

A Comprehensive Study of Map Reducing Technique in Big Data

Mr. ZULFIKAR ALI ANSARI
M.Tech Scholar
Department of CSE
Integral University Lucknow

Dr. Manish Madhav Tripathi
Associate Professor
Department of CSE
Integral University Lucknow

Abstract: Now a days our Generation are living with Social media Data like Heart beat, A lot of data are generating from Social media site like Facebook, Twitter, Linkdin, Youtube along with many other sources of data like CCTV data, Medical data. MRI Data Xray data, Shopping site etc. These data may be structured, unstructured and Semi structured. It is not possible of these data handling and managing in Traditional method smoothly. All these data are called Big Data.

Big data is defined as large amount of data which requires new technologies and architectures so that it becomes possible to extract value from it by capturing and analysis process. Due to such large size of data it becomes very difficult to perform effective analysis using the existing traditional techniques. To handling this problem Hadoop comes in Picture.

Hadoop is a framework, Hadoop is a scalable, open source, fault-tolerant Virtual Grid operating system architecture for data storage which is fault-tolerant high-bandwidth clustered storage architecture. Hadoop is an open-source framework that allows to store and process big data in a distributed environment across Thousands clusters of computers using simple programming models. Hadoop is fully dependent in a programming model or algorithm it is called Map Reduce Technique, It is also called the Heart of Hadoop Technology.

Map Reduce Technique is working Parallel Processing infrastructure that runs in contact with HDFS. MapReduce is a processing technique and a program model for distributed computing based on java. Map Reducing Algorithm mainly depend on 3 Phase Map, Reduce and Shuffle. Map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). Secondly, reduce task, which takes the output from a map as an input and combines those data tuples into a smaller set of tuples. As the sequence of the name MapReduce implies, the reduce task is always performed after the map job and also applied another phase like Shuffle it is also must in Map Reduce Technique. Map Reducing Technique work in Chunk Form or we can say it is working thousands of machine at a time on Commodity hardware, It Contains Record Reader also it should be separate a boxes data at word count job. Record Reader is working as Interface in Map Reduce.

Map Reducing is provide the functionality which is based on java Technology that's why it is provide smoothly access Big Data. Structured, Unstructured data are storing and managing very easy from Map Reducing technique.

This Paper proposed implementation of Map reducing Technology like Big Data Analysis very easily Accessing, Storing, Securing and Management. This research focusing on implementation of Map reducing Technology another field like Medical field, Shopping site, Weather climate field etc. In medical Field we can access the heart patient, cancer Patient, or any other particular patient

data fetch from Big data analysis technique using map reducing technology easily. No SQL Server use in Map Reducing Algorithm, It is used Hadoop Distributed file System(HDFS)for storing and accessing data. Hadoop contains 64 MB each block which is very effective distributed file system.

Keywords— Big Data, Hadoop, HDFS, Combiner Mapper , Reducer.

1.0 INTRODUCTION

Today's the quantity of data being generated on a daily basis is quite alarming. For Example Take data from 2018 info Graphics show that every minute, Like Facebook users share nearly 5.5 Million pieces of content, Twitter users tweet nearly 700,000 times, Instagram users post nearly 540,000 new photos every minute. YouTube users upload 10 hours of new video content, Apple users download nearly 120,000 apps. Email users send over 550 million messages. Amazon generates over \$150,000 in online sales. Also, with the internet of things, the amount of data generated by car sensors, railway sensors, CCTV, GPS devices and other devices is also at a very high level Volume. Various Medical and Research Institutions also tend to gather an enormous amount of data each day. In fact, according to IBM, every day we create an average of 3.5 quintillion bytes of data (equivalent to 3.3 trillion gigabytes). Wow! These large amount of data come in various formats ranging from images, videos, JSON strings, Audios, PDFs, WebPages, Text documents even to 3D models. Data of this nature is known as Big Data.

We can say Big data is the collection of huge amount of data sets and it becomes very complex to access using suitable traditional data processing applications and Database management tools, Faced many challenges like Acquiring, Storing, Searching, Sharing, Transferring, Accessing, Analysing, and Visualizing. The big trend of Big Data is due to other needful information which can be derivable from analysis of large data set of related data, providing correlations to be made to spot business trends, Heart disease, determine quality of research, link legal citations, contest crime and find actual roadways traffic conditions.

Big Data is used in many real world areas like Health care, Telecommunication, Financial Business, Pharmaceutical and Weather climate etc. Accessing critical business insight by querying and analyzing such a very big amounts of data is suitable the need of the hour and become a challenge to the standard data taking out approach. This data comes from a broad range of sources like sensors, digital pictures, videos, purchase transaction,

social media post, CCTV data everywhere. To Analysis and accessing the development of information technologies, organizations have had to face new challenges to analyze huge amount of information. Big Data conceptually defined from 7 V's, This concept is given by IBM.

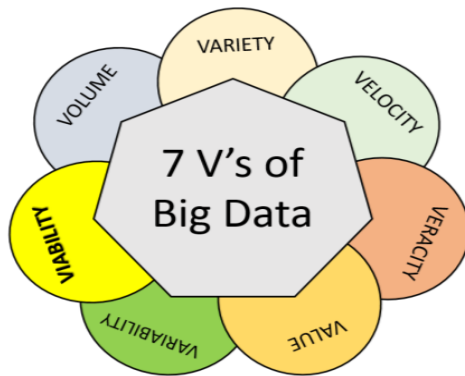


Fig (1.1) 7V's concept

Volume-Volume is refer to the huge amount of data like petabyte, zeta byte etc. needs to be processed , stored and analyzed also.

Velocity- It is an indication of how fast accessing the data needs to be analyzed so that it can provide an informed response.

Variety-The Variety is related to the types of data like structured, unstructured and semi structured data that organizations can accumulate such as tabular data, hierarchal data, images data, videos data, Text data, email data like log file and many more.

Veracity- Veracity refers to the biases of noise and abnormally in data. In Scooping out our Big Data strategy we need to work to keep our data clean and processes to keep away 'dirty data' from accumulating in our systems.

Variability- Variability causing the peaks in the data load and takes into account the data which is rapidly engulfing the system. This cause a wave effect which requires the data model and association to altered to get inferred from the growing data stock.

Value- Values of data in order to predict the behavior of its customers to formulate a policy or make a decision for maximizing the profits.

Viability- Data is multidimensional by nature. Depending on the outcome all the dimensions are not relevant during the processing, moreover considering all the dimensions of data makes the process inefficient in terms of both space and time. Thus there is a need to analyze massive data sets in real time with efficiency.

2.0 AIM AND OBJECTIVES

The aim of this study is to discover how Map Reduce technology solves the problem of Big Data Processing which is not able to efficiently solve by traditional methods and also focus on Big Data, HDFS and Hadoop Framework

Now a days the privacy and security is big issue in cloud data, Also we will discuss the very effective challenges of Big data Privacy and security, we also focus on how the improve security and privacy of the huge amount data

Privacy & Security It is the most important issue with Bigdata which is sensitive and includes conceptual, Technical as well as legal significance. The personal information of a person when combined with external large data sets leads to the interface of new facts about that person and it's possible that these kinds of facts about the person are secretive and the person might not want the data owner to know or any person to know about them.

Fault Tolerance: With the incoming of new technologies like cloud computing and Big data is always intended that whenever the failure occurs the damage done should be within acceptable threshold rather than beginning the whole task from the scratch. Fault-tolerant computing is extremely hard, involving intricate algorithms.

Scalability: The processor technology has changed in recent years. The scalability issue of Big data has lead towards cloud computing, which now aggregates multiple disparate workloads with varying performance goals into very large clusters. This requires a high level of sharing of resources which is expensive and also brings with it various jobs so that we can meet the goal of each workload cost effectively.

Quality of data: The collection of huge amount of data and its storage comes at a cost. More data if used decision making or for predictive analysis in business will definitely lead to better results. Big data basically focuses on quality data storage rather than having very large irrelevant data so that better results and conclusions can be drawn.

Heterogeneous Data: Unstructured data represents almost every kind of data being generate like social media interaction, to record meetings, to handling of PDF documents, fax transfers, to email and more. Structured data is always organized into highly mechanized and systematic way. It shows well manageable with database but unstructured data is completely raw and unorganized. Covering all this unstructured data into structured one is not feasible.

3.0 METHODOLOGY

3.1 Hadoop

Hadoop is an Apache open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models. A Hadoop frame-worked application works in an environment that provides distributed storage and computation across clusters of computers. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

Hadoop is a scalable, open source, fault-tolerant Virtual Grid operating system architecture for data storage which is fault-tolerant high-bandwidth clustered storage architecture. It runs MapReduce for distributed data processing and is works with structured and unstructured data[4]

Working Strategy in Hadoop

A User can submit the job to the hadoop for required process by identifying following steps-
The location of the 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.

The Hadoop client then submits the job like jar/executable file, and configuration to the job tracker which then assumes the responsibility of distributing the software / configuration to the slaves, scheduling task and monitoring them, providing status and diagnostic information to the job client. The TaskTracker on different nodes execute the task as per map reduce implementation and output the reduce function is stored into the output files on the file system.

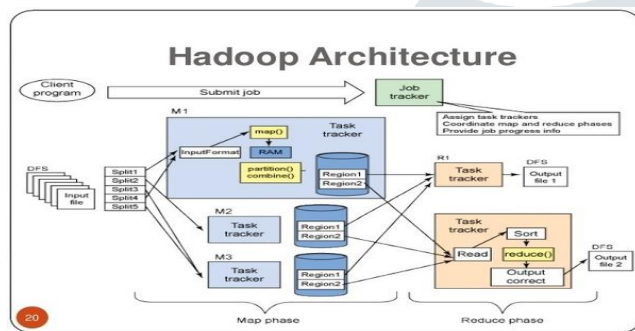


Fig (3.1) Hadoop Architecture

3.2 HDFS Architecture

Hadoop file system using Hadoop Distributed File System(HDFS). It is run on Commodity hardware. HDFS is very Haighly Fault Tolerant and design using low cost hardware. HDFS holds very large amount of data and provides efficient access. To store such huge amount of data, the files are stored across multiple nodes. These files are stored in redundant fashion to rescue the system from possible data losses in case of failure. HDFS also makes applications available to parallel processing.

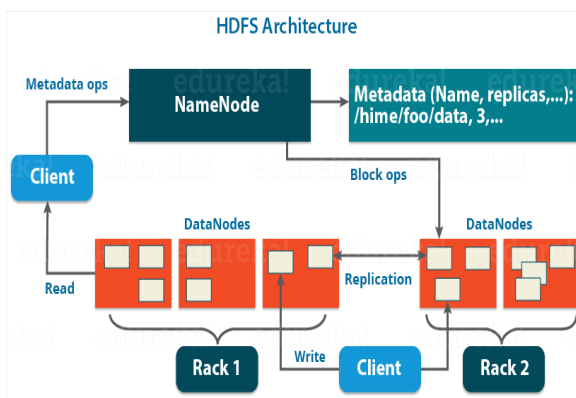


Fig (3.2) HDFS Architecture

3.3 Storing Process of HDFS

Assumptions:

Consider a file of 200 MB and file name as file.txt. Actual file size is in GB/TB/PB/ZB.For understanding purpose

consider small sized file.Block size=64 MB. As block size is 64 MB, Number of blocks=4.Out of 4,3 blocks are of 64 MB and 4th one will be about 8 MB.Let S1,S2,S3,S4 are the 4 blocks . As shown below

WordCountJob :

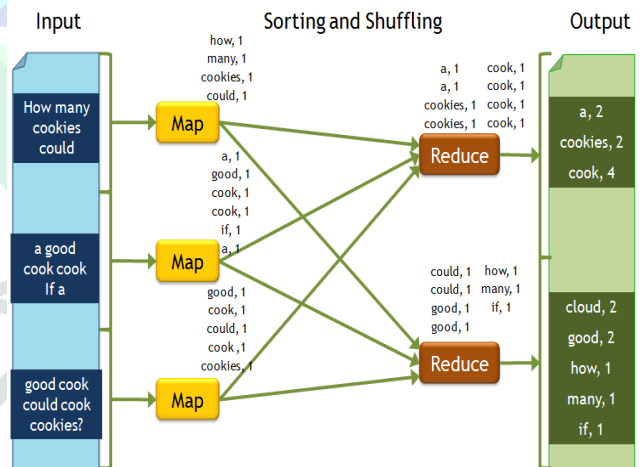
file.txt (200MB)

hi how are you how is your job	64MB
how is your family how is your brother	64MB
how is your sister what is the time now	64MB
what is the strength of hadoop	8MB

Fig (3.3) Storing Process of HDFS

3.4 Map Reduce Technology

The Map Reduce is a heart of Hadoop framework and Hadoop is a platform of Big data so they are fully dependent on Map Reduce Technology . Map Reduce is a processing technique and a program model for distributed computing based on java Technology. MapReduce consist of Three phases the Map phase, Reduce phase and Shuffle & Sort Phase.



By Manaranjan Pradhan

Fig(3.4) Sorting & Shuffling

A Map Reduce job usually splits the input data set into independent blocks of data which are assigned to Map tasks in a completely parallel manner. The output of the map is often stored and given as input to the reduce tasks to produce the final results.HadoopMapReduce works exclusively on <key, value> pairs. It views the input to the job as <key, value> pairs and produces its output from the job as a set of <key, value> pairs. It makes use of an interface known as the Writable interface to serialize the Key and Value classes. These classes have to implement the interface. Also, the key class has to implement the Writable Comparable interface in order to facilitate sorting by the framework. The input and output of the Map Reduce jobs can be depicted as

(Input) $\langle k1, v1 \rangle$ map $\langle k2, v2 \rangle \langle k2, v2 \rangle$ reduce $\langle k3, v3 \rangle$ (output). Applications which implement MapReduce usually implement the Mapper and Reducer Interfaces which provide several methods for different tasks.

Input Phase- In This phase we have a Record Reader that translates each record in an

input file and sends the parsed data to the mapper in the form of key-value pairs.

Map Phase- This phase is a user-defined function, which takes a series of key-value pairs and processes each one of them to generate zero or more key-value pairs.

Intermediate Keys- Which key-value pairs generated by the mapper are known as intermediate keys.

Shuffle and Sort Phase-The Reducer task starts with the Shuffle and Sort step. It downloads the grouped key-value pairs onto the local machine, where the Reducer is running. The individual key-value pairs are sorted by key into a larger data list. The data list groups the equivalent keys together so that their values can be iterated easily in the Reducer task.

Reducer Phase - The Reducer takes the grouped key-value paired data as input and runs a Reducer function on each one of them. Here, the data can be aggregated, filtered, and combined in a number of ways, and it requires a wide range of processing. Once the execution is over, it gives zero or more key-value pairs to the final step.

Output Phase- In the output phase, we have an output formatter that translates the final key-value pairs from the Reducer function and writes them onto a file using a record writer.

enable more broad-based health-care solutions. In most of the cases reported, we found multiple technologies that were used together, such as artificial intelligence (AI), along with Hadoop and data mining tools.

Bioinformatics research analyzes biological system variations at the molecular level. With current trends in personalized medicine, there is an increasing need to produce, store, and analyze these massive datasets in a manageable time frame. Next-generation sequencing technology enables genomic data acquisition in a short period of time. The role of big data techniques in bioinformatics applications is to provide data repositories, computing infrastructure, and efficient data manipulation tools for investigators to gather and analyze biological information. Taylor discusses that Hadoop and MapReduce are currently used extensively within the biomedical field.

In foreign countries like U.S. most of the death happens due to Chronic disease. Some of the chronic diseases are Allergy, Cancer, Asthma, Heart disease, Glaucoma, Obesity, viral diseases such as Hepatitis C and HIV/AIDS. Of all the diseases Diabetes is the most hazardous disease. Diabetes means that blood glucose (blood sugar is too high. The analysis on the data is performed using Big Data Analytics framework Hadoop.Hadoop framework is used to process large data sets. The analysis is done using map reduce algorithm[20].

Heart disease is a major health problem and it is the leading causes of death throughout the world. Early detection of heart disease has become an important issue in the medical research fields. For heart disease detection, some features are analyzed such as RR interval, QRS interval and QT interval. The classification process states whether the patient is normal or abnormal and in the detection step using map reduce technique to detect the disease and reduce the dataset. Thus, the proposed system helps to classify a large and complex medical dataset and detect the heart disease[24].

4.2 Online Marketing

Big data is very important role play in Online Marketing strategy, that's why it is widely used now a days. Big data is fully dependent on MapReduce Technology which are working on parallel programming model.

Currently Google is mainly used MapReduceTechnology for searching and finding many relevant data, also using for managing and storing data from this technology used. From MapReduce Technology Google find relevant data which are needed for any marketing strategy along with location wise.

Yahoo is also used Big Data and it is first search engine for using Map Reduce Technology. Yahoo is also provide a good platform for online marketing from using MapReduce Technology. It is also find the expressive and relevant data for managing and accessing for Digital marketing platform, that's why it is also very useful for Online Marketing.

Now a days Digital Marketing is very important role play for enhancing business, Facebook provide a very interesting platform for various business strategy, Facebook is also using Big data along with MapReduce Technology which is working on various location wise data finding, accessing, managing, and storing for good business purposes. Facebook is finding the data location wise and gender wise and also provide age group wise from using MapReducing Technology.

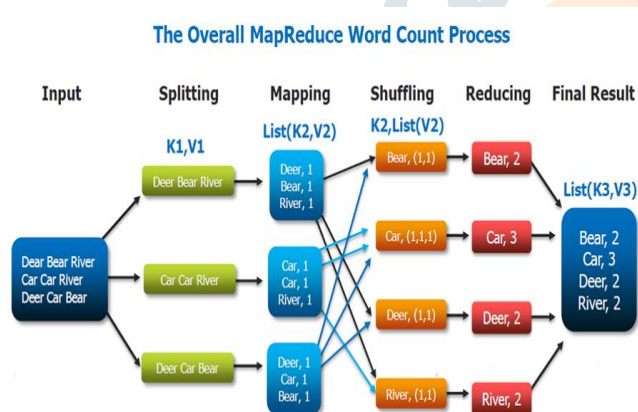


Fig (3.5 Map Reduce word count process)

4.0 APPLICATIONS OF MAP REDUCE TECHNIQUE

HadoopMapReduce is applicable in all areas of life where data is being generated. This is because the data being generated will certainly need to be stored and processed. Now a days we define applications of mapReduce in many fields like-

4.1 Medical Field

In the Biomedical informatics domain, The MapReduce Technology is a new paradigm and an ecosystem that transforms case based studies to large scale, data driven research. Biomedical scientists are facing new challenges of storing, managing, and analyzing huge amounts of datasets. The Features of Map Reducing data require powerful and novel technologies to extract useful information and

Amazon is an online shopping website which is also used MapReduce Technology data sets. From MapReduce Technology Amazon finds the affordable customer data and also provides the notifications of the relevant data which are very powerful tools for enhancing the good business and marketing. Now a days Amazon is a Top website for online shopping. Like Amazon eBay also used the MapReduce Technology for enhancing the business strategy.

4.3 Scale Weather Radar Data in Real-time

The MapReduce approach is massively scalable, and is able to create high-resolution 3D radar mosaics over the Continental United States in real-time. There are advantages

to the MapReduce approach beyond its massive scalability – it is also the most efficient way of parallelizing the creation of a 3D radar mosaic. In other words, if a domain is so large that the mosaic can not be created on a single machine, the MapReduce approach requires less hardware than either the SRC or intelligent agent approaches[16].

The MapReduce approach, therefore, permits for the creation of radar mosaics without compromises on the quality, timeliness or resolution of the mosaic.

4.4 Online Fraud Detection and Prevention

Online fraud detection is based on the principle that every user exhibits a particular pattern of behavior in his online transactions. The patterns are created as per his past online purchases and it may vary from user to the user, i.e. the patterns are unique for every online user or customer. These patterns are generated from the existing online transaction data store, based on the features of the online transaction of the customers.

Map Reduce Technology can be implemented to prevent the online fraud transactions before it happens. This novel framework uses both big data analytics and machine learning concepts for the best performance. The online transaction data are collected, cleaned and features are extracted for each user. These features are reduced using the principal component analysis technique and trained using the self-organizing maps[6]

5.0 BENEFITS OF MAPREDUCING TECHNIQUE

Cost-proficiency - Map Reduce requires ware equipment to work. Likewise, its adaptation to non-critical failure is programmed. Henceforth, less administrators are required on the system.

Effortlessness – Developers proposing to actualize Map Reduce can compose applications with their dialect of decision, for example, Java, C++ or Python. Additionally, MapReduce occupations are anything but difficult to run.

Versatility – Map Reduce can process pet bytes of information which are put away in Hadoop Distributed File System (HDFS) in one group.

Speed – With parallel preparing, Map Reduce can take issues that used to take days to comprehend and understand them in hours or minutes. For example, in July 2008, one of Yahoo's Hadoop groups arranged 1 terabyte

of information in 209 seconds, which beat the past record of 297 seconds in the yearly universally useful (Daytona) terabyte sort benchmark.

Recuperation – Map Reduce handles disappointments. Because of repetition in HDFS, if a machine with one duplicate of the information is inaccessible, another machine has a duplicate of a similar key/esteem combine, which can be utilized to illuminate a similar sub-assignment. The Job Tracker monitors everything.

Negligible Data Motion – With Map Reduce, forms are moved to information and not information to forms. Preparing typically happens on the same physical hub where the information lives. This is known as information territory. This decreases the system input/output examples and adds to the handling speed.

6.0 LIMITATIONS OF HADOOP MAPREDUCE

Despite the amazing benefits of Map Reducing Technique there are some limitations to it. Some are listed below:

The development of efficient Map Reduce applications requires advanced programming skills and also a deep understanding of the architecture of the system. Analysts who are used to SQL-like or declarative languages may view Map Reduce programming model as too “low-level” because MapReduce doesn't require implementing relational operations such as joins.

MapReduce has a batch nature. Data needs to be uploaded to the file system always and even when the same dataset needs to be analyzed multiple times, it still needs to be read every time.

The master node (where the JobTracker is) has to be more sophisticated than other nodes because it can easily become a single point of failure.

Since Google has been granted patent, it raises the question of the long-term viability of using the mechanism in open environments.

The “One Way Scalability” of its design. Map Reduce allows a program to scale up to process very large datasets but constrains a program's ability to process smaller data items.

7.0 CONCLUSION

Map Reducing Technique is a programming model which processes huge amount of structured, Unstructured and semi structured data files across the massive data sets. It splits the processing into small chunk of work, these splits processes can be executed in parallel across several nodes. It takes very less time to accessing and managing huge amount of data sets as compare to other traditional technique.

In This paper we mainly focused on Map reducing Technique in Big data and its efficiency and performance to reduced the accessing time which is directly related to the future work because the data generation is very fast.

In This paper we cover the using map reducing technique in other field also like it provide Online fraud detection, and how to detect heart disease patient from using Map reducing technology. Also covered the many applications of Map Reducing Technique like efficiently handling Big Data which is very needful at the current time.

In This paper we are also trying to focus of advantages of map Reduce Technology along with its Limitations now a days. Also taken out classified and its improvements. Finally we can say the Map Reduce Technology is simple

but provides good scalability and fault tolerant for massive data processing on a large commodity hardware machines.

7.1 FUTURE WORK

Now a days Map Reduce Technique using many fields like Google is also using, this technique is inspired by Google File System(GFS), facebook, Amazon, Twitter or we can say it is using almost all respective field which is generating huge amount of data, also covered many field in this paper. After concluding all literature review we can say the Map Reduce Technology still need to be addressed for a successful implementations.

In Future we can improve more efficiency for Map reducing technique, because the data generation is increasing very fast, and also we should focus on the commodity hardware.

Also conclude from this paper we can say the different implementations of Map Reduce Technology are evaluated using different datasets and respective applications and there is lack of standard benchmark or workload for comparing other implementations and extensions. In Future needs to define what a typical Map Reduce workload should be.

REFERENCES

- [1]SubramaniaswamyVa, VijayakumarVb, LogeshRc and IndragandhiVd “Unstructured Data Analysis on Big Data using Map Reduce” 2nd International Symposium on Big Data and Cloud Computing (ISBCC’15)-2015.
- [2] Rama Satish K. V., Dr. N. P. Kavya “Big Data Processing with harnessing Hadoop - MapReduce for Optimizing Analytical Workloads” 978-1-4799-6629-5/14/\$31.00c 2014 IEEE-2014.
- [3] Jeffrey Dean and Sanjay Ghemawat “MapReduce: Simplified Data Processing on Large Clusters”To appear in OSDI 2004.
- [4]M. Dhavapriya, N. Yasodha “Big Data Analytics: Challenges and Solutions Using Hadoop, Map Reduce and Big Table” International Journal of Computer Science Trends and Technology (IJCST) – Volume 4 Issue 1, Jan - Feb 2016.
- [5]S.Ancy, M.Maheswari “Locality Based Data Partitioning in Map Reduce” International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - 2016
- [6]Balasupramanian.N,Ben George Ephrem,ImadSalim Al-Barwani “User Pattern Based Online Fraud Detection and Prevention using Big Data Analytics and Self Organizing Maps” 2017 International Conference on Intelligent Computing,Instrumentation and Control Technologies (ICICT)-2017
- [7]Punam Bedi1, Vinita Jindal2, Anjali Gautam3 “Beginning with Big Data Simplified” 978-1-4799-4674-7/14/\$31.00 ©2014 IEEE.
- [8]SubhankarDhar, SouravMazumdar “Challenges and Best Practices for Enterprise Adoption of Big Data Technologies” 978-1-4799-3312-9/14/\$31.00 ©2014 IEEE.
- [9]VaibhavFanibhare and Vijay Dahake “SmartGrids: MapReduce Framework using Hadoop” 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN-2016
- [10] Mr. Zulfikar Ali Ansari1, Mr.Afsaruddin2 “Different Data Analysis Technique Used in Big Data” ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962-2018.
- [11]PoojaBatra Nagpal1, Sarika Chaudhary2, PreetishreePatnaik “Emerging Clustering Techniques on Big data” GE-INTERNATIONAL JOURNAL OF ENGINEERING RESEARCHVOLUME -3, ISSUE-6 (June 2015) IF-4.007 ISSN: (2321-1717-2015.
- [12]Anurag Sarkar1, Abir Ghosh2 “MapReduce: A Comprehensive Study on Applications, Scope and Challenges” International Journal of Advance Research in Computer Science and Management Studies -2015.
- [13]C.V.S.Adithya, N.VijayaSai “Weather Analysis Using Map Reducing Technique” International Journal of Civil Engineering and Technology (IJCIET) Volume 8, Issue 4, 2017.
- [14]Trupti V. Kenekar, Prof.A.R.Dani “An Efficient Private FIM OnHadoopMapReduce” 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT).
- [15]SeemaMaitrey, C.K. Jha “Handling Big Data Efficiently by using Map Reduce Technique” 2015 IEEE International Conference on Computational Intelligence & Communication Technology-2013.
- [16]Valliappa Lakshmanan1;2, Timothy W. Humphrey “A MapReduce Technique to Mosaic Continental-Scale Weather Radar Data in Real-time” IEEE JOURNAL OF SELECT TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING-2016.
- [17] A. Munar, E. Chiner, I. Sales “A Big Data Financial Information Management Architecture for Global Banking” 2014 International Conference on Future Internet of Things and Cloud-2014.
- [18] Maria ViorelaMuntean, Raul Boldea “Intelligent Agents based Improved Map-Reduce Method for Multiple Databases” ECAI 2016 - International Conference – 8th Edition-2016.
- [19]Vidya V. Pol, Prof.S.M.Patil “Implementation of on-process aggregation for Efficient Big Data Processing in HadoopMapReduce Environment” International Conference on Inventive Computation Technologies(ICICT)/IEEE-2016.
- [20]Mr.N.Ramkumar, Dr.S.Prakash, Mrs. K Sangeetha “Data Analysis for Chronic disease –Diabetes using Map Reduce Technique” 2016 International Conference on Computer Communication and Informatics (ICCCI - 2016), Jan. 07 – 09, 2016, Coimbatore, INDIA-2016.
- [21]Pulkit Sharma, KomalMahajan, Dr. Vishal Bhatnagar “Analyzing Click stream Data using Hadoop” 2016 Second International Conference on Computational Intelligence & Communication Technology-2016.
- [22]PujasumanTripathy,SiddharthSwarupRautaray,Manju shaPandey “Parallel Support Vector Machine used in Map-Reduce For Risk analysi” Copyright.978-1-5090-3239-6/17/\$31.00©2017IEEE-2017.
- [23] K. Ramana, A. Venkataramana “International Journal of Advanced Research in Computer Science and Software Engineering” International Journal of Advanced Research in Computer Science and Software Engineering-2015.
- [24] G.Vaishali1, V.Kalaivani2 “Big Data Analysis Heart Disease Detection System Using map Reduce Technique” International Conference on computing Technologies and Intelligent Data Engineering(ICCTDE'16)-2016.