

A COMPARISON OF AWS, GOOGLE CLOUD, AND AZURE SERVICES PERCEPTIVITY FOR CLOUD DATA ENGINEERING

Abstract

The cloud data engineering has revolutionized the way organizations process, manage and analyze large datasets. This area solves business problems by applying the scale, adaptability, and processing capacity of modern platforms, led by powerful cloud-based capabilities. This study focuses on the three major platforms: Amazon Web Services (AWS), Google Cloud Platform (GCP) and Microsoft Azure. They also examine the unique offerings and features of each platform. We examine the appropriateness of these CHP models for real-life prediction tasks like the detection of financial dependencies.

Keywords: Cloud Data Engineering, AWS, Google Cloud, Azure, Server-less Computing, Data Analytics, ETL, Scalability.

Author

Ayush Pant
B.Tech student,
Department of Computer Sciences,
Arya College of Engineering and IT,
Kukas, Jaipur, Rajasthan
pantayush92@gmail.com

Article details:

Published in: IIP Series
Volume: 5 | Year: 2025
ISSN: 3070-7617 (Online)
DOI: <https://www.doi.org/10.58532/nbennurFTHPSW9N1>
Publisher: Iterative International Publishers (IIP)

Cite this:

Ayush Pant , A COMPARISON OF AWS, GOOGLE CLOUD, AND AZURE SERVICES PERCEPTIVITY FOR CLOUD DATA ENGINEERING, In: *Futuristic Trends in High Performance Computing and Software Development V5B23, IIP Series, Volume 5, September, 2025, pp. 66-74, Iterative International Publishers IIP, E-ISBN: 978-93-7020-131-6 , DOI: <https://www.doi.org/10.58532/nbennurFTHPSW9N1>*

I. INTRODUCTION

There is an urgent need for effective data management and engineering solutions due to the overwhelming growth of data creation. This has made cloud data engineering the de factor solution for most organizations seeking to derive useful insights from massive datasets. As organizations evolve their needs over time, AWS, Google Cloud Platform, and Microsoft Azure are vying to be the top player in the market by offering a vast array of tools and services.

The AWS works because it has a full spectrum of services ranging from storage, computing, analytics to machine learning. This allows enterprises to use different kinds of workloads.

What makes GCP unique is its ability to effortlessly connect within its ecosystem of tools and stand out mostly for advanced analytics, machine learning, and data processing capabilities. In contrary Azure is more focused on enterprise-grade solutions and excels in joining cloud resources with Microsoft applications and on-premise infrastructure solutions. Such a comprehensive comparative analysis will help focus mainly on the distinctive characteristics along with scalability options and real-world applications of such platforms so that decision-makers can find the best solution based on specific requirements. Through capturing each platform's pros and cons, this paper shall enlighten how organizations may obtain the utmost levels of cloud data engineering.

II. PIVOTAL FEATURES OF CLOUD DATA ENGINEERING PLATFORMS

Cloud data engineering is a full stack of tools and services designed to make processes such as data ingestion, transformation, storage, and advanced analytics easier. Below, we consider some of the core offerings of AWS, GCP, and Azure in the context of their strengths and differences.

1. Amazon Web Services (AWS)

- a) **AWS Glue :** This is an AWS service for ETL which serves to automate data integration. Supporting a wide range of data sources, AWS Glue reduces schema discovery and management time by leaps and bounds, thereby reducing time taken for complex ETL workloads.
- b) **Amazon Redshift:** It's a high-performance, horizontally scalable data warehouse solution optimized for analytical queries. Redshift natively integrates with AWS's broader ecosystem, which supports advanced business intelligence workflows.
- c) **Amazon S3 (Simple Storage Service):** S3 brings durability and great cost-efficient capability at scale on unstructured data and is most compatible with analytical and other machine learning offerings thus laying foundation to Amazon's data ecosphere.
- d) **Amazon Kinesis:** Best suited to implement near-real-time streams of data ingest processing, along with analytics across any high velocity data streams: it enforces applications designed in both IoT and the evented architecture.

2. Google Cloud Platform (GCP)

- a) **BigQuery:** Is a fully managed, serverless data warehouse that can handle large-scale real-time analytics. Its in-built support for machine learning and data visualization tools enhances its utility for predictive analytics.
- b) **Dataflow:** A tool that allows the development of scalable ETL pipelines, Dataflow enables the creation of both stream and batch processing. It has an integration with Apache Beam that gives flexibility to developers across diverse programming environments.
- c) **Cloud Storage:** This highly durable and secure object storage service supports multi-regional availability for guaranteed data availability for analytics, AI, and archival.
- d) **Cloud Pub/Sub:** This distributed messaging system is for event-driven architectures that ensure low-latency communication, thereby enabling real-time data

3. Microsoft Azure

- a) **Azure Synapse Analytics:** This service is a unification of big data processing and data warehousing capabilities in one unified platform. It supports integration with Power BI and other Azure services, thereby enabling seamless data visualization and analysis.
- b) **Azure Data Factory:** This hybrid data integration service makes it easier to create and schedule ETL workflows. Its visual design tools and extensive connectors make complex data engineering tasks much easier.
- c) **Azure Blob Storage:** A general-purpose, highly available storage for unstructured data, Blob Storage supports a range of applications from archiving to AI model training with strong security and scalability.
- d) **Azure Event Hubs:** Designed for real-time data streaming, Event Hubs enables high-throughput data ingestion, making it a critical component for telemetry and event-driven applications.

Integration of Microservices and Scalability Modern businesses require platforms that are scalable, resilient, and able to adapt to shifting workloads. AWS, GCP, and Azure use serverless computing and microservices architecture to achieve scalability, as shown below:

4. Serverless Computing AWS Lambda

Event-driven applications can scale cost-effectively because of Lambda, which dynamically allocates resources according to workload demands. This also makes it suitable for use in dynamic workflows due to its close integration with AWS services.

5. Google Cloud Functions

Cloud Functions is a lightweight, event-driven application framework that enables seamless scalability and effective execution. It integrates quite well with the GCP ecosystem to offer comprehensive solutions.

6. Azure Functions

Azure Functions support serverless computing across diverse applications, including API management and real-time data processing. Deployment is easily handled through compatibility with other Azure services.

7. Microservices and Container Orchestration AWS Elastic Kubernetes Service (EKS)

A managed Kubernetes service that lets you manage and deploy scalable containerized applications. AWS EKS integrates well with other AWS services to offer comprehensive solutions.

8. Google Kubernetes Engine (GKE)

For container workloads, automated scaling and cluster management make GKE ideal for high-performance use. Added to GCP's AI services, it brings great value.

9. Azure Kubernetes Service

AKS allows Kubernetes to be used directly within the Azure ecosystem. Hybrid support carries hybrid capabilities between cloud-based and on-premise workloads. One of the most defining traits of any cloud platform is scalability, which essentially allows businesses to change according to different workloads and provide consistent service in reaction to varying levels of demand.

The primary cloud platforms—AWS, Google Cloud, and Azure—offer rich sets of tools designed to accommodate diverse operational needs and enable seamless scaling.

One of cloud platforms' most critical features is scalability, which allows companies to adjust to changing workloads and ensure consistent service delivery to customers based on varying demand levels. The major cloud platforms-AWS, Google Cloud, and Azure-provide advanced solutions designed to meet diverse operational needs and support seamless scaling.

10. Serverless Computing

Serverless computing enables developers to create and run applications without worrying about the underlying infrastructure. These services increase the cost-effectiveness and efficiency of an event-driven architecture by allocating resources automatically in response to real-time demands.

11. AWS Lambda

A brief description of what each platform offers follows AWS Lambda allows users to run code in reaction to events, automatically scaling according to workload demands.

It supports several programming languages, including Python, Node.js, and Java. It also provides integrations with S3, DynamoDB, and API Gateway among other AWS services. Since it bills in milliseconds, customers only pay for the execution time tracked. It is ideally suited for processing data streams; managing real-time data workflows and performing backend tasks such as file transformations.

12. Google Cloud Functions

Google Cloud Functions: lightweight, event-driven execution environment for building modular applications. Key Features: It supports all the major programming languages, including Go, Python, and JavaScript. It optimizes zero-downtime scaling plus native integrations with GCP services such as Pub/Sub, Firestore, and Cloud Storage to scale real-time operations forever in operations that require real-time processing of files, lightweight APIs, and bot-driven applications.

Microservices and Container Orchestration Microsoft managed serverless offering is Azure Functions which allows quite seamless running across a wide variety of workload types. It's significantly integrated within many Microsoft services.

Providing flexible event bindings and triggers helps accommodate a huge variety of use cases. This calls for robust management tools for stateful workflows.

Use Case: Most commonly applied in business logic workflows, IoT data processing, and task automation without human intervention. Microservices architecture allows developers to focus on independent, deployable components.

Cloud providers primarily based on Kubernetes to expand deployment and management capabilities for these services through powerful container orchestration capabilities.

13. AWS Elastic Kubernetes Service (EKS)

AWS EKS makes it easier to deploy, manage, and scale Kubernetes clusters in the cloud.

Key Features: Fully managed Kubernetes service that integrates AWS tools, such as IAM and CloudWatch. Supports EKS Anywhere for hybrid and on-premises environments. High availability; automated updates and patches.

Use Case: Perfect for businesses that want to use Kubernetes within the AWS ecosystem of services.

One of the original providers of managed Kubernetes services, Google Kubernetes Engine has a deep expertise in Kubernetes.

Key Features: Automatic scaling and updating of Kubernetes clusters. Anthos for hybrid cloud and multicloud deployment integrations.

14. Google Kubernetes Engine (GKE)

AI tools are highly integrated with GCP services like BigQuery & Pub/Sub. Use case: famous for managing workloads in hybrid or multi-clouds requiring complex orchestration.. Microsoft's Azure Kubernetes Service is an easy-to-use fully managed enterprise-grade security-focused service for K8s.. Key Features:

15. Azure Kubernetes Service (AKS)

Azure Monitor for Kubernetes complements RBAC's seamless integration with Azure Active Directory to provide complete visibility over all resources in the cluster.

CI/CD is natively integrated with Azure Pipelines, thus empowering DevOps processes seamlessly across teams.. Use case:. Organizations running containerized applications and microservices on Microsoft's Ecosystem. Operations in Real-Time The Cloud Platforms have become crucial lately in addressing challenges posed by intricate real-time data spanning diverse sectors. With the capability to process tremendous volumes of data almost real-time, organizations could achieve operational levels hitherto considered impractical.

Below is an in-depth exploration of real-time applications facilitated by AWS, Google Cloud Platform (GCP), and Microsoft Azure in financial fraud detection, e-commerce recommendations, and healthcare analytics.

III KEY TOOLS AND TECHNIQUES

- 1. Google BigQuery:** A serverless, highly scalable data warehouse, BigQuery allows companies to process large datasets in near real-time. Its integration with machine learning tools allows for advanced anomaly detection algorithms to be applied to transactional data.
- 2. Google Dataflow:** A fully managed stream and batch data processing service, Dataflow handles real-time ingestion, transformation, and analysis of financial data streams. With its unified programming model, organizations can detect irregularities as they occur.

IV. WORKFLOW IN PRACTICE

- 1. Data Ingestion:** All transactional data from various payment gateways, ATM channels, and banking systems is ingested directly into Dataflow in real-time through Pub/Sub and is subsequently analyzed.
- 2. Anomaly Detection:** Using statistical and machine learning models within BigQuery, patterns in fraudulent behavior are detected for anomalies, such as irregular locations of transactions, sums, or frequencies.

- 3. Alerts and Mitigation:** Real-time alerts are sent out from anomaly detection to the risk management team in real-time, allowing transaction blocking or further verification.

Microsoft Azure is a strong integration solution for enterprise ecosystems, especially for organizations already invested in Microsoft software solutions. Azure Synapse Analytics, Data Factory, and Blob Storage offer strong tools for handling structured and unstructured data.

- 4. Unified Data Solutions:** Azure Synapse Analytics brings together big data and data warehousing, making analytics workflows more streamlined.
- 5. ETL Automation:** Azure Data Factory makes it easy to create and manage ETL pipelines, thereby increasing productivity.
- 6. Secure Storage:** Azure Blob Storage guarantees scalable and cost-effective storage for unstructured data.
- 7. Scalability and Microservices Integration :** All three platforms offer advanced scalability solutions and support microservices architectures through managed Kubernetes services.

V CONCLUSION

In summary, the comparison among AWS, Google Cloud, and Azure in cloud data engineering illustrates that each has its own distinct set of strengths, innovations, and value propositions. The choice of a cloud provider thus becomes a very important decision that can result in a major factor in the success of data engineering endeavors with more and more organizations depending on cloud technology to process, manage, and analyze big data. AWS, Google Cloud, and Azure are the three leading players in this space. They each serve different needs and applications and continue to develop to keep up with the growing needs of contemporary data engineering.

As the originator of cloud computing, AWS has a mature and wide range of services. It is a flexible solution for companies of all sizes because of its large global infrastructure and many tools for data storage, processing, and analysis.

Its advantages are scalability, reliability, and the depth of AWS's set of services for data warehousing (Redshift), storage (Amazon S3), ETL (AWS Glue), and storage (Amazon S3). AWS provides the flexibility and solidity needed to design and operate advanced data pipelines for companies with complex data engineering workflows or extensive customizations.

Nonetheless, it might be challenging for smaller teams or those who have little cloud experience due to its sophistication and pricing scheme. On the other hand, Google Cloud excels because of its innovative capabilities in real-time data processing, machine learning, and data analytics.

BigQuery, a serverless and massively scalable data warehouse, and Dataflow, a fully managed stream and batch processing service, are two of Google Cloud's offerings that are

best suited for companies that highly regard speed, imagination, and data-driven decision making.

It is an excellent option for companies that desire to leverage sophisticated analytics and predictive modeling due to its extensive integration with AI and machine learning services such as TensorFlow and Vertex AI. Moreover, Google Cloud's global network infrastructure and sustainability focus contribute to its popularity.

But it might have fewer third-party integrations and a less extensive partner ecosystem as a result of its lower market share compared to AWS and Azure. Azure, Microsoft's cloud platform, has an edge when it comes to integrating with other parts of the Microsoft ecosystem, including Office 365, Power BI, and SQL Server.

Consequently, it is a wonderful choice for companies that are already heavily invested in Microsoft solutions. Organizations are able to effectively bridge on-premises and cloud environments courtesy of Azure hybrid cloud capabilities provided by services such as Azure Arc and Azure Synapse Analytics.

Its reliability as a platform for highly regulated industries is further enhanced by its robust emphasis on enterprise-level security, compliance, and governance. While Azure also offers a large array of data engineering tools, though, some users find that its documentation and interface are not as user-friendly as those of its competitors. It is important to consider costs, usability, performance, and the specific needs of the data engineering tasks in question when comparing these platforms.

For example, companies that need real-time analytics and machine learning will likely favor Google Cloud, while companies that need hybrid cloud capabilities and deep integration with Microsoft offerings will favor Azure. AWS is nevertheless a favorite among companies in search of one-stop-shop coverage for an assortment of data engineering needs due to its unmatched extent of services. Ultimately, determining a cloud hosting provider is never a matter of one size that fits all. It requires scrupulous examination of the firm's objectives, the existing infrastructure, the talents of the crew, and strategic objectives in the long term.

AWS, Google Cloud, and Azure are likely to introduce new tools and services that push the boundaries of what is possible in cloud data engineering as the cloud computing landscape continues to change.

However, a company's ability to effectively utilize these tools, cultivate a culture of data-driven decision-making, and continuously adapt to the ever-changing technological landscape will all play a role in the success of any data engineering initiative. Attaching technology to business strategy and equipping teams with the ability to leverage the strength of data in driving innovation and growth is what unlocks the entire potential of cloud data engineering within this dynamic environment.

REFERENCES

- [1]. Abuqabita., Al-Omoush, R. and Alwidian, J. (2019). A Comparative Study on Big Data Analytics Frameworks, Data Resources and Challenges. *Mod. Appl. Sci.*, 13(7), 1–14.
- [2]. Acharjya, D.P. (2016). A Survey on Big Data Analytics : Challenges, Open Research Issues and Tools. *Int. J. Adv. Comput. Sci. Appl.*, 7(2), 511–518.
- [3]. Ahmed, N., Andre L.C. Barczak, Teo Susnjak. and Mohammed A. Rashid. (2020). A Comprehensive Performance Analysis Of Apache Hadoop And Apache Spark For Large Scale Data Sets Using HiBench. *Journal of Big Data*7(110), 1–18, Springer.
- [4]. Al-Barznji, K. and Atanassov, A. (2016). A Survey of Big Data Mining: Challenges and Techniques. in *Proceedings of 24th International Symposium "Control of Energy, Industrial and Ecological Systems"*, 113–117, Bankia, Bulgaria. Al-Barznji, K. and Atanassov, A. (2016). A MapReduce Solution for Handling Large Data
- [5]. Efficiently. *Int. Sci. J. "MACHINES. Technol. Mater.*, 23 (12), 20–23.
- [6]. Al-Barznji, K. and Atanassov, A. (2018). Big Data Sentiment Analysis Using Machine Learning Algorithms in *Proceedings of 26th International Symposium "Control of Energy, Industrial and Ecological Systems"*, Bankia, Bulgaria, 53–58.
- [7]. Amazon Kinesis (Received on April 2022). <https://aws.amazon.com/kinesis/>
- [8]. Apache Drill (Received on April 2022). <https://drill.apache.org/architecture/>
- [9]. Apache Flink (Received on April 2022). <https://flink.apache.org/>
- [10]. Apache Spark (Received on April 2022). <https://spark.apache.org/docs/latest/cluster-overview.html>
- [11]. Apache Samza (Received on April 2022). <https://engineering.linkedin.com/samza/apache-samza-linkedins-stream-processing-engine/> Ellingwood J. (2016), Hadoop, Storm, Samza, Spark, and Flink: Big Data Frameworks Compared. Retrieved 25 Feb 2019, from <https://www.digitalocean.com/community/tutorials/hadoop-storm-samza-spark-and-flink-big-data-frameworks-compared>
- [12]. flink-big-data-frameworks-compared
- [13]. Zaharia, M.; Das, T.; Li, H.; Hunter, T.; Shenker, S.; Stoica, I. Discretized streams: Fault-tolerant streaming computation at scale. In *Proceedings of the 24th ACM Symposium on Operating Systems Principles (SOSP '13)*, Farmington, PA, USA, 3–6
- [14]. November 2013; pp. 423–438.