



Lecture Takeaways 7

1. Which partitioning scheme would you use?

- Block Partition
- Range Partition on the following column(s): _____
- Hash Partition on the following column(s): _____

2. Is shuffling necessary when aggregating over local data already ...

... hash-partitioned?

- Yes, necessary
- No, no need to shuffle

... range-partitioned on A?

- Yes, necessary
- No, no need to shuffle

3. When would an RDMS choose a broadcast join over a partitioned hash join?



4. How would you implement argmax in MapReduce?

Implement your map() function

Implement your reduce() function



5. How would you implement an inner join in MapReduce?

Implement your map() function

Implement your reduce() function