Cloud Resource Management: Monitoring

Anshika Rawat*
Department of Computer Science and Engineering, Amity School of Engineering and Technology Lucknow, Amity University Uttar Pradesh, India
anshika0405rawat@gmail.com

Pawan Singh
Department of Computer Science and Engineering, Amity School of Engineering and Technology Lucknow, Amity University Uttar Pradesh, India
pawansingh51279@gmail.com

Shubham Singh
Department of computer science and engineering, GLA University Mathura, Uttar Pradesh, India
singhshubham070@gmail.com

ABSTRACT

The management and use of computer resources within enterprises has changed as a result of the emergence of cloud computing as a paradigm shift. With its diverse and comprehensive portfolio of cloud services, including AWS CloudWatch for monitoring, Amazon Web Services (AWS) is at the vanguard of this revolution. In order to fully utilize the capabilities of these platforms and achieve cost-effectiveness, scalability, and optimal performance, proper cloud resource management is essential. This, study offers a thorough investigation of AWS's cloud resource management, with a focus on Particle Swarm Optimization (PSO) integration as a cutting-edge optimization technique. The dynamic and ever-changing workloads seen in cloud computing systems present a special set of issues for resource allocation and optimization. However, finding the ideal compromise between performance and cost-efficiency is still a challenging task. This study examines the changing environment of cloud resource management with an eye toward the future. In order to enable even more intelligent resource allocation decisions, we take into account the possible integration of machine learning and predictive analytics into our PSO framework. We also look at the broader implications of PSO-driven resource management.

KEYWORDS

Amazon Web Services (AWS), CloudWatch, Particle Swarm Algorithm (PSO)

ACM Reference Format:
Anshika Rawat, Pawan Singh, and Shubham Singh. 2023. Cloud Resource Management: Monitoring. In *International Conference on Information Management & Machine Intelligence (ICIMMI 2023), November 23–25, 2023, Jaipur, India.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3647444.3647899

1 INTRODUCTION

Large-scale data centres that handle a sizable amount of today's Internet applications and backend processing are now feasible thanks to advancements in commodity computing and virtualization technologies. The resulting economies of scale made it possible to rent out data centre infrastructure to third parties commercially. Thus, the cloud computing paradigm was created, in which programs that are accessed via the internet share a common pool of computing resources. The term "cloud computing" refers to both a) the leasing and use of resources by third parties as well as b) the private infrastructures maintained and used by specific businesses. Public clouds and private clouds have different names; cloud bursting, which happens when a business rents resources from a public cloud to expand the capacity of its private cloud, results in a hybrid cloud. Community clouds are another new category. The architecture of a cloud computing environment is the plan for all the different parts and how they work together. It includes both the services and applications that are provided through the cloud as well as the actual hardware, software, and networking infrastructure. Depending on the level of abstraction at which the service is provided, public cloud environments are often categorized as Infrastructure-as-a- Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS). Cloud computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services. The services themselves have long been referred to as Software as a Service (SaaS) [19]. Cloud computing is a technology that facilitates tasks by allocating virtual machine (VM) dynamically. Users charge resources as they use based on their demands [1]. A more adaptable and scalable strategy has replaced the conventional model of owning and controlling physical gear. In its simplest form, cloud computing allows users to access and use a variety of computing resources, such as servers, storage, and software, over the internet. This pay-as-you-go and on-demand model does away with the need for significant infrastructure investments up front, enabling organizations and individuals to scale their computing capabilities up or down as necessary. Cloud Computing is the result of over five decades of research and development in virtualization, distributed computing, networks, and solutions and services that have enabled the implementation of services and amenities offered by this solution [11]. It has evolved into the foundation of contemporary IT operations, encouraging creativity, cost-effectiveness, and agility across numerous industries. The traditional IT landscape has undergone a significant transformation as a result of cloud computing. Companies now rely on cloud service providers like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) to supply scalable and flexible resources on-demand rather than investing in and managing on-premises

*corresponding author.

ICIMMI 2023, November 23–25, 2023, Jaipur, India

ACM ISBN 979-8-4007-0941-8/23/11
https://doi.org/10.1145/3647444.3647899

infrastructure. However, in order to maximize the rewards of this paradigm change, efficient cloud resource management is crucial.

2 CLOUD RESOURCE MANAGEMENT

Resource management in a cloud environment is a hard problem, due to: the scale of modern data centers; the heterogeneity of resource types and their interdependencies; the variability and unpredictability of the load; as well as the range of objectives of the different actors in a cloud ecosystem [8]. Cloud resource management, which includes resource provisioning, allocation, and monitoring, is a key component of cloud computing. The resources available in the cloud include servers, memory, storage, networks, CPUs, application servers, and virtual machines. For the cloud's underlying hardware to be utilized effectively, resource management is crucial. Modern information technology infrastructure's key component, cloud resource management, has completely changed how companies and organizations store, use, and access computing resources. It basically means strategically allocating, optimizing, and keeping an eye on cloud- based resources such virtual servers, storage, and networking resources to ensure effectiveness, cost-effectiveness, and flawless performance.

One of the key aspects of cloud computing and virtualization is Resource Management (RM). RM is a process that deals with the procurement and release of resources [13]. There are various resource management systems in the cloud computing environment, each with a particular function based on the needs. The inter-cloud resource management system is an example of such a system, which unites multiple distinct clouds into a single fluid mass for on-demand operations. The inter-cloud makes sure a cloud can use resources outside of its coverage area by making use of existing contracts with other cloud service providers. Virtual machines installed on a cluster of cloud nodes are allocated physical resources by resource management systems.

Figure 1: Cloud Resource Management

Resource provisioning is an important aspect of cloud resource management. This entails figuring out how much processing power, storage, and network bandwidth are needed for a particular workload. Cost optimization is still another crucial aspect. Tools and best practices for resource management make it easier to keep track of utilization, spot inactive or underutilized resources, and put tactics like auto-scaling to work to match demand and, ultimately, cut costs. In cloud computing, cloud providers can offer cloud consumers two provisioning plans for computing resources, namely reservation and on-demand plans. In general, cost of utilizing computing resources provisioned by reservation plan is cheaper than that provisioned by on-demand plan, since cloud consumer has to pay to provider in advance. With the reservation plan, the consumer can reduce the total resource provisioning cost [18]. In cloud resource management, security and compliance are of the utmost importance. To safeguard data and applications housed in the cloud, organizations must put strong security measures in place. To maintain data integrity and regulatory compliance, these entails setting access controls, encryption, and constant monitoring.
In cloud resource management, monitoring and performance optimization are constant processes. Real-time monitoring systems keep tabs on resource usage, spot performance bottlenecks, and start scaling up or down operations to keep performance at its peak.

3 AMAZON WEB SERVICES(AWS)

In the face of fierce market competition and a dramatically changing business environment, companies are increasingly trying to integrate business processes with existing IS applications and build Internet-based technologies to conclude business transactions with trading partners [11]. Amazon Web Services (AWS) is a leading cloud computing platform and service provider offered by Amazon.com. It offers a wide range of cloud-based services, including computing power, storage, databases, machine learning, analytics, content delivery, and more. AWS is well known for its scalability, flexibility, and reliability, making it a popular choice for businesses, startups, and enterprises worldwide. AWS is appropriate for companies of all sizes and industries wishing to innovate, expand, and optimize their digital operations in the cloud because it offers low-latency access to services over a global network of data centres.

3.1 Role of AWS in Cloud Resource Management

AWS contributes significantly to the administration of cloud resources by offering a variety of services and tools that assist enterprises in:

- **Provision and Manage Resources:** AWS provides a wide range of cloud computing services, including storage, networking, databases, analytics, machine learning, and more. AWS also offers a variety of tools and services, including AWS CloudFormation, AWS Systems Manager, and AWS Resource Access Manager, to assist businesses in effectively provisioning and managing their resources.
- **Automate Task:** Task automation is possible with the help of a number of AWS services and technologies, including AWS Systems Manager Automation, AWS Step Functions, and AWS Code Pipeline. This can aid businesses in lowering the possibility of human error and enhancing the effectiveness of their cloud operations.
- **Monitor and Optimize Resources:** Resources can be monitored and optimized using a number of services and tools offered by AWS, including AWS CloudWatch, AWS Trusted Advisor, and AWS Cost Explorer. This can assist businesses in locating and fixing performance and cost problems as well as ensuring that their resources are being used effectively.
- **Secure resources:** AWS offers many security tools and services that can be used to safeguard cloud resources, including AWS CloudTrail, AWS Identity and Access Management, and AWS Key Management Service. By carrying out this, firms may be able to comply with their security and compliance obligations.

3.2 AWS Tools in Cloud Resource Management: Cloud Watch

It is possible to monitor, diagnose, and improve your AWS infrastructure and applications with the aid of the monitoring and management service known as AWS CloudWatch. It gathers and combines monitoring data in the form of logs, metrics, and events and gives you a range of tools and features to examine and visualize this data. Elastic Load Balancers, EC2 instances, RDS databases, and other AWS resources may all be monitored with CloudWatch. Furthermore, you may use it to keep an eye on logs and specific data from your own apps.

3.2.1 *Creating a Dashboard using AWS Cloud Watch.*

Step 1: Create an EC2 Instance for the dashboard

- Select Compute EC2 in order to create an EC2 instance.
- Choose the AWS Region for the EC2 server.
- Click on the 'Launch Instance' button.
- Choose Amazon Machine Image, select the default Windows (64 bit) AMI.
- Choose EC2 Instance Type
- Choose t3 micro which is a 2 vcpu and 1GB memory.
- Click "Configure Instance details." And specify information such as the number of instances, the purchasing options, networking information, storage additions, tagging your instance, and security group configuration

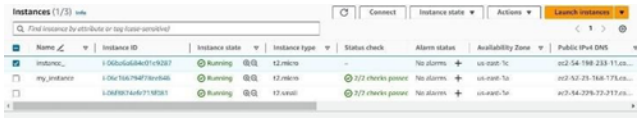

Figure 2: Created Instance

Figure 2 shows that instance by name 'my_instance' has been created. The link of this created instance will be copied and used further for making the dashboard.

Step 2: Creating dashboard Launch the CloudWatch interface.

- Select Dashboards from the navigation pane, and then select Create dashboard.
- Give the dashboard a name in the Create new dashboard dialog box, then click Create dashboard.
- Choose the widget type you want to add to your dashboard in the Add to this dashboard dialog box, and then click Configure.
- Choose Create widget after selecting the metric or metric(s) you want to display in the widget from the widget Settings dialog box.

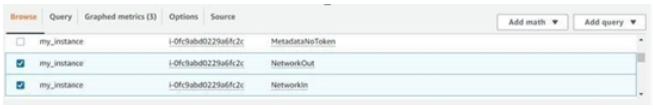

Figure 3: Selected Widges

Figure 3 shows the selected widgets which will be indicated in the dashboard.

- To add more widgets to your dashboard, repeat steps 4 and 5.
- Select Save dashboard once you've added all the widgets you want to your dashboard.

Figure 4: Created Dashboard

Figure 5: CPU Utilisation of Linux and Windows Respectively

Figure 5 shows dashboard of two instances created, here colour Blue for CPU Utilisation , Orange for NetworkIn and green for NetworkOut. Here it shows the CPU utilisation of created instances 'my_instance' and 'instance_' as shown in figure 2. The instances are created on different parameters. Instance one which is 'my_instance' is created on linux platform with instance type t2. micro while another instance which is 'instance_' is created on windows platform with instance type t2. micro. The figure 5 shows the impact on CPU Utilisation between two instances created with different parameters, Figure 5 (a) has CPU Utilisation of 0.33% while Figure 5 (b) has CPU Utilisation of 21.3%. Thus, this monitors the CPU Utilisation hence helps in resource management.

3.2.2 *Creating an Alarm for CPU Utilisation using Cloud Watch.*

- Launch the CloudWatch interface.
- Pick Alarms in the navigation pane then select Set alarm.
- Choose the CPU Utilization metric for the EC2 instance you want to track in the Select metric dialog box.
- Enter the alarm's threshold and comparison operator in the Configure condition dialog box. For instance, an alarm goes off when the CPU usage reaches 70% for longer than five minutes.
- Enter the actions you want CloudWatch to do when the alarm triggers in the Specify actions dialog box. You may, for instance, send a notification through email to a certain email address.
- Review the alarm configuration and select Create alarm in the Review and create alarm dialog box.

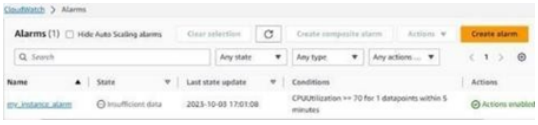

Figure 6: Alarm is created

3.2.3 Creating a notification on email address for the alarm created using SNS Console.

- Launch the Amazon SNS interface.
- Go to Topics in the navigation pane and choose the subject as you wish to create the notification.
- Choose Subscriptions and then create subscription.
- Go to Protocol and select Email.
- Enter the Email address and verify it to get the notification.
- Click on Create Subscription.

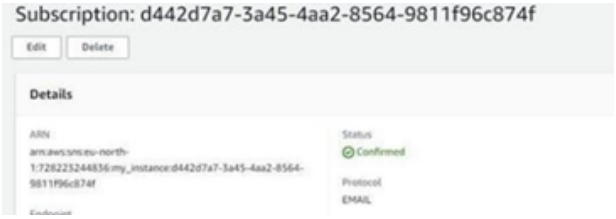

Figure 7: Subscription is created

Figure 7 shows that the subscription is created. When the CPU Utilization will exceed the limit that is 70% it will notify on the given email address.

4 CONCLUSION AND FUTURE SCOPE

With an emphasis on the use of Amazon Web Services (AWS) technologies like AWS CloudWatch and AWS CloudFormation, this article has examined the crucial elements of cloud resource management. We illustrated the significance of real-time monitoring and alerting for improving resource consumption and maintaining system dependability by creating customized dashboards and alerts in AWS CloudWatch. Cloud resource management has a very bright future as cloud computing continues to develop. The incorporation of cutting-edge optimization methods like Particle Swarm Optimization (PSO) is one such path. By dynamically changing resource provisioning and allocation depending on shifting workloads and user needs, PSO can play a crucial role in optimizing cloud resource allocation, workload distribution, and cost-effectiveness. As a result, businesses are able to improve performance, cost savings, and efficiency in their cloud settings.

A metaheuristic optimization technique called particle swarm optimization (PSO) imitates the group behaviour of social swarms like fish schools and bird flocks. It has been demonstrated that PSO works well in resolving a variety of optimization issues, including cloud resource management. PSO's potential in cloud resource management is bright. PSO is a viable technique for usage in large-scale cloud systems since it is easy to parallelize and implement. PSO is also resistant to noise and uncertainty, which is crucial in cloud environments where resources are frequently unexpected and dynamic.

REFERENCES

[1] Atul Adya, Paramvir Bahl, Jitendra Padhye, Alec Wolman, and Lidong Zhou. 2004. A multi-radio unification protocol for IEEE 802.11 wireless J. Acharya, M. Mehta, and B. Saini. 2016. Particle swarm optimization-based load balancing in cloud computing. In 2016 International Conference on Communication and Electronics Systems (ICCES).

[2] Constantin Adam and Rolf Stadler. 2007. Service middleware for self-managing large-scale systems. IEEE Trans. Netw. Serv. Manag. 4, 3 (2007), 50–64. DOI: https://doi.org/10.1109/tnsm.2007.021103.

[3] Orna Agmon Ben-Yehuda, Muli Ben-Yehuda, Assaf Schuster, and Dan Tsafrir. 2011. Deconstructing Amazon EC2 spot instance pricing. In 2011 IEEE Third International Conference on Cloud Computing Technology and Science, IEEE.

[4] J. Ahn, C. Kim, Y. R. Choi, and J. Huh. Dynamic virtual machine scheduling in clouds for architectural shared resources. In Proceedings of 4th USENIX Workshop on Hot Topics in Cloud Computing.

[5] Seema A. Alsaidy, Amenah D. Abbood, and Mouayad A. Sahib. 2022. Heuristic initialization of PSO task scheduling algorithm in cloud computing. J. King Saud Univ. - Comput. Inf. Sci. 34, 6 (2022), 2370–2382. DOI: https://doi.org/10.1016/j.jksuci.2020.11.002.

[6] J. Breddan and S. Rolf. 2013. Resource management in cloud: survey and research challenges. In Journal of network and system management.

[7] David Breitgand, Rami Cohen, Amir Nahir, and Danny Raz. 2010. On cost-aware monitoring for self-adaptive load sharing. IEEE J. Sel. Areas Commun. 28, 1 (2010), 70–83. DOI: https://doi.org/10.1109/jsac.2010.100108.

[8] Brendan Jennings and Rolf Stadler. 2015. Resource management in clouds: Survey and research challenges. J. Netw. Syst. Manag. 23, 3 (2015), 567–619. DOI: https://doi.org/10.1007/s10922-014-9307-7.

[9] S. Koush, R. Sohan, A. Rice, A. Moore, and A. Hopper. 2010. Predicting the performance of virtual machine migration. In Proceedings of 2010 IEEE International Symposium on Modeling, Analysis Simulation of Computer and Telecommunication Systems (MASCOTS 2010). IEEE, 37–46.

[10] Young Choon Lee and Albert Y. Zomaya. 2012. Energy efficient utilization of resources in cloud computing systems. J. Supercomput. 60, 2 (2012), 268–280. DOI: https://doi.org/10.1007/s11227-010-0421-3.

[11] Sean Marston, Zhi Li, Subhajyoti Bandyopadhyay, Juheng Zhang, and Anand Ghalsasi. 2011. Cloud computing — The business perspective. Decis. Support Syst. 51, 1 (2011), 176–189. DOI: https://doi.org/10.1016/j.dss.2010.12.006.

[12] Arabinda Pradhan and Sukant Kishoro Bisoy. 2020. A novel load balancing technique for cloud computing platform based on PSO. J. King Saud Univ. - Comput. Inf. Sci. (2020). DOI: https://doi.org/10.1016/j.jksuci.2020.10.016.

[13] M. Saad, N. Babar, H. Amir, A. Ur Rehman, and M. Sajjad. 2015. Resource managemnet in cloud computing:Taxonomyprospects and challenges. Elsevier.

[14] Shally, Sanjay Kumar Sharma, and Sunil Kumar. 2016. Energy efficient resource management in cloud environment: Progress and challenges. In 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), IEEE.

[15] Simeen Sheikh, G. Suganya, and M. Premalatha. 2020. Automated resource management on AWS cloud platform. In Proceedings of 6th International Conference on Big Data and Cloud Computing Challenges. Springer Singapore, Singapore, 133–147.

[16] S. Sindhu and Saswati Mukherjee. 2011. Efficient task scheduling algorithms for cloud computing environment. In High Performance Architecture and Grid Computing. Springer Berlin Heidelberg, Berlin, Heidelberg, 79–83.

[17] Jaspreet Singh, Bharti Duhan, Deepali Gupta, and Neha Sharma. 2020. Cloud resource management optimization: Taxonomy and research challenges. In 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), IEEE.

[18] C. Sivadon, L. Sung, and D. Niyato. 2012. Optimization of resource provisioning cost in cloud computing. IEEE Computer Society 5, (2012).

[19] A. View, Michael Cloud Computing, Armando Armbrust, Rean Fox, Anthony D. Griffith, Randy Joseph, Andy Katz, and Gunho Konwinski. Gunho Lee.

[20] J. Wd. 2016. Iaasmon: Monitoring architecture for public cloud computing data centers. Journal of grid computing (2016), 1–5.

[21] Andrew J. Younge, Gregor von Laszewski, Lizhe Wang, Sonia Lopez-Alarcon, and Warren Carithers. 2010. Efficient resource management for Cloud computing environments. In International Conference on Green Computing, IEEE.

[22] Qi Zhang, Lu Cheng, and Raouf Boutaba. 2010. Cloud computing: state-of-the-art and research challenges. J. Internet Serv. Appl. 1, 1 (2010), 7–18. DOI: https://doi.org/10.1007/s13174-010-0007-6.

[23] 2010. Resource pricing on federated clouds Cluster Cloud and Grid Computing. In Proceedings of the 10th IEEE/ACM International Conference on IEEE Computer Society. 513–517.

[24] 2017. Comparison of resource optimization algorithms in cloud computing. International Journal of Pure and Applied Mathematics (2017), 847–854.