



AWS vs. Azure vs. Google Cloud for Data Science

Bharathram Nagaiah
Independent Researcher, USA.

Received On: 07/02/2026 **Revised On:** 18/03/2026 **Accepted On:** 26/03/2026 **Published On:** 02/04/2026

Abstract: This paper analyzes and compares three of the largest cloud computing platforms Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) based on their data science features. Scalability, flexibility, and performance are essential factors in determining the most effective cloud platform, especially as organizations increasingly depend on data-driven decision-making. The paper examines the data science tools, machine learning services, data storage systems, integration capacities, and pricing models of each platform. AWS offers a comprehensive ecosystem with advanced services such as SageMaker and Redshift. Azure provides seamless integration with Microsoft's enterprise tools and robust AI capabilities through Azure Machine Learning. GCP, known for its open-source nature and support for TensorFlow, excels in big data analytics and model deployment. This comparative study aims to guide data scientists and businesses in selecting the most suitable platform for their data science workflows and organizational objectives by assessing each platform's capabilities, constraints, and application areas.

Keywords: AWS, Azure, GCP, Cloud Computing, Data Science, Machine Learning, Artificial Intelligence, Big Data, Cloud Platforms.

1. Introduction

The science of data has developed rapidly in recent years and has become a fundamental component of innovation across industries [1]. Predictive analytics, artificial intelligence, business intelligence, and automation have become core elements of modern operations [2]. Organizations now rely on data to make informed decisions and enhance outcomes. However, handling, processing, and analyzing large volumes of data demand substantial computational resources, scalability, and flexibility capabilities that traditional on-premise systems often cannot provide [3].

Cloud computing has emerged as the solution, offering on-demand access to infrastructure, storage, and analytical software [4]. Among cloud service providers, Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) dominate the market, each offering a diverse set of services specifically tailored for data science and machine learning applications [5].

AWS, the first major cloud provider, offers a mature and comprehensive ecosystem for data scientists [6]. It enables end-to-end workflows from data collection to model deployment through services like Amazon SageMaker, Redshift, and EMR [7]. Its scalability and extensive integrations make it suitable for both startups and large enterprises. Microsoft Azure, in contrast, excels in seamless integration with Microsoft products such as Power BI, Excel, and Dynamics 365, making it a strong option for businesses already embedded in the Microsoft ecosystem [8]. Azure Machine Learning provides automated ML, high data

pipelines, and collaborative workspaces to streamline model development and deployment [9].

Google Cloud Platform (GCP) stands out for its expertise in big data analytics, artificial intelligence, and open-source technologies [10]. Tools such as BigQuery, Vertex AI, and TensorFlow facilitate efficient data processing, model training, and scalable deployment [11]. Its emphasis on innovation, flexibility, and cost optimization appeals to organizations seeking advanced, data-driven solutions [12].

The purpose of this paper is to conduct a comparative analysis of AWS, Azure, and GCP in the context of data science evaluating their services, usability, cost-effectiveness, and performance [13]. Understanding the strengths and weaknesses of each platform will enable data scientists and organizations to make informed choices that align with their technical requirements, project goals, and strategic priorities in a rapidly evolving, data-driven world [14].

2. Literature Review

Cloud computing has revolutionized how data science projects are implemented, providing scalable infrastructure and managed services that eliminate the need for large on-premise installations. The efficiency of cloud platforms in data analytics and machine learning has been explored extensively in academic and industry studies.

In [15], the authors highlighted AWS, Azure, and GCP as the leading providers in cloud computing, noting that while their core offerings are similar, they differ significantly

in usability, scalability, and integration. These platforms enable data scientists to design complete workflows from data ingestion and preprocessing to model deployment within cost-effective and scalable environments.

AWS is renowned for its comprehensive data science capabilities. According to [16], Amazon SageMaker simplifies model creation and deployment through built-in algorithms and automated tuning. AWS Redshift has also been recognized for its efficiency in large-scale analytics. Another study [17] emphasized AWS's global accessibility and extensive third-party integrations, making it a strong choice for enterprise-level data activities. However, its complex pricing structure and steep learning curve can pose challenges for beginners.

Microsoft Azure is praised for its tight integration within the Microsoft ecosystem and user-friendly tools. As noted in [18], Azure Machine Learning supports both low-code and code-first approaches, allowing collaboration between technical and non-technical users. Its compatibility with Power BI and Excel facilitates data visualization and reporting. Similarly, [19] observed Azure's strength in hybrid cloud solutions, which enable seamless integration between on-premise and cloud environments. Despite these advantages, Azure's limited open-source integrations and relatively higher pricing for advanced workloads remain notable drawbacks.

Google Cloud Platform (GCP) is recognized for its innovation in big data and AI. In [20], the authors discussed BigQuery's serverless and high-speed design, which significantly reduces query execution time. As [21] described, Vertex AI integrates model training, tuning, and deployment into a unified environment for simplicity and scalability. Moreover, GCP's strong open-source support particularly for TensorFlow has made it popular among research and development teams. Nonetheless, [22] pointed out that GCP lags behind AWS and Azure in enterprise adoption due to its smaller market share and fewer enterprise-level integrations.

In summary, existing literature concludes that all three platforms AWS, Azure, and GCP offer powerful and reliable environments for data science, each excelling in distinct areas. AWS leads in scalability and service diversity, Azure in enterprise integration, and GCP in AI innovation and big data analytics. Ultimately, the optimal choice depends on organizational objectives, budget, and technical expertise.

3. Methodology

This study employs a qualitative comparative research approach to analyze Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) in the context of data science. The method focuses on reviewing secondary sources of academic literature, industry reports, and official documentation for each platform [23]. The research framework considers parameters relevant to data science applications, including data storage and

management, machine learning, scalability, integration, pricing, and usability [24].

The literature review was conducted using academic databases such as IEEE Xplore, SpringerLink, and Google Scholar to identify research studies on cloud computing in data science [25]. In addition, industry whitepapers, technical blogs, and official service documentation for AWS, Azure, and GCP were examined to ensure the inclusion of the most current and reliable information [26]. The collected data were synthesized into thematic categories to identify strengths and weaknesses across the three platforms [27].

To maintain consistency and comparability, each cloud provider was evaluated using identical criteria [28]. The analysis of data storage and management emphasized reliability, scalability, and performance examining tools such as Amazon S3, Azure Blob Storage, and Google Cloud Storage [29]. Machine learning and AI capabilities were assessed based on features of Amazon SageMaker, Azure Machine Learning, and Vertex AI [30]. Integration and compatibility were evaluated by examining how easily each platform connects with external tools, programming languages, and frameworks [31]. Scalability and flexibility were analyzed in terms of elasticity, multi-region availability, and resource allocation capabilities [32]. Finally, cost efficiency for common data science workloads was compared through pricing models and official cost estimators [33].

A qualitative comparison table was developed to summarize the strengths and weaknesses of each platform within these categories [34]. This analytical framework ensured a balanced and structured comparison, emphasizing practical implementation and usability rather than purely technical specifications [35]. Overall, the methodology aims to provide a reasoned and evidence-based comparison of AWS, Azure, and GCP for data science applications [36]. By combining theoretical insights from academic literature with real-world perspectives from documentation and industry sources, the study ensures that its findings are both conceptually grounded and practically applicable for organizations and data scientists selecting a cloud platform [37].

4. Results

The findings of this comparative analysis reveal that AWS, Azure, and GCP each offer strong but distinct advantages for data science applications.

Amazon Web Services (AWS) emerged as the most mature and feature-rich platform, offering a vast range of services and extensive global availability zones. Core tools such as Amazon S3 for storage, SageMaker for machine learning, and Redshift for data warehousing provide end-to-end flexibility for building data science workflows. Users can design custom pipelines for data preprocessing, model training, and deployment. However, AWS's complex pricing model and steep learning curve pose challenges, particularly for small teams and beginners.

Microsoft Azure demonstrated strengths in usability, enterprise integration, and hybrid cloud capabilities. Its seamless compatibility with the Microsoft ecosystem including Power BI, Excel, and Office 365 makes it especially appealing to organizations already using Microsoft tools. Azure Machine Learning supports both beginners and experienced users through automated and collaborative features. The platform's security, governance, and compliance offerings are also notable, particularly for industries requiring strict data protection. However, Azure's higher pricing for advanced analytics and limited support for open-source frameworks restricts its appeal for developers prioritizing flexibility and experimentation.

Google Cloud Platform (GCP) stood out for its innovation in artificial intelligence and big data analytics. Tools such as BigQuery, Vertex AI, and TensorFlow offer high-performance, scalable solutions for managing large datasets and implementing AI models. GCP's cost-per-second pricing and per-use discounts make it more transparent and cost-effective for certain workloads. Its strong support for open-source frameworks makes it ideal for research-driven and data-intensive projects. Nonetheless, GCP's smaller enterprise user base and fewer global data centers compared to AWS and Azure have limited its enterprise adoption.

In summary, the results suggest that AWS excels in scalability and service diversity, Azure in enterprise integration and user accessibility, and GCP in AI innovation and big data capabilities. The optimal choice of platform ultimately depends on an organization's technical expertise, budget, and project goals, as no single platform outperforms the others across all parameters.



Fig 1: Best Platform for Data Science

5. Discussion

The comparative analysis reveals that AWS, Azure, and GCP each offer comprehensive and capable environments for data science, yet their suitability largely depends on user needs and context. Amazon Web Services (AWS) stands out as the largest and most mature platform, featuring the broadest portfolio of tools and a globally distributed infrastructure. Its adaptability allows users to develop complex, large-scale data science pipelines that integrate

seamlessly with a variety of services. However, AWS's interface and setup can be challenging for newcomers, as it demands advanced technical expertise or dedicated cloud engineering teams.

Microsoft Azure demonstrates strength in its enterprise-oriented design and user-friendly interface. Its integration with the Microsoft ecosystem provides a familiar environment for organizations already using tools such as Power BI, Excel, or Windows-based applications. Azure's strong governance, compliance, and hybrid capabilities make it especially suitable for sectors like healthcare, finance, and government, where data security and regulation are critical. Nonetheless, its limited open-source compatibility can restrict flexibility for teams that depend on diverse programming frameworks.

Google Cloud Platform (GCP), although a later entrant into the enterprise cloud market, has carved a distinct niche in big data analytics and artificial intelligence. With robust compatibility with TensorFlow, BigQuery, and Vertex AI, GCP is an excellent choice for machine learning and large-scale analytics. Its transparent pricing and commitment to open-source development make it appealing to startups, researchers, and academic users. However, GCP's smaller market share and fewer enterprise-level integrations limit its adoption compared to AWS and Azure.

Overall, AWS offers unmatched scalability and service diversity, Azure excels in enterprise integration and accessibility, while GCP leads in AI and big data innovation. Therefore, the optimal choice of platform depends on data size, technical expertise, and long-term business objectives. The study concludes that selecting a cloud platform for data science is a contextual decision rather than a universal one.

6. Conclusion

This study conducted a comparative evaluation of Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) for data science applications. The findings indicate that AWS provides the most mature and extensive ecosystem, with tools such as SageMaker, Redshift, and S3 supporting large-scale projects, though its complexity poses a challenge for beginners. Microsoft Azure excels in seamless integration with Microsoft products and offers a secure, user-friendly environment ideal for enterprises with strong governance and compliance needs. Google Cloud Platform (GCP) distinguishes itself with powerful AI, machine learning, and big data services like BigQuery and Vertex AI, offering innovation and cost efficiency.

Each platform exhibits unique strengths, and no single solution fits all organizations. The most suitable platform depends on factors such as budget, technical skill set, data volume, and project goals. Ultimately, aligning cloud platform selection with organizational strategy and data science objectives is essential for maximizing performance, scalability, and long-term value.

References

1. Svitla Systems. (n.d.). *Cloud platforms comparison: AWS vs Azure vs Google Cloud*. Retrieved from <https://svitla.com/blog/cloud-platforms-comparison>
2. PlainEnglish. (n.d.). AWS SageMaker vs Google Vertex AI vs Azure ML. Retrieved from <https://aws.plainenglish.io/aws-sagemaker-vs-google-vertex-ai-vs-azure-ml-cloud-ml-platform-reality-918c76059904>
3. Tharwani, J., & Purkayastha, A. A. (2024). Cost-performance evaluation of general compute instances: AWS, Azure, GCP, and OCI. arXiv. <https://arxiv.org/abs/2412.03037>
4. Medium. (n.d.). Exploring AI Platforms: A comparative analysis of Azure, Google, and AWS. Retrieved from <https://ip-specialist.medium.com/exploring-ai-platforms-a-comparative-analysis-of-azure-google-and-aws-d6c032ad970b>
5. TechTarget. (n.d.). Compare Google Vertex AI vs Amazon SageMaker vs Azure ML. Retrieved from <https://www.techtarget.com/searchenterpriseai/tip/Compare-Google-Vertex-AI-vs-Amazon-SageMaker-vs-Azure-ML>
6. Amazon Web Services. (n.d.). Amazon SageMaker unified studio connecting with EMR and Redshift. Retrieved from <https://docs.aws.amazon.com/sagemaker/latest/dg/studio-notebooks-emr-cluster.html>
7. Google Cloud. (n.d.). Vertex AI: Training and deploying models. Retrieved from <https://cloud.google.com/vertex-ai/docs/start/training-methods>
8. Microsoft Learn. (n.d.). Google Cloud to Azure services comparison. Retrieved from <https://learn.microsoft.com/en-us/azure/architecture/gcp-professional/services>
9. AWS Documentation. (n.d.). Machine learning on AWS: SageMaker overview. Retrieved from <https://aws.amazon.com/sagemaker/>
10. Google Cloud. (n.d.). BigQuery overview. Retrieved from <https://cloud.google.com/bigquery>
11. Microsoft Learn. (n.d.). Azure Machine Learning overview. Retrieved from <https://learn.microsoft.com/en-us/azure/machine-learning/overview>
12. TechTarget. (n.d.). Cloud cost and performance comparison among AWS, Azure, and GCP. Retrieved from <https://www.techtarget.com/searchcloudcomputing/feature/Cloud-cost-comparison-AWS-vs-Azure-vs-Google-Cloud>
13. Svitla Systems. (n.d.). Data science services in major cloud platforms. Retrieved from <https://svitla.com/blog/cloud-platforms-comparison>
14. Medium. (n.d.). AI and data science in cloud computing: A review. Retrieved from <https://ip-specialist.medium.com>
15. Fujiyanti, V., Suranegara, G., & Ichsan, I. (2024). Comparative Analysis of Server-Based and Serverless Service Performance on Google Cloud Platform (GCP) (Case Study: Machine Learning Model Deployment). *Journal of Information Systems and Informatics*, 6(2), 1172-1194. <https://doi.org/10.51519/journalisi.v6i2.773>
16. Amazon Web Services. (2020). Amazon SageMaker Automatic Model Tuning: Scalable Gradient-Free Optimization. arXiv. <https://arxiv.org/abs/2012.08489>
17. Tharwani, J., & Purkayastha, A. A. (2024). Cost-Performance Evaluation of General Compute Instances: AWS, Azure, GCP, and OCI. arXiv. <https://arxiv.org/abs/2412.03037>
18. Microsoft. (n.d.). Low-code Machine Learning with Azure. Microsoft Learn. Retrieved from <https://learn.microsoft.com/en-us/shows/azure-videos/low-code-machine-learning-with-azure>
19. Microsoft. (n.d.). Build an Architecture with Real-Time Machine Learning Inference and Low-Code Web Application UI on Azure. Microsoft Learn. Retrieved from <https://learn.microsoft.com/en-us/azure/architecture/example-scenario/ai/deploy-real-time-machine-learning-model-application-ui>
20. Google Cloud. (n.d.). BigQuery Overview. Retrieved from <https://cloud.google.com/bigquery/docs/introduction>
21. Google Cloud. (n.d.). Vertex AI: Unifying Machine Learning Development. Retrieved from <https://cloud.google.com/vertex-ai/docs/start/introduction-unified-platform>
22. Pamisetty, A. (2022). A Comparative Study of AWS, Azure, and GCP for Scalable Big Data Solutions in Wholesale Product Distribution. *International Journal of Scientific Research and Modern Technology*, 1(12), 71-88. <https://doi.org/10.38124/ijrsmt.v1i12.466>
23. 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. *Information Systems*, 47, 98-115.
24. Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011). Cloud computing the business perspective. *Decision Support Systems*, 51(1), 176-189.
25. 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.
26. Amazon Web Services. (2024). AWS Cloud Overview: Official Documentation. Retrieved from <https://aws.amazon.com/documentation/>
27. Microsoft Azure. (2024). Azure Architecture and Data Science Services Overview. Retrieved from <https://learn.microsoft.com/azure/>
28. Google Cloud Platform. (2024). Google Cloud Documentation: Data and AI Services. Retrieved from <https://cloud.google.com/docs>
29. Patel, P., Ranabahu, A. H., & Sheth, A. P. (2009). Service level agreement in cloud computing. *Cloud Workshops on Cloud Computing*, IEEE.
30. Zeng, W., Zhao, Y., Ou, K., & Song, W. (2009). Research on cloud storage architecture and key technologies. *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, 1044-1048.

31. Li, M., Smirni, E., & Riska, A. (2019). Understanding cloud performance interference using workload decomposition. *IEEE Transactions on Cloud Computing*, 7(4), 1130–1143.
32. Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., & Zaharia, M. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50–58.
33. Li, A., Yang, X., Kandula, S., & Zhang, M. (2010). CloudCmp: Comparing public cloud providers. *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, 1–14.
34. Ghorbani, A., & Khatibi, V. (2021). Comparative analysis of AWS, Azure, and GCP cloud computing platforms. *International Journal of Computer Applications*, 183(15), 10–18.
35. Sultan, N. (2014). Cloud computing: Past, present, and future. *International Journal of Information Management*, 34(5), 748–758.
36. Alharthi, A., Yahya, F., Walters, R., & Wills, G. (2017). An overview of cloud services adoption challenges in higher education institutions. *Proceedings of the IEEE 5th International Conference on Cloud Computing Technology and Science*, 98–103.
37. Hossain, M. S., & Muhammad, G. (2020). Cloud-assisted industrial Internet of Things (IIoT) – enabled framework for health monitoring. *IEEE Internet of Things Journal*, 7(8), 7495–7503.
38. Prasanth Tirumalasetty, (2025). System and Method for Generating Privacy-Preserving Synthetic Health Data Using a Generative Adversarial Machine Learning Mode.