Session5 Assignment 2- answers

(Job Execution part )

## Name node

1. It’s the Master in the HDFS cluster which maintains the file system namespace – filesystem tree and metadata.
2. NameNode is responsible for making block replicas.
3. NameNode executes filesystem namespace operations like opening, closing and renaming files and directories.
4. It also determines the mapping of blocks to datanodes

## Data node

1. They are the workhorses of HDFS, they store and retrieve data blocks upon instruction from NameNode - serving the read and write requests.
2. Usually there are one datanode for each of the nodes in a cluster and perform creation, deletion and replication of blocks upon instruction from namenode.

## Resource Manager

1. A long-running daemon of YARN which manages the use of resources across the cluster
2. There is only one resource manager per cluster
3. Upon the job submission from client, co-ordinates the allocation of compute resources on the cluster. It generates AppId or Application ID as an acknowledgement.
4. It has 2 main component s- Scheduler and AppicationsManager.

## Node manager

1. Runs on all the nodes in a cluster to launch and monitor containers (like CPU, usage, memory,disk & network)
2. All container processes running on a slave node are initially provisioned, monitored and tracked by that slave node’s Node manager.