

Assignment 3: Install Hadoop on a cluster

Task:

Use shell script programming (aka shell scripting) to write a script to Install Hadoop 3.3.0 in ubuntu 20.04.2.0-desktop-amd64 on a cluster (the cluster will consist of one master node and 2 slaves/workers). Use the script files that we used in class to install Hadoop on a single machine.

Overview:

This assignment is designed to support your in-class understanding of how HDFS works and to get some hands-on experience in using it. You will need to deploy Apache Hadoop (HDFS as the underlying file system and MapReduce and Yarn as the execution engine).

Learning Outcomes:

After completing this programming assignment, you should have experience in deploying and configuring Apache Hadoop (HDFS, MapReduce, and Yarn).

Deliverables:

You should submit a ZIP file that consists of a brief report and the code of each step. Put a README file in for each task and provide the instructions about how to run your code. Also, include a run.sh script that can re-execute your code on a similar Virtual Machines cluster.