**NAME: Ashish Bisht**

**STUDENT ID: 801168390**

**Cloud Computing for Data Analysis**

**VIDEO CASE 02 : Hadoop HDFS**

Watch following videos:

**Video 1**: https://youtu.be/cC6sS80sZYQ

**Video 2:** https://youtu.be/yDV0EE9DCJ0

**Video 3:** https://youtu.be/nbOagGnIMiY

Video 1 introduces you to HDFS

Video 2 describes about Name Node in HDFS

Video 3 provides knowledge about file read and write operations in HDFS

**Video Case Questions:**

**Give some description about data nodes and name node**?

Data Nodes:

• Nodes which store data are called data nodes.

• They work as slave nodes or worker nodes.

• Data node does the actual groundwork for storing the data inside and are stored in racks present in clusters.

Name Node:

• Name nodes are the master nodes.

• They are responsible for the management of files that are distributed across the cluster.

• In order to store the files in the data nodes, initially the file is replicated three times and then the it is stored in the forms of blocks in the data nodes. This distribution of the data blocks is maintained by name node.

• Incase of any failure in data nodes or racks, then the name node is still able to put the file together using the replicated data blocks.

**What is the main purpose of secondary name node?**

• Secondary name node combines the edit log with the namespace images, so the name node memory doesn’t increase with increasing edit logs.

• It also creates check points of name space image and edit logs combined, on a consistent basis and writes them to a file.

• The name node then releases the main memory of the edit log till the latest checkpoint.

• Secondary name node is developed using java language.

**• What are all the steps followed by HDFS for write operation?**

Steps followed by HDFS for write operation:

• STEP 1: Write request

Initially the HDFS client sends the write request to the name node as it wants to write into name node.

• STEP 2: Various checks

The name node then performs various checks such as where the file exits or not, the client has actual permissions or not etc. to perform the activity.

• STEP 3: Lists all the nodes

If all the checks are correct, then the name node returns to the client with list of the all the nodes available in the name node (for copying).

• STEP 4: Write blocks

Here the client will connect to the first data node and asks it to connect to all the subsequent nodes using a pipe connection so the client can write the data into them.

• STEP 5: Acknowledgement

Here the data nodes acknowledge to the client as they successfully copy the block and process continues until the complete file is written.

• STEP 6: Complete

The client ends the process with a completion message that is sent to the name node.

**Explain the steps that Hadoop follows for reading the data during a data node failure.**

• Initially the HDFS client gets the block location of the data using the name node and reads the data from the nearer data block.

• If the data node is failed, then the next node in the list is picked

• Then the data is read form node that is newly picked from the list.

• Then that node is not tried for the later nodes and the activity continues.