International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 5, Sep-Oct-2025, ISSN (Online): 2395-566X

The Future is Cloud: Modernizing Big Data for the Cloud Era

Khaleel Khan Mohammed
MBA in Management Information Systems - Concordia University Wisconsin, United States,
Centene Corporation, Lead Data Engineer

Abstract- Data generation is increasing at an unprecedented pace across industries and the world. The challenge lies not only in storing and managing this massive "big data," but also in analyzing it to extract meaningful insights. To address this, various methods are employed for data collection, storage, processing, and analysis. This paper provides an overview of the layered architecture of Big Data management and highlights the key challenges within these layers that limit its practical applications across industries. In addition, the study explores different cloud-based architectural models that are designed to support diverse industrial requirements, emphasizing their role in enhancing scalability, flexibility, and efficiency. Furthermore, the paper discusses data migration strategies in detail, outlining how these approaches address the inherent limitations of Big Data systems by enabling seamless transfer, integration, and optimization of data in cloud environments.

Keywords – Big Data Architecture, Cloud Computing Acritecture, mining, Social network, Product ratin, Product recommendation.

I. INTRODUCTION

In the initial stages of condition monitoring technology, most systems were designed for specific equipment types [1,2], leading to fragmented and isolated implementations [3]. These standalone systems essentially acted as "information silos," lacking data sharing and interaction capabilities, which limited effective management and comprehensive analysis of monitoring information [4]. Moreover, the inability to share hardware resources—such as networking, computing power, and storage—resulted in inefficient use and waste of IT infrastructure [5]. To address these challenges, centralized management systems located in main control rooms were developed, enabling the unified processing of data collected from multiple monitoring devices [6]. With the advancement and adoption of high-speed optical fiber networks and wireless transmission technologies in the power sector, future monitoring centers are expected to gather real-time, large-scale condition data from geographically dispersed equipment, evolving into hubs for data integration and information sharing [6]. Consequently, the volume of data generated will become enormous, far exceeding the storage and processing capabilities of traditional monitoring systems [7]. Serial processing methods, in particular, are no longer adequate for handling such massive datasets [8].

From a business standpoint, Big Data was recognized early on as a transformative force for innovation, competitiveness, and productivity. The McKinsey report [9] highlighted its potential to increase business revenue streams and unlock new opportunities. For instance, [10] presented a four-phase framework for monetizing Big Data, while Monsanto's Climate Corporation applied geospatial analytics to model complex weather dynamics, providing farmers with insights to adapt to climate change. McKinsey further projected that Big Data could enhance performance in 60% of existing businesses and generate billions in new value over the following decade [9]. More broadly, Big Data has been reshaping the digital landscape, influencing diverse domains such as personalized medicine [12], customized product recommendations, and tailored travel planning [11]. In recent years, many of these possibilities have shifted from theory to practice through rapid technological innovation.

Research on Big Data processing has primarily focused on distributed and stream-based models [13]. Meanwhile, cloud computing—introduced slightly earlier than the Big Data boom—emerged as a revolutionary paradigm by offering computation as the "fifth utility" alongside water, electricity, gas, and telephony [14]. Its key features include elasticity, resource pooling, on-demand access, self-service, and a pay-as-you-go model [14]. These capabilities support various service layers: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

II. BIG DATA ARCHITECTURE

Broadly Big data management was structured into following set of layers [15, 16].

Layer	Purpose / Responsibilities
Data Sources / Ingestion (Data Acquisition)	Getting data from all origins. This includes structured, semi-structured, unstructured; batch or real-time; internal systems, external feeds, IoT, logs, etc. Ensuring connectivity, reliability, streaming or batch capturing.
Storage / Management	Upon ingestion, data must be stored in appropriate formats, possibly cleaned or transformed somewhat, with metadata, schema, indexing, partitioning. This layer deals with big-storage systems (data lakes, distributed file systems, NoSQL, etc.). Data governance, schema, metadata, data cataloging often start here.
Processing / Compute	Transforming, cleaning, aggregating, enriching data. Includes batch processing, stream processing, ETL/ELT, possibly also real-time analytics. May include feature extraction, data wrangling, etc.
Analytics / Query / Insight Generation	Once data is processed and in a usable form, this layer involves querying, advanced analytics, machine learning, BI tools, statistical modeling, predictive analytics, etc. Tools for exploring data.
Visualization / Consumption	Presentation of results to stakeholders: dashboards, reports, interactive visualizations, alerts; also APIs or applications that consume data or analytics outputs. Enables decision making.

Above architecture is facing many issues that direct;y impats the industries operation some of major issues are:
High Infrastructure Cost

- On-premises big data systems (Hadoop clusters, HDFS storage) require heavy upfront investment in servers, networking, and storage.
- Maintenance and hardware refresh cycles add ongoing costs.

Limited Scalability

- Scaling on-premise big data clusters means buying and configuring new hardware, which is slow and expensive.
- Traditional architectures often fail to handle sudden spikes in workload or massive IoT data streams.

Complexity in Setup and Maintenance

- Deploying and managing frameworks like Hadoop, Spark, and Kafka on-premises is complex.
- Requires skilled administrators for cluster management, fault-tolerance, job scheduling, and upgrades.

Rigid and Inflexible

- Traditional architectures are not agile; they cannot easily adapt to changing data formats (structured, semi-structured, unstructured).
- Integrating new data sources or analytics tools requires significant re-engineering.

Storage Limitations

- Local storage and HDFS are not as cost-efficient or durable as cloud object storage (e.g., Amazon S3, Azure Blob).
- Data replication across multiple data centers is difficult and costly.

Performance Bottlenecks

- On-premises big data systems often struggle with latency in real-time analytics.
- Hardware constraints limit compute power, I/O throughput, and network speed.

Lack of Elasticity

- Resources are fixed; clusters cannot shrink during low demand or expand instantly for peak workloads.
- Leads to underutilization during idle times or failure during high-load periods.

Challenges in Data Integration

- Traditional big data architectures struggle with multi-source integration (IoT devices, logs, streaming, cloud SaaS).
- ETL pipelines become bulky and inefficient.

Security and Compliance Burden

- Organizations are responsible for securing infrastructure, data governance, and compliance.
- No shared responsibility model like in the cloud.

Slow Innovation Cycle

- New tools and frameworks emerge rapidly, but integrating them into existing on-prem systems is slow and costly.
- Cloud-native services provide ready-to-use big data analytics, AI/ML pipelines, and serverless options.

III. INDUSTRY ARCHITECTURAL DEVELOPMENT

Hybrid Cloud: According to Spectrum Enterprise, most organizations today adopt a hybrid Information Technology (IT) infrastructure, unless their systems are built entirely on physical servers or are fully cloud-based [17]. A Gartner study predicted that by 2017, nearly half of mainstream enterprises would operate within a hybrid environment. Hybrid cloud models allow organizations to integrate and manage on-premise private cloud solutions with public cloud services, offering greater flexibility. This model enables businesses to evaluate the specific requirements of each task and deploy the most suitable environment for execution [18]. For instance, applications with fluctuating demands for network resources are better supported on the public cloud, while workloads requiring consistent, high-level resources are better suited to private clouds. Hybrid cloud designs can vary in complexity; for example, some enterprises integrate SaaS-based expense tracking tools with internal billing systems, reflecting a hybrid implementation.

Private Clouds The distinction between private Infrastructure-as-a-Service (IaaS) clouds and traditional server virtualization was discussed in [18]. Virtualization forms the foundation of all IaaS-based clouds, and management tools can transform virtualized environments into cloud-like systems. However, several differences set IaaS cloud infrastructures apart from virtual data centers [19]:

- Standardized environments reduce costs and streamline operations.
- Automation is a core element of cloud systems, with standardized tasks executed using automated tools.
- Self-service access is integral, giving cloud consumers workflow-based permissions without direct administrative involvement.
- Multitenancy allows for isolation in private clouds, maximizing resource use while maintaining cost efficiency.
- Clouds can be understood as part of an IT portfolio, with varying levels of automation, standardization, and deployment models. Enterprises should expect only about 15% of their applications to be cloud-ready initially, with gradual expansion over time.
- Building a private cloud should begin on a smaller scale, enabling organizations to recognize its advantages and ensure optimal resource utilization before scaling.
- Since cloud environments are shared, expansion should be demand-driven, with customers justifying investment growth.

Cloud Computing Survey [20]: Survey results show that organizations are increasingly shifting IT operations to the cloud, using a combination of public, private, and hybrid solutions. On average, 45% of IT workloads had already been migrated, with 23% on private clouds, 15% on public clouds, and 7% on hybrid models. By the end of 2017, companies were projected to migrate 59% of their IT activities, distributed as 28% private, 22% public, and 10% hybrid. Budget allocations reflected this trend: 45% was directed to Software-as-a-Service (SaaS), 30% to Infrastructure-as-a-Service (IaaS), 19% to Platform-as-a-Service (PaaS), and 6% to emerging models such as Network-as-a-Service (NaaS) or Database-as-a-Service (DBaaS). Large enterprises planned to dedicate about 21% of their budgets to PaaS compared to 17% among small and medium-sized businesses (SMBs). Conversely, SMBs were expected to spend over 75% of their budgets on SaaS adoption.

Cloud Migration Report [21]: Statistics indicated a significant year-over-year rise in cloud adoption, with public cloud usage expected to grow by 28% and private cloud usage by 15%. For larger enterprises with more than 1,000 employees, adoption rates were projected to be even higher, with public cloud increasing by over 49% and private cloud by 18%.

IV. RELATED WORK

Yibin Li et al. [22] proposed a distributed storage approach integrated with strong security-awareness mechanisms to enhance the safety of big data storage in cloud computing environments. Their method employed techniques such as alternative data distribution, secure data distribution, and efficient data conflation. In this system, files were divided into segments and distributed separately across different cloud servers. The alternative distribution strategy was introduced to determine whether data packets should be split, thereby reducing overall operational time. This security-aware framework not only strengthened cloud defenses against potential threats but also lowered computational overhead. Nevertheless, the use of encryption for all data through security algorithms introduced notable performance challenges, particularly in sensitive domains such as healthcare organizations.

Similarly, Zhicheng Zhou et al. [23] designed a cloud-based framework for big data processing with the aim of optimizing multimedia interaction. Their method integrated radial basis function neural networks within the cloud, applying MapReduce techniques to handle computation. By combining MapReduce with the error back-propagation algorithm, the system enabled efficient mapping of multi-layer neural networks, facilitating multimedia data processing. The model demonstrated advantages such as reduced iterations, faster convergence, and improved acceleration. However, the parallelization technique used with the neural network showed limitations when applied to large-scale distributed data environments.

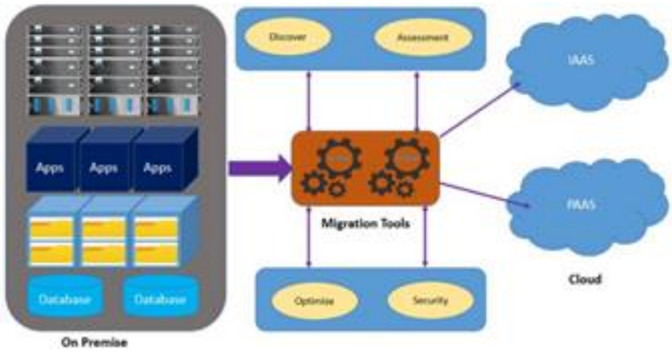

Fig. 1 Big Data to Cloud Migration architecture.

As illustrated in Fig. 1, developing a well-structured migration architecture is a critical prerequisite before transferring big data systems to the cloud. Such an architecture must clearly define cloud requirements, allocate resources, and outline strategies for seamless data transfer. Since cloud migration requires detailed technical planning and systematic design, organizations need to establish the role of a migration architect. This specialist is tasked with creating the architectural blueprint, overseeing resource planning, and coordinating implementation steps. Ultimately, the migration architect ensures a successful transition by formulating essential strategies and supervising their execution [24].

V. CONCLUSION

Big Data has emerged as a transformative force across industries, yet its adoption is often hindered by architectural challenges such as storage complexity, data silos, inefficient processing methods, and limited scalability. The layered architecture of Big Data management provides a structured framework for data collection, storage, processing, and analysis, but existing limitations reduce its effectiveness in large-scale and real-time applications. To overcome these barriers, cloud computing offers a more adaptive solution through its flexible architectures, resource pooling, and scalable infrastructure.

By leveraging private, public, and hybrid cloud models, organizations can address the shortcomings of traditional Big Data systems while enhancing efficiency and performance. Furthermore, effective data migration strategies enable seamless integration of existing data ecosystems into cloud environments, reducing overhead costs and unlocking the potential for advanced analytics. In summary, the convergence of Big Data and cloud computing is not merely a technological trend but a necessity for industries seeking innovation, operational efficiency, and sustainable growth. The findings of this paper highlight that while Big Data alone faces inherent limitations, its integration with cloud architectures paves the way for scalable, secure, and value-driven applications in the digital era.

REFERENCES

1. Lin, W.; Wu, G.; Wang, X.; Li, K. An artificial neural network approach to power consumption model construction for servers in cloud data centers. IEEE Trans. Sustain. Comput. 2019, 5, 329–340.
2. Liu, Y.; Liang, S.; He, C.; Zhou, Z.; Fang, W.; Li, Y.; Wang, Y. A Cloud-computing and big data based wide area monitoring of power grids strategy. IOP Conf. Ser. Mater. Sci. Eng. 2019, 677, 042055.
3. Raza, M.Q.; Khosravi, A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renew. Sustain. Energy Rev. 2015, 50, 1352–1372.
4. Santos, M.A.G.; Munoz, R.; Olivares, R.; Filho, P.P.R.; Del Ser, J.; de Albuquerque, V.H.C. Online heart monitoring systems on the internet of health things environments: A survey, a reference model and an outlook. Inf. Fusion 2020, 53, 222–239.
5. Sharma, S.; Kotturu, P.K.; Narooka, P.C. Implication of IoT components and energy management monitoring. In Swarm Intelligence Optimization: Algorithms and Applications; John Wiley & Sons, Inc.: New York, NY, USA, 2020; pp. 49–65.
6. Manimegalai, C.T.; Kalimuthu, K.; Gauni, S. Toward integrating bidirectional multiband data and power transmission using double clad optical fibers for the next generation disaster resilient managing communication systems. Microw. Opt. Technol. Lett. 2022, 64, 816–820.
7. Jia, X. Research on network abnormal data flow mining based on improved cluster analysis. Distrib. Parallel. Databases 2021, 40, 797–813.
8. Ahamed, R.; Habeeb, A.; Nasaruddin, F.; Gani, A.; Abaker, I.; Hashem, T.; Ahmed, E.; Imran, M. International Journal of Information Management Real-time big data processing for anomaly detection: A Survey. Int. J. Inf. Manag. 2018, 45, 289–307.
9. Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. 2011. Big data: The Next Frontier for Innovation, Competition, and Productivity. Accessed November 25, 2015.
10. Najjar, M. S., and W. J. Kettinger. 2013. "Data Monetization: Lessons from a Retailer's Journey." MIS Quarterly Executive 12 (4): 1–13.
11. Mayer-Schönberger, V., and K. Cukier. 2013. Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt. (Chapter 1).
12. Alyass, A., M. Turcotte, and D. Meyre. 2015. "From Big Data Analysis to Personalized Medicine for All: Challenges and Opportunities." BMC Medical Genomics 8: 1–33.

13. Zikopoulos, P., and C. Eaton. 2011. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill Osborne Media (Chapter 2).
14. Mell, P. M., and T. Grance. 2011. The NIST Definition of Cloud Computing. Special Publication 800-145, National Institute of Standards and Technology, Gaithersburg, MD.
15. Naamane, Z., & co. (2023). Benefits and challenges of cloud-based big data analytics. Issues in Information Systems, 24(1), 291-304.
16. Ismail, F. N., et al. (2025). Big Data Architecture for Large Organizations. arXiv preprint arXiv:2505.04717.
17. Spectrum Enterprise (2016) "10 Things You Need to Know About Hybrid Cloud", TechNavio, "Global Hybrid Cloud Market," available at http://www.technavio.com/report/global-hybrid-Cloud-market-2015-2019.
18. Marissa Comeau (2014) "Hybrid cloud: What it is, what it does and hybrid cloud use cases" TechTarget Publication. Available at http://searchservervirtualization.techtarget.com/tip/Hybrid-cloud-What-it-is-what-it-does-and-hybrid-cloud-use-cases.
19. C. W. Buyer, "Private cloud: get it right and exploit its energy economics," pp. 1–6, 2012.
20. IDG Enterprise. (2017). 2016 IDG Enterprise Cloud Computing Survey. [online] Available at: https://www.idgenterprise.com/resource/research/2016-idg-enterprise-cloud-computing-survey.
21. CloudEndure, "2017 Cloud Migration Survey Report Benchmarks, Trends," 2017.
22. Li, Y.; Gai, K.; Qiu, L.; Qiu, M.; Zhao, H. Intelligent cryptography approach for secure distributed big data storage in cloud computing. Inf. Sci. 2017, 387, 103–115.
23. Zhou, Z.; Zhao, L. Cloud computing model for big data processing and performance optimization of multimedia communication. Comput. Commun. 2020, 160, 326–332.
24. Look, L. (2021, February 10). A roadmap for migrating legacy tape storage to the cloud