JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

"Integraion of Big Data With Scalable Computing"R

Rudra Girishkumar Patel, Hara Yashpalbhai Raval, Riya Amitbhai Raval
Student
GLS University

,

Abstract

Hara Yashpalbhai Raval Rudra Girishkumar PatelRiya Amitbhai Raval

Rapid advancement in new technologies in the last few years has led to the growth of knowledge in various forms such as knowledge generation, communication, business intelligence and medicine. The concept of big data aims to capture the essence of this new vision. Compared with traditional data, big data has other characteristics in addition to big data. For example, large files are completely distributed and unnecessary. This new version allows us to create new models for data collection, distribution, storage and analysis. In this paper, we propose a big data analytics framework in the form of a scalable system that aims to provide a holistic dimension to big data systems to solve big data problems. First, we introduce the definition and background of big data.
We then propose a framework to describe the architecture of big data. Finally, we describe the challenges facing big data and the need for future work. Key words- big data analysis, data collection, data storage, data transmission, data analysis.The abstract appears to be about computing, specifically using cloud computing as the infrastructure for processing and analyzing big data. The authors discuss the challenges of dealingwith large amounts of data, known as big data, and how traditional processing tools are unable to handle them effectively. They introduce big data analytics as a crucial process in many fields and sectors, and discuss the issues and benefits of this process. The abstract also touches on the exponential growth of data generated daily, and the need for new technologies and ways of processing and analyzing this data. The authors discuss the challenges of dealingwith large amounts of data, known as big data, and how traditional processing tools are unable to handle them effectively. They introduce big data analytics as a crucial process in many fields and sectors, and discuss the issues and benefits of this process. The authors use Google's BigQuery as a case study to perform experiments on different sizes of datasets. The abstract also touches on the exponential growth of data generated daily, and the need for new technologies and ways of processing and analyzing this data. The authors aim to provide an overview of how analytics of big data in cloud computing can be done using Google's BigQuery platform.

I. INTRODUCTION

Over the next decade, digital data willgrow in India40,000 petabytes to 2.3 millionpetabytes, twice as fast as global pace and will continue to grow exponentiallyand attract different attentions. theterm "big data" was created to capture this data explosion trend and data was certainly considered a buzzword who is expected to explore our company. According to various industry estimates, the big data industry has exceeded 8 billion USD by 2012 and will reach 16.9 billion USD by 2015 worldwide, growing seven times faster than the growth rate of the ICT industry. An IDC report [1] predicts that the "digital universe" will growth to 2.7 ZB in 2012, 48% more than in 2011 and skyrocketing almost 8 ZB by 2015. According to NASSCOM [2], the big data market in India will grow by 83% annually 1 billion USD by 2015. Huge potential associated with big data has led to an emerging field of research that has immediately attracted huge interest from various industries, for example, government, industry and research commonwealth. Wide interest first illustrates understanding for both industry news [3] and the public media (e.g. The Economist [4], [5], New York Times [6]). The role of the government played a vital role promote new ideologies
[7] and stimulate progresssolving big data problems. Finally,nature and Scientific journals published special issues consider the big data phenomenon and its challenges, flourishing its impact beyond the technological sphere.As a result, this emerging problem in the field of big data z different areas require a clear and intuitive understanding its definition, architecture, technology and its various challenges. This overview paper focuses on scalable big data systems, which include a set of techniques to investigate, evaluate, validate, implement and deploy heterogeneous data using massive multiprocessor power perform a comprehensive analysis. Although, uniqueness big-data lies beneath it and suggests scalable big-data the system faces many technical challenges, including: First, disparate data sources, associated high costs and the main problem is bottlenecks in the infrastructure identified by companies for effective management unstructured data streams.Second, it needs to be captured and managed heterogeneous datasets while providing guaranteed services such as scalability, privacy and speed search. Third, the datasets must be at different levels in real time effectively mined by analyzing big data including prediction, modeling, visualization and optimization.Big data is a term used to describe the large volume of structured, semi-structured, and unstructured data that organizations collect, which can be used in machine learning projects, predictive modeling, and other advanced analytics applications. The three V's of big data are volume, velocity, and variety, which were first identified in 2001 by Doug Laney. Big data systems are used to improve operations, provide better customer service, create personalized marketing campaigns, and increase revenue and profits. [i]Scalable computing is the ability of a system to handle increasing amounts of work by adding resources to the system. In the context of big data, scalable computing is essential to process and store the vast amounts of data being collected. Scalable computing systems often involve distributed architectures, where processing workloads are distributed across hundreds or thousands of commodity servers. Integrating big data with scalable computing involves using distributed computing frameworks like Hadoop and Spark to process and analyze big data. These frameworks can handle the volume, velocity, and variety of big data by distributing processing workloads across a cluster of servers. Big data processing places heavy demands on the underlying compute infrastructure, and the cloud is a popular location for big data systems due to its cost-effectiveness and scalability. Big data processing involves several steps, including data preparation, data science, and advanced analytics. Data preparation involves profiling, cleansing, validation, and transformation of data sets. Data science and advanced

analytics disciplines, such as machine learning, predictive modeling, data mining, statistical analysis, streaming analytics, and text mining, can be applied to run different applications. There are several big data platforms and managed services offered by IT vendors, including Amazon Web Services, Microsoft Azure, Google Cloud Platform, IBM, and Oracle. These platforms combine many big data technologies in a single package, primarily for use in the cloud. For organizations that want to deploy big data systems themselves, tools such as Hadoop, Spark, NoSQL databases, and data warehouses can be used. Managing big data systems requires new skills compared to traditional relational databases. Designing a big data architecture is a common challenge for users, and using a managed cloud service can ease these issues. Other challenges in managing big data systems include making the data accessible to data scientists and analysts, integrating sets of big data, and ensuring data quality and governance. In conclusion, integrating big data with scalable computing is essential for organizations to process and analyze the vast amounts of data being collected. Distributed computing frameworks like Hadoop and Spark, cloud-based big data platforms, and tools such as NoSQL databases and data warehouses can be used to handle the volume, velocity, and variety of big data. Managing big data systems requires new skills and best practices, including data governance, data quality management, and focusing on business needs forinformation over the available technologies.

Literature Review

Paper Title	Author	Abstract
Big Data Analytics: A Literature ReviewPaper [1]	Nada Elgendy University of Oulu Ahmed Elragal Luleå University ofTechnology	This abstract addresses thechallenges of big data, emphasizing its scale and complexity. It explores innovative solutions for managing and extracting value from diverse datasets, with a focus on analytics methods applied to decision domains.
A Survey on Big DataAnalytics: Challenges,Open Research Issues and Tools [2]	D. P. Acharjya VIT University Kauser Ahmed PVIT University	This paper discusses the challenges arising from the substantial daily data generation in information systems, IoT, and cloud computing, emphasizing the importance of multi-level efforts.
Big Data: A Review [3]	Vaibhav Khanna Jagan Institute of management studies	This abstract introduces big data, emphasizing data mining's pivotal role in extracting insights from diverse and complex datasets. It explores challenges in big data analysis and offers insights into future considerations for managing and leveraging large-scale data.

Big Data Analytics: A Literature Review Perspective [4]	Sarah Al-Shiakhli Luleå University ofTechnology	This thesis delves into the multifaceted realm of big data, emphasizing its crucial role in decision-making through an exploration of extraction, storage, management, and analytical processing. It highlights the urgency to develop efficient solutions for data handling,showcasing the widespread adoption of big data analytics across sectors and examining its applications through diverse analytical techniques and tools.
Big Data Analytics for Complex Systems [5]	Ashraf MohamedAbou Tabl University of Windsor	This study introduces an integrated system employing big data analytics, machine learning, supercomputing, health machines to emulate human intelligence.

An Explorative Perceived Challenges and Remediating Strategies for Big Data among Data Practitioners [6]	Patrick Pilipiec on Olga Soprano Linnaeus University	The surge in global data, driven by the Internet of Things, has sparked considerable investment in big data analytics for business value creation. Despite extensive literature on challenges, a gap remains in understanding the practical experiences of data practitioners, highlighting the necessity for a more comprehensive exploration of their perspectives on big data analytics.
The Evolution of Big Data and Its Business Applications [7]	Marwah Ahmed Halwani University Of North Taxes	This research provides a fresh perspective on defining Big Data, with a focus on social media data. It explores the untapped potential of social media for organizational effectiveness and examines its impact on the 2016 U.S. presidential election, particularly among millennials. The dissertation also clarifies the skills needed for Big Data professionals, contributing theoretical and methodological insights to refine the analysis of Big Data.
Investigating the Impact of Big Data Supply Chain Operation: Case Studies from the UK Private Sector [8]	By Ruaa Hasan Brunel University London 2021	This research introduces a novel perspective on Big Data, focusing on social media data inconsistencies. The dissertation comprises three essays, uncovering untapped potential in organizational use of social media data, exploring its impact on millennial voting decisions in the 2016 U.S. presidential election, and providing insights into required skills for Big Data professionals. The study contributes valuable theoretical and methodological enhancements to the analysis of Big Data.
A Framework for Big Data Analytics as a Scalable Systems[9]	Himanshu Shekhar Manoj Sharma	This research introduces a novel perspective on Big Data, focusing on social media datainconsistencies.
Big Data: An Overview with Legal Aspects and Future Prospects [10]	Mohammad Nazmul Alam Vakil Singh Ms. Ripendeep Kaur Md.Shahin Kabir	This research paper provides a thorough examination of big data, encompassing applications, challenges, and processing tools across various sectors.
Automated debugging of big data analytics indata -intensive scalable	Mohammad Ali Gulzar Siman Wang Miryung Kim	BIGSIFT, a novel data localization approach, integrates automated fault isolation and data provenance in Big Data Analytics. Leveraging

computing[11]		a redefined data provenance with a test oracle function and unique optimizations, BIGSIFT achieves significantly improved fault localizability accuracy by orders- of-magnitude and boosts performance by up to 66x, localizing fault-inducing data within 62% of the original job running time for each faulty output.
Closed Loop Big Data Analysis with Visualization and Scalable Computing[12]	Guangchen RuanHui Zhang	Many scientific investigations require data-intensive research where big data are collected and analyzed. To get big insights from big data, we need to first develop our initial hypotheses from the data and then test and validate our hypotheses about the data.
Big Data and Cloud Computing: Current state and future opportunities [13]	Divyakant AgrawalSudipto Das	Scalable database management systems (DBMS)---both for update intensive application workloads as well as decision support systems for descriptive and deep analytics---are a critical part of the cloud infrastructure and play an important role in ensuring the smooth transition of applications.
The Rise Of "Big Data"On Cloud computing [14]	Ibrahim Abker TargioHashem Abdullah Gani	Cloud computing is a powerful technology to perform massive-scale and complex computing. It eliminates the need to maintain expensive computing hardware, dedicated space, and software.
Big data Computing[15]	Rajendra Akerkar	Big Data Computing refers to the processing and analysis of vast and complex datasets that exceed the capabilities of traditional data processing applications.
Scalable machine learning algorithms for big data analytics [16]	Preeti gupta Arun sharmaRajin Jinda	Cloud computing is a powerful technology to perform massive-scale and complex computing. It eliminates the need to maintain expensive computing hardware, dedicated space, and software
Big Data Processing in Cloud Computing Environments [17]	Yu Li Chagqing Ji	With the rapid growth of emerging applications like social network analysis, semantic Web analysis and bioinformatics network analysis, a variety of data to be processed.
Big Data Computing:Creati-ng Revolutionary breakthrough in commerce and society[18]	Randal Brayant Randy Katz Edward Lazowska	Advances in digital sensors, communications, computation, and storage have created huge collections of data, capturing

		information of value to business, science, government, and society.
Challenges of Big Data Analysis [19]	jianqing FanFang Han Han Liu	Big Data bring new opportunities to modern society and challenges to data scientists. On the one hand, Big Data hold great promises for discovering subtle population patterns and heterogeneities that are not possible with small-scale data.
Big Data: ScalabiltyStorage [20]	Aruna MailavaramB Padmaja Rani	As the growth of enterprise data accelerates, the task of protecting data becomes morechallenging.

i Conclusion

The prevalence of big data currently brings urgency the need for advanced data collection, management and analytical mechanisms. In this post, we have introducedthe concept of big data and highlightedthe value of big datachain. The big datavalue chain consists of four stages: datadata generation, data acquisition, data storage and data analysis .Furthermore, we presented a numberof mechanisms in different phases. We have listed several of them in the generation phase data sources and data attributes. In the data acquisition phasecollection methods and subsequently data were reviewed data transfer and preprocessing methods. In warehousephase No SQL stores were introduced .Finally in a big way in the data analysisphase, we examined different data analytical methods organized by data Properties.In conclusion, the integration of big data with scalable computing technologies is crucial for organizations to leverage the power of data and make informed, data-driven decisions. By harnessing the power of scalable computing technologies, organizations can process and analyze large volumes of data in real-time or near real-time, leading to improved decision-making, increased efficiency, and enhanced customer experiences.

References

[1] Elgendy, Nada & Elragal, Ahmed. (2014). Big Data Analytics: A Literature Review Paper. Lecture Notes in Computer Science. 8557. 214-227. 10.1007/978-3-319-08976-8_16.

[2] Saxena, Aarush. (2022). A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools. International Journal for Research in Applied Science and Engineering Technology. 10. 1899-1908. 10.22214/ijraset.2022.48294.

[3] Aburawi, Yousef & Albaour, Abdulbaset. (2021). Big Data: Review Paper. International Journal Of Advance Research And Innovative Ideas In Education. 7. 2021.

[4] Elgendy, Nada & Elragal, Ahmed. (2014). Big Data Analytics: A Literature Review Paper. Lecture Notes in Computer Science. 8557. 214-227. 10.1007/978-3-319-08976-8_16.

[5] Hassanien, Aboul Ella & Azar, Ahmad Taher & Snasel, Vaclav & Kacprzyk, Janusz. (2015). Big Data in Complex Systems: Challenges and Opportunities. 10.1007/978-3-319-11056-1.

[6] Sejahtera, Feliks & Wang, Wei & Indulska, Marta & Sadiq, Shazia. (2019). Factors Influencing Effective Use of Big Data: A Research Framework. Information & Management. 57. 10.1016/j.im.2019.02.001.

[7] Batistic, Sasa & van der Laken, Paul. (2018). The History, Evolution, and Future of Big Data and its Relationship to Performance in Organizations. Academy of Management Proceedings. 2018. 16317. 10.5465/AMBPP.2018.16317abstract.

[8] Ionica, Holban & Bunget, Ovidiu & Türkeș, & Capusneanu, Sorinel & Topor, Dan & Tamaș, & Rakos, Lili & Hint, Mihaela. (2019). The Impact of Big Data Analytics on Company Performance in Supply Chain Management. Sustainability. 11. 4864. 10.3390/su11184864.

[9] Hu, Han & Wen, Yonggang & Chua, Tat-Seng & Li, Xuelong. (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. Access, IEEE. 2. 652-687. 10.1109/ACCESS.2014.2332453.

[10] Nazmul Alam, Mohammad & Singh, Vakil & Kaur, Ms & Kabir, Md Shahin. (2023). Big Data: An Overview with Legal Aspects and Future Prospects. 10. 476-485.

[11] Gulzar, Muhammad & Wang, Siman & Kim, Miryung. (2018). BigSift: automated debugging of big data analytics in data-intensive scalable computing. 863-866. 10.1145/3236024.3264586.

[12] Ruan, Guangchen. (2017). Closed-loop Big Data Analysis with Visualization and Scalable Computing. Big Data Research. 8. 10.1016/j.bdr.2017.01.002

[13] Agrawal, Divyakant & Das, Sudipto & Abbadi, Amr. (2011). Big Data and Cloud Computing: Current State and Future Opportunities. ACM International Conference Proceeding Series. 530-533. 10.1145/1951365.1951432

[14] Hashem, Ibrahim & Yaqoob, Ibrar & Anuar, Nor & Mokhtar, Salimah & Gani, Abdullah & Khan, Samee. (2014). The rise of "Big Data" on cloud computing: Review and open research issues. Information Systems. 47. 98-115. 10.1016/j.is.2014.07.006.

[15] Kune, Raghavendra & Konugurthi, Pramod & Agarwal, Arun & Rao, C. & Buyya, Rajkumar. (2015). The Anatomy of Big Data Computing. Software: Practice and Experience. 46. 10.1002/spe.2374.

[16] Gupta, Preeti & Sharma, Arun & Jindal, Rajni. (2016). Scalable machine-learning algorithms for big data analytics: a comprehensive review: Scalable machine-learning algorithms for big data analytics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 6. 10.1002/widm.1194.

[17] Ahmad, Noraziah & Fakherldin, Mohammed & Adam, Khalid & Majid, Mazlina. (2017). Big Data Processing in Cloud Computing Environments. Advanced Science Letters. 23. 11092-11095. 10.1166/asl.2017.10227.

[18] Guo, Huadong & Wang, Lizhe & Chen, Fang & Liang, Dong. (2014). Scientific big data and Digital Earth. Chinese Science Bulletin (Chinese Version). 59. 1047. 10.1360/972013-1054.

[19] Fan, Jianqing & Han, Fang & Liu, Han. (2013). Challenges of Big Data Analysis. National Science Review. 1. 10.1093/nsr/nwt032.

[20] Mailavaram, Aruna & Rani, B.. (2019). Big Data: Scalability Storage. 10.1007/978-981-13- 7082-3_54.