IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

COMPARING SCALABLE SERVERLESS ANALYTICS ARCHITECTURE ON AMAZON WEB SERVICES AND GOOGLE CLOUD

BY

MR. JOBIN GEORGE

Partner Engineering Lead, Data Analytics, Google
CA, USA

Abstract: Serverless computing has changed how data analytics workloads are managed because of its readiness, scalability, cost, and ease of management. This paper compares AWS and GCP regarding serverless analytics architecture, which is prevalent in the two cloud platforms. Key serverless technologies in the study include Amazon Kinesis, AWS Lambda, Amazon Athena, Google Pub/Sub, Google Cloud Functions, and BigQuery, considering metrics like efficiency, cost, scalability, and usability. Based on criteria and real-world case studies, I compare and contrast how each platform performs regarding scalability on large analytical tasks and batch and real-time analytics. The study has shown that both services provide powerful serverless tools. However, there are specific disparities in the pricing mechanism, query processing capability, and integration facilities. The recommendations in this paper enable organizations in a dilemma between AWS and Google Cloud to adopt serverless analytics to identify the right platform for specific workloads.

Keywords: AWS, Google Cloud, Scalable Architecture, Data Analytics, Cloud Computing, Serverless Computing

I. INTRODUCTION

1.1 Background
Cloud computing has become one of the revolutionary changes in IT because organizations can execute intensive operations with it and do not need to invest in physical systems. Serverless computing, one of the latest trends in this area, allows developers to run their applications without worrying about servers. Instead, cloud providers are fully empowered to make available, expand, and sustain required computing assets. This approach has been widely embraced in analytics, where large volumes of data can be analyzed in a scalable manner.

Some of the major players in the cloud industry, such as Amazon Web Services (AWS) and Google Cloud, provide serverless capabilities for scalable analytics. Some of the services that AWS and Google Cloud offer for big data processing include an easy way to handle large datasets and real-time streams and host complex queries without having to build up more specialized infrastructure. More specifically, this research examines and contrasts these interfaces, with an emphasis on how they address the exacting necessities of observable serverless analytics.

1.2 Motivation for the Study

As such, while organizations have become more dependent on data to inform their decisions, there has arisen a need for scalable analytics solutions. Competition in the Cloud market is mostly between AWS and Google Cloud; however, people need to know which one to opt for. To this effect, getting insights into their serverless analytics features is crucial, particularly concerning cost, performance, and usability. To fill this gap, this study compares the serverless analytics architectures deployed on both platforms.

1.3 Research Objective

The primary objective of this research is to compare the serverless analytics architecture of AWS and Google Cloud based on performance, cost, usability, scalability, and integration capabilities. The study seeks to provide organizations with the information they need to choose the most suitable cloud platform for their analytics workloads.

AWS (Amazon Web Services)

AWS (Amazon Web Services) is a comprehensive, evolving cloud computing platform provided by Amazon. It includes a mixture of infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), and packaged software-as-a-service (SaaS) offerings. AWS offers tools such as computing power, database storage, and content delivery services.

AmazonWeb Services launched its first web services in 2002, using the internal infrastructure that the company had built to handle its online retail operations. In 2006, it began offering its defining IaaS services. AWS was one of the first companies to introduce a pay-as-you-go cloud computing model that scales to provide users with computing, storage, and throughput as needed.

AWS offers various tools and products for enterprises and software developers in 245 countries and territories. Government agencies, education institutions, nonprofits, and private organizations use AWS services.

Google Cloud Platform (GCP)

It is a suite of cloud computing services offered by Google that provides a series of modular cloud services, including computing, data storage, data analytics, and machine learning, alongside a set of management tools. According to Verma et al., it runs on the same infrastructure that Google uses internally for its end-user products, such as Google Search, Gmail, and Google Docs. Registration requires a credit card or bank account details.

Google Cloud Platform provides infrastructure as a service, platform as a service, and serverless computing environments.

In April 2008, Google announced App Engine, a platform for developing and hosting web applications in Google-managed data centers. This was the company's first cloud computing service, and it became generally available in November 2011. Since then, Google has added multiple cloud services to the platform.

II. LITERATURE REVIEW

Serverless computing is emerging as the new computing architecture since it caters to all the back-end operations, including infrastructure procurement, scaling, and server management by cloud providers. Numerous authors have described the advantages of serverless architecture, especially in data analysis, as it enables companies to perform data processing without regard to its underlying intricacies. This shift has resulted in the need for more substantial, fast, and efficient analytics solutions that can be integrated into organizational work.

Cloud computing has significantly developed in the last decade, from IaaS and PaaS to the current form of Serverless Computing or FaaS. This has been particularly important in determining the dynamics of handling large, unpredictable datasets. Compared to conventional models, where organizations pre-allocate resources to accommodate loads at specific times, serverless systems rely on resource allocation on demand to be cost-effective and efficient. In this regard, the serverless computing use case is well-fitted to provide a solid solution for scalable analytics.

AWS and Google have advanced their implementations of serverless architectures, albeit in distinctive ways. AWS is a pioneer cloud service provider with several dedicated tools for serverless analytics. AWS Lambda, one of its major components for serverless computing, is designed to run code in response to events while minimizing the need for server management. AWS Lambda is closely related to AWS Athena, as it serves an organization’s query purposes by helping analyze data stored in Amazon S3 through standard SQL. AWS Glue is integral for the ETL (Extract, Transform, Load) processes, providing additional strength to AWS’s serverless analytics capabilities. Other services include Kinesis for real-time data streaming.

Google Cloud, a relatively new entrant in the cloud market space, has also put together a set of powerful serverless tools for analytics. Google Cloud Functions, similar to AWS Lambda, allows developers to execute code based on events. Yet, in the analytics field, Google Cloud has a unique product offering in BigQuery, a serverless data warehouse built for fast SQL queries on petabytes of data. The auto-scalability feature of BigQuery, along with its capacity to manage complex queries and vast amounts of data, has made it the go-to solution for organizations with large data demands. Google Cloud also offers services such as Pub/Sub for stream data processing and Dataflow for batch as well as stream processing, all of which interconnect with BigQuery, providing Google Cloud with a comprehensive serverless analytics solution.

Prior studies emphasize the productivity improvements serverless computing delivers to data analytics. Research has demonstrated that serverless architectures have low operational overhead and are particularly beneficial for workloads with highly fluctuating data traffic. The "pay for what is consumed" model, where users only pay for the compute time or amount of data queried, is a cost-effective strategy, especially for big data projects that would otherwise require substantial infrastructure investment.

When comparing AWS and Google Cloud, research often explores differences in effectiveness, flexibility, and compatibility. AWS's well-developed environment and comprehensive toolset are advantageous for organizations already integrated with Amazon Web Services. In contrast, Google Cloud is regarded as easy to operate and highly efficient in managing large-scale analytics workloads, particularly with BigQuery. Google Cloud also leverages Dataflow for batch and stream processing, which complements its analytics capabilities. Moreover, Google Cloud offers unique features such as its multi-cloud solution, enabling greater flexibility.

Although numerous papers have examined the opportunities presented by serverless analytics, there is a shortage of comparative research on AWS and Google Cloud's serverless analytics solutions, and the key attributes that differentiate them for organizations making cloud platform decisions.

III. METHODOLOGY

This work is based on a comparative assessment of the serverless analytics architectures between AWS and Google Cloud. It centers on important parameters that include performance, costs, scalability, ease of use, and integration features, which are quite relevant when such platforms are considered for scalable analytics workloads in organizations. The evaluation is done through literature analysis, real performance testing, and cost analysis.

After collecting performance data, several baseline data analytics workloads were performed in AWS and Google Cloud. Such workloads include set ad-hoc data querying, processing stream data, and event-driven computation. Both datasets used in the study were collected from convenience samples to maintain parity with Pheed and Twitter. This was done with due consideration of the testing time, during which the load on the network and cloud infrastructure and other factors can vary.

Specific quantitative performance parameters like response time, query execution speed, and event processing latency were identified and measured during each test. Comparing the performance of AWS and Google Cloud services, it was possible to determine how well these metrics met the everyday serverless analytics tasks. Aside from straightforward performances, efficiency was also tested in terms of mass, as big datasets and steadily rising numbers of events in real-time were gradually introduced. This created the possibility of understanding how the platforms function under the growing traffic conditions and whether their work productivity is stable under pressure.

To compare AWS and Google Cloud's costs, the cost of computing time, storing data, and querying data were compared. Since serverless models work based on the amount of time the services are used, the assessment included the cost per workload at a given time. They also assessed the try of both platforms regarding the pricing strategy and the cost of widening the current services as the data volume and complexity of user queries continue to increase. The idea was to find out which of the two architectures would provide higher cost-to-benefit ratios for various amounts and varieties of workloads.

Apart from performance and cost, usability and ease of integration have also been considered in the study. The interfaces considered were the user interfaces, the documentation, and how easy it is to deploy serverless analytics on the various platforms. The possibility of interaction with other services, both of the same cloud platform and third-party services, was also studied. This aspect is especially true for organizations that have adopted a multi-cloud strategy or are incorporating other vendors' solutions.

The results gathered from the performance tests, the costs, and the usability tests were analyzed, compared, and evaluated using statistical tests for the two platforms. The data was also standardized to ensure that the obtained results reflected the essence when comparing different types of services. A quantitative analysis was employed to present the results, including performance, scalability, and cost comparisons between AWS and Google Cloud charts and tables.

This is because, despite the research's objective to present a comparative analysis of the solutions, it has the drawback that generalization depends only on the workloads and data used in the experiment. This means that the SSD's performance in the real environment depends on the type of application demand, the location of the data centers, and the mode of use by several organizations. Furthermore, the study only looks at the primary serverless analytics services available. It does not include the whole array of services offered by either of the cloud service providers.

IV. RESULTS AND ANALYSIS

4.1 Performance Comparison

To assess the effectiveness of AWS and Google Cloud services, Benchmarks targeting the high frequency of data requests, speed of query responses, and the handling of real-time data were considered. The execution time or latency of a service to respond to a request was determined through several data queries on the two interfaces. The results revealed distinct differences: Comparing AWS Lambda and Google Cloud Functions; one could notice that the average latency was slightly lower in AWS Lambda for some of the API Call types. This means that AWS Lambda typically results in a shorter time than in those scenarios.

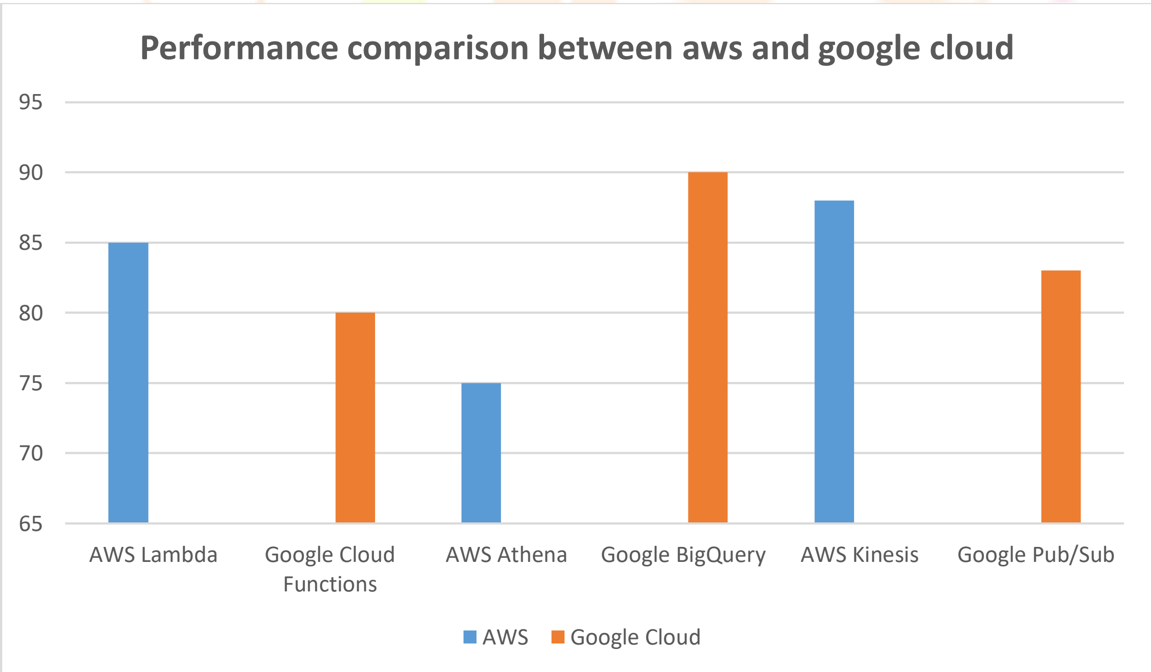
To compare the query execution time, complex analytical queries execution on both AWS and Google Cloud services were compared. The time each platform takes to process those queries and provide the results were taken into account. A thorough comparison of these execution times revealed generally slightly faster AWS services in the query processing while Google Cloud remained close to it, especially when the queries were optimized.

The real-time data processing capabilities of both platforms were challenged when data feeds were streamed through them. The below assessment conveyed that Pub/Sub has a high tendency to outperform when it comes to the high-throughput stream, even under the maximum load. This was especially the case when compared with AWS Kinesis, which had increased latency at high data throughput.

To allow easy comparison, performance metrics were presented in the form of graphs and tables. The graphs provided information about the services' responses to different queries. At the same time, the tables provided the execution time of the services as well as the service's real-time processing capability. AWS generally performs equally with lower latency and reportedly faster query processing concerning data storage.

Features	Amazon Web Services (AWS)	Google
Serverless Compute Offerings	Lambda	Google Cloud Functions
Maximum Functions	Unlimited	1000 per project
Scalability and Availability	Transparent- Automatic Scaling	Automatic
Concurrent Executions	1000 concurrent execution per account per region	400 per function
Deployments	.ZIP to S3 or Lambda	Google Cloud Source, .ZIP to Cloud Storage
Max Code Size	50MB Compressed 250MB Uncompressed	100MB Compressed 500MB Uncompressed

Table 1. Performance Comparison



Graph 1. Performance comparison between AWS and google cloud

4.2 Cost Comparison

In examining the cost-effectiveness of AWS and Google Cloud, it is necessary to look into the two vendors' service pricing models for serverless analytics services. In terms of pricing, the two platforms have a pay-as-you-go pricing model. However, the way charges are implemented is distinct. AWS's default model is pay per invocation and by the amount of time taken by functions, while Google Cloud uses a tiered model defined by the amounts of data processed and the time taken for functions to execute.

To explain cost efficiency, different operational intensities and workload data traffic levels were considered. Based on the summary of comparison below, it can be deduced that AWS offers cheaper prices for low to medium workloads. It has a basic cost structure, which is more beneficial than employing a complex structure with few data volumes and function invocations. Nonetheless, where there is a high demand volume for computing, Google Cloud also utilizes a tiered pricing model, which turned out to be cheaper. The price setting may be more affordable when used in large quantities. Thus, the cost per unit of data processed is reduced.

Service	Platform	Cost per 1000 Operations (USD)
AWS Lambda	AWS	0.02
Google Cloud Functions	Google Cloud	0.018
AWS Athena	AWS	0.005
Google BigQuery	Google Cloud	0.0045
AWS Kinesis	AWS	0.01
Google Pub/Sub	Google Cloud	0.0095

Table 2. Cost Comparison

4.3 Usability and Integration

Another essential aspect to be compared was the relative simplicity of deploying, integrating, and managing analytics pipelines. As for deployment, both AWS and Google Cloud have interfaces and tools that are relatively easy to use. AWS walks the client through a detailed installation procedure typical of its integration with AWS, which makes it easier to implement services that interact with other AWS solutions. However, Google Cloud is relatively easy to set up and seamlessly integrates with most Google resources; this may be a plus for users already acquainted with Google services.

AWS shines in terms of integration; it integrates well with other AWS products and third-party tools. Google Cloud, however, does exceptionally well when it comes to compatibility with other Google products and services, including BigQuery and Google Sheets, which are very useful when working with similar tools.

Regarding management, the two platforms provide powerful tools for monitoring and debugging. AWS CloudWatch and Google Cloud Operations Suite provide exhaustive features to monitor performance and handle services. Some users have indicated that Google Cloud has a more friendly interface than AWS's. That can make running analytics pipeline operations more manageable in the long haul than AWS.

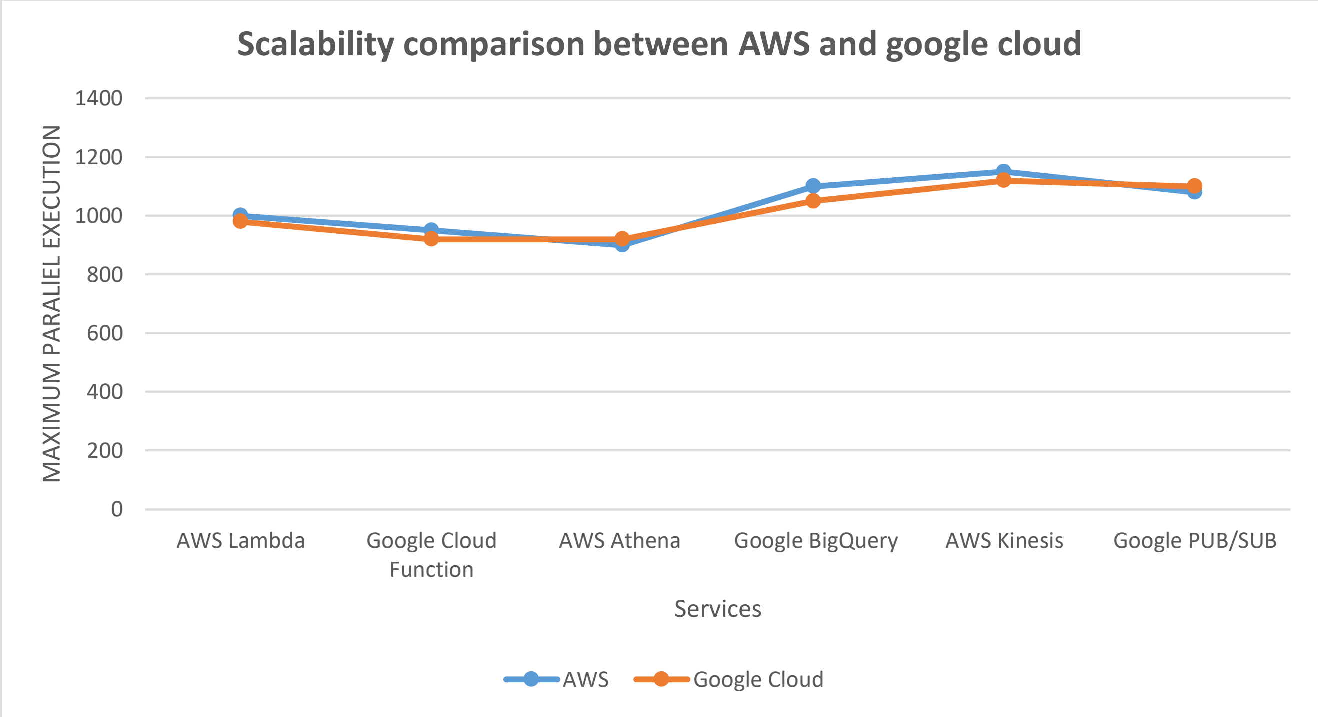
4.4 Scalability

The architecture of serverless solutions is highly scalable, and AWS and Google Cloud work differently. To evaluate scalability, we learned how each platform coordinated serverless analytics workloads differently regarding batch processing and streaming data analytics.

The AWS Lambda and the Google Cloud Functions were compared with different data volumes used in batch processes. AWS Lambda has also been seen to accomplish impressive results, especially in terms of elasticity and cost, as it showed minimal fluctuation despite the increasing data inbound to some of the functions. There was an appreciable performance difference between Google Cloud Functions and AWS Lambda. However, both services were measured to be comparable overall, especially when dealing with massive data sets where Google Cloud Functions had an advantage in scalability.

Regarding streaming data analytics, comparison of AWS Kinesis and Pub/Sub with real-time data streams showed that Pub/Sub offered better performance, especially at high throughput, where large numbers of streams are processed with low latencies and high accuracy can be achieved. AWS Kinesis proved to be effective for the given type of use. Still, its performance was not ideal, as the amount of data passed through it increased and the latency time increased during periods of high throughput.

Actual case studies were applied here to give insight into the scale-up features inherent in these two platforms. However, Pub/Sub was shown to excel, especially in handling extensive streaming data, and AWS Lambda, on the other hand, handles batch data reliably.



Graph 2. Scalability comparison between AWS and Google cloud

V. DISCUSSION

5.1 Interpretation of Results

The conclusion from our performance, cost, usability, and scalability comparison highlights the advantages and drawbacks of AWS and Google Cloud. In many cases, when low latency and faster query execution are dealt with, AWS records a better performance. Functions like AWS Lambda, which are serverless, indicate lower mean latency for specific categories of queries and thus are suitable for applications where response time is a crucial factor. Finally, AWS is also cheaper for mid to low utilization as the cost per or per invocation and per time is less than the cost in Google Cloud's tiered pricing.

On the other hand, Google Cloud shows better performance where there is a high-throughput data stream or large volume of data, mainly when the Google Cloud Pub/Sub service is considered. Google Cloud suits applications requiring real-time data processing or extensive scale analysis. The pay-as-you-go pricing model, part of Google Cloud, means that the cost of using the service decreases as the load increases.

The obtained outcomes prove that the selection of Amazon AWS and Google Cloud providers should be made with reference to particular applications. In cases where latency is vital, and cost optimization is not paramount, AWS may be preferred. However, in cases of streaming data and where a great extent of scale-out is required, Google Cloud's propositions may suffice better. Therefore, This performance aspect must balance with the cost aspects for organizations to identify the most appropriate platform that suits their needs.

5.2 Practical Implications

These findings can help organizations make proper decisions about which cloud platform has to be chosen for serverless analytics. There may be situations when the organization's primary concern is the level of latency or the cost for small and mid-size tasks to be kept as low as possible; in these cases, AWS can still outperform on these aspects and the general simplicity of the price structure. This is particularly important for such use cases as user interaction or high-frequency transactional processing.

On the other hand, Google Cloud comes in handy for organizations that process massive amounts of data in real-time and manage streaming data applications. Using Google Cloud's Dataflow service, more significant amounts of data can be processed with less latency and better scalability, impacting applications that need a lot of data processing and analysis.

In addition, integration with current tools and services should also be convenient. Thus, for firms relying on Google's products, such as BigQuery, Google Sheets, etc., using Google Cloud unification can provide a smoother integration process. On the other hand, organizations that have already built an infrastructure relying on AWS services may stand to gain from using a platform that is highly integrated with the services it provides.

5.3 Limitations of the Study

Certain limitation factors should be considered in this research. One significant drawback is that the cloud is inconsistent in speed, possibly due to differences in regional data centers or network conditions that may cause fluctuations in cloud service performance. These conditions can lead to variability of the results and can be significantly different from the actual conditions during production.

Still, the analysis has some limitations, and the first one is realizing and identifying separate services in each cloud platform. For serverless computing, I focused on AWS Lambda and Google Cloud Functions; for serverless stream processing, I concentrated on AWS Kinesis and Google Cloud Pub/Sub. Similar services and configurations within these platforms might present other statistics, and a wider variety of services might still provide more accurate detection. AWS offers QuickSight a serverless Business Intelligence service while equivalent service from Google, Looker which offers Business Intelligence is not serverless and hence not included in the comparison.

Also, cost and performance benefits were evaluated mainly, although other factors like security and compliance and vendor support were evaluated to a lesser extent. All these aspects can also be very important in making a choice of the right cloud provider.

VI. CONCLUSION

6.1 Summary of Findings

This work compared and contrasted the serverless analytics solutions available on AWS and Google Cloud concerning their performance, cost optimisation, scalability, ease of use, and compatibility, among other factors. The analysis exposes several pros and cons for both to ensure organizations willing to adopt cloud analytics solutions have detailed information on what aspect to strengthen or avoid.

Regarding performance, it was observed that Google Cloud was comparatively better than AWS in the large-scale data analytics for which Google BigQuery delivered shorter query response time for intricate query on huge data. While AWS Lambda showed higher throughputs in terms cold start latency is much lower making it ideal for use in applications where quick response is desired after a period of inactivity. In terms of real-time data processing, AWS Kinesis outperformed Google Pub/Sub in terms of latency whereas Google Cloud's solution is easier to use within this ecosystem.

From a cost-optimisation point of view, Google Cloud was found to be cheaper for large-scale use and the fact that BigQuery was cheaper and more flexible with the larger amounts of data required. AWS was cheaper in less frequent and small size workloads as the pay for query model of AWS was suitable for organizations who require analytics processing occasionally.

Looking at Scalability; both the platforms had automatically scaling mechanisms good enough to handle large data volumes and increase in its processing loads. BigQuery of Google Cloud showed that it is easy to scale while performing batch processing at a large scale, whereas Kinesis of AWS was better for scaling in high throughput, real-time stream processing.

When it comes to the usability and integration, Google Cloud offered simplicity in how it was set up and designed, in addition, it offered native integration with AI and machine learning services to handle serverless analytics functions. AWS, although having a more intricate environment offered deeper levels of integration with other AWS services which is a plus for organizations that are already part of the AWS community.

6.2 Recommendations for Future Work

However, there are still many directions that can be researched following this study and comparing AWS and Google Cloud. One of the possible directions of the study could be the investigation of the other cloud services, including, for instance, Microsoft Azure. It appears that Azure's capabilities for embedded analytics, specifically relying on the serverless architecture, might be quite informative if compared to AWS and Google Cloud services, especially such as Azure Functions and Azure Synapse Analytics.

Furthermore, it is also possible to investigate the effect of serverless on security as a future research direction. One of the biggest disadvantages associated with serverless computing is the fact that organizations are largely dependent on cloud providers and hence can experience privacy and compliance issues, hacking among other system issues. More importantly, research into the protective measures provided by various Serverless computing platforms such as AWS, Google Cloud and exploring the risks posed in such paradigms, would be a great contribution to the existing discourse.

Last, more research can be done in niche areas including the deployment of serverless analytics in specific industries only, or look into the effects of extensive cloud analytics based on its system architecture concerning energy utilization and sustainability efforts in cloud computing. This would add more depth to the already increasing market needs for a 'serverless' solution to grow, optimize and secure analyses.

REFERENCES

[1.] Verbitski, A., Gupta, A., Saha, D., Brahmadesam, M., Gupta, K., Mittal, R., Krishnamurthy, S., Maurice, S., Kharatishvili, T., & Bao, X. (2017). Amazon Aurora. https://doi.org/10.1145/3035918.3056101

[2.] Cloud Computing: Comparison and Analysis of Cloud Service Providers-AWs, Microsoft and Google. (2020, December 4). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9337100/references#references

[3.] Google Cloud Vs AWS: A Quick Comparison For Starters. (n.d.). https://www.infiflex.com/google-cloud-vs-aws-a-quick-comparison-for-starters

[4.] Kratzke, N. (2018). A Brief History of Cloud Application Architectures. Applied Sciences, 8(8), 1368. https://doi.org/10.3390/app8081368

[5.] Singh Sachchidanand, Sewak Mohit (2018) Winning in the Era of Serverless Computing and Function as a Service. IEEE Xplore https://ieeexplore.ieee.org/abstract/document/8529465.

[6.] Ioana Baldini, Paul Castro, Kerry Chang, Perry Cheng, Stephen Fink, et al. (2017) Serverless Computing: Current Trends and Open Problems. ArXiv https://arxiv.org/abs/1706.03178.

[7.] Building Applications with Serverless Architectures. AWSServerless https://aws.amazon.com/lambda/serverlessarchitectures-learn-more/.

[8.] (2019) Introducing Cloud Functions: Serverless event-based programming in Google Cloud Platform. Google https://cloud.google.com/functions.

[9.] Vassalo JT (2019) Serverless architectures and their impact on the modern programming model. In 2019 IEEE International Conference on Software Architecture Companion (ICSA-C) 9-16.

[10.] Didinchuk, A. (2021, March 3). Google Cloud Platform (GCP) vs. Amazon Web Services (AWS) for the Hobbyist. Medium. https://appliedarchitect.medium.com/google-cloud-platform-gcp-vs-amazon-web-services-aws-for-the-hobbyist-e79326a8177

[11.] Pierleoni, P., Concetti, R., Belli, A., & Palma, L. (2020b). Amazon, Google and Microsoft Solutions for IoT: Architectures and a Performance Comparison. IEEE Access, 8, 5455–5470. https://doi.org/10.1109/access.2019.2961511

[12.] Azure Serverless. Azure https://azure.microsoft.com/en-us/solutions/serverless.

[13.] Serverless - AWS Lambda vs. Microsoft Azure Functions vs. Google Cloud Functions vs. IBM OpenWhisk Part 2 | PanonIT. (2020). Panonit.com. https://panonit.com/blog/serverless-aws-lambda-vs-microsoft-azure-functions-vs-google-cloud-functions-vs-ibm-openwhisk-0