

Big Data in Cloud Computing: Reviews and Opportunities

Krishan Pareek¹, Chaitanya Prakash², Prateek Raghav³, Umang Raj⁴ and Navjot Kaur⁵

¹⁻⁵Lovely Professional University / Department of Computer Science & Engineering, Jalandhar, India
Email: krishanpareek809@gmail.com, chaitanyaprakash000@gmail.com, prateekraghav226@gmail.com, umangraj7725@gmail.com, navjot.20506@lpu.co.in

Abstract—Big Data and Cloud Computing are pivotal technologies in today's landscape, addressing the exponential growth of data and providing flexible, cost-effective platforms for users. Integrating these technologies offers significant efficiency and knowledge advantages across industries like Finance, Management, and supply chain. This paper explores the fusion of Big Data and Cloud Computing, leveraging the 4 V's (Volume, Velocity, Variety, Veracity) of Big Data to enhance performance, insights, detail, and security while reducing costs. It delves into analytics, technologies, challenges, trends, and security considerations, showcasing the potential and benefits of this integration.

Index Terms— Big data, cloud computing, efficiency, analytics, database, data warehouse.

I. INTRODUCTION

The rise of mobile devices and multimedia content has led to a massive increase in data generation, both structured and unstructured. This data requires preprocessing before analysis. Industries such as healthcare, finance, engineering, e-commerce, and sciences leverage big data for better decision-making. Cloud computing has transformed big data storage and analysis by providing improved parallel processing, scalability, accessibility, data security, and resource virtualization. This eliminates the need for heavy investments in hardware or data centers. Virtualization enables the creation of virtual platforms for servers and storage devices, optimizing data access, management, analysis, and computation. This paper explores the potential opportunities and challenges of big data in cloud computing while proposing design principles for efficient data processing [1][2].

II. BIG DATA

Big Data refers to extremely huge large datasets that are challenging to store, manage, and analyze using traditional database methods. It originates from diverse sources like smartphones, social media, sensors, and more. To handle such vast data effectively, scalable architectures are necessary for storage, manipulation, and analysis. Integrating various technologies helps uncover valuable insights and gain a competitive edge through improved decision-making. Cloud platforms play a crucial role in providing the necessary data architecture, analytical methods, and tools for real-time capabilities, benefiting businesses significantly [3].

The concept of big data encompasses several important characteristics that differentiate it from traditional data sources. These characteristics are velocity, volume, variability, variety, value, veracity, validity, vulnerability,

visualization and volatility.

Volume pertains to the huge and vast amount of generated data from various sources.

Velocity pertains to the speed at which data is created, collected, and analyzed.

Variety means to the different data forms and structures.

Variability pertains to inconsistent flow of data, which can be affected by various factors.

Value pertains to the benefits that can be derived from analysis of big data.

Veracity pertains to the accuracy and reliability of the source of data, which affects its usefulness.

Validity pertains to the accuracy of the data for its intended purpose.

Vulnerability pertains to the data security, as it can be vulnerable to breaches and cyber-attacks.

Volatility pertains to how long data remains relevant and useful before it becomes outdated.

Visualization pertains to the creation of visual representations of data, which can help and utilize the insights derived from big data analysis.

Proper management of these characteristics is essential for successfully leveraging big data to drive business growth and improve decision- making processes [3].

III. CLOUD COMPUTING

The concept of cloud computing is based on delivering computing services like software, databases, servers, storage, networking, analytics and intelligence via the internet. This facilitates accelerated innovation, adaptable resource allocation, intensive computation, concurrent data processing, and economies of scale. This allows organizations to focus on their core business by abstracting computation, storage, and network resources to workloads as required, and leverage a wide range of pre-built services. Figure 3 illustrates the distinctions between the three main cloud service models: IaaS, PaaS and SaaS [5].



Fig. 1 : V's of Big Data

PaaS vs SaaS vs IaaS comparison			
	PaaS	SaaS	IaaS
Stands for	Platform as a Service	Software as a Service	Infrastructure as a Service
Target Users	Developers	Business users	System managers
User controls	Application Data	Nothing	Operating system, Runtime, Middleware, and Application data
Knowledge required	Knowledge of basic setup	No specific knowledge is required	Specific training required
Access	Runtime environment	User functionalities	Virtual machines and storage
Operational Cost	Low	Minimal	High
Security	Additional security needed to make sure applications don't expose vulnerabilities	Requires transparency in service provider's security policies to determine sensitive data	Should consider virtual and physical server security
Risk of vendor interlock	Medium	Highest	Lowest

Fig. 2 : Summary of Key Differences

IV. LITERATURE REVIEW

The combination of big data and cloud computing has sparked significant research interest. Cloud computing offers cost-effective and scalable data storage, processing, and retrieval over the internet. Big data involves vast datasets from diverse sources. Integrating these technologies addresses challenges in storing and analyzing such data efficiently. Qian et al. (2016) proposed a cloud-based big data platform that harnesses cloud computing capabilities for managing and analyzing large datasets effectively.

V. COMPARATIVE ANALYSIS

Big data and cloud computing are interconnected for managing and analyzing large datasets efficiently. Cloud computing provides scalable resources for data analysis, integrating services like SaaS, PaaS, and IaaS with big

data tools. Analytics as a service (AaaS) is a growing trend offering data analytics through the cloud for fast and scalable data processing. Virtualization optimizes data management by consolidating workloads onto virtual machines, reducing complexity and costs in data centers.

However, privacy and security concerns arise with third-party data hosting, especially in mobile health applications. Compliance with standards and agreements is crucial for data protection. Security measures are essential in cloud-based big data management to safeguard against intruders and ensure secure storage. The connection between big data and cloud computing adheres to an input-processing-output model, involving data collection, storage, cloud-based processing, and insights extraction. Cloud providers aid in data transmission and storage, while tools like Hadoop handle data processing, enabling real-time analysis of IoT data.

The relationship between cloud computing, IoT, and big data is depicted in Figure 4.

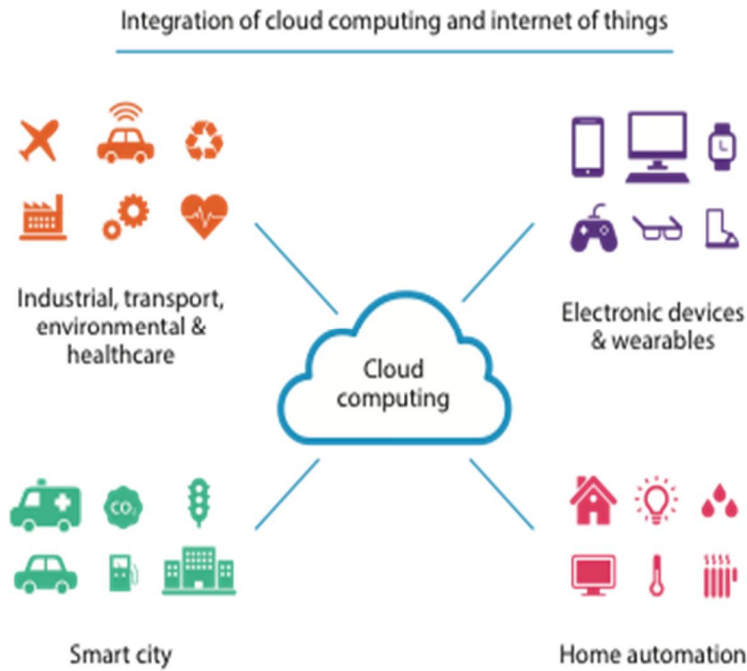


Fig. 3: Summary of IoT, Big Data processing and Cloud Computing

TABLE I. INFLUENCE OF BIG DATA IN PRESENT SCENARIO INDUSTRY WISE

Industry	Influence of Big Data	Benefits
Healthcare	Personalized medicine through data analytics, Predictive analytics for disease diagnosis and treatment planning, Monitoring patient health in real-time for proactive interventions, Fraud detection and prevention, Drug discovery and development	Improved patient outcomes, Early disease detection and prevention, Enhanced patient monitoring, Reduced healthcare fraud and abuse, Accelerated drug discovery process
Finance	Risk management and fraud detection through real-time data analysis, Algorithmic trading and investment strategies, Customer segmentation and targeted marketing, Compliance monitoring and reporting	Minimized financial risks, Improved investment decisions, Enhanced customer satisfaction and loyalty, Regulatory compliance
Retail	Customer behavior analysis for personalized marketing strategies, Inventory optimization and supply chain management, Predictive maintenance for equipment and infrastructure, Market trend analysis and forecasting	Increased sales and revenue, Reduced operational costs, Minimized downtime and maintenance costs, Competitive advantage
Manufacturing	Predictive maintenance for machinery and equipment, Quality control and defect detection through data analytics, Supply chain optimization and real-time tracking, Product lifecycle management	Improved equipment reliability and uptime, Enhanced product quality, Streamlined operations, Reduced time-to-market
Transportation	Route optimization and traffic management,	Reduced fuel consumption and emissions,

Industry	Influence of Big Data	Benefits
	Predictive maintenance for vehicles and infrastructure, Real-time monitoring of fleet operations, Demand forecasting and capacity planning	Increased safety and reliability of transportation systems, Improved logistics and delivery efficiency, Optimal resource utilization
Telecommunications	Network optimization and performance monitoring, Usage pattern analysis for service customization, Fraud detection and security enhancement	Enhanced network reliability and performance, Improved customer retention and loyalty, Tailored services and offerings, Reduced security risks and vulnerabilities
Energy	Smart grid management for efficient energy distribution, Predictive maintenance for power plants and infrastructure	Increased energy efficiency, Reduced maintenance costs and downtime, Sustainable energy practices
Agriculture	Precision farming for optimized crop yield and resource usage, Weather forecasting and risk management, Supply chain optimization for perishable goods, Pest and disease monitoring and control	Increased crop productivity and profitability, Mitigated weather-related risks and losses, Improved inventory management, Reduced crop losses and improved sustainability
Education	Personalized learning experiences based on student data, Predictive analytics for student performance and intervention, Administrative and operational efficiency through data-driven insights, Alumni engagement and fundraising strategies	Enhanced student engagement and performance, Early identification of at-risk students and targeted interventions, Streamlined administrative processes and resource allocation, Improved alumni relationships and fundraising outcomes

TABLE II. STATISTICAL COMPARISON OF BIG DATA INDUSTRY WISE

Industry	Benefit	Impact (1-5)
Healthcare	Improved patient outcomes	5
Finance	Minimized financial risks	5
Retail	Increased sales and revenue	4
Manufacturing	Improved equipment reliability and uptime	4
Transportation	Reduced fuel consumption and emissions	4
Telecommunications	Enhanced network reliability and performance	4
Energy	Increased energy efficiency	4
Agriculture	Increased crop productivity and profitability	5
Education	Enhanced student engagement and performance	4

VI. RESEARCH CHALLENGES

Cloud computing is popular for big data processing but faces challenges in data transmission, storage, transformation, quality, privacy, and governance. Issues include managing exponential data growth, real-time transmission, and adapting algorithms for volume, velocity, and variety. Storage technologies like DAS and NAS are used, yet scalability is difficult. Cloud systems need to be elastic to handle fluctuations in data and users while ensuring availability and data integrity. Security, privacy, and compliance are crucial, requiring efficient data uploading and recovery. Visualizing semi-structured and unstructured data also presents parallelization challenges.

VII. CONCLUSION AND FUTURE SCOPE

In the era of big data innovation driven by cloud computing advancements, hidden knowledge is unearthed from vast datasets. Big data and cloud computing are increasingly intertwined, with big data necessitating scalable computing platforms, which cloud computing readily provides. Conversely, cloud computing allows the processing and analysis of large datasets, yielding valuable insights. Big data revolutionizes data analysis by efficiently handling vast amounts of information. Future directions for big data in cloud computing include prioritizing data privacy and security, advancing machine learning and artificial intelligence, the emergence of edge computing, and increased adoption of cloud-native big data solutions.

REFERENCES

- [1] Konstantinou, I., Angelou, et al. On the elasticity of nosql databases over cloud management platforms. In Proceedings of the 20th ACM international conference on Information and knowledge management (pp. 2385-2388). ACM.
- [2] Labrinidis, Alexandros, et al "Challenges and opportunities with big data." Proceedings of the VLDB Endowment 5.12 (2012): 2032-2033.
- [3] Abadi, D. J. (2009). Data management in the cloud: Limitations and opportunities. IEEE Data Eng. Bull, 32(1), 3-12.
- [4] Luhn, H. P. (1958). A business intelligence system. IBM Journal of Research and Development, 2(4), 314-319
- [5] Sivarajah, Uthayasankar, et al. "Critical analysis of Big Data challenges and analytical methods." Journal of Business Research 70 (2017): 263-286.
- [6] <https://www.bmc.com/blogs/saas-vs-paas-vs-iaas-whats-the-difference-and-how-to-choose/>
- [7] Kavis, Michael J. Architecting the cloud: design decisions for cloud computing service models (SaaS, PaaS, and IaaS). John Wiley & Sons, 2014.
- [8] Sakr, S. & Gaber, et al., 2014. Large Scale and big data: Processing and Management Auerbach, ed.
- [9] Ji, Changqing, et al. "Big data processing in cloud computing environments." 2012 12th international symposium on pervasive systems, algorithms and networks. IEEE, 2012.
- [10] Zhang, L. et al., 2013. Moving big data to the cloud. INFOCOM, 2013 Proceedings IEEE, pp.405–409
- [11] http://acme.able.cs.cmu.edu/pubs/uploads/pdf/IoTBD_2016_10.pdf
- [12] Xiaofeng, Meng, et al. "Big data management: concepts, techniques and challenges [J]." Journal of computer research and development 1.98 (2013): 146-169.
- [13] Muniswamaiah, Manoj et al. (2019). Challenges of Big Data Applications in Cloud Computing. 221-232. 10.5121/csit.2019.90918.