Monitoring of Cloud Computing Environments: Concepts, Solutions, Trends, and Future Directions

Guilherme Da Cunha Rodrigues\textsuperscript{1}, Rodrigo N. Calheiros\textsuperscript{2}, Vinicius Tavares Guimaraes\textsuperscript{1}, Glederson Lessa dos Santos\textsuperscript{1}, Márcio Barbosa de Carvalho\textsuperscript{1}, Lisandro Zambenedetti Granville\textsuperscript{1}, Liane Margarida Rockenbach Tarouco\textsuperscript{1}, Rajkumar Buyya\textsuperscript{2}

\textsuperscript{1}Federal University of Rio Grande do Sul, 9500 Bento Gonçalves, Porto Alegre, Brazil
\textsuperscript{2}Cloud Computing and Distributed Systems (CLOUDS) Laboratory, The University of Melbourne, 3010 Parkville, Melbourne, Australia

gcrodrigues@inf.ufrgs.br, rncalheiros@ieee.org, vtguimaraes@inf.ufrgs.br, glsantos@inf.ufrgs.br, mbcarvalho@inf.ufrgs.br, granville@inf.ufrgs.br, liane@inf.ufrgs.br, rbuyya@unimelb.edu.au

ABSTRACT

Cloud computing is a technology that companies, universities, and research centers use to acquire computational resources on demand to improve availability and scalability of applications while reducing operational costs. In this context, resource management is an important mean to improve clouds, and resources monitoring is the key to achieve it. This paper presents an overview on cloud monitoring and a comparison among relevant cloud monitoring solutions. In complement, we analyze trends on monitoring of cloud computing environments and propose future directions.

CCS Concepts

\item Networks \rightarrow Cloud computing; Network monitoring;

Keywords

Cloud Computing; Cloud Management; Cloud Monitoring

1. INTRODUCTION

Cloud computing became a recent groundbreaking paradigm because it efficiently reduces costs of ICT (information and communication technologies) infrastructures by offering computer resources as services [39]. Companies, universities, and governments, for example, can enjoy high availability and scalability by leasing cloud services offered by providers, with reduced prices if compared to traditional ICT models.

Typical cloud computing scenarios are composed of infrastructure providers (InPs), service providers (SPs), and customers [38]. InPs offer computing resources (e.g., processing, storage, networking) that can be leased by SPs. SPs, in turn, take into account the needs of customers and offer service applications to address these needs [22] [39]. Finally, customers inform what kind of services they need and their expectations on the quality of service, which are usually expressed through Service Level Agreements (SLAs).

To properly manage the complex scenarios resulted from the adoption of the cloud computing paradigm, some crucial tasks take place; cloud monitoring is one of them. Key to this paper. Cloud monitoring offers to both InPs and SPs means to observe the granted/allocated, virtual/physical resources. Through monitoring, SPs can show cloud information to customers. Information retrieved from cloud monitoring is also basis for decisions towards, for example, reducing energy consumption, enhancing systems’ dependability, or tuning cloud response time [41]. The literature has addressed cloud monitoring in several aspects such as requirements/properties, and tools/solutions.

In terms of requirements/properties, cloud monitoring has a plethora of proposed requirements [13] [14] [15] [22] [25] [33]. However, the current researches just show a list of cloud monitoring requirements without concerns about questions such as difference between requirements and abilities, and a definition of monitoring focus. Requirements/properties are specific features of a cloud computing environment that monitoring solutions have to support. Abilities are specific features that a monitoring system has to improve its duties on particular points such as accuracy, and autonomy. In addition, regardless of requirement or ability, design and deployment of monitoring solutions depend on what kind of resource or service will be monitored. For this reason, a definition of monitoring focus is an important issue. In terms of tools/solutions, cloud monitoring is currently performed not only by solutions specifically designed for clouds, but also by solutions originally conceived for general purpose monitoring. Cloud-specific solutions include Accelops [1], and Amazon CloudWatch [2]. More generic solutions include MRTG [9], and Nagios [10]. It is reasonable to assume that there is a varying quality in cloud monitoring solutions. However, there is a lack in the literature of a proper evaluation of cloud monitoring solutions that could offer to InPs, SPs, and customers a better view over such solutions.

In this paper, we present an overview on cloud monitoring and a comparison among relevant cloud monitoring solutions. In complement, we analyze trends on monitoring.
cloud computing environments, and propose future directions. The main contributions of this paper are:

- It distinguishes the concepts of cloud monitoring requirements and cloud monitoring abilities;
- It introduces the concept of cloud monitoring focus;
- It compares cloud monitoring solutions;
- It analyzes monitoring trends and future directions in cloud computing environments.

The remaining of this paper is organized as follows. Section 2 discusses relevant concepts for cloud monitoring. Section 3 introduces cloud monitoring focus. A comparison among cloud monitoring solutions is presented in Section 4. Section 5 presents cloud monitoring trends and future directions. Section 6 concludes the paper.

2. MONITORING CONCEPTS

Cloud computing environments have introduced features that must be addressed by cloud monitoring systems through requirements. In this section, we highlight some cloud features along with requirements that must be supported by cloud monitoring systems [13] [22] [33]:

- **Scalability**: Scalability is the capacity to improve the performance of the system by increasing the computational resources. In order to fulfill this feature, the monitoring system needs to keep monitoring efficient with a potentially large number of probes;
- **Elasticity**: Elasticity is the competence to increase and decrease computational resources on demand, according to the goal of a specific application or system. Elasticity aims to improve a cloud computing environment in terms of performance and cost. To support elasticity, the monitoring system needs to track virtual resources created/destroyed by expanding/contracting a cloud and to correctly handle expansion/retraction of the system;
- **Migration**: Migration is the capacity to change the location of computational resources according to the goals of a specific application or system. Migration has provided improvements to users in terms of performance, energy consumption, and costs. In migration, any virtual resource that moves from one physical host to another must be monitored correctly to ensure that no information is lost upon migration, and that the monitoring system is not negatively affected by the potential migration of a monitored resource;
- **Accuracy**: Accuracy is the ability of monitoring systems to measure without making mistakes. In cloud computing environments, accuracy is important because SLAs are an intrinsic part of the system; thus, poor performance can lead to financial penalties to InPs and SPs and loss of customers confidence that may damage the reputation of the company and lead to permanent reduction of the customer base;
- **Autonomy**: In clouds, dynamicity is a key factor because changes are intense and frequent. Autonomy is the ability of a monitoring system to self-manage its configuration to keep itself working in a dynamic environment. Enabling autonomy in a cloud monitoring system is complex, since it requires the ability to receive and manage inputs from a plethora of probes;
- **Comprehensiveness**: Cloud computing environments encompass several types of resources (e.g., different virtualization resources, different physical resources) and information. Comprehensiveness is the ability of a monitoring system to support several types of resources and information. Therefore, the monitoring system must have the ability to retrieve updated status from different types of resources, several types of monitoring data, and a large number of users.

3. MONITORING STRUCTURE

Usually, a cloud has a large number of resources on data centers that are geographically spread. Such resources must be continuously monitored, since cloud entities (e.g., SPs, InPs) need information related to these resources, mainly for two reasons. Firstly, to evaluate the status of services hosted in the cloud. Secondly, to use information about resources to perform control activities (e.g., allocation, migration).

In general, cloud services are offered in different service models and are composed of different types of resources (e.g., processing, network). The effective management of a cloud depends on complete monitoring of its structure. To provide a complete monitoring, we consider that a cloud monitoring structure is divided into three components defined as: cloud model, monitoring view, and monitoring focus.

- **Cloud Model**: Clouds are offered on service models. They are Software as a Service (SaaS), when applications ready to be used are provided to customers; Platform as a Service (PaaS), when SPs are offered a platform where applications can be deployed. The InPs controls the allocation of underlying resources, and SPs have only to concern about writing the application; and Infrastructure as a Service (IaaS), where SPs have access to virtual machines where they can install their own platforms and applications [33];
- **Monitoring View**: The view of resources depends on who wants to obtain the information, i.e., InPs, SPs, or customers. InPs are the owners of the infrastructure, and normally are concerned about the infrastructure’s correct operation and efficient utilization. InPs may get information about both virtual and physical layers. Besides, InPs can make control activities over the layers. SPs are the guiding support to customers. SPs, in general, can obtain information about the virtual layer, such as response times and latencies observed across different elements of the platform, and how it relates to the performance observed by customers. Customers, in turn, can see information about the high level application/services they are using. Thus, monitoring view must be set so as to cope with different visions to different InPs, SPs, and customers [33];
- **Monitoring Focus**: Design and implementation of monitoring solutions depend on the type of resource (e.g.,
processing, network) or service (e.g., SLA, QoS) to be monitored. Monitoring focus is the goal (resource type or service) defined by a specific monitoring solution or group of monitoring solutions so as to attend the specific requirements of InPs, SPs, and customers. Monitoring focus can be divided using two methods: by cloud model or by goal. The first one refers to the service model: SaaS, PaaS, or IaaS. The second refers to the goal/objective of the monitoring performed by InPs, SPs, or customers (e.g., SLA, billing).

Figure 1 shows a cloud monitoring structure, depicting the cloud models that compose a cloud, monitoring views to both SPs and customers, and monitoring focus. In this scenario, monitoring focus has several goals. In general, these goals are reached by monitoring solutions that are developed to address specific monitoring necessities (specific goal), i.e., monitoring the cloud models and/or achieve monitoring requirements (e.g., SLA, QoS).

In the next subsections, we highlight the main goals of cloud solutions in order to present a landscape to the development of monitoring solutions. The main goals of the solutions are defined according to their cloud model. Furthermore, we discuss research challenges based on specific goals (e.g., basic metrics, self-configuration) of each model.

### 3.1 Infrastructure—IaaS

In the IaaS, cloud resources are created on top of the bare hardware, which is often performed with the use of virtualization technologies. At IaaS, monitoring solutions acting on behalf of InPs monitor the actual hardware supporting the infrastructure, whereas SPs aim to get information about the virtual resources that are rented by them. IaaS is either offered by public IaaS cloud providers such as Amazon EC2 [3], or is built as private clouds by using solutions such as Eucalyptus [34] and OpenNebula [38]. Resources offered at IaaS are typically in the form of virtual machines (e.g., Xen [16]). Virtual machines are composed of resources such as processing, and storage. Therefore, at IaaS, cloud monitoring solutions have goal of monitoring basic metrics (e.g., processor load and network usage) [22] [37].

### 3.2 Platform—PaaS

The PaaS is composed of both, programming environments and execution environments. Commercially, Google App Engine [6], and Heroku [7] are examples of PaaS. PaaS aims to provide an environment to applications development (e.g., APIs, programming language). Besides, at PaaS services are provided to support the deployment and execution of applications, including features such as fault tolerance, auto scaling, and self-configuration [19] [26] [35]. At PaaS, cloud monitoring provides information to assist a given InP to deal with issues such as self-configuration and fault tolerance management. From a SP perspective, monitoring has the goal of ensuring that the platform is supporting a responsive application, as observed by customers.

### 3.3 Software—SaaS

At SaaS, there are applications of interest to potentially millions of users that are geographically spread. An example of this is online alternatives for typical office applications such as word processors and spreadsheets [39].

Besides, the diversity of applications in clouds is growing. To handle the diversity at SaaS, a cloud monitoring system needs to have unusual abilities such as coping with heterogeneous APIs and coping with different monitoring slices [23]. Additionally, SPs and customers have defined SLAs to regulate the agreement between both. Therefore, SLAs need to be respected and accomplished [18] [19].

### 4. MONITORING SOLUTIONS

We divide monitoring solutions for cloud computing environments in three types: generic solutions, cluster and grid solutions, and cloud-specific solutions.

Generic solutions have been created to monitor computational systems without concerns about specific peculiarities relating to each type of system. These solutions are widely used in computational systems to retrieve information about global hosted resources. However, generic solutions may not be suitable in regards to some specific features of clouds, such as virtualization and server consolidation. Cluster and grid solutions, in turn, where created with these specific domains in mind but they also lack support for cloud-specific features. For this reason, it is required the design and development of cloud-specific solutions.

In the next subsections we review some monitoring solutions. Table 1 shows a summary of the main goal of monitoring solutions presented in this section. Goals are defined according to the most important purpose described in the literature related to each research/solution. Additionally, main ability to each monitoring solution is described.

#### 4.1 Generic Monitoring Solutions

Generic monitoring solutions are designed often without a specific context, and therefore are suitable to be used in computational systems in general. Although generic solutions have been usually created before the emergence of cloud computing environments, we can find initiatives to explore the utilization of those solutions in clouds. At clouds, generic solutions can be used to monitor basic metrics (e.g., memory, network). That is the case, for example, of Cacti [4], MRTG [9], and Nagios [10].

**Cacti** [4] and MRTG [9] are solutions to create RRD-Tool graphs that are usually used to show bandwidth consumption in network links, but that can both plot graphs to any monitoring metric such as processing and storage. Cacti and MRTG, however, do not provide features such as self-configuration or support for discovery to help in their own configuration. In clouds, Cacti and MRTG are used
to build graphs from basic metrics and aggregated metrics (e.g., number of cloud slices, amount of available resources).

**Nagios** [10] is a monitoring tool widely employed in traditional environments. One of Nagios main features is the support for plugins that are used to collect monitoring information from the monitored objects. These plugins can be easily developed and leverage the Nagios ‘flexibility’ that allows monitoring virtually any type of environment. Therefore, Nagios ‘flexibility’ allows the development of plugins to collect basic metrics and aggregated metrics in clouds. Cacti, MRTG, and Nagios do not handle cloud monitoring requirements/abilities such as elasticity and autonomy. For instance, these solutions do not support self-configuration. Thus, new solutions focusing on the requirements of cloud monitoring are required.

### 4.2 Cluster and Grid Monitoring Solutions

Many monitoring systems were proposed in the literature to handle cluster and grid systems. Monitoring systems specific to clusters include PARMON [20] and RVision [27], whereas grid monitoring systems include GridEye [28] and Ganglia [31].

There are clear overlaps between cluster and grid requirements and cloud requirements. For example, clusters, like clouds, are composed of many machines connected in local networks. However, clusters do not have SLA as a key priority like clouds have; furthermore, clouds strongly depend on virtualization, whereas this is not always the case for clusters. Grids tend to be geographically distributed and belong to autonomous management domains, whereas clouds have a large scale infrastructure managed by a single organization. This is a reality in emerging InterCloud approaches where multiple (potentially distinct) cloud services are aggregated to provide a service to the final customer. Similarly to the cluster case, grid systems do not have SLAs as their key priority, and they tend to be cooperative environments rather than financially-driven services.

### 4.3 Cloud Specific Monitoring Solutions

Cloud specific monitoring solutions have been created to be used in cloud computing environments. Currently, cloud specific monitoring solutions are designed by academic research or commercial efforts.

**Amazon CloudWatch** [2] is a monitoring solution for Amazon Web Services (AWS). Amazon CloudWatch allows easy handling of basic metrics such as processing and storage. Additionally, it presents several types of statistics and self-configuration. Thus, Amazon CloudWatch is a good solution for users and managers of Amazon clouds, however, it is restricted to AWS products. Other commercial cloud monitoring solutions include Accelops [1], Copperegg [5], Zennoss [12], Monitis [8], and Rackspace Cloud Monitoring [11], that, like CloudWatch, focus on specific proprietary platforms/solutions.

**Private Cloud Monitoring Systems** (PCMONS) [24] is a monitoring solution for private clouds. PCMONS is an open source solution that uses a layer called Integration to provide homogeneous access to users and managers that manipulate resources in a cloud. It provides a uniform monitoring of infrastructure, independently of type of resource hosted in a cloud. In addition, other monitoring solutions can be used as support and complement PCMONS, promoting an integration among monitoring solutions. On the other hand, the configuration of monitoring must be done manually, which compromises cloud monitoring requirements/abilities such as scalability, and autonomy.

**Cloud Management System** (CMS) [29] aims to provide a monitoring solution based on RESTful Web Services. CMS employs REST to allow the development and integration of monitoring solutions. The REST system can design monitoring elements (e.g., network). The Get method in REST can replace the operations of monitoring. Because REST is widely used for web services in research centers and commercial applications [32], CMS can be integrated with other solutions, services, and technologies that use REST.

**Runtime Model for Cloud Monitoring** (RMCM) [37] aims to monitor resources through abstract models, making possible homogeneous handling of heterogeneous resources. In this way, it is possible to work with different resources, such as platforms and virtual systems in a same approach. In addition, it generates customized models according to the needs of each agent that integrates a cloud. These models have been defined as model for operators, model for developers, and model for users. However, it requires a constant update of monitoring resources in order to maintain the model consistent. The main disadvantage of this solution is related to the manual installation and configuration of specific agents. For this reason, cloud monitoring requirements/abilities such as scalability, migration, and autonomy are compromised.

The System of monitoring/warning that operates over the mOSAIC [35] platform generates warnings when a SLA is in risky conditions, in other words, the monitoring system observes SLA rules and when they are close to being violated, a message is send to managers. It is a simple and efficient method to maintain control over monitoring resource.

**Flexible Automated Cloud Monitoring Slices** (FlexACMS) [23] aims to integrate several monitoring solutions to provide a comprehensive cloud solution. FlexACMS allows cloud administrators to automatically configure monitoring solutions, providing flexibility and dynamicity to cloud providers. Flexibility and dynamicity improve operational aspects of cloud providers such as billing and SLA. To reach such automatic configuration, FlexACMS is based on rules defined by cloud administrators that determine the metrics that must be monitored on each cloud slice and what monitoring solutions must be used to monitor each slice.

<table>
<thead>
<tr>
<th>Solution</th>
<th>Main Goal</th>
<th>Main Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloudwatch</td>
<td>Basic Metrics</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Zennoss</td>
<td>IaaS</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Accelops</td>
<td>Self-Config</td>
<td>Autonomy</td>
</tr>
<tr>
<td>Copperegg</td>
<td>SaaS</td>
<td>Autonomy</td>
</tr>
<tr>
<td>Monitis</td>
<td>SaaS</td>
<td>Autonomy</td>
</tr>
<tr>
<td>Rackspace</td>
<td>SaaS</td>
<td>Autonomy</td>
</tr>
<tr>
<td>PCMONS</td>
<td>Integrated</td>
<td>Comprehensiveness</td>
</tr>
<tr>
<td>CMS</td>
<td>Integrated</td>
<td>Comprehensiveness</td>
</tr>
<tr>
<td>mOSAIC</td>
<td>SLA</td>
<td>Accuracy</td>
</tr>
<tr>
<td>RMCM</td>
<td>Integrated</td>
<td>Comprehensiveness</td>
</tr>
<tr>
<td>MRTG</td>
<td>Basic Metrics</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Cacti</td>
<td>Basic Metrics</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Monitis</td>
<td>Basic Metrics</td>
<td>Accuracy</td>
</tr>
<tr>
<td>FlexACMS</td>
<td>Integrated</td>
<td>Comprehensiveness</td>
</tr>
</tbody>
</table>
5. TRENDS AND FUTURE DIRECTIONS

Monitoring plays an important role in cloud computing and enables the delivery of services meeting agreed contracts/SLAs. Currently, there are several solutions with different purposes of monitoring such as target application, basic metrics, and target infrastructure. However, some areas in cloud computing environments are growing, and represent trends of research on monitoring. Therefore, research opportunities for monitoring in clouds have emerged. In this scenario, we highlight as open research opportunities in cloud computing integrated monitoring and energy efficiency.

Integrated monitoring is a trend for two main reasons. First, there are several consolidated monitoring solutions that integrate themselves in a natural way. Second, there are sundry goals to monitoring solutions depending on the service model and the role of the monitor (e.g., InP, or SP). Thus, design and development of an overall monitoring solution is a hard task. Additionally, integrated monitoring is a goal of recent developments in the area of cloud monitoring (e.g., PCMONS [24] and RMCM [37]).

Furthermore, there is a trend in cloud computing towards energy management, green computing, and reduction of operational costs. Actually, it confirms a trend towards energy efficiency. This is evidenced by recent work such as Beloglazov et al. [17], Rodero et al. [36], and Wang et al. [40]. In this context, monitoring is paramount to enable reduced energy consumption without compromising application performance and SLAs.

The aforementioned issues are generic, i.e., they are enforced for all cloud models. To be more specific, we present below trends and future directions to each cloud model.

In the IaaS model, monitoring issues such as energy efficiency [17] [21] and integrated monitoring [24] [30] are especially important. Energy efficiency concerns about reasonable consumption of power to operate the service. It aims to reduce the power waste. Integrated monitoring has concerns about integration of several monitoring solutions that operate in a cloud computing environment. It aims to develop comprehensive cloud monitoring solutions or promote integration among monitoring solutions.

In the context of IaaS monitoring, open challenges include translation of higher services objectives in effective lower-level metrics as observed in the infrastructure layer, and unintrusive accomplishment of SLAs.

In the PaaS model, a gap exists in self-configuration. Self-configuration has addressed adaptation of cloud monitoring solutions to features such as dynamicity and elasticity. However, to support these features, cloud monitoring solutions demand more resources from infrastructure to cater for SLAs. Therefore, an emerging issue in the area of PaaS monitoring concerns improved techniques for self-configuration to reach a suitable balance between consumption of infrastructure resources and SLA constraints.

In the SaaS model, because of the diversity of customers, applications, and SLAs, cloud monitoring systems must handle different scenarios. There are different customer profiles with different SLAs using different applications. The monitoring system needs to correctly identify the customers and ensure that privileged customers are actually getting better service than regular customers.

In this scenario, emerging research questions include how to fulfill SLAs without being invasive (impairing other SLAs), and how to improve the profit of SPs without compromising quality of service to customers.

Additionally, when a cloud monitoring system aims to accomplish a specific requirement, it is usually negatively or positively affected by other requirements. Therefore, the development of monitoring systems focused on clouds aims at improving specific aspects of cloud operation, providing partial solutions for cloud monitoring. Thus, the balance among cloud monitoring requirements is a challenging and important trend. Regarding abilities, it is important to develop solutions that can be integrated. Thereby, cloud monitoring solutions designed to cope with specific abilities could be complemented by other solutions, resulting in a complete solution for cloud monitoring.

Finally, several gaps remain and there are several challenging research directions to be explored such as creation of comprehensive monitoring solutions, improved techniques for self-configuration, translation of higher service objectives in effective lower level metrics as observed in the infrastructure layer, and unintrusive accomplishment of SLAs.

6. CONCLUSIONS

In this paper, we presented an overview on cloud monitoring aiming to distinguish the concepts of cloud monitoring requirements and cloud monitoring abilities. We also introduced the concept of cloud monitoring focus.

Moreover, we presented a comparison among cloud monitoring solutions and discussed trends and future directions in the area to predict a future landscape in order to assist the design and development of new cloud monitoring solutions.

7. REFERENCES

Research dimensions, design issues, and state-of-the-art. CoRR, abs/1312.6170.


