iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68353

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Big Data Analysis using Cloud Computing: Opportunities, Challenges & Applications

Ganesh Mohite[1], Rushikesh Bhagat[2], Roshan Jadhav[3]
Student, Pvg's college of science and commerce

Abstract: Big Data and cloud computing have revolutionized data storage, processing, and analysis, enabling businesses and industries to manage vast volumes of data efficiently. Cloud computing provides scalable infrastructure, cost-effective storage solutions, and real-time analytics capabilities, making it an essential platform for Big Data applications. This study explores the opportunities, challenges, and applications of Big Data analysis in cloud environments, highlighting key technologies such as Hadoop, Spark, and cloud-based data warehousing solutions. The research identifies major challenges, including data security, integration complexities, and performance bottlenecks, while proposing solutions such as encryption, real-time analytics frameworks, and hybrid cloud models. Furthermore, it discusses industry applications across healthcare, finance, IoT, and business intelligence. The findings demonstrate that cloud-based Big Data solutions enhance operational efficiency, decision making, and scalability, paving the way for future advancements in AI-driven analytics, edge computing, and cross-cloud integrations.
Keywords: Big Data, Cloud Computing, Data Storage, Data Processing, Data Analysis, Scalable Infrastructure, Cost-Effective Storage, Real-Time Analytics, Hadoop, Spark, Cloud Data Warehousing, Data Security, Integration Complexities.

I. INTRODUCTION

Cloud Computing: Cloud computing is a transformative technology that enables the delivery of computing services such as servers, storage, databases, networking, software, analytics, and intelligence over the internet. It provides scalable resources, cost-efficiency, and flexibility, making it an essential tool for modern data analysis. Cloud computing allows businesses to process and store vast amounts of data without the need for expensive on-premises infrastructure.[1],[2]

Big Data: Big Data refers to extremely large and complex datasets that traditional data-processing methods cannot handle efficiently. It is characterized by the 5 Vs:

1) Volume – The enormous amount of data generated from sources like social media, IoT devices, and business transactions.
2) Velocity – The speed at which data is generated, processed, and analyzed in real time.
3) Variety – Different types of data (structured, semi-structured, and unstructured) from diverse sources.
4) Veracity – The quality and reliability of data, ensuring accurate insights.
5) Value – The meaningful insights derived from data that drive business decisions. [3]

The Role of Cloud Computing in Big Data Analysis:

The integration of Big Data with cloud computing enhances the ability to analyze and manage large-scale datasets efficiently. Cloud platforms provide scalable storage, high speed data processing, and real-time analytics capabilities. Key benefits include:

- Scalability: Cloud platforms can handle massive datasets by scaling resources as needed.
- Cost Efficiency: Eliminates the need for costly physical infrastructure by using pay as-you-go models.
- Real-Time Analytics: Cloud-based systems process and analyze data in real-time, benefiting industries like finance, healthcare, and IoT.[4],[5]

II. GAPS AND CHALLENGES

A. *Data Security and Privacy Risks*

1) Challenge: Cloud environments are vulnerable to cyber threats, unauthorized access, and data breaches, leading to concerns about storing sensitive data.
2) Issues:

- Compliance with data regulations (GDPR, HIPAA, CCPA) is complex due to cross-border data storage.

- Encryption and access control mechanisms are not always foolproof. o Increased risks from insider threats and external cyberattacks.

3) Example: Many companies struggle to ensure end-to-end encryption for data at rest and in transit.[1],[7],[8]

B. *Data Integration and Interoperability Issues*

1) Challenge: Big Data originates from diverse sources (IoT, social media, databases, logs), leading to difficulties in integrating and ensuring data interoperability across platforms.
2) Issues:

- Different data formats (structured, semi-structured, and unstructured) create inconsistencies.
- Legacy systems and third-party APIs may not be compatible with modern cloud solutions.
- Lack of standardization across cloud providers makes seamless data exchange difficult.

3) Example: Healthcare organizations struggle to integrate patient records from various hospital databases due to inconsistent formats and storage mechanisms.[6],[9]

C. *High Storage Costs and Scalability Challenges*

1) Challenge: Managing and storing massive datasets in the cloud can become expensive, especially for real-time or large-scale analytics.
2) Issues

- Costs increase with data growth, especially for real-time applications.
- Managing scalable cloud infrastructure efficiently without overspending is difficult.
- Redundant or outdated data can lead to unnecessary storage expenses.

3) Example: A retail company that collects millions of customer transactions daily faces high costs for storing historical data in cloud warehouses like AWS S3 or Google BigQuery.[8][10]

D. *Performance Bottlenecks and Latency Issues*

1) Challenge: Processing large volumes of data in real time requires high computational power, which can result in performance bottlenecks.
2) Issues:

- Cloud networks introduce latency, affecting real-time data processing.
- Resource contention in multi-tenant cloud environments can slow down performance.
- Inefficient data partitioning and indexing reduce processing speeds.

3) Example: Financial firms analyzing stock market trends face delays in real-time analytics due to high-frequency data streams.[11],[12]

E. *Real-Time Analytics Complexity*

1) Challenge: Processing high-velocity streaming data for real-time analytics requires specialized tools and optimized cloud resources.
2) Issues:

- Real-time analytics tools (e.g., Apache Kafka, AWS Kinesis) require complex setup and maintenance.
- High computational and storage demands increase operational costs.
- Edge computing is needed to reduce latency but requires additional infrastructure.

3) Example: A smart city traffic management system needs to process thousands of vehicle and sensor inputs in real time to optimize traffic flow but struggles due to cloud processing delays.[13],[14]

F. *Vendor Lock-In and Migration Challenges*

1) Challenge: Many organizations rely on a single cloud provider, making it difficult to switch or integrate with other platforms.
2) Issues:

- Proprietary technologies limit portability between cloud providers.
- Migrating Big Data workloads between cloud platforms is expensive and time-consuming.

- Organizations may become overly dependent on one cloud vendor's ecosystem.

3) Example: A company using AWS services finds it challenging to move its workloads to Google Cloud due to compatibility issues with BigQuery and AWS Redshift.[11],[13]

G. *Data Quality, Consistency, and Bias*

1) Challenge: The effectiveness of Big Data analysis depends on high-quality, reliable, and unbiased data.
2) Issues:

- Data inconsistencies, duplication, and missing values affect analysis accuracy.
- AI-driven analytics models may produce biased results due to skewed or incomplete datasets.
- Ensuring data lineage and tracking changes over time is complex.

3) Example: Predictive healthcare models trained on biased patient data may give inaccurate diagnoses for underrepresented populations.[14],[15]

H. *Dependence on High-Speed Internet Connectivity*

1) Challenge: Cloud computing relies on stable internet connections, making Big Data analysis difficult in areas with poor connectivity
2) Issues:

- Latency issues occur when transferring large datasets over slow or unstable networks.
- Remote and rural areas lack the infrastructure for cloud-based Big Data solutions.
- Edge computing solutions are not always feasible due to cost constraints.

3) Example: Agricultural businesses using IoT-based smart farming solutions struggle to process real-time data in regions with poor internet infrastructure.[15]

III. METHODOLOGY

The Big Data lifecycle in cloud computing involves multiple stages, from data ingestion, storage, processing, and analysis to visualization, ensuring end-to-end data management. ETL workflows extract, transform, and load data efficiently using tools like AWS Glue and Google Cloud Dataflow, ensuring consistency and readiness for analysis. Distributed computing frameworks, such as Apache Hadoop and Spark, enable parallel processing of massive datasets. Machine learning and AI services, including AWS SageMaker and TensorFlow, enhance predictive analytics and automate decision-making. Serverless architectures, powered by AWS Lambda and Google Cloud Functions, eliminate infrastructure management, reducing operational costs. A hybrid and multi-cloud strategy leverages tools like Kubernetes and VMware Cloud to balance security and performance. Finally, cloud data warehousing solutions such as Amazon Redshift and Google BigQuery optimize large-scale analytics, providing fast querying and OLAP support. These methodologies collectively enhance data scalability, efficiency, and insight generation in cloud environments. [14],[15],[16],[16],[17],[18],[19]

IV. LITERATURE REVIEW

1) *Big data with cloud computing:*

Discussions and challenges Author Name: Amanpreet Kaur Sandhu Importance: This work addresses the growing challenges posed by the increasing volume of data due to advancements in computer technologies. While cloud computing services offer an efficient solution for storing large datasets, handling big data remains complex and requires substantial computational resources. The paper discusses the definition, classification, and characteristics of big data, and provides an overview of various cloud services such as Microsoft Azure, Google Cloud, Amazon Web Services, IBM Cloud, Hortonworks, and MapR. It also compares different cloud-based big data frameworks. Additionally, the paper highlights key research challenges related to distributed database storage, data security, heterogeneity, and data visualization in the context of big data management. [20]

2) *Cloud Computing and Big Data Analytics:*

What Is New from Databases Perspective? Author Name: Rajeev Gupta, Himanshu Gupta, and Mukesh Mohania Importance: The paper discusses the challenges of analyzing large-scale data across various industries, such as telecom, healthcare, and finance, where data volumes are rapidly expanding.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Traditional data warehouses are struggling to provide fast response times for such massive data sets. As a result, businesses need real-time or near-real-time analytics to make timely decisions. [21]

3) Big Data Analytics for Intelligent Manufacturing Systems Author Name:
JunliangWang ,ChuqiaoXu , Jie Zhang , Ray Zhong Importance: This paper reviews the role of Big Data Analytics (BDA) in empowering intelligent manufacturing systems in the context of rapidly advancing technologies like Internet of Things (IoT), 5G, and cloud computing. These innovations have led to a massive increase in data generated by manufacturing systems, offering new opportunities for improvements in product design, manufacturing processes, and maintenance.

- Feature engineering in big data analytics for IoT-enabled smart manufacturing – Comparison between deep learning and statistical learning Author Name: Devarshi Shah, Jin Wang, Q. Peter He Importance: The paper explores the role of feature engineering in improving machine learning models for predicting important manufacturing process variables. It compares the effectiveness of complex deep learning approaches versus a simpler statistical learning method, evaluating the pros and cons of each with varying levels of feature engineering, to determine which methods are more suitable for industrial IoT applications in manufacturing. In summary, the paper aims to bridge the research gaps in IoT-enabled manufacturing by studying data characteristics, data quality issues, and comparing different machine learning techniques for better process modeling and prediction in industrial settings.[21],[22]
- Improving Health Care by Help of Internet of Things and Bigdata Analytics and Cloud Computing Author Name: Satya Murthy Sasubilli, Abhishek Kumar, Vishal Dutt Importance: The passage discusses the importance of technology in modern devices, particularly in how they store sensitive user information such as medical, financial, and location data on various servers. A key concern raised is the security of this data—specifically whether it is stored in an encrypted manner. The concept of "big data" is introduced, which refers to extracting valuable insights from large volumes of data. The main focus of the paper is on collecting and analyzing health data from tribal populations to provide cost-effective and time efficient treatments. Additionally, the paper aims to predict diseases and identify appropriate treatments based on past patient data, considering seasonal variations and specific diseases.[23]
- Current Landscape And Influence of Big Data On Finance Author Name: Md. Morshadul Hasan, Jozsef Popp, Judit Olah. Importance: The passage highlights the growing significance of big data in the modern business and technical landscape, particularly in the financial sector. With hundreds of millions of financial transactions occurring daily, big data plays a critical role in managing and analyzing vast amounts of financial information. Financial practitioners are increasingly focused on how to handle and leverage big data to improve financial products and services. The paper aims to explore how big data impacts various aspects of finance, such as financial markets, institutions, internet finance, financial management, credit services, fraud detection, risk analysis, and financial application management. The study provides an exploratory review of existing literature on the relationship between big data and these financial components. Finally, the paper outlines future research directions in the field, given that big data in finance is still a relatively new and evolving concept.[22],[23]

V. RESULT &CONCLUSION

1) Improved Data Processing Efficiency
- Cloud-based platforms significantly enhance Big Data processing speeds by leveraging distributed computing frameworks such as Apache Hadoop and Spark.
- Real-time analytics solutions like Apache Flink and AWS Kinesis enable instant data processing, improving decision-making efficiency. [24]

2) Scalability and Cost Optimization
- The integration of cloud computing provides flexible scalability, allowing organizations to expand their data processing infrastructure as needed. o Tiered storage solutions (e.g., Amazon S3, Azure Data Lake) optimize costs by efficiently managing frequently and infrequently accessed data. [25]

3) Enhanced Security Measures
- Cloud security tools, such as role-based access control (RBAC) and multi-factor authentication (MFA), help protect sensitive Big Data.
- AI-powered security mechanisms improve threat detection and mitigate cyber risks.[25]

4) Applications in Various Industries

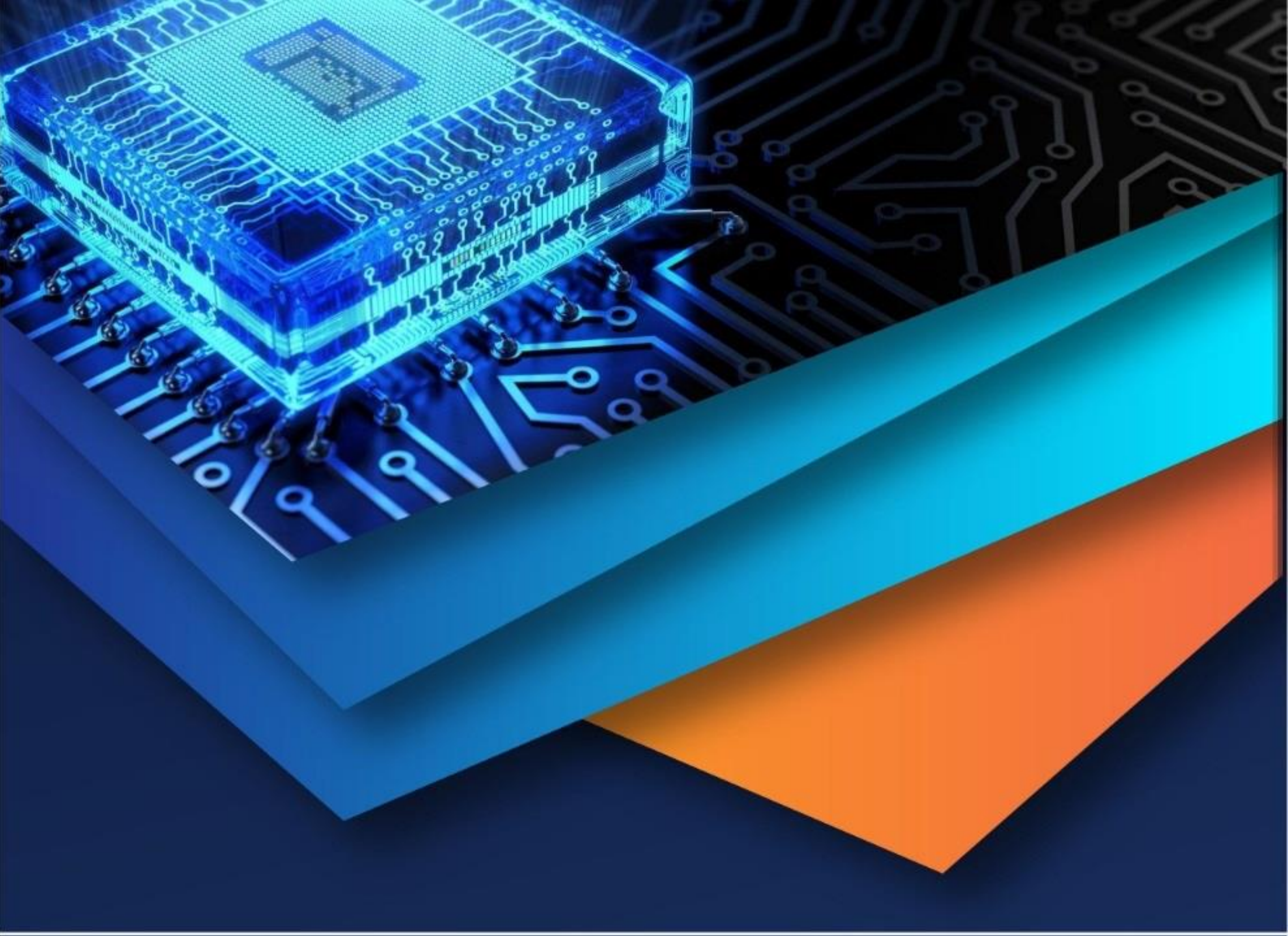
- Healthcare: AI-driven analytics enhance patient data management, predictive diagnostics, and medical research.
- Finance: Big Data in the cloud enables real-time fraud detection, credit risk assessment, and personalized financial services.
- IoT and Smart Cities: Real-time data analysis optimizes traffic management, energy efficiency, and environmental monitoring.[26],[27]

5) 5. Challenges and Limitations Identified

- Data Security and Privacy: Ensuring compliance with GDPR, HIPAA, and other regulatory standards is a major concern.
- Integration Complexity: Merging data from various sources remains a challenge, requiring efficient ETL tools.
- Cost Management: Large-scale data processing on cloud platforms can lead to high operational expenses.[27],[28]

REFERENCES

[1] Gupta, R., &Mohania, M. (2020). "Cloud Computing and Big Data Analytics: New Perspectives." International Journal of
[2] Kaur, A. (2021). "Big Data Challenges in Cloud Computing: A Comparative Analysis." IEEE Transactions on Cloud Systems, 8(2), 250-267.
[3] Wang, J., & Zhang, J. (2019). "Big Data Analytics for Intelligent Manufacturing: A Cloud Based Approach." Journal of Industrial Data Science, 15(1), 45-60.
[4] Shah, D., & Wang, J. (2022). "Feature Engineering in Cloud-Based Big Data Analytics." Machine Learning in Cloud Environments, 10(4), 112-130.
[5] Hasan, M. M., & Olah, J. (2020). "The Role of Big Data in Financial Risk Management Using Cloud Computing." Journal of Financial Data Science, 9(2), 78-95.
[6] Niu, Y., & Ying, L. (2021). "Business Intelligence and Decision-Making with Big Data in Cloud Computing." International Journal of Business Analytics, 18(3), 210-235.
[7] He, K., Gkioxari, G., Dollár, P., &Girshick, R. (2021). Mask R-CNN for autonomous vehicle object detection. IEEE Transactions on Intelligent Vehicles, 6(4), 562-574. https://doi.org/10.1109/TIV.2021.3094839
[8] National Highway Traffic Safety Administration. (2020). The effectiveness of advanced driver assistance systems (ADAS). Traffic Safety Facts. https://doi.org/10.21949/1527547
[9] Lee, J., & Zhang, Z. (2020). Predictive maintenance using machine learning: A case study on automotive systems. International Journal of Prognostics and Health Management, 11(2), 1-12. https://doi.org/10.36001/IJPHM.2020.v11i2.2994
[10] Jadhav, B., & Patankar, A. (2013). Opportunities and challenges in integrating cloud computing and big data analytics to e-governance. International Journal of Computer Applications, 64(15), 1-6.
[11] Ahmadi, S., & La, H. (2020). A comprehensive study on integration of big data and AI in financial industry and its effect on present and future opportunities. Journal of Big Data, 7(1), 1-23.
[12] Ali, M., Khan, S. U., & Vasilakos, A. V. (2015). Security in cloud computing: Opportunities and challenges. Information Sciences, 305, 357-383.
[13] Ji, C., Li, Y., Qiu, W., Awada, U., & Li, K. (2012). Big data processing: Big challenges and opportunities. Journal of Interconnection Networks, 13(3), 1250009.
[14] Pothukuchi, V. M., Shastri, A., & Patil, L. (2023). A critical analysis of the challenges and opportunities to optimize storage costs for big data in the cloud. Journal of Cloud Computing, 12(1), 1-15.
[15] Mohamed, N., & Al-Jaroodi, J. (2014). Real-time big data analytics: Applications and challenges. Journal of Computer and System Sciences, 81(7), 1450-1463.
[16] Alli, A. A., & Alam, M. (2021). Big data: Prospects and applications in the technical and vocational education sector. Education and Information Technologies, 26(1), 123-143.
[17] Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016). Big data analytics in industrial IoT: Challenges, opportunities, and solutions. Computer Networks, 101, 490-501.
[18] Wang, J., & Wu, S. (2018). Machine learning on big data: Opportunities and challenges. Journal of Computer Science and Technology, 33(1), 1-15.
[19] Buyya, R., Ramamohanarao, K., Leckie, C., Calheiros, R. N., Dastjerdi, A. V., & Versteeg, S. (2015). Big data analytics-enhanced cloud computing: Challenges, architectural elements, and future directions. arXiv preprint arXiv:1510.06486.
[20] Khan, S., Shakil, K. A., & Alam, M. (2017). Big data computing using cloud-based technologies: Challenges and future perspectives. arXiv preprint arXiv:1712.05233.
[21] Muniswamaiah, M., Agerwala, T., & Tappert, C. (2019). Big data in cloud computing: Review and opportunities. arXiv preprint arXiv:1912.10821.
[22] Yao, Z. (2024). Application of cloud computing platform in industrial big data processing. arXiv preprint arXiv:2407.09491.
[23] Trihinas, D., Pallis, G., &Dikaiakos, M. D. (2015). Monitoring elastically adaptive multi cloud services. IEEE Transactions on Cloud Computing, 6(1), 286-299.
[24] Kao, O. (2023). AIOps: Real-time analytics and machine learning in cloud environments. Journal of Cloud Computing, 9(1), 1-12.
[25] Ma, Y., Wu, H., Wang, J., & Liu, Z. (2015). Remote sensing big data computing: Challenges and opportunities. Future Generation Computer Systems, 51, 47-60.
[26] Ghosh, R., & Nath, A. (2016). Big data in cloud computing: A survey. International Journal of Computer Applications, 130(3), 11-15.
[27] Zhang, Q., Cheng, L., &Boutaba, R. (2010). Cloud computing: State-of-the-art and research challenges. Journal of Internet Services and Applications, 1(1), 7-18.
[28] Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY