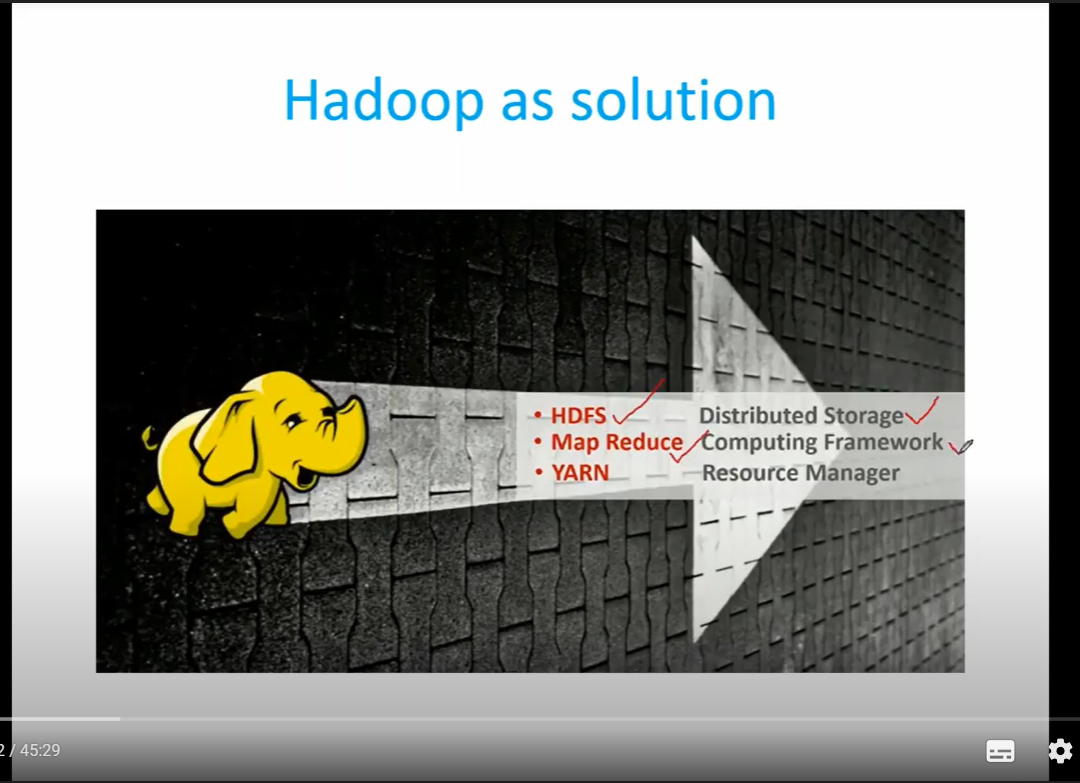
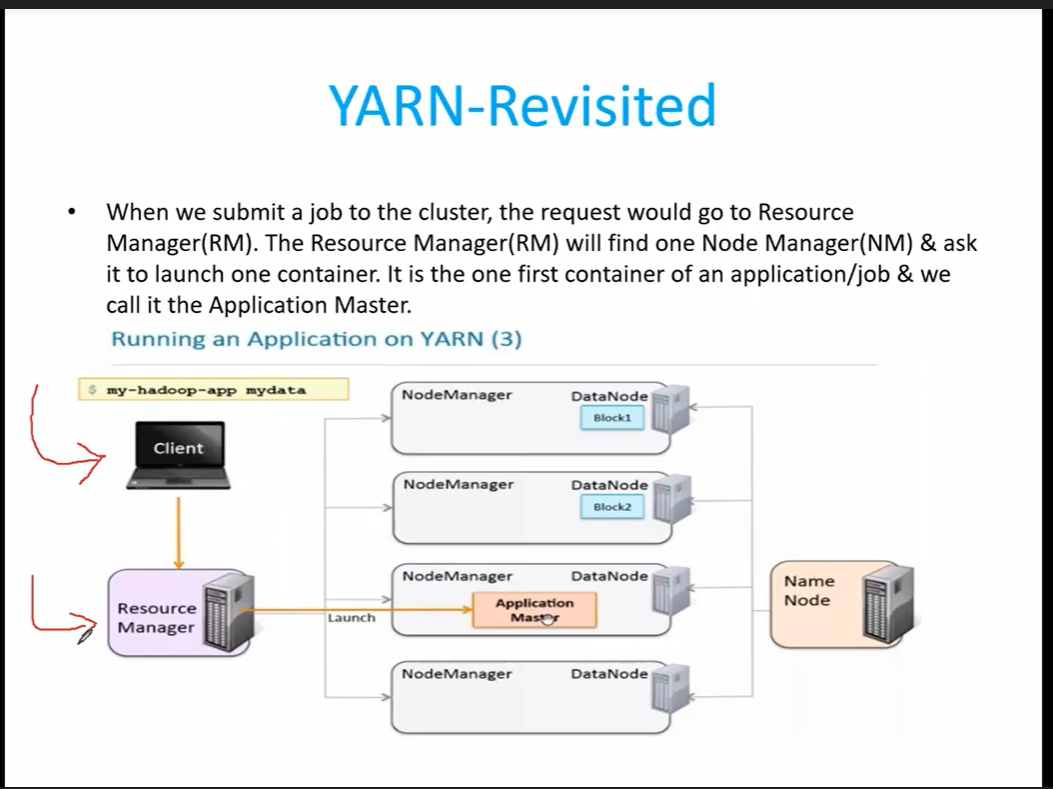
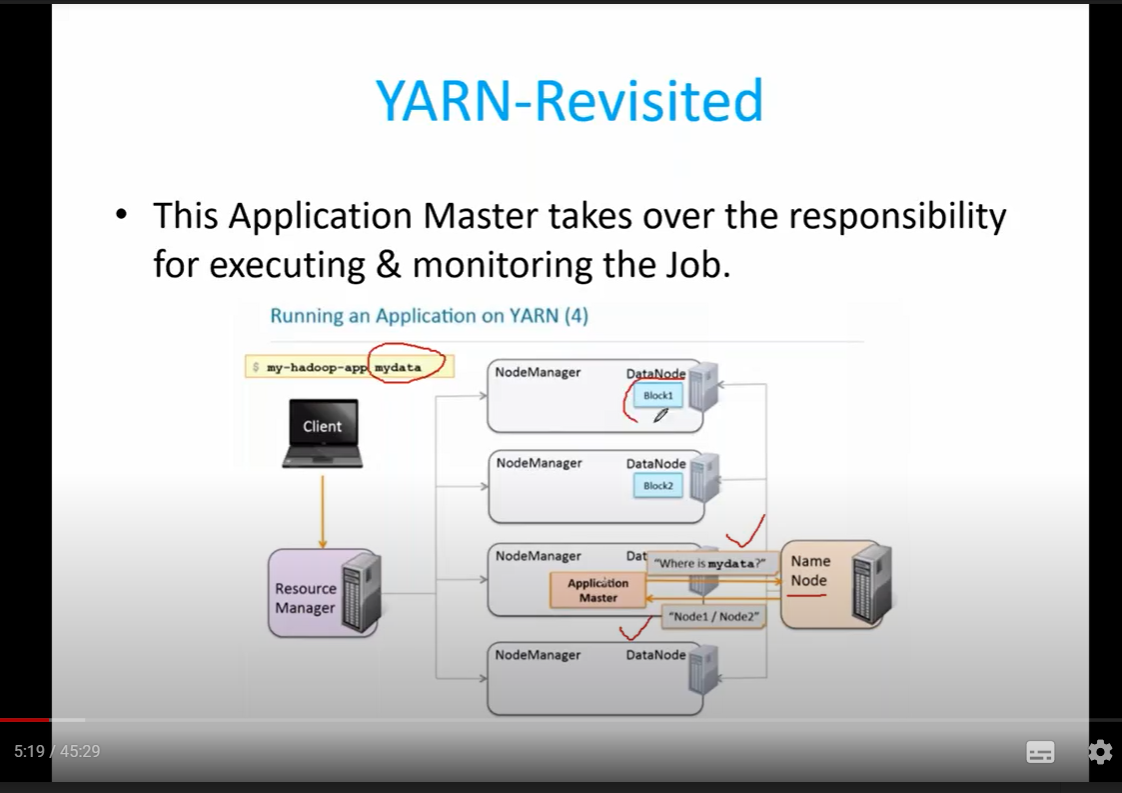
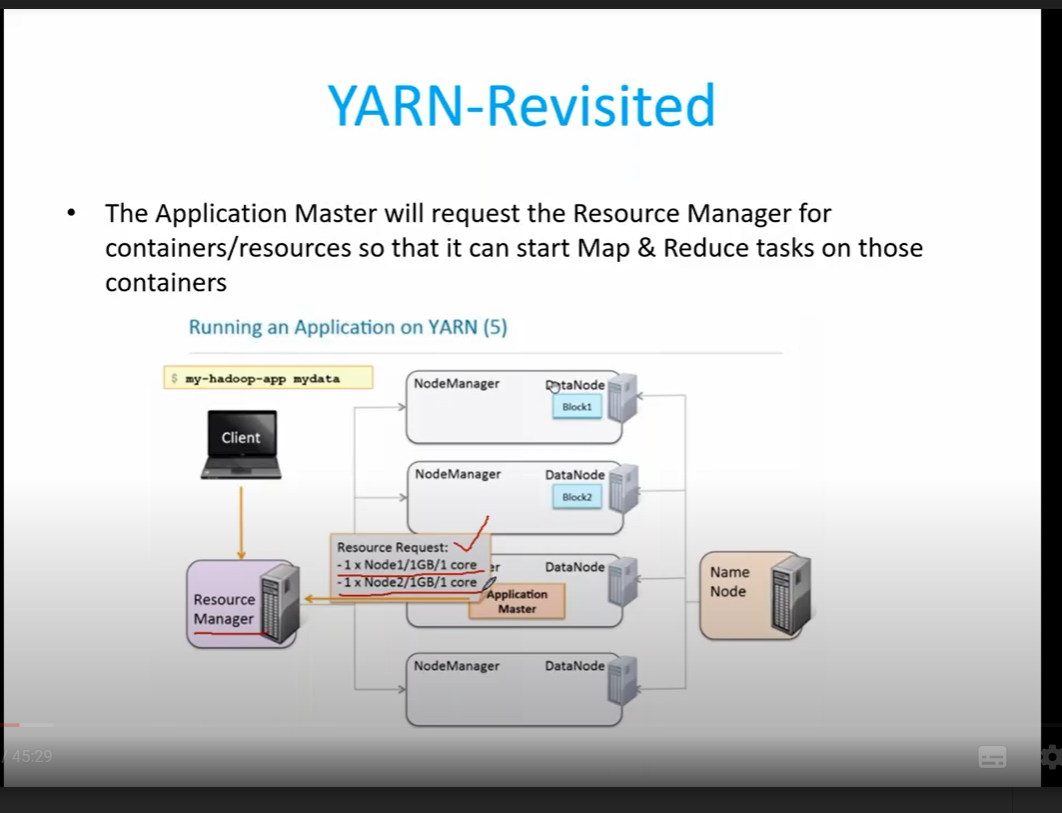
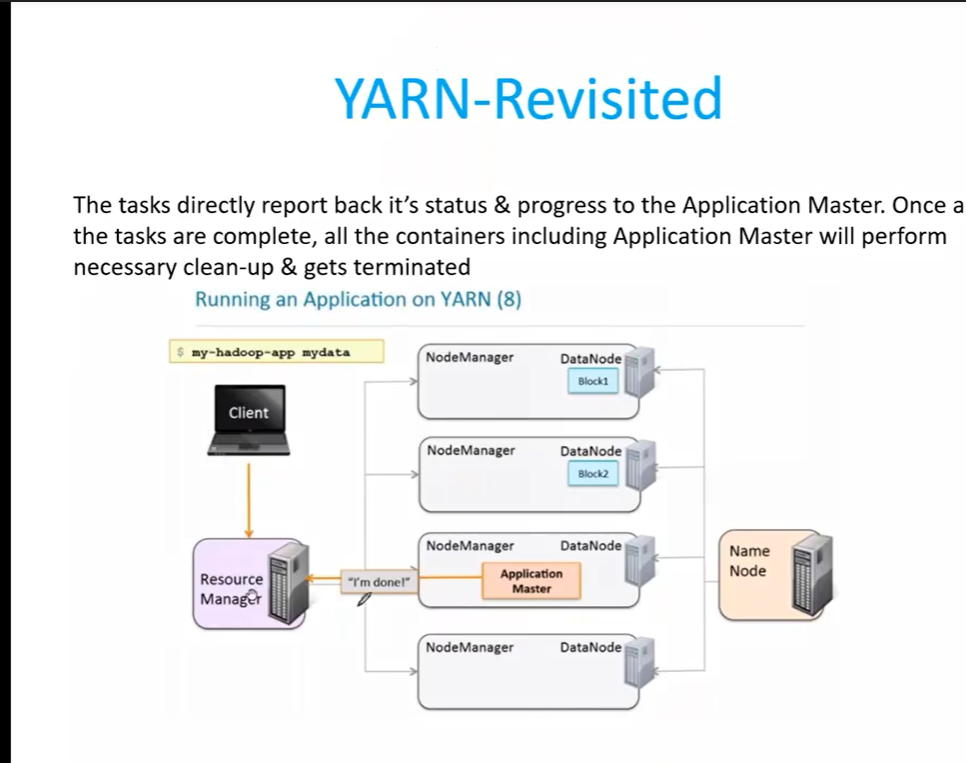
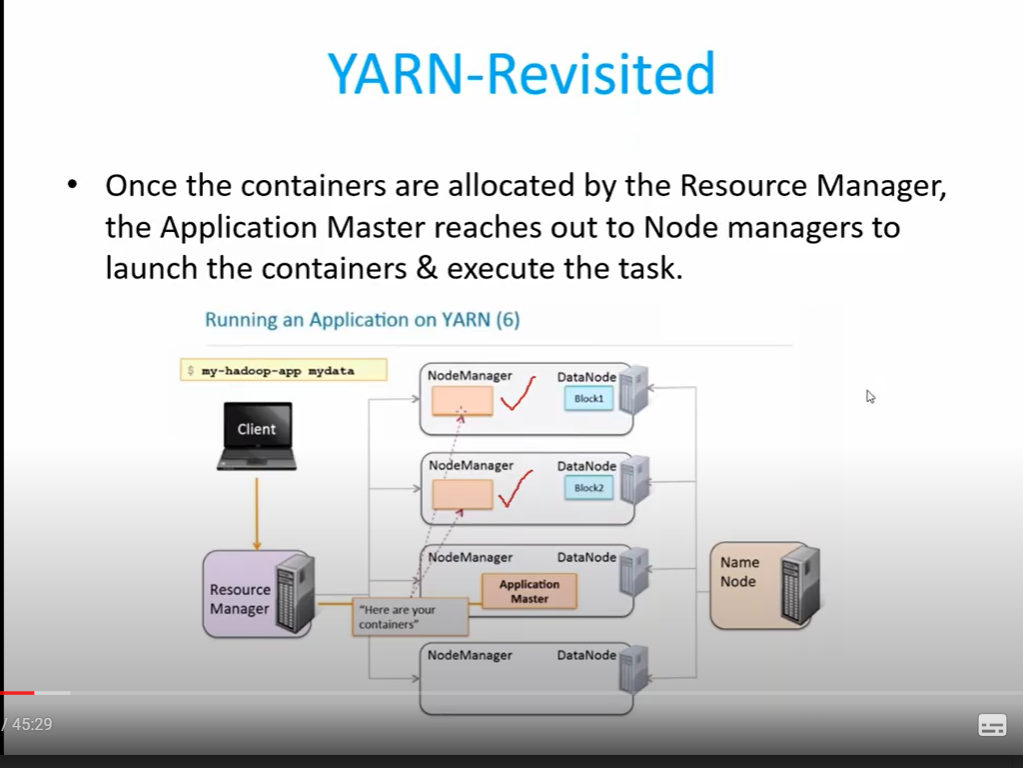
Todays problem is data is so huge that we cannot save in one system. So we came up with a sloution as haoop.

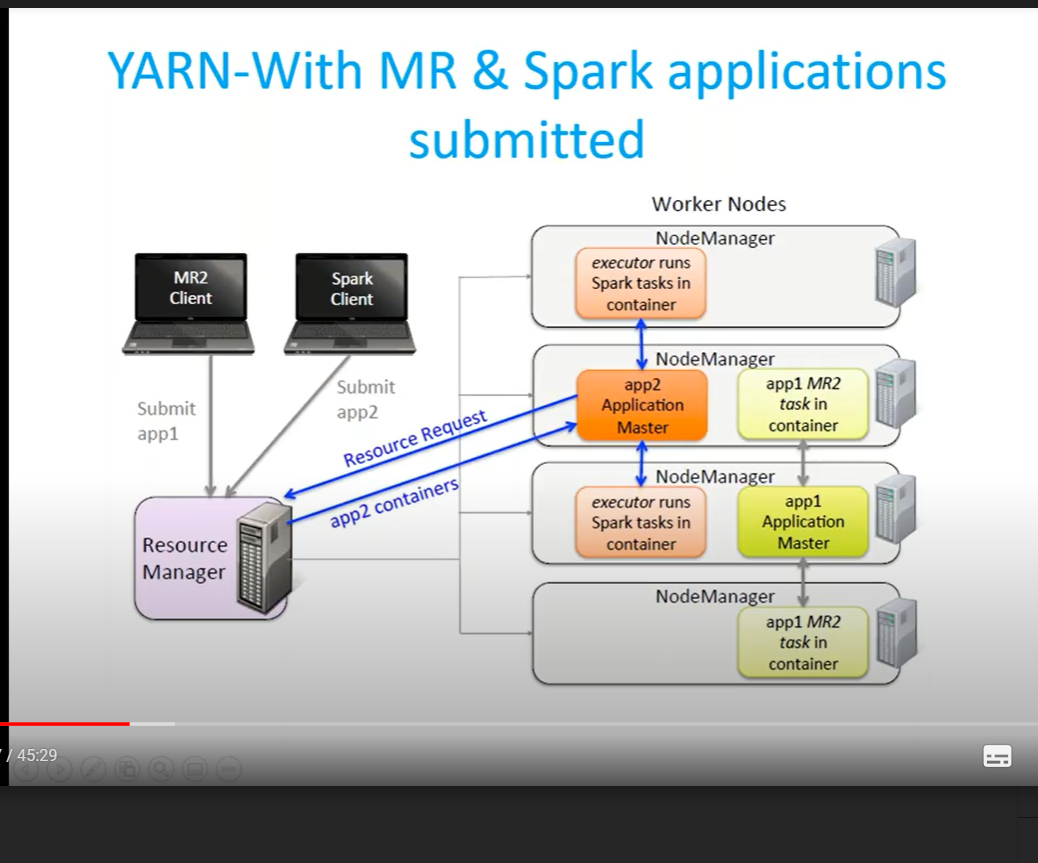


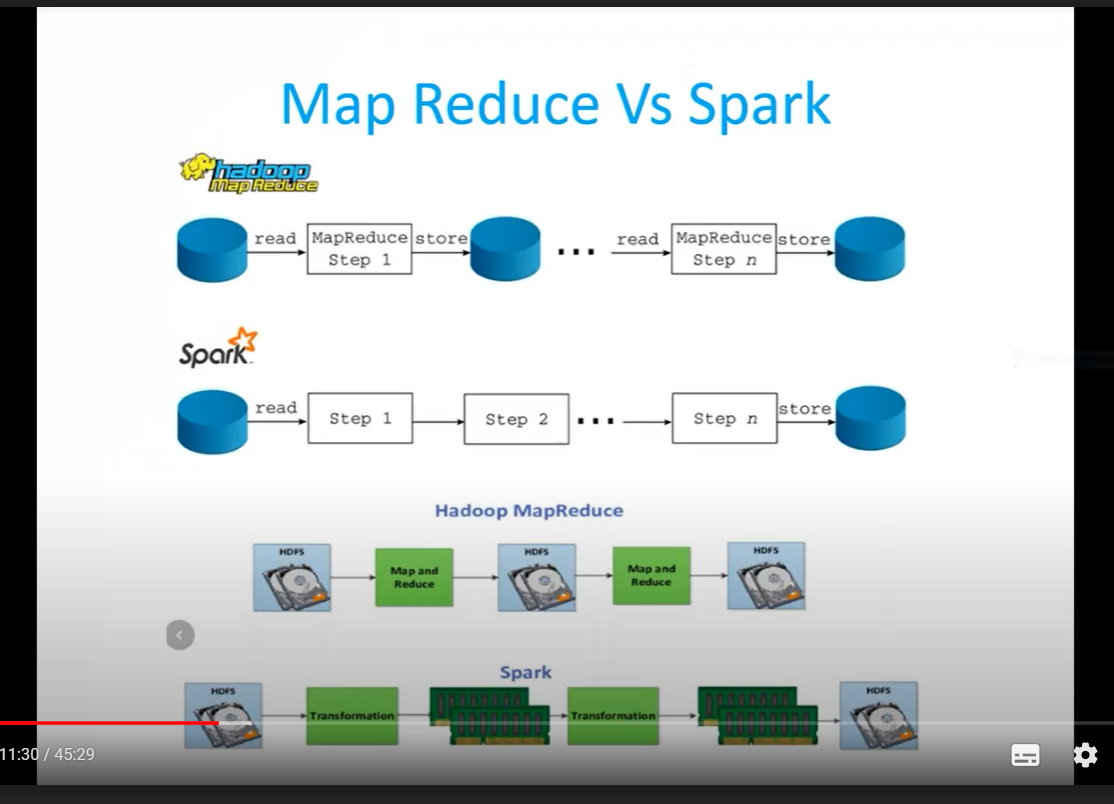




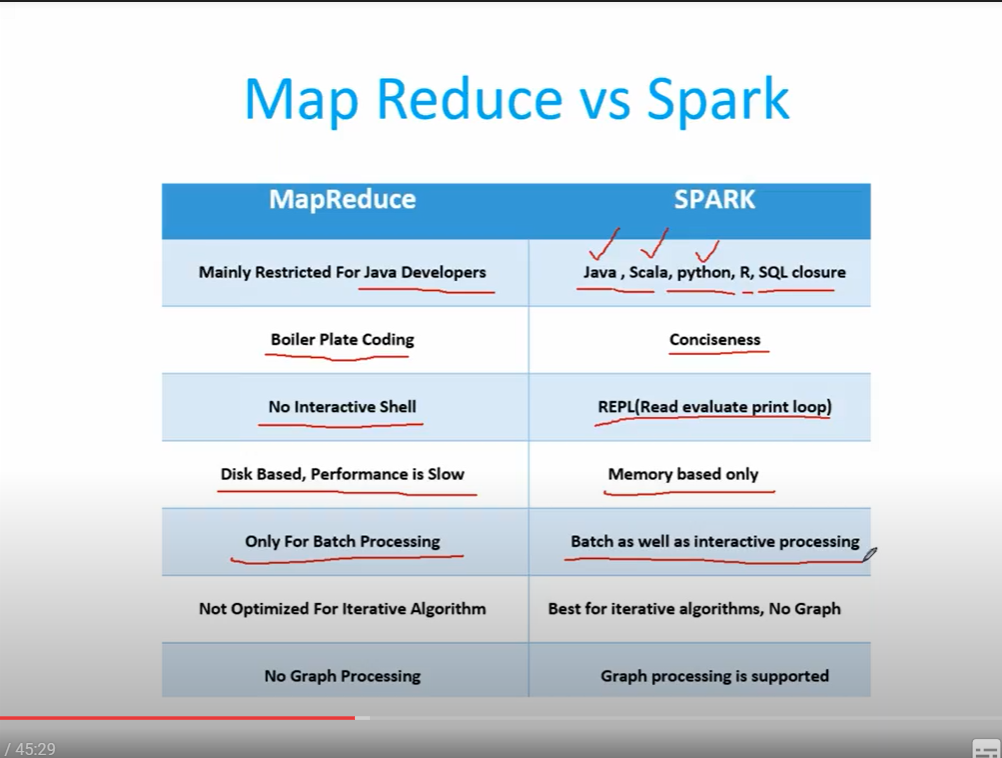






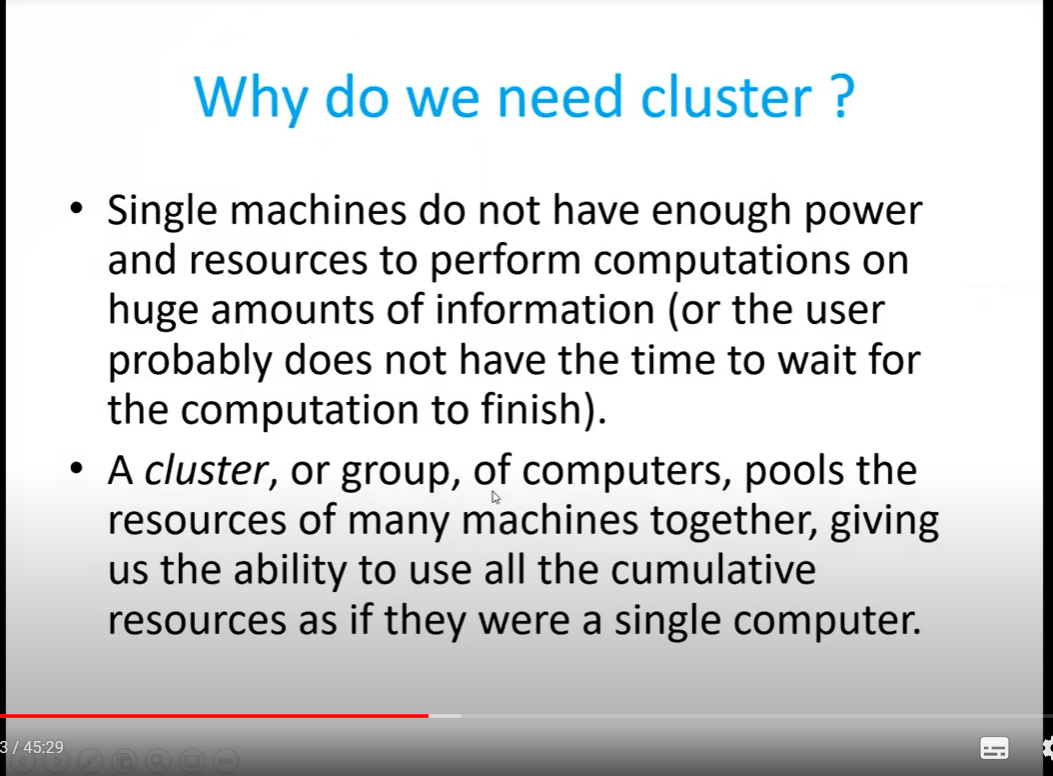


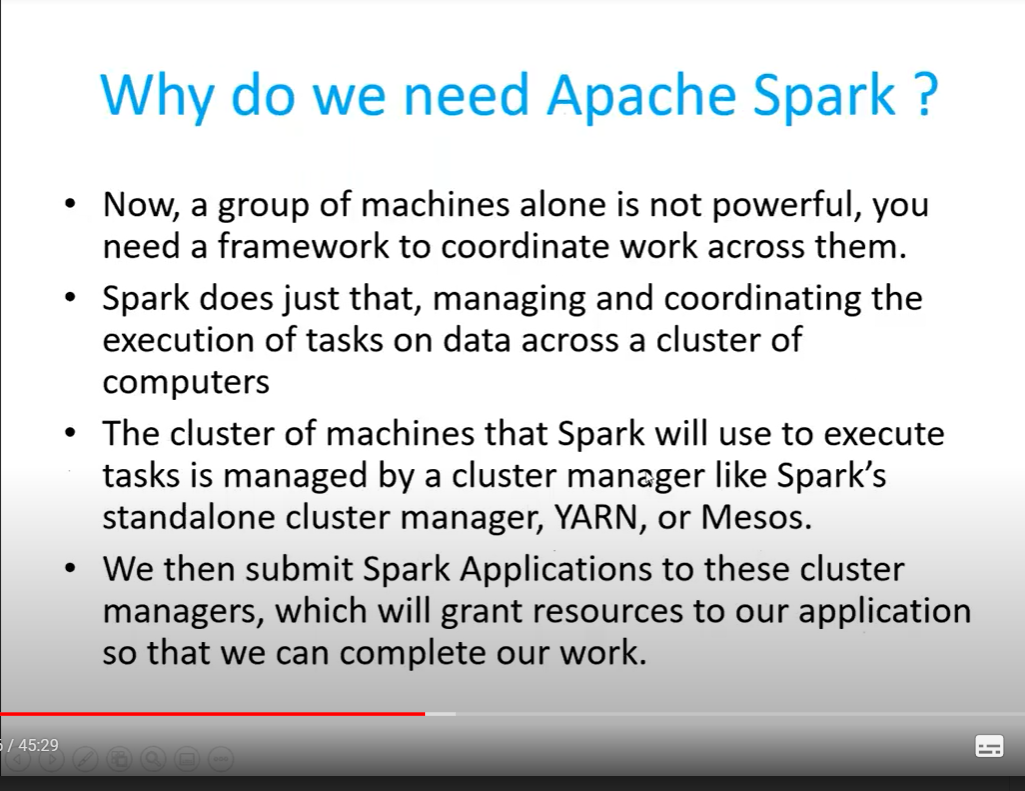
IN hadoop map reduce we read from hdfs and do map reduce and store in hdfs. But in spark we read from hdfs and transform the data and store in ram not in hdfs. So spark is fasteer than hadoop. That is why we can spark as inmemory.

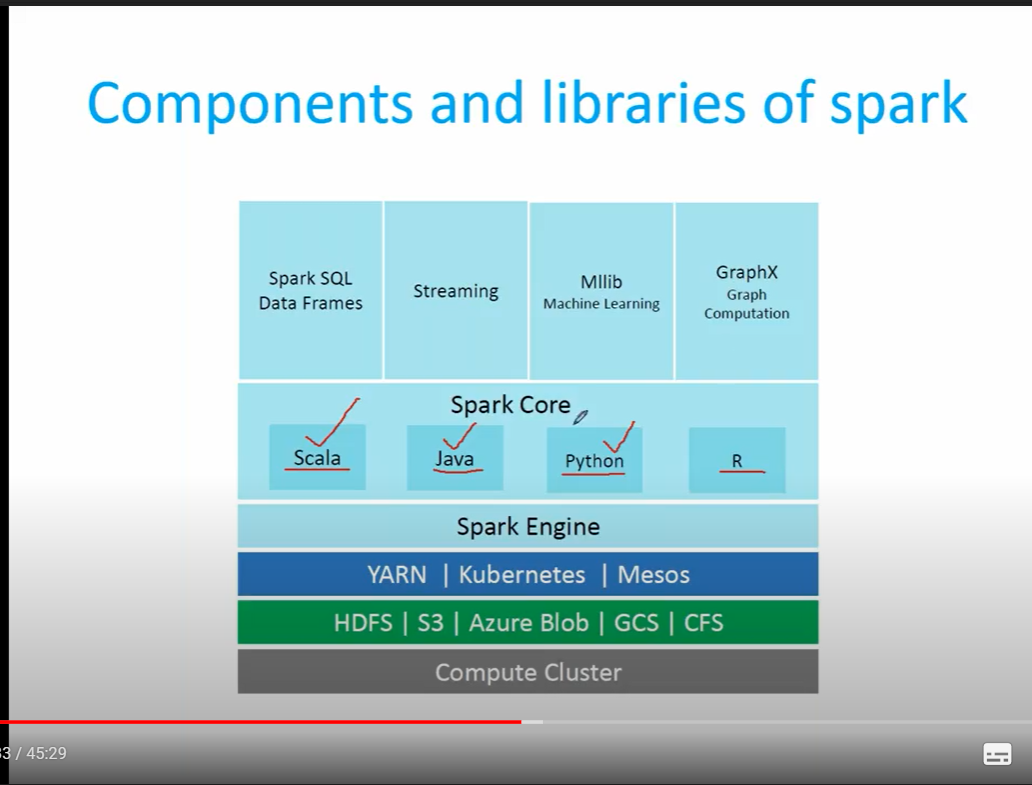


What is spark?

So as we can not store bigdata in a single machine. We came up with a solution as storing data in a mutiple systems. So to coordinate these systems we need a framework called Spark.

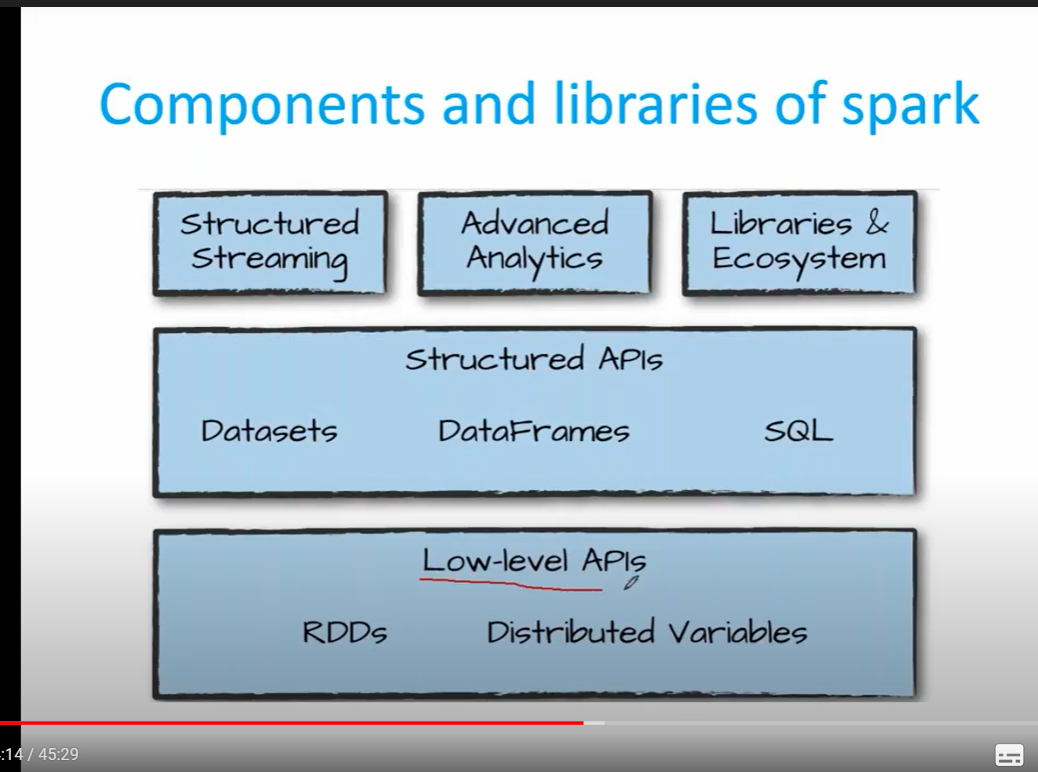


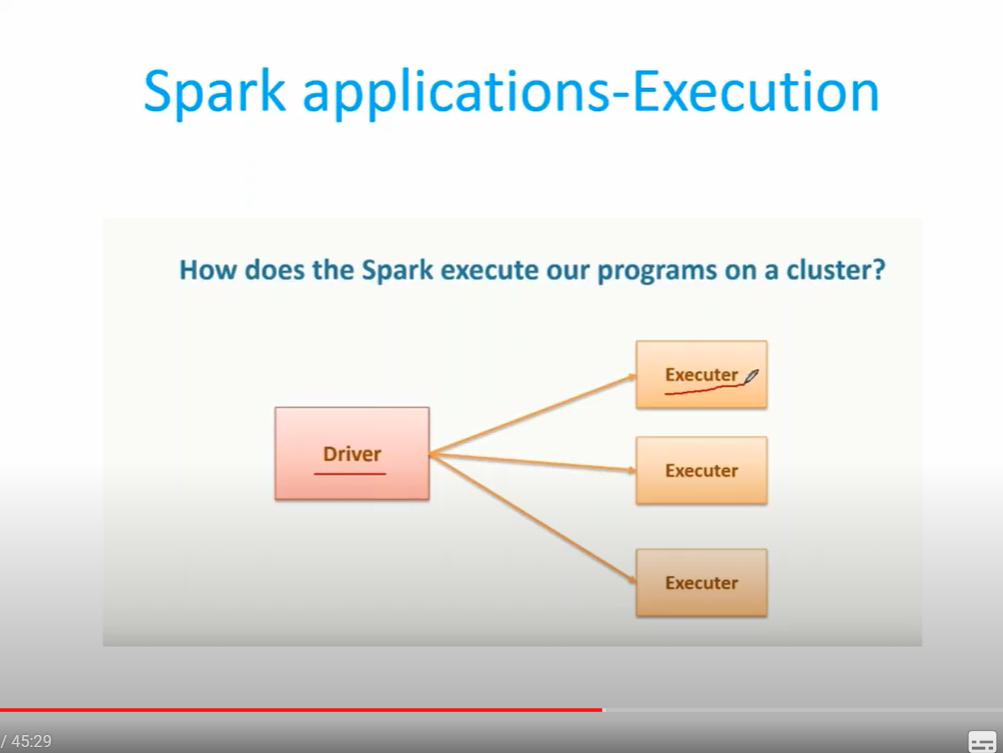




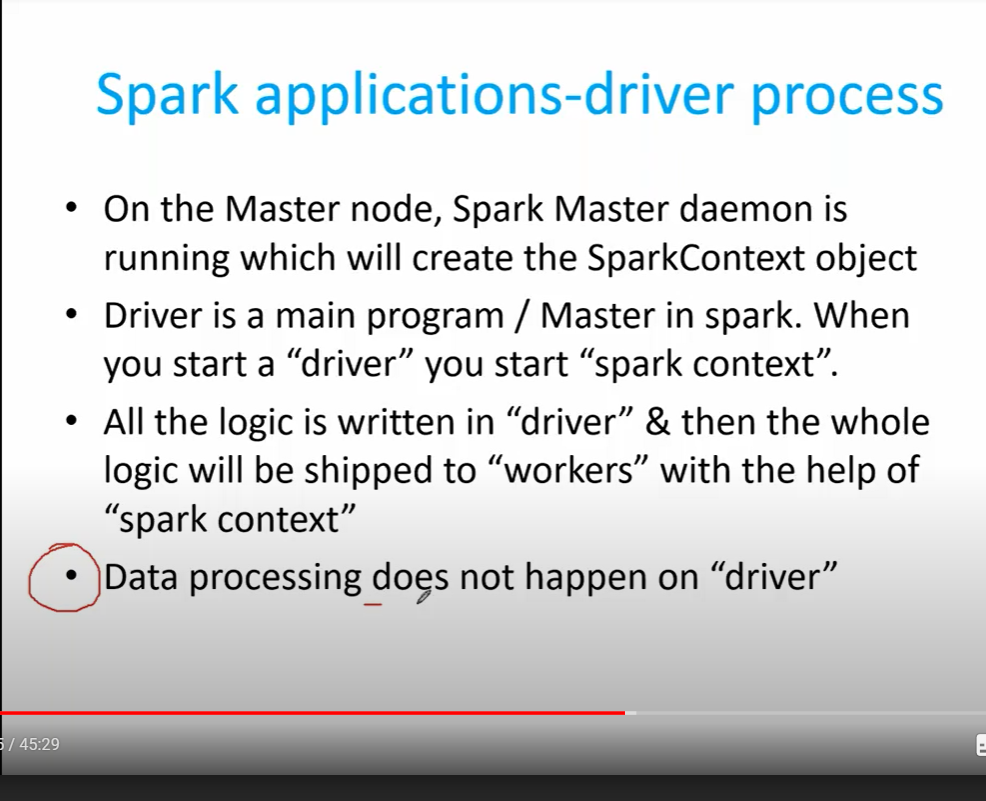
Most of companies are using HDFS and S3 for storage and Yarn for resource manager. Kubernetes and mesos is still in research phase.

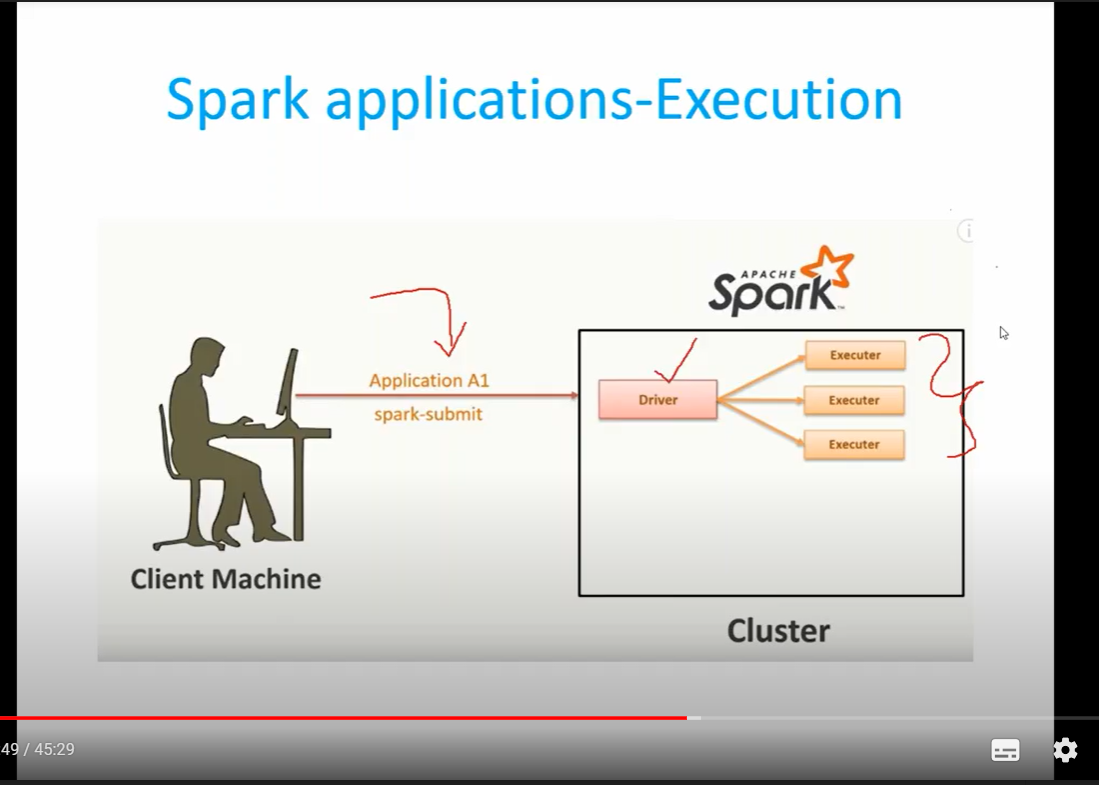
Datastrucutres availabe in spark.



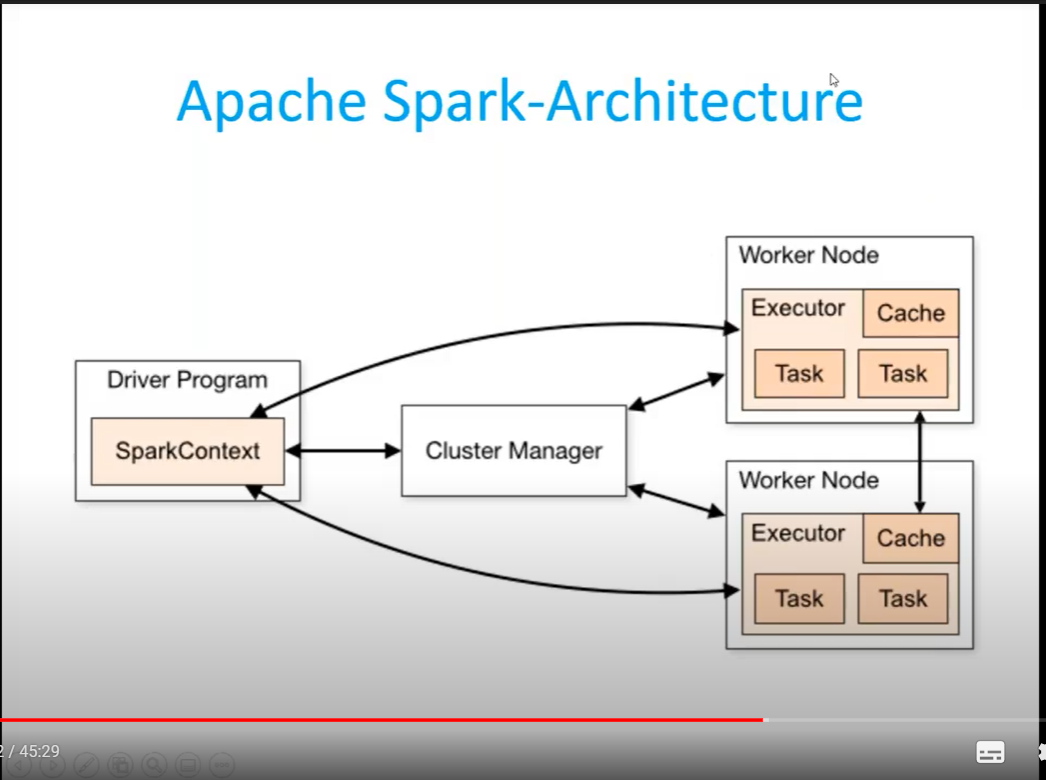


Similarly as in hadoop, in spark we have Drive a master and executer as slave.





We submit spart jobs using the command spark-submit. Once we submit Drive and executers will take of the job.



Cluster manager will act as bridge between drive programm and cluster manager. The code we write is send to cluster manager and this programm will get executed by the tasks in the executers.

