**Hadoop Cluster Overview: What it is and how to setup one?**

## **What is a Hadoop Cluster?**

In general, a computer cluster is a collection of various computers that work collectively as a single system.

“A [hadoop](https://www.dezyre.com/article/hadoop-explained-how-does-hadoop-work-and-how-to-use-it-/237" \o "What is Hadoop?" \t "_blank) cluster is a collection of independent components connected through a dedicated network to work as a single centralized data processing resource. “

“A hadoop cluster can be referred to as a computational computer cluster for storing and analysing big data (structured, semi-structured and unstructured) in a distributed environment.”

“A computational computer cluster that distributes data analysis workload across various cluster nodes that work collectively to process the data in parallel.”

Hadoop clusters are also known as “Shared Nothing” systems because nothing is shared between the nodes in a hadoop cluster except for the network which connects them. The shared nothing paradigm of a hadoop cluster reduces the processing latency so when there is a need to process queries on huge amounts of data the cluster-wide latency is completely minimized.

If you would like more information about Big Data and Hadoop Training, please click the orange "Request Info" button on top of this page.

## **Advantages of a Hadoop Cluster Setup**

* As big data grows exponentially, parallel processing capabilities of a [Hadoop cluster](https://www.dezyre.com/hadoop-wiki#hadoop-cluster) help in increasing the speed of analysis process. However, the processing power of a hadoop cluster might become inadequate with increasing volume of data. In such scenarios, hadoop clusters can scaled out easily to keep up with speed of analysis by adding extra cluster nodes without having to make modifications to the application logic.
* Hadoop cluster setup is inexpensive as they are held down by cheap commodity hardware. Any organization can setup a powerful hadoop cluster without having to spend on expensive server hardware.
* Hadoop clusters are resilient to failure meaning whenever data is sent to a particular node for analysis, it is also replicated to other nodes on the hadoop cluster. If the node fails then the replicated copy of the data present on the other node in the cluster can be used for analysis.