Assignment	**02**
Name	**Murtaza Jamali**
Course	**Big Data Analytics**
Instructor	**Moeed Tariq**
Submission date	**21/11/2023**

What is the difference between put and copyFromLocal?

The difference between put and copyFromLocal is that the “CopyFromLocal” command will help copy the file from local file system to HDFS, while the “Put” command will copy from anywhere (local or network) to anywhere (HDFS or local file system).

Example:

PUT:

1) hadoop fs -put /root/input1/sample1.txt /user/hduser/output3/

2) hadoop fs -put /root/input1/sample1.txt /root/input2/sample2.txt /root/input3/sample3.txt /user/hduser/output4/

copyFromLocal:

1) hadoop fs -copyFromLocal /root/input1/sample1.txt /user/hduser/output1/

2) hadoop fs-copyFromLocal /root/input1/sample1.txt /root/input1/sample2.txt /root/input1/sample3.txt /user/hduser/output2/

CopyFromLocal and Put:

The difference between these two is that the “CopyFromLocal” command will help copy the file from the local file system to HDFS, while the “Put” command will copy from anywhere (local or network) to anywhere (HDFS or local file system).

moveFromLocal:

the command is employed to transfer files or directories from the local file system to HDFS. It requires specifying a source path in the local file system and a destination path in HDFS. For instance, one might use the following syn

moveToLocal:

On the other hand, the moveToLocal command is utilized for copying files or directories from HDFS to the local file system. It necessitates specifying a source path in HDFS and a destination path in the local file system. An illustrative example is as follows:

Why 3 Name Nodes in Hadoop?

In Hadoop's three-NameNode setup for High Availability, an active NameNode, a standby NameNode, and shared JournalNodes collaborate to ensure fault tolerance. This architecture minimizes downtime by swiftly transitioning control to the standby NameNode in case of an active NameNode failure, enhancing the overall reliability of the Hadoop cluster.

Block Size should be configured with a multiple of 512kb why?

Configuring Hadoop's block size with a multiple of 512KB is recommended because it aligns well with the typical size of disk sectors, which are often 512 bytes. This alignment enhances data storage and retrieval efficiency, optimizing read and write operations by matching the block size with the disk's natural I/O unit.

COGNITIVE CLASS.ai

This is to certify that

Murtaza Jamali

successfully completed and received a passing grade in

Hadoop 101

(BD0111EN, provided by IBM)

A course on **cognitiveclass.ai**
Powered by **IBM Developer Skills Network.**

Issued by
Cognitive Class

November 19, 2023

Authenticity of this certificate can be validated by going to:
https://courses.cognitiveclass.ai/certificates/bf2af03aeb1c4f2485c9570835c3976f

Certificate ID Number: **bf2af03aeb1c4f2485c9570835c3976f** **November 19, 2023**