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 **4600 Cities** in **180 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We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the areas of Computer Science & Applications;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

CASE STUDY ON MINING BIG DATA

M.DHANAMALAR
ASST. PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE
KRISTU JAYANTI COLLEGE
BANGALORE

B.AYSHWARYA
ASST. PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE
KRISTU JAYANTI COLLEGE
BANGALORE

ABSTRACT

Big Data is a broad term for datasets used to describe the exponential growth and availability of data, both structured and unstructured., we cannot handle them with our existing methodologies or data mining tools. Big Data is the capability of extracting valuable information from these large datasets or streams of data, that due to its volume, variability, and velocity, it was not possible before to do it.

KEYWORDS
tools, 5Vs, application.

I. INTRODUCTION

In Recent years there are lot of tools & technologies available to collect data from various devices in different formats, from independent or connected applications. This data overflow has capability to process, analyze, store and understand these datasets. Consider the Internet data. The web pages indexed by Google were around one million in 1998, but quickly reached 1 billion in 2000 and have already exceeded 1 trillion in 2008. As per Google in 2013 it reached 30 trillion unique individual pages. It increased 30 times in five years. To store all that data, it need over three million 32GB USB thumb drives. This rapid expansion is accelerated by the dramatic increase in acceptance of social networking applications, such as WhatsApp, Linkeldn, Google plus, Face book, Twitter, Instagram, etc., that allows individuals to create a public profile, to create a list of users with whom to share connections, and view and cross the connections within the system. In addition, with smart phones a device that combines a cell phone with a hand-held computer, typically offering Internet access, data storage, email capability, etc. In day to day life lot of bill payments (recharge, credit card, bus booking etc.,) are doing through smart phones. People and devices (from home, shopping mall to cars, to buses, railway stations and airports) are all loosely connected. Trillions of data and valuable information must be discovered from the data to improve value of life and build our world a better place.

II. BIG DATA

Big data is a popular search query. Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. Input data to big data systems could be chatter from social networks, web server logs, sensor technology and networks, satellite imagery, broadcast audio streams, banking transactions, MP3s of rock music, the content of web pages, scans of government documents, GPS trails, telemetry from automobiles, financial market data etc., Google now processes over 40,000 search queries every second on average which translates to over 3.5 billion searches per day. Number of Twitter search engine queries every day - 2.1 billion. Hours of video watched per month on you tube is 6 billion. We need new algorithms, and new tools to deal with all of this data.

To characterize different type of big data the five Vs of volume, velocity variety, veracity and variability are commonly used.

Volume- It is a scale of data. Big data implies enormous volumes of data. Now that data is generated by machines, networks and human interaction on systems like social media the volume of data to be analysed is very big. For example, 1) It is estimated that 2.5 trillion of gigabytes of data are created every day. 2) World population is 7 billion in that 6 billion of people using cell phones

Velocity – Velocity is analysis of streaming data. Data is being generated fast and need to be processed fast. It deals with speed at which data flows from sources like business processes, machines, networks and human interaction like social media sites, mobile devices, etc. The flow of data is massive and continuous. For example 1) Modern cars have close to 100 sensors that monitor item such as fuel level and tire pressure. 2) New york stock exchange captures 1TB of trade information during each trading session.

Variety – Variety is a different types of data both structured and unstructured. It is used to store data from sources like spreadsheets and databases. Now data comes in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. This variety of unstructured data creates problems for storage, mining and analysing data. For example, 1) Every month 30 billion pieces of content are shared in Facebook. 2) Every month 4 billion hours of videos are watched in you tube.

Veracity- It refers to uncertainty of data. Accuracy of analysis depends on the source data. For example, the company is losing due to poor data management

Variability- There are changes in the structure of the data and the process of being able to handle and manage the data effectively.

The application of Big data are as per following:

Internet of things*:* It involves IoT-connected devices managed by hardware, sensor, and information security. "These devices are sitting in their customers' environment, and they phone home with information about the use, health, or security of the device".

A 360 degree view of the customer**:** Online dealers want to find out what customers are doing on their sites – which pages they visit, where they remain, how long they stay, and when they leave.

Operation analysis*:* Analyze a variety of data to improved business results. By using big data for operations analysis, organizations can gain real-time visibility into operations, customer experience, transactions and behaviour.

Information Security*:* Vendor is looking for an efficient way to store petabytes machine data. In the past, companies would store this information in relational databases. "These traditional systems weren't scaling, both from a performance and cost "so big data is a better option for storing machine data.

III. BIG DATA ANALYSIS AND TOOLS

Hadoop*:* It is an open source framework and distributed processing of very large data sets on computer clusters. Hadoop consists of a storage part (Hadoop Distributed File System (HDFS)) and a processing part (Map Reduce). The Apache distributed data processing software is "Hadoop" and "big data" are used. Map Reduce is a programming model and software framework for writing applications that rapidly process vast amounts of data in parallel on large clusters of compute nodes. It's used by Hadoop, as well as many other data processing applications.

STORM: It's highly scalable, robust, fault-tolerant and works with nearly all programming languages. It is owned by twitter.

MangoDB: An open-source document database, mongoDB is model for developers to control over the final results and processes for handling Big Data. It is a Nosql database with document oriented, full index support, it has a flexibility to index any attribute and functionality. Rich, document-based queries and GridFS for storing files of any size without the risk of compromising your stack, mongoDB is a scaleable, flexible, and powerful solution for Big Data.

HBase: HBase is the non-relational data store for Hadoop. Features include linear and modular scalability, strictly consistent reads and writes, automatic failover support

CouchDB: CouchDB stores data in JSON documents can access via the Web or query using JavaScript. It offers distributed scaling with fault-tolerant storage.

OrientDB: This NoSQL database can store and can load graphs. It combines the flexibility of document databases with the power of graph databases, while supporting features such as ACID transactions, fast indexes, native and SQL queries, and JSON import and export.

Guavus: Guavus drives decision making with powerful analytics to combined with advanced data science and the ability to handle data in real time to derive actionable insights at the precise moment of opportunity.

IV. OPERATIONAL VS ANALYTICAL

Two classes of Technology in big data are system that provide the real time interactive workload where data is captured and stored is operational big data and system that complex analysis of data is analytical big data.

OPERATIONAL DATA: For operational Big Data workloads, NoSQL Big Data systems such as document databases have emerged to address a broad set of applications, and other architectures, such as key-value stores, column family stores, and graph databases are optimized for more specific applications. NoSQL technologies, which were developed to address the shortcomings of relational databases in the modern computing environment, are faster and scale much more quickly and inexpensively than relational databases.

ANALYTICAL DATA: Analytical Big Data workloads, on the other hand, tend to be addressed by MPP database systems and MapReduce. These technologies are also a reaction to the limitations of traditional relational databases and their lack of ability to scale beyond the resources of a single server. Furthermore, MapReduce provides a new method of analyzing data that is complementary to the capabilities provided by SQL.

As applications gain traction and their users generate increasing volumes of data, there are a number of retrospective analytical workloads that provide real value to the business. Where these workloads involve algorithms that are more sophisticated than simple aggregation, MapReduce has emerged as the first choice for Big Data analytics. Some NoSQL systems provide native MapReduce functionality that allows for analytics to be performed on operational data in place. Alternately, data can be copied from NoSQL systems into analytical systems such as Hadoop for MapReduce.

OPERATIONAL DATA

A company's operations are supported by applications that automate key business processes. These include areas such as sales, service, order management, manufacturing, purchasing, billing, accounts receivable and accounts payable. These applications require significant amounts of data to function correctly. This includes data about the objects that are involved in transactions, as well as the transaction data itself. For example, when a customer buys a product, the transaction is managed by a sales application. The objects of the transaction are the Customer and the Product. The transactional data is the time, place, price, discount, payment methods, etc. used at the point of sale. The transactional data is stored in On-Line Transaction Processing (OLTP) tables that are designed to support high volume low latency access and update.

ANALYTICAL DATA

Analytical data is used to support a company's decision making. Customer buying patterns are analyzed to identify churn, profitability, and marketing segmentation. Suppliers are categorized, based on performance characteristics over time, for better supply chain decisions. Product behavior is scrutinized over long periods to identify failure patterns. This data is stored in large Data Warehouses and possibly smaller data marts with table structures designed to support heavy aggregation, ad hoc queries, and data mining. Typically the data is stored in large fact tables surrounded by key dimensions such as customer, product, supplier, account, and location.

V. CONCLUSION

The study concluded that Big Data is a broad term for datasets used to describe the exponential growth and availability of data, both structured and unstructured., we cannot handle them with our existing methodologies or data mining tools. Big Data is the capability of extracting valuable information from these large datasets or streams of data, that due to its volume, variability, and velocity, it was not possible before to do it.

REFERENCES

1. "Big Data" http:// en.wikipedia .org/ wiki / Big_data
2. "Big Data and Analytic hub",http://www. ibm bigdatahub .com/infographic/four-vs-big-data.
3. "Big data use case watch" http:// www. informationweek.com/big-data/big-data-analytics /5-big-data-use-cases-to-watch/d/d-id/1251031
4. "Bigdata" http://www.mongodb.com/big-data-explained
5. "Google searches query per day", http://www.internetlivestats.com/google-search-statistics/
6. "Issue of bigdata veracity" http:// inside big data.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/
7. "Mining Big Data: Current Status, and Forecast to the Future"- Wei Fan, Albert Bifet
8. "Open source tools for bigdata" http :// www . datamation.com/data-center/50-top-open-source-tools-for-big-data-1.html
9. "Operational vs analytical" –https:// blogs . oracle..com/mdm/entry/operational_vs_analytical_master_data_management
10. "Twitter", https://www.convertwithcontent.Com /twitter-search-advanced/
11. "Youtube", http:// expande dramblings . com / index.php/youtube-statistics/

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 to supply 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 you have 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

The information and opinions presented in the Journal reflect the views of the authors and not of the Journal or its Editorial Board or the Publishers/Editors. Publication does not constitute endorsement by the journal. Neither the Journal nor its publishers/Editors/Editorial Board nor anyone else involved in creating, producing or delivering the journal or the materials contained therein, assumes any liability or responsibility for the accuracy, completeness, or usefulness of any information provided in the journal, nor shall they be liable for any direct, indirect, incidental, special, consequential or punitive damages arising out of the use of information/material contained in the journal. The journal, neither its publishers/Editors/ Editorial Board, nor any other party involved in the preparation of material contained in the journal represents or warrants that the information contained herein is in every respect accurate or complete, and they are not responsible for any errors or omissions or for the results obtained from the use of such material. Readers are encouraged to confirm the information contained herein with other sources. The responsibility of the contents and the opinions expressed in this journal are exclusively of the author (s) concerned.

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