

Running Hadoop 2.7.3 single node cluster on Ubuntu 16.04 LST

Ziad Muhaisen and Mohammad Nassar

Apache Hadoop has the potential to revolutionize the processing of very large data sets. Due to its importance, Hadoop is receiving an increasing attention among research literature in recent years, where many implementations and algorithms have been proposed to overcome constrains and challenges in Hadoop framework in order to improve its data availability and overcome heavy load failures. Hadoop consists of multiple nodes which offer transparency and redundancy by allowing the distribution of storage and processing over multiple different locations. This paper gives a brief outline to the storage part of apache Hadoop, known as “Hadoop Distributed File System (HDFS)”, and describes installation procedures, the environment requirements needed to run the Hadoop approach on a single node cluster, and how to overcome the problems during the installation.

Muhaisen, Ziad and Nassar, Mohammad, (2017), Running Hadoop 2.7.3 single node cluster on Ubuntu 16.04 LST, 2ndInternational Computer Sciences and Informatics Conference (ICSIC 2017).