**Session 5: MapReduce**

**Assignment 2**

Q 1. Explain the importance of below 4 demons in job execution with minimum of 5 points

1. Name node:

2. Data node:

3. Resource Manager:

4. Node manager:

**Solution:**

**1. Name node:**

a) It is master node of HDFS file system .

b) It keeps the directory tree of all files in the file system, and tracks where across the cluster the file data is kept. That is it stores only metadata.

c) Does not store of data of these files.

d) Client applications talk to the NameNode whenever they wish to locate a file, or when they want to add/copy/move/delete a file.\

e) The NameNode is a [Single Point of Failure](https://wiki.apache.org/hadoop/Single%20Point%20of%20Failure) for the HDFS Cluster.

**2. Data node:**

a) DataNode is responsible for storing the actual data in HDFS.

b)DataNode is also known as the Slave

c) Data node has more than one replication , so that one node gows down than there is no lost to data .

d) DataNode is usually configured with a lot of hard disk space. Because the actual data is stored in the DataNode.

e) Data node sends heartbeat to namenode ,to show that it’s alive.

**3. Resource Manager:**

a) It is master node.

b) Tracks the resource in system.

c) Scheduling the resource.

d) It works together with the **NodeManagers (NMs)** and the per-application **ApplicationMasters (AMs)**.

**4. Node Manager:**

a) It is slave node.

b) There is one node manager per node.

c) Task execution is performed by node manager. There is no slots so node manager is efficient as it can take any type of task for execution.

d) Node manager send ping(heartbeat) to resource manager so that it came to know how many resources are available