Cloud Service Level Agreements and Resource Management

Isaac Odun-Ayo[*], Blessing Udemezue, Abiodun Kilanko

Department of Computer and Information Sciences, Covenant University, 112233, Nigeria

A R T I C L E I N F O	A B S T R A C T
	Cloud computing is a technical "as-a-service" usage model utilizing virtualization. Virtual machines are the core of cloud computing that runs as independent machine grouped into different networks within the hypervisor. Practically done by deployment of clouds computing hosts enterprise servers in virtual machines on an array of high-end servers. Managing cloud resources has to do with controlling and limiting access to the pool of available resources. This brings about the conception of an agreement between the resource services cloud providers and the customers otherwise known as Service Level Agreement, in other to restrict access to provisioned resources. Resource provisioning is a flexible on-demand pay-as-you-go package that is negotiated; and signed based on SLAs between customers and cloud providers. SLAs enable cloud providers to evade costly SLA consequences payable when there are violations, optimizing the performance of customers' applications and professionally manage resources to reduce cost. These documents are the agreements called Service Level Agreements (SLAs). SLAs outline the expectations which are: terms, conditions, and services of the clients from their service providers regarding availability, redundancy, uptime, cost, and penalties in cases of violations. These ensure clients' confidence in the services offered. Managing resource is an on-going major issue in cloud computing. Considering the limitation of resources, it poses a challenge for cloud service providers to make provision for all resources as needed. This paper seeks a solution to problems that relates to the present trends and developments of what cloud Service Level Agreements are and it ensures so by reviewing current literature. Thus, this research is a study of cloud Service Level Agreements, cloud resource management and their challenges. This paper has made provision to act as guidance for future research and it is anticipated to be beneficiary to potential cloud end users and cloud service providers.

1. Introduction

This paper is an extension of work originally presented in 2018 International Conference on Next Generation Computing and Information Systems [1]. This expands on resource management on Cloud computing and its challenges. Cloud computing makes provision for services by cloud service providers to simplify computing, storage and applications development by an end user using the Internet. Cloud computing is a computing method that enables ubiquitous service which is accessible to a mutual group of computerised resources. . Cloud computing is not limited to networks, services, storage and applications which can be swiftly made provision for and made available with minimal supervision involved [2].

In a cloud environment, resource management is challenging because of the scalability of present data centers, the heterogeneity in types of resources, the interconnection amongst these resources, the inconsistency and irregularity of the load, and the variety of aim of varied end users [3]. Cloud computing and management of resource provisioning are actualized at:

a. Infrastructure-as-a-Service (IaaS) - resources not limited to physical machines, virtual machines, and storage devices are offered as a service to customers.

b. Platform-as-a-Service (PaaS) - offers pre-installed software to customers for application development and testing.

c. Software-as-a-Service (SaaS) - offers a service instance with customizable interfaces to many customers in a cost-efficient manner [2].

Cloud resources management and cloud services are made available based on established Service Level Agreement (SLA) indicated and contracted between Cloud service providers and their end users describing the relations of what is agreed, which

[*]Isaac Odun-Ayo, Covenant University, +2348028829456
isaac.odunayo@covenantuniversity.edu.ng

also does not exclude non-functional requests, such as quality of service (QoS). In [4], the authors discussed that cloud end users prefer to choose a cloud service appropriate for them from the major service providers that can offer services with adequate quality of service guaranteed.

Lock manager is an independent algorithm that runs on cloud for one manager per server. Algorithm is a set of steps used to solve computing problems such as consistency in maintaining distributed files in cloud computing. There are many client requests made to different servers, requesting for granting simultaneous access to the same file. These requests run independently of each other and pool resources with one another to maintain uniformity of the files in the cloud. This algorithm contributes solution to the major concerns for availability and scalability of file resources [5]. Making the provisioning of resource management flexible and reliable and SLAs are very important to Cloud providers and their clients. Flexible and dependable provision of resource management and SLA agreements in SaaS, PaaS and IaaS are important to the Cloud providers and their clients. This enables cloud providers to evade exorbitant SLA penalties billed for infringements, to professionally manage resources, and lessen charge while elevating the performance of clients' applications. To assure SLA of customer applications, cloud providers need refined resource and monitoring of SLA approaches to accomplish sufficient data for the management procedure [6].

Cloud computing makes use of the concept of virtualization and multi-tenancy to provide service for their end users. Enterprises can migrate some of their applications and utilize services based on appropriate agreements. Cloud computing is made up of four types of deployment: the public, private, community and hybrid clouds. Organizations own private cloud. The amenities can either be in-house or off and the management is by a third party. Private clouds are often perceived to be more secure. Service providers that offer their services with a cost to cloud users majorly own public clouds. Public cloud make use of big data centres most times across many geographical locations. They are perceived to be less secure. Community cloud are run by some firms that have shared mutual interest. The management could be a community or a third party. The combination of either private, community or public cloud is known to be Hybrid cloud. Hybrid computing is cloud utilizes on the benefits of the different cloud types.

Service Level Agreement (SLA) is an official agreement between service providers and clients to assure consumers the quality service expectancy can be realised [6], Service Level Agreement lifecycle involves six phases: ascertain the cloud service providers, outline the Service Level Agreement, institute an agreement, observe likely SLA violators, end SLA and implement consequences [6]. Service Level Agreements are used to officially define the services that offered, the Quality of Service anticipated from cloud provider, duties of both parties involve and possible fines [7]. According to [8], Service Level Agreements ought to comprise of five simple matters. They are: set of services that are made available by the cloud service providers; an explicit and imminent evidence terms of services offered; set of Quality of Service metrics for assessing service delivery points; a way of monitoring these metrics; and means of make a decision during disagreements that may arise as a result of failure in meeting the Service Level Agreement terms. It is imperative for cloud end users to delight in the resources and service assured by the cloud

service provider, as the cloud service provider should gain optimally from the resources being provided [9].

The aim of this paper discusses cloud service computing, SLA and resource management. The paper discusses SLA in depth, looks at issues that relate to cloud computing and resource management from the industry perspective. The remaining paper is as follows. Section 2 studies related work. Section 3 converses Service Level Agreement in terms of resource management. Section 4 concludes the paper.

2. Related Works

Managing resources for cloud computing environments is an active field in research. We now present relevant work in Service Level Agreement, resource monitoring and management, framework, method, and model.

In [9], the authors presented dynamic data-centred virtual reality method that prevents Service Level Agreement defilements in combined cloud settings. This research offered an architecture that enables the release of end users resources without problems or violation of the Service Level Agreement. This achieved was through the use of several cloud service providers consumed by end users. SLA in Cloud Computing was offered in [10]; whereby the authors scrutinized SLAs between end users and cloud service providers. A Web Service Level Agreement structure was projected, applied, and confirmed. In [8], a structure for dialoguing SLA of Cloud-based Services was offered. The framework decides the best appropriate cloud service provider for cloud end users. These writers as well emphasised the significant characters of cloud dealers for optimal resource consumption by the cloud end user. The role of supremacy and other Service Level Agreement concerns in cloud environment were made known in [11]. The work untaken, includes main characteristics to be well thought-out when drafting up cloud service legal contracts in cloud computing. The problem of cloud domination was also evaluated with an understanding to determine right ways of handling information with least consequences. In [12], the writers proposed a method that concurrently provided for Quality of Service observing and resource optimal usage in cloud data centres. These authors offered a methodology which gathered user workloads into groups and used these groups to avert Service Level Agreement violation and at the time make better resource consumption, thus profiting both the cloud consumers and cloud providers.

In [2][13], the authors presented a resource monitoring framework which is push-pull technique for Cloud computing environments. This method is established in addition to dominant push and pull methodology monitoring method in Grids to Cloud computing with the aim to create a capable monitoring framework to be able to share resources in Clouds. In [14], the authors defined cloud federalism as being able to bring together services from different cloud vendors to provide a solution; an example is Cloud Burst. According to [15], on resource management, performing live migration is a recent feature concept within Cloud systems for active fault tolerance by flawless movement of Virtual Machines from waning hardware to unwavering hardware with the consumer not being aware of any change in a virtualized environment. Resource consolidation and management can be performed though the use of virtualization technologies. With the use of hypervisors contained by a cluster environment, it enables the consolidation of some standalone physical machines into a virtualized environment, therefore needing fewer physical resources when compared to a system without virtualization capabilities. Nonetheless, this

improvement is still inadequate. Big Cloud deployments have need of physical machines in thousands and megawatts of power.

Data-driven models can be used to manage resource cloud complex systems. The author in [16] discussed his dissertation and argued the advancements in machine learning for the management and optimization of today's resource systems through arising insights from the performance and utilization data these systems create. In other to bring to reality the vision of model-based resource management, there is need for the key challenges data-driven models raise to be dealt with such as uncertainty in predictions, cost of training, generalizability from benchmark datasets to real-world systems datasets, and interpretability of the models. Challenges of data-driven models includes:

a. Robust modelling procedures being able to work with noisy data, and to develop methods for model uncertainty estimates [17].

b. Data-driven model has to incorporate domain knowledge into the modelling process in order to increase the model accuracy and trust in the modelling results [17].

c. Transforming or migrating data from a relational schema to schema-less is difficult to accomplish due to data modeling challenges [18].

In [2], the authors designed a run-time monitoring framework name LoM2HiS to perform monitoring fixated towards increasing the scalability and practice in distributed and parallel environments. The run-time system observer target is to observe the services established on the negotiated and contracted Service Level Agreements by mapping rules; Cloud Service Provider's (CSP) clients request the provision of a contracted service; and the run-time monitor loads the service Service Level Agreement from the decided Service Level Agreement depository. Infrastructure resources are based on the service provisioning that represents hosts and network resources in a data center that host Cloud computing services. The design of this run-time monitor framework is aimed at being highly scalable.

There is a practice of dynamically allocating and de-allocating resources on physical machines in the public cloud environments [16]. As all kind businesses move to cloud, this leads to increase in varied workloads running in the Cloud; and each has diverse performance requirements and financial trade-offs [19]. According to [16], there are varied collection of Virtual Machine (VM) cloud providers, and the main problem of accurately and economically selecting the best VM for a specified workload resources is addressed. Considering the speed of growth in Cloud computing, some VM management platform solutions including OpenStack [13] has come into play.

Apache VCL in [20] is an open-source system for flexible resource providing and reserving of computation resources for different applications in a data center by the use of simplified web interface. In [16], the author discussed about Manjrasoft Aneka a service platform to build and deploy distributed applications on Clouds that makes provision for many applications for the transparency, exploiting distributed resources; and Service Level Agreement focused on allocating resource that distinguishes requested services based on their available resources and hence provides for consumers' needs.

3. Cloud Computing SLA and Resource Management

3.1. Service Level Agreement Concepts

1) National Institute of Standards and Technology (NIST)

The NIST describes cloud computing synopsis and recommendations in its publication [21], viewed from a viewpoint of cloud service provider. Based on this report, a usual profit-making cloud service level agreement has to consist of:

a. These agreements which should clearly include an agreement addressed to the consumer regarding constant availability of service, resolution of dispute steps, preservation of data, and legal protection of consumers' private information.

b. Cloud service provider has limitation of service offered due to effects of natural disasters that are beyond the control of cloud providers', outages in service, and updates. Cloud service providers are nonetheless obligated to provide realistic warning to their end users.

c. The anticipations from cloud consumers, which consist of receiving of terms and conditions including a fee for used services.

This NIST's point of view to Service Level Agreement is quite inflexible and baised in support of the Cloud providers. For a case in point, it does not take into contemplation choice for consumers to converse on the amendment to service level agreements with service cloud providers if the default SLA terms do not address all of the consumers' needs. The report also appeals consumers to know that Service Level Agreements may change at the providers' choice with practical reasons ahead of time notification. In order to be ready to transfer workloads to alternate cloud service providers in case of changes that might be undesirable. This though is not an easy activity, because of vendor lock-in and lack of standardization that could enable interoperability between cloud providers is still an on-going issue in cloud computing.

3.2. Service Level Agreement Levels

Some of the important levels related with cloud Service Level Agreements discussed:

a. Facility Level: The cloud service provider delivers on Service Level Agreement that covers the data centre services needed to conserve the client generated data and/or applications. These consist of things including electrical power, onsite generator, and refrigerating device. Service Level Agreements do assure high availability, fault tolerance, and data replication services.

b. Platform Level: There is need for physical servers, virtualization infrastructure, and network associated hardware possessed by the cloud provider and consumed by the cloud clients. Service Level Agreement at this level would contain information regarding physical security for the rejection of illegal accessibility to the building, facility, and computing resource. Checking background and analysing the character of staff should be carried out before employment by cloud service providers. [22] Every end user has control over his or her applications and some resources, while the CSP has complete control over the platform.

c. Operating System Level: A cloud service provider typically would convey certain volume of services managed to their clients. This additional service authorizes

the cloud provider to assure that the Operating System is properly continuous in operation so that it is reliably available. Service Level Agreements would enclose information regarding updates on security, system patches, confidentiality, or encryption, user authorization, and audit trails log files.

d. Application Level: This level provides safety against problems that associated with application level data. At this point, the cloud provider ensures the readiness, stability, and performance of their cloud user software which they are hosting. This is sometimes difficult to assure specifically in Infrastructure-as-a-Service and Platforma-as-a-Service whereby the end-user is only responsible for the application they put on the cloud.

3.3. Cloud Provider Service Level Agreement

The usual Service Level Agreement of cloud service providers include:

a. Service Assurance: This metric states the service level which a provider is obligated to over a contracted time frame.

b. Service Assurance Time Period: This explains the period of a service assurance ought to take place. The duration can be billed per month, or as contracted upon by mutual parties.

c. Service Assurance Granularity: This outlines the measure of resources for cloud providers to specify services assurance. For instance, the granularity can be as per service, per data centre, per instance, or per transaction basis. In relation to time of the assured service it can be fixed if granularity of service assurance is fine–grained. Service guarantee granularity can also be set up as accumulative of the considered resources such as contacts.

d. Service Guarantee: Oversights are occurrences which are not included in the service guarantee metric calculations. The oversight usually will contain mismanagement of the system by a consumer or an interruption connected with the programmed care.

e. Service Recognition: Is the account attributed to the customer or geared in the direction of forthcoming expenses if the service guarantee does not get realised. The total can be an inclusive or regulated acknowledgement of the consumer reimbursement for an overestimated service.

f. The Service Violation Measurement and Reporting: This explains how and who processes and informs of the violation of service assured respectively. The users typically do this but cloud service provider could also perform this role. In certain instances, a regulating service-monitoring third party could be assigned this.

3.4. Service Level Agreement Benefits and Challenges

2) Service Level Agreement Benefits

It is important to have comprehensive contract between the client and cloud provider to guarantee trust and confidence on both sides. The following are a number of the benefits of having a cloud Service Level Agreement [11]:

a. SLA supports solid understanding of cloud service and responsibilities of all stakeholders involved.

b. SLA assists the consumer to realise their understanding.

c. SLA emboldens clarity, accountability, and trustworthiness.

d. Make available information on team performance, capabilities, and staffing judgment.

e. SLA sees to the establishment of reassuring and communal functioning.

3) Service Level Agreement Challenges

All single agreement has certain challenges that requires both parties to deal with. These challenges include:

a. Service Level Agreements are difficult to achieve in cloud computing for the reason that some infrastructure and circumstances such as network and force majeure are beyond the control of both the cloud service provider and consumer, therefore challenging to draft a Service Level Agreement.

b. In circumstances where multi-tenancy is in use, Service Level Agreements relating to service separation and high obtainability may be challenging to for the cloud service provider for assurance.

c. Regarding cloud Software-as-a-Service model, Service Level Agreements are challenging to achieve due to the fact that it is almost impossible for the cloud service provider to assess all potential user software/application with diverse system configuration ahead of time.

d. It is challenging to come to an agreement on a cloud Service Level Agreement that is covering security. Distributed Denial of Service (DDoS), keystroke timing and side-channel attacks have been identified as some of the greatest common attacks in cloud environment [15] and continue to be a continuous challenge. As a result of the dynamic and always changing nature of these attacks, a cloud service provider can only at best give a general security-based Service Level Agreement and may not be capable of guaranteeing it.

e. In distributed and multi-cloud application deployment, Service Level Agreements are challenging to agree on as the several providers apply various standards. These standards are often proprietary and unable to interoperate therefore single Service Level Agreement for all parties concerned may be difficult to realise.

3.5. Resource Management and Challenges

4) Resource Management Models

Cloud provides a working tool towards achieving high efficiency in resource management and low cost service provisioning [2]. Sophisticated monitoring techniques are needed to cope with resources and the enforcement of SLAs by Cloud providers. Presently, most Cloud providers offer multi-tenant software as a service where provisioning of an application instance is made available to many customers. This has presented problems to monitor especially on the problem of how to ensure Service Level Agreements for diverse customers. The implementation of any application may affect the presentation of the others when Cloud providers make use of sharing resource for provisioning customer applications as a result of different economic reasons. Monitoring a Cloud environment can move to large number of nodes in a data center or distributed within geographical locations which is not a minor undertaking considering that the Cloud provider makes sure the monitoring processes does not make weak the performance of the applications being provisioned. In [23] [24] the authors discussed about ant colony algorithm, a job

programmed algorithm of cloud computing. Considering the cloud computing architecture and the that of ant colony algorithm analysis, managing cloud computing resources is intended to have varied rights management for its users, be it admins. It involves inspecting and the application status of host VM, monitor the task application and task performance scheduling, and the submission of computed tasks.

Sharing resources in cloud computing technology is established on virtualization of physical machines in order to attain most importantly security and performance separation. Virtual machines do share resources of physical machines when they are operational. These virtual machines provide individual customer services, and are shared among customers. It is quite challenging to effectively handle shared resources in order to decrease cost and achieve high utilization. According to [25], Autonomic resource management makes sure that Cloud providers serves huge number of demands without breaking terms of the SLA. With dynamism it manages the resources though the use of VM migration and consolidation. The upcoming computing paradigm is projected to be *utility computing;* that is, making computing services available when users require them, that is making computing services to become commodity utilities, just as internet, electric power, fuel, and telephone.

In [16], the authors pointed out that workload resource requirements are impervious, and during various points at production, it is tough finding out the resources critical for utmost performance. It has been difficult for hosted computing services like AWS Lambda for the load handled as black box functionality. It is understood that the performance detected may be due to primary environmental implementation, the allocation policies, issues performance isolation, or non-linear dependency of performance on resources accessibility. Monitoring resources expended while running task might aid in ascertaining resources for the utilization for that run, but will not indicate performance impacts in controlled set ups, or changed hardware or software.

The author Singh in [24], presented resource allocation model that evaluates users' request centered on resource requirements and resources that are available in order to get least run time. Information on resource represent if resources are available and configured. The resource allocation proposed algorithm assigns resource to negotiate the requirements for minimal run time. Client needs varies and heterogeneous therefore it is considerable to allocate using optimal solution. Availability of resource decides the allocation decision and performance, which brings about pool of various resources. Workload linked with the computing power available of the central processing unit is considered as below demand and present consumption of resources. In uncertain cloud environment, manage resources efficiently; allocation issue executed with swarm optimization tactic made better, the likelihood to look for best appropriate resource for task with least run time possible. Swarm practice is built on nature motivated and artificial intelligence system which involves of self-sufficient agent with combined activities with the use of decentralized system[24] [26].

These authors in [27] presented Ant Colony Optimization (ACO) based energy aware solution to address the problem of allocating resource. This proposed Energy Aware Resource Allocation EARA methodology aims to optimize the allocating of resources so as to make energy efficiency better in the cloud infrastructure while meeting up to the QoS requirements of the end users. The allocation of resources to jobs is based on their QoS requirements. EARA uses ACO at two levels which are: the level where ACO allocates virtual machines resources to jobs, and the level where ACO allocates physical machines resources to Virtual Machines. EARA is handled from the cloud service provider's and end users' point of view. This reduces: the overall energy consumption for the benefit of cloud service provider, and the overall execution cost and entire time of execution for the clients' satisfaction. Benefits of ant colony algorithm includes [28]:

a. Making use of decentralized approach to provide good solution to its single point of failure

b. Collecting information faster

c. Reduces load on networks.

Challenges of ant colony algorithm includes [28]:

a. Dispatching a lot of ants which result to network overloads.

b. Some operating parameters are not taken into consideration which might result to poor performance.

c. Points of initiating ants and the number of ants are not established.

In [29], the authors Zhang and Li proposed a user utility oriented queuing model to handle task scheduling in cloud environment. These authors modelled task scheduling as an M/M/1 queuing system, classified the utility into time utility and cost utility, and built a linear programming model to maximize total utility for time utility and cost utility. They also proposed a utility oriented algorithm to make the most of the total utility. They described the randomness of tasks with the $M/M/1$ model of queuing theory. This utility model involves one server, some schedulers, and some computing resources. When users' tasks are submitted, the server analyzes and schedules the tasks to different schedulers, and adds them to local task queue of the corresponding scheduler. This model can reschedule remaining tasks dynamically to get the maximised utility. Finally, every scheduler schedules its local tasks to obtainable computing resources. The main aim of utility computing is to make provision for the computing resources, which includes: computational power, storage capacity and services to its users while charging them for their usage [30]. In [31], the authors stated that utility computing makes provision for services and computing resources to customers, such as storage, applications, and computing power. This utility computing model services is the foundation of the shift to on demand computing, software as a service and cloud computing models. Benefits of utility computing:

a. The end user have no need to buy all the hardware, software and licenses needed in order to conduct business with cloud service provides.

b. Companies have the option to subscribe to a single service and use the same suite of software throughout the whole client organization.

c. It provides compatibility of all the computers in large companies.

d. It offers unlimited storage capacity

Disadvantages of utility computing:

a. Reliability issue. This means the service could be halted from the utility computing organization for any reason such as a financial trouble or equipment problems.
b. There are many servers across various administrative domains

Some utility computing driving factors [30]:

a. Virtualization enables the abstraction of computing resources so that a physical machine can function as a set of multiple logical virtual machines.
b. Resource pricing in other to develop a pricing standard that can gainfully support economy based utility computing paradigm.
c. Standardization and commoditization are important use of utility services. Standardized information technology services facilitate transparent use of services and make services ideal candidate for utility computing. The commoditization of hardware and software applications such as office suites is increasing with the acceptance of the open standards.
d. Resource allocation and scheduling handle user requests to satisfy user quality of service and maximize profit. A scheduler determines the resources based on what is required of a job which is a user request, performs the resource allocation, and then maps the resources to that job.
e. Service Level Agreement defines service constraints required by the end users. Cloud service providers know how users value their service requests, hence it provides feedback mechanisms to encourage and discourage a service request.

In [32], the author Kim proposed a resource management model that CloudStack supports. CloudStack is an effective resource model that enables scalability, multi-tenant, and it is a cloud computing software that manages the creation and deployment of cloud infrastructure. The system manager is unable to monitor VMs computing resources, but increases visibility into resource allocation and can look up the status of Virtual Machines such as if it is on or off at the console of cloud management. Comprehensive information regarding the status of the Virtual Machine chosen by the system manager can be monitored; examples are the central processing unit use rate, disk, and network activities. The main aim of this management model for resource is to evaluate unprocessed data, then report the statistical information about the Virtual Machine usage volume for the computing resources contained by the resources constraint. What this resource management model specifically does is that:

a. It enables system manager to look into the statistical analysis and the status from the log analysis framework from communication among VMs.
b. It aids system manager in generation of rules for the present constraint of obtainable computing resources gotten from the log analysis framework. An instance is when users make request and create Virtual Machines using a quad core central processing unit, the system manager has the choice of scaling down from a quad core to a dual core for the sum of central processing units

constructed on the statistical report from the log analysis framework.
c. Based on firewall policy, it provides the chance to enable the system manager to script enabling the cloud management console to automatically, measure the virtual machines upward or downward.
d. Enables the system manager to produce policies during emergencies for service stability of the virtual machines.

5) **Resource Management Challenges**

Traditional resource management methods are not enough for cloud computing because they are on virtualization technology with distributed nature [33]. Considering the growth in Cloud computing, it has challenges for managing resource because of heterogeneity in hardware capabilities, per request service model, pay-as-you-use model and assurance to satisfy quality of service. Managing resources dynamically will reduce the resource controversy, scarcity of resources, fragmentation of resource, and overflow / underflow problems. To circumvent these difficulties resources should be adjusted with dynamism to manage resources efficiently in the cloud architecture [34].

There are challenges in the capacity to monitor resource and to manage SLAs in cloud computing, they include [2]:

a. Distributed Application Execution – NetLogger is a distributed system that monitors in order to observe and collect information of computer networks. In [2], the authors discussed ChaosMon, an application that monitors and displays performance information for parallel and distributed systems. This monitoring system is aimed to support application developers to enable them specify performance metrics and to also observe them visually in order to identify, evaluate, and comprehend performance logjams. This system is a distributed monitoring that has a centralized control.
b. Enforcement of the Agreed SLA – Enforcing SLAs is challenging and it involves complex processes whereby both parties might have to keep going back and forth until a mutual agreement is achieved. Implementation of different technologies poses challenges to evaluate Service Level Agreement documents, multifaceted deployment procedures, and scalability concerns. Ensuring uninterrupted monitoring and dynamic resource provision can be employed to put into effect agreed SLA.
c. Energy Management - Managing the consumption of energy is becoming challenging. Statistics and report show unchecked energy consumption might be a key problem in few years' time. According to [14], Stanford University made known that there is an annual increase of 56% in the consumption of electricity in data centers between 2005 and 2010.
d. Multi-Tenancy Application – This provision is becoming standard amongst Cloud providers to cost efficiently provide their sole instant application to numerous consumers. The technology approach is still challenging in its realization. In [2], the authors discussed proposed SLAs that is multi-tenant oriented in monitoring performance, in detection, and to schedule architecture with their clients. The writers aimed towards separating and guaranteeing optimized performance of each client in the Cloud space. This is complex and challenging to manage such resource because the cloud provider is making provision for every tenant with the same

application instance whereby the Service Level Agreements for the applications are contacted on per tenant basis. In situations like this, the workload of some tenants' applications will affect the unusual resource consumptions by other tenants considering the Cloud hardware, and all tenants share software resources. To solve this issue, a dynamic SLA mechanism architecture is designed, which monitors application grades, discerns what is not normal, and dynamically schedules shared resources to make sure of tenants' SLAs and optimize system performance.

e. High Availability - This is the server's availability for use for its intended purpose. High availability main objective is to ensure a system continues to function regardless of the failure of any components. This can be achieved through redundancy and elimination of single point of failure [14].

f. The Scalability of the Monitoring Mechanism - Key factors of cloud computing includes scalability of resource and application provisioning on-demand in a pay-as-you-go method. Cloud environments measure thousands of computing devices in large-scale environments; hence, the need for scalable monitoring tools. In [35], the authors proposed some methods to handle the challenges in setting up virtualized datacenter management systems that are scalable. The plan is to strategise a high performance and strong management tool that will be able to scale from some hosts to a large-scale Cloud datacenter. These authors indicated how they came about a scalability management tool by noting these challenges:

a. Performance / Fairness;
b. Security;
c. Robustness;
d. Availability; and
e. Backward compatibility.

3.6. Trending Practices in Cloud Resource Management

Daily growing attributes and advancements in cloud computing are promising the future and luring organizations and stakeholders towards the use of cloud computing. This fast growth of cloud computing is in need of stricter security and large quantity of resources like memory, process and storage to meet the Service Level Agreement standards. The trending practices include:

a. Live Migration – The use of live migration of virtual machines is an indispensable attribute of virtualization, enabling a seamlessly and transparent migrating of virtual machines from a source host to another destination host, even though the VM is still running without interrupting VMs [36]. Cloud live migration is on the increase in cloud computing with the aim to gain the cloud benefits from non-cloud situated applications. Live migration is a recent concept within cloud computing that is utilized during practical fault tolerance for instance by faultlessly moving Virtual Machines away from faulty hardware to stable hardware without the consumer noting any adjustment in a virtualized space. This is so because in the use of live migration, the end user is unaware there is only a delay of 60 mini seconds to 300 mini seconds [15]. Virtual Machine live migration can be performed via static migration or Dynamic migration. In order to fulfill

Service Level Agreement, static Virtual Machine migration is used as required resources gets carefully chosen as per requirement in SLA, but in dynamic Virtual Machine provisioning, the resources change dynamically to handle unforeseen workload changes [37]. Major benefit for live VM migration includes [36]:

a. Load balancing, the moving of over-loaded servers to light-loaded servers in order to release congested hosts.
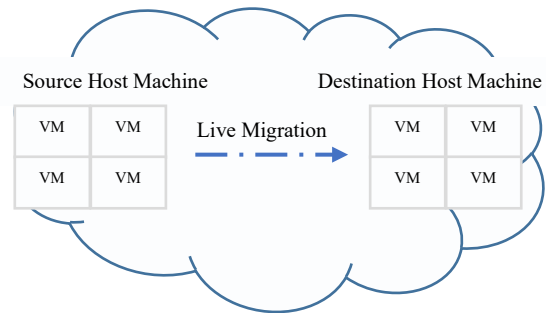
c. Proactive fault tolerance and online maintenance

Figure 1: Live Migration [27]

b. Power Saving in Green Computing - the authors in [38] discussed a proposed resource-utilization- aware energy saving server consolidation algorithm for the provision of improved utilization of resources as it reduces the total number of virtual machines during live migrations. Experimental results demonstrated its ability to reduce the consumption of energy and violations in Service Level Agreement in cloud data center. In [39], the authors describes green data center as a depository of a competent management of the system with a reduced amount of power consumed environment. The power consumption by an average data center can be so enormous that it can serve as a power source for approximately 25,000 private homes. In [40], Jena presented an optimization algorithm that arrange job with the aim of achieving an optimized energy usage and overall computation time using Clonal Selection Algorithm (CSA). Dynamic Voltage Frequency Scaling (DVFS) technique designed to reduce the usage of cloud resources by managing the power voltage as well as the regularity of the processor without ruining the performance. In [41], the authors presented an optimization framework of two modules that run concurrently: the Datacenter Energy Controller that reduces the use of datacenter energy without degrading the Quality-of-Service (QoS) and; the Green Energy Controller that brings about renewable sources. This Green Energy Controller computes the anticipated energy budget for the datacenter by the use of the lithium-ion battery as an added energy reserve or the grid of both banks in which the Hybrid Electric Systems can drain. For cloud providers to make their services become green computing, they have to spend in renewable energy sources to generate power from renewable sources of energy, like wind, solar, or hydroelectricity [42]. Merging

resources can advance utilization and make provision for increase in space, power and cooling capacity within the same facility environment. Energy preserving approaches includes [43]:

    a.   Nano Data Center is a distributed computing platform that makes provision for computing and storage services and adopts a managed peer-to-peer model to form a distributed data center infrastructure.

    b.   Reduction of Central Processing Unit Power Dissipation adapts free cooling so that power dissipation reduces.

    c.   Server Consolidation is the migration of server roles from various underutilized physical servers to VMs for the reduction of hardware and energy consumption.

    d.   Energy Saving Strategy of System Software is a dynamic energy usage, so the operating system can dynamically manipulate the system unit for obtaining the minimal power consumption without degrading already assigned the task or any job in performance.

    e.   Live Migration of Virtual Machine is about the movement of a running Virtual Machine from one host to another.
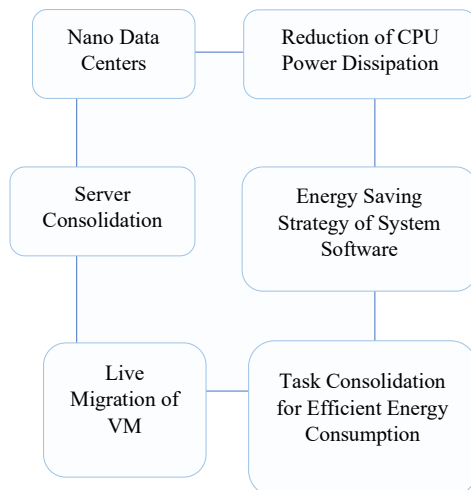

Figure 2: Energy Preserving Approaches [34]

3.7. Resource Grid Architecture Management

In [34], the authors proposed a Resource Grid Architecture for Multi Cloud Environment (RGA) for allocation and management of the resources with dynamism in virtual manner. Resource Management in Single Cloud Architecture is an ongoing area in research and RGA for Multi Cloud Environment is a current practice that has high scalability to manage resource at single cloud environments. This architecture presents a resource layer with logical resource grid and the usages of the VMs to project the resources in contrast to physical systems of a

cloud. Experimentations have showed that their Resource Grid Architecture can efficiently manage the resources against numerous clouds and support the green computing. Logically, RGA is grouped into four layers to manage resource efficiently:

    a.   Cloud Layer - This layer investigates the assigned resources to cloud application in order to keep log of record for resource usage and directs the information to cloud Resource Prediction Unit for hot spot detection and cold spot detection. Local Resource Manager is a component that verifies the resource allocation and utilization only at cloud level and is a proxy of Global Resource Manager.

    b.   Network Layer - This layer supports high-speed commutation among cloud and is the backbone for the experimented architecture for the reason that intra cloud communication is very much in demand in the implementation for resource sharing and process sharing.

    c.   Virtual Machine Layer - This layer is a separate group of virtual machine for sole cloud architecture to accomplish the operational feasibility. Each group of virtual machine is a set of virtual machines for physical system representation and are constantly accessible to resource grid layer monitoring and resource mapping.

    d.   Resource Grid Layer – This layer is an assembly of management modules of this architecture such as Global Resource Manager; these modules are interwoven for the monitoring, analyzing, allocating and tracking of resource information of multiple clouds at a centralized virtual location.

4. Conclusion

Cloud computing makes provision for scalable, on demand, elastic, multi-tenant and virtualization of services to customers on the Internet through cloud service providers. The types of service are the SaaS which provides applications with customizable interfaces, PaaS provides platform for application development and testing, and IaaS which makes provision of storages and computing infrastructure to users and extends to physical machines, virtual machines. These services provided are based on the service level agreements between the cloud service providers and the client. The service level agreement stipulates the terms of the services makes provision for mutually beneficial transaction between the parties. This paper presented a survey of recent developmental trends and issues in resource management in cloud computing. Cloud computing is aimed at providing the best resource to its users. Many researchers are focused on request allocation and resource management for Cloud Computing. All-encompassing use of learning empowers systems to be dynamic in handling workloads and execution environments. This paper documented techniques for creating scheduling and resource allocation choices for cloud computing and the datacenters. This paper concluded with a review of resource management on Cloud computing and cloud SLAs.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] I. Odun-ayo, A. Olasupo, and N. Omoregbe, "Cloud Service Level Agreements – Issues and Development," Int. Conf. Next Gener. Comput. Inf. Syst., pp. 1–6, 2017.

[2] V. C. Emeakaroha and R. N. Calheiros, "Achieving Flexible SLA and Resource Management in Clouds," Achiev. Fed. self-manageable cloud infrastructures theory Pract., pp. 266–287, 2012.

[3] N. M. Gonzalez, T. C. M. de B. Carvalho, and C. C. Miers, "Cloud resource management: towards efficient execution of large-scale scientific applications and workflows on complex infrastructures," J. Cloud Comput. Adv. Syst. Appl., vol. 6, no. 1, 2017.

[4] I. Odun-ayo, O. Ajayi, and A. Falade, "Cloud Computing and Quality of Service : Issues and Developments," Proc. Int. MultiConference Eng. Comput. Sci., vol. I, pp. 179–184, 2018.

[5] A. Koci and B. Cico, "ADLMCC-Asymmetric Distributed Lock Management In Cloud," Int. J. Inf. Technol. Secur., vol. 10, 2018.

[6] L. Wu, S. K. Garg, and R. Buyya, "Journal of Computer and System Sciences SLA-based admission control for a Software-as-a-Service provider in Cloud computing environments," J. Comput. Syst. Sci., vol. 78, no. 5, pp. 1280–1299, 2012.

[7] K. Buck, D. Hanf, and D. Harper, "Cloud SLA Considerations for the Government Consumer," Syst. Eng. Cloud Comput. Ser., pp. 1–65, 2015.

[8] R. El-awadi and M. Abu-rizka, "A Framework for Negotiating Service Level Agreement of Cloud-based Services," Int. Conf. Commun. Manag. Inf. Technol., no. 65, pp. 940–949, 2015.

[9] F. Faniyi, R. Bahsoon, and G. Theodoropoulos, "A dynamic data-driven simulation approach for preventing service level agreement violations in cloud federation," Int. Conf. Comput. Sci. ICCS, pp. 1167–1176, 2012.

[10] S. B. Dash, H. Saini, T. C. Panda, A. Mishra, and A. B. N. Access, "Service Level Agreement Assurance in Cloud Computing : A Trust Issue," Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 3, pp. 2899–2906, 2014.

[11] M. Cochran and P. D. Witman, "Governance and Service Level Agreement Issues in A Cloud Computing Environment," J. Inf. Technol. Manag., vol. XXII, no. 2, 2011.

[12] Olasupo O. Ayayi, F. A. Oladeji, and C. O. Uwadia, "Multi-Class Load Balancing Scheme for QoS and Energy Conservation in Cloud Computing," West African J. Ind. Acad. Res., vol. 17, pp. 28–36, 2016.

[13] Y. Jiang, H. Sun, J. Ding, and Y. Liu, "A Data Transmission Method for Resource Monitoring under Cloud Computing Environment," Int. J. Grid Distrib. Comput., vol. 8, no. 2, pp. 15–24, 2015.

[14] O. Ajayi and F. . Oladeji, "An overview of resource management challenges in cloud computing," UNILAG Annu. Res. Conf. Fair, vol. 2–3, pp. 554–560, 2015.

[15] A. J. Younge, G. Von Laszewski, L. Wang, S. Lopez-alarcon, and W. Carithers, "Efficient Resource Management for Cloud Computing Environments," Int. Conf. Green Comput. IEEE , 2010.

[16] N. Yadwadkar, "Machine Learning for Automatic Resource Management in the Datacenter and the Cloud," Ph.D. Diss. Dept. Electr. Eng. Comput. Sci. Electr. Eng. Comput. Sci. Univ. California, Berkeley, U.S. Accessed Jan. 6, 2019. Available http//digitalassets.lib.berkeley.edu/etd/ucb/text/Yadwadka, 2018.

[17] A. Ostfeld, "Data-Driven Modelling : Some Past Experiences and New Approaches approaches," J. Hydroinformatics, vol. 10, no. 1, 2008.

[18] A. Schram and K. M. Anderson, "MySQL to NoSQL Data Modeling Challenges in Supporting Scalability," Proc. 3rd Annu. Conf. Syst. Program. Appl. Softw. Humanit., pp. 191–202, 2012.

[19] C. Reiss, G. R. Ganger, R. H. Katz, and M. A. Kozuch, "Heterogeneity and Dynamicity of Clouds at Scale : Google Trace Analysis," Proc. Third ACM Symp. Cloud Comput., 2012.

[20] G. Blanchard, G. Lee, and C. Scott, "Generalizing from Several Related Classification Tasks to a New Unlabeled Sample," Proc. 24th Int. Conf. Neural Inf. Process. Syst., pp. 2178–2186, 2011.

[21] L. Badger, R. Patt-corner, J. Voas, L. Badger, R. Patt-corner, and J. Voas, "Cloud Computing Synopsis and Recommendations," Natl. Inst. Stand. Technol. Spec. Publ. 800-146, 2012.

[22] I. Odun-ayo, N. Omoregbe, and B. Udemezue, "Cloud and Mobile Computing – Issues and Developments," Proc. World Congr. Eng. Comput. Sci., vol. I, 2018.

[23] F. G. Fang, "Study on Resource Management Based on Cloud Computing," MATEC Web Conf., vol. 63, 2016.

[24] H. Singh, "Efficient Resource Management Technique for Performance Improvement in Cloud Computing," Indian J. Comput. Sci. Eng., vol. 8, no. 1, pp. 33–39, 2017.

[25] R. Buyya, S. K. Garg, and R. N. Calheiros, "SLA-Oriented Resource Provisioning for Cloud Computing : Challenges, Architecture, and Solutions," Int. Conf. Cloud Serv. Comput., 2011.

[26] E. Pacini, C. Mateos, and C. García Garino, "Distributed job scheduling based on Swarm Intelligence: A survey," Comput. Electr. Eng., vol. 40, no. 1, pp. 252–269, 2014.

[27] K. Ashok, K. Rajesh, and S. Anju, "Energy aware resource allocation for clouds using two level ant colony optimization," Comput. Informatics, vol. 37, pp. 76–108, 2018.

[28] B. Ghutke, "Pros and Cons of Load Balancing Algorithms for Cloud Computing," Int. Conf. Inf. Syst. Comput. Networks, 2014.

[29] Z. Zhang and Y. Li, "User utility oriented queuing model for resource allocation in cloud environment," J. Electr. Comput. Eng., vol. 2015, pp. 1–8, 2015.

[30] I. Chana and T. Kaur, "Delivering IT as A Utility- A Systematic Review," Int. J. Found. Comput. Sci. Technol. (IJFCST)., vol. 3, no. 3, 2013.

[31] S. Mawia, I. Omer, and A. B. A. Mustafa, "Comparative study between Cluster, Grid, Utility, Cloud and Autonomic computing," IOSR J. Electr. Electron. Eng., vol. 9, no. 6, pp. 61–67, 2014.

[32] A. Kim, J. Lee, and M. Kim, "Resource management model based on cloud computing environment," Int. J. Distrib. Sens. Networks, vol. 12, no. 11, 2016.

[33] S. Parikh, N. Patel, and H. B. Prajapati, "Resource Management in Cloud Computing : Classification and Taxonomy," Comput. Res. Repos., 2017.

[34] C. Aruna and R. Prasad, "Resource Grid Architecture for Multi Cloud Resource Management in Cloud Computing," Emerg. ICT Bridg. Futur. - Proc. 49th Annu. Conv. Comput. Soc. India, vol. 1, pp. 631–640, 2015.

[35] S. Vijayaraghavan and K. Govil, "Challenges in Building Scalable Virtualized Datacenter Management," ACM SIGOPS Oper. Syst. Rev., vol. 44, no. 4, pp. 95–102, 2012.

[36] M. Noshy, A. Ibrahim, and H. A. Ali, "Optimization of live virtual machine migration in cloud computing : A survey and future directions," J. Netw. Comput. Appl., vol. 110, no. March, pp. 1–11, 2018.

[37] S. B. Rathod and K. R. Vuyyuru, "Secure Live VM Migration in Cloud Computing : A Survey," Int. J. Comput. Appl., vol. 103, no. April 2016, 2014.

[38] B. Saha, "Green Computing : Current Research Trends," Int. J. Comput. Sci. Eng., vol. 6, no. June, pp. 3–6, 2018.

[39] S. Sharma and G. Sharma, "A Review on Secure and Energy Efficient Approaches for Green Computing," Int. J. Comput. Appl., vol. 136, no. May, 2016.

[40] R. K. Jena, "Energy Efficient Task Scheduling in Cloud Environment," 4th Int. Conf. Power Energy Syst. Eng., vol. 141, pp. 222–227, 2017.

[41] A. Pahlevan, R. Maurizio, G. D. V. Pablo, B. Davide, and A. David, "Joint computing and electric systems optimization for green datacenters Handbook of Hardware / Software Codesign Soonhoi Ha and Jürgen Teich," Jt. Comput. Electr. Syst. Optim. Green Datacenters. Ha S., Teich J. Handb. Hardware/Software Codesign. Springer, Dordr., no. November, 2016.

[42] M. K. Aery, "Green Computing : A Study on Future Computing and Energy Saving Technology," An Int. J. Eng. Sci., vol. 25, no. November 2017, 2018.

[43] A. D. Borah, D. Muchahary, and S. K. Singh, "Power Saving Strategies in Green Cloud Computing Systems," Int. J. Grid Distrib. Comput., vol. 8, no. February, pp. 299–306, 2015.