



# ARCHITECTURAL MODELS FOR FAULT TOLERANCE WITHIN CLOUDS AT INFRASTRUCTURE LEVEL

J. K. R. Sastry<sup>1</sup>, K. Sai Abhigna<sup>1</sup>, R. Samuel<sup>1</sup> and D. B. K. Kamesh<sup>2</sup>

<sup>1</sup>Department of ECM, KL University, Vaddeswaram, Guntur District, Andhra Pradesh, India

<sup>2</sup>Department of Computer Science Engineering, St. Martin Engineering College, Dupally, Hyderabad, India

E-Mail: [drsastri@kluniversity.in](mailto:drsastri@kluniversity.in)

## ABSTRACT

Cloud computing technologies are being used aggressively these days to enable use of shared resources. However the confidentiality and availability of the data stored on the cloud is still a serious problem. In a cloud, several faults do occur which adversely hamper the continuous availability of service to the end customer. Faults could be hardware, software or network related. Infrastructure installed on the clouds does get affected due to all kinds of faults. The infrastructure supported on the clouds must be made available to the clients even during the occurrence of the faults to provide continuous service. In this paper architectural models have been proposed using which the infrastructure related services are made available to the clients even during the occurrence of the faults making the entire process of cloud computing reliable and effective.

**Keywords:** cloud computing, fault tolerance, high availability of IaaS.

## 1. INTRODUCTION

The accessibility of information processing resources has become simple and less complex with the advent of internet and the recent introduction to cloud computing technologies. This most recent introduction lead to a critical change in IT environment through another new model known as "cloud computing", prevalently called as "Cloud". The cloud computing model moved calculation and information far from the desktops to gigantic servers which are found somewhere else. It leases assets to the clients for general utility as and when demanded by the clients.

The prime expectation of cloud computing is the better usage of dispersed assets, coordinate to accomplish extreme throughput and resolve expensive scale of computational inconveniences. The cloud computing technologies express trustworthy strength, adaptable foundation, speedy provisioning, versatility, solid element and green arrangement.

In cloud environment, the conventional roles of providers is classified into: *infrastructure providers* who maintain the platform and rent resources in accordance with usage-based pricing model, and *service providers* who host one or many infrastructure providers to rent resources to serve the end users.

In cloud, the users/pumps-in the data, applications, or any other services over the internet through browser irrespective of the device or location of the user. This is possible because of the infrastructure that is usually provided by the third-party organizations. These vendor parties give access and lease their services following the usual delivery models of cloud-private, public, hybrid or community. The Cloud providers deploy their services in accordance with the need of the end user. The vendor provides the computing services in three ways, namely, *Infrastructure-as-a-Service* (IaaS), *Platform-as-a-Service* (PaaS) and *Software-as-a-Service* (SaaS).

Cloud computing is a model for enabling present, convenient, on-demand network access to a shared pool of

configurable computing resources (Networks, Servers, Storage, Applications and Services). The cloud model consists of 5 essential characteristics, 3 service models and 4 deployments models. Whereas fault tolerance is a crucial key issue in cloud computing and it's connected with all the techniques. The main edges of implementing fault tolerance in cloud computing embrace failure, recovery, lower price etc.

The cloud computing technology is segmented into 3 layers that include hardware layer, virtualization layer (Infrastructure Layer) and an application layer. The physical resources of cloud computing are managed at the hardware layer that is often maintained at the data centres. The data centres consist of many physical servers that are situated in racks and connected with routers, switches and alternative parts. The virtualization layer or infrastructure layer partitions the physical resources into many virtual machines which can be assigned to several users. Virtual machines provide the kind of computing resources required by the user. The resources can be added dynamically to the virtual machines at run time

Platform layer which is on top of the virtualization layer provides the framework for deploying the applications into Virtual machines. The application layer which is in the top of hierarchy contains all the applications that the top user utilizes from the cloud. The cloud is adaptive. Each layer and also along with other layers services the user requests. Clouds are designed to support varied application demands while minimizing the management and maintenance expenses.

Infrastructure is the lowest layer and it provides processing, storage, networks and other resources over the network. It saves the memory or the files when the faults occur at infrastructure. Infrastructure as a Service (IaaS) is a means of delivering Cloud Computing resources which include servers, storage, and network, operation systems as and when demanded by the users. IaaS can be used to get the resources that are deployed on the WEB as shared resources (Publicly owned) or privately owned or a



combination of both of them generally known as Hybrid Cloud. The infrastructure desired by the end user is made available as a virtual machine.

Faults do occur within the clouds and when they happen will hamper adversely the services to be provided to the end user in reasonable time. A mistake within the software package or defect within the hardware could lead to disturbance in the way the cloud is expected to operate. It is necessary to recognize different faults occurring on the infrastructure and find ways and means of countering the same with the main aim of providing continued, uninterrupted service to the end customer. It is necessary to analyse cause and effects of infrastructure faults and find the ways to mitigate them. The programs which are part of the operating system are generally designed for recovering back to the normal operation in the case of occurrence of the faults. In no recovery is possible, the program dies and nothing of the ongoing session is stored leading to heavy disruption in providing the services to the end-user.

Faults can happen anywhere starting from physical servers to Virtual machines due to various reasons that include, Overflow, Lack of memory, VM failure, CPU faults, Loss of files, Disk storage, execution faults or Network faults. Several clients are dependent on cloud for various applications and software. The occurrence of any infrastructure fault may result in the containment of the general functioning of the client works. The clients send their requests to cloud that are directed to the Middleware which communicates between the end user and the service provider. An infrastructure fault may result into the termination or failure of the servers releasing all the current work status, session details and uncommitted work done, drawing the client to a great loss. Whenever a server fails, the middleware diverts the requests to the remaining servers. At the point where the demand reaches to a critical stage, the alternate servers may also breakdown unable to handle the requests. To avoid a situation wherein the client suffers from the infrastructure faults, a fault tolerant system should be designed to keep the infrastructure function normally even if one of its servers fails. In this paper several architectures have been presented that cater for fault tolerance at infrastructure Level.

## 2. LITERATURE SURVEY

Saikia *et al.* [1] studied the inclusion of Fault-tolerance in the cloud computing framework. They have provided a survey of different kinds of methods that can be used within cloud computing for making the entire services fault tolerant. They have not however deliberated on the accuracy and exactness with which a fault tolerant method could be implemented

To reduce the faults that occur on the devices, they must be foreseen and are to be managed such that the system works efficiently even when the failure occurs. Anju Bala *et al.* [2] presented fault tolerant techniques which actually dependent on system behaviour. They have presented Cloud Virtualized System architecture, which implements autonomic fault tolerance method. Their preliminary outcomes indicate that their suggested system

takes into account different software failures for server programs. To implement their technique, they have used tools such as AZURE, Hadoop, SHelp. To execute Linux based applications, Amazon EC2 presents a virtual environment.

AlianTchana *et al.*, [3] have explained that cloud technologies use layered approach, actually divided into three layers. The three layers fit servers, digital system and programs. The three layered approach makes the cloud computing platform complex especially when fault tolerance has to be included in the cloud computing system. They have proposed online repair mechanism when faults happen. They have discussed the implications of implementing fault tolerance.

Rejinapaul, *et al.*, [4] have presented the way a real-time system can be implemented using cloud computing infrastructure. A real-time system can be designed to take advantage of huge computing power and scalable virtualized environment. Real-time systems that are implemented using cloud computing infrastructure, suffers from various errors especially related to management of latency of processing which may not match the real-time requirements. Most of the real-time systems are safety critical systems and cannot lose out on latency times. A smart architecture has been presented based on checkpoint infrastructure for virtualized server model with built-in fault tolerance model for achieving real-time computing. Hadoop file management system has been used for storing the check points. These check points are replicated and stored on all the servers on the cloud making it possible to restart executing a program that got crashed due to a fault. They have presented that storing the checkpoints in a file systems supported by HADOOP is faster with low interference on task execution and efficient task recovery after the failure. The method proposed by them uses effectively the dynamic resource provision supported by Cloud computing infrastructure.

Cloud computing architecture is complex as the computing is supported at three different levels of abstractions which include infrastructure, platforms and applications. Jasbir Kaur *et al.*, [5] have shown how faulttolerance can be implemented through focusing on FCFS, SJF along with MPIL. The algorithm proposed by them implements reactive fault tolerance among the servers. The faulty server tasks are relocated to a new server having minimum load at the instance of occurrence of the fault. They have made a comparison with MPI and MPIL.

One of the major challenges of cloud computing is reliability and availability of the contracted services and applications. Rajesh *et al.*, [6] have proposed a model using which one can analyse how a system can tolerate a fault on the basis of assessment of reliability of a processing node which is a virtual machine. The reliability is assessed in terms of executing a task in time and correctly. The increase and decrease in reliability is assessed based on whether the task is executed in time and correctly. A node which is continuously failing is removed and a new node is added. If a virtual machine gains reliability less than the threshold value, then the same are



backward recovered and the VM is migrated to a new server. They have proposed a method based on designing diverse variants on multiple virtual machines. They have proposed both forward and backward recovery systems.

Cloud computing is dynamic. It is adaptable to the new technologies as the model is based on integrating the hardware and software dynamically. The cloud providers have to guarantee the quality of service and on-interest administration. Within clouds the integration of hardware and software is scalable. Faults do occur in any dynamic environment so is the case with clouds as well. Fault tolerance of a system is the ability to react gracefully to an unexpected malfunctioning of either hardware or software. If cloud computing must be robust and dependable, faults must be identified and handled when they occur. Zeeshan Amin *et al.*, [7] have presented various fault detection methods and architectural models using which one can achieve the fault tolerance. They have presented an algorithm based on Artificial Neural networks.

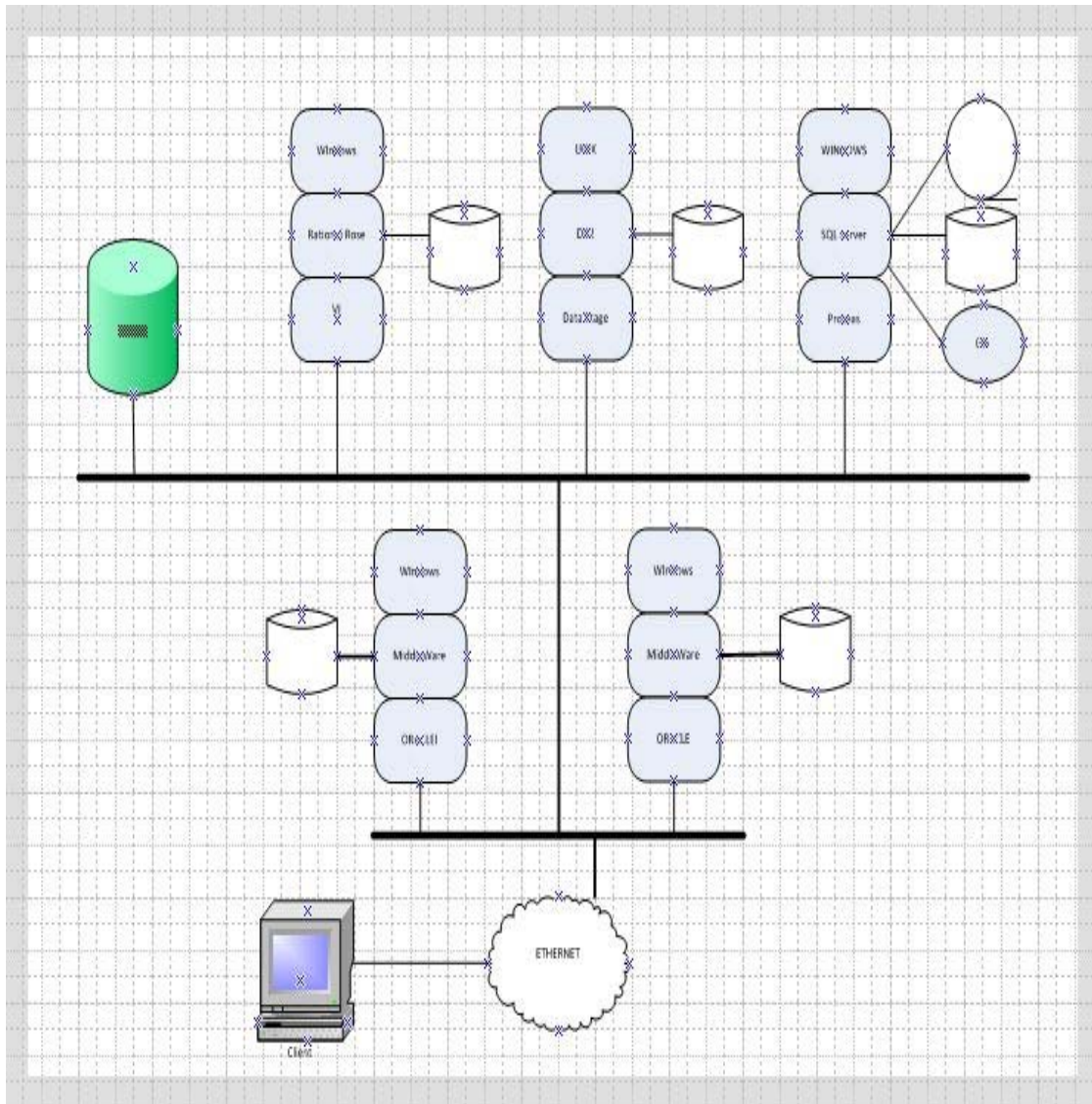
Seyyed Mansur Hosseini *et al.*, [8] have presented that every organization these days are considering one form or other of using the cloud computing as applicable to one's own business. Faults can occur within clouds. It is necessary to find and evaluate different types of faults that can occur within a cloud. They have presented the concept of fault tolerance in the context of cloud computing, some fault tolerant techniques are presented along with comparative analysis. From the

survey one can see that very few techniques have been proposed relating to infrastructure being fault free.

### 3. CLOUD COMPUTING ARCHITECTURE WITH INFRASTRUCTURE AS A SERVICE

The IaaS model of cloud computing provides the user with the backend services along with the infrastructure, through SaaS and PaaS support. The clients make use of all the services by requesting their demands to the cloud which has many servers working for rendering infrastructure as a service. To address all the requests to the specified server, keep track of all the requests and to keep up the small print of each session, all the client-server transactions are managed by a primary server that could be seen as middleware. The middleware accepts the requests from the client and assigns the request to the related servers, as well as keeps a log file which carries the details of the client-server connection and specifications of the server. Infrastructure includes a network, storage, and GPS which are connected to servers and the same or managed through processors resident on the server.

Generally many virtual machines are created on a physical server. Failure of a physical server shall lead to failure of all the virtual machines that are created on the server. Sometimes failure of Physical server or physical resources may disturb the operations of many virtual machines which are created through resources connected to several physical machines. The architectural model that supports infrastructure as a service is shown in Figure-1.



**Figure-1.** Architectural model - Infrastructure as a service within cloud.

As shown in the Figure-1, client sends a request through TCP/IP to middleware server for want of a service for a infrastructural requirement, such as storage, memory or an application software such as Rational rose, db2, sql server etc. The middleware creates a virtual machine of the infrastructural facilities required by the user and informs the client of the details of the virtual machine

#### 4. FAILURE ANALYSIS OF INFRASTRUCTURE AS SERVICE

Even though the architecture and components for delivering infrastructural services are employed wisely, the occurrence of infrastructure failures as discussed earlier may tend an instance or server to terminate or fail completely. If the number of client requests exceeds the handling capacity of the infrastructure server it may fail due to overflow, too much switching or CPU fault. The cause of an instance to fall apart maybe due to execution

fault, loss of files or any runtime errors. This drives the server to temporarily stop the execution of a client request until it redirects the request details and session details to any other server which can now act as the alternative to the other. However if any server processing a user request fails then the data that it is being processed by the server may be lost due to which reason it may not be possible to effect a migration to some other server. The feature of cloud computing to divert the traffic to an unloaded server as such will not happen due to loss of data. The recovery from a failure and restart of the service from where it failed is not possible due to lack of transaction data.

#### 5. AN EFFICIENT FAULT TOLERANCE METHOD FOR ENHANCING THE AVAILABILITY OF INFRASTRUCTURE AS A SERVICE

Fault tolerance concept which began to gain its importance since two decades, has at least seen its success





in imbibing this concept at the database servers. The probability that the underlying database server is corrupted is minimal. Nevertheless with striding technological advancements, the software faults do have seen high inclination. To save the software from losing its essential data, the concept of virtual memory is being used. Virtual memory is a feature of an operating system. Insufficiency of physical memory is managed through transferring files or data temporarily to a secondary storage generally through disk storage. If any server fails, all the client requests to the resources connected to the server have to be kept on hold and subsequently rolled back if the fault on the server cannot be serviced as fast as possible.

If a server is corrupted, all the client transactions can be diverted to other servers which will process the request afresh. The concept of Mutual Buffers

maximizes the availability of database system to the servers or cloud. Figure-2 shows the Virtual Memory which can be used as a shared resource. Disk storage can be used for storing the content existing in the primary memory. A strategy can be evolved using which flushing of the primary memory can be undertaken. At any time the state of processing can be stored and backed up on to the storage device. The transaction data as such can be stored in a Disk while simultaneously holding into the memory. In the situations involving failure of a server, the transaction data is still held in the storage device. The transaction processing as such can be diverted to a different server where the user request can be processed. This way of using the mutual buffers will help the transaction processing system fault tolerant.

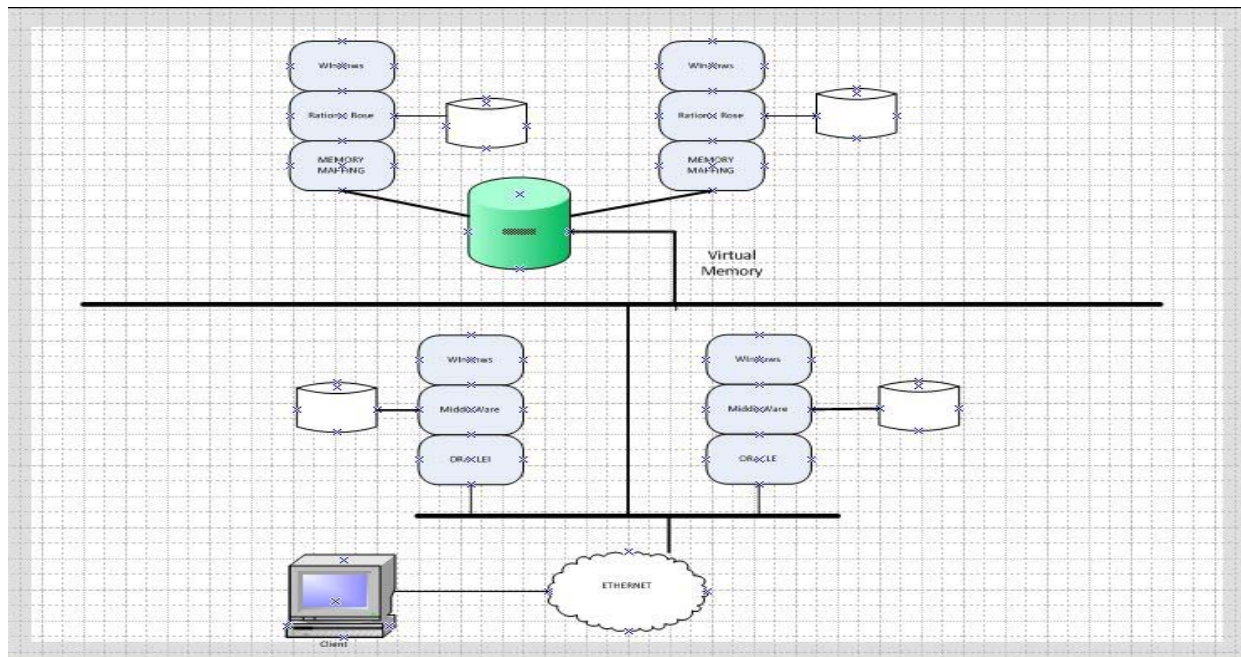


Figure-2. Virtual memory as infrastructural resources.

## 6. ADVANCING FAULT TOLERANCE WITHIN CLOUDS @ IAAS LAYER

While the technology is advancing every day, the faults too are advancing in their own way. So it is important that the fault tolerance system be taken to a next level. Recently, every virtual memory server has been maintaining a transaction log file which incrementally updates the sequence of changes made by the transaction in the virtual memory. In the previous method introduced by us, the client request is passed on to another server which processes the request afresh but in the present scenario, doing as such may mean a waste of time.

Mirroring data helps in removing the complicity of using stored data storage that holds the transaction data which in a way do not reflect the state of processing at which time a fault has occurred. Manteca of redo log files in different disk storage devices located on the same

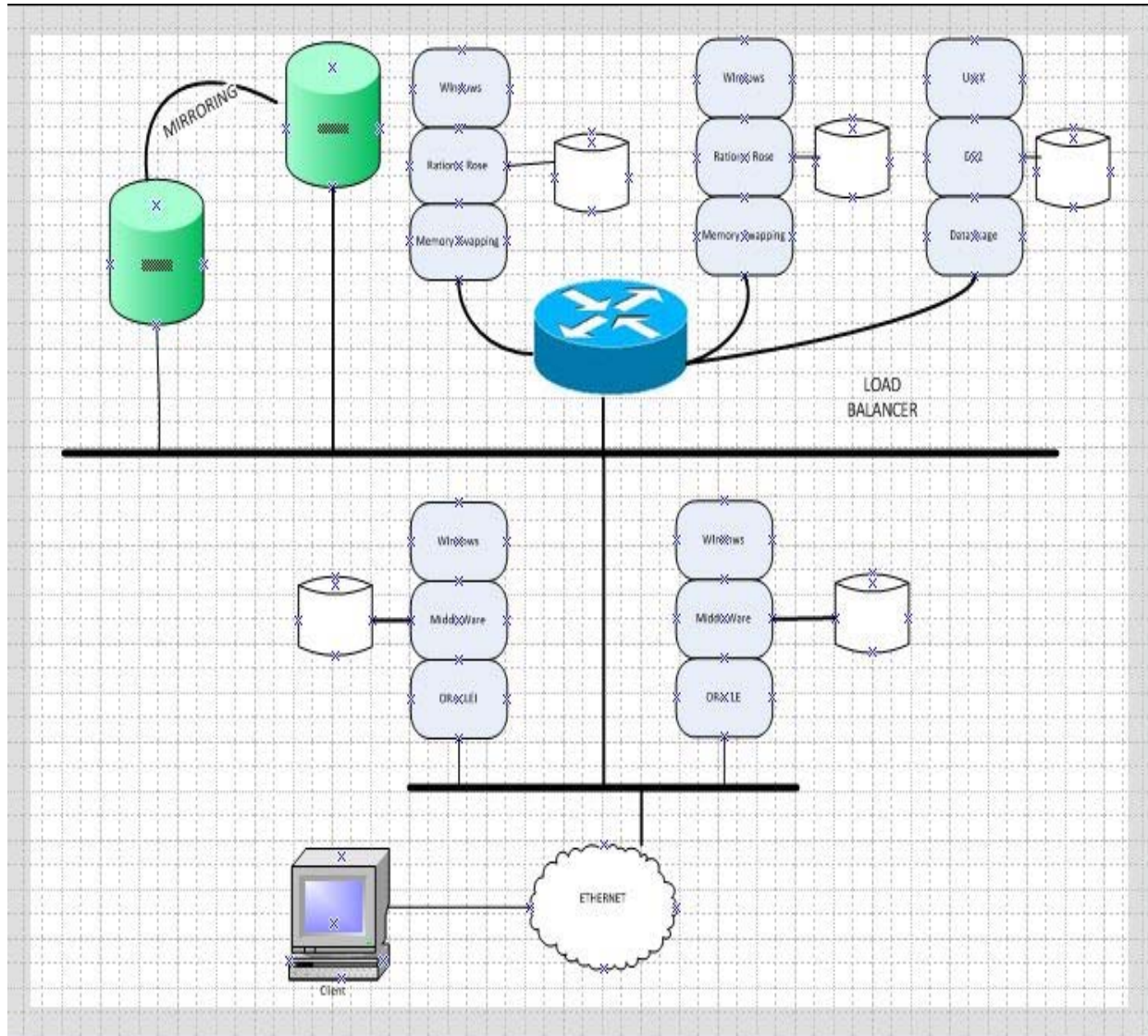
machine or on different machine makes available current state of processing of a transaction. Data mirroring can be undertaken thorough application specific development or can be achieved through invoking mirroring function supported within a database management system. Mirroring can also be achieved by configuring the process with an operating system. In cloud computing configuration of operating system can be undertaken through an application specific interface.

Mirroring helps to recover from failure of storage systems. If one disk fails, recovery from a secondary disk where mirror of the data is stored can be done. Fault tolerance of storage systems can also be achieved through using clustered parallel processing with the help of a load balancer. Redundant server systems can be used for processing the same transaction while one processing can be considered as primary, the second processing is



considered as secondary. Same transaction is processed through two servers. In the event one processor fails, the second processor still provides the service without any interruption. When the failed server is recovered, redo log

files are copied and the service is commenced by the primary server. The architecture of a fault tolerance system that uses disk mirroring and load balancing is shown in Figure-3.



**Figure-3.** Fault tolerances through disk mirroring and load balancing.

#### 4. COMPARATIVE ANALYSIS

Many authors have recommended different types of fault tolerance models. Table-1 shows the comparisons of the methods proposed by the authors. Many authors have banked on implementing the fault tolerance using single criteria. A single fault tolerance system will not cater for protecting all kinds of faults that might happen.

Faults might happen in relation to many of the infrastructural facilities that are supported on the cloud. None of the methods presented so far has considered the fault tolerance with respect to disk storage which is resident either on the server or on the network. Storage level fault tolerance can be effectively achieved through DISK mirroring and middleware services effectively.

**Table-1.**Comparison of fault tolerance models.

Parameters	Contributors to fault tolerance at infrastructure level					
	Proposed model	Anju Bala	Lakshmi Prasad	N. R Rejinapaul	RajeshS.	Ravi Jhawar
Mirroring	√	X	X	X	X	X
Middleware service	√	X	X	X	X	X
Virtualization	√	√	X	√	√	X
Multiple process communication	√	X	X	X	X	X
Check pointing	X	X	√	√	X	X
System level fault tolerance	X	X	X	X	X	√
Process level redundancy	√	X	X	X	√	X
File sharing	√	X	X	X	X	X

## 5. CONCLUSIONS

Cloud computing technologies are making radical changes in the way computing is undertaken in support of several levels and sizes of the business establishments. The cost of computing is being reduced quite drastically due to the use of cloud computing technologies. The clouding computing infrastructure while is leading to many great economies but still are suffering from different kinds of faults due to which the users are not being given with satisfactory services.

The services being provided for making available infrastructure as required by the user is most crucial for supporting cloud computing. Disk storage is most crucial part for which users are generally dependent on. The infrastructure made available on the cloud must be fault free and must be high available. Several methods and mechanisms must be inbuilt to cloud computing so as to make the resources fault tolerant so that the users can be given with satisfactory services. Disk Mirroring, Load balancing, mutual buffers are the some of the excellent initiatives for making the infrastructure as service most satisfactory with high throughput and response.

## REFERENCES

- [1] Lakshmi Prasad, Saikiyumnam Langlen Devi. 2008. Fault Tolerance Techniques and Algorithms in Cloud Computing. International Journal of Computer Science & Communication Networks. 4(1): 6278-6281.
- [2] Anju Bala, Inderveer Chana. 2012. Fault Tolerance-Challenges, Techniques and Implementation in Cloud Computing. IJCSI International Journal of Computer Science Issues. 9(1): 288-293.
- [3] AlianTchana, Laurent Broto, Daniel Hagimont. 2012. Fault Tolerant Approaches in Cloud Computing Infrastructures. The Eight International Conferences on Autonomic and Autonomous Systems. pp. 43-48.
- [4] N.RRejinapaul, Maria Michael Visuwasam. 2012. Checkpoint-based Intelligent Fault Tolerance for Cloud Service Providers. International Journal of Computers and Distributed Systems. 2(1): 59-65.
- [5] Jasbir Kaur, SupriyaKinger. 2014. Efficient Algorithm for Fault Tolerance in Cloud. International Journal of Computer Science and Information Technologies. 5(5): 6278-6281.
- [6] Rajesh.S, Kanniga Devi.R. 2014. Improving Fault Tolerance in Virtual Machine Based Cloud Infrastructure. International Journal of Innovative research in Science, Engineering and technology. 3(3): 2164-2168.
- [7] Zeeshan Amin, NishanSethi, Harpreet Singh. 2015. Review on Fault Tolerance Techniques in Cloud Computing. International Journal of Computer Applications. 16(18): 12-17.
- [8] Sayed Mansur Hosseini and Mostafa GhobaeiArani. 2015. Fault-Tolerance Techniques in Cloud Storage: A Survey. International Journal of Database Theory and Application. 8(40): 183-190.