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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 10, October 2013

Failure Management In Cloud : An Overview

R. Jemina Priyadarsini[1] and L. Arockiam[2]

1. Research Scholar, St. Joseph's College (Autonomous), Tiruchirappalli, India[1]

2 .Associate Professor in Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli, India [2]

Abstract: Cloud Computing provides an on-demand network access to computing resources. Estimation of failures are done by predictors. An efficient solution on failure management is highly required in cloud storage system. Failure management approach as a whole comprises of mechanisms like prediction, detection, and correction. The task of resource manager is to identify fatal events and to provide effective resource provisioning so as to improve reliability. This paper analyses failure issues in Cloud environment as a part of Cloud computing services to satisfy reliability. Failure occurs when the system deviates from fulfilling its normal execution without satisfying the customer constraints. Cloud environments come across three types of failures, the hardware, software, and Virtual Machine (VM) failure. The main aim is to enhance the system reliability which is one of the Quality of Service (QoS) issues and thereby decreasing the failure events. Failure Occurrences directly affects the systems reliability. This paper is a survey which is more specific to exposes the failure management models and issues in cloud environments with respect to enhancement of reliability.

Keywords: Failure, Failure management, Reliability, Quality of Service (QoS), Virtual Machine (VM).

I. INTRODUCTION

Cloud Computing system provides access to large amount of data and computational resource that is acquired and released on-demand. Cloud makes it possible for you to access your information from anywhere at any time. Cloud Computing is a special way of delivering IT services and resources over the Web, using rapid, self-service provisioning, while insulating the user of the services or resources from procuring and implementing hardware and software to use IT resources. Distributed storage system generally use data replication to support data reliability [1][2],for example data storage system like Amazon S3,Google File system and Hadoop Distributed File System[3],apply 3-replica data replication strategy. The cloud providers are responsible for the cloud resources and services. The aim of Cloud Computing is to deliver reliable, fault-tolerant, sustainable and scalable infrastructure for hosting services. Cloud resource is categorized into two; *the hardware and the Virtual Machine (VM).*Virtual machines provide great advantages over the installation of OS and software directly on

physical hardware. VMs can also be easily moved, copied, and reassigned between host servers to optimize hardware resource utilization. A single failure will cause multiple computers to be unavailable, which will directly affect resource utilization [4]. Three types of failures identified in cloud environments are the Hardware failure, VM failure, and Application failure. Failure management includes three distinct phases: Prediction, Detection and Correction. In the above stated three failures, the *Hardware failure* can only be detected and corrected by cloud provider; *VM failure* can be detected by customer or provider and repaired only by cloud provider; and the Application failure can only be detected by customer and corrected by customer or provider. Study of failure instances, cause of failure and system responses to failures is of very much importance in designing a good fault tolerant system. Thus, failure is an event in which the system fails to perform according to its specified functions [16]. Fault-tolerance is an important research topic in the study of cloud efficiency [5].This paper describes the failure management models and also virtualization which is one of the key attribute for resource provisioning. Section II provides an overview of cloud computing, section III contains the QoS issues, section IV explains the background and related works, section V elaborates the virtualization concepts and section VI focus on the issues and challenges related to failure management and also highlights the importance of virtualization and resource provisioning. We conclude on section VII along with references.

II. CLOUD COMPUTING

Internet is considered to be the foundation for cloud environment. But Cloud is something more than the internet; wherever the user goes he can use the technology. User need not pay for the technology when he is not using it. Three main criteria which are indispensable in cloud environment are,

- Services are delivered through a web browser
- Zero capital investment

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 10, October 2013

- Ubiquitous
- Pay only for what we use [6].

The following figure shows the three basic service models namely Software as a Service (SaaS)**,** Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) along with the three basic types of cloud public, private and hybrid.

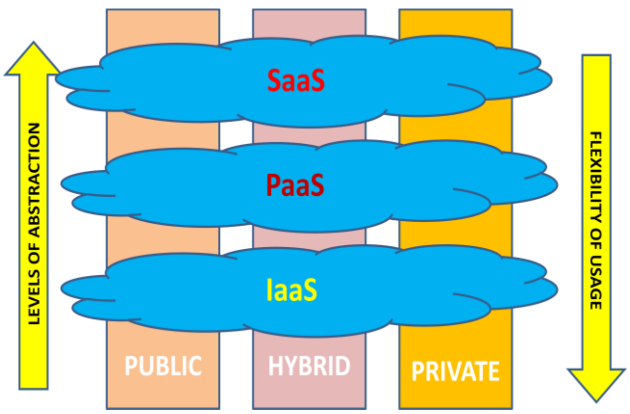

Figure 1: Types of cloud and service models

A. Types of Cloud

Cloud Computing is basically classified into three types as shown in Fig(i).

***1*. Public cloud:**
In Public cloud the vendor hosts the computing infrastructure at his premises. The customer is not given visibility and control over the computing infrastructure. organizations share the computing infrastructure between themselves

***2*. Private cloud:**
The private cloud computing infrastructure is totally given to a particular organization and not shared with other organizations. Sometimes the private clouds are not considered as examples of cloud computing. They are more expensive and more secure when compared to public clouds.

***3*. Hybrid cloud**:

The usage of both private and public clouds together is called hybrid cloud. It is also referred to as Cloud Bursting [7].

B. Service Models:

- Software as a Service - **SaaS**
- Platform as a Service - **PaaS**
- Infrastructure as a Service- **IaaS**

1 . SaaS:

Consumer can use the providers' applications running on a cloud infrastructure. These applications are accessible from client devices through web browser. In SaaS also, the consumer cannot manage the underlying infrastructure [23].As given in Fig.(i) the flexibility of usage is less when compared to other service models. Especially SaaS doesn't offer the flexibility of creating custom software like IaaS.

2. PaaS:

In cloud computing, Platform as a Service(PaaS) is the capability provided to the consumers to utilize the hardware, operating system, storage and network over the internet. Consumers rent virtualized servers and associated services with the help of service delivery models. The consumer is not given authority to manage the underlying cloud infrastructure, but can control the deployed applications [23].

3. IaaS:

Infrastructure layer is the set of hardware resources. The basic functions of infrastructure layer are resource abstraction, resource monitoring, load management, data management, resource deployment, security management and billing management. Infrastructure layer abstracts physical resource by virtualization technology and achieves the automation within the process by optimizing resource management to provide dynamic and flexible infrastructure services to external resource users [8].As shown in Fig.(i) IaaS provides more flexibility. The user is responsible for maintaining and updating the system and not the provider.

C. Related Technologies

Cloud computing typically has the characteristics of grid computing, utility computing, autonomic computing and virtualization [7]. Section 5 explains the virtualization technology which is one of the key factors in effective resource provisioning to enhance reliability [16].

III. ISSUES IN CLOUD QoS

The main aim of cloud computing is to provide a quality service to its consumers.The major QoS issues in cloud environments are scalability, security, availability, trust and reliability.

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 10, October 2013

A. Scalability:

Cloud computing helps organizations, to scale their computing resources whenever needed. This is done by either increasing or decreasing the required resources. Payment is based on the usage of resource and the organization need not pay for the unutilized resource. Scalability is the ability of the system to expand itself by adding resources, either making hardware stronger (scale up/vertical scalability) or adding additional nodes (scale out/horizontal scalability).

B. Security:

Cloud computing security refers to the set of procedures, processes and standards designed to provide information security assurance in computing environment. Cloud security addresses both physical and logical security issues across all the different service models of software, platform and infrastructure. It also emphasizes about the delivery of services through service models like public, private or hybrid delivery model. Cloud security comprises a broad range of security constraints from an end-user and provider's perspective, where the end-user will always prioritize the provider's security policy. That is the user always has a concern over how and where their data is stored and who has access to that data etc. A cloud provider, on the other hand, takes care of the security issues ranging from physical security of the infrastructure to the access control mechanism and also the execution and maintenance of security policy.

C. Availability:

Availability is the degree to which a system or component is operational and accessible when required for use. In software engineering, availability is measured in terms of mean time between failures and mean time to repair .High availability typically is addressed by means of replicating servers and storage [22]. When a job is submitted to a cloud resource, the resource is said to be unavailable in one of the following situations:

- A part of service of the resource is denied to the user.
- The resource is shut down.

D. Trust:

Trust is considered as the ‘assurance’ and confidence that people, data, entities, information or processes to function or behave in expected ways. Trust may be human to human, machine to machine or machine to human . At a deeper level, trust is regarded as a criteria towards security or privacy objectives.

E. Reliability

Reliability is an important component of trust. It is also called as success rate. Reliability is the ability of a system or component to perform the stated functions under stated conditions for a specified period of time. Reliability of a cloud resource is a measure of successful completion of accepted jobs by the cloud resource [22]. The key issue is to enhance reliability by minimizing failures [8]. Reliability is also defined as the probability that the data is available in the system for a certain period. Data replication supports data reliability. Replication strategy is adopted in cloud environments to enhance reliability.

IV. BACKGROUND

This section presents the background information on various elements that form the basis cloud computing systems. To evaluate the performance of cloud service provider, we need to set an objective to represent the quality experienced by the user. One way to achieve this is by having a study over failure issues and its cause in cloud environments. In general a *failure is an event in which the system fails to operate according to its specifications*. Cause of system failure is due to deviation of system from its normal system functions. An error is also a part of the system state which is liable to subsequent failure [16]. Further in this section we shall discuss on the existing failure prediction, detection and correction methods.

A. Related works on failure management issues

Failure prediction determines the possible occurrences of fatal events in the future [9]. To predict failure health related data’s are to be collected. Health data’s are nothing but, systems performance data. Existing prediction methods are based on statistical learning techniques also called as supervised learning approach. One of the best ways to prevent a failure from spreading in the system is identifying signs of the failure before its occurrence and deal with it proactively before it causes serious problems. Although several approaches have been proposed to predict failures by analyzing past system message logs and identifying the relationship between the messages and the failures, it is still difficult to automatically predict the failure for several reasons such as various types of log message formats or time gaps between message pattern learning and application of the identified patterns in real systems.

Most of the failure prediction models are based on statistical learning methods. Statistical learning methods help in failure diagnosis. Two types of statistical learning methods are

- Supervised learning
- Un-supervised learning.

A Failure prediction mechanism by two learning approaches unsupervised and supervised has also been proposed by

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 10, October 2013

Song Fu et al[11]. Unsupervised failure detection method uses an Ensemble of Bayesian models and supervised learning method uses decision tree classifiers to predict failure occurrences in the cloud.

The former approach uses Bayesian method to predict failure dynamics in cloud environments. It works in an un-supervised learning manner and deals with unlabelled data set. Experimental results show that the method can achieve high true positive rate and low false positive rate for failure prediction. Reactive failure management technique such as check-pointing and redundant execution can be exploited to avoid mis-predictions. The integration of proactive and reactive failure management is kept as the future work to enhance cloud dependability.

Song fu et al [10], proposed a frame work for autonomic failure management with hierarchical failure predictor functionality for coalition clusters. It analyses node, cluster and system wide failure behavior and forecast the failure dynamics. The frame work consists of *event sensor, failure predictor, failure manager*. Also failure correlations are expressed through temporal (spherical covariance model) and spatial (stochastic model) failure correlations.

A failure prediction framework for autonomic management of networked computer system was also proposed by Song Fu et al [10]. The prediction framework includes five time-series algorithms: MEAN, LAST, AVE (n), AR (m), and ANN (n). It also has 3 input neurons to receive temporal spatial correlation of data and failure measures. Weka machine learning software is used for implementation.

Derek smith et al [12], proposed an autonomic mechanism for anomaly detection. Data transformation is applied to format health related data. Clustering and outlier detection are explored to identify nodes with anomalous behavior. This method detects faulty nodes with high accuracy and low computational overhead. Naive Bayes and Decision tree analysis are performed against the dataset to determine the discriminating factor.

Song Fu [13], proposed a failure aware node selection strategy for construction and reconfiguration of reconfigurable distributed virtual machine (RDVM).Random select, First-fit, Optimistic Best-fit, and Pessimistic Best-fit strategies are used to evaluate the performance in improving the productivity of clusters. An experimental result shows enhanced system performance by using Best-fit strategies.

Mao-Lun Chiang in [14] , describes about the FCA (Fast Cloud Agreement) protocol which is proposed to enhance the reliability of the cloud computing environment. The FCA can reach an agreement and detect faulty processors by using a minimum number of messages simultaneously and efficiently. The maximum number of faulty processors can be detected by the FCA in a cloud computing environment. One of the important issues surrounding fault-tolerance is the Byzantine Agreement (BA) problem. It requires a set of healthy processors to reach an agreement, even if some components are faulty.

To provide a high performance, resource-friendly and scalable Checkpoint-Restart mechanism BlobCR (BlobSeer-based Checkpoint-Restart), a dedicated checkpoint repository that is able to take live incremental snapshots of the whole disk attached to the Virtual Machine (VM) instances was proposed. The approach supports both application-level and process-level check pointing and includes the unique ability to implicitly roll back file system changes [15].

V. VIRTUALIZATION

The increasing availability of VM technologies has enabled the creation of customized environments on top of physical infrastructures. The use of VMs helps in (i) server consolidation (ii) the ability to create VMs to run legacy code without interfering in other applications' APIs; (iii) improved security so as to enhance reliability;(iv) dynamic provision of VMs to services, allowing resources to be allocated to applications on the fly; and (v) performance isolation, thus allowing a provider to offer better quality of service to customers' applications. Virtualization technology minimizes some security concerns inherent to the sharing of resources among multiple computing sites [21]. Cloud computing and Virtualization are technologies that were developed to maximize the use of computing resources. Virtualization helps reduce complexity by reducing the number of physical hosts but it still involves purchasing servers and software and maintaining infrastructure. The greatest benefit is reducing the cost of infrastructure for companies by maximizing the usage of the physical resources.

Resource provisioning leads to increased reliability of the system. Resource failures leads to performance degradation, premature termination, data corruption, data loss, violation of SLA, loss of customer and revenue[16].Virtualization is the pillar for multiplexing computing resources. Virtual Machine Monitor (VMM), resides on a physical hardware and provides computing resource for VMs of different cloud user. Malware detection is difficult because all VM's are in the same cloud. If one VM is infected the malware may extend to other benign VM [17].

Roberto Di Pietro et al, [18] proposed an architecture that can trace, analyze and control live virtual machine activity as well as intervened code and data modifications possibly due to either malicious attacks or software faults. Virtualization maximizes the use of physical resources.

Resource failure is handled by two approaches

- Knowledge Based approach and
- Knowledge Free approach.

Knowledge Based approach focuses on statistical model and later focuses on mathematical model. *Knowledge-Free* approach does not need any statistical information about the failure model. Dynamic provisioning of virtualized resources is possible in cloud environments. These resources are delivered through VM. Resource allocation is an important key issue affecting the performance in cloud computing environment. Hongli et al, [19] proposed a novel method that uses Particle Swarm Optimization (PSO) based strategy to schedule applications. PSO is self adaptive global search based optimization technique. It is a Meta heuristic method. The existing algorithms like Best Resource Selection (BRS) and Genetic Algorithm (GA) could not give best results when compared to PSO. It is found that the result obtained by PSO is 3 times cost saving when compared to BRS [20].

VI. ISSUES AND CHALLENGES

Cloud computing is a new technology which is built on the success of previous developments like internet and others. Cloud computing has shown a quick change in the progress of technology. Companies can expand their computing resource with a highly scalable environment by utilizing resource virtualization. This paper exposes the various research issues and challenges related to failure management in cloud environments. It also addresses the importance of virtualization and resource provisioning. Some of the issues identified are

- The failure prediction models are based on statistical learning methods (Knowledge Based).
- To propose a failure prediction model for cloud environments
- Statistical learning method helps in failure diagnosis.
- The integration of proactive and reactive failure management helps to enhance cloud dependability.
- Effective resource utilization helps to achieve reliable cloud environment.
- Optimization Techniques like PSO can be used to achieve good resource utilization.
- Existing optimization algorithms can be combined to give a hybrid model with better results.
- New evolutionary algorithms to be proposed for resource scheduling.

VII. CONCLUSION

Cloud computing has its own merits and demerits. Generally services are managed by the provider. Among the three basic services IaaS provides higher flexibility when compared to PaaS and SaaS. The user need to pay for the resource utilized. Without installation the applications can be used. Some of the demerits which are experienced in cloud environments are its expensiveness and inability of user to have control over the remote servers and their software. Also it is very difficult to migrate large amount of data from provider.

This paper gives a brief survey on failure management and reliability enhancement through resource provisioning in cloud environments. The study shows that failure instances directly affect the systems reliability. Thus a failure free environment is to be achieved. By maximizing the fault tolerance and minimizing the failure incident increased reliability is possible. Efficient utilization of resource helps to achieve high performance. Thus the study illustrates that effective resource utilization can be achieved with the help of optimization technique like PSO. Also evolutionary algorithms can be proposed to optimize resource scheduling

REFERENCES

[1] Wenhao Li, Yun Yang, Jinjun Chen, Dong Yuan “A Cost-Effective Mechanism for Cloud Data Reliability Management Based on Proactive Replica Checking” IEEE/ACM *proceedings Cluster, Cloud and Grid Computing* (ccgrid), 2012.

[2] Ifeanyi P.Egwutuoha, Shinping Chen, David Levy, Bran Selic “A Fault Tolerance Framework for High Performance Computing in Cloud” IEEE/ACM proceedings on Cluster, Cloud and Grid Computing 2012.

[3] Wenhao Li,Yun Yang,DongYuan, “A Novel Cost-effective dynamic data replication stratergy for reliability in cloud data centres”, IEEE International conference on dependable, autonomic and secure computing, 2011.

[4] Bahman Javadi, Jemal Abawajy, Rajkumar Buyya, “Failure-aware resource provisioning for hybrid Cloud infrastructure”, Elsevier, Journal of Parallel and Distributed Computing, Vol.72, Issue 10, October 2012, pp 1318–1331.

[5] Alain Tchana, Laurent Broto, Daniel Hagimont “Approaches to Cloud Computing Fault Tolerance”, IEEE International conference proceedings,2012, pp.1-6.

[6] Cloud Computing by George Reese

[7] http:// thecloudtutorial.com-15.10.2013

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 10, October 2013

[8] XiaoJun Chen , Jing Zhang , Junhuai Li , Xiang Li "Resource virtualization methodology for on-demand allocation in cloud computing systems" Springer- service oriented computing and application, Vol.7, Issue 2, June 2013, pp. 77-100.

[9] Z. Zhang and S. Fu. "Failure prediction for autonomic management of networked computer systems with availability assurance", IEEE Workshop on De-pendable Parallel, Distributed and Network-Centric Systems in conjunction with IEEE International Parallel and Distributed Processing Symposium (IPDPS), IEEE, 2010.

[10] Ziming Zhang and Song Fu, "Proactive Failure Management for High Availability Computing in Computer Clusters", Third International Joint Conference on Computational Science and Optimization, IEEE 2010, pp. 377-381.

[11] Qiang Guan, Ziming Zhang , and Song Fu, "Ensemble of Bayesian Predictors for Autonomic Failure Management in Cloud Computing", Computer Communications and Networks (ICCCN), International Conference on Computer Communication and Networking, 2011,pp.1-6.

[12] Qiang Guan, Derek Smith, and Song Fu, "Anomaly Detection in Large-Scale Coalition Cluster for Dependability Assurance", IEEE International Computer Software and Applications Conference, COMPSAC Workshops 2010, Seoul, Korea, 19-23 July 2010.

[13] Song Fu "Failure-Aware Construction and Reconfiguration of Distributed Virtual Machines for High Availability Computing" IEEE International proceedings, 2009 pp. 372-379.

[14] Mao-Lun Chiang, "Efficient Diagnosis Protocol to Enhance the Reliability of a Cloud Computing Environment", Springer Science Business Media, vol.20, 2012, pp. 579-600.

[15] Bogdan Nicolae, Franck Cappello, "BlobCR Virtual disk based checkpoint-restart for HPC application on IaaS Clouds", Elsevier Inc, Vol. 73, Issue 5, 2013, pp. 698–711.

[16] Deqing Zou, et al., "Design and implementation of a trusted monitoring framework for cloud platforms", Elsevier, Future generation computer systems, 2013.

[17] Roberto Di Pietroa, et al., "CloRExPa: Cloud resilience via execution path analysis", Elsevier, Future generation computer systems, 2012.

[18] Hongli Zhang, et al.,"A PSO-Based Hierarchical Resource Scheduling Strategy on Cloud Computing", Trustworthy Computing and Services Communications in Computer and Information Science, Vol.320, (2013) pp.325-332.

[19] Suraj Pandey, et al., "A Particle Swarm Optimization –based heuristic for scheduling workflow applications in cloud computing environments", International conference on Advanced Information Networking and Application(AINA) –April 2010.

[20] Rajkumar Buyya, et al., "A Cost-benefit analysis of using cloud computing to extend the capacity of clusters", Springer Science Business Media, Cluster Computer 2010, pp. 335–347.

[21] Paul Manue, "A trust model of cloud computing based on Quality of Service", Springer, Annuals of operations research, vol.204, 2013.

[22] Peter Mell and Timothy Grance, "NIST -National Institute of Standards and Technology", Special Publication 800-145, September 2011.

BIOGRAPHIES

Jemina Priyadarsini. R is doing her research in the Department of Computer Science, St.Joseph's College (Autonomous), Tiruchirappalli, Tamil Nadu, India. She has 12years of teaching experience. She has attended many International and National Conferences, and also various Seminar and Workshops. Her broad area of research is 'Cloud Computing'. She is presently working on Reliability Issues in Cloud Computing. Her area of interest is Software Engineering, Data Structures and Algorithms, Unified Modeling Language, Computer Graphics and Mobile Computing.

Dr. Arockiam. L is working as Associate Professor in the Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli, Tamil Nadu-India. He has 24 years of experience in teaching and 17 years of experience in research. He has published more than 150 research articles in the International / National Conferences and Journals. He has also presented 2 research articles in the Software Measurement European Forum in Rome. He has chaired many technical sessions and delivered invited talks in National and International Conferences. He has authored a book on "Success through Soft Skills" and C#.net. His research interests are: Software Measurement, Cognitive Aspects in Programming, Data Mining, Mobile Networks Cloud Computing and Big Data Analytics. He has been awarded "Best Research Publications in Science" for 2010, 2011, & 2012, "Best Teacher Award" for 2012-13 and ASDF Global Awards for "Best Academic Researcher" from ASDF, Pondicherry for the academic year 2012-13.