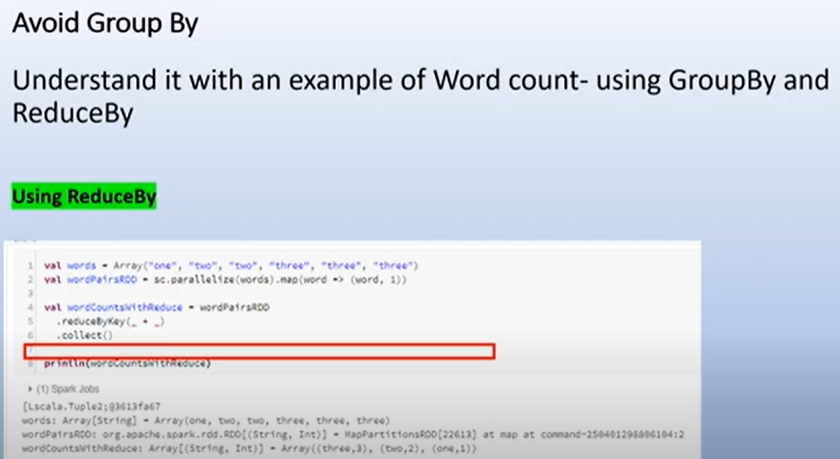
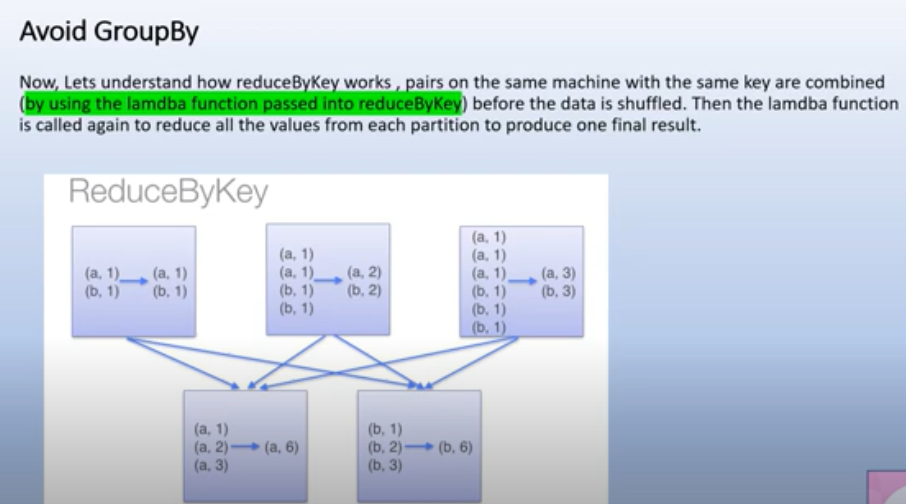
SPARK PERFORMANCE TUNING(AVOID GROUPBY )





In reducebykey operation, the reduce operation performs at partition level itself. Now shuffle is done , which is less due to reduce operation occur at partition level



But if you see in group by operation key a is shuffled across all the partition, as all keys are shuffled around all partitions. Reduce is not happen before the partition in group by

