ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Big Data Mining Platforms: A Survey

Tejaswini U. Mane, Asha M. Pawar

M.E. Student, Dept. of Computer, ZCOER, Savitribai Phule Pune University, Pune, India

Asst. Prof., Dept. of Computer, ZCOER, Savitribai Phule Pune University, Pune, India

ABSTRACT: Big information came into existence once the traditional relative info systems weren't ready to handle the unstructured data (weblogs, videos, photos, social updates, human behavior) generated today by organisation, social media, or from any other information generating source. Data that is therefore giant in volume, so diverse in selection or moving with such speed is referred to as Big data. Analyzing Big information is a difficult task because it involves large distributed file systems that ought to be fault tolerant, flexible and climbable. The technologies used by big information application to handle the huge information are Hadoop, Map Reduce, Apache Hive, No SQL and HPCC. These technologies handle massive quantity of information in MB, PB, YB, ZB, KB and TB. In this research paper varied technologies for handling big information on with the advantages and disadvantages of every technology for catering the issues in hand to deal the huge information has discussed.

KEYWORDS: Big Data, Hadoop, Map Reduce, Apache Hive, No SQL.

I. INTRODUCTION

With the growth of technological development and services, the large quantity of knowledge is made which will be structured and unstructured from the different sources in different domains. Massive information of such type is terribly difficult to method that contains the data of the records of million people that includes everyday massive quantity of information from social sites, cell phones GPS signals, videos etc. Big information is a largest buzz phrases in domain of IT, new technologies of personal communication driving the big information new trend and internet population grew day by day however it ne'er reach by 100%. The need of huge information generated from the massive companies like facebook, yahoo, Google, YouTube etc for the purpose of study of enormous quantity of knowledge which is in unstructured kind or even in structured form. Google contains the large quantity of data. So; there is the necessity of huge Data Analytics that's the processing of the complicated and huge datasets This data is totally different from structured information (which is hold on in relational information systems) in terms of 5 parameters –variety, volume, value, veracity and speed (5V's). The five V's (volume, variety, velocity, value, veracity) are the challenges of huge information management area unit [1]:

1. Volume: Data is ever-growing day by day of all types ever MB, PB, YB, ZB, KB, TB of information. The data results into giant files. Excessive volume of data is main issue of storage. This main issue is resolved by reducing storage cost. Data volumes area unit expected to grow fifty times by 2020.
2. Variety: Data sources (even in the same field or in distinct) are very heterogeneous [1]. The files comes in various formats and of any kind, it may be structured or unstructured like text, audio, videos, log files and more. The varieties are endless, and the data enters the network while not having been quantified or qualified in any way.
3. Velocity: The data comes at high speed. Sometimes one minute is too late therefore huge information is time sensitive. Some organisations data speed is main challenge. The social media messages and credit card transactions drained millisecond and data generated by this putt in to databases. 4. Value: Which addresses the would like for valuation of enterprise data? It is a most significant v in big
data. Value is main buzz for huge information as a result of it's important for businesses, IT infrastructure system to store large quantity of values in information.
5. Veracity: The increase within the range of values typical of a large information set. When we have a tendency to handling high volume, velocity and selection of information, the all of data aren't going 100% correct, there will be dirty data. Big information and analytics technologies work with these types of information.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

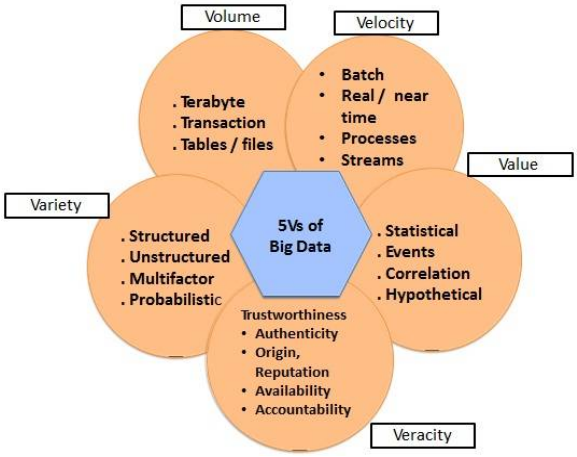

Fig 1. Big Data Parameters

Huge volume of knowledge (both structured and disorganized) is management by grouping, control and governance. Unstructured data is a data that is not gift in a very info. Disorganized data may be text, verbal data or in another kind. Textual unstructured data is like power purpose presentation, email messages, word documents, and instant massages. Data in another format can be. png images, jpg images, audio files (.mp3, .wav, .aiff) and video files that can be in flash format, .mkv format or .3gp format. According to the "IDC Enterprise Disk Cache Expenditure Model" report released in year 2009, in which the transactional data is projected to raise at a composite yearly growthrate (CAGR) of 21.8%, it's far outpaced by a sixty one.7% CAGR calculation for disorganized data [3].

From last twenty years, the data is mounting day by day across the world in every domain. Some specific facts about the data are, there are regarding 277,000 tweets per minute, 2 million queries or so on Google each minute in all domains, 75 hours of new videos in contrasting formats are uploaded to YouTube, More than 100 million emails are sent via Gmail, yahoo, rediff mail and many a lot of, 350 GB of data is dealing out on facebook every day and quite 576 websites are created every minute. During the year 2012, 2.5 quintillion bytes of knowledge were created daily. Big data and its depth analysis is the core of recent science, research space and business areas. Huge quantity of data is generated from the distinct numerous sources either in structure or unstructured form. Such form of data keep in databases and then it become terribly complex to extract, transform and create in use [8]. IBM indicates that 2.5 Exabyte knowledge is created everyday which is terribly tough to investigate in numerous aspects. The estimation about the generated knowledge is that until year 2003 it was represented regarding five Exabyte, then until year 2012 is 2.7 Zettabyte and until 2015 it is expected to boost up to three times [10]. This paper is organized as follows. In section I Related Work have been described alongside advantages and disadvantages of the paper. In section II the various huge knowledge techniques has been mentioned. Future Scope has been discussed in section III for direction to emerging researchers and Final section gives a conclusion of the paper.

II. RELATED WORK

1. John A. Keane [2] in 2013 suggested a structure in which massive information applications will be developed. The framework consist of three stages (multiple information sources, data analysis and modelling, data organization and analysis) and seven layers (visualisation/presentation layer, service/query/access layer, modelling/ statistical layer, processing layer, system layer, data layer/multi model) to divide big information application into blocks. The main motive of this paper is to administer and architect a large quantity of huge information applications. The advantage of this paper is big information handles heterogeneous information and data sources in timely to get high performance and Framework Bridge the gap with business needs and technical realities. The disadvantage of this paper is too difficult to integrate existing information and systems.

2. Xin Luna Dong [5] in 2013 explained problems of big information assimilation (schema mapping, record linkage and data fusion). These challenges are explained by using examples and techniques for
data integration in addressing the new challenges raised by big information, includes volume and number of sources, velocity, variety and truthfulness. The advantage of this paper is identifying the information source issues to integrate

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

existing information and systems. The disadvantage of this paper is big information integration such as integrating information from markets, integrating crowd sourcing information, providing an exploration tool for data sources.
3. Jun Wang [17] in 2013 suggested the Data-g Rouping-Aware (DRAW) data placement theme to improve the issues like performance, readiness, execution and latency. It could cluster many sorted information into a little variety of nodes as compared to map reduce/hadoop. the three main phases of DRAW characterized in this paper are: cluster the data-grouping matrix, learning data organization information from system logs and recognizing the grouping data. The advantage of the paper is improve the throughput up to fifty nine.8%, reduce the execution time up to 41.7% and improve the overall performance by 36.4% over the Hadoop/map reduce.
4. Yaxiong Zhao [7] in 2014 proposed information aware caching (Dache) framework that made minimum change to the original map scale back programming model to increment processing for massive information applications using the map scale back model. It is a protocol, data aware cache description theme and architecture. The advantage of this paper is, it boost the completion time of map reduce jobs.
5. Jian Tan [6] in 2013 author talks about the theoretical assumptions, that improves the performance of Hadoop/map reduce and purposed the optimal scale back task assignment schemes that minimize the fetching value per job and performs he both simulation and real system preparation with experimental evolution. The advantage of this paper is improves the performance of large scale Hadoop clusters. The disadvantage of this paper is environmental factors such as network topologies effect on a scale back task in map scale back clusters.
6. Thuy D. Nguyen [4] (2013) author solve the multilevel secure (MLS) environmental issues of Hadoop by using security increased UNIX system (SE Linux) protocol. In which multiple sources of Hadoop applications run at different levels. This protocol is an extension of Hadoop distributed file system (HDFS). The advantage of this paper is solving environmental problems while not requiring complex Hadoop server parts.
7. Keith C.C. Chan [15] 2013 author describes large
amount of structured and unstructured information assortment, processing and analysis from hospitals, laboratories, pharmaceutical, companies or even social media and also discus regarding however to collect or analyse huge volume of information for drug discovery. The advantage of this paper is how massive information analytics contributes to better drug safety efficaciousness for pharmaceutical regulators and companies. The disadvantage of this paper it needs the algorithms that are straightforward, scalable, efficient and effective for data discovery method.
8. Sagiroglu, S. [8] (2013) offered the big knowledge content, its scope, functionality, data samples, advantages and disadvantages together with challenges of big knowledge. The critical issue in relation to the Big knowledge is that the privacy and protection. Big data samples describe the review concerning the environment, science and research in biological area. By this paper, we will conclude that any association in any domain having big knowledge will take the benefit from its careful investigation for the problem finding principle. Using information Discovery from the Big knowledge convenient to induce the information from the sophisticated knowledge records. The overall appraisal describe that the information is mounting day by day and becoming advanced. The challenge is not only to assemble and handle the information but conjointly however to extract the helpful data from that collected data records. In accordance to the Intel IT Center, there are many challenges related to massive knowledge that area unit speedy knowledge growth, data infrastructure, and variety of knowledge, visualization and knowledge speed.
9. Garlasu, D. [10] (2013) discussed the enhancement concerning the storage capabilities, the processing power on with handling technique. The Hadoop technology is widely used for the simulation purpose. Grid Computing provides the notion of distributed computing using HDFS. The benefit of Grid computing is that the most storage capability and the high processing power. Grid Computing makes the big help among the scientific research and facilitate the man of science to analyze and store the large and complicated knowledge in various formats.
10.Mukherjee, A. [11] (2012) The Big knowledge analysis define the massive quantity of knowledge to retrieve the useful data and uncover the hidden information. Big knowledge analytics refers to the Map Reduce Framework that is discovered by the Google. Apache Hadoop is the open source platform which is used for the aim of simulation of Map Reduce Model. In this the performance of SF-CFS is compared with the HDFS with the help of the SWIM by the facebook job traces. SWIM contains the workloads of thousands of jobs with complex and large knowledge arrival and computation patterns.
11. Aditya B. [12] (2012) defines big knowledge drawback using Hadoop and Map Reduce" reports the experimental research on the massive knowledge issues in various domains. It describe the optimal and efficient solutions exploitation Hadoop cluster, Hadoop Distributed File System (HDFS) for storage data and Map Reduce framework for parallel processing to method large knowledge sets and records.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

III. BIG DATA MINING PLATFORMS

Big information is a new thought for handling huge data so the bailiwick description of this technology is very new. There are the totally different technologies which use nearly same approach i.e. to distribute the data among varied native agents and reduce the load of the main server so traffic are often avoided. There are endless articles, books and periodicals that describe Big information from a technology perspective so we tend to can instead focus our efforts here on setting out some basic principles and also the minimum technology foundation to help relate massive information to the broader IM domain.

A. Hadoop

Hadoop is a framework that may run applications on systems with thousands of nodes and terabytes. It distributes the file among the nodes and allows to system continue work in case of a node failure. This approach reduces the risk of catastrophic system failure. In which application is broken into smaller components (fragments or blocks).Apache Hadoop consists of the Hadoop kernel, Hadoop distributed file system (HDFS), map reduce and connected comes are zookeeper, Hbase, Apache Hive. Hadoop Distributed File System (HDFS) consists of three Components: the Name Node, Secondary Name Node and Data Node [15]. The multilevel secure (MLS) environmental issues of Hadoop by using security increased Linux (SE Linux) protocol. In which multiple sources of Hadoop applications run at different levels. This protocol is an extension of Hadoop distributed file system (HDFS) [12]. Hadoop is commonly used for distributed batch index building; it is desirable to optimize the index capability in near real time. Hadoop provides components for storage and analysis for massive scale processing [1]. Now a day's Hadoop used by a whole bunch of companies. The advantage of Hadoop is Distributed storage & Computational capabilities, extremely climbable optimized for high output, large block sizes, tolerant of software and hardware failure.

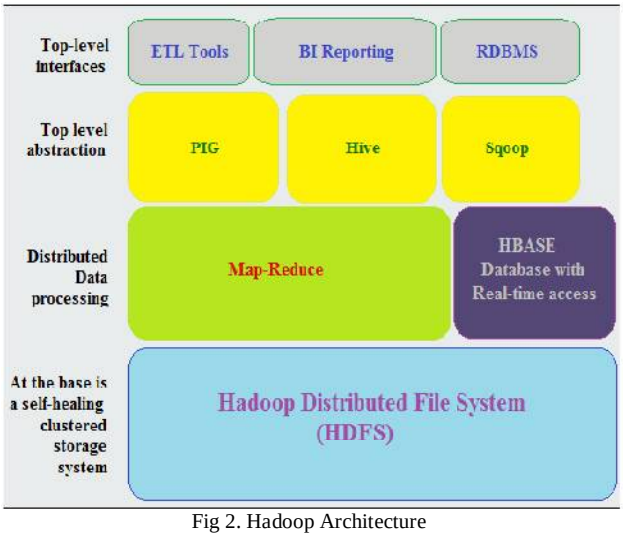

Fig 2. Hadoop Architecture

The disadvantage of Hadoop is that it is master processes are single points of failure. Hadoop does not offer storage or network level encoding, inefficient for handling small files.

Components of Hadoop [8]:

- HBase: It is open source, distributed and Non relational database system enforced in Java. It runs above the layer of HDFS. It can serve the input and output for the Map Reduce in well mannered structure.
- Oozie: Oozie is a web-application that runs in an exceedingly java servlet. Oozie use the database to gather the information of progress that is a assortment of actions. It manages the Hadoop jobs in a mannered method.
- Sqoop: Sqoop is a command-line interface application that provides platform which is employed for converting information from relative databases and Hadoop or vice versa.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

- Avro: It is a system that gives functionality of data publishing and service of knowledge exchange. It is basically utilized in Apache Hadoop. These services can be used along in addition as independently according the information records.
- Chukwa: Chukwa is a framework that's used for data assortment and analysis to method and analyze the massive quantity of logs. It is built on the higher layer of the HDFS and Map Reduce framework.
- Pig: Pig is high-level platform where the Map Reduce framework is created that is employed with Hadoop platform. It is a high level processing system where the information records area unit analyzed that occurs in high level language.
- Zookeeper: It is a centralization based service that provides distributed synchronization and provides group services on with maintenance of the configuration information and records.
- Hive: It is application developed for data warehouse that provides the SQL interface also as relational model. Hive infrastructure is built on the top layer of Hadoop that facilitate in providing conclusion, and analysis for respective queries.

B. Map Reduce

Map-Reduce was introduced by Google in order to process and store giant datasets on goods hardware. Map Reduce is a model for process largescale data records in clusters. The Map Reduce programming model is based on 2 functions that are map() perform and reduce() perform. Users can simulate their own processing logics having well defined map() and reduce() functions. Map function performs the task as the master node takes the input, divide into smaller sub modules and distribute into slave nodes. A slave node further divides the sub modules again that lead to the stratified tree structure. The slave node processes the base problem and passes the result back to the master Node. The Map Reduce system organize along all intermediate pairs based on the intermediate keys and refer them to reduce() function for manufacturing the final output. Reduce function works as the master node collects the results from all the sub problems and combines them along to form the output [19].

Map(in_key,in_value)---
>list(out_key,intermediate_value)
Reduce(out_key,list(intermediate_value))---
>list(out_value)
The parameters of map () and reduce () perform is
as follows:
map (k1,v1) ! list (k2,v2) and reduce (k2,list(v2))
! list (v2)

A Map Reduce framework is based mostly on masterslave architecture wherever one master node handles a number of slave nodes [18]. Map Reduce works by initial dividing the input information set into even-sized data blocks for equal load distribution. Each information block is then assigned to one slave node and is processed by a map task and result is generated. The slave node interrupts the master node when it is idle. The scheduler then assigns new tasks to the slave node. The scheduler takes
data section and resources into thought once it disseminates data blocks.

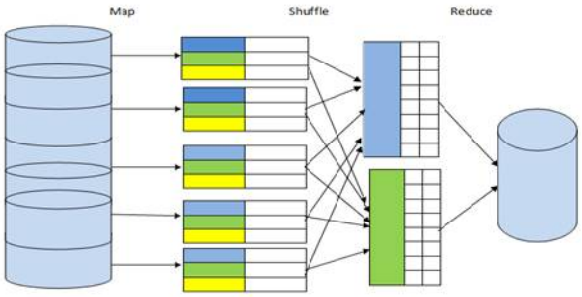

Fig 3: Mapreduce Architecture

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Figure 3 shows the Map scale back design and Working. It always manages to allot a native information block to a slave node. If the effort fails, the scheduler will assign a rack-local or random information block to the slave node instead of local information block. When map() function complete its task, the runtime system gather all intermediate pairs and launches a set of condense tasks to produce the ultimate output. Large scale information process is a difficult task, managing hundreds or thousands of processors and managing parallelization and distributed environments makes is more tough. Map Reduce provides solution to the mentioned problems, as is supports distributed and parallel I/O scheduling, it is fault tolerant and supports scalability and it has inherent processes for status and observance of heterogeneous and large datasets as in massive information [18]. It is way of approaching and *solving a given downside. Using Map Reduce framework the potency and the time to retrieve the data is sort of* manageable. To address the volume aspect, new techniques have been proposed to enable parallel process victimization Map scale back framework [13]. Data aware caching (Dache) framework that made slight amendment to the original map reduce programming model and framework to enhance processing for massive information applications victimization the map reduce model [16]. The advantage of map reduce is a massive form of problems area unit simply speak able as Map scale back computations and cluster of machines handle thousands of nodes and fault-tolerance. The disadvantage of map reduce is period of time processing, not always terribly simple to implement, shuffling of data, batch processing.

Map Reduce Components:

1. Name Node: manages HDFS metadata, doesn't deal with files directly.
2. Data Node: stores blocks of HDFS—default replication levels for each block: three.
3. Job Tracker: schedules, allocates and monitors job execution on slaves—Task Trackers.
4. Task Tracker: runs Map Reduce operations.

C. HIVE

Hive is a distributed agent platform, a decentralized system for building applications by networking local system resources [8]. Apache Hive data deposition component, an component of cloud-based Hadoop ecosystem that offers a question language known as HiveQL that translates SQL-like queries into Map Reduce jobs mechanically. Applications of apache hive are SQL, oracle, IBM DB2. Architecture is divided into Map-Reduce-oriented execution, Meta data data for data storage, and an execution half that receives a query from user or applications for execution.

The advantage of hive is more secure and implementations are sensible and well tuned. The disadvantage of hive is only for unintentional queries and performance is less as compared to pig.

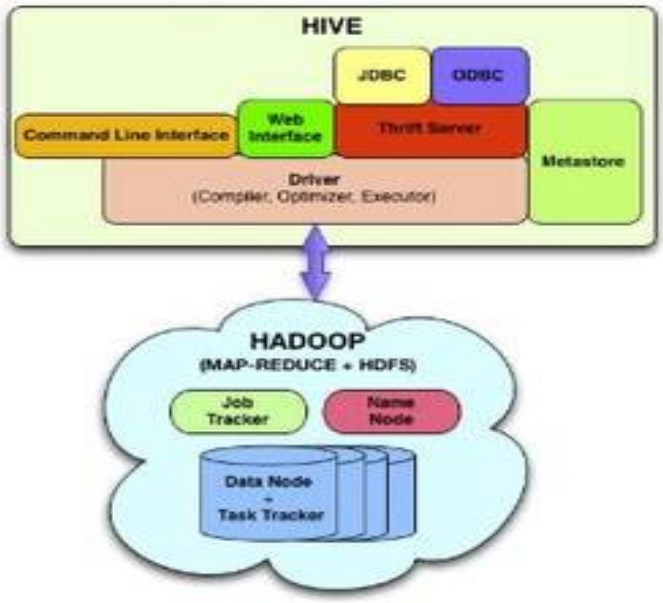

Fig. 4 Hive Architecture

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

D: NO SQL

No-SQL knowledgebase is AN approach to data management and data style that's helpful for terribly giant sets of distributed data. These databases are in general part of the period events that ar detected in method deployed to inbound channels however will additionally be seen as AN enabling technology following analytical capabilities such as relative search applications. These are solely made possible as a result of of the elastic nature of the No- SQL model where the spatiality of a question is evolved from the data in scope and domain instead of being fixed by the developer in advance. It is useful when enterprise would like to access immense quantity of unstructured data. There are a lot of than 100 No SQL approaches that specialize in management of different multimodal knowledge sorts (from structured to nonstructured) and with the aim to solve very specific challenges [5]. Data mortal, Researchers and Business Analysts in specific pay more attention to agile approach that leads to prior insights into the info sets that may be hid or forced with a a lot of formal development process. The most popular No- SQL database is Apache prophetess. The advantage of No-SQL is open source, Horizontal scalability, Easy to use, store complex knowledge types, Very quick for adding new knowledge and for easy operations/queries. The disadvantage of No-SQL is Immaturity, No indexing support, No ACID, Complex consistency models, Absence of standardization.

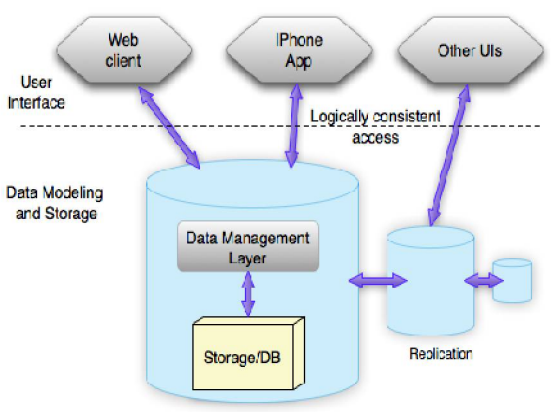

Fig: 5 NO SQL Architecture

E. HPCC

HPCC is an open supply platform used for computing and that provides the service for handling of massive huge information work flow. HPCC data model is defined by the user finish according to the wants. HPCC system is proposed and then more designed to manage the most complex and data-intensive analytical related issues. HPCC system is a single platform having one architecture and a single programming language used for the data simulation. HPCC system was designed to analyze the large amount of knowledge for the purpose of solving advanced drawback of huge information. HPCC system is based on enterprise management language which has the declarative and on-procedural nature programming language the most components of HPCC are:

- HPCC Data Refinery: Use parallel ETL engine mostly.
- HPCC Data Delivery: It is massively supported structured query engine used.

Enterprise Control Language distributes the workload between the nodes in applicable even load.

IV. FUTURE SCOPE

There is nothing concealed that huge information significantly influencing IT corporations and through development new technologies only we tend to will handle it in a managerial method. Big information altogether amendment the method of organizations, government and academic establishment by using range of tools to create the management of massive data. In future Hadoop and NoSQL database can be highly in demand moving forward. The amount of knowledge produced by organizations in next 5 years can be larger than last 5,000 years. In the upcoming years cloud will play the necessary role for personal sectors and organisations to handle the big information with efficiency.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

V. CONCLUSION

In this paper we have surveyed numerous technologies to handle the big information and there architectures. In this paper we have conjointly mentioned the challenges of Big information (volume, variety, velocity, value, veracity) and various blessings and a disadvantage of these technologies. This paper discussed Associate in Nursing architecture using Hadoop HDFS distributed information storage, real-time NoSQL databases, and MapReduce distributed data processing over a cluster of artifact servers. The main goal of our paper was to make a survey of various big information handling techniques those handle a huge amount of information from totally different sources and improves overall performance of systems.

REFERENCES

1. Yuri Demchenko "*The Big Data Architecture Framework (BDAF)"* Outcome of the Brainstorming Session at the University of Amsterdam 17 July 2013.
2. Tekiner F. and Keane J.A., Systems, Man and Cybernetics (SMC), "*Big Data Framework*" 2013 IEEE International Conference on 13–16 Oct. 2013, 1494–1499.
3. Margaret Rouse, "*unstructured data*" April 2010.
4. Nguyen T.D., Gondree M.A., Khosalim, J.; Irvine, "*Towards a Cross Domain MapReduce Framework*" IEEE C.E. Military Communications Conference, MILCOM 2013, 1436 – 1441
5. Dong, X.L.; Srivastava, D. Data Engineering (ICDE)," *Big data integration*" IEEE International Conference on , 29(2013) 1245–1248
6. Jian Tan; Shicong Meng; Xiaoqiao Meng; Li ZhangINFOCOM, "*Improving ReduceTask data locality for sequential MapReduce*" 2013 Proceedings IEEE ,1627 - 1635
7. Yaxiong Zhao; Jie Wu INFOCOM, "*Dache: A Data Aware Caching for Big-Data Applications Using the MapReduce Framework*" 2013 Proceedings IEEE 2013, 35 - 39 (Volume 19)
8. Sagiroglu, S.; Sinanc, D.,"*Big Data: A Review*",2013,20-24
9. Minar, N.; Gray, M.; Roup, O.; Krikorian, R.; Maes, "*Hive: distributed agents for networking things*" IEEE CONFERENCE PUBLICATIONS 1999 (118-129)
10. Garlasu, D.; Sandulescu, V.; Halcu, I.; Neculoiu, G,"*A Big Data implementation based on Grid Computing*", Grid Computing, 2013, 17-19
11. Mukherjee, A.; Datta, J.; Jorapur, R.; Singhvi, R.; Haloi, S.; Akram, "*Shared disk big data analytics with Apache Hadoop*", 2012, 18-22
12. Aditya B. Patel, Manashvi Birla, Ushma Nair, "*Addressing Big Data Problem Using Hadoop and Map Reduce*", 2012, 6-8
13. Jefry Dean and Sanjay Ghemwat, "*MapReduce:A Flexible Data Processing Tool*",Communications of the ACM, Volume 53, Issuse.1,2010, 72-77.
14. Chan,K.C.C. Bioinformatics and Biomedicine (BIBM), "*Big data analytics for drug discovery*" IEEE International Conference on Bioinformatics and Biomedicine 2013,1.
15. Kyuseok Shim, "*MapReduce Algorithms for Big Data Analysis*", DNIS 2013, LNCS 7813, pp. 44–48, 2013.
16. Wang, J.; Xiao, Q.; Yin, J.; Shang, P. Magnetics, DRAW: A New Data-gRouping-AWare Data Placement Scheme for Data"*Intensive Applications With Interest Locality*" IEEE Transactions (Vol: 49), 2013, 2514 – 2520
17. HADOOP-3759: Provide ability to run memory intensive jobs without affecting other running tasks on the nodes
18. Tejaswini U. Mane, Mrs. Asha M. Pawar, "*A Survey On Big Data And Its Mining Algorithm*", IJIRCCE, Vol. 3, Issue 12, December 2015.

BIOGRAPHY

Ms. Tejaswini U. Mane: She is Student of M.E. Computer, in Zeal College of Engineering and Research, Narhe, Pune.

Mrs. Asha M. Pawar: she is Asst.Professor in Computer department of Zeal college of Engineering and Research, Narhe, Pune, of Savitribai Phule Pune University. She has done M.E (CSE) from belgaon. And also published various papers.