International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 12 | December - 2022 Impact Factor: 7.185 ISSN: 2582-3930

THE STUDY ON HADOOP ECOSYSTEM

Dr C K Gomathy, Assistant Professor, Department of CSE, SCSVMV Deemed to be University, India

Ms. S.Tarini, Ms.P. Sree Mahi, Ms.R. Naga Mallika, Ms.R.V.S.Tejaswini

UG Scholars- SCSVMV Deemed to be University, India

ABSTRACT

The Hadoop Ecosystem is a platform or framework that addresses big data issues; it is neither a programming language nor a service. You can think of it as a collection of services that includes consuming, storing, analysing, and managing data. This article defines the concept of big data, which is a collection of enormous data sets that typical computational methods cannot handle. Hadoop is a system created to process large amounts of data. Businesses use Hadoop as their platform for processing big data. In a distributed computing context, Hadoop is an open source, Java-based programming platform that facilitates the processing and archiving of very massive data collections. By solving the challenges that are typically encountered when managing Big Data, it supports Big Data analytics. Hadoop may fail.

Keywords: Hadoop, hadoop ecosystem, Distributed Computing, Business Intelligence

I. INTRODUCTION

The Hadoop Ecosystem is a platform or collection of tools that offers a range of services to address big data issues. It consists of Apache projects as well as a number of paid tools and services. Hadoop is made up of four main components: HDFS, MapReduce, YARN, and Hadoop Common. Most of the time, these important components are supplemented or supported by tools or solutions. Together, these instruments can offer services including data absorption, analysis, storage, and maintenance.

Big data is an enormous, unstructured data set for which typical data processing application software is insufficient. Data collection, storage, analysis, data search, sharing, transfer, visualisation, querying, updating, and information privacy are all big data challenges. Big data's three dimensions, Volume, Variety, and Time.

II.HADOOP STAGES

Stage 1: The user or application sends a task to Hadoop with the following parameters for the necessary process:

- The location of input and output files in the distributed file system.
- The java classes in the form of jar file containing the implementation of map and reduce functions.
- The job configuration by setting different parameters specific to the job.

Stage 2: The Hadoop job client then sends the job (jar/executable) and configuration to the JobTracker, which is in charge of scheduling tasks, distributing the configuration to the slaves, and monitoring. The JobTracker also gives status and diagnostic information to the job-client.

III. METHODOLOGY

HDFS and Map Reduce are used in Hadoop's master-slave architecture design for data storage and distributed data processing. Hadoop HDFS is the Name Node's master node for data storage, while Job Tracker is the Name Node's master node for Hadoop Map Reduce's concurrent data processing. The additional computers in the Hadoop cluster that store data and carry out sophisticated computations are referred to as slave nodes in the Hadoop architecture. The Task Tracker and Data Node on each slave node synchronise the running processes with the Job Tracker and Name Node. The master or slave systems in the Hadoop architectural implementation can be put up in the cloud.

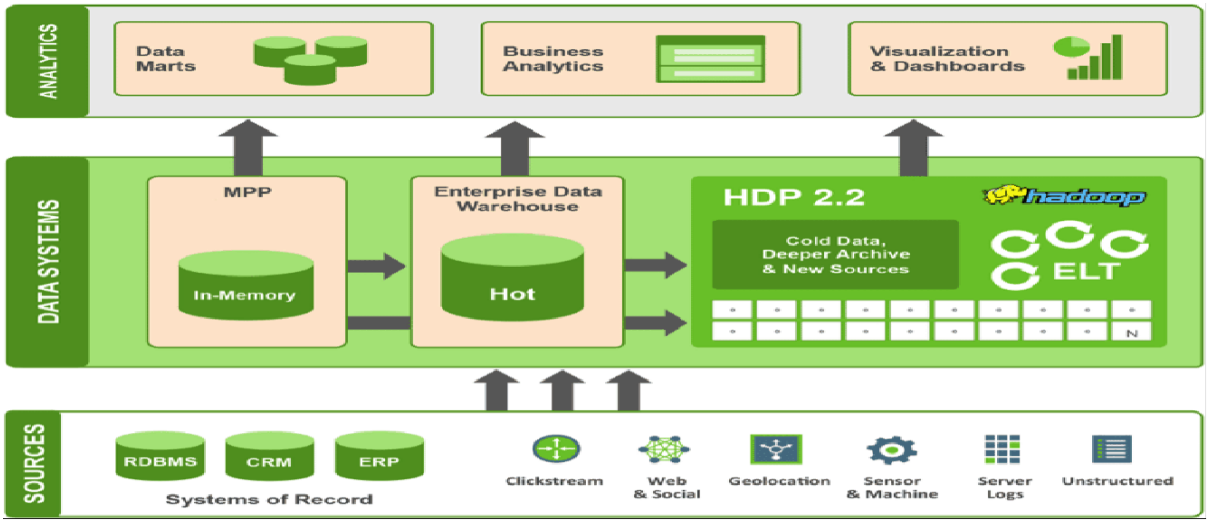

Fig 1: Architecture of Optimized Data with Hadoop

IV. IMPLEMENTATION

Implementation Steps:

A. Installation of cloudera CDH 5.8

https://www.cloudera.com/downloads/quickstart_vms/5-8.html

After download tar file of cloudera CDH5.8 vm image file.

usingvmware player install it given below video link.

https://www.youtube.com/watch?v=4XBXJpYPkUk

B.Ubuntu Installation:

1. Downloadubuntu 16.04 from this link

http://www.ubuntu.com/download/desktop/contribute?version=16.04.1&architecture=amd64

https://www.youtube.com/watch?v=KfOt2As6apQ

Hadoop Installation:

2.InstallHadoop 2.7 using following given steps below:

3.Open command line terminal using ctrl+shift+ T

•Installing Oracle Java 8: run the below command on $ shell

```
sudo add-apt-repository ppa:webupd8team/java
sudo apt-get update
sudo apt-get install oracle-java8-installer
```

•Installing SSH

```
sudo apt-get install openssh-server
```

•Configuring SSH

```
ssh-keygen -t rsa -P ""
cat$HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
```

4.Download latest Apache Hadoop source from Apache

mirrors

First you need to download hadoop 2.7.3 binary file from the

give path given below

http://hadoop.apache.org/releases.html

•Copy the Hadoop 2.7.3 folder tar file in home directory->

/home/username/Work

5.User profile :sudonano ~/.bashrc

```
# -- HADOOP ENVIRONMENT VARIABLES START -- #
export JAVA_HOME=/usr/lib/jvm/java-8-oracle
export HADOOP_HOME=/home/username/Work
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export
HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HO
ME/lib/native
export HADOOP_OPTS="-
Djava.library.path=$HADOOP_HOME/lib"
# -- HADOOP ENVIRONMENT VARIABLES END -- #
```

6.Commit the changes of .bashrc

Source ~/.bashrc

7.Configuration file : hadoop-env.sh

To edit file, fire the below given command

hduser@pingax:/home/username/Work/hadoop2.7.3/hadoop/e

tc/hadoop$ sudogedit hadoop-env.sh

Update JAVA_HOME variable,

JAVA_HOME=/usr/lib/jvm/java-8-oracle

8.Configuration file : core-site.xml

To edit file, fire the below given command
hduser@pingax:/home/username/Work/hadoop2.7.3/hadoop/e
tc/hadoop$ sudogedit core-site.xml
Paste these lines into <configuration> tag
<property>
<name>fs.default.name</name>
<value>hdfs://localhost:9000</value>
</property>

V.CONCLUSION

This paper is about Hadoop ecosystem and has explored its major components as well as Hadoop setup. Various aspects of data storage is focused like HDFS and its architecture. The process of installation of Hadoop setup is analyzed. HDFS ensures data integrity throughout the cluster considering features like maintaining transaction logs. Another feature is validating checksum-an effective error detection technique wherein numerical value is assigned to a transmitted message on the basis of number of bits. HDFS maintains replicated copies of data blocks to avoid corruption of file due to failure of server. This paper also deals with MapReduce framework, which is an integration of different functions to sort,process and analyze bigdata. The future research includes implementing various technologies for optimizing and improving performance on large data set. The experimental results to be analyzed using various tools and experimental setup.

VI.REFERENCES

[1] DR.C.K.Gomathy , V.Geetha , S.Madhumitha , S.Sangeetha , R.Vishnupriya Article: A Secure With Efficient Data Transaction In Cloud Service, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, March 2016, ISSN: 2278 – 1323.

[2] Dr.C.K.Gomathy,C K Hemalatha, Article: A Study On Employee Safety And Health Management International Research Journal Of Engineering And Technology (Irjet)- Volume: 08 Issue: 04 | Apr 2021

[3] Dr.C K Gomathy, Article: A Study on the Effect of Digital Literacy and information Management, IAETSD Journal For Advanced Research In Applied Sciences, Volume 7 Issue 3, P.No-51-57, ISSN NO: 2279-543X,Mar/2018

[4] Dr.C K Gomathy, Article: An Effective Innovation Technology In Enhancing Teaching And Learning Of Knowledge Using Ict Methods, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) E-Issn: 2395-5325 Volume3, Issue 4,P.No-10-13, April ’2017

[5] Dr.C K Gomathy, Article: Supply chain-Impact of importance and Technology in Software Release Management, International Journal of Scientific Research in Computer Science Engineering and Information Technology (IJSRCSEIT) Volume 3 | Issue 6 | ISSN : 2456-3307, P.No:1-4, July-2018.

[6] C K Gomathy and V Geetha. Article: A Real Time Analysis of Service based using Mobile Phone Controlled Vehicle using DTMF for Accident Prevention. International Journal of Computer Applications 138(2):11-13, March 2016. Published by Foundation of Computer Science (FCS), NY, USA,ISSN No: 0975-8887

[7] C K Gomathy and V Geetha. Article: Evaluation on Ethernet based Passive Optical Network Service Enhancement through Splitting of Architecture. International Journal of Computer Applications 138(2):14-17, March 2016. Published by Foundation of Computer Science (FCS), NY, USA, ISSN No: 0975-8887

[8] C.K.Gomathy and Dr.S.Rajalakshmi.(2014), "A Software Design Pattern for Bank Service Oriented Architecture", International Journal of Advanced Research in Computer Engineering and Technology(IJARCET), Volume 3,Issue IV, April 2014,P.No:1302-1306, ,ISSN:2278-1323.

[9] C. K. Gomathy and S. Rajalakshmi, "A software quality metric performance of professional management in service oriented architecture," Second International Conference on Current Trends in Engineering and Technology - ICCTET 2014, 2014, pp. 41-47, doi: 10.1109/ICCTET.2014.6966260.

[10] Dr.C K Gomathy, V Geetha ,T N V Siddartha, M Sandeep , B Srinivasa Srujay Article: Web Service Composition In A Digitalized Health Care Environment For Effective Communications, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, April 2016, ISSN: 2278 – 1323.

[11] C.K.Gomathy.(2010),"Cloud Computing: Business Management for Effective Service Oriented Architecture" International Journal of Power Control Signal and Computation (IJPCSC), Volume 1, Issue IV, Oct - Dec 2010, P.No:22-27, ISSN: 0976-268X .

[12] Dr.C K Gomathy, Article: A Study on the recent Advancements in Online Surveying , International Journal of Emerging technologies and Innovative Research (JETIR) Volume 5 | Issue 11 | ISSN : 2349-5162, P.No:327-331, Nov-2018

[13] Dr.C.K.Gomathy,C K Hemalatha, Article: A Study On Employee Safety And Health Management International Research Journal Of Engineering And Technology (Irjet)- Volume: 08 Issue: 04 | Apr 2021

[14] Dr.C K Gomathy, V Geetha , T.Jayanthi, M.Bhargavi, P.Sai Haritha Article: A Medical Information Security Using Cryptosystem For Wireless Sensor Networks, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) E-Issn: 2395-5325 Volume3, Issue 4, P.No-1-5,April ’2017

[15] C.K.Gomathy and Dr.S.Rajalakshmi.(2014), "Service Oriented Architecture to improve Quality of Software System in Public Sector Organization with Improved Progress Ability", Proceedings of ERCICA-2014, organized by Nitte Meenakshi Institute of Technology, Bangalore. Archived in Elsevier Xplore Digital Library, August 2014, ISBN:978-9-3510-7216-4.

[16] Parameshwari, R. & Gomathy, C K. (2015). A Novel Approach to Identify Sullied Terms in Service Level Agreement. International Journal of Computer Applications. 115. 16-20. 10.5120/20163-2253.

[17] C.K.Gomathy and Dr.S.Rajalakshmi.(2014),"A Software Quality Metric Performance of Professional Management in Service Oriented Architecture", Proceedings of ICCTET’14, organized by Akshaya College of Engineering, Coimbatore. Archived in IEEE Xplore Digital Library, July 2014,ISBN:978-1-4799-7986-8.

[18] C.K.Gomathy and Dr.S.Rajalakshmi.(2011), "Business Process Development In Service Oriented Architecture", International Journal of Research in Computer Application and Management (IJRCM) ,Volume 1,Issue IV, August 2011,P.No:50-53,ISSN : 2231-1009

19. http://www.hadooptpoint.com/introduction-hive/

AUTHOR'S PROFILE:

Dr. C.K. Gomathy is Assistant Professor in Computer Science and Engineering at Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Enathur, Kanchipuram, India. Her area of interest in Software Engineering, Web Services, Knowledge Management

Ms. S. Tarini, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa MahaVidyalaya Enathur, Kanchipuram, India. Her area of interest: Machine Learning and Cyber Security.

Ms.P.Sree Mahi, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya Enathur, Kanchipuram, India. Her area of interest: Artificial Intelligence.

Ms.R.Naga Mallika, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa MahaVidyalaya Enathur, Kanchipuram, India. Her area of interest: Artificial Intelligence.

Ms.R.V.S.Tejaswini, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa MahaVidyalaya Enathur, Kanchipuram, India. Her area of interest: Cloud Computing.