Q1. It allows a user to update the metastore to add partitions to a table. Ex: msck repair table RoseDynamicEmployeesManualAdd; will update the metastore of partitions added in previous commands.
Q2. Order by guarantees sorted order of the entire data set while sort by sorts per reducer so if there are more than one reducer it could be a partially ordered answer. Ex: sort by could return 3,4,1,2 if 3 and 4 are in one reducer and 1 and 2 are in another reducer while order by will always return 1,2,3,4
Q3. Distribute by determines which rows go to which reducer based on a column. If you implement distribute by correctly then do sort by, you will get the entire collection sorted correctly. The equivalent of distribute by and sort by is cluster by.
Q4. A) Bucketing is when you specify the number of groups the data can be put in. Bucketing is good for when you have values in a column that can have many possibilities and generally follow a uniform distribution. Ex: Make 4 buckets based on employee_id. If the company has a million employees 250,000 employees will be in each bucket.

B) Union all will join two tables together into one table and not remove duplicates. Ex: you could do a union all of CS students and SE students. Even with quite a few students doing a double major (and therefore be listed in both) you can do a union all on the data