Responsible for HA in hadoop/distributed systems

Elects name node when active namenode goes down

Zookeeper runs on more then one node

Leader and follower type cluster

Journal node has all metadata and namenode will take data from here

HA means more then one namenode and not secondary namenode

Datanode and task tracker/node manager are always on a single node and cannot be separated

Namenode and job tracker/ resource manager are on one node but they can be separated

Namenode and job tracker together constitute master node

Minimum 5 nodes ideally : 1 for master, one for secondary and 3 for datanodes, this is without HA

With HA we need 1 namenode, atleast one passive namenode, 3 zookeeper nodes, 3 journal nodes and 3 data nodes.

Secondary namenode is not required in HA since passive namenode does that job

Yarn: