1. Difference between Hadoop 1 and Hadoop2 ?

Ans Hadoop 1

In this we have map reduce.

In this we have Namenode, Datanode , Secondary Namenode,Job Tracker , Task Tracker

In this map reduce is used for resource management and data processing.

Performance is not very good.

Hadoop 2

In this we have YARN.

In this we have Namenode, Datanode , Secondary Namenode, Resource Manager, Node

Manager

In this YARN is used for resource management and Map reduce is also used.

Performance is very good.

1. Why block size is 128 MB in Hadoop ?

In Hadoop block size is 128 MB because if we take lesser size then we need to create many blocks to store large amount of data and due to this huge size of meta data to be stored in Name node, it reduces the processing time.

1. 10 PB -> Data node

Meta data has each object 200 bytes.

How much RAM or Memory needed for name node in a cluster to manage the meta data ?

10 PB/(128MB \*3) \*200 B =10\*10^15/ (128 \*10^6 \*3) \* 200B = 5.2 GB

1. If Name node and Data node fails?

We recover Data node first.

1. Name node is given relay on memory rather than Data node why ?

Name Node only store metadata which is related to the different blocks and because of this reason it needs high memory space. Data Nodes don’t need large memory space