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

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

Efficient Methodology for Emerging and Trending of Big Data Based Applications

R.Karthikeyan MCA., M. Phil. [1], Dr. T. Geetha.,M.C.A.,M.Phil.,Ph.D [2,]

P. Lavanya [3],.S. Kalpana [4]

Assistant Professor, Department. of MCA, Gnanamani College of Technology, Namakkal, India[1]

Head of the Department, Department. of MCA, Gnanamani College of Technology, Namakkal, India[2]

PG Scholar, Department. of MCA, Gnanamani College of Technology, Namakkal, India[3]

PG Scholar, Department. of MCA, Gnanamani College of Technology, Namakkal, India[4]

ABSTRACT: Big data is a main buzz phrase and new curve for IT today. Big data is driven data with high velocity, volume, variety, veracity and value. It comes from different sources like mobile devices, internet, social media, sensors, geospatial devices and other machine-generated data. Traditional data processing and analysis of structured data using RDBMS and data warehousing no longer satisfy the challenges of Big Data. Therefore, big data analysis is a current area of research and development. The basic objective of this paper is to explore the potential impact of big data challenges, open research issues, and various tools associated with it. As a result, this article provides a platform to explore big data at numerous stages. Today's we have an overwhelming growth of data in terms of volume, velocity and variety on web. Moreover this, from a security and privacy views, both area have an unpredictable growth. So Big Data challenge is becoming one of the most exciting opportunities for researchers in upcoming years.

KEYWORDS: Big data, Hadoop, HDFS, Map Reduce, NoSQL.

I. INTRODUCTION

In digital world, data are generated from various sources and the fast transition from digital technologies has led to growth of big data. It provides evolutionary breakthroughs in many fields with collection of large datasets. In general, it refers to the collection of large and complex datasets which are difficult to process using traditional database management tools or data processing applications. These are available in structured, semi-structured, and unstructured format in petabytes and beyond. Formally, it is defined from 3Vs to 4Vs. 3Vs refers to volume, velocity, and variety. Volume refers to the huge amount of data that are being generated everyday whereas velocity is the rate of growth and how fast the data are gathered for being analysis. Variety provides information about the types of data such as structured, unstructured, semistructured etc. The fourth V refers to veracity that includes availability and accountability. The prime objective of big data analysis is to process data of high volume, velocity, variety, and veracity using various traditional and computational intelligent techniques [1]. Some of these extraction methods for obtaining helpful information was discussed by Gandomi and Haider [2]. The following Figure 1 refers to the definition of big data. However exact definition for big data is not defined and there is a believe that it is problem specific. This will help us in obtaining enhanced decision making, insight discovery and optimization while being innovative and cost-effective.

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

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

II. ARCHITECTURE OF GRID COMPUTING

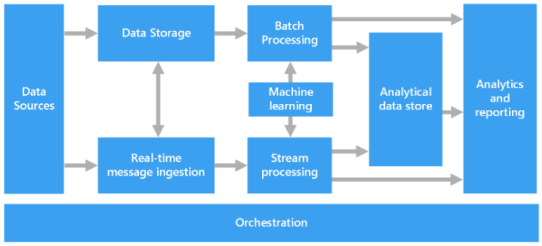

Figure 1: Architecture

Most big data architectures include some or all of the following components:

Data sources:

All big data solutions start with one or more data sources. Examples include: Application data stores, such as relational databases. Static files produced by applications, such as web server log files. Real-time data sources, such as IoT devices.

Data storage. Data for batch processing operations is typically stored in a distributed file store that can hold high volumes of large files in various formats. This kind of store is often called a *data lake*. Options for implementing this storage include Azure Data Lake Store or blob containers in Azure Storage.

Batch processing. Because the data sets are so large, often a big data solution must process data files using long-running batch jobs to filter, aggregate, and otherwise prepare the data for analysis. Usually these jobs involve reading source files, processing them, and writing the output to new files. Options include running U-SQL jobs in Azure Data Lake Analytics, using Hive, Pig, or custom Map/Reduce jobs in an HDInsight Hadoop cluster, or using Java, Scala, or Python programs in an HDInsight Spark cluster.

Real-time message ingestion. If the solution includes real-time sources, the architecture must include a way to capture and store real-time messages for stream processing. This might be a simple data store, where incoming messages are dropped into a folder for processing. However, many solutions need a message ingestion store to act as a buffer for messages, and to support scale-out processing, reliable delivery, and other message queuing semantics. This portion of a streaming architecture is often referred to as stream buffering. Options include Azure Event Hubs, Azure IoT Hub, and Kafka.

III. APPLICATIONS OF BIG DATA

Big data applications solve and analyze real world problems using Hadoop and associated tools. Internet users and machine-to-machine connections are causing the data growth. Real time areas are defined following in which big data is used: A. Big data in healthcare Healthcare practices and policies differ tremendously around the world, there are three objectives regarding healthcare system [10]. The first objective is to improve the patient experience (including quality and satisfaction). Second, improving overall population health and reducing the cost of health care and third is traditional methods have fallen short to manage healthcare and create modern technology to analyze large quantities of information. It is time consuming for clinical staff to Collecting massive amounts of data in healthcare. High-performance analytics are new technologies making easier to turn massive amounts of data into relevant and critical insights used to provide better care. Analytics helps to predict negative intervene and reactions. Unstructured data can be captured through text mining from patient records. It means information can be collecting without causing additional work for clinicians. Transparent, information can thus improve encourage and quality innovation. As information becomes increasingly available, comparable and transparent, patients will also be empowered and more involved in their own treatment through online health applications, which can integrate patient information with their health records and make it available to clinicians. A massive amount of data collected from different sources provides the best

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

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

practices for today, and will help healthcare providers identify trends so they can achieve better results to improve medical facilities all around the world..

IV. TECHNOLOGIES USED FOR BIG DATA ANALYTICS

NoSQL database can handle unstructured and unpredictable data.The data stored in a NoSQL database is typically of a high variety.A NoSQL database provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases. Relational and NoSQL data models are very different. The relational model takes data and separates it into many interrelated tables that contain rows and columns. But document-oriented NoSQL database takes the data into documents using the JSON format. JSON is Javascript Object Notation.Another major difference is that relational technologies have rigid schemas while NoSQL models are schemaless. Many NoSQL databases have excellent integrated caching capabilities. So,the frequently used data is kept in system memory

V. OPEN RESEARCH ISSUES IN BIG DATA ANALYTICS

Big data analytics and data science are becoming the research focal point in industries and academia. Data science aims at researching big data and knowledge extraction from data. Applications of big data and data science include information science, uncertainty modeling, uncertain data analysis, machine learning, statistical learning, pattern recognition, data warehousing, and signal processing. Effective integration of technologies and analysis will result in predicting the future drift of events. Main focus of this section is to discuss open research issues in big data analytics. The research issues pertaining to big data analysis are classified into three broad categories namely internet of things (IoT), cloud computing, bio inspired computing, and quantum computing. However it is not limited to these issues. More research issues related to health care big data can be found in Husing Kuo et al. paper.

VI. TOOLS: OPEN SOURCE REVOLUTION THE BIG DATA

phenomenon is intrinsically related to the open source software revolution. Most of the large companies like as Facebook, Yahoo!, Twitter, and LinkedIn etc. benefit and contribute working on open source projects. Big Data infrastructure deals with Hadoop, and other related software as: Apache Hadoop [3]: this software is used for• data-intensive distributed applications; based in the MapReduce programming model and a distributed file system called Hadoop Distributed File system (HDFS). Hadoop allows writing applications that rapidly process large amounts of data in parallel on large clusters of compute nodes. A MapReduce job divides the input dataset into independent subsets that are processed by map tasks in parallel [38]. This step of mapping is then followed by a step of reducing tasks. These reduced tasks use the output of the maps to obtain the final result of the job. Apache S4 [2, 3, and 16]: it provides a• platform for processing continuous data streams. S4 is designed specifically for managing data streams. S4 apps are designed combining streams and processing elements in real time [38]. Apache Hadoop related projects [3, 16]:• Apache Pig, Apache Hive, Apache HBase, Apache ZooKeeper, Apache Cassandra, Cascading, Scribe and many others [38]. Storm [1, 3]: this software is used for• streaming data-intensive distributed applications, similar to S4, and developed by Nathan Marz at Twitter.

VII. CONCLUSION

Due to increasing the size of data day by day, Big Data is going to continue growing during the next years and becoming one of the exciting opportunities in future. This paper insights about the topic, and controversy, and the main challenges etc. for the future. Hence Big Data is becoming the new Final Frontier for scientific data research and for business applications. And on challenging side, Securely Management of Big Data with today's threat spectrum is a big issue. Because today's we have an overwhelming growth of data in terms of volume, velocity and variety. So from a security and privacy standpoint, the threat landscape and security and privacy risks have also seen an unprecedented growth. So as for future research is needed to build a generic architectural framework towards addressing these security

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

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

and privacy challenges in a holistic manner. Now we are in a new era where Big Data mining will help us to discover knowledge that no one has discovered before. So everybody is warmly invited to participate in this intrepid journey to discover the future views.

REFERENCES

1. R.Karthikeyan, & et all "Biometric for Mobile Security" in the international journal of Engineering Science & Computing, Volume7,Issue6, June 2017, ISSN(0):2361-3361,PP No.:13552-13555.
2. R.Karthikeyan, & et all "Data Mining on Parallel Database Systems" in the international journal of Engineering Science & Computing, Volume7,Issue7, July 2017, ISSN(0):2361-3361,PP No.:13922-13927.
3. R.Karthikeyan, & et all "Ant Colony System for Graph Coloring Problem" in the international journal of Engineering Science & Computing, Volume7,Issue7, July 2017, ISSN(0):2361-3361,PP No.:14120-14125.
4. R.Karthikeyan, & et all "Classification of Peer –To- Peer Architectures and Applications" in the international journal of Engineering Science & Computing, Volume7,Issue8, Aug 2017, ISSN(0):2361-3361,PP No.:14394-14397.
5. R.Karthikeyan, & et all "Mobile Banking Services" in the international journal of Engineering Science & Computing, Volume7,Issue7, July 2017, ISSN(0):2361-3361,PP No.:14357-14361.
6. R.Karthikeyan, & et all "Neural Networks for Shortest Path Computation and Routing in Computer Networks" in the international journal of Engineering and Techniques, Volume 3 Issue 4, Aug 2017, ISSN:2395-1303,PP No.:86-91.
7. R.Karthikeyan, & et all "An Sight into Virtual Techniques Private Networks & IP Tunneling" in the international journal of Engineering and Techniques, Volume 3 Issue 4, Aug 2017, ISSN:2395-1303,PP No.:129-133.
8. R.Karthikeyan, & et all "Routing Approaches in Mobile Ad-hoc Networks" in the International Journal of Research in Engineering Technology, Volume 2 Issue 5, Aug 2017, ISSN:2455-1341, Pg No.:1-7.
9. R.Karthikeyan, & et all "Big data Analytics Using Support Vector Machine Algorithm" in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 6 Issue 9, Aug 2018, ISSN:2320 - 9798, Pg No.:7589 - 7594.
10. R.Karthikeyan, & et all "Defense and Confidentiality Smart Access for High Throughput in Mobile Internet of Things " in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 6 Issue 9, Aug 2018, ISSN:2320 - 9798, Pg No.:7222 - 7228.
11. R.Karthikeyan, & et all "Data Security of Network Communication Using Distributed Firewall in WSN " in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 6 Issue 7, July 2018, ISSN:2320 - 9798, Pg No.:6733 - 6737.
12. R.Karthikeyan, & et all "An Internet of Things Using Automation Detection with Wireless Sensor Network" in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 6 Issue 9, September 2018, ISSN:2320 - 9798, Pg No.:7595 - 7599.
13. R.Karthikeyan, & et all "Entrepreneurship and Modernization Mechanism in Internet of Things" in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 7 Issue 2, Feb 2019, ISSN:2320 - 9798, Pg No.:887 - 892.
14. R.Karthikeyan ,& etall "Efficient Methodology and Applications of DynamicHeterogeneous Grid Computing" in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 7 Issue 2, Feb 2019, ISSN:2320 - 9798, Pg No.:1125 -1128.

BIOGRAPHY

1. R.KARTHIKEYAN
Assistant Professor,
Department of MCA
Gnanamani College of Technology
Namakkal (Dt),

2 Dr. T. GEETHA,
Head of the Department,
Department of MCA
Gnanamani College of Technology
Namakkal (Dt),

3. Ms.P. LAVANYA
PG Scholar
Department of MCA
Gnanamani College of Technology
Namakkal (DT),

4. Ms.S. KALPANA
PG Scholar
Department of MCA
Gnanamani College of Technology
Namakkal (DT),