5 known facts in the world

1. Client is more important than internal team
2. Every laptop has RAM and Hard disk
3. Ram is faster than hard disk
4. If you save large file from RAM to HDD, the device might go to hanged state
5. If you don’t save the data in RAM and device gets restarted, the data in RAM will be lost

Hadoop 2.0

The important components of Hadoop 2.0 are

1. Active Name node
2. Standby Name node
3. Data nodes
4. Journal node
5. Zookeeper

Active Name node

1. The Name node contains a file called in\_use.lock which indicates that it is Active Name node

Standby Name node

1. It does not contain in\_use.lock file as it is not active
2. Every one hour, read the edit logs from journal node for merging
3. Wait for my instructions

Data nodes

1. You will be connected with both Active name node and Standby name node
2. The connection with Active name node is a complete connection.
3. The connection with the Standby name node is a partial connection waiting to get connected whenever the Active name node crashes.
4. But the transaction information is sent only to the name node which contains the in\_use.lock file.

Zookeeper

1. Zookeeper maintains the Hadoop cluster.
2. The name is given because Doug Cutting thought of the Hadoop cluster as Zoo and the nodes as animals

Journal node

1. Journal node create a temporary file called editlog.sh and helps doing checkpointing

Checkpointing

1. Checkpointing is the process of merging the FS image and edit log.
2. It is generally done by Standby name node but if the cluster goes into safe mode, Active name node will do the checkpointing itself.