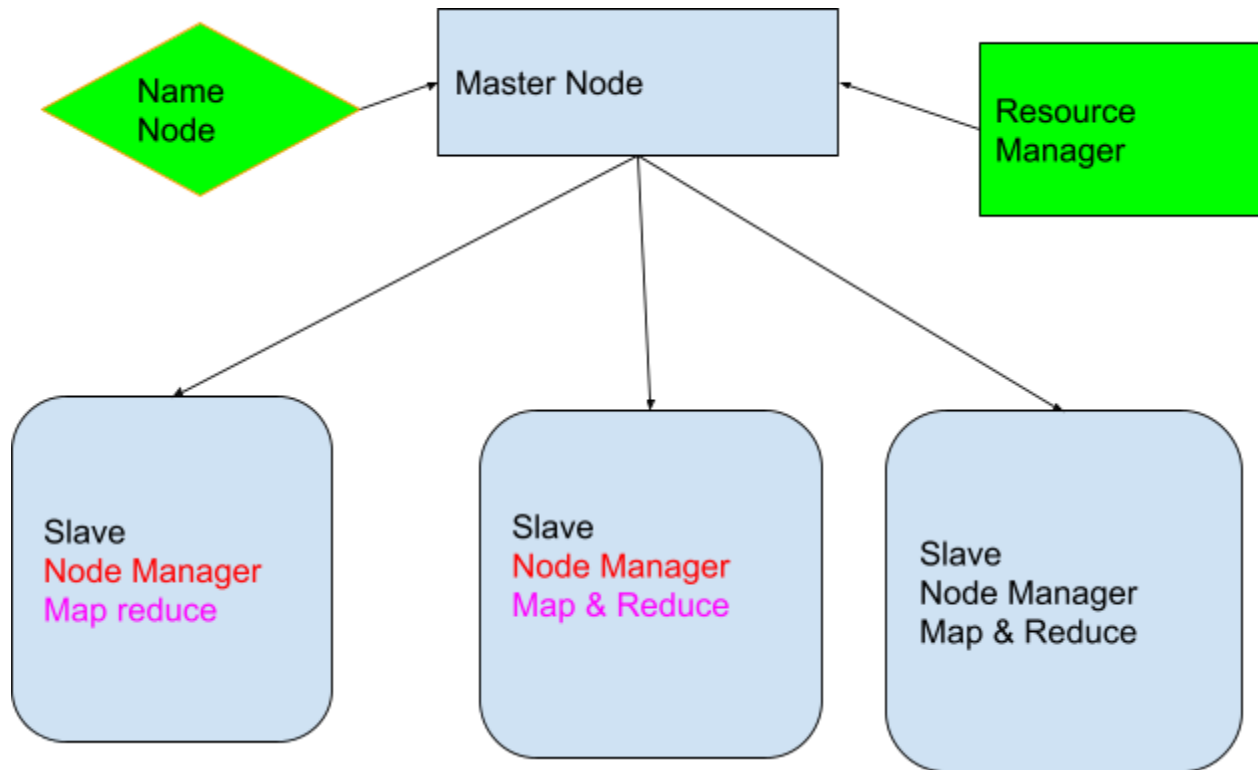1)

2) HDFS is a distributed file system that handles large sets of data and runs on commodity hardware. HDFS has fast recovery from hardware failure. It accommodates large sets of data.

3) A block in HDFS is a small piece of a huge file. The default block size is 128MB for hadoop 2.

4) Json , Arvo, and PARQUET are different file formats.