VOLUME NO. 5 (2015), ISSUE NO. 10 (OCTOBER) *ISSN 2231-1009*

INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

IJRCM

A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

Indexed & Listed at:

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.,
Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)],
Index Copernicus Publishers Panel, Poland with IC Value of 5.09 & number of libraries all around the world.
Circulated all over the world & Google has verified that scholars of more than **4456 Cities** in **177 countries/territories** are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

http://ijrcm.org.in/

CONTENTS

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the areas of Computer Science & Applications; Commerce; Business; Finance; Marketing; Human Resource Management; General Management; Banking; Economics; Tourism Administration & Management; Education; Law; Library & Information Science; Defence & Strategic Studies; Electronic Science; Corporate Governance; Industrial Relations; and emerging paradigms in allied subjects like Accounting; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Rural Economics; Co-operation; Demography: Development Planning; Development Studies; Applied Economics; Development Economics; Business Economics; Monetary Policy; Public Policy Economics; Real Estate; Regional Economics; Political Science; Continuing Education; Labour Welfare; Philosophy; Psychology; Sociology; Tax Accounting; Advertising & Promotion Management; Management Information Systems (MIS); Business Law; Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labour Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; International Relations; Human Rights & Duties; Public Administration; Population Studies; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism & Hospitality; Transportation Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic; Web Design and emerging paradigms in allied subjects.

Anybody can submit the **soft copy** of unpublished novel; original; empirical and high quality **research work/manuscript** ***anytime*** in ***M.S. Word format*** after preparing the same as per our **GUIDELINES FOR SUBMISSION**; at our email address i.e. infoijrcm@gmail.com or online by clicking the link **online submission** as given on our website (***FOR ONLINE SUBMISSION, CLICK HERE***).

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

DATED: _____________

THE EDITOR

IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF ______________________.**

(e.g. Finance/Mkt./HRM/General Mgt./Engineering/Economics/Computer/IT/ Education/Psychology/Law/Math/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled '____' for possible publication in one of your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the co-authors of this manuscript have seen the submitted version of the manuscript and have agreed to their inclusion of names as co-authors.

Also, if my/our manuscript is accepted, I agree to comply with the formalities as given on the website of the journal. The Journal has discretion to publish our contribution in any of its journals.

NAME OF CORRESPONDING AUTHOR :

Designation :

Institution/College/University with full address & Pin Code :

Residential address with Pin Code :

Mobile Number (s) with country ISD code :

Is WhatsApp or Viber active on your above noted Mobile Number (Yes/No) :

Landline Number (s) with country ISD code :

E-mail Address :

Alternate E-mail Address :

Nationality :

NOTES:

a) The whole manuscript has to be in ***ONE MS WORD FILE*** only, which will start from the covering letter, inside the manuscript. ***pdf. version is liable to be rejected without any consideration***.

b) The sender is required to mention the following in the **SUBJECT COLUMN of the mail**:

New Manuscript for Review in the area of (e.g. Finance/Marketing/HRM/General Mgt./Engineering/Economics/Computer/IT/ Education/Psychology/Law/Math/other, please specify)

c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any **specific message** w.r.t. to the manuscript.

d) The total size of the file containing the manuscript is expected to be below **1000 KB**.

e) **Abstract alone will not be considered for review** and the author is required to submit the **complete manuscript** in the first instance.

f) ***The journal gives acknowledgement w.r.t. the receipt of every email within twenty four hours*** and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending a separate mail to the journal.

g) The author (s) name or details should not appear anywhere on the body of the manuscript, except the covering letter and the cover page of the manuscript, in the manner as mentioned in the guidelines.

2. **MANUSCRIPT TITLE:** The title of the paper should be **bold typed**, **centered** and **fully capitalised**.

3. **AUTHOR NAME (S) & AFFILIATIONS:** Author (s) **name**, **designation**, **affiliation** (s), **address**, **mobile/landline number** (s), and **email/alternate email address** should be given underneath the title.

4. **ACKNOWLEDGMENTS:** Acknowledgements can be given to reviewers, guides, funding institutions, etc., if any.

5. **ABSTRACT:** Abstract should be in **fully italicized text**, ranging between **150** to **300 words**. The abstract must be informative and explain the background, aims, methods, results & conclusion in a **SINGLE PARA**. ***Abbreviations must be mentioned in full***.

6. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of **five**. These should be arranged in alphabetic order separated by commas and full stop at the end. All words of the keywords, including the first one should be in small letters, except special words e.g. name of the Countries, abbreviations.

7. **JEL CODE:** Provide the appropriate Journal of Economic Literature Classification System code (s). JEL codes are available at www.aeaweb.org/econlit/jelCodes.php, however, mentioning JEL Code is not mandatory.

8. **MANUSCRIPT:** Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. *It should be free from any errors i.e.* ***grammatical, spelling*** *or* ***punctuation.*** *It must be thoroughly edited at your end*.

9. **HEADINGS:** All the headings must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

10. **SUB-HEADINGS:** All the sub-headings must be bold-faced, aligned left and fully capitalised.

11. **MAIN TEXT:**

THE MAIN TEXT SHOULD FOLLOW THE FOLLOWING SEQUENCE:

INTRODUCTION

REVIEW OF LITERATURE

NEED/IMPORTANCE OF THE STUDY

STATEMENT OF THE PROBLEM

OBJECTIVES

HYPOTHESIS (ES)

RESEARCH METHODOLOGY

RESULTS & DISCUSSION

FINDINGS

RECOMMENDATIONS/SUGGESTIONS

CONCLUSIONS

LIMITATIONS

SCOPE FOR FURTHER RESEARCH

REFERENCES

APPENDIX/ANNEXURE

The manuscript should preferably range from ***2000 to 5000 WORDS***.

12. **FIGURES & TABLES:** These should be simple, crystal **CLEAR**, **centered**, **separately numbered** & self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. *It should be ensured that the tables/figures are referred to from the main text*.

13. **EQUATIONS/FORMULAE**: These should be consecutively numbered in parenthesis, horizontally centered with equation/formulae number placed at the right. The equation editor provided with standard versions of Microsoft Word should be utilised. If any other equation editor is utilised, author must confirm that these equations may be viewed and edited in versions of Microsoft Office that does not have the editor.

14. **ACRONYMS**: These should not be used in the abstract. The use of acronyms is elsewhere is acceptable. Acronyms should be defined on its first use in each section: Reserve Bank of India (RBI). Acronyms should be redefined on first use in subsequent sections.

15. **REFERENCES**: The list of all references should be alphabetically arranged. ***The author (s) should mention only the actually utilised references in the preparation of manuscript*** and they are supposed to follow Harvard Style of Referencing. **Also check to make sure that everything that you are including in the reference section is duly cited in the paper**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parenthesis.
- ***Headers, footers, endnotes** and **footnotes** should **not be used** in the document.* However, **you can mention short notes to elucidate some specific point**, which may be placed in number orders after the references.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–23

UNPUBLISHED DISSERTATIONS

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITES

- Garg, Bhavet (2011): Towards a New Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

BIG DATA USING HADOOP MAP REDUCE

TAYYABA HASHMI
STUDENT
ME INFORMATION TECHNOLOGY
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
CHEMBUR

PRAMILA SHINDE
ASST. PROFESSOR
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
CHEMBUR

ABSTRACT

Big data is an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process them using traditional data processing applications. Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big data size is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data. Big data is a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale. Big data can also be defined as "Big data is a large volume unstructured data which cannot be handled by standard database management systems like DBMS, RDBMS or ORDBMS".

KEYWORDS

big data analysis, HDFS, map reduce.

1. INTRODUCTION

1.1 OVERVIEW

Big Data may well be the Next Big Thing in the IT world. Big data burst upon the scene in the first decade of the 21st century. The first organizations to embrace it were online and startup firms. Firms like Google, eBay, LinkedIn, and Facebook were built around big data from the beginning. Like many new information technologies, big data can bring about dramatic cost reductions, substantial improvements in the time required to perform a computing task, or new product and service offerings.

Big Data' is similar to 'small data', but bigger in size but having data bigger it requires different approaches. It is an aim to solve new problems or old problems in a better way. Big Data generates value from the storage and processing of very large quantities of digital information that cannot be analyzed with traditional computing techniques. Due to its specific nature of Big Data, it is stored in distributed file system architectures[1].

Scientists regularly encounter limitations due to large data sets in many areas, including meteorology, genomics, connectomics, complex physics simulations, biological and environmental research, and in e-Science in general. The limitations also affect Internet search, finance and business informatics. Data sets grow in size in part because they are increasingly being gathered by ubiquitous information-sensing mobile devices, aerial sensory technologies (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers, and wireless sensor networks.

Big data is difficult to work with using most relational database management systems and desktop statistics and visualization packages, requiring instead "massively parallel software running on tens, hundreds, or even thousands of servers". What is considered "big data" varies depending on the capabilities of the organization managing the set, and on the capabilities of the applications that are traditionally used to process and analyze the data set in its domain.

Big Data is a moving target; what is considered to be "Big" today will not be so years ahead. "For some organizations, facing hundreds of gigabytes of data for the first time may trigger a need to reconsider data management options. For others, it may take tens or hundreds of terabytes before data size becomes a significant consideration."

FIG. 1: BIG DATA SCENARIO

1.2 GENESIS OF BIG DATA

In a 2001 research report and related lectures, META Group (now Gartner) analyst Doug Laney defined data growth challenges and opportunities as being three-dimensional, i.e. increasing volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources). Gartner, and now much of the industry, continue to use this "3Vs" model for describing big data. In 2012, Gartner updated its definition as follows: "Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization." Additionally, a new V "Veracity" is added by some organizations to describe it. One of the fundamental characteristics of the Big Data is the huge volume of data represented by heterogeneous and diverse dimensionalities[2].

If Gartner's definition (the 3Vs) is still widely used, the growing maturity of the concept fosters a more sound difference between big data and Business Intelligence, regarding data and their use:

- Business Intelligence uses descriptive statistics with data with high information density to measure things, detect trends etc.
- Big data uses inductive statistics and concepts from nonlinear system identification to infer laws (regressions, nonlinear relationships, and causal effects) from large sets of data with low information density to reveal relationships, dependencies and perform predictions of outcomes and behaviors.

2. REVIEW OF LITERATURE

2.1 THREE KEY TECHNOLOGIES FOR EXTRACTING BUSINESS VALUE FROM BIG DATA

Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery and/or analysis. Furthermore, this analysis is needed in real time or near-real time, and it must be affordable, secure and achievable. NASA uses advanced information systems to mature big data capacity in order to support future Earth observation missions[3]. Here are three key technologies that can help you get a handle on big data – and even more importantly, extract meaningful business value from it.

- Information management for big data. Manage data as a strategic, core asset, with ongoing process control for big data analytics.
- High-performance analytics for big data. Gain rapid insights from big data and the ability to solve increasingly complex problems using more data.
- Flexible deployment options for big data. Choose between options for on premises or hosted, software-as-a-service (SaaS) approaches for big data and big data analytics.

2.2 INFORMATION MANAGEMENT FOR BIG DATA

Since recent developments (in technology, science, user habits, businesses, etc.) gave rise to production and storage of massive amounts of data, not surprisingly, the intelligent analysis of big data has become more important for both businesses and academics[4]. SAS provides the management and governance capabilities that enable organizations to effectively manage the entire life cycle of big data analytics, from data to decision. SAS provides a variety of these solutions, including data governance, metadata management, analytical model management, run-time management and deployment management. With SAS, this governance is an ongoing process, not just a one-time project. Proven methodology-driven approaches help organizations build processes based on their specific data maturity model.

2.3 HIGH-PERFORMANCE ANALYTICS FOR BIG DATA

High-performance analytics makes it possible to analyze all available data (not just a subset of it) to get precise answers for hard-to-solve problems and uncover new growth opportunities and manage unknown risks – all while using IT resources more effectively. System elasticity or hardware heterogeneity and multigrain parallelism cannot be taken into account easily: instead, a new programming model would have to be chosen, and the development process restarted[5]. Accelerated processing of huge data sets is made possible by following technologies:

- Grid computing: A centrally managed grid infrastructure provides dynamic workload balancing, high availability and parallel processing for data management, analytics and reporting. Multiple applications and users can share a grid environment for efficient use of hardware capacity and faster performance, while IT can incrementally add resources as needed.
- In-database processing: Moving relevant data management, analytics and reporting tasks to where the data resides improves speed to insight, reduces data movement and promotes better data governance. Using the scalable architecture offered by third-party databases, in-database processing reduces the time needed to prepare data and build, deploy and update analytical models.

2.4 FLEXIBLE DEPLOYMENT OPTIONS FOR BIG DATA

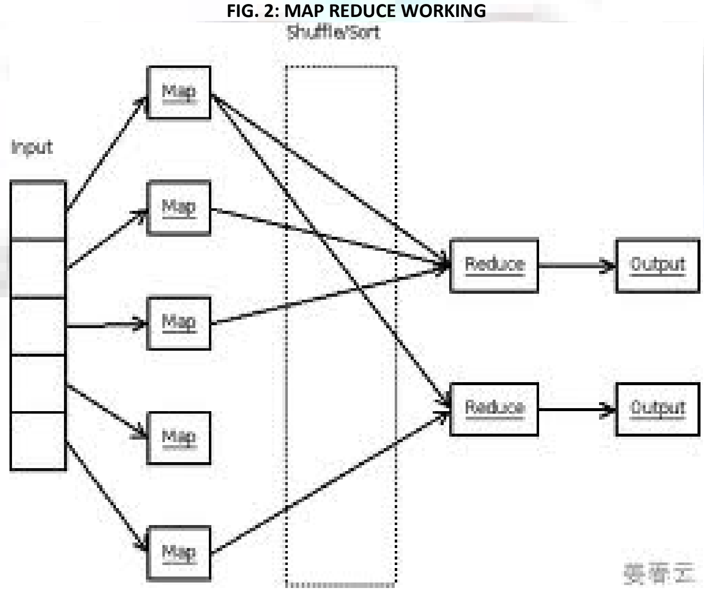
For some organizations, it won't make sense to build the IT infrastructure to support big data, especially if data demands are highly variable or unpredictable. Those organizations can benefit from cloud computing, where big data analytics is delivered as a service and IT resources can be quickly adjusted to meet changing business demands. Our purpose is to utilize Apache CloudStack to manage infrastructure resource by constructing a private cloud, and then integrate relational and NoSQL database for storing multiple datasets[6].

2.5 TOOLS USED IN BIG DATA - MAP REDUCE

MapReduce [1-2] is a programming model for processing large-scale datasets in computer clusters. The MapReduce programming model consists of two functions, map() and reduce(). Users can implement their own processing logic by specifying a customized map() and reduce() function. The map() function takes an input key/value pair and produces a list of intermediate key/value pairs. The MapReduce runtime system groups together all intermediate pairs based on the intermediate keys and passes them to reduce() function for producing the final results.

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

FIG. 2: MAP REDUCE WORKING



2.6 ADVANCEMENTS

The power of MapReduce derives from providing an abstraction that allows developers to harness the power of large clusters but abstractions manage complexity by hiding details and presenting well-defined behaviours to users of those abstractions. This process makes certain tasks easier, but others more difficult, if not impossible .To reduce the computation and to focus the mining for the latter situations, we propose a data science solution that uses MapReduce to mine uncertain Big data for frequent patterns satisfying user-specified anti-monotonic constraints[7]. MapReduce is certainly no exception to this generalization, even within the Hadoop/HDFS/MapReduceecosystem; it is already observed the development of alternative approaches for expressing distributed computations. For example, there can be a third merge phase after map and reduce to better support relational operations. Join processing mentioned n the paper can also tackle the Map Reduce tasks effectively. The future directions in Big Data analysis gives a very encouraging picture paradigm of HDFS and Hadoop, overcoming the existing drawback of the present systems and the advantages it provides.

3. REPORT ON PRESENT INVESTIGATION

3.1 PROBLEM

Big data is set to offer companies tremendous insight. More recently, Big Data has made its appearance in the shared mindset of researchers, practitioners, and funding agencies, driven by the awareness that concerted efforts are needed to address 21st century data collection, analysis, management, ownership, and privacy issues[8]. But with terabytes and petabytes of data pouring in to organizations today, traditional architectures and infrastructures are not up to the challenge. IT teams are burdened with ever-growing requests for data, ad hoc analyses and oneoff reports. Decision makers become frustrated because it takes hours or days to get answers to questions, if at all. More users are expecting self-service Data visualization is becoming an increasingly important component of analytics in the age of big data. access to information in a form they can easily understand and share with others. This begs the question: How do you present big data in a way that business leaders can quickly understand and use? This is not a minor consideration. Mining millions of rows of data creates a big headache for analysts tasked with sorting and presenting data. Organizations often approach the problem in one of two ways: Build “samples” so that it is easier to both analyze and present the data, or create template charts and graphs that can accept certain types of information. Both approaches miss the potential for big data. Instead, consider pairing big data with visual analytics so that you use all the data and receive automated help in selecting the best ways to present the data. This frees staff to deploy insights from data. Think of your data as a great, but messy, story. Visual analytics is the master filmmaker and the gifted editor who bring the story to life. Grouping data together, or “binning,” can help you easily visualize large quantities of data, including outliers. 1 Meeting the need for speed In today’s hypercompetitive business environment, companies not only have to find and analyze the relevant data they need, they must find it quickly. Visualization helps organizations perform analyses and make decisions much more rapidly, but the challenge is going through the sheer volumes of data and accessing the level of detail needed, all at a high speed. The challenge only grows as the degree of granularity increases. One possible solution is hardware. Some vendors are using increased memory and powerful parallel processing to crunch large volumes of data extremely quickly. Another method is putting data in-memory but using a grid computing approach, where many machines are used to solve a problem. Both approaches allow organizations to explore huge data volumes and gain business insights in near-real time. 2 Understanding the data It takes a lot of understanding to get data in the right shape so that you can use visualization as part of data analysis. For example, if the data comes from social media content, you need to know who the user is in a general sense – such as a customer using a particular set of products – and understand what it is you’re trying to visualize out of the data. Without some sort of context, visualization tools are likely to be of less value to the user. One solution to this challenge is to have the proper domain expertise in place.

3.2 SOLUTION

Apache Hadoop is an open-source software framework written in Java for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware. All the modules in Hadoop are designed with a fundamental assumption that hardware failures (of individual machines, or racks of machines) are commonplace and thus should be automatically handled in software by the framework. HDFS helps in efficient management of Big data.

The core of Apache Hadoop consists of a storage part (Hadoop Distributed File System (HDFS)) and a processing part (MapReduce). Hadoop splits files into large blocks and distributes them amongst the nodes in the cluster. To process the data, Hadoop MapReduce transfers packaged code for nodes to process in parallel, based on the data each node needs to process. This approach takes advantage of data locality[3]—nodes manipulating the data that they have on hand—to allow the data to beprocessed faster and more efficiently than it would be in a more conventional supercomputer architecture that relies on a parallel file system where computation and data are connected via high-speed networking.

3.3 IMPLEMENTATION

The user configures and submits a MapReduce job (or just *job* for short) to the framework,which will decompose the job into a set of map tasks, shuffles, a sort, and a set of reduce tasks.The framework will then manage the distribution and execution of the tasks,collect the output, and report the status to the user. The job consists of the parts shown in Figure 3 and listed in Table 1.

FIG. 3: PARTS OF A MAPREDUCE JOB

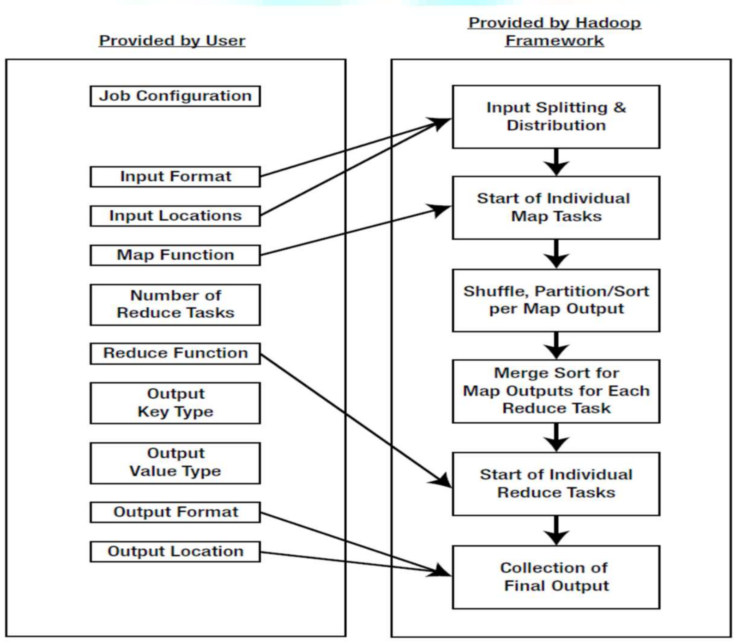


TABLE 1: PARTS OF A MAPREDUCE JOB

Part	Handled By
Configuration of the job	User
Input splitting and distribution	Hadoop framework
Start of the individual map tasks with their input split	Hadoop framework
Map function, called once for each input key/value pair	User
Shuffle, which partitions and sorts the per-map output	Hadoop framework
Sort, which merge sorts the shuffle output for each partition of all map outputs	Hadoop framework
Start of the individual reduce tasks, with their input partition	Hadoop framework
Reduce function, which is called once for each unique input key, with all of the input values that share that key	User
Collection of the output and storage in the configured job output directory, in *N* parts, where *N* is the number of reduce tasks	Hadoop framework

The user is responsible for handling the job setup, specifying the input location(s), specifyingthe input, and ensuring the input is in the expected format and location. The frameworkis responsible for distributing the job among the TaskTracker nodes of the cluster; running themap, shuffle, sort, and reduce phases; placing the output in the output directory; and informing the user of the job-completion status.

4. RESULTS AND DISCUSSIONS

Big data are datasets that grow so large that they become awkward to work with using on-hand database management tools. Difficulties include capture, storage, search, sharing, analytics, and visualizing. This trend continues because of the benefits of working with larger and larger datasets allowing analysts to "spot business trends, prevent diseases, combat crime." Though a moving target, current limits are on the order of terabytes, exabytes and zettabytes of data. Scientists regularly encounter this problem in meteorology, genomics, biological research, Internet search, finance and business informatics. Data sets also grow in size because they are increasingly being gathered by ubiquitous information-sensing mobile devices, "software logs, cameras, microphones, RFID readers, wireless sensor networks and so on." Cloud computing helps organizations store, manage, share, and analyze their Big Data in an affordable and easy-to-use way. Today's cloud Infrastructure-as-a-Service (IaaS) providers such as Microsoft, GoGrid, Amazon, Google, Rackspace and Slicehost, supported by the on-demand analytics solution vendors, make Big Data analytics very affordable. As humans explore the real world through scientific research, humans unravel the mysteries in the information world through big data and data mining, which are attracting much attention from academia[9]. Java and other JVM-based languages play an important role in the Hadoop and Big Data ecosystem[10].

5. FUTURE SCOPE

We regard Big Data as an emerging trend and the need for Big Data is arising in all sciences and engineering fields. With Big Data technologies, we will hopefully be able to provide most relevant and most accurate social sensing feedback to better understand our society. We can further stimulate the participation of the public audiences in the data production circle for societal and economical places. The development and efficient use of big data extends the scope of human activities. It requires proper attention from academia, industry and government. The world has been integrating on a global scale. Human is enforced to change mode from the local to the global in their everyday life . It redefines the relationship among individuals, businesses, governments, and societies through networked thinking and further to improve the human living environment, to enhance the quality of public services, to improve performance, efficiency and productivity through the intelligentized interactive operating. The technological progress and industrial upgrading of big data will create new markets, new business models and new industry rules, and more importantly it demonstrates the collective will of a country that looking for strategic and economic advantage. Although there is still a large gap to gain data intelligence like human wisdom big data is a promising topic and it certainly helps us to understand the world.

6. CONCLUSIONS

Big Data analysis tools like Map Reduce and HDFS, promises to help organizations better understand their customers and the marketplace, hopefully leading to better business decisions and competitive advantages. The need to process enormous quantities of data never been greater. Not only are terabyte- and petabyte-scale datasets rapidly becoming common place, but there is consensus that great value lies buried in them. In the commercial sphere, business intelligence, driven by the ability to gather data from an array of sources. For engineers building information processing tools and applications, large and heterogeneous datasets which are generating continuous flow of data, lead to more effective algorithms for a wide number of tasks, from machine translation to spam detection. In the natural and physical sciences, the ability to analyse massive amounts of data may provide the key to unlocking the secrets of the cosmos . MapReduce can be exploited to solve a variety of problems related to text processing at scales that would have been unthinkable few years back.

REFERENCES

1. Puneet Singh Duggal ,Sanchita Paul , "Big Data Analysis : Challenges and Solutions", International Conference On Cloud, Big Data and Trust 2013 , Nov 13-15 , RGPV.
2. Xindong Wu , Fellow , IEEE , Xingquan Zhu , Senior Member , IEEE , Gong-Qing Wu , and Wei Ding , Senior Member, IEEE , "Data Mining with Big Data", IEEE Transactions on knowledge and Data Engineering, VOL. 26, NO. 1 , January 2014.
3. Shuliang Wang, Hanning YUAN, "Spatial Data Mining in the Context Of Big Data" 2013 International Conference on Parallel and Distributed Systems.
4. Sandy Moens ,EminAksehirli , Bart Goethals , "Frequent Itemset Mining for Big Data", 2013 IEEE International Conference on Big Data.
5. Ana Lucia Varbanescu, Alexandra Iosup, "On Many-Task Big Data Processing: From GPUs to Clouds".
6. Feng Ye, Zhijian Wang, Fachao Zhou, Yapu Wang, Yuanchao Zhou, "Cloud –based Big Data Mining &Analyzing Services Platform integrating R", 2013 International Conference on Advanced Cloud and Big Data.
7. Carson Kai-Sang Leung, Fan Jiang, "A Data Science Solution for Mining Interesting Patterns from Uncertain Big Data", 2014 IEEE Fourth International Conference on Big Data and Cloud Computing.
8. Pascal Hitzler, Krzysztof Janowicz, "Linked Data, Big Data, and the 4th Paradigm", Semantic Web 0(0) 1 IOS Press.
9. Wang Shuliang, DING Gangyi, ZHONG Ming, "Big Spatial Data Mining", 2013 IEEE International Conference on Big Data.
10. Xueyuan Su, Garret Swart, Brian Goetz, Brian Oliver, Paul Sandoz, "Changing Engines in Mindstream: A Java Stream Computational Model for Big Data Processing".

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Computer Application & Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you tosupply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail**infoijrcm@gmail.com** for further improvements in the interest of research.

If youhave any queries please feel free to contact us on our E-mail **infoijrcm@gmail.com**.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-
Co-ordinator

DISCLAIMER

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals