Shodhbodhalaya: An International Peer reviewed
Multidisciplinary Journal
Volume 4 | Issue 1 | ISSN: 2584-1807 | (https://shodhbodh.com/)

A COMPARATIVE STUDY OF BIG DATA PLATFORMS AND CLOUD COMPUTING PLATFORMS: EVALUATING SCALABILITY, SECURITY, AND ANALYTICAL CAPABILITIES

Shweta Kashyap[1], Dr. Gopal Jee[2]

Research Scholar[1], Professor and Head of the Department[2]

Faculty of Science (Computer Application)[1], Department of Physics[2]

Babasaheb Bhimrao Ambedkar Bihar University[1], Langat Singh College, Muzaffarpur, Bihar[2]

Abstract

The rapid growth of digital technologies and data-driven decision-making has significantly increased the importance of efficient data processing platforms. Organizations across industries generate massive volumes of structured and unstructured data that require advanced systems for storage, management, and analysis. Big Data platforms and Cloud Computing platforms have emerged as two prominent technological solutions that support large-scale data processing and analytics. While Big Data platforms are designed specifically to handle large and complex datasets through distributed computing frameworks, Cloud Computing platforms provide flexible and scalable infrastructure that enables organizations to store, process, and access data remotely. This study presents a comparative analysis of Big Data platforms and Cloud Computing platforms with a focus on their scalability, security mechanisms, and analytical capabilities. The research examines how these platforms support organizations in managing high-volume data environments and enabling advanced analytics for informed decision-making. Furthermore, the study evaluates the adaptability of these platforms in handling large-scale data processing tasks across different business contexts. By identifying the strengths and limitations of each platform, the paper aims to provide insights that can assist organizations in selecting the most suitable technological framework for their data analytics requirements. The findings contribute to a better understanding of how modern enterprises can leverage data platforms to enhance operational efficiency, innovation, and strategic planning.

Keywords

Big Data Platforms; Cloud Computing; Data Analytics; Scalability; Data Security; Distributed Computing; Large-Scale Data Processing; Business Intelligence

Introduction

The rapid advancement of information technology has led to an unprecedented growth in the volume, velocity, and variety of data generated across the globe. Organizations in sectors such as healthcare, finance, retail, education, and government produce massive amounts of digital information through online transactions, sensors, social media platforms, and enterprise systems. Managing and analyzing such large-scale data has become a major challenge for traditional data processing systems. As a result, advanced technological frameworks such as Big Data platforms and Cloud Computing platforms have gained considerable attention in recent years.

Big Data platforms are designed to process and analyze extremely large datasets that cannot be efficiently handled by conventional database systems. Technologies such as distributed storage systems, parallel computing frameworks, and advanced analytics tools enable organizations to extract valuable insights from complex datasets. These platforms support large-scale data processing by distributing computational tasks across multiple nodes, thereby improving performance and efficiency.

On the other hand, Cloud Computing platforms provide on-demand access to computing resources such as storage, servers, networking, and applications over the internet. Cloud technologies allow organizations to scale their infrastructure according to demand while reducing the need for expensive hardware investments. The flexibility and cost-effectiveness of cloud services have made them an attractive solution for organizations seeking to manage large volumes of data efficiently.

Both Big Data and Cloud Computing platforms play a crucial role in modern data-driven environments. However, they differ in terms of architecture, processing capabilities, security frameworks, and analytical performance. Understanding these differences is essential for organizations that aim to adopt the most appropriate technological platform for their data analytics needs.

This study aims to compare and analyze Big Data platforms and Cloud Computing platforms by examining their scalability, security mechanisms, and analytical capabilities. The research also explores how these platforms adapt to large-scale data processing requirements and how they support organizations in making data-driven decisions. By providing a comprehensive comparative analysis, the study contributes to a deeper understanding of the role of advanced computing platforms in modern business environments.

1. Background of Big Data and Cloud Computing

Big Data emerged as a response to the growing inability of conventional database systems to manage the scale, speed, and complexity of modern digital information. It is generally associated with datasets whose size, diversity, and rate of generation demand new storage models, distributed processing methods, and advanced analytical techniques beyond traditional computing environments (Chang et al., 2019). As the digital economy expanded, organizations began generating data from social media, sensors, business transactions, mobile applications, and internet-based platforms at an unprecedented rate. This rapid growth made it difficult for traditional systems to capture, store, and analyze information efficiently. Consequently, Big Data became an important technological paradigm for extracting meaningful patterns, trends, and insights from massive datasets. Its relevance has increased significantly in sectors such as healthcare, finance, retail, education, and governance, where data-driven decision-making has become essential. Cloud Computing developed as a service-oriented model that allows users to access computing resources such as servers, storage, and applications over networks on demand. Its rise has transformed the way organizations acquire and use information technology by shifting from fixed infrastructure ownership to flexible, service-based consumption (Mell & Grance, 2011). This model enables businesses to reduce capital investment in physical hardware while improving operational efficiency and accessibility. Cloud services also provide elasticity, meaning that organizations can scale resources up or down according to workload requirements.

2. Evolution of Data Processing Technologies

The historical evolution of data processing moved from centralized mainframes to client-server systems and then toward distributed architectures capable of supporting large-scale computation. A major milestone in this transition was the development of distributed storage systems such as the Google File System, which demonstrated that large datasets could be stored reliably across clusters of low-cost machines (Ghemawat et al., 2003). Earlier computing systems were suitable for limited and structured data, but they lacked the flexibility and scalability needed for rapidly increasing digital workloads. As organizations began handling more complex and voluminous data, distributed architectures provided a practical alternative by improving storage reliability, fault tolerance, and processing efficiency. This transition laid the foundation for modern large-scale data infrastructures used in analytics and enterprise computing environments today. The next major step came with distributed programming models such as MapReduce, which simplified the parallel processing of massive datasets. This shift enabled organizations to process data across many machines efficiently, making large-scale analytics more practical, affordable, and robust than in earlier computing models (Dean & Ghemawat, 2004). By dividing tasks into smaller sub-processes and executing them simultaneously, distributed programming significantly reduced computational time and improved system performance.

3. Concept and Characteristics of Big Data Platforms

Big Data platforms are specialized environments designed to ingest, store, manage, and analyze very large datasets across distributed systems. Their defining characteristics include fault tolerance, parallel processing, high-throughput storage, and the ability to operate on heterogeneous structured and unstructured data sources. Hadoop's HDFS architecture is one of the best-known examples of this design logic. HDFS is specifically designed to run on commodity hardware, provide high-throughput access to application data, and support very large datasets through a distributed file system model. This makes Big Data platforms highly suitable for organizations that need reliable and scalable storage for massive and continuously growing data resources (Apache Hadoop, 2023a). These platforms usually include an ecosystem of components for storage, resource management, and batch or stream processing. In the Hadoop model, for example, HDFS supports distributed storage, YARN handles resource coordination, and MapReduce or related engines execute large computational jobs, together forming a scalable big data framework. YARN further strengthens this architecture by separating resource management from job scheduling and monitoring, thereby improving flexibility and cluster efficiency.

4. Concept and Architecture of Cloud Computing Platforms

Cloud Computing platforms are built around the principle of delivering configurable computing resources as services over the internet. Their core characteristics include on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service, which together distinguish cloud environments from conventional in-house infrastructure models. These features enable organizations to access computing power as needed without maintaining extensive physical infrastructure of their own. This service-based model has become especially attractive for businesses seeking flexibility, scalability, and cost efficiency in their digital operations (Mell & Grance, 2011). From an architectural perspective, cloud platforms are organized around service providers, consumers, brokers, auditors, and carriers, with service orchestration and management functions coordinating the delivery of infrastructure, platform, and software services. This layered design supports flexibility, standardization, and efficient service governance in cloud ecosystems. The NIST cloud reference architecture also emphasizes service deployment, orchestration, cloud service management, and security and privacy as central architectural elements. Such a structure helps organizations understand roles, responsibilities, and operational control within cloud-based environments, making cloud platforms more systematic and manageable for enterprise adoption (Liu et al., 2011).

5. Role of Big Data in Modern Data Analytics

Big Data platforms play a central role in modern analytics because they allow organizations to work with batch, streaming, interactive, and machine learning workloads in a unified environment. Systems such as Apache Spark extended the earlier big data model by enabling faster in-memory computation and broader support for analytics applications. This integration has made it possible for organizations to process massive datasets more efficiently and derive insights in near real time. As a result, big data platforms have become essential for firms that need speed, flexibility, and analytical depth in complex data environments (Zaharia et al., 2016). In contemporary organizations, big data analytics is no longer limited to retrospective reporting; it increasingly supports predictive and prescriptive decision-making. Recent systematic reviews show that predictive analytics built on large and complex datasets has become important for forecasting, optimization, and strategic planning across many domains. This shift reflects the growing importance of data-driven intelligence in helping organizations anticipate trends, improve performance, and make more accurate strategic decisions. Thus, Big Data now serves as a foundation for advanced analytical practices in both public and private sectors (Jamarani et al., 2024).

6. Importance of Cloud Computing in Data Management

Cloud Computing has become highly important in data management because it offers standardized, remotely accessible infrastructure for storing, moving, and processing data. It supports a utility-like model in which organizations can provision resources according to operational demand rather than maintaining permanently overbuilt local systems. This approach improves cost efficiency, operational agility, and accessibility, especially for businesses with changing workload requirements. It also reduces the need for large upfront investment in hardware and maintenance, making digital infrastructure more flexible and manageable (NIST, 2013). Cloud environments also support modern data management architectures such as data lakes, where structured and unstructured data can be stored at scale before being processed for dashboards, real-time analytics, and machine learning. This makes cloud platforms especially useful for organizations dealing with continuously growing and diverse data assets. By centralizing data in scalable repositories, cloud systems enable better integration, faster analysis, and improved support for innovation and business intelligence. Such capabilities are particularly valuable in data-intensive industries that rely on timely and accurate insights for decision-making (AWS, n.d.).

7. Scalability Features in Big Data and Cloud Platforms

Scalability is one of the most important criteria in comparing Big Data and Cloud Computing platforms. In cloud-based analytics environments, scalability is often achieved through elastic resource allocation, automation, and repeatable architectural patterns that allow organizations to expand workloads without redesigning the whole system. This flexibility helps organizations respond quickly to increasing data volumes, user demand, and analytical complexity. It also supports cost optimization because resources can be matched more closely with actual workload requirements rather than fixed long-term provisioning. In a broader platform sense, scalability also refers to the ability to add or remove resources in response to demand while maintaining service quality and availability. This elastic behavior is central to both cloud-native data systems and large-scale analytics environments where workload intensity changes over time. Effective scalability therefore improves not only technical performance but also business continuity, responsiveness, and resource efficiency. It enables organizations to maintain stable operations even during spikes in traffic, storage demand, or computational load (Oracle, 2025).

8. Security Challenges in Big Data and Cloud Environments

Security remains one of the most critical issues in cloud environments because outsourcing infrastructure and applications introduces concerns related to control, privacy, trust, multi-tenancy, and compliance. NIST guidelines emphasize that cloud adoption must be accompanied by careful consideration of security and privacy requirements throughout the system life cycle. These concerns become especially significant when sensitive organizational or customer data is stored and processed outside traditional local infrastructure. As a result, cloud adoption requires strong governance, risk assessment, and continuous security monitoring to ensure dependable and compliant service use (Jansen & Grance, 2011). Big Data environments face additional security and privacy complexities because of the scale, diversity, and distributed nature of the data they handle. NIST's Big Data security framework highlights the need for risk management, privacy-aware architectures, access control, and domain-specific safeguards across the data lifecycle. Because big data systems often combine multiple data sources and analytical processes, vulnerabilities may arise at the stages of collection, storage, transmission, processing, and sharing. Therefore, security in big data platforms must be treated as an integrated architectural concern rather than a separate technical add-on (NIST, 2019).

9. Analytical Capabilities of Modern Data Platforms

Modern data platforms are valued not only for storage and processing but also for the analytical capabilities they provide. Cloud security and governance frameworks increasingly recognize that analytics platforms must support secure operations while enabling monitoring, policy enforcement, and workload management across complex digital environments. This reflects a shift from viewing data systems as passive repositories to treating them as intelligent environments that generate insights while maintaining control and accountability. As organizations become more data-driven, analytical capabilities are now considered a core requirement of platform design (Cloud Security Alliance, 2017). In addition, major cloud architecture guidance now treats security, governance, and analytics as integrated design concerns rather than separate functions. This is important because organizations increasingly expect data platforms to combine performance, intelligence, and compliance within a single operational framework. Such integration helps improve decision quality, regulatory alignment, and overall platform resilience. It also ensures that analytics initiatives remain sustainable and manageable as systems grow more complex and business dependence on data intensifie.

10. Integration of Big Data and Cloud Computing Technologies

The integration of Big Data and Cloud Computing has created powerful hybrid models in which scalable cloud infrastructure supports distributed analytics engines. Studies on cloud-based big data analytics show that platforms such as BigQuery and related services allow organizations to process very large datasets through managed, serverless, and highly scalable environments. This integration reduces infrastructure management burdens while improving access to advanced querying, storage, and analytical functions. As a result, organizations can focus more on insight generation and less on maintaining complex hardware and software stacks (Google Cloud, 2026). This integration is also visible in hybrid cloud models, where public cloud, private cloud, and on-premises systems are combined into a unified environment. Such arrangements are increasingly used when organizations need flexibility, local control, regulatory compliance, and large-scale analytical capacity at the same time. Hybrid integration is particularly useful for enterprises that cannot move all operations to a single cloud environment due to legacy systems, security requirements, or data sovereignty concerns. In practice, it offers a balanced model for modern digital transformation and enterprise-scale data management.

11. Applications of Big Data and Cloud Platforms in Business Organizations

In business organizations, Big Data and cloud platforms are applied across customer relationship management, e-commerce, accounting, human resources, and operational planning. Literature reviewing business applications of big data shows that firms use these technologies to gain both financial and non-financial benefits through better insight generation and process improvement. These platforms support more accurate forecasting, stronger customer understanding, and better coordination across business units. They also help organizations improve decision speed and operational visibility in competitive and data-intensive markets (Alnoukari, 2016). Their importance is also increasing in supply chains and other competitive business functions, where analytics capabilities can strengthen visibility, responsiveness, and decision quality. Research indicates that digital technologies such as big data analytics help firms build competitive advantage and improve operational performance when adopted strategically. This means that the value of these platforms lies not only in technical efficiency but also in their contribution to innovation, adaptability, and long-term organizational competitiveness. Businesses that align analytics capabilities with strategic goals are more likely to achieve sustained performance benefits (Korayim et al., 2024).

REFERENCE

Alnoukari, M. (2016). Business intelligence and analytics: The case of big data. *International Journal of Business Intelligence Research, 7*(1), 1–15.

Apache Hadoop. (2023a). *HDFS architecture guide*. Apache Software Foundation.

Chang, W. L., Grady, N., & NIST Big Data Public Working Group. (2015). *NIST Big data interoperability framework: Volume 1, definitions* (NIST Special Publication 1500-1). National Institute of Standards and Technology.

Cloud Security Alliance. (2017). *Security guidance for critical areas of focus in cloud computing v4.0*. Cloud Security Alliance.

Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified data processing on large clusters. In *Proceedings of the 6th Symposium on Operating Systems Design and Implementation* (pp. 137–150).

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM, 51*(1), 107–113.

Ghemawat, S., Gobioff, H., & Leung, S.-T. (2003). The Google file system. In *Proceedings of the 19th ACM Symposium on Operating Systems Principles* (pp. 29–43).

Google Cloud. (n.d.). *BigQuery documentation*. Google Cloud. Retrieved March 16, 2026, from Google Cloud website

Jamarani, A., Haddadi, S., Akbarzadeh, M., Moradi, S., & Mousavi, S. M. (2024). Big data and predictive analytics: A systematic review of applications. *Artificial Intelligence Review, 57*, Article 299.

Jansen, W., & Grance, T. (2011). *Guidelines on security and privacy in public cloud computing* (NIST Special Publication 800-144). National Institute of Standards and Technology.

Korayim, A. S., El Baz, J., Ghanem, M. M., & Salem, M. A. (2024). Big data analytics and competitive advantage: The role of operational performance and strategic adoption. *Technological Forecasting and Social Change, 199*, 123050.

Liu, F., Tong, J., Mao, J., Bohn, R., Messina, J., Badger, L., & Leaf, D. (2011). *NIST cloud computing reference architecture* (NIST Special Publication 500-292). National Institute of Standards and Technology.

Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing* (NIST Special Publication 800-145). National Institute of Standards and Technology.

National Institute of Standards and Technology. (2013). *NIST cloud computing standards roadmap* (NIST Special Publication 500-291r2). U.S. Department of Commerce.

National Institute of Standards and Technology. (2019). *NIST Big data interoperability framework: Volume 4, security and privacy* (NIST Special Publication 1500-4r2). U.S. Department of Commerce.

Simmon, E., Reynolds, M., & NIST Cloud Computing Program. (2018). *Evaluation of cloud computing services based on NIST SP 800-145* (NIST Special Publication 500-322). National Institute of Standards and Technology.

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J. (2016). Apache Spark: A unified engine for big data processing. *Communications of the ACM, 59*(11), 56–65.