Session 5

# Assignment 2

**NameNode:**

1. NameNode is the centerpiece of HDFS and it is also known as Master.
2. NameNode only stores the metadata of HDFS – the directory tree of all files in the file system, and tracks the files across the cluster.
3. NameNode does not store the actual data or the dataset.
4. NameNode is so critical to HDFS and when the NameNode is down, HDFS/Hadoop cluster is inaccessible and considered down.
5. NameNode is a single point of failure in Hadoop cluster.

**DataNode:**

1. DataNode is responsible for storing the actual data in HDFS.
2. DataNode is also known as the Slave.
3. When a DataNode starts up it announce itself to the NameNode along with the list of blocks it is responsible for.
4. When a DataNode is down, it does not affect the availability of data or the cluster. NameNode will arrange for replication for the blocks managed by the DataNode that is not available.
5. DataNode is usually configured with a lot of hard disk space. Because the actual data is stored in the DataNode.

**Resource Manager:**

1. This daemon process runs on master node (may run on the same machine as name node for smaller clusters)
2. It is major component of YARN.
3. It is responsible for getting job submitted from client and schedule it on cluster, monitoring running jobs on cluster and allocating proper resources on the slave node.
4. It communicates with Node Manager daemon process on the slave node to track the resource utilization.
5. It uses two other processes named Application Manager and Scheduler for MapReduce task and resource management.

**Node Manager:**

1. This daemon process runs on slave nodes (normally on HDFS Data node machines).
2. It is major component of YARN.
3. It is responsible for coordinating with Resource Manager for task scheduling and tracking the resource utilization on the slave node.
4. It also reports the resource utilization back to the Resource Manager.
5. It uses other daemon process like Application Master and Container for MapReduce task scheduling and execution on the slave node.