Session 5: MapReduce

Assignment 2

# Problem Statement

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

* **Name node**

**Ans:**

The importance of NameNode in Job execution:

1. The Name node receives the request from Client to locate a file or to add, copy, move or delete a file. As soon as the request is received the Namenode starts the Jobtracker or Resource Manager.
2. Namenode periodically gets a heartbeat signal from Datanode indicating that the Datanode is fine and working properly.
3. When the Namenode does not receive a signal from a particular Datanode after a certain amount of time the Namenode send the task to another Datanode and make sure that all the tasks gets completed.
4. The NameNode contains the list of all the metadata about the data that is used at the time of client request completion.
5. Namenode keep all the track of all the Jobtracker send to each and every Datanode and after getting the response of request from Datanode gives the output to the Client.

* **Data node**

**Ans :**

The importance of Datanode in Job execution:

1. Each Datanode has one Tasktracker that performs the task assigned by the JobTracker from Namenode.
2. Sends Heartbeat Signal to the Namenode so that the NameNode can come to know that the datanode is alive and working properly.
3. A DataNode stores data in the Hadoop File System HDFS. There is only One DataNode process run on any hadoop slave node. DataNode runs on its own JVM process. On startup, a DataNode connects to the NameNode. DataNode instances can talk to each other, this is mostly during replicating data.

* **Resource Manager**

**Ans:**

1. Resource Manager is the master in Hadoop V.2 architecture. It arbitrates all the available resources among all the applications running on the system.
2. Listen to the demands from Application Manager and tries to fufill it if possible.
3. It works together with the per-node NodeManagers (NMs) and the per-application ApplicationMasters (AMs).
4. Listen to the Client Services such as application submission, application termination, obtaining queue information etc.
5. It is responsible for registration of new nodes, rejecting requests from any invalid/decommissioned nodes, obtain node-heartbeats and forward them over to the YarnScheduler.

* **Node manager**

**Ans :**

1. Takes instructions from the resource manager and manages resources available in a single node.
2. Node manager works with the application manager and starts the containers.
3. Node Manager registers with the Resource Manager and send the information to resource manager about the resources available in a single node.
4. Node manager and Resource Manager communicates with each other to provide the updates on containers status such as new containers running on the node, completed containers etc.