DISTRIBUTE BY with SORT BY

- Check the number of reducers in use -
    - **set mapreduce.job.reduces;**
- Sort data using SORT BY clause -
    - **select info.gender, job.salary from emp_50 sort by salary;**
    - The output isn't ordered.
- Using DISTRIBUTE BY to ensure the records with same key go to the same reducer -
    - **select info.gender, job.salary from emp_50 t1 distribute by gender sort by salary;**
    - Now the output is in order. This is because all the records were distributed by the gender. So, all the records having gender as male, went to the same reducers. And only then the records were sorted based on their salary.
- *But one thing to keep in mind is that, even DISTRIBUTE BY and SORT BY together may not guarantee global ordering. As you can see, even though the same gender records are together because of the equal distribution, the global ordering is not there.*