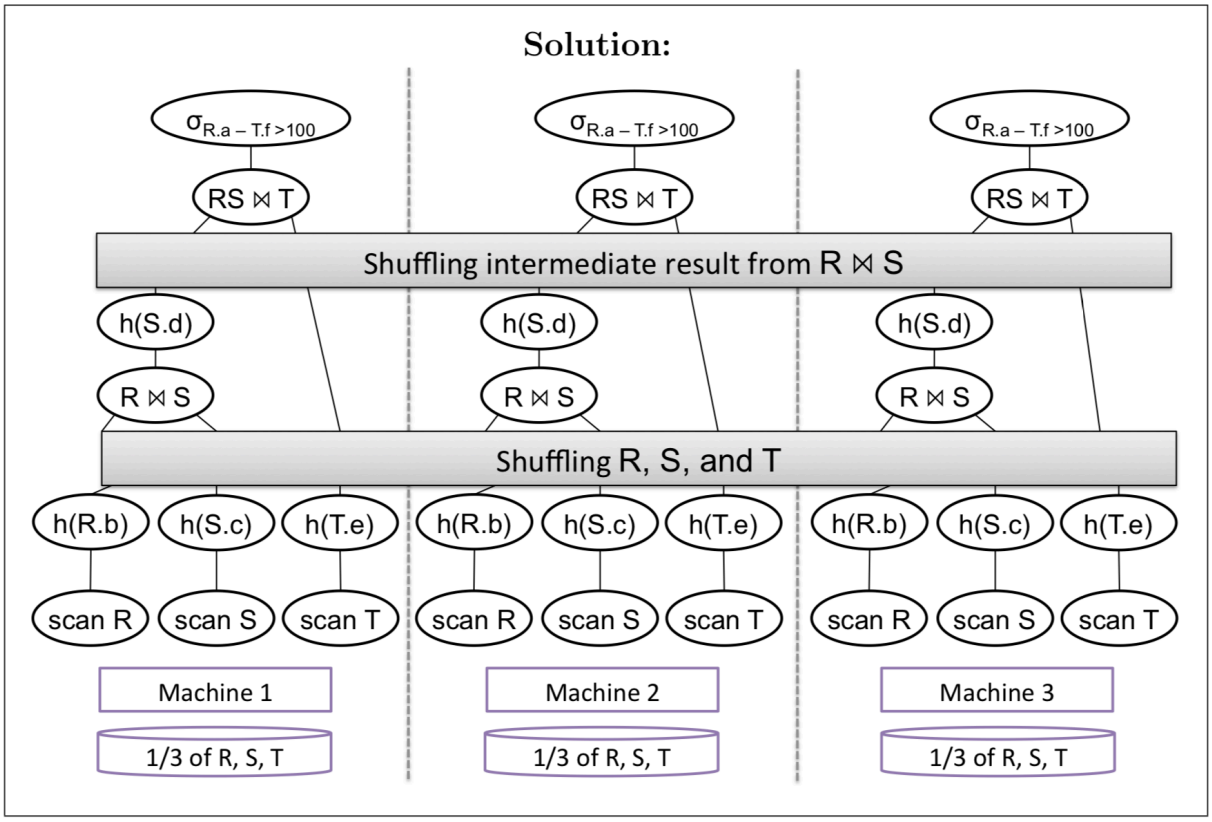
1. Scan R and S locally
2. Select on R.b locally
3. Hash on R.a and S.c and shuffle
4. Local join on R.a = S.c
5. Group By R.a, and make the aggregate
6. (Optional) Project and return final results

Problem 3 (Adapted from 344 15AU Final)

a) Consider relations R(a,b), S(c,d), and T(e,f). All three are horizontally random block partitioned across N = 3 machines (N1, N2, N3). Show a relational algebra plan for the following query and how it will be executed across the N = 3 machines. Use hash-join (a.k.a shuffle-join) operators. Include operators that need to re- shuffle data and add a note explaining how these operators will re-shuffle that data.

```
SELECT *
FROM R, S, T
WHERE R.b = S.c AND
      S.d = T.e AND
       (R.a - T.f) > 100
```

(Solution reference only)



b) Now consider the case where the R relation is very large and both S and T are very small. Show a plan that uses broadcast joins to compute the result of the query.
(Solution reference only)