DA-IICT
IT 508, **Winter 2021-2022**
Lab Exercise 3
Date: 08/03/2022, **Expected by: 22/03/2022**

This lab focuses on two things: (1) learning basic HDFS commands, (2) installing PIG and running a few basic commands.

Lab Problems:

1. Configure a Hadoop cluster to run in pseudo-distributed mode and do the following tasks:

   (a) Run the commands (1) mkdir, (2) rm, (3) copyFromLocal, (4) copyToLocal and observe the output.

   (b) Read "Managing Files with the Hadoop File System Commands" in chapter 5 from [3].Try out any 5 commands from Table 5-2 other than those asked for in (1-a).

   You can refer to {a}-{c} for a general know-how on working with HDFS commands.

2. Refer to chapter 11 (Pig) from [2], and read up to "Generating Examples". Try out the installation and other commands given in there.

   Under the topic "An Example", with regard to working in "grunt" shell on the weather data set example, you have two options: (1) download the data files for years 1901 and 1902 from `https://github.com/tomwhite/hadoop-book/tree/master/input/ncdc/all` (directed from `http://www.hadoopbook.com/code.html`) and run all the commands for these years only, or (2) You go the website of NCDC and look out for free datasets. If available, you download one and work with it. For help in downloading the data set from NCDC website, you can look into `https://serc.carleton.edu/hydromodules/steps/downloading_ncd.html`.

   If none of the options work out, you may try out the commands on some random dataset/files of your choice. Our purpose in this exercise is to learn the basic working in Pig.

References - books for perusal:

[1] *Hadoop In Action*, Chuck Lam, Manning Publications Co.
[2] *Hadoop The Definitive Guide*, Tom White, O' Reilly.
[3] *Hadoop for Dummies*, D. deroos, P. C. Zikopoulos, R. B. Melnyk, B. Brown, R. Coss, John Wiley.

References - online for perusal:

{a} *Basics of Hadoop Distributed File System (HDFS)*, available on: `https://www.youtube.com/watch?v=ZYQhHrq5kj8`.
{b} *Top Hadoop HDFS Commands with Examples and Usage (Part - 1)*, available on: `https://www.youtube.com/watch?v=8vrUaMVG4po`.
{c} *Top Hadoop HDFS Commands with Examples and Usage (Part - 2)*, available on: `https://www.youtube.com/watch?v=HOT2wfvQ-GQ`.

Disclaimer: For the video links above, I, or DA-IICT, do not endorse any of the online learning platforms and the video creators. Links are provided since the content is freely available and I assume that it might help in learning.

General Instructions:

- There is a lot of help available online. You should definitely search for your queries online to get an early and a better resolution.
- Your lab report must contain a list of steps you took to run the programs for the two problems above and the output. For putting the output, use the screen shot. Although it is desired that you solve the problems completely, but if this does not happen, you can give the output up to the stage you could reach while solving the problems.
- The lab is intentionally made from the text books and refers to a lot of online content, so that you have ample resources to refer to and learn.