

DATA MINING IN CLOUD COMPUTING (A PROACTIVE ERROR DETECTION PROGRAM IN CLOUD COMPUTING TO IDENTIFY THE ERRORS)

¹K. Arun Raj Bapuji, ²Prof K. Mruthyunjaya Rao
^{1,2} Acharya Nagarjuna University, Guntur

Abstract: The huge income of data into a data center's servers through intranet and internet cloud requests, cloud computing which is rapidly developing very efficient technology in the present era. Cloud computing on demand services of multiple process of virtual server's. Cloud computing uses & handles the concepts of big data and data mining area's for pure deployment and services of it. Arises of error detection that occurs in the data centers at the time of updating of existing servers. Here, we are providing the error detection techniques to identify the problems statements at the level of servers through the EDP program. The responsibility that is given to the program that which executes on the different types of the servers and notify the updating problems and reboot issues of the servers, the errors will be stored in an error data base server and try to resolve, if not report to the related team before server maintenance slot.

Keywords: CloudComputing, Fault Tolerance, virtual, EDP, Data Mining, Patching, updating, IaaS, proactive.

I. Introduction

Cloud Computing is a general term for the delivery of hosted services over the internet. Cloud computing enables companies to consume a computer resource, such as a virtual machine, storage or an application, as a utility- just like electricity- rather than having to build and maintain computing infrastructure in house.

Cloud computing is the delivery of computing services – Servers, Storages, databases, Networking, Software's, analytics and many more – Over the internet [“the Cloud “]. Companies offering these computing services are called cloud providers and typically charges for cloud computing services based on usage, similar to how you are billed for water [or] electricity at home.[16][17]

A. Uses Of Cloud Computing

You are probably using cloud computing here now, even if you don't realize it. If you use an online service to send email, edit documents, watch movies [or] TV, listen to movies, play games etc..., it is likely that cloud computing is making it all possible behind the scenes. The first cloud computing services are barely a decade old, but already a variety of organizations – from tiny startups to global corporations, government agencies to non-profits – are embracing the technology for all sorts of reasons. Here are few of the things you can do with the cloud.[16]

1. Creation of new apps and servers.
2. Store, backups and recover data.
3. Host website and blogs.
4. Analyses data for patterns and make predictions

B. Top Benefits Of Cloud Computing

Reasons for usage of cloud computing services:-

1. **Cost:** cloud computing eliminates the capital expense of buying hardware and software.
2. **Speed:** services are provides self service and on demand, resources can be provisioned in minutes.[16]
3. **Global scale:** Services are includes the ability elastically.
4. **Productivity:** On site data centers typically requires a lot of “stacking and racking”. Hardware set up, software patching and other timing – consuming IT management chores.[16]
5. **Performance:** The biggest cloud computing services run on a worldwide network of secure data centres, which are regularly upgraded to the latest generation of fast & efficient computing hardware.
6. **Reliability:** cloud computing make's data backup, disaster recovery and business continuity easier & less expensive, because data can be mirrored at multiple redundant sites on cloud provider's network.[16]

C. Cloud Computing Models

Cloud computing is based on service models. These are categorized into three basic service models which are – [18]

1. **Infrastructure-as-a-Service [IaaS]:** IaaS provides access to fundamental resources such as physical machines, virtual machines, virtual storages, etc.....

2. **Platform-as-a-Service [Paas]:** Paas provides the runtime environment for applications, development tools, etc....
3. **Software-as-a-Service[Saas]:** Saas model allow to use software applications as a service to end users.

D. Deployment Models In Use

Deployment models says the type of access mode to the cloud, Cloud have the four types of access:

1. **Public:** Allows systems and services to be easily accessible to the general public. Public cloud may be less secure because of its openness.
2. **Private:** Allows systems and services to be accessible within an organization. It is more secured because of its private nature.
3. **Hybrid:** It a mixture of public and private cloud, in which the critical activities are performed using private cloud while the non-critical activities are performed using public cloud
4. **Community:** Allows systems and services to be accessible by a group of organizations.

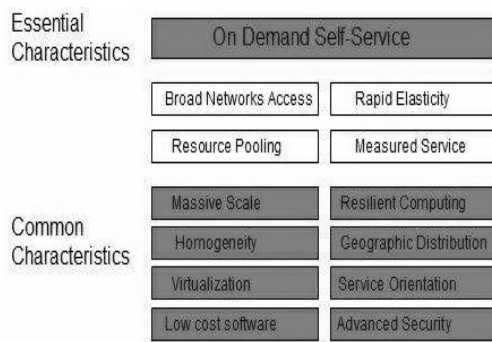


Figure: 1:Cloud Computing characteristics.

II. Data Mining

Data mining is defined as the process of withdrawal information from vast sets of data. In other words the practice of examine huge pre-existing databases in order to generate new information. It a process of sorting through large sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allows enterprises to predict future trends.



Figure 2:KnowledgeDiscovery Process.

The knowledge extracted so can be used for any of the following applications –

1. Market Analysis
2. Fraud Detection
3. Customer Retention
4. Production Control
5. Science Exploration

A. Data Mining Applications:

Data mining is highly useful in the following domains:

- Market Analysis and Management
- Corporate Analysis & Risk Management
- Fraud Detection

Apart from these, data mining can also be used in the areas of production control, customer retention, science exploration, sports, astrology, and Internet Web Surf-Aid

III. Fault TolarenceIn Cloud Computing

There are various faults which can occur in cloud computing which are very critical in real time which may lead to SLA (Service level agreement). Some of them discussed below.

- **Network fault:** A Fault occur in a network due to network partition, Packet Loss, Packet corruption, destination failure, link failure etc.
- **Physical faults:** This Fault can occur in hardware like fault in CPUs, Fault in memory, Fault in storage, etc.
- **Media faults:** Fault occurs due to media head crashes.
- **Processor faults:** fault occurs in processor due to operating system crashes, etc.
- **Process faults:** A fault which occurs due to shortage of resource, software bugs, etc.
- **Service expiry fault:** The service time of a resource may expire while application is using it.

Also, a fault can be categorized on the basis of computing resources and time. A failure occurs during computation on system resources can be classified as: omission failure, timing failure, response failure, and crash failure. The Fault may be

1. **Permanent:** These failures occur by accidentally cutting a wire, power breakdowns and so on. It is easy to reproduce these failures. These failures can cause major disruptions and some part of the system may not be functioning as desired.

2. **Intermittent:** These are the failures appears occasionally Mostly these failures are ignored while testing the system and only appear when the system goes into operation. Therefore, it is hard to predict the extent of damage these failures can bring to the system
3. **Transient:** These failures are caused by some inherent fault in the system. However, these failures are corrected by retrying roll back the system to previous state such as restarting software or resending a message. These failures are very common in computer systems.
- 6) **Self-healing:** In this technique the job will be assigned and receive the errors from the existing machines from the several parts of data centers in the cloud. This type of technique is used for a better performance. When the application has been running on a different machines and got failed of all machines. In that scenario every machine will be handle automatically in the cloud for updation or reboot.
- 7) **Safety bug checks:** In this technique the commands will be blocked when the commands are not meeting the desired safety properties in the cloud.

IV. Techniques Of A Fault Tolerance

The different types of fault tolerance techniques can be used at either task level or workflow level to solve the problems. In cloud computing two different types faults can be occurred.

- **Reactive Fault Tolerance:**

There are various types of techniques provided in error detection shown below:

- 1) **Task re-submission:** when any job has been failed du to any reason, in that scenario the same task job assigned to the same existing server in the same manner and different resources used to completing the job assigned.
- 2) **Check point restart [21]:** It is a technique mostly used for updating for the server which is not updated for a long time and for huge data base servers. This type of technique is when a job assigned to update of a server then the task has been failed at any level of execution the same job will be assigned to the same server from the starting point or from the stopped execution point. This technique is a better one for a huge database servers and long running servers and also for the servers which are used to run different applications.
- 3) **Replication:** The replica of various jobs will be assigned and run on different resources and have to wait until the replica's task to complete without any failure and for the desired output. This type of technique implemented in HA-Proxy, Amazon, Hadoop tools.
- 4) **Job migration:** Whenever a job has been assigned to the a client server during the execution, that job has failed due to various reasons. In this case the job can be assigned to any of the connected existing cloud machines in multiple data centers.
- 5) **S-Guard [2]:** It is a less turbulent to normal stream processing and it can create more resources. It is based on roll back recovery and can be implemented in Hadoop and Amazon Ec2.

- **Proactive Fault Tolerance:**

It will identify the errors and faults before it comes [or] suspected to come. There are different types of techniques used in this detection of errors as shown below:

- 1) **Software rejuvenation:** It is a technique designed for periodical reboot and updates of a server. With this technique the server will start in clean state.
- 2) **Proactive fault tolerance – self healing:** This technique is used in an application is running in multiple virtual servers at a time, when any failure occur then the technique will automatically handle in resolving the problem.
- 3) **Proactive fault tolerance using premature migration [12]:** In this technique migration relies on a feedback-loop control mechanism and application will be continuously monitored for a better performance and saving time.

Based on proactive error detection technique we have found some errors in software **rejuvenation**, we found below errors when we reboot/update the server. We have “n” of issues and few of them discussed below:

- Server may not reboot for a long time more than 1]] days.
- Duplicate package issues
- Compatibility package issue
- Common kernel errors
- Application level errors
- Incompatibility of application provided by library and gcc.lib files and other packages and files in kernel level.
- Hardware level [memory, processor, etc.,]
- Kernel crash

- Service level errors
- Bug fixing

When will reboot the server?

In proactive fault tolerance to get better performance and security to fixing of bugs the existing server need to be updated every quarterly and depends on requirement of new version of applications and request of client.

This reboot make causes issue before without checking any status of existing server, in this case the server may not come-up.

V.Existing System

The existing cloud provides Fault Technique mechanism for all the physical and virtual servers. This technology will address the errors after or when it completed [example server may not respond in rebooting file system issues, application related issues etc]. This cause’s server’s long time down without maintenance in certain cases as prescribed.

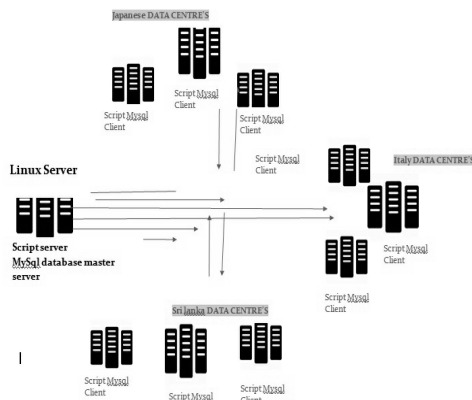


Figure 3

Error Detection:

Error detection in the cloud computing via **data mining**. Error detection is one of the major concept in cloud computing. When we think about cloud computing it deals with “n” members of servers means hundreds or thousands of servers connected to master server.

Error detection is classified into two parts one is proactive error detection and [2] reactive error detection. Proactive error detection tries to avoid errors and failures, by predicting them proactively suspected components will be detected and detected the problem before it actually comes [7]. Reactive error detection is one of the techniques which will removes errors [or] faults after it occurs while the updation process. This technique will reduces the number of errors efficiently when it executed the error detection program on the client servers.

There are various kinds of errors can be occurred when we working with cloud environment. Below some error factors are shown:

- 1) **Kernel related errors:** After updation / patching of a server can be crashed due to application, library [or] gcc++ libs may not compatible to the existing server.
- 2) **Duplicate packages:** The proposed application was installed previously in the client server, the same application try to install in the same servers in cloud then replication will be occurred and the task cannot be executed.
- 3) **Package conflict errors:** In the cloud one package was installed long back and an advanced package of the same application is trying to install in client server, then the library files and others system files are not allow the advance version to be installed until unless the previous application un-installed.
- 4) **Process errors:** An error which occur due to some reasons like software bugs.

VI. Related Work Of Methodology

There are different types of error detection methods and models are available in cloud completing.

A failure may happen due to occurrence of error, that is due to reaching a invalid system state. The hypothesized cause for an error is a fault which represents a fundamental impairment in the system. The motion of faults errors and failures can be represented using the following chain [S.2]]4, HH.1997]

- 1] **AFTRC [Adaptive fault tolerance model]:** It is purely used in Real time cloud computing. In this model the errors proactively and makes decisions on the basis of the reliability of the processing nodes.
- 2] **Candy [15]:** Candy is a component based availability model. It is based on the high availability assurance of cloud service is one of the main characteristic of cloud service and also one of the main critical and challenging issue for cloud.
- 3] **FTM [14]:** It is a model to overcome the limitation of existing methodologies and achieve the reliability and flexibility. An inventive is proposed perspective on creating and managing errors. By this particular methodology user can specify and apply the desire level of error tolerance. FTM architecture can primarily be viewed as an assemblage of several web services components, each with a specific functionality.
- 4] **LLFT [13]:** A propose model which contains a low latency error tolerance middleware for providing error tolerance for distributed applications deployed within the cloud computing environment. This middleware replicates application by the use of semi-active replication or semi-

passive replication process to protect the application against various types of errors.

5] **FTWS [19]:** It is an error/fault tolerant work flow scheduling algorithm for providing error tolerance by using replication and resubmission of task based on based on the priority of the task. This model is based on the fact that workflow is a set of tasks processed in some order based on data and control dependency. Scheduling the workflow along with the task failure consideration in a cloud environment is very challenging. FTWS schedule and replicates the tasks to meet the deadline.

VII. Proposed Model [An Error Detection Programme]

This project aims to be focus on proactive technique in cloud environment. Proactively identifying the errors on physical or virtual servers in cloud environment and rectifying them. In case of errors not rectified, the errors message will be passed to configuration management tools.

Software rejuvenation is a mechanism which will handle the error before it fails, when error occurs this mechanism will proactively resolves the error, if not we need to manually resolve it by sending to configuration management tool.

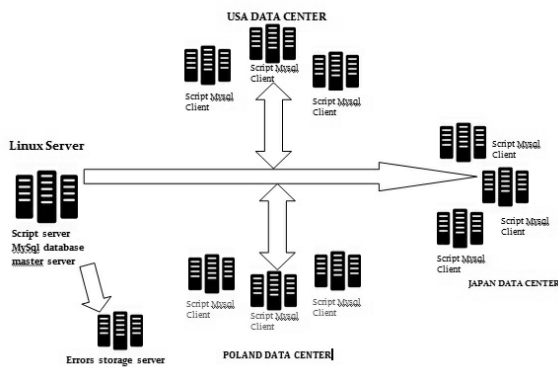


Figure 4

The proposed program mechanism will proactively identify the error and stores in a separate **data base for each and every server**. Before proposed to reboot/update a server in cloud, in the proactive mechanism the **EDP** will check the existing active server status and errors, when it found any errors, the errors will be stored in a **database server**. Even it will solve those errors, if not it will send the error to configuration management tool. Where the errors ware solved then these servers ready for reboot/updation. To store the error data the encryption and decryption method will be used for privacy and security depends on the existing environment.

As per the software rejuvenation techniques, a server can be rebooted by checking suspected errors. But after reboot the server it may not come up [or] not?

Our proposed program model help to reboot the server and make it online 100% before updating/patching or rebooting.

This EDP program will easily interacts with kernel and log messages and various types of file systems and also packages, it identifies the status of server, it can be rebooted or not and specific information about the reboot of server. The EDP program completely desired in Ruby and shall program it can easily interact with the kernel and machine.

The EDP program can easily avoid fault tolerance and errors and it helps to validate the service of critical importance to ensure the SLA’s are properly adhere. The program will be executed depends on the server.

Availability of the server = Uptime server

$$\text{Uptime server} + \text{Server down time}$$

This EDP program will run on each and every server in the cloud environment while running the program, the program will find out different kinds of errors and those errors will be sent to the **database client and the client will try to rectify all the errors which are stored in the database server**. The proposed EDP program will try to resolve the errors while running, if not it will sent to the configuration tool to get resolved.

If errors are not found, then the program was executed the client server will generate a message as “the existing server is ready to update”. This message will be sent to **database server**. The EDP Program helps in future updating of a server to overcome with the same error while update the servers or reboot. This EDP will help the admins in forecasting of repeated error in future and they can easily imagine and with the knowledge of the **error database** they will easily resolve any type of errors by using this program.

Resolved:

Fault / error →detected→ resolved

Not Resolved :

Fault / error→ detected → configuration management tool.

VIII. Conclusion And Future Work

In Cloud computing fault Tolerance techniques are very familiar to the real world. In this paper we present a solution for the proactive fault tolerance before occurrence of the failure to the data centers. Especially this research provides the clear information of the servers before patching whether servers up’s or not .However, here in this

paper we define the solution for proactive fault tolerance and future work will be extends the reactive fault tolerance.

References

- [1]. ImadM.Abbadi “self-management services, conceptual model in trustworthy clouds infrastructures”, 2]11].
- [2]. Vikas Kumar, Shiva Prakash “A cloud balances based cloud computing techniques and challenges” IJSRM, volume, Issue 5, May 2]14.
- [3]. SonalRana, Aditi Sharma, AmandeepKapoor, “Fault tolerance in cloud computing by efficient load balance mechanism based on Ant. Colony”, IJARCSSE, volume 5, issue 1, January 2]15.
- [4]. Seyyed Mansur Hosseini and MostafaGhobaeiArani, “Fault tolerance technique in cloud storage: A Survey”. International Journal of Database Theory and Application volume 8, No.4 [2]15], pp. 183-19].
- [5]. N.Chandrakala and Dr.P.Sivaprakashan “Analysis of Fault Tolerance approaches in dynamic cloud computing” international journal of advanced research in computer science and software engineering volume 3, issue 2, February 2]13.
- [6]. Anjali D.Meshram, A.S.Sambare and S.D.Zade “Fault Tolerance Model for Reliable Cloud Completing”, international journal on Recent and Innovation Trends in Computing and Communication,ISSN 2321-8169, volume 1, issue 7, July 2]13.
- [7]. Prasenjit Kumar Patra, Harshpreet Singh &Gurpreet Singh, “Fault Tolerance Techniques and Comparative Implementation in Cloud computing” International Journal of Computer Applications. Vol.64, No.14, February 2]13.
- [8]. Sweta Patel, Ajaya Kumar Singh. “Fault Tolerance Mechanisms and its Implementation in Cloud Computing –A Review”. IJARCSSE, Volume 3, Issue 12, December-2]13.
- [9]. Review on Fault Tolerance Techniques in cloud computing. IJCA Volume 116-NO.18,April 2]15 Efficient fault Tolerant Strategies in Cloud Computing Systems.
- [10]. Efficient Fault Tolerant Strategies in cloud computing System, IARJSET, Volume 2, issues 2, February 2]15
- [11]. Review on Fault Tolerance Techniques in cloud computing, IJCA Volume 116,no 18, April 2]15.
- [12]. Novel steps of authentication Technique for privacy presentation in cloud computing infrastructure
- [13]. A Survey of Gloud computing Fault Tolerance:Techniques and Implementation, IJCA Volume 138, March 2]16
- [14]. Effective Fault Tolerance Management in cloud computing, IJETAE, volume 6, Issuse 3, March 2]16.
- [15]. Review on Fault Tolerance Techniques in Cloud Computing, IJEMR, Volume7, Issuse3 , June 2]17.
- [16]. <https://connect2compute.wordpress.com/tutorials/what-is-cloud-computing/>
- [17]. <https://www.imagineiti.com/the-cloud/costs-cloud>
- [18]. <http://www.x-centric.com/solutions/cloud-datacenter/>