Problem 2 (10 points)

You are using the hill climbing strategy to optimize the join $R(A,B) \bowtie S(A,C) \bowtie T(A,D) \bowtie V(A,E)$. Each of these relations is stored on a different site. Your goal is to minimize the amount of data transmitted over the network. The sizes of the relations are as follows:

Relation	number of tuples	bytes per tuple
R	100	100
S	200	80
T	100	50
V	150	100

We estimate the number of tuples in the result of $X \bowtie Y$ to be $0.1\times$ MIN$\{$ number of tuples in X, number of tuples in $Y\}$. The size of each result tuple is the sum of the tuple sizes of X and Y. (That is, we assume that the size of common attributes is negligible.)

For this problem we ignore the cost of sending the final answer to the site where it is needed, since all plans will have that final cost.

Your initial plan is to send R, T and V to the S site and perform the join there.

1. What is the total amount of data transmitted in the initial plan?

   ANSWER: $R \rightarrow S$ requires transmitting $100 \times 100 = 10,000$ bytes. $T \rightarrow S$ costs $5,000$ bytes and $V \rightarrow S$ costs $15,000$ bytes, for a total of $30,000$ bytes.

2. From the initial plan you look for improvements, by considering sending one of the relations to a site other than S, for joining with the local relation there. This is as we discussed in class. List all the transfers that should be considered in this first step. Use the notation $X \rightarrow Y$ for the transfer of relation X to the site that holds relation Y.

   ANSWER: $R \rightarrow T$, $T \rightarrow R$, $R \rightarrow V$, $V \rightarrow R$, $V \rightarrow T$, $T \rightarrow V$.

3. Of all the transfers considered, which one reduces the cost most? What is the new *total* cost for this plan?

   ANSWER: $T \rightarrow V$ reduces the cost by $13,500$ bytes, which is the maximum reduction. (Instead of sending $V \rightarrow S$ at a cost of $15,000$, we send $(T \bowtie V) \rightarrow S$ at a cost of $0.1 \times 100 \times 150 = 1,500$.) The total cost of the new plan is $16,500$ bytes.

4. Assume you perform a second hill climbing step from the plan of question (3) above. What is the resulting plan and what is its total cost?

   ANSWER: We repeat the optimization process, as if we had three relations only, S, R, and $T \bowtie V$. The two transforms to consider are $(T \bowtie V) \rightarrow R$ and $R \rightarrow (T \bowtie V)$. The first one gives us the largest reduction, so we use it. The final plan is $T \rightarrow V$; $(T \bowtie V) \rightarrow R$; $((T \bowtie V) \bowtie R) \rightarrow S$. The total cost is $6,750$ bytes.