

## A Survey on SLA Management for Cloud Computing and Cloud-Hosted Big Data Analytic Applications

Radhya Sahal<sup>1</sup>, Mohamed H. Khafagy<sup>2</sup> and Fatma A. Omara<sup>3</sup>

<sup>1,3</sup> Department of Computer Science, Faculty of Computers & Information, Cairo University, Egypt

<sup>2</sup> Department of Computer Science, Faculty of Computers & Information, Fayoum University, Egypt

<sup>1</sup>radhya.sahal@grad.fci-cu.edu.eg, <sup>2</sup>mhk00@fayoum.edu.eg

<sup>3</sup>f.omara@fci-cu.edu.eg

### Abstract

Cloud computing is a new generation of computing based on the layered model which provides different services to cloud customer. Conceptually, cloud computing offers a scalable platform for Big Data Analytic Applications (BDAAs) which can elastically provision resources based on data growth and complexity of analytic applications. This complexity of analysis increases with vast volumes of data in big enterprises which prefer effective and fast decision-making. Therefore, the importance of Service Level Agreement (SLA) is appeared which clarifies the roles between a customer (i.e., cloud user or Big Data analyst) and a provider for particular service provision comes. The main components of the SLAs are Quality of Service parameters which must be monitored to achieve a set of Service Level Objectives (SLOs) and detect violations. Indeed, many SLA management approaches have been developed as solutions for preventing SLA violations to avoid costly penalties. Consequently, many interesting solutions developed works of SLA violation management in cloud technology and cloud-hosted BDAAs. A survey about the existed works in terms of idea, strengths and weaknesses is introduced in this paper. Meanwhile, the challenges and new research directions in this area which require further investigation will be discussed to provide a comprehensive overview and big-picture.

**Keywords:** Cloud Computing, Big Data Analytic Applications, Service Level Agreement, Quality of Service QoS, Violation, Penalty

### 1. Introduction

Cloud computing is a new generation of computing. It allows users to use computational resources and services of data centers (i.e., machines, network, storage, operating systems, application development environments, application programs) over the Internet to deploy and develop their applications [1]. The main feature of cloud computing is providing self-service provisioning, which allows the users to deploy their own sets of computing resources. Generally, cloud computing services are divided into three parts; SaaS (Software as a Services), PaaS (Platform as a Services), and IaaS (Infrastructure as a Services) [2].

On the other hand, Big Data has been grown rapidly in many domains such as social networks and other information systems. The need for distributed computing is become an important issue due to the increasing of workstations power and the data sets sizes. The development and implementation of distributed system for Big Data applications are considered a challenge [3, 4]. One of the famous frameworks that have emerged for Big Data processing is MapReduce. It was first introduced by Google in 2004 [5]. The main concept of MapReduce is to abstract the details of a large cluster of machines to facilitate the computation on large datasets.

Fundamentally, a Service Level Agreement (SLA) represents an agreement between a customer and a provider (s) to receive a particular service provision. SLAs contain Quality of Service (QoS) parameters that must be maintained by a provider (*e.g.* response time, bandwidth, storage, reliability, deadline, throughput, delay, and cost) [6]. Generally, these QoS parameters are defined as a set of Service Level Objectives (SLOs) [7, 8]. These parameters need to be measurable and must be monitored during the provision of the service to detect and solve violations that have been agreed in the SLA. The SLA must also contain a set of penalty clauses when service providers fail to deliver the pre-agreed SLOs. Therefore, the SLA management plays a role in scaling and automating SLA to adapt to dynamic environmental changes and heterogeneous resources [9].

This survey paper reviews SLA management frameworks from two points of view; cloud computing environment among the Software/Platform/Infrastructure (SPI) layered model and cloud-hosted Big Data Analytic Applications (BDAAs). Essentially, the term SLA and QoS are not new and early emerged before cloud computing and Big Data era. The importance of these two concepts comes from the sensitivity of time for BDAAs such as social network and recommendation systems [10-11]. The cloud-hosted applications need strong QoS guarantees to meet SLA real time response-constrained and deadline-constrained.

According to this survey paper, the identifying of SLA violation in cloud computing is focused. Then, the SLA-based management of BDAAs in cloud computing environments is discussed. Significantly, the SLA-based provisioning and resource allocation with respect to price, latency, and resource availability time which adopted for all layers of cloud computing or Big Data computing hosted in clouds are argued. Moreover, the approaches which proposed to manage SLA violation according to the majority of researches' contribution theme in this area have been classified.

The rest of this paper is organized as follow; the foundational concepts in SLA are summarized in Section 2. The SLA violation management approaches for cloud computing and cloud-hosted BDAAs are discussed in Section 3. The survey paper is concluded in Section 4. Finally, a brief outlook and recommendations for the future work are given in Section 5.

## **2. The Service Level Agreement Foundations**

### **2.1 SLA Definition**

Wikipedia defines SLA as: "The SLA records a common understanding about services, priorities, responsibilities, guarantees and warranties". Each area of service scope defined the level of service according to its point of view (*e.g.*, availability, security, performance, operation, billing, IT resources, recovery, *etc.*).

Generally, SLA is a crafted contract between the customer and provider(s). It contains a list of completely defined services which the provider(s) will deliver them to the customer. Also, it contains the responsibilities of the provider and the customer and remedies available to both if the terms of the SLA are not met. Furthermore, the contract describes how the SLA will change over time. Indeed, the SLAs may be between two parties, for instance, a single client and a single provider, or between multiple parties, for example, a single client and multiple providers [7-8].

On the other hands, SLAs illustrate the provider and customer understandings of services which be delivered. Also, SLAs remove all doubts about what the provider's services will be delivered and what the customer expectations of these services. Thus, the provider and customer must define the SLA criteria. For example, a list of important SLA criteria are described in Table 1 [12].

**Table 1. SLA Criteria [12]**

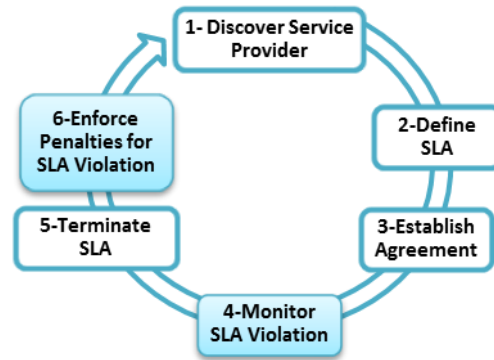
SLA Criteria	Example
Availability	99.99% during work days, 99.9% for nights/weekends
Performance	minimum response times
Security / privacy of the data	encrypting all stored and transmitted data
Disaster Recovery expectations	worse case recovery commitment
Location of the data	consistent with local legislation
Access to the data	data retrievable from a provider in a readable format
Portability of the data	Data movment among different providers
Identifying process and resolution expectations problems	call center
Change Management process	changes – updates or new services
Dispute mediation process	escalation process, consequences

## 2.2 SLA Life Cycle

Generally, the SLA can be broadly defined through various phases. These phases include; discovering providers, defining the SLA, agreeing to the terms of the SLA, monitoring SLA violations, terminating an SLA, and penalties for SLA violation. More clearly explanation of SLA phases is illustrated as follows [8-9]:

- 1- Discovering service provider phase involves choosing possible partner(s) to interact with from known registry list of providers which is located according to customer's needs.
- 2- Define SLA phase which includes a definition of services, parties, penalty policies and QoS parameters. In this phase, it is possible to negotiate between parties to reach a mutual agreement.
- 3- Establish agreement phase, in which an SLA template is established and filled in by specific agreement, and parties are starting to commit to the agreement.
- 4- Monitor SLA violation phase, in which the provider's delivery performance is measured against to the contract.
- 5- Terminate SLA phase, in which SLA terminates due to timeout or any party's violation.
- 6- Enforce penalties for SLA violation, if there is any party violating contract terms, the corresponding penalty clauses are invoked and executed.

These phases are illustrated in Figure 1. The shadowed phases are considered in this survey. So, in the next subsections, SLA violation monitor and SLA penalties will be explained in details.



**Figure 1. SLA Life Cycle Six Phases, as Defined by Sun Microsystems Internet Data Center Group (2002) [8-9]**

### 2.3 SLA Violation Monitor

Monitoring plays an important role in determining whether an SLA has been violated or not. It helps the SLA parties to determine which penalty clause should be invoked. Most violations likely happen during load fluctuations. The SLA violation monitoring begins once an agreement has been established. More specifically, monitoring plays a critical role in determining whether SLA is achieved or violated. There are three main concerns. Firstly, which party should pay in case of violation occurs. Secondly, how fairness can be assured between parties (if there are more than one part of the contract). Thirdly, how the boundaries of SLA violation are defined. There are three types of SLA violation according to SLA monitoring in Table 2 [7-9].

**Table 2. SLA Violations Types**

Violation Type	Description	Solution
All-or-nothing Provisioning	- Provisioning all of SLO constraints must be satisfied for a successful delivery of a service. -occur when complete services failure.	- New SLA needs to be negotiated, possibly with another service provider, which requires additional effort on both the client and the service provider
Partial Provisioning	- Provisioning of a service meets some of the SLOs. -Some of the SLO constraints must be satisfied for a successful delivery of a service.	- Identify the strict SLOs from the first negotiation. - Clients and service providers can then react differently depend on the importance of the violated SLO.
Weighted Partial Provisioning	-Provision of a service meets SLOs that have a weighting greater than a threshold (identified by the client).	- Instead of terminating the SLA altogether it might be possible to renegotiate, <i>i.e.</i> , with the same service provider, the part of the SLA that is violated.

### 2.4 SLA Penalties

These penalty terms may also be negotiated between a customer and a provider or a fixed set of penalty terms may be used. The SLA must also contain a set of penalty clauses specifying what happens when service providers fail to deliver the pre-agreed

quality [8]. One challenge of the cloud is that cloud providers do not offer on-demand service adaptability in SLA but they may pay a penalty if there is any violation. For example, Amazon EC2 only provides the minimum terms for service guarantees with an “Annual Uptime Percentage of at least 99.95% during the Service Year” and a penalty model based on service credits [13].

Actually, there are different types of penalties for SLA violation; a decrease in the agreed payment for using the service (*i.e.*, during the negotiation phase the parties may choose an ‘agreed payment for non-performance’) and a decrease of the reputation of the provider (*i.e.*, change or decreases the provider service rank) [8].

### 3. SLA Violation Management Approaches

In this survey paper, the SLA violation management approaches broadly classified into two parts; SLA on cloud computing and SLA for cloud-hosted BDAAAs. Each part can be sub-classified into groups according to the research contribution themes which discussed in the following subsections (See Figure 2).

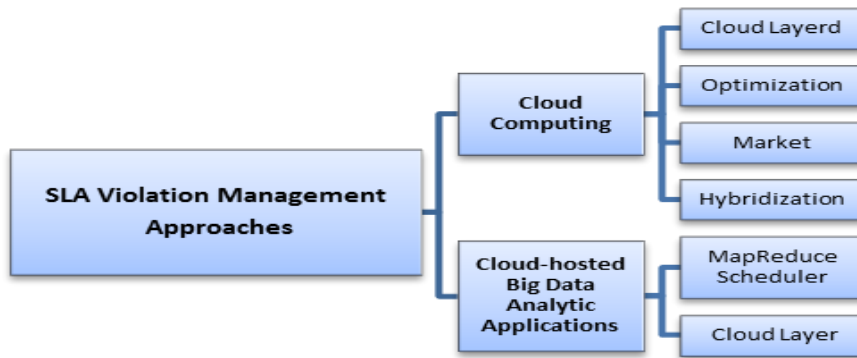


Figure 2. SLA Violation Management Approaches Classification

#### 3.1 Cloud Computing

Recent works consider various strategies to develop the models, mechanisms and frameworks for SLA violation. Some of the researches highlighted the SLA violation approaches according to cloud layer-based (*i.e.*, SaaS, PaaS, and IaaS) [14-18]. Furthermore, some of the researches proposed solutions for SLA violation which able to cope with all cloud layers. However, some of the researches consider the problem as an optimization problem in terms of minimizing violations (*i.e.*, increase customer satisfaction) and maximize provider profit [19-21]. Other researchers considered SLA management as a market-based issue [22-23].

Therefore, SLA violation management in cloud computing has been discussed in research papers using different views. In this survey paper, the existed research papers are organized as follows; 1) Cloud Layered-based approaches, which focus on violation in one layer only, 2) Optimization-based approaches which present the violation problem as optimization (*i.e.*, considering QoS parameters), 3) Market-based approaches which based on cloud services as a subject to the business, and iv) Hybridization-based approaches which apply more than one technique to design end-to-end solution for SLA violation. Figure 2 depicts the classification of SLA violation management approaches.

### 3.1.1 Cloud Layered-Based

In order to depict the whole pictures, some researches who proposed the SLA violation approach to one layer will be mentioned. In [14, 18], the approaches are proposed for SaaS to detect and tackle the SLA violation for provisioned services in SaaS layer. Nevertheless, some of the researches are focused on certain application type in SaaS. For example, Sakr and Liu [15] have proposed SLA-based management model for cloud-hosted database systems from the consumer perspective. However, related to SLA enforcement and provisioning in PaaS layer, Michael *et al.* [13] have presented a novel PaaS architecture being developed targeting real-time (QoS) guarantees for online interactive multimedia applications. While some researches in [16-17] are focused on cloud IaaS layer as a critical essential asset for SLA violation in terms of security, privacy, quality of hardware, availability, *etc.*

The SLA violation mechanism for SaaS layer will be addressed. Hence, the SaaS providers have to use their own hardware or rent it from infrastructure providers, the cost of hardware will increase. So, the SaaS providers need to minimize the cost with respect to SLA satisfaction to avoid violations penalties. Buyya *et al.* [14] have proposed some algorithms of resource allocation for SaaS providers to balance the cost of hardware and SLA violations. Their proposed algorithm considers certain QoS parameters (including request type, product type, account type, contract length, and a number of accounts). In order to satisfy the customers (*i.e.*, minimizing the SLA violation) and minimize the cost of hardware resources. In fact, the proposed algorithm reused the already created VMs in order to minimize the cost of hosted hardware. But, this solution may cause security problems for customers. Due the residual information in the VMs, this can ease the attack attempts to gain unauthorized access.

### 3.1.2 Optimization-Based

Generally, resource utilization makes the management and monitoring tasks be more beneficial for a cloud provider. As a result, a cloud provider continuously needs optimization algorithms to provide optimal resource allocation/reallocation solutions in order to maximize utilization and profit [24-25]. Thus, the scheduling strategies, the multiple SLA parameters and an efficient allocation of resources are being necessary to be considered.

In [19], a scheduling heuristic (*i.e.*, optimization-based technique) has been proposed which considers multiple SLA parameters for application deployments in the cloud. The considered attributes (*i.e.*, multiple SLA parameters) include CPU time, network bandwidth and storage capacity for deploying applications. These parameters have limited application in real world systems as they need to be considered only during deployment. Once the applications have been ready to use, these parameters will be replaced. The customers will be concerned with performance parameters such as response time, processing time, throughput, delay, latency, *etc.* Moreover, this heuristic solution suffers from the problem of being trapped in local optimums and may not have much practical significance in real world business environments.

On the other hand, there are some studies have discussed the different optimization-based approaches of resources allocation in the cloud environment. These approaches deal with service request scheduling considering multiple SLA objectives and high resource utilization. Some studies such as [20] use Ant Colony optimization-based technique to utilize the SLAs to manage the provider resources. While, a resource scheduling scheme for cloud computing environment, that provides QoS guarantees, is presented in [21]. This scheme is based on improved Electromagnetism-like Mechanism algorithm (EM).

### 3.1.3 Market-Based

Cloud providers face a challenge of resource allocation to meet the users' QoS constraints according to the pre-agreed SLAs to serve million users' workloads requirements. Therefore competition between the cloud providers becomes a big issue to maximize their market and their total profit by maximizing their resource utilization. Thus, market-based resource management becomes a challenge for cloud provider's reputation which affected by SLA violation.

In [22], a framework is described for SLA management with customer-driven service to manage QoS requirements. The proposed framework successfully integrates the market-based resource provisioning with virtualization technologies for flexible resource allocations to the user applications which used Aneka platform. Aneka acts as a market-based cloud platform for building and deploying distributed applications on clouds. It monitors the resources in terms of pricing, accounting, and QoS/SLA services in private and/or public cloud environments. Actually, the most important note of the proposed framework is that it does not support different cloud services such as IaaS, PaaS and SaaS at the same time.

### 3.1.4 Hybridization-Based

According to the mentioned researches which are targeted only one approach, some authors combined more than one technique to design end-to-end solution for SLA violation. They used the integration of theoretical application workload prediction, service performance models, machine learning techniques and optimization techniques to propose an end-to-end monitoring and automated SLA management over cloud environments.

Alhamazani *et al.*, [26] have highlighted the importance of dynamically monitoring the QoS of virtualized services. Their focus was to apply dynamically monitored the application components and cloud services for automatic QoS control under uncertainties. The monitoring of the services would help both the cloud provider and application developer to maximize the return on their investments. Also, the monitoring techniques and services can help a cloud provider or application developer in regards to keeping the cloud services and hosted applications to be operated at peak efficiency, detect changes in service and application performance, SLA violations, failures of cloud services and other dynamic configuration changes. The researchers mainly concentrate on SNMP based QoS monitoring.

## 3.2 Cloud-Hosted Big Data Analytics Applications

As Big Data represents large sets of complex data that need special data processing applications, the importance of cloud-hosted BDAA comes. Meanwhile, big enterprises adopt cloud-hosted BDAA for investigating the vast of created information from different sources. Generally, there are rules which maintained for cloud-hosted Big Data such as capacity, scalability, security, privacy, reliability, analytics and operational [27]. More specifically, there are some unique QoS metrics that fit the requirements of cloud-hosted BDAA such that throughput and latency in distributed messaging system, response time in batch processing platform for queries in a transactional system. In this survey paper, the issues related to data storage and data growth in data analytics context are concentrated. The main three challenges which are posed for cloud-hosted BDAA namely; 1) type of cloud resources to be dynamically provisioned for incoming workloads, 2) selection of private infrastructure or public cloud provider to serve a given request with respect to budget and deadline constraints, 3) geographically distributed, which direct effect data transfer ,as well as, processing costs.

Consequently, the existed researches are organized as follows; 1) MapReduce scheduler-based approaches which present the SLA violation problem in MapReduce

scheduler, 2) Cloud Layered-based approaches which focus on violation according to cloud layers.

### 3.2.1 MapReduce Scheduler-Based

Primary, SLA-based Big Data computing and MapReduce scheduling have been studied in the cloud context. These prior works toward SLA-based approaches for cloud-hosted BDAAAs do not provision cloud resources dynamically, where the resources already pre-provisioned statically on a private cloud (*i.e.*, a virtual cluster).

Firstly, several solutions have been proposed for achieving SLAs of MapReduce jobs to meet deadline-constrained [28-31]. Verma *et al.*, [28] have proposed a framework called ARIA, to address SLA violation with respect to the performance of MapReduce jobs. This proposed framework is based on profiling the routine critical performance characteristics (*i.e.*, soft deadline) and estimating the amount of resources required for job completion within the deadline. Their SLO-scheduler in Hadoop determines job ordering and the amount of resources to be allocated to meet the job deadlines.

Regarding cloud computing environment, Lema *et al.*, [29] have proposed AROMA system which auto reconfigures Hadoop parameters and provisions cloud resources to meet deadline-constrained for MapReduce ad-hoc jobs. It includes two-phase optimization framework; an offline phase which applies machine learning techniques to cluster Hadoop jobs to feed the online operations and an online phase uses an optimization technique to allocate resources and configure Hadoop. However, Hwang *et al.*, [30] have proposed a MapReduce scheduling algorithm with SLA deadline-constrained and pricing policies on the cloud where the MapReduce scheduler left the resources (VM) selection to the user.

On the other hand, Wang and Shi [31] have proposed MapReduce task-level scheduling algorithms for iterative applications with two SLA practical constraints; budget and deadline in the heterogeneous cloud. The goal of the proposed greedy optimization algorithm was to select a proper machine from a fixed pool of heterogeneous machines for each task stage in MapReduce job to minimize the job completion, as well as, the cost.

Most of the above research works are tailored for MapReduce task scheduling in cloud infrastructure with regardless the dynamic resource provisioning consideration.

### 3.2.2 Cloud Layered-Based

Fundamentally, SLA can be defined according to a well-understood cloud layer stack as SaaS-level (*e.g.*, event detection delay minimization and availability maximization), PaaS-level (*e.g.*, stream processing latency minimization and batch processing latency minimization), and IaaS-level (*e.g.*, CPU utilization, network throughput and latency minimization) [32]. According to these types of SLAs-based cloud layers, the research works are discussed in this section.

Nita *et al.*, [33] have addressed the tricky of Big Data transfer towards the needed cloud resources by virtual machine migration inside cloud infrastructure. The transportation of Big Data as a multi-criteria optimization problem to find the optimized cost of cloud resource replacement is formalized. A scheduling policy included two greedy algorithms is proposed for data transfer. It manages an SLA-based network-constrained which affect cloud performance. The proposed scheduling policy is validated by simulated experiments which considered as SLA for IaaS-level.

On the other hand, Alrokayan *et al.*, [34] have proposed a SaaS-level SLA, cost-aware resource provisioning and task scheduling approach tailored for Big Data applications in the cloud. They concentrated on the dynamic resource provisioning for Big Data applications in cloud computing context by model the user request, which consist of analytics jobs with budget and deadline running on Hadoop. Although, they incurred



SLA-based for dynamic resource provisioning of cloud-hosted Big Data applications, they only considered SLA-based budget-constrained and deadline-constrained.

Worthwhile, the author in [10] has proposed a cross-layer SLA management for cloud-hosted BDAA's. The multi-layered SLAs framework developed cloud simulation model (*i.e.*, CloudSim) and benchmarked to study the impact of SLA constraints for cloud-hosted BDAA's. More specifically, they extended CloudSim framework to support SLA-oriented evaluation. The extensible environment can simulate the multiple Big Data processing (*i.e.*, MapReduce) and hardware resources at the same time.

Finally, the summary of the strengths and the weaknesses of selected proposed solutions for SLA violations in cloud computing and cloud-hosted BDAA's are described in Table 3.

**Table 3. Summary of Strengths and Weaknesses of Selected Proposed Solutions for SLA Violations in Cloud Computing and Cloud-hosted Big Data Analytic Applications**

Ref	Work	Potential contributions	Management Approach Type	Improvement/Strengths	Limitations/Weakness
[14]	<b>Buyya.</b> SLA-Based Resource Allocation for SaaS Provider	Proposed two algorithms for resource allocation in SaaS providers to balance the cost of hardware and SLA violations.	Cloud Computing / Layer-based <i>e.g.</i> , SaaS	reduce the cost of SaaS providers with keeping the QoS of customers stratification	Cause security problems for customers. The residual information in the VMs can ease the attack attempts.
[19]	<b>Emeakaroha</b> SLA-Aware Application Deployment Using Scheduling Heuristic	A scheduling heuristic that takes multiple SLA parameters when deploying applications in the Cloud Considers deployment attributes such as CPU time, network bandwidth, storage capacity <i>etc.</i> , before installation of applications in the cloud system.	Cloud Computing / Optimization-based <i>e.g.</i> , heuristic	Considers deployment attributes such as CPU time, network bandwidth, storage capacity <i>etc.</i> , before installation of applications in the cloud system.	- Does not consider performance parameters such as response time, performance time <i>etc.</i> - Heuristic solution trap in local optimums so may not have much practical significance in real world business environments.
[22]	<b>Calheiros</b> SLA-Oriented Resource	The proposed a framework is described for SLA management	Cloud Computing / Market-based	successfully integrates the market-based resource	Suffers from the integration services

	Provisioning for Cloud Computing	with customer-driven service to managing QoS requirements.		provisioning with virtualization technologies for flexible resource allocations to user applications.	problem of IaaS, PaaS and SaaS in the same time.
[30]	<b>Hwang</b> Minimizing Cost of Virtual Machines for Deadline-Constrained MapReduce Applications in the Cloud	The proposed a MapReduce scheduling algorithm with SLA deadline-constrained on cloud where the MapReduce scheduler left the resources (VM) selection to the user	Cloud-hosted BDAAAs / MapReduce Scheduler-based	Cloud MapReduce scheduler with deadline-constrained.	Selection of resource left to the user which incur burden on him.

#### 4. Conclusions

In this paper, an overview and state-of-the-art of SLAs in the cloud computing and Big Data (*i.e.*, representative applications such as BDAAAs) are reviewed. Also, SLA management has been investigated because it plays a critical role in determining whether SLOs are achieved or violated. Moreover, it surveyed the latest advances approaches (*i.e.*, models, mechanisms, frameworks, algorithms) which have been proposed to handle SLA violation which affects the performance, profit, and cloud provider's reputation. These approaches are classified according to the majority of contribution theme. According to SLA in the cloud environment, the hybridization-based approach, through the combination of approaches which, is considered better than other approaches to tackle SLA violation. . On the another side, some researches have proposed efficient scheduling techniques for cloud resources hosted BDAAAs which satisfy SLA constrains such as budget and deadline. The conceptual cross-layer SLA management can offer significant advantages for eliminating SLA violation for cloud-hosted BDAAAs.

#### 5. Future Work

Although SLA violation is a challenging task, several researchers have undertaken to develop mechanisms, which could guarantee the QoS requirements of different services. More specifically, different proposed approaches for SLA violation in cloud computing are addressed in this survey paper. In spite of the resource allocation management is used to choose the proper resources for cloud customer requests and provider profit and cloud-hosted BDAAAs, it has not considered enough carefully. According to this survey, two aspects should be considered in the future; 1) Deep investigation of SLA-based elastic resource allocation for cloud (on the provider side) according to demand of services (on the consumer side), in order to improve the system efficiency, minimize violation of SLAs, and improve the profitability of service providers. 2) Exploring the SLA management for resource allocation strategies to optimize response time of the multi-query applications running on the cloud computing.

## References

- [1] L. C. Q. Zhang and R. Boutaba, "Cloud computing: state-of-the-art and research challenges", *Journal of Internet Services and Applications*, vol. 1, no. 1, (2010), pp. 7-18.
- [2] C. Computing, "Handbook of Cloud Computing", Available: <http://www.springerlink.com/index/10.1007/978-1-4419-6524-0>, (2010).
- [3] R. Akerkar, "Big data computing", CRC Press, (2013).
- [4] A. G. Divanis and A. Labbi, "Large-Scale Data Analytics", Springer, (2014).
- [5] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters", *Communications of the ACM*, vol. 51, (2008), pp. 107-113.
- [6] M. Firdhous, S. Hassan and O. Ghazali, "Monitoring, Tracking and Quantification of Quality of Service in Cloud Computing", *International Journal of Scientific & Engineering Research*, vol. 4, (2013).
- [7] S. A. Baset, "Cloud SLAs: present and future", *ACM SIGOPS Operating Systems Review*, vol. 46, (2012), pp. 57-66.
- [8] O. F. Rana, M. Warnier, T. B. Quillinan, F. Brazier and D. Cojocarasu, "Managing violations in service level agreements", in *Grid Middleware and Services*, ed: Springer, (2008), pp. 349-358.
- [9] L. Wu and R. Buyya, "Service Level Agreement (SLA) in utility computing systems", *Performance and Dependability in Service Computing: Concepts, Techniques and Research Directions*, (2011), pp. 1-25.
- [10] Z. Xuezhai, R. Ranjan, P. Strazdins, S. K. Garg and W. Lizhe, "Cross-Layer SLA Management for Cloud-hosted Big Data Analytics Applications", in *Cluster, Cloud and Grid Computing (CCGrid)*, 2015 15th IEEE/ACM International Symposium on, (2015), pp. 765-768.
- [11] R. Sahal and S. a. A. ElKorany, "An Adaptive Framework for Enhancing Recommendation Using Hybrid Techniques", *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 6, (2014).
- [12] (18/12/2013 8:00 AM). *Service Level Agreements in the Cloud: Who cares?* Available: <http://www.wired.com/insights/2011/12/service-level-agreements-in-the-cloud-who-cares/>, (2012).
- [13] M. Boniface, B. Nasser, J. Papay, S. C. Phillips, A. Servin, X. Yang, Z. Zlatev, S. V. Gogouvitis, G. Katsaros and K. Konstanteli, "Platform-as-a-service architecture for real-time quality of service management in clouds", in *Internet and Web Applications and Services (ICIW)*, 2010 Fifth International Conference on, (2010), pp. 155-160.
- [14] L. Wu, S. K. Garg and R. Buyya, "SLA-Based Resource Allocation for Software as a Service Provider (SaaS) in Cloud Computing Environments", in *Proceedings of the 2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, (2011), pp. 195-204.
- [15] S. Sakr and A. Liu, "Sla-based and consumer-centric dynamic provisioning for cloud databases", in *Cloud Computing (CLOUD)*, 2012 IEEE 5th International Conference on, (2012), pp. 360-367.
- [16] V. C. Emeakaroha, M. A. Netto, R. N. Calheiros, I. Brandic, R. Buyya and C. A. De Rose, "Towards autonomic detection of sla violations in cloud infrastructures", *Future Generation Computer Systems*, vol. 28, (2012), pp. 1017-1029.
- [17] V. C. Emeakaroha, R. N. Calheiros, M. A. Netto, I. Brandic and C. A. De Rose, "DeSVi: An architecture for detecting SLA violations in cloud computing infrastructures," in *Proceedings of the 2nd International ICST Conference on Cloud Computing (CloudComp'10)*, (2010).
- [18] V. C. Emeakaroha, T. C. Ferreto, M. Netto, I. Brandic and C. A. De Rose, "Casvid: Application level monitoring for sla violation detection in clouds", in *Computer Software and Applications Conference (COMPSAC)*, 2012 IEEE 36th Annual, (2012), pp. 499-508.
- [19] V. C. Emeakaroha, I. Brandic, M. Maurer and I. Breskovic, "SLA-Aware application deployment and resource allocation in clouds", in *Computer Software and Applications Conference Workshops (COMPSACW)*, 2011 IEEE 35th Annual, (2011), pp. 298-303.
- [20] H. Liu, D. Xu and H. Miao, "Ant Colony Optimization Based Service flow Scheduling with Various QoS Requirements in Cloud Computing", in *Software and Network Engineering (SSNE)*, 2011 First ACIS International Symposium on, (2011), pp. 53-58.
- [21] S. Yun, "Guaranteed QoS Resource Scheduling Scheme Based on Improved Electromagnetism-Like Mechanism Algorithm in Cloud Environment", in *Intelligent Networking and Collaborative Systems (INCoS)*, 2013 5th International Conference on, (2013), pp. 353-357.
- [22] R. Buyya, S. K. Garg and R. N. Calheiros, "SLA-oriented resource provisioning for cloud computing: Challenges, architecture, and solutions", in *Cloud and Service Computing (CSC)*, 2011 International Conference on, (2011), pp. 1-10.
- [23] C. Redl, I. Breskovic, I. Brandic and S. Dustdar, "Automatic sla matching and provider selection in grid and cloud computing markets", in *Proceedings of the 2012 ACM/IEEE 13th International Conference on Grid Computing*, (2012), pp. 85-94.
- [24] S. M. K. R. Sahal and F. A. Omara, "GPSO: An improved search algorithm for resource allocation in cloud databases", in *Computer Systems and Applications (AICCSA)*, 2013 ACS International Conference on, (2013), pp. 1-8.
- [25] F. A. Omara, S. M. Khattab and R. Sahal, "Optimum Resource Allocation of Database in Cloud Computing", *Egyptian Informatics Journal*, vol. 15, (2014), pp. 1-12.

- [26] K. Alhamazani, R. Ranjan, F. Rabhi, L. Wang and K. Mitra, "Cloud monitoring for optimizing the QoS of hosted applications", in Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on, **(2012)**, pp. 765-770.
- [27] A. B. Alnafoosi and T. Steinbach, "An integrated framework for evaluating big-data storage solutions - IDA case study", in Science and Information Conference (SAI), **(2013)**, pp. 947-956.
- [28] A. Verma, L. Cherkasova and R. H. Campbell, "ARIA: automatic resource inference and allocation for mapreduce environments", in Proceedings of the 8th ACM international conference on Autonomic computing, **(2011)**, pp. 235-244.
- [29] P. Lama and X. Zhou, "Aroma: Automated resource allocation and configuration of mapreduce environment in the cloud", in Proceedings of the 9th international conference on Autonomic computing, **(2012)**, pp. 63-72.
- [30] H. Eunji and K. Kyong Hoon, "Minimizing Cost of Virtual Machines for Deadline-Constrained MapReduce Applications in the Cloud", in Grid Computing (GRID), 2012 ACM/IEEE 13th International Conference on, **(2012)**, pp. 130-138.
- [31] Y. Wang and W. Shi, "On Optimal Budget-Driven Scheduling Algorithms for MapReduce Jobs in the Heterogeneous Cloud", Technical Report TR-13-02, Carleton University, **(2013)**.
- [32] H. Mohanty and S. Vaddi, "Big Data Service Agreement", in Big Data, ed: Springer, **(2015)**, pp. 137-160.
- [33] M.-C. Nita, C. Chilipirea, C. Dobre and F. Pop, "A SLA-based method for big-data transfers with multi-criteria optimization constraints for IaaS", in Roedunet International Conference (RoEduNet), 2013 11th, **(2013)**, pp. 1-6.
- [34] M. Alrokayan, A. Vahid Dastjerdi and R. Buyya, "SLA-aware Provisioning and Scheduling of Cloud Resources for Big Data Analytics", in Cloud Computing in Emerging Markets (CEEM), 2014 IEEE International Conference on, **(2014)**, pp. 1-8.