











Wide transformation:

Data belonging to one key will be shuffled to the same executor. So, every transformation of this league provides an overloaded function, which allows the user to provide the partitioner of his/her choice.

Apache Spark uses **Hash Partitioner** by default.

As per HashPartitioner, the resulted partition for a given key is decided by the following formula:

{hashCode of the key}%{number of partitions}

As Spark cannot decide the number of partitions by itself, therefore it keeps: *number of partitions =*

*number of partitions of largest upstream RDD,* where the largest upstream RDD is the RDD with most number of partitions in the upstream DAG.To avoid *Out Of Memory* errors.