



## Assignment 5 clarification about parallel join

Venkata Vamsikrishna Meduri

All Sections

In today's class I discussed parallel sort and parallel join that you are supposed to implement for assignment 5. There is a slight clarification about the implementation of parallel join that you may follow.

**1) Parallel sort** - For parallel sort, whatever we discussed in class holds good. You create five equi-range (equi-width) partitions using min and max from the sort column and run the sort individually on each partition and concatenate them into an output table.

**2) Parallel join** - In class, I raised a clarification question as in whether you can assume that both the left and the right tables have the same minimum and maximum (i.e., same join column range for both the left and the right tables). You CANNOT make that assumption. Therefore, following are the alternative methods in which you may implement the parallel join.

**Alternative 1:** You have to partition only one of the tables. In this case, you can choose to partition the left table into five equi-width range partitions with respect to the join column but not the right table. So this will be the equivalent of nested loop join, where only the left table is being partitioned but not the right table. So effectively, each left partition will either be L1, L2, L3, L4 or L5. Right table is not partitioned. So, you run the join between each partition from the left table and entire right table R.

**Alternative 2:** If you want to co-partition both the left and the right tables, make sure you use a hash function on the join column. The simplest hash function is join column % 5 which will create five partitions. This way you have five pairs of partitions <L1, R1>, <L2, R2>, <L3, R3>, <L4, R4> and <L5, R5>. You will run the joins independently on each pair of partitions.

Either of these alternative approaches should yield the same result, although I assume that w.r.t. efficiency, alternative 2 should be faster. You will get full points for any of these approaches.

I will discuss more about this in the coming week's lecture as well. Meanwhile if you have any questions, you can either post on Canvas or visit my office hours next Tuesday from 4:30 pm to 6:30 pm.

This announcement is closed for comments

Unread

