IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 12, Issue 12, December 2024

DOI: 10.17148/IJIREEICE.2024.121207

Serverless Data Analytics with GCP Cloud Functions

Jakkoju Spanditha

Department of Management Studies, National Institute of Technology, Warangal, India

Abstract: Data analytics workflows expressed through Directed Acyclic Graphs run commonly today in serverless computing environments such as Google Cloud Platform because these settings provide better scalability and efficient cost benefits. Serverless workflows help scientists and business professionals deal with complex analytics requirements across astronomy, science, and social science, as well as bioinformatics and neuroscience. The data pipelines require specific programming models and runtime environments to handle their complex requirements. This paper investigates execution methods for data analytics DAGs through Google Cloud Platform services. This research demonstrates both the designed architecture and the achieved results of workflow implementations that will provide instrument pipelines as a service.

Keywords: Data analytics, serverless computing, Directed Acyclic Graphs, Google Cloud Platform, Cloud Functions, data pipelines, workflow management.

I. INTRODUCTION

The innovative Serverless computing framework of cloud computing enables developers to operate code without needing to worry about the administration of servers [15] [16]. The operational complexities vanish through this approach, which enables developers to write code and develop applications while cloud providers manage infrastructure deployments and scalability and ensure system availability [1]. Serverless architectures adapt to various use cases [2] because of their flexibility, which enables them to solve multiple computing requirements.

The various technologies of serverless computing include Functions as a Service, which represents the leading model as described [7][8]. The model allows developers to compose individual triggerable functions that respond to HTTP requests as well as database updates or message queue events [14]. Cloud providers scale the resources upon demand triggers while executing these functions to ensure the application remains available and responsive during peak usage periods [9][10].

Serverless computing operates with a native billing procedure where users pay for the exact duration their functions execute, thereby delivering cost efficiency. The server-based method contrasts with this approach by requiring users to pay for unused resources, which leads to potential cost inefficiencies [14].

II. LITERATURE REVIEW

The serverless approach enables scientific researchers to access cloud resources through a platform that streamlines the management tasks of physical servers. Cloud computing giants Amazon, together with Google, Microsoft, IBM, and Alibaba, introduced elastic computing services following the serverless concept in recent years [11]. The platforms let users access auto-scaling features through their pay-as-you-go pricing model, which provides economic computational services [13, 33,34]. Through this flexibility, scientists remain free to work on their research goals; thus, they can bring about faster discoveries that generate greater scientific advancement.
‘
The serverless computing industry shows rapid growth while researchers across multiple disciplines research its possibilities and solutions for existing challenges. New surveys about serverless computing bring essential knowledge about the field by showing researchers' activities together with statistical patterns alongside practical applications [17]. Figure 1 talks about the uses of Serverless Computing for different purposes, such as Research activity or to find statistical patterns.

Figure 1: Serverless computing uses

Surveys function as fundamental information tools because they provide an extensive overview of the field to both researchers and practitioners who track both current and future industry needs. The serverless computing movement demonstrates continued growth in industrial usage since more organizations choose the paradigm for different applications [2].

III. DATA ANALYTICS DAG EXECUTION IN GCP

Serverless computing solutions have become increasingly popular across commercial cloud platforms and open-source projects since they prove adaptable to different use cases. Serverless computing began life as a solution for event-triggered stateless applications before its adoption for running DAG workflow-based data analytics applications [5]. Scientific workflows find an effective execution solution through serverless containers since these systems provide flexible and scalable cloud instance provisioning according to demand patterns [4].

DevOps, together with Continuous Integration/Continuous Delivery (CI/CD) pipelines, use serverless computing in their approach to speed up the development of cloud-native applications [7]. Serverless computing offers excellent scale capabilities and flexibility, which drive many organizations to select this approach to manage their increasing internet service needs [11]. In serverless computing frameworks, the fundamental concept enables developers to build applications by writing code because infrastructure management is abstracted away from them [17].

Serverless computing brings distinctive features that create both development problems and deployment solutions for serverless applications [16]. Learning both the external environment and inner workings of serverless solutions offers better possibilities to deploy them effectively for fulfilling computational requirements [7][8][9][14].

IV. DISCUSSION

Scientific applications benefit from serverless computing, but several challenges, such as vendor lock-in and observability difficulties alongside performance-cost balances and distributed state management and caching requirements alongside tooling needs, continue to exist [12]. The successful implementation of serverless computing for scientific applications demands solutions to minimize vendor lock-in problems and enhance observability, along with better cost-performance optimization and easier state management systems.

DOI: 10.17148/IJIREEICE.2024.121207

When conducting serverless data analytics projects, workflow management plays a crucial role because it determines how tasks are executed in the system. Start by designing the DAG, which will break down the overall analytics procedure into individual components. Each serverless function needs to be implemented as its own individual function.

The workflow orchestration tool (like Google Cloud Composer and Apache Airflow) enables users to build and handle the execution of the DAG. The specific implementation for individual functions depends on the particular data analytics objective, yet the core concept remains unchanged because code modularization should produce separate, independent functions that function independently of each other. Figure 2 shows the cost of serverless computing, which seems to be a cost-effective way to compute systems.

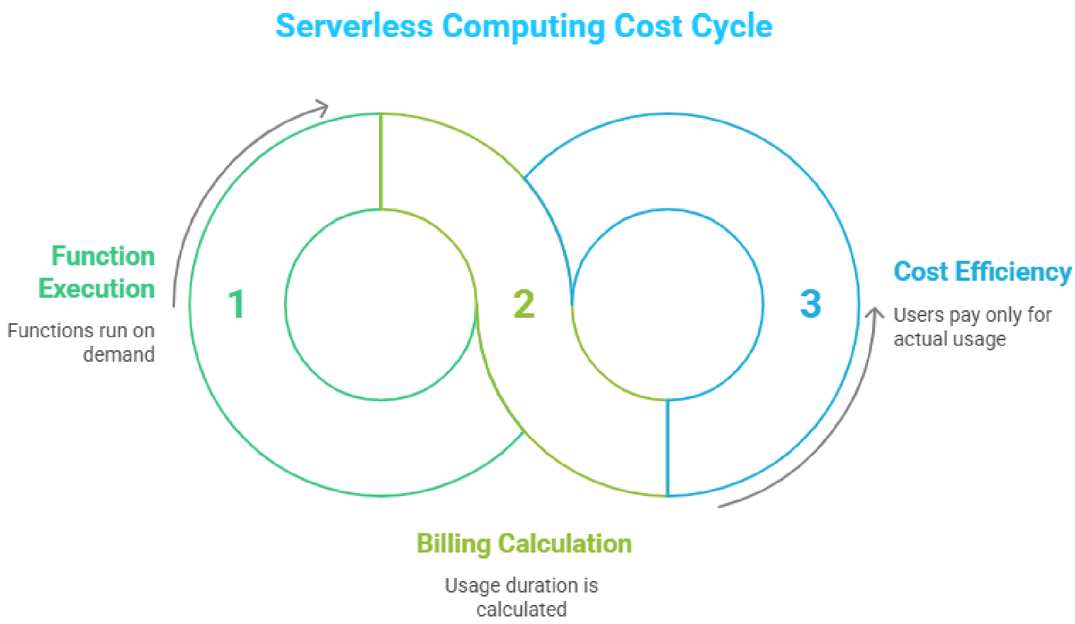

Figure 2: Serverless Computing Cost Cycle

The implementation of robust monitoring systems should include logs that enable staff to track DAG execution performance and progress points. Individual function execution tracking and logging enable administrators to find and resolve problems, maximize system performance, and guarantee reliable data analytics operation.

The strategic application of these planned steps allows organizations to use serverless computing systems for running data analytics DAGs, which brings forth scalability, cost performance, and enhanced agility. The documentation shows existing issues with DAG parallel job management, which primarily revolve around accessibility difficulties, complicated cluster management requirements, provisioning challenges, and configuration issues [6].

Guiding analysis at a functional level provides organizations with precise resource deployment capabilities and resourceable code libraries combined with independent scalability features that optimize budget use and resource efficiency. The implementation of approximate computing techniques, such as precision reduction techniques or data down sampling and step skipping, helps minimize both costs and response times, but it may also cause prospective quality impacts [19]. Caching represents a proven efficiency enhancement method that prevents tasks from redoing work that has recently occurred [28,29].

Application efficiency increases in serverless computing through proactive reuse approaches because they combine similar concurrent functions and requests into a merged operation even before existing tasks finish [7, 30]. The function orchestration system should adopt a data-centric method that matches data processing instructions to the stream of information [19]. The correct data order processing, together with transformation, establishes an efficient system that minimizes data movement while optimizing computational performance [31,32].

For serverless computing development, developers have the capability to build event-driven applications by creating functional sets that perform logical tasks one at a time [18]. The workflow engine AWS Step Functions enables developers to link their functions and implement control logic [3].

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁞ Peer-reviewed & Refereed journal ⁞ Vol. 12, Issue 12, December 2024

DOI: 10.17148/IJIREEICE.2024.121207

V. CONCLUSION

The second-generation approach of cloud technology seeks to ease programming stress through serverless computing technology, which helps developers and solution architects. The serverless market holds a projected value of $22 billion for the year 2024. Cloud programming becomes simpler through this solution because it enables users to acquire computing resources when needed and avoids the need for upfront financial commitments. Development teams dedicated to their business logic can focus on their work because the cloud provider takes charge of server provisioning responsibilities and maintenance tasks.

Serverless computing includes various technologies that base their operations on the Functions as a Service production model. A user function activated in serverless computing gets deployed and executed from a single virtual environment known as a sandbox. The cloud environment receives the function source code and dependencies and keeps them ready for execution when triggers start the process. The developer configures predefined triggers that automatically invoke applications. The diverse range of events includes uploading files to storage cloud systems, triggering HTTP requests, and managing timer schedules. The abstraction toolkit simplifies operational complexity. Serverless computing is driven by the potential for computational reuse and approximate computing, which can significantly enhance the efficiency of serverless cloud systems from both the user's Quality of Service and the system's perspectives.

REFERENCES

[1]. Addula, S. R., & Sajja, G. S. (2024, November). Automated Machine Learning to Streamline Data-Driven Industrial Application Development. In 2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES) (pp. 1-4). IEEE.
[2]. A. Chahal et al., “Systematic analysis based on Conflux of machine learning and internet of things using bibliometric analysis,” Journal of Intelligent Systems and Internet of Things, vol. 13, no. 1, pp. 196–224, 2024. doi:10.54216/jisiot.130115.
[3]. Baldini, I., Castro, P. M. E., Chang, K. S.-P., Cheng, P., Fink, S. J., Ishakian, V., Mitchell, N., Muthusamy, V., Rabbah, R., Slominski, A., & Suter, P. (2017). Serverless Computing: Current Trends and Open Problems. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.1706.03178.
[4]. Konda, B. (2024). Explore Data Mining (DM) Techniques That Data Scientists Adopt in IT.
[5]. Kasula, V. K. (2022). Empowering Finance: Cloud Computing Innovations in the Banking Sector. International Journal of Advanced Research in Science Communication and Technology, 2(1): 877-881.
[6]. Burkat, K., Pawlik, M., Baliś, B., Malawski, M., Vahi, K., Rynge, M., Silva, R. F. da, & Deelman, E. (2021). Serverless Containers – Rising Viable Approach to Scientific Workflows (p. 40). https://doi.org/10.1109/escience51609.2021.00014.
[7]. Kumar, D., Pawar, P. P., Ananthan, B., Indhumathi, S., & Murugan, M. S. (2024, May). CHOS_LSTM: Chebyshev Osprey optimization-based model for detecting attacks. In 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT) (pp. 1-6). IEEE.
[8]. S. Menon et al., “Streamlining task planning systems for improved enactment in contemporary computing surroundings,” SN Computer Science, vol. 5, no. 8, Oct. 2024. doi:10.1007/s42979-024-03267-5.
[9]. Yenugula, M. (2022). Google Cloud Monitoring: A Comprehensive Guide. Journal of Recent Trends in Computer Science and Engineering (JRTCSE), vol. 10, no. 2, pp. 40-50.
[10]. Denninnart, C., Chanikaphon, T., & Salehi, M. A. (2023). Efficiency in the serverless cloud paradigm: A survey on the reusing and approximation aspects. In Software Practice and Experience (Vol. 53, Issue 10, p. 1853). Wiley. https://doi.org/10.1002/spe.3233.
[11]. Yadulla, A. R. (2023). Leveraging Secure Multi-Party Computation and Blockchain for Collaborative AI in IoT Networks on Cloud Platforms. Journal of Recent Trends in Computer Science and Engineering (JRTCSE), 11(2), 54–59. https://doi.org/10.70589/JRTCSE.2023.2.9.
[12]. M. Yenugula et al., “Enhancing mobile data security with zero-trust architecture and federated learning: A comprehensive approach to prevent data leakage on smart terminals,” J. Recent Trends Comput. Sci. Eng. (JRTCSE), vol. 11, no. 1, pp. 52–64, 2023.
[13]. Kumar, D., Pawar, P. P., Ananthan, B., Rajasekaran, S., & Prabhakaran, T. V. (2024, May). Op-timized support vector machine based fused IOT network security management. In 2024 3rd In-ternational Conference on Artificial Intelligence For Internet of Things (AIIoT) (pp. 1-5). IEEE.
[14]. Hassan, H. B., Barakat, S., & Sarhan, Q. I. (2021). Survey on serverless computing. In Journal of Cloud Computing Advances Systems and Applications (Vol. 10, Issue 1). Springer Nature. https://doi.org/10.1186/s13677-021-00253-7.

DOI: 10.17148/IJIREEICE.2024.121207

[15]. Jonas, E., Schleier-Smith, J., Sreekanti, V., Tsai, C.-C., Khandelwal, A., Pu, Q., Shankar, V., Carreira, J., Krauth, K., Yadwadkar, N. J., Gonzalez, J. E., Popa, R. A., Stoica, I., & Patterson, D. A. (2019). Cloud Programming Simplified: A Berkeley View on Serverless Computing. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.1902.03383.

[16]. B. Konda et al., "Homomorphic encryption and federated attribute-based multi-factor access control for secure cloud services in integrated space-ground information networks," Int. J. Commun. Inf. Technol., vol. 3, no. 2, pp. 33–40, 2022.

[17]. Malawski, M., & Baliś, B. (2022). Serverless Computing for Scientific Applications. In IEEE Internet Computing (Vol. 26, Issue 4, p. 53). IEEE Computer Society. https://doi.org/10.1109/mic.2022.3168810.

[18]. Kasula, V. K. (2023). AI-driven banking: A review on transforming the financial sector. World Journal of Advanced Research and Reviews, 2023, 20(02), 1461-1465.

[19]. Daniel, V. A. A., Vijayalakshmi, K., Pawar, P. P., Kumar, D., Bhuvanesh, A., & Christilda, A. J. (2024). Enhanced affinity propagation clustering with a modified extreme learning machine for segmentation and classification of hyperspectral imaging. e-Prime-Advances in Electrical Engineering, Electronics and Energy, 9, 100704.

[20]. Misra NK, Das S, Satpathy S, et al (2024a) Covid-19 pandemic: A worldwide critical review with the machine learning model-based prediction. Journal of The Institution of Engineers (India): Series B 106:339–349. doi: 10.1007/s40031-024-01155-3.

[21]. S. Almotairi et al., "Personal data protection model in IOMT-blockchain on secured bit-count transmutation data encryption approach," Fusion: Practice and Applications, vol. 16, no. 1, pp. 152–170, 2024. doi:10.54216/fpa.160111.

[22]. A. R. Yadulla, M. Yenugula, V. K. Kasula, B. Konda, S. R. Addula, and S. B. Rakki, "A time-aware LSTM model for detecting criminal activities in blockchain transactions," Int. J. Commun. Inf. Technol., vol. 4, no. 2, pp. 33–39, 2023.

[23]. Wen, J., Chen, Z., Jin, X., & Liu, X. (2022). Rise of the Planet of Serverless Computing: A Systematic Review [Review of Rise of the Planet of Serverless Computing: A Systematic Review]. arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2206.12275.

[24]. Wen, J., Chen, Z., & Liu, X. (2022). Software Engineering for Serverless Computing. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2207.13263.

[25]. Wen, J., & Liu, Y. (2021). An Empirical Study on Serverless Workflow Service. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2101.03513.

[26]. V. K. Kasula, A. R. Yadulla, B. Konda, and M. Yenugula, "Enhancing financial cybersecurity: An AI-driven framework for safeguarding digital assets," 2022.

[27]. Yu, M., Cao, T., Wang, W., & Chen, R. (2021). Following the Data, Not the Function: Rethinking Function Orchestration in Serverless Computing. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2109.13492.

[28]. Kumar, D., Pawar, P., Gonaygunta, H., & Singh, S. (2023). Impact of federated learning on industrial iot-A Review. Int. J. Adv. Res. Comput. Commun. Eng, 13(1), 1-12.

[29]. Carver, B., Zhang, J., Wang, A., & Cheng, Y. (2019a). In Search of a Fast and Efficient Serverless DAG Engine. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.1910.05896.

[30]. Kumar, D., & Singh, S. (2024). Analyzing the impact of machine learning algorithms on risk management and fraud detection in financial institutions. International Journal of Research Publication and Reviews, 5(5), 1797-1804.

[31]. S. Almotairi et al., "Personal data protection model in IOMT-blockchain on secured bit-count transmutation data encryption approach," Fusion: Practice and Applications, vol. 16, no. 1, pp. 152–170, 2024. doi:10.54216/fpa.160111.

[32]. Pawar, P. P., Kumar, D., Meesala, M. K., Pareek, P. K., Addula, S. R., & KS, S. (2024, November). Securing Digital Governance: A Deep Learning and Blockchain Framework for Malware Detection in IoT Networks. In 2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS) (pp. 1-8). IEEE.

[33]. K. Meduri et al., "Leveraging federated learning for privacy-preserving analysis of multi-institutional electronic health records in rare disease research," Journal of Economy and Technology, vol. 3, pp. 177–189, 2024.

[34]. Shafiei, H., Khonsari, A., & Mousavi, P. (2022). Serverless Computing: A Survey of Opportunities, Challenges, and Applications [Review of Serverless Computing: A Survey of Opportunities, Challenges, and Applications]. ACM Computing Surveys, 54, 1. Association for Computing Machinery. https://doi.org/10.1145/3510611.