**#Assignment\_8.2**

**Q1-Explain the core changes made in Hadoop 2.x.**

ANSWER:

1. The numbers of nodes in earlier version were 4000 are now changed to 10000.
2. In case of failure in hadoop1.x admin had to explicitly come and resolve the issue while hadoop2.x there are secondary name node and passive node present which takes the control known as high availability.
3. It supports processing models like MPI message passing interface along with map reduce.
4. It allows integration of other applications with HDFS.
5. Better cluster utilization is provided as the resources are dynamic and fine grained.
6. It supports multi-tenancy as more clusters can simultaneously work in a framework.
7. Block size was scaled up from 64mb to 128mb.

**Q2- Explain the difference between MapReduce 1 and MapReduce 2 / Yarn**

ANSWER:

|  |  |
| --- | --- |
| MapReduce 1 | MapReduce 2 / Yarn |
| Need of job tracker is to track of resource utilization and monitoring | Resource Manager take care of utilization and Node Manager, whereas job monitoring is taken care by Application Master |
| Block size where data is saved in chunks is max **64MB** | The block size in yarn is **128MB** |
| **Low cluster utilization-**  Task Tracker configuration is static slots, while a map tasks cannot run on reduce slot. | **Better cluster utilization-**  Resources are fine grained and dynamic. |
| Maximum 4000 nodes and 40000 tasks can be performed in MapReduce 1 | It count is scaled up to 10000 nodes and 100000 tasks in yarn. |
| It is not possible to run other framework than MapReduce | Multi frame clustering is possible in yarn. |
| It uses concept of slots ,it can either map or reduce. | It uses concept of containers, to run generic tasks |