



**Spark vs. MapReduce**

1. 1. Spark does not need a file server, whereas MapReduce can store files in a Hadoop distributed file system.
2. 2. Spark outperforms MapReduce by up to 100 times when it comes to running operations.
3. 3. Using MapReduce MapReduce writes the remaining data to disc for each Map (here, the input data is processed and stored in an HDFS after which the mapper method produces small chunks of data) and Reduce (here, the input data from the map stage is processed and produced a new set of output for storage in the HDFS) procedure. whereas The majority of the data is loaded into memory after each Spark shift.
4. If the memory in Spark runs out, it will overflow onto the disc.
5. MapReduce writes the majority of the data to disc after each map and reduces operation.
6. Spark retains the bulk of the data in memory after each transformation.

