Survey Paper on Cloud Computing & Cloud Monitoring: Basics

Bhaswati Hazarika[1], Thoudam Johnson Singh[2]
[1](Computer science & Engineering, DBCET/ Assam Don Bosco University, India)
[2](Computer science & Engineering, DBCET/ Assam Don Bosco University, India)

ABSTRACT *: Cloud computing has been used widely for delivering services over the internet, both for technical and economical reasons. The number of services based on cloud has increased strongly and rapidly in the past years. Excellent facilities have been provided to business entrepreneurs by flexible infrastructure. To manage such infrastructures properly effective and efficient monitoring is constantly needed.*

In this paper ,we will provide a survey on Cloud computing & some of its monitoring techniques. We will start analyzing motivations for Cloud monitoring, which will also provide definitions and background for the following contributions. Then, we will carefully analyze and also discuss the properties of a monitoring system for the Cloud, how they have been tackled in literature. Here, we also describe services for Cloud monitoring, underlining how they relate with the properties identified before.

Keywords *- Cloud Computing, Data centers, State Monitoring*

1. Introduction

Cloud Computing has rapidly become a widely adapted paradigm for delivering services over the internet.Giant companies like Google, Microsoft and Amazon are attracted by Cloud computing and is considered as a great influence in today's Information Technology industry.

Business owners are attracted to cloud computing concept because of several features. These are:

- Lower initial investment
- Easier to manage
- Scalability
- Deploy faster
- Location independent
- Device independent
- Reliability
- Security

Cloud services are on-demand, elastic and scalable, and the following main features are therefore needed by a Cloud system: availability, concurrency, dynamic load balancing, and independence of running applications, security, and intensiveness. Data centers for Cloud Computing continue to grow in terms of both hardware resources and traffic volume, thus making Cloud operation and management more and more complex.

In this scenario, accurate and fine-grained monitoring activities are required to efficiently operate these plat forms and to manage their increasing complexity. In this paper, we will also provide a survey of Cloud monitoring techniques. Our aim is to provide a better understanding of Cloud Computing and focus on the ongoing research in this tremendously flourishing arena of computer science.

2. Cloud Computing Overview

2.1 Definitions

Cloud Computing is a way of leveraging the Internet to consume software or other IT services on demand. Users share processing power, storage space, bandwidth, memory, and software. The resources are shared and so are the costs. Users can pay as they go and only use what they need at any given time, keeping cost to the user down. Cloud Computing is very much a business model as well. Providers of Cloud Computing solutions, whether they are platform, hardware, software, or storage providers, deliver their offerings over the Internet. Cloud providers typically charge monthly recurring fees based on the usage.

According to NIST (National Institute of Standards & Technology) Cloud Computing is defined as[1]: "Model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (net works, servers, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.''

A simple example of cloud computing is Yahoo email, Gmail, or Hotmail etc. For this we don't need software or a server to use them. All that a consumer would require is just an internet connection and they can start sending emails. The server and email management software is all on the cloud (internet) and is totally managed by the cloud service provider e.g Google etc. The consumer can use the software alone and enjoy the benefits.

Cloud computing is further broken down into three segments: "application" "storage" and "connectivity." Each segment serves a different purpose and offers different products for businesses and individuals around the world.

2.2 Types of Cloud

Cloud computing is typically classified into two ways [2]:

1. Location of the Cloud Computing
2. Types of services offered

Location of the cloud

Cloud computing is typically classified in the following ways:

1. **Public cloud**: In Public cloud the computing infrastructure is hosted by the cloud vendor at the vendor's premises. The customer has no visibility and control over where the computing infrastructure is hosted. The computing infrastructure is shared between any organizations.

2. **Private cloud**: The computing infrastructure is dedicated to a particular organization and not shared with other organizations. Some experts consider that private clouds are not real examples of cloud computing. Private clouds are more expensive and more secure when compared to public clouds.

3. **Hybrid cloud:** Organizations may host critical applications on private clouds and applications with relatively less security concerns on the public cloud. The usage of both private and public clouds together is called hybrid cloud. A related term is Cloud Bursting. In Cloud bursting organization use their own computing infrastructure for normal usage, but access the cloud for high/peak load requirements. This ensures that a sudden increase in computing requirement is handled gracefully.

4. **Community cloud:** It involves sharing of computing infrastructure in between organizations of the same community. For example all Government organizations within the state of California may share computing infrastructure on the cloud to manage data related to citizens residing in California.

2.3 Classification based on service provided

1. **Infrastructure as a service (Iaas)**involves offering hardware related services using the principles of cloud computing. These could include some kind of storage services (database or disk storage) or virtual servers. Leading vendors that provide Infrastructure as a service are Amazon EC2, Amazon S3, Rackspace Cloud Servers and Flexiscale.

2. **Platform as a service (Paas)** involves offering a development platform on the cloud. Platforms provided by different vendors are typically not compatible. Typical players in PaaS are Google's Application Engine, Microsofts Azure, Salesforce.com's force.com .

3. **Software as a service (Saas)** includes a complete software offering on the cloud. Users can access a software application hosted by the cloud vendor on pay-per-use basis. This is a well-established sector. The pioneer in this field has been Salesforce.coms offering in the online Customer Relationship Management (CRM) space. Other examples are online email providers like Googles gmail and Microsofts hotmail, Google docs and Microsofts online version of office called BPOS(Business Productivity Online standard Suite).

The above classification is well accepted in the industry. David Linthicum describes a more granular classification on the basis of service provided. These are:

1. Storage-as-a-service
2. Database-as-a-service
3. Information-as-a-service
4. Process-as-a-service
5. Application-as-a-service
6. Platform-as-a-service
7. Integration-as-a-service
8. Security-as-a-service
9. Management/Governance-as-a-service
10. Testing-as-a-service

3. Cloud Computing: The need for Monitoring

Monitoring of Cloud is an important task for both Providers and Consumers. On the one side, it is a key tool for controlling and managing hardware and soft- ware infrastructures; on the other side, it provides information and Key Performance Indicators (KPIs) for both platforms and applications. The continuous monitoring of the Cloud and of its SLAs (for example, in terms of availability, delay, etc.) supplies both the Providers and the Consumers with information such as the workload generated by the latter and the performance and QoS offered through the cloud, also allowing to implement mechanisms to prevent or recover violations (for both the Provider and Consumers).Monitoring is clearly instrumental for all the activities covered by the role of Cloud Auditor. In more general terms, Cloud Computing involves many activities for which monitoring is an essential task.

In this Section we carefully analyze such activities, underlining the role of monitoring for each of them.

3.1 Data Center Management

Cloud services are provided through large scale datacenters, whose management is a very important activity. Actually, this activity is a part of resource management and we reported it here because of its importance and of its peculiar requirements. Data center management activities (e.g. data center control) imply two fundamental tasks: (i) monitoring, it keeps track of desired hardware and software metrics; and (ii) data analysis, that processes such metrics to infer system or application states for resource provisioning, troubleshooting, or other management actions [3]. In order to manage such data centers properly, both monitoring and data analysis tasks must support real-time operation and scale up to tens of thousands of heterogeneous nodes, dealing with complex network topologies and I/O structures. In this context energy-efficiency is a major driver of monitoring data analysis for planning, provisioning and management of resources.

3.2 SLA Management

The unprecedented flexibility in terms of resource management provided by Cloud Computing calls for new programming models in which Cloud applications can take advantage of such new feature [4], whose underlying premise is monitoring. Moreover, monitoring is mandatory and instrumental in certifying SLA compliance when auditing activities are performed to respect regulation [5] (e.g. when government data or services are involved). Finally, monitoring may allow Cloud providers to formulate more realistic and dynamic SLAs and better pricing models by exploiting the knowledge of user-perceived performance[6].

4. Cloud Monitoring: Basics

Cloud monitoring is needed to continuously measure and assess infrastructure or application behaviors in terms of performance, reliability, power usage, ability to meet SLAs, security, etc. [7], to perform business analytics, for improving the operation of systems and applications [8], and for several other activities.

5. Cloud Monitoring: Properties

In order to operate properly, a distributed monitoring system is required to have several properties that, when considered in the Cloud Computing scenario, introduce new issues. Some of the properties are listed below:

5.1 Scalability

A monitoring system is scalable if it can cope with a large number of probes [9]. Such property is very important in cloud computing scenarios due to the large number of parameters to be monitored about a huge number of resources. The measurements required to obtain a comprehensive view on the status of the cloud lead to the generation of a very large volume of data coming from multiple distributed locations. hence, a scalable monitoring system should be able to efficiently collect, transfer and analyze such volume of data without impairing the normal operations of the cloud.

5.2 Elasticity

A monitoring system is elastic if it can cope with dynamic changes of monitored entities, so that virtual resources created and destroyed by expansion and contractions are monitored correctly [9]. Such property, also referred to as dynamism [10], implies scalability and adds to it the requirement of supporting on-line upsizing or downsizing of the pool of monitored resources.

5.3 Adaptability

A monitoring system is adaptable if it can adapt to varying computational and network loads in order not to be invasive (i.e. impeding for other activities) [9]. Due to the complexity and the dynamism of the Cloud scenarios, adaptability is fundamental for a monitoring system in order to avoid as much as possible a negative impact of monitoring activities on normal Cloud operations, especially when active measurements are involved. In fact, the workload generated by active measurements, together with the collection, processing, transmission and storage of monitoring data and the management of the monitoring subsystem, require computing and communication resources and therefore constitute a cost for the Cloud infrastructure.

5.4 Timeliness

A monitoring system is timely if detected events are available on time for their intended use [3]. Monitoring is instrumental to activities related with core goals of a Consumer or a Provider, hence failing to get the necessary information on time for the appropriate response (e.g. to raise an alarm, to provision more resources, to migrate services, to enforce a different policy) would void the usefulness of monitoring itself. Timeliness is interdependent.

5.5 Autonomicity

An autonomic monitoring system is able to self-manage its distributed resources by automatically reacting to unpredictable changes, while hiding intrinsic complexity to Providers and Consumers [28]. Supporting autonomicity in a Cloud monitoring system is not trivial, since it requires to implement a control loop that receives inputs from a huge number of sensors (i.e. the monitoring data) and propagates control actions to a large number of distributed actuators.

5.6 Accuracy

We consider a monitoring system to be accurate when the measures it provides are accurate, i.e. they are as close as possible to the real value to be measured. There are two main issues related to the accuracy of Cloud monitoring systems: the workload used to perform the measurements, and the impact of virtualization systems that add additional layers between applications and physical resources.

6. Cloud Monitoring: Issues

6.1 Effectiveness

Main issues reside in the possibility to have a clear view of the Cloud and to pinpoint the original causes of the observed phenomena. To achieve this, improvements are needed in terms of: (i) custom algorithms and techniques that provide effective summaries, filtering and correlating information coming from different probes (ii) very importantly accurate measures in an environment dominated by virtualized resources. we believe that more effort is required for currently available Cloud monitoring systems in order to be also reliable. Even if implicitly addressed in Scalability and Adaptability issues, Timeliness in itself is explicitly considered and evaluated only in [19]. This is a fundamental property that can be effectively used to quantitatively evaluate a Cloud monitoring system and objectively compare it with alternatives (e.g. by defining a specific kind of monitored event and measuring the time needed for the information to reach the managing application).

Future proposals and comparisons of Cloud monitoring systems should include the use of the related metric, Time to insight, and further research is needed in this field to model the relations among the parameters involved in Timeliness. Similar considerations can be made about the property of Availability of a monitoring system: though it is closely related with Scalability and Reliability, at the best of the knowledge of the authors there are no evaluations in terms of percentage of missed events, unanswered queries and similar failures in employing the monitoring subsystem and no explicit design constraints in ensuring a given level of availability – possibly 100% as monitoring is a critical functionality.

6.2 Efficiency

Main improvements in terms of efficiency are expected for data management. In particular, algorithms and techniques more and more efficient are needed to manage, quickly and continuously, the large volume of monitoring data necessary to have a comprehensive view of the Cloud, without putting too much burden on the Cloud and monitoring infrastructures both in terms of computing and communication resources. The monitoring system should be therefore able to do several operations on data (collect, filter, aggregate, correlate, dissect, store, etc.) respecting strict requirements in terms of time, computational power, and communication overhead. These requirements become more and more strict with the increasing spread of Cloud Computing and therefore, the increasing number of users and resources.

7. State Monitoring in Cloud Datacenter

State monitoring is widely used for detecting critical events and abnormalities of distributed systems. As the scale of such systems grows and the degree of workload consolidation increases in Cloud datacenters, node failures and performance interferences, especially transient ones, become the norm rather than the exception. Hence, distributed state monitoring tasks are often exposed to impaired communication caused by such dynamics on different nodes. Unfortunately, existing distributed state monitoring approaches are often designed under the assumption of always online distributed monitoring nodes and reliable inter-node communication. As a result, these approaches often produce misleading results which in turn introduce various problems to Cloud users who rely on state monitoring results to perform automatic management tasks.

State monitoring is a fundamental building block for many distributed applications and services hosted in Cloud datacenters. It is widely used to determine whether the aggregated state of a distributed application or service meets some predefined conditions [11]. For example, a web application owner may use state monitoring to check if the aggregated access observed at distributed application-hosting servers exceeds a pre-defined level [14].

Most existing state monitoring research efforts have been focused on minimizing the cost and the performance impact of state monitoring. For example, a good number of state monitoring techniques developed in this line of works focus on the threshold based state monitoring by carefully partitioning monitoring tasks between local nodes and coordinator nodes such that the overall communication cost is minimized [3][2][4][1]. Studies along this direction often make strong assumptions on monitoring-related communications, such as 100% node availability and instant message delivery.

7.1 Efficient Continuous State Violation Detection

Most existing works on distributed state monitoring employ an instantaneous monitoring model, where the state is evaluated based on the most recent collected results, to simplify algorithm design. Such a model, however, tends to introduce false state alerts due to noises and outliers in monitor monitoring data. To address this issue, we proposed WISE, window based state monitoring which utilizes temporal windows to capture continuous state violation in a distributed setting. WISE not only delivers the same results as those of a centralized monitoring system with a distributed implementation, but also decouples a global monitoring task into distributed local ones in a way that minimizes the overall communication cost.

7.2 Fault-Tolerant State Monitoring

While we often assume monitoring results are trustworthy and monitoring services are reliable, such assumptions do not always hold, especially in large scale distributed environments such as datacenters where transient device/network failures are the norm rather than the exception. As a result, distributed state monitoring approaches that depend on reliable communication may produce inaccurate results with the presence of failures. Thus Crystal Ball, a robust distributed state monitoring approach that produces reliable monitoring results by continuously updating the accuracy estimation of the current results based on observed failures. It also adapts to long-term failures by coordinating distributed monitoring tasks to minimize accuracy loss.

8. Conclusion

The advancement of cloud computing is dramatically changing the horizon of information technology and ultimately turns the utility computing into a reality. However, it provides a large array of benefits, but many challenges in this domain, including automatic resource positioning, energy management, information security are only attracted the research community. There are still so many issues to be explored. Opportunities are enough in this arena for some groundbreaking contribution and bring significant development in the industry. a reliable state monitoring approach enables estimation of monitoring accuracy based on observed messaging dynamics, and self-adaption to disruptions we have discussed the main activities in Cloud environ ment that have strong benefit from or actual need of monitoring. To contextualize and study Cloud monitoring, we have provided background and definitions for key concepts . We have also derived the main properties that Cloud monitoring systems should have. Cloud computing is at an early stage of research and development, we believe our paper will provide a better understanding of the cloud computing.

References

[1] P. Mell, T. Grance, *The NIST Definition of Cloud Computing, NIST Special Publication* 800-145, 2011.

[2]http://thecloudtutorial.com/cloudtypes.html

[3] C. Wang, K. Schwan, V. Talwar, G. Eisenhauer, L. Hu, M. Wolf, *A flexible architecture integrating monitoring and analytics for managing large-scale data centers, in: Proceedings of ICAC,* 2011

[4] M. Rak, S. Venticinque, T. Mahr, G. Echevarria, G. Esnal, *Cloud application monitoring: the mOSAIC approach*, in: 2011.

[5] P. Massonet, S. Naqvi, C. Ponsard, J. Latanicki, B. Rochwerger, M. Villari, *A monitoring and audit logging architecture for data location compliance in federated cloud infrastructures, in: 2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum* (IPDPSW), 2011.

[6] A. Khurshid, A. Al-Nayeem, I. Gupta, *Performance evaluation of the illinois cloud computing testbed*, Unpublished, Tech. Rep., 2009.

[7] M. Kutare, G. Eisenhauer, C. Wang, K. Schwan, V. Talwar, M. Wolf, *Monalytics: online monitoring and analytics for managing large scale data centers, in: Proceedings of the 7th International Conference on Autonomic Computing*, Ser. ICAC '10, ACM, NewYork, NY, USA, 2010, pp. 141–150.

[8] V. Kumar, Z. Cai, B.F. Cooper, G. Eisenhauer, K. Schwan, M. Mansour, B. Seshasayee, P. *Widener, Implementing diverse messaging models with self-managing properties using IFLOW*, in: ICAC, 2006.

[9] S. Clayman, A. Galis, L. Mamatas, *Monitoring virtual networks with lattice, in: Network Operations and Management Symposium Workshops (NOMS Wksps)*, 2010 IEEE/IFIP, 2010, pp. 239–246.

[10] P. Hasselmeyer, N. d'Heureuse, *Towards holistic multi-tenant monitoring for virtual data centers, in: Network Operations and Management Symposium Workshops (NOMS Wksps)*, 2010 IEEE/IFIP, 2010, pp. 350–356.

[11] S. Meng, L. Liu, and T. Wang. *State monitoring in cloud datacenters. IEEE Trans. Knowl. Data Eng. (TKDE), 23(9)*:1328–1344, 2011.

[12] S. Meng, T. Wang, and L. Liu. *Monitoring continuous state violation in datacenters: Exploring the time dimension. In ICDE*, pages 968–979,2010.

[13] S. Meng, A. Iyengar, I. Rouvellou, L. Liu, K. Lee, B. Palanisamy, and Y. Tang. *Reliable state monitoring in cloud datacenters. In IEEE CLOUD*, pages 951–958, 2012.

[14] B. Raghavan, K. V. Vishwanath, S. Ramabhadran, K. Yocum, and A. C. Snoeren, "*Cloud control with distributed rate limiting," in SIGCOMM07.*

[15] Mohiuddin Ahmed1, Abu Sina Md. Raju Chowdhury2, Mustaq Ahmed3, Md. Mahmudul Hasan Rafee4 "*An Advanced Survey on Cloud Computing and State-of-the-art Research Issues"*

[16] Giuseppe Aceto, Alessio Botta, Walter de Donato, Antonio Pescapè *"Survey Paper Cloud monitoring: A survey"*.

[17] Shicong Meng Arun K. Iyengarz Isabelle M. Rouvellouz Ling Liu Kisung Lee Balaji Palanisamyy Yuzhe Tang *"Reliable State Monitoring in Cloud Datacenters"* College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA IBM Research T.J. Watson, Hawthorne, NY 10532, USA

[18] Pankesh PatelWright State University - Main Campus Ajith H. RanabahuWright State University - Main Campus Amit P. Sheth,Wright State University - Main Campus.*Service Level Agreement in Cloud Computing*

[19] C. Wang et al., *"A flexible architecture integrating monitoring and analytics for managing large-scale data centers",* Proceedings of ICAC, 2011.

[20] Giuseppe Aceto, Alessio Botta, Walter de Donato, Antonio Pescapè University of Napoli Federico II (Italy), {giuseppe.aceto,a.botta, walter.dedonato, pescape}.*Cloud Monitoring: definitions, issues and future directions.*