**Cloud Computing for Data Analysis**

**VIDEO CASE 02 : Hadoop HDFS**

**Video Case Questions:**

1. Give some description about data nodes and name node

## **Answer: Data Nodes:** It is responsible for storing the actual data in HDFS. Since, the actual data is stored in the DataNode, it is usually configured with a lot of hard disk space. It is a daemon that runs on the SlaveNode in a Hadoop Cluster, hence often called as the Slave.

## **Name Node:**

A NameNode is the centerpiece of an HDFS file system. It holds the metadata for [HDFS](http://data-flair.training/blogs/comprehensive-hdfs-guide-introduction-architecture-data-read-write-tutorial/) **like** [block](http://data-flair.training/blogs/data-blocks-hdfs-hadoop-distributed-file-system/) information, size etc. It’s function is to keep the directory tree of all the files in the file system and track where the file is actually kept, cross the cluster. NameNode does not store the data of the files itself. It’s a single point of failure for HDFS systems. The file system would go completely offline if the NameNode goes down.

1. What is the main purpose of secondary name node?

**Answer:**

## Secondary NameNode is a specially dedicated node in HDFS cluster and its main function is to take checkpoints of the file system metadata present on namenode. It checkpoints namenode’s file system namespace and is not a backup namenode. It’s a helper to the primary NameNode but not not its replacement.

1. What are all the steps followed by HDFS for write operation?

**Answer:**Steps that are followed for write operation –

1. Write Request: HDFS client sends this request to the NameNode.
2. Checks: The NameNode checks for the validity of the request.
3. List of Nodes: The NameNode returns the list of nodes which are available to write.
4. Write Block: The block is written according to the list of nodes given.
5. Acknowledgement: A final acknowledgement is sent to client when the writing is successful.
6. Explain the steps that Hadoop follows for reading the data during a data node failure.  
   **Answer:**

Steps that are followed for the read operation –

1. Read request:The client sends a read request.
2. Get block location: The list of block locations is sent, arranged according to distance.
3. Read blocks: The blocks are read according to the order given
4. Failure handling
5. Next set blocks: After a block is read successfully, next set is read.