

{tag}

{/tag}

IJCA Proceedings on Recent Innovations in

Computer Science and Information Technology

© 2016 by IJCA Journal

RICSIT 2016 - Number 1

Year of Publication: 2016

Authors:

Dilbag Singh

Chirag Goyal

{bibtex}ricsit2019.bib{/bibtex}

Abstract

In this modern era, analysis of enormous amount of data is becoming a big challenge to the decision makers. Big data is the datasets in size as well as high in variety, velocity and volume.

So there is a need of the mean to handle and extract valuable insights from these datasets for better precision. It is very tedious rather impossible in some cases to handle enormous data using traditional databases and techniques their being the need for massive parallel processing and scalability which is not supported by the existing methods. Hadoop supports the scalability as it provides big storage and distribute big data sets over large no of servers operating in parallel. Traditional relational database systems don't scale to process the big data.

Scaling of traditional RDBMS to such big data increases cost in many folds which is not affordable. Making efforts to reduce cost, the organizations have had to down-sample data and classify the data on assumptions by deleting raw data that may be useful only for a short term. Hadoop is designed as a scale out architecture and can affordably store company's data for use in future. In the present paper the Big Data Analytics has been carried out using experimental research method. Structured Queries are executed by setting up Hadoop Cluster and RDBMS environment using secondary datasets. The response time of RDBMS with Hadoop framework will be compared.

Refer

ences

- Tom White "Hadoop Definitive Guide" , Second Edition, O'Reilly Media, pp 1-9, October-2010.
- Shiqi Wu , Big Data Processing with Hadoop, pp. 13-16, June 2015.
- A Review Paper on Big Data and Hadoop, International Journal of Scientific and Research Publications, Volume 4, Issue 10, October 2014.
- Mark Kerzner and Sujee Maniyam, "Hadoop Illuminated", GitHub, pp. 28-30, 2014.
- Apache Hadoop, MapReduce Tutorial, 2013. https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html, accessed April 2014.
- Ketaki Subhash Raste, "Big Data Analytics-Hadoop Performance Analysis", pp. 18-22, 2014.
- Rui Xue, "SQL Engines for Big Data Analytics: SQL on Hadoop", pp 31-41, Nov 20,2015.
- Jefferey Shafer ,"A Storage Architecture for Data-Intesive Computing" , pp. 87-100, May 2010.

Index Terms

Computer Science

Information Sciences

Keywords

Big Data Hadoop Cluster Hdfs Map Reduce.