Hadoop software for large amount of data

Technology, Information Technology

Its a platform managed under the apache software foundation and its an open source and its deal with big data with any data type structer semi structer or unstructers and give the result in very short time it allows to work with structured and unstructured data arrays of dimension from 10 to 100 gb and even more.

V. Burunova and its structer is a group of clusters or one each of them contains groups of nodes too and each cluster has two type of node name node and data node name node is a unique node on cluster and it knows any data block location on cluster and data node is the remining node in cluster and that have done by using a set of servers which called a cluster.

Hadoop has two layers cooperate together first layer is mapreduce and it task is divided data processing across multiple servers and the second one is hadoop distributed file system hdfs and its task is storing data on multiple clusters and these data are separated as a set of blocks. Hadoop make sure the work is correct on clusters and it can detect and retrieve any error orfailurefor one or more of connecting nodes and by this way hadoop efforts increasing in core processing and storage size and high availability. Hadoop is usually used in a large cluster or a public cloud service such as yahoo.

Facebook twitter and amazon hadeer mahmoud 2018 hadoops features: Scalable: Hadoop able to work with huge applications and it can run analyze store process distribute large amount of data across thousands of nodes and servers which handle thousands terabytes of data or more also it can add additional nodes to clusters and these servers work parallel. Hadoop better

than traditional relational database systems because rdbms cant expand to deal with huge data.

Single write multiple read the data on cluster can be read from multiple source at the same time data avalibility: When data is sent to a data node that hadoop creates multiple copies of data on other nodes in the cluster to keep data available if there a failure on one of nodes on cluster.