Comparative Analysis of Best Cloud Service Providers for High Availability

Grigor Khachatryan [1]

Director of Infrastructure Engineering Lyve Global LTD (https://lyveglobal.com/en), Abu Dhabi (Headquarters) 30th Floor, Capital Plaza Business Tower Abu Dhabi United Arab Emirates

Summary
Using a cloud framework makes one able to easily create and manage cloud-based applications by combining various cloud computing aspects, such as development and deployment tools, middleware, and data storage, all of which are essential for cloud computing. Utilizing Big Data Analytics (BDA), businesses may store, analyze, and manipulate enormous amounts of information on the cloud. HA solutions are increasingly being used in the cloud because of their scalability, on-demand delivery, and flexibility. Many operators offer HA solutions that include floating internet protocols (IPs)along with the load balancing, but the examination of public cloud service disruption in past years suggests that a combo of numerous regions or virtual servers is needed to ensure "five nines" availability. Using several server locations and geographically dispersed clouds may boost efficiency and save costs in addition to decreasing failure risks. Moreover, the effective and adaptive provisioning of highly available multi-tier applications in various zones, as well as multi-cloud settings, presents different issues. Automated implementation and management of high-availability multi-tier infrastructure are made possible by launching innovative affinity processes like VM-to-location as well as role-based propagation or dispersion of modules among numerous clouds, and their correlating affinity rules.

Keywords:
Cloud framework, big data analytics, high availability, multi-cloud,

1. Introduction

In recent years, as the need for cloud services has grown significantly, the scalability of cloud platform users has seen a significant uptick. In the present era, cloud computing is one of the most important innovations. Customers and service providers alike benefit directly from the advantages of the cloud. In order to better serve their customers, corporations such as Microsoft, Google, and Amazon have often restructured their pricing policies [1]. Warehousing, upload, and download are only a few of the capabilities offered by cloud service platforms. Conventional data storage and management methods have been replaced by a new cloud-based approach. Cloud computing offers cost-effective and effective data management. Between the client and the service provider, multiple forms of SLA certificates are provided by the cloud. There are a variety of price options and incentives available to cloud clients.

An idea first floated by John McCarty at MIT in 1961 was that computers might be marketed as a utility like public utilities. Salesforce.com began offering apps to users through a basic website later in 1999. in the United States. The goal of computers as a utility was realized when programs were distributed to businesses over the Internet. Cloud computing became a key technical development, and many experts believe that cloud technology will have a profound impact on IT operations and the IT industry. Rather than having a local server or personal device manage an application, cloud computing delivers computing services via the internet. Servers, memory, analytics, networking, programming, statistics, and insight are all examples of computing services.

Maurizio Naldi and his colleagues did some work comparing cloud computing companies based on the prices they charge for their services, and their findings demonstrate that this is an effective way to do so [2]. In the meanwhile, Filiopoulou's study focused on computing service efficiency including computation-optimized cases, memory-optimized situations, and storage-optimized incidents, and it has been published [3]. Other Junjie Peng papers take advantage of the cloud services' functional needs [4]. As far as server location is concerned, this is tied to the country's restrictions for data center utilization, such as the regulations in Indonesia. Consequently, a comparison factor is required by the developer or user in selecting the provider and best cloud computing services, based on the demands and current rules for building an IoT solution that is connected with big data infrastructure.

1.1 Cloud Computing and Big Data

A massive amount of data (in the petabytes and Exabytes) is being disseminated every day by social media forums that includes YouTube, Facebook, Quora, WhatsApp, Stack overflow, Instagram, and several more online business apps, which are controlling the enterprise age. Big Data refers to the enormous amount of data created by messaging, satellite images, social media, email, and a host of other methods [5, 6]. We have presented the infrastructure of cloud computing in Fig. 1.

Manuscript received May 5, 2022
Manuscript revised May 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.5.59

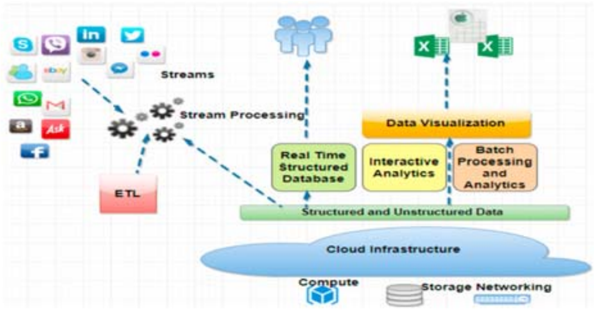

Fig. 1 Cloud Computing High-Level Architecture for Big Data

As a consequence, many businesses are struggling to keep track of and manage so much data that their expensive data warehouses are becoming clogged, resulting in a significant processing load [7]. Because of this bottleneck, a new technology known as BDA is being used by organizations to extract the Appropriate Data from these enormous amounts of unstructured information [8]. Big data analytics (BDA) has become more popular and accepted in the IT business, which has exacerbated several concerns and challenges [9]. This difficulty can be solved by using the cloud. As an example of fast internet service provisioning, cloud technology is a complete, user-friendly architecture that allows for on-demand access to a shared pool of shared resources while requiring minimum delivery administration [10, 11].

2. Related Work

In this part, three tools from the stated prominent manufacturers are picked for Data manipulation, data storage, and data warehousing & data processing, and their succinct features and advantages are presented.

2.1 Amazon Web Services (AWS)

AWS provides a broad range of BDA platforms that make it simple and rapid to develop and deploy BDA services.

As an open-source as well as an AWS-managed service, Amazon ElasticSearch (ES) provides real-time technical requirements, log analysis, and clickstream analysis in the AWS cloud. It's simple to use, has a scalable cluster, is widely available, and supports Open-Source APIs and tools, among other things [12, 13].

It is possible to store large amounts of data on Amazon S3 using a wide range of engines and several use scenarios, one of which is Big Data Analytics. Flexible regulatory and data storage options, unparalleled robustness, consistency and scalability, on-premises querying, integration with the widest possible range of vendor solutions, and the broadest possible range of security and compliance competencies are just a few advantages that come with utilizing this framework [11, 14].

One of the well-known, managed service AWS data warehouse solutions, Amazon Redshift, provides fast, simple, and cost-effective data analysis using conventional SQL and prevalent BI Tools. Amazon Redshift's advantages include lightning-fast queries, low costs, a simple and highly secure architecture, a dynamic and expandable cluster, compatibility with a wide range of SQL clients, and a large amount of extensibility [15].

2.2 Google Cloud Platform (GCP)

The Google Cloud Platform offers a wide range of powerful methods for a variety of applications, including big data analytics, data warehouses, databases, and memory.

It's obvious from the name that Google Cloud Dataproc is a cloud-based Apache Hadoop and Spark service designed to make cluster administration quick, simple, and inexpensive. Some of its advantages include quick cluster scalability, low costs, and an open-source architecture [11, 16].

Google cloud technology is integrated object storage that performs multiple jobs, from real-time data analysis to data preservation to data analytics. In addition to its cheap cost and high availability, other advantages of using this solution include improved archiving and storage, smooth and easy data migrations, low cost and better security for mission-critical resources, and partnerships with top vendor solutions [11].

Data warehouse for business intelligence in operations that is extremely scalable, fast, and inexpensive. A few of the benefits of employing this solution include fast infrastructure setup, easy scalability, effective analysis, as well as rapid insights into corporate data and investments protections [11].

2.3 Microsoft Azure

In order to build, launch, and manage applications ranging from mobile apps to ISC (Internet-Scale Computing) technologies, Microsoft Azure makes use of a global grid of data centers and provides DevOps and a unified framework to help its broad range of cloud services. In addition to data analytics and data intelligence solutions, their features and advantages are outlined below.

One virtual machine instance has a 99.999% SLA, while other industrialized services only guarantee SLA for important virtual machines. With 99.9 percent SLA, for example, Spark, Kafka, R Servers, Hadoop, and HBase Storm, maybe provisioned for optimized clusters. Features such as worldwide availability, high data security, and a highly competent framework for research and

development are among the most important aspects of this service [17].
Exabytes of unstructured data such as music, video, and picture files may be easily and cost-effectively stored in Blob Storage's three tiers, which are hot, cool, and archive. In addition to the ability to undertake modifications for application advancement and bandwidth sharp decline, a variety of blob types, such as append blobs, provide flexibility for storage optimization. It also enables automated geo-replication to enable easy autonomy of enhanced local and global connectivity [17].

2.4High Availability Computing Services

When one or more of the hardware or software components powering a service, including a cluster or a multi-tier application, fails, the service may continue to operate. Since no single system component fails, this ensures no system downtime and that the service keeps running, no matter how many components fail. Redundancy and replication are often used to decrease service downtime, and this is typically done by adding these characteristics.
HA approaches have been widely utilized to provide IT business operations by assuring a specific degree of organizational performance in the event of system or programmed downtime that includes schedule maintenance [18]. Single spots of failure are avoided by including duplicate instances of important system components, which is a common way to achieve high availability (HA).

2.5 High Availability Clustering

Multiple computer systems executing key service components are part of a high-availability (HA) cluster that may be configured in a variety of ways. If the main node fails in an active-passive arrangement, the secondary node may take over the important services. In an active-active arrangement, several cluster nodes are concurrently performing important service components. A load balancer divides the burden across multiple nodes, which typically provide the same service. Nodes in a cluster may take up the workload of a failed node. Heartbeat protocols are often used in HA clusters to keep tabs on the health of each node and alert administrators when a node goes down [19]. In many cases, a dedicated private network is used to link all of the cluster nodes, enabling this heartbeat technique to function.

The multiple cluster nodes in a HA setup are often operated on relatively autonomous embedded systems, for example on distinct physical hosts in the same data center, to minimize service interruption in the event of hardware failure. A data center outage might cause this solution to fail hence various options have been presented, like establishing the cluster in multiple data centers various time sectorsemploying the identical infrastructure, or even distinct clouds [20].

2.6 Multi-Zone HA

Cloud technology is becoming more popular as a means of deploying HA solutions due to its flexibility, on-demand deployment, and scalability, among other qualities. Cloud services and management infrastructures are now providing a variety of solutions for HA support, including bilt-in failover methods or elastic IP addresses.

In order to limit the risk of service interruption in the event of a zone failure, the availability zone functionality is an important consideration when deploying high availability clusters.

It is common practice in data centers and clouds to utilize Load Balancing methods to distribute and eliminate performance degradation in high-demand application settings.

Another noteworthy feature supplied by many cloud suppliers to help HA is the Floating IP Address.

Affinity is a crucial notion in enabling cloud-based HA setups to be deployed automatically.

3. High Availability Services

There are several ways to achieve high availability, but the most common are redundancy and replication for failover and load balancing, respectively. Clusters and multi-tier applications may benefit from these strategies.

HA cluster installations employing single and multiple cloud settings are examined and compared in this section. An example multi-tier service is discussed and some specifics regarding the deployment configuration are given in the first part of the article. Then, we examine the benefits of the multi-cloud strategy in terms of the cluster's availability. A single-cloud setup is shown in Fig. 2.

The application aspects are duplicated in two separate zones of the same cloud to increase availability. There are two different substantial clusters inside the cloud architecture, thus if one flops, the system continues to function.

A multi-cloud implementation of the identicalonline application as illustrated in Fig. 2. Replication of application modules takes place in both public and private clouds in this situation, ensuring service continuity in the event of cloud disruption.

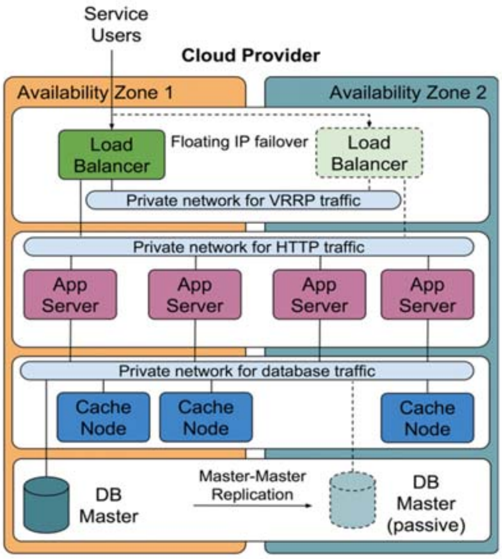

Fig. 2 Single Cloud Deployment

There are two separate availability zones for the application in this deployment, just as in the single-cloud operation.

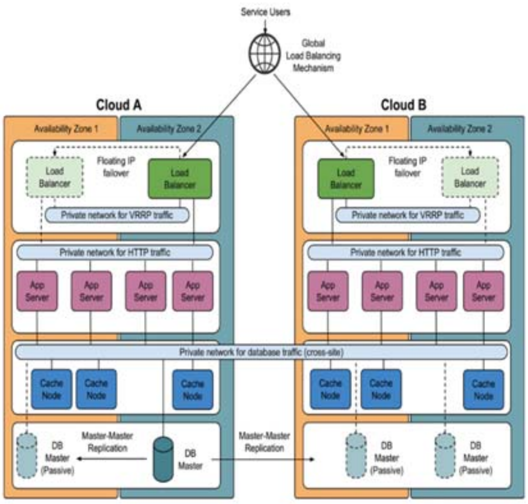

Fig. 3 Multi-Cloud Deployment

According to our findings, we presume that the application is in charge of distributing or duplicating various service components throughout both cloud environments, and we recommend that each cloud planner get a list of these aspects. The clouds have been deployed and their limits on placement (called "affinity rules") have been established, and each cloud performs its own, unique placement choices within those constraints in light of this, let us consider the issue of orchestration in additional issues in a multi-cloud situation other than those related to the multi-zone case.It is necessary to compare the costs of different multi-tier web service cloud platforms, such as single clouds and multiple clouds (as shown in Figs. 2 and 3), in order to see how much various portions of the service charge, like instances in Table 1, warehousing costs in Table 2, and operation costs in Table 3.

Table 1: Charges of Instances

Cloud Service	**Type**	**vCPU's**	**RAM**	**Price/Hour**
AWS	t2.xlarge	4	16 GB	$0.18
GCP	n1-std-4	4	15 GB	$0.19
Azure	A4-v2	4	8 GB	$0.19

Notably, only single-cloud installations based on AWS, GCP, and Azure will be considered, as will different multi-cloud configurations of these three services. This is because we are comparing the costs of these three services. As a result of the difficulty in comparing the costs of instances across multiple clouds, we've chosen specifically for cloud service a uniquely specific instance and vCPUs and at least 8 GB of RAM (shown in Table 1). It is anticipated, for the sake of simplicity, that these instance types are used to execute all service layers.

Table 2: Monthly Storage Charges

Cloud Service	**Type**	**Monthly/GB**
AWS	S3 Standard	$0.02
GCP	Multi-Region	$0.03
Azure	LRS-Hot	$0.02

Table 3: Operation Charges

Cloud Service	**Type**	**Charges**
AWS	PUT	$0.05
	GET	$0.04
GCP	PUT	$0.05
	GET	$0.04
Azure	PUT	$0.05
	GET	$0.04

Although the expenses of multi-cloud implementations are a little more expensive (approximately 15% more expensive, on average), the increased availability makes it worth the extra money to implement a multi-cloud strategy. There are a few major variations in the cost of single-cloud as well as multi-cloud implementations when evaluating the costs of different single-cloud with multi-cloud implementations.

3. Conclusion

Here, we've looked at how to manage HA service deployments across many zones and multiple clouds. a) Affinity methods and placement methodologies need to be improved to perform with zone limitations and role-based location limitations; b) global load balancing and fault-tolerance mechanisms have to be incorporated (multi-cloud scenario), and c) the requirement of implementing and organizing cross-site VPNs. In light of the fact that the last two difficulties have already been addressed, this study will concentrate on the first. An algorithm based on matching has been developed for scheduling in multitier services with multiple availability zones, and we've suggested novel affinity policies including role-to-role as well asa virtual machine to zonal affinity/anti-affinity rules to address common zone restrictions in multitier amenity. By presuming that each cloud planner is independent of the others, we've been able to expand our approach to multi-cloud situations by replicating and spreading the service over many clouds and specifying the necessary affinity rules for each one. According to our research, these multi-cloud installations boost uptime from three to six nines while increasing costs by roughly 15%.

References

[1] P. Srivastava, R. J. I. J. o. A. R. i. C. S. Khan, and S. Engineering, "A review paper on cloud computing," vol. 8, no. 6, pp. 17-20, 2018.

[2] M. Naldi and L. Mastroeni, "Cloud storage pricing: A comparison of current practices," in Proceedings of the 2013 international workshop on Hot topics in cloud services, 2013, pp. 27-34.

[3] E. Filiopoulou, P. Mitropoulou, N. Lionis, and C. J. I. T. o. C. C. Michalakelis, "On the efficiency of cloud providers: A DEA approach incorporating categorical variables," vol. 9, no. 1, pp. 272-285, 2018.

[4] J. Peng, X. Zhang, Z. Lei, B. Zhang, W. Zhang, and Q. Li, "Comparison of several cloud computing platforms," in 2009 Second international symposium on information science and engineering, 2009, pp. 23-27: IEEE.

[5] R. Math, "Big data analytics: Recent and emerging application in the services industry," in Big Data Analytics: Springer, 2018, pp. 211-219.

[6] I. Chebbi, W. Boulila, and I. R. Farah, "Big data: Concepts, challenges, and applications," in Computational collective intelligence: Springer, 2015, pp. 638-647.

[7] R. Zhang, I. Manotas, M. Li, and D. Hildebrand, "Towards a big data benchmarking and demonstration suite for the online social network era with realistic workloads and live data," in BPOE, 2015, pp. 25-36: Springer.

[8] A. Madaan, V. Sharma, P. Pahwa, P. Das, and C. Sharma, "Hadoop: Solution to Unstructured Data Handling," in Big Data Analytics: Springer, 2018, pp. 47-54.

[9] G. Skourletopoulos et al., "Big data and cloud computing: a survey of the state-of-the-art and research challenges," in Advances in mobile cloud computing and big data in the 5G Era: Springer, 2017, pp. 23-41.

[10] R. Vora, K. Garala, and P. Raval, "An era of big data on cloud computing services as a utility: 360 of review, challenges and unsolved exploration problems," in Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 2, 2016, pp. 575-583: Springer.

[11] S. Saif and S. J. P. c. s. Wazir, "Performance analysis of big data and cloud computing techniques: a survey," vol. 132, pp. 118-127, 2018.

[12] S. J. b. P. A. E. Gulabani, SQS, Kinesis,, "Practical Amazon EC2, SQS, Kinesis, and S3," vol. 3, 2017.

[13] V. A. Kumar, V. A. Kumar, H. Divakar, and R. Gokul, "Cloud-enabled media streaming using Amazon Web Services," in 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2017, pp. 195-198: IEEE.

[14] S. Krishnan and J. L. U. Gonzalez, Building your next big thing with google cloud platform: A guide for developers and enterprise architects. Springer, 2015.

[15] A. Nakhimovsky, T. Myers, and A. Nahkimovsky, Google, Amazon, and Beyond: Creating and Consuming Web Services. Springer, 2004.

[16] S. Krishnan and J. L. U. Gonzalez, "Google BigQuery," in Building Your Next Big Thing with Google Cloud Platform: Springer, 2015, pp. 235-253.

[17] M. P. Singh, M. A. Hoque, and S. J. a. p. a. Tarkoma, "A survey of systems for massive stream analytics," 2016.

[18] E. Marcus and H. Stern, Blueprints for high availability. John Wiley & Sons, 2003.

[19] A. Robertson, "{Linux-HA} Heartbeat System Design," in 4th Annual Linux Showcase & Conference (ALS 2000), 2000.

[20] N. Grozev, R. J. A. T. o. A. Buyya, and A. Systems, "Multi-cloud provisioning and load distribution for three-tier applications," vol. 9, no. 3, pp. 1-21, 2014.

Grigor Khachatryan (Գրիգոր Խաչատրյան), Director of Infrastructure Engineering Lyve Global LTD (https://lyveglobal.com/en), Abu Dhabi (Headquarters) 30th Floor, Capital Plaza Business Tower Abu Dhabi United Arab Emirates, ORCID: 0000-0002-6735-7149