IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

Different Techniques to Ensure High Availability in Cloud Computing

Wejdan Bajaber1, Manahil AlQulaity[2], and Fahd S. Alotaibi[3]

Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia[1-3]

Abstract: Cloud computing is subject to failures which emphasize the need to address user's availability. Availability refers to the system uptime and the system capability to operate continuously. Different techniques can be implemented to increase the system availability. Indeed, cloud service provider is taking the lead in providing highly available and cost effective services. However, these solutions are not always working as it claims. This paper conducts a literature review to investigate different techniques used to ensure the system availability. An analysis of high profile cloud service company is presented to evaluate the availability level of their services and to find out if high availability can be ensured or not.

Keywords: Cloud Computing, High Availability, Cloud service provider, Scalability.

I. INTRODUCTION

Cloud computing is a revolution in the world of technology due to the performance, accessibility, cost effectiveness, and other fancy benefits that it provides. Cloud computing is a combination of hardware and software that provide application and services over the internet [1]. The concept mainly empowers by the use of virtualization. Virtualization allows the users to save their data in a remote database instead of their local storage devices. Also, it only requires the internet to provide the connection between the user computer and the remote database; it also provides unlimited access to computing resources such as application and services with minimal management effort [2].

The central concept and purpose of cloud computing are the accessibility to files, data, programs and 3rd party services through websites via the Internet which is hosted by a 3rd party provider for free, and the user is only required to pay for the computing resources and services used [3]. Cloud computing advances include the internet backbone collection, internet broadband access adoption, powerful of servers and storage networks in the data center, the level of data center and high web performance and scalable software infrastructures. Cloud computing architecture includes many modules that built to maintain the system tools and activities, one of this module is the system resource management module which controls parallel running servers in a massive network. Also, it uses virtualization techniques to place computing resources dynamically. Some of the cloud computing advantages include the following:

- All of the computer resources provided and controlled by the 3rd party while the users only need to access the cloud system by housing space, support run electricity, and the system maintaining and administering cost, network, and database.
- The users have the ability to control the use level of the computing resources and services smoothly and clearly.
- The fees required from the user is only for using the computing resources and services.
- The users have the ability to access cloud service anytime and anywhere.

Availability is a critical factor for cloud computing as it's considered a significant requirement that needs to achieve. The users expect the system to be running 24/7 and thus, different techniques and technologies needed to be applied and implemented. High availability refers to the system uptime in which the system needs to perform and function its services in a continuous matter. Cloud providers are promising to provide highly available services and to minimize the system downtime by implementing different solutions and techniques. However, availability is always a questioning matter whether if we can ensure or is it always a subject of failure. This paper aims to investigate different techniques to ensure the system availability and to analyze different solutions and case studies to conclude the dilemma of availability in cloud computing.

II. LITERATURE REVIEW

High Availability refers to the capacity and the ability of a system to provide continuous services. Researchers are mainly concerned about discovering new technologies and different techniques that can improve security, performance, and availability of cloud computing. According to [4], high availability can be obtained at several layers of the system which include application, data, infrastructure and geographic location layer. Every single layer needs to provide a certain level of high availability but also, we need to consider the system as a whole to provide a complete and comprehensible service delivery platform. There are minimum two load balancers, web servers, and database

servers in the essential configuration at the infrastructure level. The user connected to the internet to the cloud. Also, to guarantee high availability, active and passive nodes should be deployed. One of cloud computing techniques is that nodes support each other so if one node lost connection, the other node supports the load and thus decreasing the downtime, and this process repeats at each configuration level. One of cloud computing features is dynamic scalability which can improve both load balancing and efficient control of the network traffic sudden increasing. This dynamic scalability can programmatically control through cloud servers Application Programming Interface (API). Scaling techniques have two main types; Horizontal and Vertical scaling [5]. In the environment of Windows Azure, the hardware resources are distracted and visible for the cloud applications to use them. A fabric controller control Windows Azure Fabric which is responsible for storage and computing resources detection [4]. Windows Azure announced fault domains and making fields where two VMs are in a single fault domain, so if one hardware failed, the network or power outage could put down the both other machines which mean that the Fabric Controller alert the application to use another virtual machine if one instance goes down. Windows Azure mainly perform electronic distribution of multiple instances across different fault domains which mean that a single outage does not bring down the role. This procedure guarantees that the application availability is attaining through minor influences of the downtime in a steady way. The traditional infrastructures must contain many servers in a stable and endless manner to ensure the system availability. However, there are more than a few third-party vendors who offer to supervise and cloud management tools. Though Service availability is not only about ensuring the server run time, it is much more. Also, the communication infrastructure between the cloud and the user need to be guaranteed especially it's outside of the user's control. Another technique is domains upgrade which used to updates service to a running application which means that if an in-place upgrade done to an existing application, Windows Azure rolls out one upgrade domain at a time which will ensure some instances of the service will always be available to process user requests. Therefore, fault domains and upgrade domains improve service availability of customer applications. Another side of availability is storage availability. Windows Azure stores three copies of user data in three different nodes that use different fault domains which will decrease the impact of hardware failures. And that is the reason of the customers' ability to open the same instance through various application by using different fault domains and upgrade domains [6].

Some authors in various papers determined that availability should use present information and predicting usage patterns and dynamic resource scaling [7]. defined load balancing as a technique that assists networks and resources by specifying the maximum throughput and the minimum response time. Traffic allocation between servers helps data to be sent and received without delay. Load balancers work for two pressing needs, mainly to increase the availability of cloud resources and secondarily to promote performance. As a result, load balancing will prevent the unreliable arrival of tasks at the cloud, and the cloud provides fast scaling up or down of resources to its clients, this will help to reduce the number of failures of the cloud system. In case, if failure existed in one part of the scheme the load balancer will move or switch to the other available resources in the system. This technique also will allow dividing servers' traffic which will reduce delay. Therefore, load balancers increase availability and performance as well [8]. According to [8] two main categories of load balancing algorithms that help traffic packed among available servers. Static Algorithm-distribute the traffic equally between servers. It's also known as round robin algorithm where the time allocated in equal shares for each process and in circular order "the process is handled without priority." This approach makes the situation imperfectly, many issues occurred in this algorithm. Furthermore, weighted round robin algorithms developed to improve the critical challenges related with round robin for efficient handling of servers with different processing capacities. Higher weights Servers receive new connections and more connection first, then fewer mass servers, and servers with equal weights receive steady traffic. Dynamic Algorithm- Required a real-time connection with the network to search for the lightest weight server to balance the traffic. Nevertheless, the real-time connection with the network to select the proper server will lead to additional traffic on the system [7]. The dynamic algorithm used to reallocate running tasks through available resources dynamically which will enable the functions to use the highest resource capacity [8].

In contrast between these algorithms, we can notice that round robin algorithm based on the basic rule, more loads considered on servers and as a result imbalanced traffic revealed. The dynamic algorithm based on the query that can regularly make on servers, but sometimes overcome traffic will prevent these questions to be answered, thus more added overhead can be illustrious on the network [7]. Cloud computing implements load balancing services to increase the number of CPU'S or memories to handle the growing number of scaled tasks [8].

Amazon Web Services has provided Elastic Load Balancing techniques that perform automatic distribution of incoming application traffic through multiple VM and scales its request handling capacity in response to incoming application traffic thus avoiding that multiple Load balancer's needs to be deployed for High Availability and thus to prevent additional cost [9].

Load balancer applies many models and rules based upon the reason of this applies. The network structure must be considered when creating the logical rules for the load balancer [7].

In [10], the authors advised a solution to increase cloud computing service availability which is by replicating the big data to multiple appropriate locations. Data replication is a well-known technique that enables data (e.g.,

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

database) to be available to the user in different and closer cloud applications. The mechanism used to minimize network delays, bandwidth usage, and increasing the availability. Dynamic data replication and Static data replication are two main classifications for replication algorithms. The main issues to be addressed in data replication is which data to replicate when to replicate, how many replicates to create, and where to allocate these replicates. All of the authors presented effective data duplication strategy to address data duplication issues powerfully. A mathematical model was generated to explain the relationship between system availability and the number of replicates. Identifying the characteristic data, and; it Also, duplication process will be caused when the popularity of a data file passes a dynamic starting point [10]. In [8], the authors mentioned binary weighted tree which used to decide which node highly available in the cluster for the next process. Replica placement method involves the placement of replicas among data nodes. Dynamic replication used based on access frequency, to fix the issues of node failure and access frequency.

Efficient replication method helps to increase the elasticity of the cloud system. According to [8], the number of replicas decided by replication number decision engine which prevents further increase of replication if a certain limit has reached. Replicas are located through data nodes in a consistency way, and the approach was evaluated and approved the efficiency of the improved system presented by the proposed strategy in a cloud [10].

Reference [11] Approach to cloud computing environment is clustering virtual machines in data centers; it mainly concerns about the resource higher availability besides the improved scalability. The problem to be solved is the performance in term of resource utilization that occurs due to poor virtual machine placement and migration in data centers. Clustering VM block contains a cluster manager that use K-means algorithm which contributes to collect similar types of object into one group and helps to know the number of the cluster in advance. The constructed architecture with multiple clustering in the same domain can solve the VM migration problem for the same user through different data centers in addition to reconfiguring and easy scheduling. Their proposed solution resulted in resource sharing optimization with increase usability in private cloud computing [11]. In reference [8], the authors defined another mechanism to ensure cloud computing availability "Deployment Choices." It is the method of deploying application components into virtual machines which will improve availability by the placement of these on physical. First, applying an analysis of some process which includes the best deployment strategy for the available software components selection process, the components replication number, and the placed components on the same machine. Implementation choices play a significant role in determining the availability of cloud applications. Executor refers to a component or a service that runs on a VM. There is some approach of deployment choice, the first approach includes combining all the task executors into one node, the second approach includes that one node for a job and the third method involves task executor groupings. The selection of executors for their corresponding VMs should be decided first, to maintain the placement process without any failure. The task executors are hosted on VMs that holds a collection of resources.

According to them the middleware architecture in multi-master pattern contains four components:

- User/ Client: where the user of cloud system needs to interact with the cloud manager then with the interface.
- Cloud Manager CM: connects through full-mesh topology, and they considered as master nodes. Their function as an interface for the user that allows him to request a list of the available hardware resources and to put the connection on hardware components.
- Cluster Controller (CC) - Node Controller (NC): They are connected to exactly one of the master nodes. They work individually to announce and update a list of all known cloud managers in the system.
- Backup Manager (BM): it is responsible for an automatic self-healing that fix the failure of CMs and when the CCs lose the connection to the cloud interface.

Cloud Manager is responsible for forwarding the available hardware components' request to all of the connected Cluster Controller which will transmit the request to the Node Controllers(NC), the NC will execute the request and returns the response to the CM that will return it back to the user. This architecture prevents an utter cloud failure in case of a failed master node [8].

Network scalability is also a necessary element of the infrastructure layer. An improved mechanism is considered to define the actual network usage. One could regularly specify the amount of the actual network usage for each application and let applications temporarily use other others allocated bandwidth. Statistical multiplexing improves the bandwidth allocation if some others are not using. By applying this mechanism, the actual bandwidth will be assigned to applications needs which mean that the users will only pay for the actual bandwidth consumption. The flow control, distributed rate limiting, and network slicing techniques support the cloud network provision paradigm. Statistical multiplexing used to optimize the network usage rate which indeed will measure the final bandwidth for each application [5]. The authors in [12] mentioned that other factors combined with statistical multiplexings such as decreasing the cost of electricity, network bandwidth, operations, software, and hardware available which can be used to increase system utilization that result in lower cost and good profit. Computation, storage and communications model needed for all application. The statistical multiplexing required to reach elasticity and the deception of unlimited capacity require virtualized resources to hide the multiplexing and sharing implementation.

Fault tolerance is a necessary technique to adapt system with software errors. There is two type of fault tolerance, Reactive fault tolerance that reduces the failures effects on the application execution like Check pointing/ Restart - When a task fails, it allows the user to restart from the recent point rather than from the beginning, it's a very useful technique, especially for long running tasks. Task Resubmission which considered as the commonly used fault tolerance technique currently in running systems. It allows submitting the task when detecting the failure to either the same or to a different resource at runtime, and so many other techniques. Proactive fault tolerance is implemented to prevent fault recovery, early prediction for errors and failures and proactively replace the failed components by other working components such as Self- Healing fault tolerance technique. Automatic failure handling will run if multiple application instances are running on multiple VM and pre-emptive migration which depends upon the mechanism of feedback-loop control where the application continually monitored and analyzed [13].

The authors in [14], conducted a research paper on how critical is adaptive resource management for fault tolerance of applications in Cloud computing. Cloud computing is subject to failures which emphasize the need to address user's availability, performance and security issues. The authors extended the concept of fault tolerance management with an online controller that realizes a heuristic-based algorithm to restore application's requirements at runtime in failures and recovery event. Markov chains and queuing networks algorithms are used to estimate the availability and performance attributes of a different implementation. By using models and simulation, they prove that their proposed approach was able to increase the availability and lower the degradation of system response times compared to traditional static schemes [14].

Kim in [3] provided a historical review of service availability of cloud computing; the authors mentioned some technical issues such as Amazon S3 which experienced two outages in 2008, 2 hours in February and 8 hours in August, also Google Gmail suffered an outage of 2 hours twice in August. In the author's opinion, it is impossible to provide 100% availability, but by adopting availability architecture, and applying a complete test for the platform and services applications, it might reach. According to [12], the organizations concern about utility computing services which should have acceptable availability, while SaaS products have high standards for availability techniques. Another important point is to ensure the desired level of availability by the service level agreement (SLAs) if 100% availability required the users need to take a combination of precautionary measures. Reference [3] mentioned that it would be better to keep on-premises storage backup, use a backup cloud, or not to store mission-critical data on the cloud, and for the applications. The users will need to keep an on-premises version of the application which means that they may need to work offline if the cloud is down. Another cloud computing critical issue is when the vendor changes out of business, or when they failed to provide the service, and that's why the users must select trustable vendors that have a consequence set of plans. To ensure cloud computing some technologies should be adopted which include the following:

- Cloud computing software platform.
- Collaboration applications.
- Application and data integration across clouds.
- Ongoing work on transferring multimedia and data mining
- service management

The authors expect cloud computing to become the computing paradigm core for the following years by reaching hardware and software consequences models which will make it easier for different size of enterprises to develop new services, as they are already established SaaS services on Amazon's Web Services [3]. In [12], they found out ten obstacles and opportunities to the growth of cloud computing, one of them in the service availability and they also mentioned some its opportunities which are the use of multiple cloud providers and the use of elasticity to prevent DDOS.

The authors said one of the solutions to avoid service availability failure which exists by using multiple Cloud Computing providers, so if one of them get off or out of service other will keep running the services. Another obstacle of the service availability is the distributed denial of service (DDoS) attacks where criminals decrease SaaS incomes by making their service unavailable. They solve this by using quick scale-up which gives SaaS providers the opportunity to prevent DDoS. In the author's opinion, cloud computing moves the attack target from SaaS provider to Utility Computing provider, who can understand it and treat DDoS protection as a core competency [12].

Reference [15] mentioned cloud computing as a computing infrastructure key that geographically spreads to supports computing organization. They attempted briefly to evaluate the fundamental techniques and techniques' challenges that improve system availability and they focused on the main technologies that are dependability of multicore processors, dependability of virtual machines, the reliability of the storage system and cloud infrastructure

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

and service assessment. The first technique to improved is by specific methods such as permanent failure tolerance and transient error tolerance; some technical challenges include increasing integration levels at the node level and increasing the CPU usage. The second technique is used to improve virtual machine by the fault and failure detection and recovery where the gap between those different systems can be solved by knowing the structure and understanding the internal implementation of those affected systems. Although the third technique can manage the increasing volume of data while maintaining stored data availability and consistency, there are several challenges such as achieving scalability, improving performance and deploying large scale failure detection. The fourth technique mainly concerns about the cloud service of real-world behavior that may differ from the service level agreement, some technical challenges are performance, consistency, availability, fault tolerance, and the cost [15].

III. PROBLEM STATEMENT

The conducted literature review investigated different techniques that were made to avoid a particular type of failure within the system. Most of the researchers focused on finding a solution to increase the system availability, and most of the techniques have resulted positively toward the system performance. Commercial solutions were not highly presented as a solution for highly available cloud systems, although they have existed for a while. With the huge budget invested in the industry section, cloud providers are now leading the market with highly implement solutions. Insight will be brought on some cloud computing providers, investigating their solution toward highly available systems. Different Case studies will be provided to illustrate real life examples.

IV. OBJECTIVES

This paper aims to investigate different techniques and different solutions for achieving high availability in cloud computing thus we could determine the answer to the following question:

- Can we ensure high availability in cloud computing?

V. METHODOLOGY

With the high rate of technology growth, a system may need sooner or later to expand and scale its capacity to accommodate different types of requests and acquired services. Developing a highly-available system encompasses an approach that goes through every layer of the system and identifies failure that should be reliable and available. With the fund and time required for research, development, and implementation of the emergent techniques and technologies; cloud providers seem to be the fastest solution. Reducing the cost of establishing hardware and software infrastructure, deploying, integrating, and maintaining the system is another reason. The services level agreement (SLA) between the client and the cloud provider guarantees a high standard of system availability with reasonable subscription fees. Cloud providers Industries, like Amazon Web Services, Google App, Microsoft Azure, Salesforce and much more has taken the lead in developing highly efficient and cost effective solutions that provide software, platform, and infrastructure services. The cloud service layer architecture will be illustrated as well as the availability applies to every layer with respect to the SLA agreement. Some case study from the cloud industry market will be provided to examine their approach and techniques in providing highly available cloud services to their clients as well as the obstacles and challenges they have encountered.

A. Cloud computing service layers' architecture

The cloud computing service layer consist of 3 layers, see Fig. 1

- Infrastructure as a Service "IaaS."

Infrastructure layer is the underlying layer that comes at the bottom of the cloud architecture diagram. It contains all the cloud physical resources such as servers, storage, and network that are virtualized to offer computational resources to the consumer [16].

- Platform as a Service "PaaS":

PaaS delivers software systems that are less flexible but easier to use by developers. It allows them to use the provider platform for developing and deploying their application without worrying about the operating system maintenance. One of the most famous examples of PaaS is Google App Engine.

- Software as a Service "SaaS":

It's the high layer of the cloud service layers' architecture. Allows the user to use applications and resources available in the cloud "online" without any required installation of an application on their physical devices. Users are limited to the functionality provided by the software with no control to add or manipulate the data from servers "e.g. Gmail" [17].

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

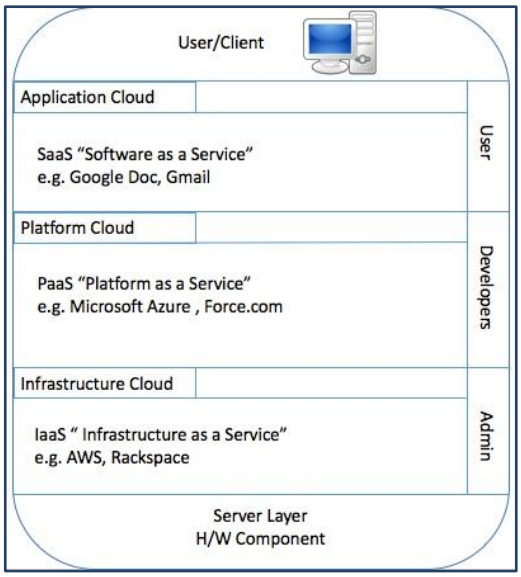

Figure (1): Cloud Computing Service layers' architecture

B. Service level agreement (SLA)

Are the contract and the agreement between the cloud service provider and the client that outline each party's role in delivering and using the system? It creates the metrics for assessing the performance of the system and the target level of availability and scalability.

Availability concept of cloud service layer in the aspect of SLA agreement:

- IaaS- The availability applies only to the infrastructure layer, platform, and software layer are not included in the SLAs.
- PaaS-The availability applies on the infrastructure on which the platform run on as well as the platform availability.
- SaaS- The availability on this layer applies on all of the cloud service layers. [17].

C. Measuring the availability

High availability concerns with the capacity of the system to provide continuous service, avoiding downtime, and the system uptime. Vendors describe availability as a given number of nines which represent the estimated number of minutes or seconds of system downtime over a specified period (monthly/annually). See Table. 1. Measuring the uptime is a direct function of SLA. The following equation describes the association between uptime and downtime:
A = 100 – (100*D/U) D = unplanned downtime, U = uptime; D, U expressed in minutes (ciurana, n.d).

Table I: Availability Measurement

Availability Percentage	Downtime in Min	Total Downtime/ Year	Vendor Jargon
99.99	52.56	53 min	Four nines
99.999	5.26	5.3 min	Five nines

VI. CASE STUDY

A. Salseforce.com

Saleceforce.com is one of the leader company that provides cloud software as a service (SaaS). It was established in 1999 as the first provider for customer relationship management (CRM) solution. They offer a complete integration of services and product to manage the interaction between the subscriber and their clients. Their products and application solutions run entirely on the cloud which means they are available online. In their annual report of 2016, the company stated that they have more than 150,000 customers and partners, more than 20,000 employees around the world, and achieved 6 billion$ in annual revenue before other competitors' company.

Their journey was not quite a smooth one, as they faced several incidences during their continuous growth and refined business. In 2009, an outage occurred due to a network device failing caused by memory allocation errors and prevented data in Europe, Japan, and North America from being processed for 38 minutes. Despite the customer expectation, that the cloud will be available almost 24/7, almost over 177 million transactions were affected. A few months later, a similar incidence happened, affecting Europe and North America for few hours which cause the customer to question the cloud availability of services [19].

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

Recently in 2016, huge outages occurred and lasted for two business days. One of their instances of North America called "NA14" was affected and the number of the affected customers was not released. Comfortable refers to the server that Salesforce organizations live on. The service disruption caused by a database failure on the NA14 instance. Due to a file integrity issue. The problem solved by restoring NA14 from a prior backup; that was not affected by the file integrity issues. Analysts commented that any IT system, in the cloud or on traditional premises will always be subject to service outages.

Therefore, cloud provider needs to develop a real hardware/software architecture and to implement robust techniques to ensure the continuity and the availability of their services. Jhonson, a Lead Site Reliability Engineer at salesforce.com, explained the structure that enables them to handle the Nemours amount of daily transactions. As first as Logging into salesforce.com a group of servers is required to handle login traffic for all instances. Once the traffic occurs in the data center of a returned IP address, the load balancer will be directed to which that IP exists. The load balancer directs the traffic to the application tier of the given instance in which they service standard web page traffic and API traffic. The core app tier contains 10 to 40 app servers; each server runs a single Hotspot JVM. Customers' data are backed up weekly in a single archive file format, besides a batch server which is responsible for scheduled running and the automated processes on the database tier. For handling, asynchronous processes on the content application tier, content search server, and a content batch server used. Salesforce.com is a database-driven system; Database performance is important. The data typically flows between two tires (the database tire and the core app server tier". The load on this tire need to be reduced thus, ACS -- API Cursor Server was developed. ACS considered as a cursor cache that runs on a pair of servers; it provides a method that offloads cursor processing from the database tier thus, allowing them to solve two main problems: first, cursors were stored in the database but unfortunately deleting them can impact the performance. Second, the DDL overhead became a negative impact in case of moving to database tables to hold the cursors; ACS enabled them to improve their database performance significantly. The development of Lifeforce system helped in reducing the load on the database tire too. All Binary Large Objects (BLOBs) "larger than 32 Kbit" are migrated to store on the Fileforce system rather than storing them directly in the database. Also, it supports a bundler function to reduce the disk seek load on the Fileforce servers. A process runs on the app server that gathers every 100+ object smaller than 32 Kbit into one single file while keeping the reference to the bundle file in the database. Other support servers are also implemented such as debugging application servers and application servers in the app tier [20].

B. Google App engine

One of the most famous PaaS examples is Google App Engine that targets web applications; it improves the separation relation between the dynamic computation tier and the static storage tier. Although that elasticity and virtualization in PaaS layer, especially in App Engine, play a major role, they are almost entirely invisible. To fill the capacity requirements changes, this model has automatic elasticity. Although Google provides a Secure Data Connector, it offers App Engine as a public service. It provides Secure Data Connector (SDC) which ensures secure access to the private data which called cloud bursting. The number of Google users increased through the past years which required them to get more powerful data center. Google built a scalable software infrastructure with almost 450,000 servers. In 2008 Google started PaaS type platform, started building web application by using runtime environment, they offered utility computing service and development tools.

Google App Engine offers different functions which include capacity, storage, and networking. It supported by three distinct environments: Python, Java, and Go. App Engine response to the user request as a request-response manner in the web applications where the program executes after receiving the user request and returns as a response, which assumes the CPU utilization time. As a result, for ever request Google perform ration CPU time separately. App Engine's aim to guarantee the mechanism of platform scaling and its high-availability, and also the private data storage such as Big Table. App Engine also run scheduled programs by using the Cron service which uses Big Table databases which are different than relational databases to store application data. App Engine manipulates Google Accounts by a user authentication mechanism where the user can sign into third party applications via their Google account. Also, Google offers open source set of tools to the developers known as an SDK (Software Development Kit). Application portability can be involved through its features which cause issues while moving service from App engine to another environment. Data extraction tool is needed and it was not available before to solve this problem, but in April 2009, Google announced that there would be a future edition. The dedicated application needed. To address the issue of the third parties in data migration.

Google also considered another possible platform lock-in which is related to Google Accounts Also by allowing the use of Google's authentication mechanism. Also, when Google's' users use their accounts by an external application, Google become more valuable because it registers more users on the system which therefore will improve Google's applications adoption. The final benefit of this will be gained by the third parties of this service because they will start using the service immediately, with no need of local registering to Google.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

The main standards that Google concern was about the infrastructure scalability and ease of use. App Engine sets a higher level of abstraction and concern about the smooth scaling, and the load balancing which means that they are responsible for the maintenance and the third parties are free of any maintenance issues. According to the openness cost and a potential platform lock-in which is related to the private database, Google prefers to retain its IP and to open the SDK code, which will influence other work in improving the development tools.

Another case study of Google service showed that some of the possible errors of the service availability can the caused by human error. It happened when one of their employees put some Gmail servers offline to perform maintenance, which caused routers to became pussy with traffic because there wasn't enough capacity to control the traffic increasing. In that time, Google brought additional request routers online for more capacity. Also, Google added more routers to separate the failures in data centers which will prevent the traffic overloads to affect another process

C. Menumate

Menumate is a sale point provider for hospitality software and hardware industry across Australasia. Menumate deployed Force.com PaaS advantages. Some of its applications are:

- License Key Generation: Which used to activate the customer paid features.
- Enhanced Case Management: Because of that most of the support case was done by consumables with separated DOS application which enables the client to place new with a receipt.
- Label Printing: it is a legacy application responsible for generating freight labels to send consumables and hardware to the clients.

According to Menumate case study, Daniel Fowlie and Abhinav Keswani are Directors of development house Trineo, which is a Menumate development boutique company. Fowlie stated that Force.com platform allowed centralization, modernization, and integration of all other disparate in-house software toolkit. Also, Keswani said that no needed infrastructure, connectivity or security to deploy Force.com PaaS, Force.com platform generate provides needed non-functional requirements which will allow Menumate and Trineo to develop the functions they need. In addition, by implementing PaaS approach, Trineo can reach the advantages of the automated deployment tools and the existing integrations. PaaS cloud has some benefits on their applications.

- The case of Key License generation: Quick code port to Force.com and the linked license keys to the saleforce.com CRM customer record, this will allow the sales and support staff check the license status.
- The case of Enhanced Case Management: allow PaaS Menumate to add a product, maintenance, and care of case and to send the receipt to their accounting programs by using current integration product.
- The case of Label Printing: enabling printing the freight labels from the customer record.

Utilizing a PaaS development environment produces a faster application. Also, according to PaaS absence, the cost of developing the application can be unaffordable.

D. Aire

London-based Aire started by the beginning of 2014, it has a credit score that can help people who have financial barriers such as living outside of the UK to get fair access to banks, mobile operators, and lenders for monetary products and services. Users who are willing to gain this help will first go through an exclusive, entirely online process, giving perceptions about their background and approaches to ensure the side of risk and credit.

Jon Bundy, the founder of Aire, mentioned that with AWS they get the redundant architecture that works well at all levels which prevent errors and need for other alternatives such as backing up databases. Furthermore, they only focus on raising a new product that should be noted in the market.

In the beginning, Aire was trying to create models and start the analysis of proof-of-concept to check the advantage in their idea. Aire founders aimed to start their process in a special environment that matches the flexibility of their operation at the beginning manner. Aire wants to build an infrastructure that can handle quick test and develop in an unconstrained way. Furthermore, Aire didn't build this architecture traditionally because this will prevent them to start their service quickly.

Aire founders already had experience using Amazon Web Services (AWS), and that is why they decided to start their system with AWS because they believe that AWS provides the quickest and the most cost-effective running system. Tim Kimball, head of engineering at Aire mentioned that Aire needed both highly available databases, applications, with security features which will enable them to log, monitor, and track access to their infrastructure. Because AWS gives their customers access to some of its technologies by AWS Activate program, Aire got quick access to AWS technologies which prevent significant upfront investment in IT infrastructure. Aire mainly concerns about Security and availability which already got higher priority in the startup.

In Aire founder opinion, getting a redundant infrastructure would be easy with AWS. AWS deploy high secured systems and this is one of the reasons, also they expect that AWS will deploy a platform that can anticipate customer requirements for future needs. AWS has different services such as Amazon Relational Database Service (Amazon RDS) which gives the firm resizable capacity which frees their users from any database management and any

administration tasks. Elastic Load Balancing which automatically distributes the executing application traffic through multiple Amazon Elastic Compute Cloud (Amazon EC2). All of the data and process through all services include Amazon RDS, Amazon EC2, and Elastic Load Balancing stored on Amazon Virtual Private Cloud (Amazon VPC) which by the end provide a highly available architecture that the client aims to.

AWS also guarantee security by Identity and Access Management (IAM) which allows creating, tracking, and managing users within AWS. For security perspective, AWS has different service such as AWS CloudTrail that records API calls, Amazon CloudWatch which allows Aire to control log files and alerts, and AWS Key Management Service (KMS) which offering logs to permit the client to check the submission needs, also to manipulate encryption keys and for some purpose integrates them with Amazon RDS and CloudTrail.

AWS gives Aire the ability to access technical expertise by their support for any needed information toward their system. Using AWS improve Aire reliability through the disruptive approach and the service runs reliably among the required security which helps to meet the requirement of the financial-services clients. By working in this flexible environment, the Aire innovation becomes cheaper and faster in contrast with traditional IT environment. AWS gives Aire user the ability to innovate, automate, develop through agile approach quicker than the traditional environment. Also, AWS enables Aire to perform cost-effective and repeatable way to meet the client requirements.

Aire team mentioned that the deployment start-up was satisfied with primary security requirements along the way, AWS keep every financial-services client with a set of third-party security requirements. Because of AWS ease starting and the working effortlessness tools, the operation starts in a lean and agile way with no need for any specialist to run the system. In Aire team opinion, AWS allows them to focus only on the main task of the system and the requirements of their business with no worries of any technical issues and the maintenance of the platform. Another feature is that AWS tools considered easy tools that can be handled by all type of employees. Besides, AWS allows Aire to deliver reliable infrastructure that customers need with redundant architecture in all, so there was no need for a backup database to cover any failure. AWS offers tools scalability with services reliably and securely universal service delivering. Because Aire grows, the ability to scale rapidly needed to increase their server, and also their service is worldwide, and that is why they need to consider regulations requirements outside of UK and for that, they chose AWS, because of their infrastructure is worldwide. They conclude by saying that the most significant advantage they gained by using AWS is by creating a highly available system, and the secured infrastructure for their customers and their ability to use AWS other features such as AWS big data workloads.

Another case study showed that some of the system failures were because of human mistakes, it was about configuration error in the scheduled update of the network. moving traffic from primary server to secondary server that used for data backup caused misconfiguration. Thus, overloaded the backup network because the network had more traffic to handle which indeed forced the software to launch a massive recovery effort, to avoid such mistake it is recommended to take more protection while updating the configuration. Also, all employees should be trained enough for configuration and updating. Another protection technique achieved by giving all of the employees the authority for manual shift traffic from one network to another. Furthermore, the cloud environment should be configured to move traffic from one network to another automatically according to the bandwidth needs.

VII. RESULT & ANALYSIS

The study and the analysis of the provided cases discussed the availability of the services offered by cloud service provider. High availability must be tackled in different areas of the cloud architecture layer. Each layer of the system has a different level of availability that needs to be achieved. Cloud computing companies have to address this area according to the service layer they are providing. Salesforce.com has encountered several outages that affect the availability of their services. Failures varied between network device caused by a memory allocation error, database failures resulting from a file integrity issue, and another kind of failures that were not relieved. However, there continuous improvement of their services and the used different techniques and solutions resulted in customers and revenue increase as shown in their annual reports. Google App Engine case study clarified different techniques that Google followed to improve their service availability and the errors that might occur during implementing or developing methods. Some of the applied techniques include application portability and lock-in. However, these improvement causes mistakes that can't be fixed as final system edition because the service runs through the time with increased number of users under the possible threats. Aire company chose AWS to host their services as they have implemented a different technique that improves the system reliability such as redundant architecture, security, and the reliable infrastructure, thus fewer errors will be encountered and database backup will be eliminated. Indeed, AWS offers scalability tools with reliable and secured services.

The analysis of the different high profile cloud provider shows both active and negative side of cloud computing service availability, as there are lots of improvement and highly available services, there is also room for system failure and outages. Failures can occur either on traditional IT environment or the cloud platform. Systems are exposed to different kinds of failures and challenges such as network vulnerability, resource management, human mistakes, server failure, storage failure, Power failures, and many others. Misconfiguration of the directorate services

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

can cause cloud outages. Thus, automatic configurations must be implemented to add validation checks which improve the detection mechanisms and service failure recovery. Cloud providers need to be prepared to avoid that kind of failures especially in the cloud infrastructure; that is hardware such as a server, storage, network, and other components that must be of high quality and undergo a moderate maintenance program. Some of the recommendations include data redundancy, failure detection, recovery, backup, auto scaling, (SDC), using BigTable for data storage, infrastructure scalability, high level of abstraction,
lock-in platform, redundant architecture, and Better Transparency of SLA Agreement. As a result, high availability cannot be ensured, but it can be increased and improved by avoiding common system failures through the implementation of different solutions and techniques.

VIII. CONCLUSION

Cloud computing is one of the major revolutions in the world of technology. In order for the user to optimize the benefits that it provides, the availability concerns associated with cloud resources have to be addressed. Different techniques to increase the availability of the cloud performance has been documented throughout this paper. Some of the techniques and their algorithms have been discussed in detail such as Fault tolerance, Dynamic scalability, load balancing, data replication, clustering VMs, and others. Cloud service provider has many solutions to improve the availability of the cloud resources. The analysis of some of the high-profile company such as AWS, Google App Engine, and Salesforce showed a different side of availability in the cloud. Most of the time cloud providers succeed in delivering highly available services yet; failures and outages were something they have to face at a time. Failures that might occur include but not restricted to the following: network vulnerability, human mistakes, server, storage or Power failures need to be avoided. Some of the solutions of the discussed cases to recover from some of the outages were: the high quality and the regular maintenance of the hardware component, data redundancy, failure detection, backup, auto scaling, using BigTable for data storage, infrastructure scalability, high level of abstraction, lock-in platform, and redundant architecture. As a conclusion, the cloud will remain subject to failure and failures can occur in the cloud as well as the IT traditional environment. Thus, high availability cannot be ensured, but it can be increased and improved, by avoiding common system failures through the implementation of different solutions and techniques.

REFERENCES

[1] Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I. and Zaharia, M., 2010. A view of cloud computing. Communications of the ACM, 53(4), pp.50-58.
[2] Mell, P. and Grance, T., 2011. The NIST definition of cloud computing.
[3] Kim, W., 2009. Cloud computing: Today and tomorrow. Journal of object technology, 8(1), pp.65-72.
[4] Ahuja, S.P. and Mani, S., 2012. Availability of services in the era of cloud computing. Network and Communication Technologies, 1(1), p.2.
[5] Vaquero, L.M., Rodero-Merino, L. and Buyya, R., 2011. Dynamically scaling applications in the cloud. ACM SIGCOMM Computer Communication Review, 41(1), pp.45-52.
[6] Sitaram, D. and Manjunath, G., 2011. Moving to the cloud: Developing apps in the new world of cloud computing. Elsevier.
[7] Chaczko, Z., Mahadevan, V., Aslanzadeh, S. and Mcdermid, C., 2011, September. Availability and load balancing in cloud computing. In International Conference on Computer and Software Modeling, Singapore (Vol. 14).
Vancouver
[8] Vani, B. and Priya, R.C.M., 2014. Availability in Cloud Computing.
[9] Amazon. (2016). Get started with Amazon AWS.. [online] Available at: http://docs.amazonaws.cn/en_us/aws/latest/userguide/aws-ug.pdf [Accessed 17 Mar. 2016].
[10] Sun, D.W., Chang, G.R., Gao, S., Jin, L.Z. and Wang, X.W., 2012. Modeling a dynamic data replication strategy to increase system availability in cloud computing environments. Journal of computer science and technology, 27(2), pp.256-272.
[11] Chavan, V. and Kaveri, P.R., 2014, August. Clustered virtual machines for higher availability of resources with improved scalability in cloud computing. In Networks & Soft Computing (ICNSC), 2014 First International Conference on (pp. 221-225). IEEE.
[12] Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R.H., Konwinski, A., Lee, G., Patterson, D.A., Rabkin, A., Stoica, I. and Zaharia, M., 2009. Above the clouds: A berkeley view of cloud computing (Vol. 17). Technical Report UCB/EECS-2009-28, EECS Department, University of California, Berkeley.
[13] Bala, A. and Chana, I., 2012. Fault tolerance-challenges, techniques and implementation in cloud computing. IJCSI International Journal of Computer Science Issues, 9(1), pp.1694-0814.
[14] Jhawar, R. and Piuri, V., 2013, July. Adaptive resource management for balancing availability and performance in cloud computing. In Security and Cryptography (SECRYPT), 2013 International Conference on (pp. 1-11). IEEE.
[15] Pham, C., Cao, P., Kalbarczyk, Z. and Iyer, R.K., 2012, June. Toward a high availability cloud: Techniques and challenges. In Dependable Systems and Networks Workshops (DSN-W), 2012 IEEE/IFIP 42nd International Conference on (pp. 1-6). IEEE.
[16] Zhang, Q., Cheng, L., and Boutaba, R., 2010. Cloud computing: state-of-the-art and research challenges. Journal of internet services and applications, 1(1), pp.7-18.
[17] Bahrami, M. and Singhal, M., 2015. The role of cloud computing architecture in big data. In Information granularity, big data, and computational intelligence (pp. 275-295). Springer International Publishing.
[18] Czarnowski, I. and Jędrzejowicz, P., 2014. Ensemble classifier for mining data streams. Procedia Computer Science, 35, pp.397-406.
[19] Laudon, K.C., Laudon, J.P., Brabston, M.E., Chaney, M., Hawkins, L. and Gaskin, S., 2012. Management Information Systems: Managing the Digital Firm, Seventh Canadian Edition (7th. Pearson.
[20] Johnson, C., 2014. Salesforce architecture-how they handle 1.3 billion transactions a day.
[21] Leymann, F., and Fritsch, D., 2009. Cloud computing: The next revolution in IT. Proceedings of the 52nd Photogrammetric Week, pp.3-12.
[22] Leavitt, N., 2009. Is cloud computing really ready for prime time?. Growth, 27(5), pp.15-20.

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

[23] Buyya, R., Ranjan, R., and Calheiros, R., 2009. Modeling and simulation of scalable Cloud computing environments and the CloudSim toolkit: Challenges and opportunities. High-Performance Computing & Simulation, 2009. HPCS'09. International Conference on, IEEE, pp1-11.
[24] Youseff, L., Butrico, M., and Da Silva, D., 2008. Toward a unified ontology of cloud computing. Grid Computing Environments Workshop, 2008. GCE'08, IEEE, pp.1-10.
[25] Ciurana, E. (2009). Scalability & High Availability - Dzone Refcardz. [online] Available at: https://dzone.com/refcardz/scalability [Accessed 20 Mar. 2016].
[26] case study: Aire. (n.d.). Retrieved from: https://aws.amazon.com/solutions/case-studies/aire/
[27] Koakowski, B., 2009. Platform Leadership in Software as a Service: How Platforms Facilitate Innovation.
[28] Gedymin, A., 2011. Cloud computing with an emphasis on PaaS and Google app engine. Barcelona: FIB.

DOI 10.17148/IJARCCE.2017.61102