

Scientific Linux Cluster for BIG-DATA Analysis (SLBD): A Case of Fayoum University

Authors : Hassan S. Hussein, Rania A. Abul Seoud, Amr M. Refaat

Abstract : Scientific researchers face in the analysis of very large data sets that is increasing noticeable rate in today's and tomorrow's technologies. Hadoop and Spark are types of software that developed frameworks. Hadoop framework is suitable for many Different hardware platforms. In this research, a scientific Linux cluster for Big Data analysis (SLBD) is presented. SLBD runs open source software with large computational capacity and high performance cluster infrastructure. SLBD composed of one cluster contains identical, commodity-grade computers interconnected via a small LAN. SLBD consists of a fast switch and Gigabit-Ethernet card which connect four (nodes). Cloudera Manager is used to configure and manage an Apache Hadoop stack. Hadoop is a framework allows storing and processing big data across the cluster by using MapReduce algorithm. MapReduce algorithm divides the task into smaller tasks which to be assigned to the network nodes. Algorithm then collects the results and form the final result dataset. SLBD clustering system allows fast and efficient processing of large amount of data resulting from different applications. SLBD also provides high performance, high throughput, high availability, expandability and cluster scalability.

Keywords : big data platforms, cloudera manager, Hadoop, MapReduce

Conference Title : ICECECE 2016 : International Conference on Electrical, Computer, Electronics and Communication Engineering

Conference Location : Jeddah, Saudi Arabia

Conference Dates : January 26-27, 2016