**Cloud Computing for Data Analysis**

**VIDEO CASE 02 : Hadoop HDFS**

**Student Name: Balasundaram Avudai Nayagam**

**Student ID: 801104903**

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:**

1. Give some description about data nodes and name node
2. - The name node and data nodes follow Master- Slave relationship.

* The Name node acts as the master and is responsible for management of the storage space on the data nodes. Name node performs a function of keeping a track of the complete file system by managing two: Namespace image and Edit Logs. These two are maintained in the main memory of the Name node.
* The data nodes perform the functionality of storing the data blocks.

1. What is the main purpose of secondary name node?
2. - Merge the edit log regularly and create check points.

* Generates the merged checkpoints of the namespace image and Edit log, and saves it in a file.

1. What are all the steps followed by HDFS for write operation?
2. - Client needs to interact with master i.e. name node (master) and the name node provides the address of the data nodes on which client will start writing the data.

* The first data node will copy the block to another data node, which will copy it to other data node. Once it creates the replicas of blocks, it sends back the acknowledgment.

1. Explain the steps that Hadoop follows for reading the data during a data node failure.
2. - The HDFS client will choose the next available node in the data list and continue the reading process.

* The node that failed will not be visited again to get the data blocks.