The Apache Hadoop Project Page **1** of **1**

**LAB: Loading Data into HDFS**

# Copying Data to the Hadoop File System (HDFS)

In this exercise, you copy the exercise data from the Linux file system to the Hadoop file system (HDFS) using HDFS command-line tools. The data is required in HDFS to support subsequent queries and analysis using HiveQL and Pig Latin. Below are the specifics of the exercise:

* + Using the Client node login, navigate to the **/home/student** folder, if it is not already open.
  + Verify that the **DIHPS** directory exists with content.
  + Verify that the HDFS **/user/student** directory exists.
  + Create the HDFS directory /**user/student/DIHPS/data**.
  + Copy the contents of **/home/student/DIHPS/data** on the Client node to the

**/user/student/DIHPS/data** HDFS location.

* + Verify the contents of **/user/student/DIHPS/data**.

**Question**: Using the HDFS command line, how many files were copied to the data subfolder under DIHPS in HDFS?

**Answer**:

**Question**: What HDFS command would you use to view the contents in the HDFS location,

# /user/student/DIHPS/data/numbers.txt?

**Answer**:

Copyright © 2016, Mohamed Rahal.