1-29-2012

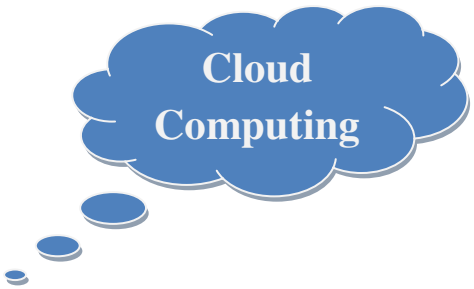
Proceedings of International Conference on Cloud Computing

Prof.Srikanta Patnaik Mentor
IRNet India, patnaik_srikanta@yahoo.co.in

Proceedings

Of

International Conference on



ICCC-2012

29th, January 2012.

Editor-in-Chief

Prof. (Dr.) Srikanta Patnaik
President, IRNet India and Chairman IIMT
Intersceince Campus, Bhubaneswar
Email: patnaik_srikanta@yahoo.co.in

Organized by:

IRNet
INTERSCIENCE RESEARCH NETWORK

About **ICCC-2012**

Cloud computing refers to computing with a pool of virtualized computer resources and is driven by economics of scale. A cloud can host a variety of different workloads, and allow workloads to be deployed and scaled-out quickly on-demand by rapid provisioning of virtual machines or physical machines. A cloud supports redundant, self-recovering, highly scalable programming models and allows workloads to recover from many unavoidable hardware/software failures. A cloud also monitors resource use in real time to enable rebalancing of allocations when needed. The idea is to move desktop computing to a service-oriented platform using server clusters and huge databases at datacenters. Cloud computing leverages its low cost and simplicity that benefits both users and the providers through providing cost-effective services and pay-per-use pricing model. In cloud computing, everything including software, platform, and infrastructure is as a service. Cloud computing makes data truly mobile and a user can simply access a chosen cloud with any internet-accessible device. Cloud computing overlaps some of the concepts of cluster, distributed, grid, service, ubiquitous, utility and virtual computing; however it has emerged from these computing domains and now has its own meaning. In cloud computing, a user does not care much what is in the cloud or what goes on there apart from being able to receive service from it. Cloud computing is now associated with a higher level abstraction of the cloud. Instead of there being application software, routers and servers, there are now services. The underlying hardware and software of networking is of course still there but there are now higher level service capabilities available to build applications. Hidden behind the services are data and computer resources. There have been many cloud computing platforms built so far including Google Cloud, IBM BlueCloud and Amazon Elastic Cloud. Programming models for cloud computing have been developed including MapReduce by Google and Hadoop by Yahoo group. Using these cloud computing models and toolsets, IT-related capabilities are provided as services, accessible without requiring detailed knowledge of the underlying technology. Of course, many mature technologies are used as components in cloud computing, but there are still many unresolved and open problems due to its unique characteristics which are different from distributed computing, cluster computing, grid computing, utility computing and service computing.

ICCC invites papers from all areas of innovative developments, research issues and solutions in cloud computing and its related technologies.

AREA OF COVERAGE

Although the possible set of topics is large and we encourage submission on any area within the scope of cloud computing, the following areas are particularly suitable (but not exhaustive):

- Auditing, monitoring and scheduling
- Automatic reconfiguration
- Autonomic computing
- Cloud architecture and modelling

- Cloud-based services
- Consistency models
- Data grid and semantic web
- Fault tolerance and reliability
- Hardware as a service (HaaS)
- High-performance computing
- Integration of mainframe and large systems
- Innovations in IP (esp. open source) systems
- IT service and relationship management
- Load balancing
- Middleware frameworks
- New and innovative pedagogical approaches
- Novel programming models for large computing
- Optimal deployment configuration
- Peer to peer computing
- Power-aware profiling, modelling, and optimisation
- Resource registration and discovery
- Scalable fault resilience techniques for large computing
- Scalable scheduling on heterogeneous architectures
- Security and risk
- Self healing
- Self monitoring
- Service integration and management
- Service level agreement definitions
- Software as a service
- Trustworthiness, security and privacy
- Utility computing
- Virtualisation technologies
- Web services

EDITORIAL BOARD

Dr. K.S. Sreedhara,
Professor & Chairman
Dept. of Computer Science & Engineering
University BDT College of Engineering
Davanegere University
Davengere 577004

Dr. A. Sreenivasan
Professor, PG Studies
Dayananda Sagar College of Engineering,
Kumaraswamy Layout
BANGALORE-560078

Dr. Ramaswamy
Professor, Dept. of computer Science
JAIN UNIVERSITY, Kanakapura Road
BANGALORE, Karnataka

Dr. N V Subba Reddy,
Director,
Sikkim Manipal Institute of Technology,
Majitar, Rangpo, East Sikkim

Dr. S.V. Sathyanarayana
Prof & Head
Dept. of Electronics & communication
J.N.N. College of Engineering
SHIMOGA
Karnataka

Dr. K. N. Hari Bhat,
Prof & Head,
Dept. of E & C,
Nagarjuna College of Engg., & Tech.,
Bangalore.

Dr. K.T.Veeramanju
Professor, Dept. of Electrical Engineering
J. C. Engg. College, MYSORE

Organizing Committee

Chief Patron:

Prof. (Dr.) Srikanta Patnaik
President, IRNet India
Intersceince Campus, Bhubaneswar

Programme Chair

Prof. K. Karbasappa,
Professor & Head,
Department of Electronics and Communication Engineering
Dayananda Sagar College of Engineering
Bangalore, Karnataka

Secretary IRNet:

Prof. Pradeep Kumar Mallick
IIMT, Bhubaneswar
Email:pradeepmallick84@gmail.com
Mobile No: 08895885152

Organising Coordinator

Mr. Sanjay Sharma
Asst. Professor.
Department of Infomation & Technology
IIMT, Bhubaneswar
Bhubaneswar- 752054 , India

Conference Coordinator:

Prof. Suchana Mishra
Dayananda Sagar College of Engineering
Bangalore, Karnataka

Bibhu Prasad Mohanty
IRNet, Bhubaneswar

Event Manager
Prof. Sharada Prasad Sahoo.
IIMT, Bhubaneswar

Publication

Prof. Sushanta Kumar Panigrahi
IIMT, Bhubaneswar
Prof. Mritunjay Sharma
IIMT, Bhubaneswar

Post Conference Coordinator

Bibhu Prasad Mohanty
IRNet, Bhubanesawr
Mob:08895995279

Head (System & Utilities)
Prof. Sanjay Sharma
IIMT, Bhubaneswar

Members of IRNet

Rashmi Ranjan Nath
IRNet , Bhubaneswar,India

Miss. Ujjayinee Swain
IRNet , Bhubaneswar,India

Akash Kumar Bhoi
IRNet, Bhubaneswar, India

Bikash Kumar Rout
IRNet, Bhubaneswar

Pritika Mohanty
IRNet, Bhubaneswar

Mr. Jayabrata Banarjee.
IOAJ , Bhubaneswar

First Impression : 2012

Proceedings of International Conference On **Cloud Computing**

DISCLAIMER

The authors are solely responsible for the contents of the papers complied in this volume. The publishers or editors do not take any responsibility for the same in any manner. Errors, if any, are purely unintentional and readers are requested to communicate such errors to the editors or publishers to avoid discrepancies in future.

ISBN : 978-93-81693-17-9

Published by :

IPM PVT. LTD., Interscience Campus
At/PO.: Kantabada, Via: Janla, Dist: Khurda, Pin- 752054
Publisher's Website : www.interscience.in
E-mail: ipm.bbsr@gmail.com

Typeset & Printed by :
IPM PVT. LTD.

TABLE OF CONTENTS

Sl. No.	Topic	Page No.
	Editorial - *Prof. (Dr.) Srikanta Patnaik*	
1	**Cloud Computing in South African SMMEs (Risks and Rewards for Playing at Altitude)** - *Charles Hinde & Jean-Paul Van Belle*	1 - 10
2	**Cloud Adoption Lifecycle** - *Kumar Anand*	11 - 14
3	**Three Tier Data Storage Security In Cloud Using Face Fuzzy Vault** - *Vrushali Joshi, Payal Sanghvai & Yogita Bhargude*	15 - 20
4	**Challenges on Amazon Cloud in Load Balancing, Performance Testing and Upgrading** - *Himanshu Shah, Paresh Wankhede & Anup Borkar*	21 - 26
5	**Enhancement for Data Security in Cloud Computing Environment** - *Gawali M. B., R. B. Wagh & S. P. Patil*	27 - 33
6	**Improving Accessibility of Mobile Cloud Computing Using Multidimensional Data Structure** - *Rinku V. Rajankar & R.W. Jasutkar*	34 - 36
7	**Packet Monitoring Approach to Prevent DDoS Attack in Cloud Computing** - *Vikas Chouhan & Sateesh Kumar Peddoju*	37 - 41
8	**Challenges in Cloud Environment** - *Himanshu Shah, Paresh Wankhede & Anup Borkar*	42 - 46
9	**QoS Based Scheduling of Workflows in Cloud Computing** - *Jayadivya S K & S. Mary Saira Bhanu*	47 - 53
10	**Stepping into the Cloud** - *Venubabu Kunamneni*	54 -64
11	**Dynamic Load Balancing for the Cloud** - *Venubabu Kunamneni*	65 - 69

Cloud Computing in South African SMMEs

Risks and Rewards for Playing at Altitude

Charles Hinde & Jean-Paul Van Belle

Department of Information Systems, University of Cape Town, Cape Town, South Africa

***Abstract* -** Cloud computing can offer an excellent value proposition to small, medium and micro enterprises (SMMEs), for whom cash flow, physical IT assets, and human resources are at a premium. However, recent industry surveys among SMMEs are not very conclusive in how its benefits and barriers are perceived and experienced by SMMEs. This study investigates the adoption and perceptions of cloud computing by SMMEs in an emerging economy, South Africa, to gain a deeper insight into two particular questions. The first is to investigate whether the (perceived or actual) benefits and risks associated with cloud computing differ from that of those in the developed world. The second research question is whether these perceived factors differ between adopters and non-adopters of cloud computing. These questions were pursued using a survey-based study. The emerging profile indicates that company size, industry sector and owner/manager involvement all shape businesses' cloud adoption policies. Adoption benefits and perceived risks differ to some extent with those factors uncovered in empirical research in developed countries. Interestingly, there are statistically significant differences in perceptions on cloud computing between adopters and non-adopters. These findings offer concrete and practical intervention options if governments and/or vendors wish to encourage the adoption of cloud computing among SMMEs in emerging countries in order to increase their competitiveness and innovation.

***Keywords* -** *cloud computing; small business, SMMEs; South Africa.*

I. INTRODUCTION

Cloud computing has fundamentally changed the information technology (IT) landscape for small, medium and micro enterprises (SMMEs). Traditional software models in which businesses are tied to applications which are vendor specific, restricted by user-licenses, incur annual license fees, and require software patches or upgrades, are being replaced by cloud-based solutions. Cloud computing can change the way business interacts with software, hardware and computer services [1]. Some of the cloud computing benefits are more acutely felt by SMMEs: they often do not have the financial or human resources to invest in IT. Cloud computing can level the playing field by spreading the economies of scale in respect of software and hardware platforms [2] [3]. Identifying the actual perceived benefits and issues associated with cloud computing remains a critical research issue [4] [5]. Furthermore, the research on how businesses are approaching and perceiving cloud technology has mostly been conducted in Europe and the United States with very few studies having been carried out in emerging economies, such as South Africa.

Gaining a better understanding of the cloud computing benefits experienced and barriers faced by SMMEs is a very important research question. SMMEs are an important part of most emerging economies: in South Africa, they contribute 56% of private sector employment and 36% of the gross domestic product [6]. If their survival and growth through cost savings, increased efficiencies, greater competitiveness and innovation can be sustained or enhanced by cloud computing, it will contribute directly to the development of the larger economy.

The objective of the research is to determine the level of adoption of cloud computing within South African SMMEs and develop an insight into both the advantages and disadvantages that business owners feel cloud technology offers them. In particular, these factors will be compared to those uncovered in recent studies in more developed economies, given that SMMEs in emerging countries are faced with different contextual issues such as relatively scarce IT skills, weaker internet connectivity infrastructure and different relative cost structures. A second research objective is to build a comparative profile contrasting the differences, if any, in perspectives between businesses that have adopted cloud computing and those that have not. The study thus wishes to provide original insights into the factors and possible reasons behind South African companies' disposition to adopt cloud services.

II. LITERATURE REVIEW

A. Definitions

Since cloud computing is a technology which redefines both infrastructure and software, the breadth of its scope makes it at times difficult to conceptualise. A technical definition from the National Institute of

Standards and Technology (NIST) defines cloud computing as constituting the following essentials: on-demand self-service, broad network access, resource pooling, rapid elasticity, and it must be a measured service [7]. A more mainstream understanding of cloud computing can be defined as 'internet centric software.' [8].

Cloud computing currently provides three service models. Software as a Service (SaaS) can encompass customer relationship management solutions, accounting packages, enterprise resource planning, human resource management, and content management systems – essentially any fully functional service or application delivered via the internet. Platform as a Service (PaaS) offers an environment that allows developers to design, develop, test and deploy web applications or services without the cost or complexity of setting up the hardware and software in-house. With Infrastructure as a Service (IaaS), the consumer controls the majority of their environment, and cloud vendors offer a virtual machine (VM) environment allowing control over the operating system, storage, and application deployment [5].

In addition to service models, the cloud can be categorised into deployment models: private, public, community and hybrid cloud. Here the focus is on the public cloud: an infrastructure available to the general public and owned by an organisation (Google, Amazon, Rackspace, Microsoft) selling cloud services [7].

Since this study focuses on South African SMMEs, it is envisaged that the services utilised within the cloud will span all 3 models (SaaS, PaaS, IaaS) depending on requirements, but the platforms on which they are used will be within the public cloud. As most SMMEs don't have the capital to invest in their own data centres to create a private cloud, nor do they have the infrastructure to offer a community cloud, the public cloud will by default become the chosen platform. There may be instances where an SMME has a small server farm that requires additional compute capacity during high usage times and therefore may fall into a hybrid cloud scenario.

SMMEs are defined in South Africa as consisting of micro-businesses (up to 5 employees), very small businesses (up to 20 employees), small businesses (<50) and medium sized businesses (up to 200) [9]. However, in many larger economies, such as the US, the cut-off values to qualify as an SMME are slightly higher [3], making comparisons between surveys somewhat difficult.

B. Evaluating the Cloud as an SMME

Cloud computing adoption issues are different for SMMEs than they are for larger organisations [10]. However, these authors identified the most important adoption issues for SMMEs as follows.

Outage of the service provider or connection affects the service availability which, in developing countries, depends critically on the reliability of internet infrastructure. Security of data is often perceived as an important weakness even though, from a technical and practical perspective, data in the cloud often is more secure than in-house hosted data, especially for SMMEs who often lack staff with security expertise. Performance can become an issue because communication lag reduces application responsiveness. This can be exacerbated when a larger number of users are simultaneously making data-intensive requests over a limited bandwidth channel; this is a typical circumstance for small businesses in developing world contexts. Integration with existing applications is also an important issue where the SMME already has a number of information systems implemented. Finally, although the economies of scale by the cloud computing provider can result in significant savings, internet connectivity is still very expensive in developing nations, often eroding some of the major cost advantages [10].

Given the importance and exciting potential of cloud computing, a number of empirical research studies have focussed on the perceived and actual benefits and risks associated with cloud computing adoption. Given its relatively recent ascent, much of this research is very recent (2009-2011) and on-going. However, early results show little or no consensus findings, especially if research in the UK and USA is compared with studies outside those countries.

For instance, the 2010 IDC survey [11] ranked perceived security as the highest obstacle to cloud computing, followed closely by both outage/availability and performance. This was followed by financial barriers (potentially higher costs) and technical concerns such as interoperability/integration and lack of customization. By contrast, a small but recent survey in the Czech Republic [12] identified cost reduction, rapid deployment, scalability and improved flexibility as the key drivers behind SMME cloud computing adoption. Interestingly, the increased dependence on external providers was seen as the most important barrier (17%) although it was rated only slightly more important than the potential for increased costs (15%) and security concerns (14%). Then again, [13] highlighted the lack of environments for helping businesses migrate their legacy applications to the cloud as well as the difficulties of finding and integrating different cloud services for a given set of business requirements as important cloud computing barriers for SMMEs.

Similarly, there is little consensus on the relative importance of the benefits which cloud computing

offers. Although there is general consensus that cloud computing can offer scalability, reliability, security, ease of deployment, and ease of management for customers [14], others [3] list low start-up costs, low cost for sporadic use, ease of management, device and location independence as key benefits. Other benefits identified are the lowering of entry barriers and easy opportunities to test new information system solutions and deal with seasonal fluctuations in demand. Ref [15] emphasises the financial savings resulting from the pay-per-use model, whereas [16] singles out cloud computing's potential for incremental improvement to avoid disruptive transformation of business processes. Finally, [17] focuses on the more strategic benefits such as increased competitiveness and increased capacity for innovation. Particularly relevant for SMMEs in emerging countries are the opportunities which cloud computing offers SMMEs to network and collaborate to create global competitive advantage [18].

C. *Reference Surveys*

In what follows, findings will be compared with those of a small number of recent cloud computing surveys conducted in developing countries. In order to preserve homogeneity within the reference group, only European surveys were referenced in the data analysis. The European Network and Information Security Agency (ENISA) published its "SME perspective on Cloud Computing" based on a European survey [19]; Easynet Connect conducted a study into UK small and medium sized business' readiness for cloud computing and Software as a Service [20]; GFI Software [21] published its SME Technology Report that, although not specifically aimed at cloud computing, included this as a major survey component. Furthermore, our findings will also be compared to a survey conducted by ITWeb [22] into cloud computing adoption in South Africa, although this survey was not aimed specifically at SMMEs (that make up 50% of its sample) and its sample is heavily biased in favour of the IT industry.

III. RESEARCH PROPOSITIONS AND METHODOLOGY

The objectives of the proposed research project are to determine the level of adoption of cloud computing within South African SMMEs and the factors that influence the adoption process as well as to investigate the differences between businesses that have adopted cloud computing and those that have not. We are therefore positing the following research propositions:

P1: The benefits associated with and the barriers to cloud computing, as perceived or experienced by SMMEs in South Africa, differ in relative importance from those in more developed countries.

P2: Perceptions of the benefits and issues associated with cloud computing will differ significantly between adopters and non-adopters.

The underlying philosophy of this research is positivist and aims to investigate systematically and empirically the quantitative properties of cloud computing adoption by SMMEs in South Africa. This research is both exploratory and explanatory in nature: exploratory as it attempts to build a profile of the factors that influence the adoption of cloud-based services within SMMEs, and explanatory as it examines these factors to clarify the reasons behind the adoption process.

The primary data collection was a self-administered online survey distributed to SMME owner/managers. The online survey questions were derived from multiple sources: areas of interest highlighted in the literature review process as well as current local surveys and past surveys aimed at European SMEs. The research instrument was piloted with a few SMME owners and researchers for clarity, and is available from the researchers on simple request.

The survey instrument for this study incorporated questions from similar research instruments being used in Europe [19-21] and South Africa [22]. Where available, the analysis in this section will include the findings from those surveys. Comparative analysis between the various surveys does not always provide like-for-like answers as the various categories (like company size) may differ; where necessary, discretion has been used in grouping similar answers. The answers to each survey question have been ranked from highest to lowest (percentage or actual value) according the published results. These rankings have then been compared with the rankings for answers to the like-for-like questions in this survey. It particularly references the ITWeb survey conducted by an industry trade organisation. However, half of the ITWeb respondents had more than 200 employees (the DTI cut-off for medium-sized enterprises) and the survey was heavily weighted in favour of the IT industry (47% of the responses).

Given the typical difficulties associated with obtaining a random, representative or stratified sampling frame for SMMEs in a fast-growing emerging country, we opted for a convenience sample obtained by means of the snowball sampling method. An encouraging total of 72 responses were received but not all respondents completed all questions. Where relevant, the number of respondents for a particular response will be indicated. On completion of the online survey, respondents were also invited to participate in a semi-structured, face-to-face interview. Although limited, the additional

information gathered through the semi-structured interviews did provide context and nuance to the quantitative data. Due to space limitations, much of this data is not incorporated, although some particularly relevant references or quotes will be included in the analysis contained in section five of this report.

IV. DATA ANALYSIS: THE SOUTH AFRICAN VERSUS EUROPEAN SMME PERSPECTIVE

This section provides a high-level overview of the research findings and compares them with the findings from European surveys. In the following section 5, the survey sample is separated into two distinct groups (those which have adopted cloud services and those which have not) and more in-depth analysis is undertaken. Section 5 also includes excerpts from interviews, which provide a richer source of information with which to overlay the survey data.

A. Demographic Characteristics of the Survey Sample

Although the survey was specifically aimed at SMMEs, the snowball sampling method could not guarantee that all respondents were from companies employing less than 200 employees. 21% of respondents indicated that they had more than 200 employees. However, given that the latter respondents still considered themselves as medium-sized (as per the survey introduction) and some of them employed less than 250 employees (the criterion used in the comparative international studies), we decided to keep them were included in the sample. This decision was supported after additional data analysis confirmed that the responses from these organisations did not differ markedly from the other respondents.

TABLE I : COMPANY SIZE

How many people does your company employ?			
This Survey - ZA 2011)	***ITWeb - ZA (2011)***	***GFI - UK (2010)***	***ENISA – EU (2009)***
1-5: 21%	3-20: 28%	<10: 46%	1-9: 37%
6-20: 19%	21-50: 7%	10-99: 29%	10-50: 16%
21-50: 26%	51-100: 7%	100-249: 25%	20-250: 19%
51-200: 12%	101-200: 7%	-	>250: 28%
>200: 21%	> 200: 50%	-	-

Unfortunately, differences in definitions of small, medium and micro-enterprises do not allow the surveys to be compared directly. However, our survey has slightly less micro-enterprises than the European studies but relatively more small and medium-sized businesses.

Surveys in IT tend to be aimed at the IT industry or industries that are IT- dependent. It is not surprising therefore that analysis reveals the IT sector as the dominant sector in the data with IT accounting for 37% of the responses. However this was equal to the combined financial services (20%) and general business services & consulting (17%) industry segments. A few responses came from manufacturing (3), government (2) and one response each from engineering and retail SMMEs with 12% of respondents from other industries.

TABLE II : ROLE WITHIN COMPANY

What is your job role/position within your organisation?	***This Study (2011)***	***ITWeb (2011)***	***GFI (2010)***
Owner	50%	19%	41%
Executive Management	10%	10%	18%
Middle Management	10%	25%	-
IT Specialist	24%	26%	21%
Other IT Roles	7%	11%	20%
Other	0%	8%	-

This survey attracted 66% of its respondents from companies with 1-50 employees, so it is not surprising that 50% of the respondents were business owners. This supports the literature that one executive usually makes IT-related decisions in SMMEs [23]. This survey aligns well with the 2010 GFI survey. The increase in owner respondents can be attributed to the growth in the number of SMMEs taking the survey. The ITWeb survey's weighting towards middle management will be due to the high-percentage contribution of businesses employing over 200 employees (50%).

When asked how important IT is to the business, 71% of the respondents across all sectors indicated that IT plays a "crucial" role in their business, with an additional 10% indicating it is "very important" (81% in total). This supports the literature which has found that IT plays an important role in enabling and supporting growth within the SMME sector [24].

B. Cloud Adoption

Cloud computing is a growing reality in South Africa with just over half (52%) of respondents saying they, or the company they work for, have adopted a cloud service. This is further strengthened with 65% of respondents believing it is real and here to stay although 25% think it is more suitable for larger, more mature markets. However, only 34% of respondents say their organisations have a formal cloud computing adoption model in place. This supports the findings of a recent industry report published by Microsoft [25], which questioned 3,258 SMBs that employ up to 250 employees across 16 countries worldwide, finding that small to mid-size businesses (< 250 employees) are rapidly becoming more involved in adopting cloud services. However, the actual adoption rate of 52% is

unlikely to hold much meaning due to the self-selection bias inherent in our snowball sampling method; thus no comparison with reported (and possibly similarly biased) adoption rates from other surveys is made.

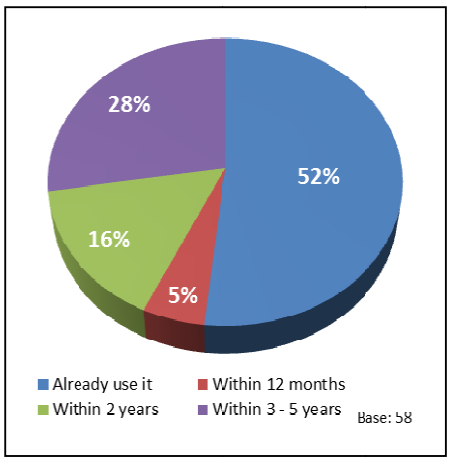


Fig. 1 : Cloud computing adoption by survey respondents (n=58).

When looking at different cloud solution models, 36% of respondents say their organisations would choose to adopt either fully hosted software services (SaaS) or infrastructure services (IaaS). Google is regarded as the most credible cloud vendor, with 76% of respondents giving it a rating of four or five out of five.

This confirms the Microsoft report that smaller companies are adopting cloud services more aggressively. It suggests that the difference between the two surveys, 52% versus 38% of companies adopting cloud services, is realistic.

C. *Cloud Views, Strategies, Layers, and Vendor Credibility*

Of particular interest was whether South African SMMEs perceived cloud computing to be more relevant to either larger organisations or more developed countries. Although almost exactly two-thirds (65%-67%) of respondents to both surveys agree that cloud computing is "real and here to stay", a substantial number of respondents, accounting for most of the remaining third of respondents, believe it is more suitable to "more mature markets". Unfortunately, due to the ADSL footprint only extending to 20% to 27% of fixed-line Telkom subscribers (Muller, 2010), there may still be the perception that the South African infrastructure is immature compared to Europe and the United States.

TABLE III : VIEW ON CLOUD COMPUTING

Cloud computing is ….	***This Study (2011)***	***ITWeb (2011)***
Real and here to stay	65%	67%
Is more suitable to larger more mature markets but will need to prove itself in SA	25%	30%
Currently mainly of interest for large South African organisations	6%	0%
Mostly hype generated by the IT industry, like Y2K	4%	3%

With 65% of respondents believing that cloud computing is "real and here to stay", and a 52% cloud services adoption rate, it is surprising that only a third of companies have a formal cloud strategy in place. A simple explanation for this low-strategy adoption may be that smaller companies have adopted cloud services like Google Docs, Dropbox or Gmail, meaning a formal cloud strategy was not required.

TABLE IV : CLOUD ADOPTION STRATEGY

Our business has or plans …	***This Study (2011)***	***Easynet Connect (2009)***
to develop, a formal strategy for moving over to cloud computing services	34%	29%
on putting in place more stringent security measures as a result of cloud computing	49%	15%
on implementing, business continuity measures as result of cloud computing	51%	22%
on, increasing internet bandwidth as a result of cloud computing	62%	19%

The upward trend indicated around security, business continuity and increased bandwidth as the result of implementing a cloud solution is positive. This contrasts starkly with data collected in the Easynet Connect survey conducted four years ago in which security and business continuity were not considered important parts of an overall cloud computing strategy.

When queried about the first layer (or type) of cloud service, there was a significant difference between the ITWeb results and this survey. Unfortunately, the latter used a multiple response so the results cannot be compared directly. There are also differences with ENISA with more South African SMMEs focussed on IaaS instead of PaaS. Our findings confirm the literature suggesting that SaaS will lead the way in market growth [17]

TABLE V : CLOUD ADOPTION LAYER (N = 72)

Which layer of the cloud would you be most likely to approach?	*This Study (2011)*	*ENISA (2009)*
SaaS	36%	34%
IaaS	36%	25%
PaaS	17%	29%
Security Services	10%	10%
Other	1%	3%

When asked to rate the credibility of cloud computing service providers, Google's score of 76% was substantially higher than that of the other vendors, with Amazon Web Services and Microsoft Azure tying with 51% in second place. The Apple iCloud (43%) and IBM Blue Cloud (32%) filled the next two slots above 30%. The cloud credibility of these companies, although not covered in this study, was more than likely enhanced by the products and services provided in other domains.

D. Influences, Concerns, and Benefits of Cloud Computing and Moving Applications to the Cloud

When asked about the factors that influence SMMEs to move applications to the cloud, it is interesting to note the differences with their European counterparts.

TABLE VI : MOST IMPORTANT CLOUD COMPUTING INFLUENCES

This Study (2011) (% scoring 4 or 5 out of 5)	*GFI (2010)*
No lock-in terms, no contracts: 81%	Better vendor terms and pricing: 43%
Standards based security (ISO/Global standards): 67%	No lock-in terms, no contracts: 40%
Better pricing: 60%	Ability to use the service on a monthly basis: 31%
Better accountability and auditability of backend processes: 60%	Guaranteed high level of security: 22%
Improved Service Level Agreements: 58%	Ability to mix and match services: 22%

It is interesting to note that global standards-based security is not the most important factor influencing respondents to adopt cloud technology, although it is the biggest reason for *not* adopting the technology (see below). This survey has security placed second (67%) to contracts and lock-in terms (81%), while the 2010 GFI survey places security joint last. Concerns around lock-in terms and contracts may be a hangover from traditional software, in which annual licensing agreements, software upgrades, and the lack of data interoperability tend to keep clients tied to a particular vendor. It seems that potential cloud clients need to be sure their data is transferable and that they are free to try before they buy, after which they will assess the security of their data.

TABLE VII : CLOUD ADOPTION CONCERNS

Rate the concerns below in terms of the influence they would have on your business NOT adopting cloud computing?	*This Study (2011)*	*ITWeb (2011)*	*GFI (2010)*	*ENISA (2009)*
Data Privacy & Security	88% (1)	62% (1)	35% (1)	94% (1)
Connectivity and downtime issues	83% (2)	58% (3)	30% (5)	82% (5)
Recovery of Data	76% (3)		29% (6)	
No off-line/ redundancy options	67% (5)		29% (6)	
Cloud vendor downtime	67% (6)			74% (6)
Other	0%	8%	-	

The four surveys offer a respondent a multitude of options for not considering adopting cloud technology. However, within the top six options in each survey the two consistent themes that will stop prospective clients from moving to cloud technology are data security, privacy and connectivity. With the recent spate of high-profile data breaches (e.g. Sony and Citibank) it is important that cloud vendors provide potential clients with unparalleled security and access to their data.

TABLE VIII : CLOUD ADOPTION BENEFITS

In your opinion what are the main benefits for using a cloud-based service as opposed to running it in-house?	*This Study (2011)*	*GFI (2010)*
Business continuity & disaster recovery	72% (1)	
Scalability	65% (2)	32% (9)
Ease of adding/removing services	57% (3)	32% (13)
Lower capital costs	55% (4)	37% (10)
Speed of deployment/implementation	55% (5)	37% (8)
Reliability	50% (6)	39% (5)
Lower operating costs	48% (7)	42% (2)
Expertise of the service providers staff	48% (8)	45% (1)

Much more surprising is the lack of agreement about the cloud computing benefits (expected or experienced) between the two surveys. There is hardly any overlap between the two samples' data – if anything they seem to be inverted. One possible reason for this is that over the course of a year, and with the maturing of the cloud services on offer, there is a greater awareness of what the cloud has to offer businesses: disaster recovery, business continuity, scalability of services, and ease of adding or removing these services. As users have become more familiar with how cloud services work, there is less reliance on the vendors' technical staff, operating costs have dropped and normalised due

to additional competition, and due to the services' maturation there is less concern about their accountability.

TABLE IX : CLOUD ADOPTION SERVICES

Which of the following applications/services would you move to the cloud?	*This Study (2011)*	*ITWeb (2011)*
Webhosting & e-commerce	94% (1)	61% (2)
Email hosting/archiving	75% (2)	70% (1)
Customer Relationship Management Systems (CRM)	58% (3)	38% (5)
Configuration and data backup	58% (4)	56% (3)
Application development	40% (6)	30% (6)

The top two services selected in both surveys can be considered 'old-school'. Email and webhosting have been offered as services since the internet boom in the mid-1990s with brands such as Hotmail, Yahoo, Google and Amazon being household names. Moving in-house services such as email and webhosting to the cloud would not be considered high-risk as the services are established and mature. The nature of the data at risk would also not be considered crucial to a business's success or failure – possibly an inconvenience if the service were temporarily to fail, but very unlikely to cause irreparable damage.

From the above analysis (Tables VI, VII and VIII), we can conclude that the relative importance attached to various influence on, benefits associated with and concerns about cloud computing, as perceived or experienced by SMMEs in South Africa, differs from the rankings given by European SMMEs. This supports our first proposition.

Section 5 will continue to analyse the sample data with the aim of refining the profile of respondents who have adopted cloud services.

V. CLOUD ADOPTERS VERSUS NON-CLOUD ADOPTERS

After the full survey sample of 58 respondents has been analysed, this section aims to give a deeper analysis of companies who have adopted cloud services in contrast with those who have not. The study aims to identify the underlying drivers that lead certain companies to adopt cloud services and to explore the role of the business owner in this process.

The survey data for this section has been split into two groups: those who already have implemented cloud technology and those who have not. The latter includes those who indicated they were considering implementing cloud technology in the next "12 months," "2 years," and "3 - 5 years".

A. Demographic Characteristics: Cloud Adopters versus Non-Cloud Adopters

A demographic analysis of the companies who have adopted cloud services into their IT infrastructure provides an interesting profile. Cloud adopting SMMEs tend to be small(er) (87% employing 1-50 people); provide services in the IT industry (52%) and, unsurprisingly, consider IT a crucial part of their business (86%). Despite the relatively small sample, these differences are all statistically significant (critical Z-score at 95% = 1.960).

TABLE X : COMPANY DEMOGRAPHICS

	Cloud Adopters	*Non-Adopters*	*Z-score*
Company Size	1-50 employees: 87% (18/21)	1-50 employees: 48% (10/21)	2.619
Company Industry	IT Industry: 52% (11/21)	IT Industry: 15% (3/20)	2.523
Role of IT in Company	Crucial: 86% (18/21)	Crucial: 57% (12/21)	2.049
Respondent's Role	Business Owner: 67% (14/21)	Business Owner: 33% (7/21)	2.160

Although this survey builds a limited profile of a company's human IT skills, the data collected suggests that business owners who did complete the online survey (67%) compare favourably to IT specialists with their own IT use and familiarity. Of the respondents who indicated that they or the companies they work for have not adopted cloud services, only 33% are business owners with 43% being either IT specialists or coming from an IT background. There is no evidence to suggest business owners passed the survey on to their "IT Department" to complete, but those businesses who have adopted cloud services seem to have owners who are more comfortable answering IT-related questionnaires.

B. Influences on the Decision to Adopt Cloud-based

Selecting only the variables with statistically significant differences ($p<0.05$), it becomes apparent that the influences and concerns of cloud adopters are based on real-world implications.

TABLE XI : DIFFERENCES IN ADOPTION INFLUENCES

What would influence you MOST in your decision to ADOPT cloud-computing services?	***Cloud Adopters***	***Non-Adopters***	***t-score (p-value)***
Better Pricing	73% (16/22)	48% (10/21)	2.114 (0.041)
Better Accountability and Auditability	77% (17/22)	43% (9/21)	2.452 (0.019)
Which concerns would influence you MOST in your decision to NOT ADOPT cloud-computing?	***Cloud Adopters***	***Non-Adopters***	***t-score (p-value)***
Total cost of ownership	67% (14/21)	38% (8/21)	2.572 (0.014)
Application Performance	86% (18/21)	43% (9/21)	2.990 (0.005)

These variables can be considered realistic requirements that respondents would assess before considering a migration from traditional, vendor-supplied applications or services to those offered in the cloud. If the cloud services are not able to provide the same accountability, auditability and performance functionality as traditional vendor software, or the implementation, running costs, and total costs of ownership are higher, then there is little use in considering adopting them.

Interviewee A questioned why a business would want to host their own Exchange Server when they could use a cloud service: *"Why have an Exchange Server in your business, why...the software is going to be out of date in 3 years time, your hard disks need to be replaced...why do it....you pay 200 grand for it, when you can pay 7 grand a month or whatever it is, get the same thing, and never have to worry about it again."*

Interviewee B continued the thread of application service and prices when discussing the cloud services they have implemented*: "You can't believe how easy it is, you get all this service for no investment...not having to set up servers and installing things, trying to back-up and fail-over...you just get it, and it's working...and you feel that in your operations, and you'll question anything you'll ever want to install again after going through that. […] Cloud services fundamentally change start-up business models...and it's the one reason we're attacking this business opportunity...strategic investment is in fact because there is so little strategic investment required."*

C. The Main Benefits of Moving to the Cloud

Identifying the variables which differed significant ly between adopters and non-adopters once again point to real-world experiences by cloud adopters.

TABLE XII : DIFFERENCES BETWEEN PERCEIVED CLOUD BENEFITS

In your opinion what are the main benefits for using a cloud-based service as opposed to running it in-house?	***Cloud Adopters***	***Non-Adopters***	***t-score (p-value)***
Accountability	46%	18%	2.383 (0.021)
Reliability	73%	23%	3.729 (0.001)
Re-deploy existing IT skills	56%	23%	2.139 (0.038)

Accountability and reliability require cloud services to match or out-perform traditional software packages. The re-deployment of in-house IT skills offers businesses the flexibility to use external skills that might otherwise have required employing these skills – a cost saving. Businesses that have adopted cloud services appear to find the services' accountability and reliability acceptable relative to traditional software, and they derive additional benefits from these services by re-deploying their existing staff IT skills elsewhere in the business.

D. Service Providers' Perceived Credibility

Respondents' replies around vendor credibility provide an interesting insight into real-world versus perceived credibility. Google Apps was the most credible vendor for both adopting and non-adopting respondents, but the rankings after this show the difference in real-world versus perceived credibility. Respondents who have adopted cloud services ranked Amazon Web Services, Rackspace and Salesforce as the next most credible vendors, while non-adopters listed Apple iCloud, Microsoft Azure and Amazon Web Services in that order.

In the literature around cloud vendor market leaders [26], the companies that are regularly mentioned are Google, Amazon Web Services, Rackspace and Salesforce, while a company like Apple, for all its market presence, has only just entered the cloud domain with their iCloud solution. About deciding on cloud vendors, Interviewee B says: *"If I think of it in the main, we've gone with industry leaders, for email we're at Rackspace...for server hosting we're at Amazon, although Rackspace are also good, but I think Amazon are now the industry leader...we use Dropbox for document management, they must be in the top two for*

that...in the main the market leaders are the market leaders for a reason."

In summary, the analysis confirms our second proposition that the perceptions of *some* influences and benefits associated with cloud computing differs significantly between adopters and non-adopters. Additionally, we uncovered also statistically significant differences between the demographic profile and the perceptions around the relative credibility of the leading cloud vendors.

VI. CONCLUSION

A. Summary of Findings

Cloud computing is redefining the IT landscape. As it becomes more pervasive, cloud computing brings with it the opportunity for SMMEs to reduce their IT infrastructure costs, adopt high-end information systems, increase flexibility, implement hugely scalable solutions and thus effect not only operational efficiencies but also improve their global competitiveness and innovation capacity.

This study revealed that even SMMEs in an emerging economy such as South Africa are adopting cloud computing solutions quite aggressively. The survey shows that most SMMEs currently adopting cloud computing fit a unique profile: they are small, work in the IT industry, and have business owners who are technologically proficient enough to drive the cloud adoption process themselves.

Interestingly, it emerged that the relative importance as well as criticality of benefits and obstacles associated with cloud computing differs markedly in South Africa when compared with previous surveys conducted in Europe. Although some of these can be ascribed to the developing world context of limited skilled IT resources and infrastructure, there also appears to be less consensus about this even between surveys held in the developed world.

Another important finding in this study was that the perceptions of cloud computing are statistically significantly different between the SMMEs that have implemented cloud computing already versus those that have not. It is clear that after adoption, operational and practical concerns and benefits quickly overtake initial misconceptions held by those who have not implemented cloud computing.

This research therefore confirmed that IT is an almost inescapable part of the business landscape, and with SMMEs traditionally under-resourced and under-funded, cloud computing offers a viable solution for expanding their current ICT infrastructure, be it application or hardware solutions.

It is hoped that SMMEs will consider cloud computing as a viable alternative to traditional software and hardware solutions, and that vendors are able offer robust solutions that are mindful of the fears of potential cloud adopters.

B. Limitations and Future Research

It must be acknowledged that, due to the snowball sampling technique and relatively small final sample size, the study's findings cannot be generalised or considered representative of all South African SMMEs. There was a significant response bias with a larger proportion of adopters responding than non-adopters. Due to the technical nature of the survey, its niche topic and its relatively short duration, the respondents also tended to be dominated by IT-centric businesses.

Future research could definitely attempt to obtain a more representative and larger sample. However, as pointed out, even the larger surveys conducted in large developed economies are inconsistent in the relative ranking of benefits and issues associated with cloud computing adoption. The differences uncovered between adopters and non-adopters suggested that a longitudinal study may be necessary. This could reveal that influencing factors, benefits and barriers to cloud computing vary depending on the maturity of the solutions offered by cloud computing providers as well as on the stage of adoption by SMMEs.

ACKNOWLEDGMENTS

We hereby gratefully acknowledge the support of the many SMMEs participating in the research. This research was supported by South Africa's National Research Foundation.

REFERENCES

[1] S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang, & A. Ghalsasi. Cloud computing - The business perspective. Decision Support Systems, 51(1), April 2011, pp.176-189.

[2] N. Sultan Reaching for the "cloud": How SMEs can manage. Intl. Jl. of Information Management, 31(3), 2011, pp.272-278.

[3] F.T. Neves, F.C. Marta, A.M. Correia, M. de Castro-Neto. The Adoption of Cloud Computing by SMEs: Identifying and Coping with External Factors. 11ª Conferência da Associação Portuguesa de Sistemas de Informação (CAPSI

2011) – A Gestão de Informação na era da Cloud Computing, Lisboa, 19-21th Oct 2011.

[4] I.L. Sriram & A. Khajeh-Hosseini. Research Agenda in Cloud Technologies. Subm. 1st ACM Symposium on Cloud Computing, Indianapolis, Indiana, USA, June 2010.

[5] M.A. Vouk. Cloud computing Issues, research and implementations. In 30th Intl Conference on Information Technology Interfaces (ITI 2008). Cavtat/Dubrovnik, Croatia, June 2008, pp.31-40.

[6] O.O. Fatoki & A.V. Smit. Constraints to credit access by new SMEs in South Africa: A supply-side analysis. African Jnl of Business Management, 5(4), Feb 2011, pp.1413-1425.

[7] P. Mell & T. Grance. The NIST Definition of Cloud Computing. Gaithersburg: National Institute of Standards and Technology, 2009.

[8] R. Cohen. Twenty-One Experts Define Cloud Computing. 2009. Retrieved: http:// cloudcomputing. sys-con.com/ node/ 612375.

[9] Department of Trade and Industry. National Small Business Act, Pretoria, 27 Nov 1996.

[10] W. Kim, S.D. Kim, E. Lee & S. Lee. Adoption issues for cloud computing. Proc. 7th Intl. Conf. on Advances in Mobile Computing and Multimedia MoMM 09 (2009). ACM Press.

[11] IDC. Cloud Computing: An IDC Update. 2010. Retrieved: http://www. slideshare.net/ JorFigOr/cloud -computing-2010-an-idc-update.

[12] G. Feuerlicht, L. Burkon & M. Sebesta. Cloud Computing Adoption: What are the issues? Systémová Integrace Vol 2 (Spring 2011), pp.187-192.

[13] H. Motahari-Nezhad, B. Stephenson, & S. Singhal Outsourcing Business to Cloud Computing Services: Opportunities and Challenges. Subm. IEEE Internet Computing, Special Issue on Cloud Computing, 2009.

[14] H. Erdogmus. Cloud Computing: Does Nirvana Hide behind the Nebula? Software, IEEE 26, 2 (2009), pp.4-6.

[15] T. Dimitrakos. Common Capabilities for Service Oriented Infrastructures and Platforms: An Overview. IEEE 8th European Conf. on Web Services (ECOWS), 1-3 Dec.2010.

[16] M. Skilton. Building Return on Investment from Cloud Computing. Whitepaper, The Open Group, 2010. Retrieved : http:// www.opengroup.org / cloud/ whitepapers/ ccroi/index.htm

[17] J. Chen, & H. Song. Industrial clusters' information based on SaaS model. Intl Conf. on Business Management and Electronic Information (BMEI), 13-15 May 2011.

[18] A. Petrakou, P. Brandt, R. Gustavsson, & P. Jokela. Collaborative e-Marketplaces Containing Clusters of SMEs: Drivers and Barriers in the Local Food Sector. 44th Hawaii Intl. Conf. on System Sciences (HICSS), 4-7 Jan 2011.

[19] ENISA. An SME perspective on cloud computing. Report, European Network and Information Security Agency, 2009. Retrieved: http://www.enisa. europa.eu/ act/rm/files/deliverables/cloud-computing-sme-survey.

[20] Easynet Connect. Is 2010 the year SMEs fully embrace cloud computing? 2010. Retrieved: http://www.easynetconnect.net/Portals/0/DownloadFiles/IndustryInsight/WhitePapers/Is-2010-the-year-SMEs-fully-embrace-cloud-computing.pdf

[21] GFI Software. The SME Technology Report. 2010. Retrieved: www.gfi.com/documents/ SME_ Technology_ Report_web.pdf.

[22] ITWeb. Cloud Computing Survey 2011. Retrieved: http://www.itweb.co.za/index.php?option=com_content&view=article&id=45782&Itemid=2480.

[23] C. Riemenschneider, D. Harrison, & P. Mykytyn. Understanding IT adoption decisions in small business: integrating current theories. Information & Management, 40(4), 2003, pp.269-285.

[24] P. Matthews. ICT Assimilation and SME Expansion. Jl. of International Development, 19(6), 30 Jul 2007, pp.817–827.

[25] Microsoft. SMB Cloud Adoption Study - Global Report. Market Research Report, EdgeStrategies, Dec 2010.

[26] J. Vance. Ten Cloud Computing Leaders. IT-Management, 26 May 2010. Retrieved: http:// itmanagement.earthweb.com/article.php/3884311

◈◈◈

Cloud Adoption Lifecycle

Kumar Anand

Global Consulting Practice, Tata Consultancy Services, Bangalore, India

***Abstract* -** This paper can help an organization in adopting the cloud platform. It shows various stages to go through, from 'What' (Foresee) to 'What Else' (Reinvigorate), in adoption of cloud based infrastructure and services. At the end of this paper an organization would know what are the various stages involved and from where can it start if it is clueless from where to begin their quest for adopting the next gen way of sourcing IT services. The paper concludes on the note that an organization can utilize the lifecycle to relate to the stage where it is currently at and identify the next stages involved in the Cloud Adoption Lifecycle.

***Keywords* -** *Cloud Computing, Lifecycle, migration, adoption.*

I. INTRODUCTION

As per the 'Hype Cycle for Emerging Technologies' released by Gartner in July 2011, Cloud Computing is placed in the 'Peak of Inflated Expectations' phase [1]. It has been in this phase for the last 3 years. If one goes by it, Cloud Computing should enter the phase of 'Trough of Disillusionment', if at all that happens. An effort should be made, by the organization contemplating to adopt cloud, to avoid falling into this trough by carefully considering all aspects of cloud adoption. On reading through what has been written and listening to experiences of many on cloud adoption, one realizes that there are many things to consider before deciding to 'fly' into the Cloud. Hence a clear cut approach is needed which ensures that the organization moves in the right direction from the first step it takes towards adoption of cloud platform.

The paper discusses the 'FASTER' cloud lifecycle which can be followed to ensure that various aspects, technical to cultural, are thought through at each of the stages.

II. CLOUD ADOPTION LIFECYCLE

The paper discusses the 'FASTER' cloud lifecycle which can be followed to ensure that various aspects, technical to cultural, are thought through at each of the stages. Following are the stages which are involved in the cloud adoption lifecycle –

- **(F)**oresee
- **(A)**sses
- **(S)**trategize
- **(T)**est
- **(E)**stablish
- **(R)**einvigorate

Each of the lifecycle stages are discussed in the subsequent sections.

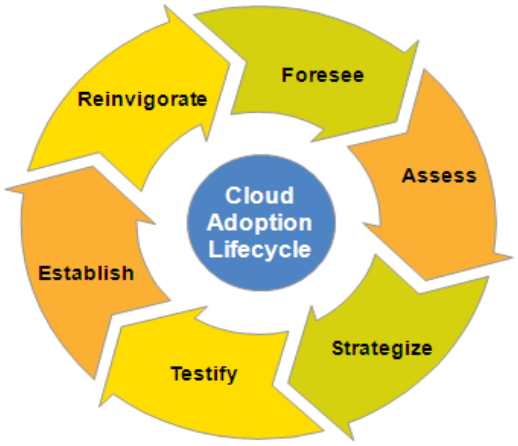


Fig. 1: Cloud Adoption Lifecycle

*A. **(F)**oresee*

1) Crystal Gaze: As a first step of this stage, organization should crystal gaze to understand the importance and need for a flexible and elastic response from IT as it grows. One should know if application and the underlying infrastructure are scalable and reliable enough to take on increasing workload and has the ability to absorb the spikes. How agile is IT to respond to the need of quicker deployment of services? To have an answer to these questions the organization should introspect to understand the challenges it is expected to face in the future. Some of the points which can be looked at are –

- Does the current delivery model enable to provision IT services at the speed at which business requires?

- Is daily operation of IT consuming and diverting valuable time of the team away from focusing on value creating activities?
- Is the organization unable to fully exploit the potential of IT, through underutilization of its assets, because of cultural issues of having a dedicated IT setup for each department

Introspecting on the above points can help in laying down the vision for IT which can enable business to grow in leaps and bounds.

Once the organization is able to foresee challenges coming up and the growing need for an IT which can support and sustain organizations' growth it can start looking for opportunities to build one.

2) Aware: Next step should draw focus on exploring the potential opportunities which can address the challenges and realize the vision laid down for IT. Cloud computing can be one such opportunity to look at. Organization should learn about what is cloud technology – What are the ongoing industry trends in cloud, what is the assessment of maturity level of cloud technology and potential value it can generate for the organization.

The whole exercise should converge to answer the question if it is beneficial to consider adopting cloud and will it be a panacea for all the challenges and secure the vision of the organization to restructure its IT by its implementation?

*B. **(A)**sses*

Once the organization answers the previous question in affirmative this stage propels the organization to the following steps:

1) Understand: Understand in detail the technicalities of cloud, to begin with, what are the various cloud services delivery and cloud architecture models. Gaining knowledge of the architecture and delivery models of cloud is critical to understanding what would suit the organization needs.

2) Assess: An insight into organization's current service delivery model and architecture for delivering IT services is also required to assess the quantum of change before migrating to a cloud setup. To get a complete picture it is imperative to have a Technical and Financial assessment done to build a business case for cloud adoption.

a) Technical Assessment – A detailed technical readiness assessment of the entire IT landscape should be conducted to evaluate the readiness of the organization to adopt cloud. To begin with following pointers can give direction to carrying out a complete technical assessment –

- Infrastructure Assessment – Percentage of virtualized environment [As a thumb rule, higher the percentage of virtualized environment (infrastructure and application), more ready is the environment to be migrated to cloud]
- Security Assessment – Critical security requirements of the organization; this includes data & regulatory security requirements, user authentication and access controls
- Application Assessment – Identification and categorization of applications based on criticality, dependencies, age, customization etc.
- Processes Assessment – Governance process for deploying services and applications, change requests, regulating policies for backup & archival, audits, monitoring, managing and agreement of SLA's

b) Financial Assessment – To get a buy in from the decision makers it is important to conduct a comparative Total Cost of Ownership (TCO) study of 'As – Is' and 'To – Be' environment over a multiyear period. This should include both direct (hardware, software) and indirect costs (energy consumption, carbon footprint) incurred of running the environment. Also tangible (quicker deployment of services) and intangible benefits (faster go to market) would also accrue over a period of time should be given due consideration.

*C. **(S)**trategize*

This is a critical stage in which actual decision is made to move ahead with adoption of cloud platform. Following are the key points which can be discussed to formulate strategy for cloud adoption –

a) With huge investments made in data center space, hadware, software, licenses etc. is it prudent to migrate organizations applications to a cloud vendor OR is it advisable to build an own internal private cloud

b) Which are the potential applications to be migrated to cloud? What are the associated risks on migrating these applications to cloud?

c) Instead of experimenting within the organization would it be better to consult a cloud vendor BUT at the same time avoid vendor lock – in and retain the freedom of shifting to another in case of unsatisfactory service

d) How to convince the business or IT teams to accept the change coming their way and provide them comfort against any unforeseaable circumstances

Answers to above questions may vary across organizations. In case of a startup organization which may not have the financial strength to setup a dedicated IT environment to run its business it might be advisable to start its operations on a cloud and consult the vendor to build its skills over a period of time.

A large organization spread across the globe having few critical applications may not want to leverage cloud because of running the risk of security breaches or SLA commitments by the cloud vendor or the internal IT team. On the other hand it can leverage the expertise available within its own IT team to gradually build a private cloud environment of its own.

Hence, strategy needs to be developed based on the needs of the organization and should be evaluated on a case by case basis.

1) Prepare a Business Case – To enable decision making and formulate a strategy the organization should build a business case to evaluate benefits of moving to cloud. Ground work for building the business case would have already been initiated in the previous stage while conducting the Technical and Financial Assessment. Objective behind cloud adoption project and its alignment to overall IT and business strategy should be considered.

2) Lay down a Roadmap – Once the business case is approved a high level roadmap for cloud adoption can be laid down. The roadmap should categorize initiatives planned for short term, medium term and long term for adoption of cloud platform.

- Short Term – In the short term an organization could look at starting off with a pilot project
- Medium Term – If successful, it can derive confidence to roll out a private cloud platform across the organization in the medium term
- Long Term – In the long term it can move to hybrid cloud model by developing its internal cloud inside out and through which it can bring its business partners as well into its fold

Additionally, roadmap plan should be strengthened by a risk and mitigation strategy which should include rollback plans in case the rollout plan does not succeed.

*D. **(T)**estify*

The stage is to testify the strategy devised in the previous stage and implement the same by executing the pilot project. The stage involves migrating applications, identified for pilot project, into cloud. Applications should undergo following checks before being migrated to cloud –

a) Applications capable of running in virtual environments. (In case of certain applications where application code is bound with the hardware need to recoded to make them compatible to run in virtual environments)

b) Redeveloping highly customized applications to make it compatible to Software supported by Cloud (SaaS)

c) Replatforming to cloud compatible platform (PaaS)

d) Reporting to cloud compatible infrastructure (IaaS)

e) Repackaging the application in case of any porting changes done to the application

f) Testing the applications before finally migrating to cloud

Once the application has gone live in the cloud environment Business and IT users should provide a feedback on the response of the applications. Any lessons learnt or valuable feedbacks can be incorporated during the roll out of the cloud across the organization. Organization now can stabilize the cloud operations and monitor performance of cloud over a period of few weeks.

Pilot project's success will be the key in instilling confidence in the mind of business and IT team. Also if a vendor has been engaged during the consulting or implementation of the pilot project, the performance can be gauged.

*E. **(E)**stablish*

At this stage upon successful completion of the pilot project organization can establish new governance processes to monitor and manage cloud infrastructure and establish cloud sourced services and applications by rolling it out as per the roadmap defined in the 'Strategize' stage.

1) Enable – New established governance processes should look into ways of enabling wider adoption of cloud across the organization by integrating applications which are not yet on cloud by creating an interface between them. This would ensure smooth functioning of daily business until majority of applications are phased into cloud. Businesses departments can now be allocated resources based on their usage pattern.

2) Train – To further strengthen the cloud culture into the organization, training sessions can be conducted across the organization for business and end users which can help them derive further benefit from cloud.

3) Govern – Organization should look with cautious eye at the benefits associated with service provisioning through cloud. Reason being increase in 'Cloud Sprawl'. Cloud gives liberty to users to provision services at their will but it may turn to be disadvantageous if a 'Control' mechanism, which checks provisioning duplicated services from cloud, is not established through a proper governance process.

4) Delineate Roles & Responsibilities – With the kind of cloud service delivery model adopted the roles and responsibilities of running the cloud should be clearly defined between the consumer and the provider of the cloud. For example, on adoption of PaaS service delivery model the responsibility of decisions taken, managing and maintaining the cloud at and below this layer (Paas + IaaS) lies with the cloud provider (internal IT team or the cloud vendor). However, on the other hand responsibility of handling any IT related issues above PaaS layer rests with the consumer of the cloud (application developers and business users).

*F. **(R)**einvigorate*

This stage is an ongoing process to continuously improve, strengthen & sustain the established cloud platform through meeting & improving SLA's commitments, new cloud based services, increased availability to business & reinventing the cloud.

To sustain and improve cloud performance, benchmarking assessment should be done to assess all round cloud performance. Following performance parameters can benchmarked to assess cloud performance –

- Service Level Agreements (high availability, impregnable security, compliance, high performance)
- Key Performance Indicators (virtualization)
- Maturity level of cloud services offered based on industry benchmarks
- Comparative study on cloud offerings currently offered across organization and those offered across industry
- Unit cost incurred on providing cloud services

III. CONCLUSION

The paper makes an attempt to bring forth a methodological approach to adopt cloud highlighting various activities involved at each stage. Organizations can utilize this paper to carefully evaluate at each stage and give due consideration to vital steps involved at each stage.

The activities described in this paper are not exhaustive but can be used as leads to trigger discussion points amongst different stakeholders involved during cloud adoption lifecycle. Activities though generic in nature can be added onto and tailored as per organization needs to exploit full potential of cloud platform. We conclude with the hope that paper will contribute towards keeping the cloud computing at 'Peak of Rising Expectations' and prevent any eventualities which may lead to disillusionment.

ACKNOWLEDGMENT

I thank Mr. Shiwanand Pathak, for encouraging me through the process of writing this paper and providing valuable suggestions, which have helped in drafting of this paper.

REFERENCES

[1] Gartner Report on 'Hype Cycle for Emerging Technologies' in year 2009, 2010 and 2011

◈◈◈

Three Tier Data Storage Security In Cloud Using Face Fuzzy Vault

Vrushali Joshi, Payal Sanghvai & Yogita Bhargude

AISSMSIOIT, Pune, Maharashtra, India

Abstract - The well-known buzzword, say Cloud Computing, attracts the attention from both academy and industry across the world, because of its potential in achieving the long-held dream, say Utility Computing and opens a new era for new spawned companies to provide services above this new concept[6],[8]. At the same time hackers have also turned their faces to cloud to breach the security. The best software solution provides utilization of cloud computing and take full advantage of it with best security solutions and assurance. To achieve this goal proposed system provides the following solution, which uses three tier security in cloud environment. Set of biometric features are first extracted from user's face image, The extracted features are then quantized and mapped to binary representation for feature point matching. The produced features and the secret key (which will restrict unauthorized access) are bound using Face fuzzy vault scheme .During authentication key will be correctly retrieved if face vault matches. Also to prevent data from cloud service provider, data is encrypted while saving it on the cloud using Confidentiality, Integrity, Availability (CIA) values which will categorize data in three rings. While providing data if user is authorized then the concerned password gets extracted from the fuzzy vault and the extracted password will help to decide which user belongs to which data ring.

Keywords - *Cloud Computing, Face fuzzy vault, CIA, Data ring, Three tier security.*

I. INTRODUCTION

Today a growing number of companies have to process huge amounts of data in a cost-efficient manner. Classic representatives for these companies are operators of Internet search engines, like Google, Yahoo, or Microsoft. The vast amount of data they have to deal with every day has made traditional database solutions prohibitively expensive. Thus the cloud is best suitable for above requirements.

One of the cloud services Data as a Service (DaaS) is capable in order to simplify the development of distributed applications. On top of such architectures, many of these companies have also built customized data processing frameworks. Cloud computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [7].Cloud provides key features such as on-demand self-service, broad network access, virtualization, Resource pooling and data sharing. Different Service Models available which are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS), Data as a Service (DaaS). The delivery of virtualized storage on demand becomes a separate Cloud service - data storage service. Deployment Models are: Private cloud: Enterprise owned or leased, Community cloud shared infrastructure for specific community, Public cloud: Sold to the public, mega-scale infrastructure, Hybrid cloud: Composition of two or more clouds. Now a day cloud computing make everything flexible and easier but there is another aspect that is the security issue. According to the survey, some companies states that due to cloud computing it becomes easier for bad guys to focus their effort and breach hundred and thousands of record. There is no security rating system in place for cloud computing, so business users can't rely on third party security mechanism. Risk factor with cloud computing are high because level of security provided by cloud provider are not same. CRYPTOGRAPHIC techniques [1] are being widely used for ensuring the secrecy and authenticity of information .The security relies on the cryptographic keys and maintaining the secrecy of keys is one of the main challenges in practical cryptosystems. However, passwords can be easily lost, stolen, forgotten, or guessed using social engineering and dictionary attacks. Limitations of password-based authentication can be alleviated by using stronger authentication schemes, such as biometrics. Biometric systems [3],[4],[5]establish the identity of a person based on his or her anatomical or behavioral traits, such as face, fingerprint, iris, voice, etc.

A critical issue in biometric systems is to protect the template of a user which is typically stored in a database or a smart card. The Face fuzzy vault construct is a biometric system that secures both the secret key and the biometric template by binding them in single template to provide unique identification. According to literature survey, we found that To provide security in private cloud environment following framework is suggested [9].As shown in figure1 when user sends request along with username to access the data to cloud provider, the cloud provider first checks in which ring requested data belong. If authentication is required, it first checks the username in its own directory for existence, if the username does not exist it ask the user to register itself. If the username matches, it redirects the request to company for authentication, then the user sends password for authentication, and after authentication it redirects the request to cloud provider to access resource.

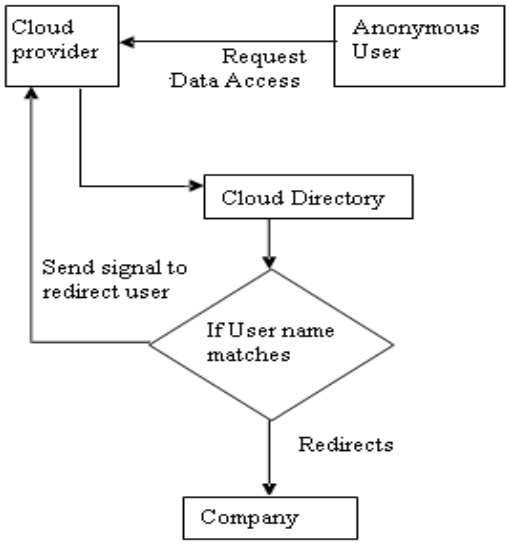


Fig. 1 : Security with redirection

Above framework introduces redirection for security which will increase time, space and maintenance cost. Proposed system will provide secure framework without redirection which is shown in figure 2.Juels and Sudan [2] proposed a cryptographic construction called a fuzzy vault that operates in the key binding mode and, in principle, can compensate for intra class variations in the biometric data. They have presented a fully automatic implementation of the fuzzy vault scheme based on fingerprint minutiae. In the SonglongYuana ‘s [10]online authentication model, at the registration stage, fuzzy vault encoding is implemented using both key and transformed template. And then the fuzzy vault is encrypted using digital signature. Face extraction can be implemented [11], [13] with PCA algorithm, but face may get injured due to any mishaps and then the user may lose his/her identity. Considering all available techniques and related issues, proposed system will provide one of the solutions to handle different challenges that we found in survey. To solve security issue proposed system will provide THREE TIER SECURITY which is as follow.

TIER 1: Focuses on the problem of data leakage the data classification is done by company before storing the data. This classification is done on the basis of CIA (Confidentiality, Integrity, Availability).The value of C is based on level of secrecy at each junction of data processing and prevents unauthorized disclosure, value of I based on how much assurance of accuracy is provided, reliability of information and unauthorized modification is required, and value of A is based on how frequently it is accessible. With the help of proposed formula, the priority rating is calculated, accordingly data having the higher rating is considered to be critical and Three Tier security is recommend which will provide higher security.

TIER 2: Clients are also categorized to avoid unauthorized access. Classified data is distributed in three rings. Clients are categorized according to these rings, same priority clients belong to same ring. To provide unique access each ring will contain a secrete key.

TIER 3: Face fuzzy vault is the technique which binds ring secrete key with user’s face feature to provide controlled data access to authorized user i.e. fuzzy vault provides unique identification. In this paper Section1 provides abstract of system.Section2 represents domain specific analysis and literature survey. Section 3 represents proposed techniques to provide security. Section 4 represents conclusion and section 5 is to represent references.

II. PROPOSEDSYSTEM:

As discussed earlier [9] general framework provide redirection (figure 1) but to handle security risk, proposed system helps without redirection. When the user requests for data access, user authentication and authorization issue gets solved at cloud only through worker role which is running at cloud and get controlled through company. Company is responsible to set CIA values for data classification while saving it on cloud proposed framework is shown in figure 2.

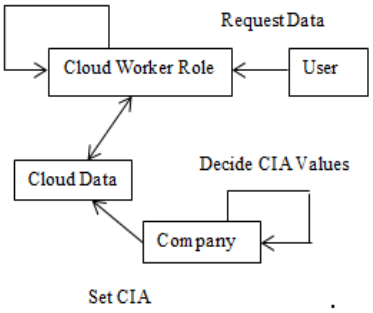


Fig. 2 : Security without redirection

2.1 CIA Technique For Ring Selection:

Let D be the set of 'n' documents to be stored on cloud database Data D= {1, 2, ……., n}Let C be the value of confidentiality where C={C|c ε N} where N is any natural number, Let I be the value of Integrity where I={I|IЄN},

Let A be the value of availability where A= {A|A Є N}1. Input: Data, protection ring, D[] array of n integer size. Array C, I, A, S, R of n integer size.2. Output: categorized data for corresponding ring. For i 1 to n C [i] = Value of Confidentiality. I [i] = Value of Integrity. A [i] =Value of Availability.

Calculate: S=Selection of Ring according to the given formula below: S (i) = [C (i) +A (i) +I (i)]/3, Let S= {1, 2, 3},Let R be ring where R= {1, 2, 3} If 0> S (i) <33Select Ring3 If 33≥ s (i) ≤66Select Ring2 Else if 66≥S(i)≤100 Select Ring1

Analysis: After applying algorithm for categorizing the data on the basis of sensitivity. Now ring rule and conflict of interest is applied in the ring to make more robust security system. Ring rules restrict data access.

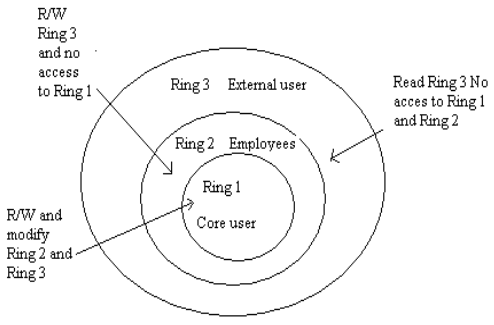


Fig. 3 : Ring Representation

Ring rule:

1. The users granted access to Ring3 is not allowed to access any R/W in lower rings.
2. The user granted access to Ring1 is allowed to access upper rings i.e. Ring 2 and Ring3 (Add, delete modify)
3. The users granted access to Ring2 is allowed to read data from ring 3 but cannot modify existing data.

2.2 Face Fuzzy Vault Encoding.

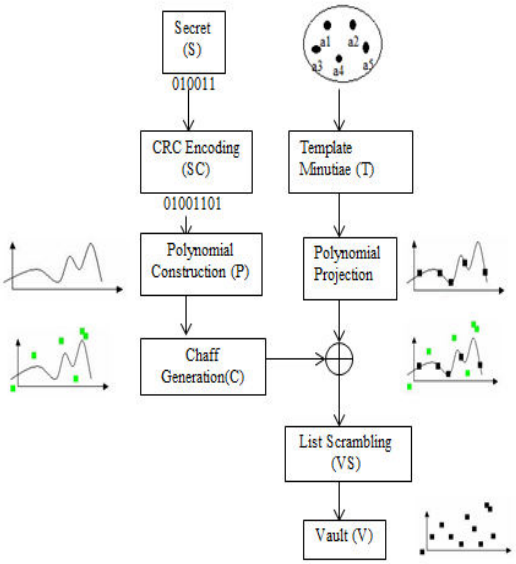


Fig. 4 : Face fuzzy vault encoding

Let Secret 'S' be 'cloud' which is to be bound with the face feature template. The characters in the secret are converted into binary format and the binary value is encoded with 16-bit Cyclic Redundancy Check (CRC) encoding. We generate 16-bit CRC data from the secret S. The 16-bit primitive polynomial we use for CRC generation, $g_{CRC}(a) = a^{16} + a^{15} + a^2 + 1$, is called "CRC-16".Appending the 16 CRC bits to the original secret S (128-bits), we construct 144-bit data SC. From this point on, all operations take place in Finite field $F(2^{16})$. We concatenate x and y coordinates of a minutia (8-bits each) as[x|y] to arrive at the 16-bit locking/unlocking data unit u. The bit string SC is used to find the coefficients of the polynomial p: 144-bit SC can be represented as a polynomial with 9 (144/16) coefficients, with degree D = 8.p(u) = $c_8u^8 + c_7u^7 +$ …… $+ c_1u + c_0$.In other words, SC is divided into non-overlapping 16-bitsegments, and each segment is declared as a specific coefficient c_i , i = 0, 1,2,…..,8. Two sets composed of point pairs are generated. The first one, called genuine set G, is found by evaluating p(u)on the template minutiae (T). Assuming that we

have N template minutiae (if we have more than N minutia, we choose the first N sorted according to ascending u values),$u_1,u_2,\ldots\ldots\ldots\ldots\ldots\ldots u_N$, we construct $G\{(u_1,p(u_1),(u_2,p(u_2),\ldots\ldots\ldots\ldots\ldots(u_n\ p(u_n)\}$.Note that the template minutiae, $u_1,u_2,\ldots\ldots$, u_N are selected to be unique, namely, $u_i \neq u_k$, if $i \neq k$, where i = 1,2,……,N, k = 1, 2, ……..,N. The second set, called the chaff set C, ensures the security of the system. To add M chaff points, we first generate M unique random points, c_1, c2,……,c_M in the field $F(2^{16})$, with the constraint that they do not overlap withu_1, u_2, ……..,u_N. Then, we generate another set of M random points, d1, d2,….,d_M, with the constraint that the pairs (c_j ,d_j), j = 1,2, …..,M do not fall onto the polynomial p(u). The chaff set C is defined as $C\{(c_1,d_1),(c_2,d_2),\ldots\ldots\ldots\ldots..(c_n,d_n)\}$The union of the genuine and chaff sets, $G \cup C$, is finally passed through a list scrambler that randomizes the list, with the aim of removing any stray information that can be used to separate chaff points from genuine points. This results in the vault set V

$S\{(v_1,w_1),(v_2,w_2),\ldots\ldots\ldots\ldots..(v_{N+M},w_{N+M})\}$.

2.3 Face Fuzzy Vault Decoding:

A user tries to unlock the vault V using N query minutiae Q ={ }. The points to be used in polynomial reconstruction are found by comparing, i =1,2,……,N with the abscissa values of the vault V , namely v_l, l = 1, 2,… (N + M): if any , i =1,2,…..,Nis equal to v_l; l = 1, 2,. .., (N + M), the corresponding vault point (v_l,w_l) is added to the list of points to be used during decoding.

Assume that this list has K points, where $K \leq N$. For decoding a degree D polynomial, (D + 1) unique projections are necessary. We find all possible combinations of (D + 1) points, among the list with size K, resulting in (K/(D+1)) combinations. For each of these combinations, we construct the Lagrange interpolating polynomial. For a specific combination set given as L{(v1,w1), (v2,w2),……., (vD+1,wD+1)}. the corresponding polynomial is

$$p^*(u)=\frac{(u\ ¡\ v2)(u\ ¡\ v3)\cdots(u\ ¡\ vD+1)}{(v1\ ¡\ v2)(v1\ ¡\ v3)\cdots(v1\ ¡\ vD+1)}w_1\ldots\ldots\ldots+\frac{(u\ ¡\ v1)(u\ ¡\ v2)\cdots(u\ ¡\ vD)}{(vD+1\ ¡\ v1)(vD+1\ ¡\ v2)\cdots(vD+1\ ¡\ vD)}w_{D+1}$$

This calculation is carried out in F(216), and yields $P^*(u)=C_8^* U^8+C_7^* U^7+\ldots\ldots+C_1^* U+C_0^*$.

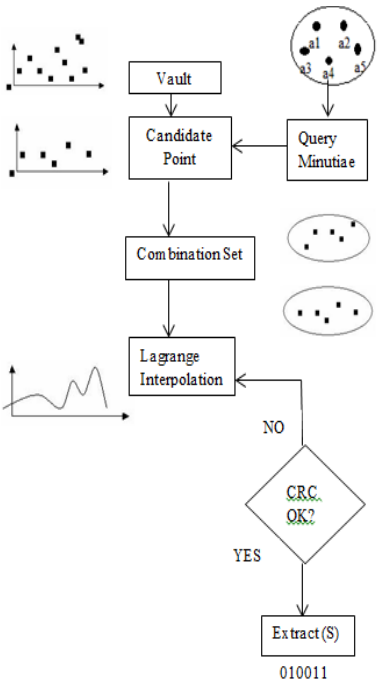


Fig. 5 : Face fuzzy vault decoding

The coefficients are mapped back to the decoded secret SC*. For checking whether there are errors in this secret, we divide the polynomial corresponding to SC*with the CRC primitive polynomial, gCRC(a) = a16 + a15 + a2 +1. Due to the definition of CRC, if the remainder is not zero, we are certain that there are errors. If the remainder is zero, with very high probability, there

2.4 Face Extraction:

Face is captured through Webcam and stored in a specified size required. The stored image is in RGB color format. The RGB format is converted to YUV chrominance format for easy extraction of face features. The skin color is detected by skin detection algorithm and the face is cropped by applying right, left, up and bottom boundaries. The cropped face is converted to grayscale so that during decoding the features are matched correctly. The features are then extracted by applying Scanning Edge Detection Algorithm to the grayscale image.

2.5 User Registration:

Company is using cloud services and data is categorized in three rings according to CIA, and the company will assign secrete key to each ring. Users are of three type Ring1 includes Core user (e.g. CEO, higher authority etc), Ring2 include company employees and Ring 3 includes external or anonymous users. When first time the user visits the company, the user has to register with the company .For that he/she provide face image and required details. Company checks user details and necessary constraints then allows the user to become part of company .According to the type of user company generates User ID(UID)which will categorize users in three layers. If UID is generated successfully then fuzzy vault encoding takes place i.e. the secret and extracted face features get combined in vault. After this users UID along with face fuzzy vault get saved on cloud database.

2.6 User Data Access:

At the time of data access user must provide the UID received during registration and face image. Firstly users face gets encrypted to ensure image gets securely travelled through internet. At the worker process, received UID gets checked with saved database. If the UID is available, saved fuzzy vault gets compared with received template, if the match is found, secret key of particular user's ring gets extracted during fuzzy vault decoding and accordingly controlled data access can be achieved. Figure represents Data access flow.

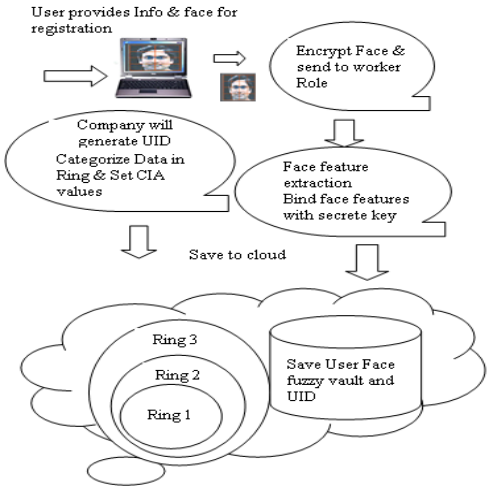


Fig. 6 : User Registration process

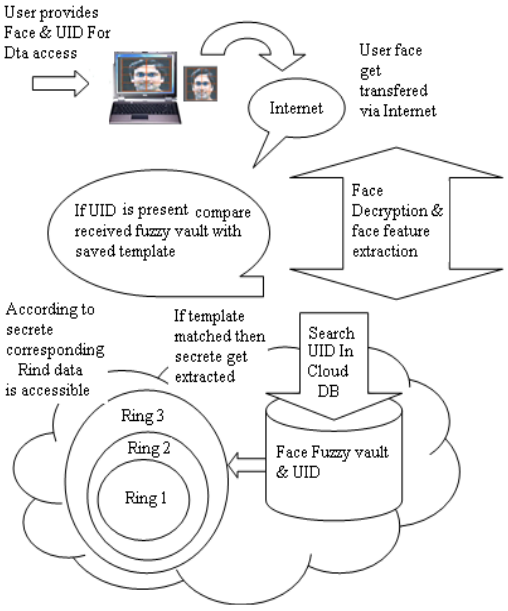


Fig.7 : User data access

III. CONCLUSION:

We will provide one of the solution to secure data stored on cloud using face fuzzy vault. The data on cloud is arranged in three layers according to CIA and accessed by authorized user of the particular layer. Hence, the data is protected from any modifications or misuse by the service provider as well as unauthorized user.

REFERENCES:

[1] Karthik Nandakumar, Student Member, IEEE, Anil K. Jain, Fellow, IEEE, and Sharath Pankanti, Senior Member, IEEE, "Fingerprint-Based Fuzzy Vault:Implementation and Performance", IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 2, NO. 4, DECEMBER 2007.

[2] A. Juels and M. Sudan. "A fuzzy vault scheme," in Proc. IEEE Int. Symp. Inform. Theory, Lausanne, Switzerland, 2002, p. 408.

[3] Dacheng Xu School of Applied Technology Soochow University Suzhou 215006, Wang School of Electronics and Information Engineering Soochow University Suzhou 215006, "A Scheme for Cancelable Fingerprint Fuzzy Vault Based on Chaotic Sequence", Proceedings of the 2010 IEEE

International Conference on Mechatronics and Automation August 4-7, 2010, Xi'an,China.

[4] Hosik Sohn, Yong Man Ro, Senior Member, IEEE, and Kostantinos N. Plataniotis, Senior Member, IEEE," Content Sharing between Home Networks by using PersonalInformation and Associated Fuzzy Vault Scheme ", IEEE Transactions on Consumer Electronics, Vol. 55, No. 2, MAY 2009.

[5] Yongjin Wang, KN. Plataniotis The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, ON, Canada, M5S 3G4, "FUZZY VAULT FOR FACE BASED CRYPTOGRAPHIC KEY GENERATION", 2007 Biometrics Symposium,IEEE-2007

[6] Yi Wei and M. Brian Blake, University of Notre Dame, "Service-Oriented Computing and Cloud Computing Challenges and Opportunities", Published by the IEEE Computer Society November/December2010

[7] Tharam Dillon, chen.wu, change Digital Ecosystems and Business Intelligence Institute Curtin University of Technology Perth, Australia, "Cloud Computing: Issues and Challenges ", 2010 24th IEEE International Conference on Advanced Information Networking&Applications

[8] Michael Kretzschmar and Sebastian Hanigk Universität der Bundeswehr München, Institut für Technische Informatik, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany "Security Management interoperability challenges for Collaborative Clouds", 2010 4th International DMTF Academic Alliance Workshop on Systems&Virtualization.Management

[9] Abhishek, utkarsh chaurasiast, badrinathojha, rajeevranjanshahi, Indian Institute of Information Technology, Allahabad U.P India. "3 Dimensional Security in Cloud Computing" 2011 IEEE.

[10] Lifang Wua,b Songlong Yuana aSchool of Electronic Information and Control Engineering, Beijing University of Technology, Beijing China 100124 Dept of Computer science and Engineering, State University of New York at Buffalo, Buffalo, New York USA 14260," A face based fuzzy vault scheme for secure online authentication", 2010 Second International Symposium on Data, Privacy, and E-Commerce

[11] V. Evelyn Brindha Associate Professor, Dept. of CSE., Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India. "Biometric Template Security using Fuzzy Vault", 2011 IEEE 15th International Symposium on Consumer Electronics

[12] Scott Dowell Computer Sciences Corporation San Diego, CA, USA. Albert Barreto III, James Bret Michael,and Man-Tak Shing Naval Postgraduate School Monterey, CA, U.S.A"Cloud to Cloud Interoperability", Proc. of the 2011 6th International Conference on System of Systems Engineering, Albuquerque, New Mexico, USA-June-27-30,2011

[13] Qian Wang, Student Member, IEEE, Cong Wang, Student Member, IEEE, Kui Ren, Member, IEEE, Wenjing Lou, Senior Member, IEEE, and Jin Li."Biometric Template Security using Fuzzy Vault", 2011 IEEE 15th International Symposium on ConsumerElectronics

◈◈◈

Challenges on Amazon Cloud in Load Balancing, Performance Testing and Upgrading

Himanshu Shah, Paresh Wankhede & Anup Borkar

Accenture Services Pvt. Ltd., Mumbai, India

Abstract **-** Web application hosting in a data centre is clouded with quite a few issues ranging from hardware provisioning, software installation and maintaining the servers. Traditional data-centre techniques need production grade hardware to test application's behavior/performance under expected peak load. This could be costly and procuring hardware could be very time consuming causing delays in software delivery. Cloud (Infrastructure-as-a-Service) can be an answer to this. Cloud Computing provides production grade server instances at very cheap rates.

This whitepaper is divided into two sub parts: first part details out the typical web application setup on Amazon Web Services cloud (AWS) [Ref 2], challenges faced during the setup and resolution for the same, while the second part talks about the observations made during load testing using Apache JMeter performance testing tool on AWS cloud. Three different application setup topologies (single tier, two tier and three tier) are tested and findings and learning from it are discussed here.

This whitepaper only highlights the pitfalls encountered and possible resolutions for each and is not a comment on performance of Amazon cloud. The whitepaper endeavors to find out the best architecture which would give maximum return on investment.

I. HOSTING A WEB APPLICATION ON CLOUD

Organizations are lured into moving onto cloud from traditional data centre to reap the benefits of its agility, elasticity, cost-efficiency and security.

A. Issues with traditional Infrastructure

There are some issues with traditional infrastructure which increases application's 'time-to-market' parameter. Those issues could be:

1) *Hardware provisioning*
2) *Installation of software*
3) *Prediction of expected usage/load.*
4) *Scale up and Scale down*
5) *Hardware maintenance*

The best way to deal with these could be to move the application on Cloud. The subsequent sections talk about deploying a typical web application on AWS cloud with various topologies. Along with it are the recommendations based on performance parameter.

B. Typical Web application topology

A typical web application could take one of the following forms:

1) *Single tier (all in one)*
2) *Multi tier*

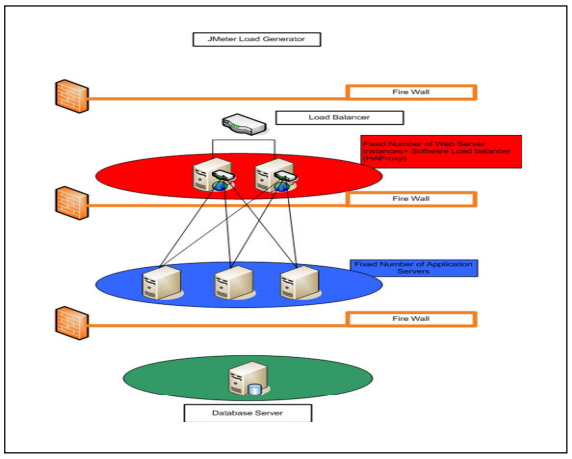


Fig. 1: Web application deployment topology in data-centre

In a typical data-centre, hardware sizing is predefined according to the expected peak request load.

This approach generally leads to over-sizing or under-sizing of the hardware. Apart from this, server hardware and software maintenance is an overhead. AWS cuts through these traditional challenges to provide elasticity and agility. AWS reduces the capital expenditure required for initial hardware setup (AWS provides pay-as-you-go model with no upfront investment).

C. Web application on AWS

Organization has been skeptical about application migration on cloud. Reasons could be one or more of the following:

1) Data security on public cloud

2) Performance issue of an application

3) Loosing control over infrastructure

4) Vendor lock-in

There are various documents and whitepapers that talk about the data security on cloud. Performance testing of a web application on cloud is something that has not been touched upon. We have addressed this topic by deploying a simple web application and testing it for variety of load with Apache JMeter load testing tool.

A simple shopping cart application is hosted on AWS in following topologies:

1) All In One: web server, application server and db server are installed on the same Red Hat Enterprise Linux (RHEL) instance. Refer row number 1 in TABLE I below.

2) Two tier: Web server and application server are installed on same instance and DB server is installed on a different instance. The web + application server instance is configured to auto-scale [Ref 4] as per the request load on the server. Refer row number 2 in TABLE I below.

3) Three tier: Web server, application server and DB server are installed on three separate instances. Both the Web server and Application server instances are configured to auto-scale according to the request load. Refer row number 3 in TABLE I below.

TABLE I : WEB APPLICATION DEPLOYMENT TOPOLOGIES

Scenario	ELB	Web Server	App Server	DB
1		All in One		
2	Yes	Combined, Auto-Scaled		One Instance
3	Yes	Auto-Scaled	Auto-Scaled	One Instance

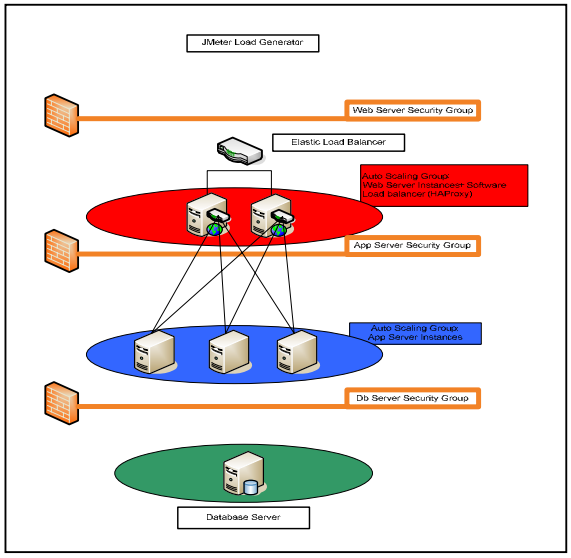


Fig. 2: Web Application Deployment (auto scaling with ELB and HAProxy)

The auto scaled environment is configured to work with minimum two instances. This can grow up to twenty instances depending on the load. The software load balancer (HAProxy) is installed on the same server instance which is hosting the web server. A custom component is deployed on each of the HAProxy server instances. This component is responsible for the discovery of newly instantiated application server instances. Once discovered the custom component will register those app server instances with HAProxy server. However there would be only one DB server instance at any given point in time. All the application servers would be pointing to the same DB server instance.

D. Components used

1) Elastic Load Balancer (ELB)

ELB [Ref1] has been used for scenario 2 and scenario 3 as a public interface for web application. It redirects the requests to web server.

2) Virtual Private Cloud (VPC)

VPC [Ref7] is required to create a virtual network subnet which is the key component for setting up the environment for Apache JMeter load testing. This has been used in all of the scenarios.

3) HAProxy

HAProxy [Ref3] is used to balance the load between web server and app server in scenario 3. This is used along with custom component for dynamic node discovery.

4) Elastic Ip

Elastic IP [Ref8] is used in scenario 2 and scenario 3. It is associated with database server and the same is configured in application server.

5) Auto-Scaling Group

Auto-Scaling Group [Ref4] is used for the purpose of application scale out and scale in. This feature is used in conjunction with ELB and HAProxy in scenario 2 and scenario 3.

6) Security Groups

Security groups [Ref 9] act as a virtual firewall. It's a collection of access rules that defines the ingress network traffic to the instance.

II. LEARNING AND RECOMMENDATIONS

Following is the learning outcome from this exercise:

A. Load Balancers

ELB does not have a private IP and cannot be wrapped around AWS security group.

Description

ELB does not have a private IP and security group cannot be wrapped around it, which exposes it and makes it open-for-all to access. This behaviour does not make a sound proposition for inter-server communication between web server and app server.

Recommendation

A software load balancer, like HAProxy [Ref 3], becomes a candidate for load balancing the application server layer. These HAProxy instances being normal EC2 instances, has private IP addresses and could be protected by security group thus making them protected from unauthorized access.

B. Auto Discovery of EC2 nodes

Task is to auto discover the newly instantiated application servers on AWS cloud.

Description

Application servers being part of auto-scaling group are configured to scale out and scale in dynamically. There are no provisions in HAProxy to find out new servers and configure itself accordingly to handle new servers.

Recommendation

To use HAProxy in tandem with auto-scaled application server a custom component was developed which discovers the newly instantiated app server instance and updates the HAProxy configuration file with app server instance information. These auto-scaled application servers, once dynamically registered with HAProxy, can be load balanced by the HAProxy instance.

III. PERFORMANCE TESTING ENVIRONMENT

Performance Testing is an important activity of software life cycle. It ensures that the delivered software meets customer's expectations - varying load, availability and scalability. In a virtualized environment such as cloud, testing the deployed application for its performance becomes an important aspect. This section talks about the observations made during load testing using Apache JMeter performance testing tool on AWS. It lists out the parameter and corresponding values on which the comparison of various web application topologies under auto-scaled environment is done.

This section delineates the steps that were followed to perform load testing. Following points are discussed:

1) Technical details about software and hardware infrastructure.

2) Strategy and parameters used while generating load for application.

3) Test results

4) Observations from test results

Following diagram (Figure 3) depicts the load testing strategy on AWS. It includes multiple Apache JMeter servers and one UI client.

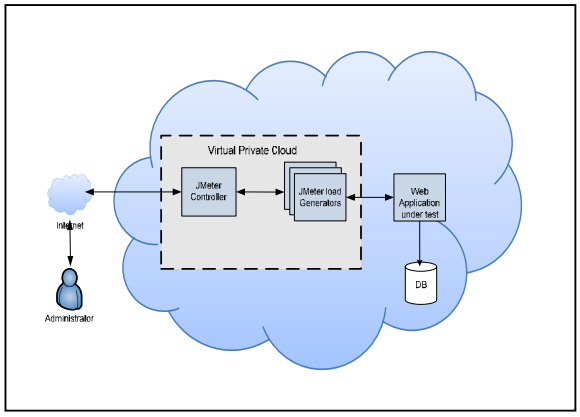


Fig. 3 : Apache JMeter setup in Amazon Cloud

A. APPLICATION SETUP

A simple web application was used for performance testing. Following are the details of software and hardware components:

1) *EC2 Instance type [Ref6]– Small (1Elastic Compute Unit, Single Core, 1.7GB RAM)*
2) *Application Server – Tomcat 6*
3) *Database server – MySql 5.5*
4) *Web Server – Apache 2*
5) *Web Application*
6) *Customization done for web application –*
7) *All application log level is set to FATAL.*
8) *App server JVM tuning*
9) *Software Load balancer - HAProxy 1.4*
10) *Performance Testing Tool – Apache JMeter*

B. LOAD GENERATION STRATEGY

Following points were taken into consideration:

1) *Number of concurrent requests –*

a) Scenario 1 (single tier)–

This scenario was tested by sending 25, 50, 75 and 100 concurrent requests to the application service.

b) Scenario 2 onwards (multi tier)–

Both these test scenarios were tested by sending 25, 50, 75 and 100 concurrent requests to the application service.

2) *Duration of test: 10 minutes.*

C. Measurable Parameters

Following are the measurable parameters that are taken into consideration:

1) *Throughput.*
2) *Overall Response Time.*

D. TEST RESULTS

Following figure (Fig 4) is a chart with Number of users plotted on X-axis vs. Response time plotted on Y-axis.

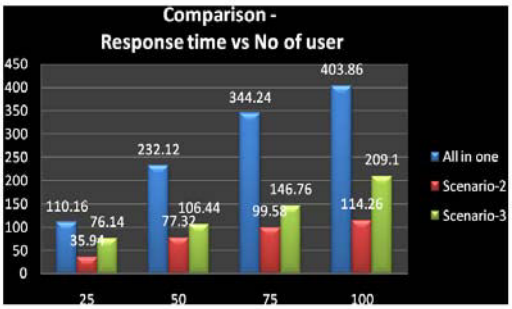


Fig. 4 : Number of users vs. response time

The chart above clearly shows that the response time is at the minimum for scenario 2 while it is at the maximum in case of scenario 1.

Following figure (Fig 5) is a chart with Number of users plotted on X-axis vs. Throughput plotted on Y-axis

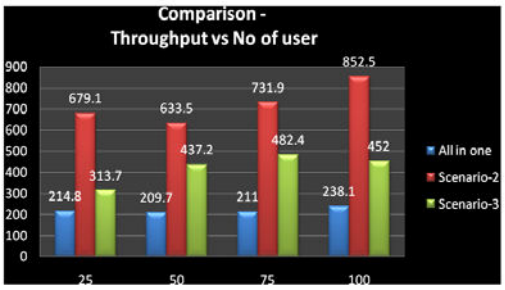


Fig. 5 : Number of users vs. Throughput

From the chart shown in Figure 5 it is clear that scenario 2 throughput > scenario 3 throughput is the best suitable topology:

scenario 2 throughput > scenario 3 throughput

Thus, keeping Apache web server and application server on the same instance and auto-scaling this server instance is capable of handling more load than the topology where web server and app server are hosted on separate instances.

TABLE II : COMPARISON CHART FOR 3 SCENARIOS

Scenario No.	Number of threads	Throughput	Avg. Throughput	Availability	Configurability	Fault-tolerance	Maintainability	Scalability
1	25	214.8	218.40	No failover capability	Easy	No	Easy	Not Scalable
	50	209.7						
	75	211						
	100	238.1						
2	25	679.1	724.25	Highly Available	Moderate	Yes	Moderate	Scalable

	50	633.5						
	75	731.9						
	100	852.5						
3	25	313.7	421.32	Highly Available	Difficult	Yes	Difficult	Scalable
	50	437.2						
	75	482.4						
	100	452						

E. Observations

1) Scenario 1: It gives moderate performance as compared to other 2 scenarios. All 3 servers are installed on same instance causing the degradation in performance and hence the response time. Though the configuration and maintenance of this setup is very easy, it is devoid of any failover capability. At any given point in time there would only be one sever which would be serving all the requests to the application and as the number of requests increases the performance of an application would go on decreasing.

2) Scenario 2 : From the table above it is clear that Scenario 2 gives the best throughput. Reason being; the database server is kept on separate instance which helps reduce the load on single server, while the web + application server instance handles the request load, which in turn increased the request handling capacity of a server. The availability for this scenario is better than scenario one. Reason being layer 1 was auto-scaled and registered with ELB. Scenario 2 is easier to configure in a sense that it does not require additional software load-balancer and a custom component to be installed for node discovery.

3) Scenario 3 : It gives reduced throughput as compared to Scenario 2. Though the demarcation of servers can be seen here the reduced throughput is due to the intricacy developed by network traffic between web server instance, application server instance and db server instance. Along with this, the time taken for distribution of load at ELB level and HAProxy level also contributes to the cause. Scenario 3 is high on availability due to auto-scaled layer 2 and 3. Software load-balancer and custom component makes configuration and administration difficult as compared to scenario 1 and 2.

F. Recommendation

For different type of web applications, the three scenarios that have been documented in this whitepaper may serve as a benchmark. Here is a list of recommendation:

TABLE III : SCENARIO RECOMMENDATIONS

Application Scenarios	Recommended Scenario	Examples
High performance	Scenario 2	Search engines, Data reporting
High availability	Scenario 3	E-commerce, High data streaming applications
High maintainability	Scenario 1	Organization intranet sites

IV. LEARNING

Following is the learning outcome from this exercise:

A. Network Subnet

The task is to create a single network subnet on AWS cloud.

Description

Apache JMeter works on RMI protocol. RMI cannot communicate across subnets without a proxy. There is no guarantee that all the instances created on AWS public cloud will be on the same subnet. This creates a road block to use JMeter.

Recommendation

Using Virtual Private Cloud [Ref7] in AWS makes sure that all the instances created within the VPC have the same subnet. JMeter load generators and controllers are placed inside the VPC so that they can generate the distributed load for web application on AWS cloud.

V. CONCLUSION

Hosting a web application on cloud does addresses the issues involved with traditional data centre, such as hardware provisioning, software installations and infrastructure maintenance. There were certain challenges that were faced while setting up an environment which were solved either by using software

or by developing custom components. The hurdle around security with Elastic Load Balancers could be tackled by implementing custom logic along with using software load balancers.

Creating a network subnet which is an integral part for Apache JMeter configuration can be handled by using AWS VPC. As far as the performance of an application is concerned, web application deployed with two tier topology gives optimum throughput on AWS cloud when compared with single tier and 3 tier setup.

Unless and until there is a specific requirement to go for three tier setup, two tier setup is optimal solution. It's fault-tolerant, highly available, easy to configure and maintain, and scalable. Scenario 3 is difficult to configure, due to additional overhead of installing HAProxy and custom component for node discovery. It also lags in performance due to network traffic and time taken by request to route thorough HAProxy.

REFERENCES AND FURTHER READING

[1] http://aws.amazon.com/elasticloadbalancing/

[2] http://aws.amazon.com/ec2/

[3] http://haproxy.1wt.eu/

[4] http://aws.amazon.com/autoscaling/

[5] http://jakarta.apache.org/jmeter/usermanual/remote-test.html

[6] http://aws.amazon.com/ec2/instance-types/

[7] http://aws.amazon.com/vpc/

[8] http://aws.amazon.com/articles/1346

[9] http://docs.amazonwebservices.com/AWSEC2/2007-08-29/DeveloperGuide/distributed-firewall-concepts.html

◈◈◈

Enhancement for Data Security in Cloud Computing Environment

Gawali M. B.[1], R. B. Wagh[2] & S. P. Patil[3]

[1,2]Computer Engineering Dept., RCPIT , Shirpur, India, [3]Computer Engineering Dept. GECA, Aurangabad, India

Abstract **-** Cloud computing, a rapidly developing information technology, has aroused the concern of the whole world. Cloud computing is Internet-based computing, whereby shared resources, software and information, are provided to computers and devices on-demand, like the electricity grid. Cloud computing is the product of the fusion of traditional computing technology and network technology like grid computing, distributed computing parallel computing and so on. It aims to construct a perfect system with powerful computing capability through a large number of relatively low-cost computing entity, and using the advanced business model like SaaS (Software as a Service) to distribute the powerful computing capacity to end users' hands. To address this longstanding limitation by building a multi-tenant system. Our system provides the environment for the user to perform his tasks, but with very high security. By using further facilities provided in this system user fill secure about his data and his account.

Keywords **-** *PaaS, IaaS, SaaS, Multi – tenant, MaaS etc.*

I. INTRODUCTION

A. The Cloud Computing

The term *cloud* has been used historically as a metaphor for the Internet. This usage was originally Derived from its common depiction in network diagrams as an outline of a cloud, used to represent the transport of data across carrier backbones (which owned the cloud) to an endpoint location on the other side of the cloud. This concept dates back as early as 1961, when Professor John McCarthy suggested that computer time-sharing technology might lead to a future where computing power and even specific applications might be sold through a utility-type business model. This idea became very popular in the late 1960s, but by the mid-1970s the idea faded away when it became clear that the IT-related technologies of the day were unable to sustain such a futuristic computing model. However, since the turn of the millennium, the concept has been revitalized. It was during this time of revitalization that the term *cloud computing* began to emerge in technology circles [1].

B. The Emergence of Cloud Computing

Utility computing can be defined as the provision of computational and storage resources as a metered service, similar to those provided by a traditional public utility company. This, of course, is not a new idea. This form of computing is growing in popularity, however, as companies have begun to extend the model to a cloud computing paradigm providing virtual servers that IT departments and users can access on demand. Early enterprise adopters used utility computing mainly for non-mission-critical needs, but that is quickly changing as trust and reliability issues are resolved. Some people think cloud computing is the next big thing in the world of IT. Others believe it is just another variation of the utility computing model that has been repackaged in this decade as something new and cool. However, it is not just the buzzword "cloud computing" that is causing confusion among the masses. Currently, with so few cloud computing vendors actually practicing this form of technology and also almost every analyst from every research organization in the country defining the term differently, the meaning of the term has become very nebulous. Even among those who think they understand it, definitions vary, and most of those definitions are hazy at best. However, when "the cloud" is combined with "computing," it causes a lot of confusion. Market research analysts and technology vendors alike tend to define cloud computing very narrowly, as a new type of utility computing that basically uses virtual servers that have been made available to third parties via the Internet. Others tend to define the term using a very broad, all-encompassing application of the virtual computing platform. They contend that anything beyond the firewall perimeter is in the cloud. A more tempered view of cloud computing considers it the delivery of computational resources from a location other than the one from which you are computing.

C. SaaS Multi Tenant-

The architecture of SaaS-based applications is specifically designed to support many concurrent users (multi tenancy) at once. This is a big difference from the traditional client/server or application service provider (ASP)- based solutions that cater to a contained audience. SaaS providers, on the other hand, leverage enormous economies of scale in the deployment, management, support, and maintenance of their

offerings. Multi-tenancy is an organizational approach for SaaS applications. Although SaaS is primarily perceived as a business model, its introduction has lead to numerous interesting problems and research in software engineering. Despite the growing body of research in this area, multi-tenancy is still relatively unexplored, despite the fact the concept of multitenancy first came to light around 2005. While a number of definitions of a multi-tenant application exist, they remain quite vague. Therefore, we define a multi-tenant application as the following:

Definition 1. A multi-tenant application lets customers (tenants) share the same hardware resources, by offering them one shared application and database instance, while allowing them to configure the application to fit their needs as if it runs on a dedicated environment.

Definition 2. A tenant is the organizational entity which rents a multi-tenant SaaS solution. Typically, a tenant groups a number of users, which are the stakeholders in the organization. These definitions focus on what we believe to be the key aspects of multi-tenancy:

1. The ability of the application to share hardware resources.
2. The offering of a high degree of configurability of the software.
3. The architectural approach in which the tenants (or users) make use of a single application and database instance[2].

II. RELATED WORK

The cloud is a next generation platform that provides dynamic resource pools, virtualization, and high availability. Today, we have the ability to utilize scalable, distributed computing environments within the confines of the Internet, a practice known as cloud computing[3]. Cloud Computing is becoming a well-known buzzword nowadays. Many companies, such as Amazon, Google, Microsoft and so on, accelerate their paces in developing Cloud Computing systems and enhancing their services to provide for a larger amount of users. However, security and privacy issues present a strong barrier for users to adapt into Cloud Computing systems. In this paper, we investigate several Cloud Computing system providers about their concerns on security and privacy issues. We find those concerns are not adequate and more should be added in terms of five aspects (i.e., availability, confidentiality, data integrity, control, audit) for security. Moreover, released acts on privacy are out of date to protect users' private information in the new environment (i.e., Cloud Computing system environment) since they are no longer applicable to the new relationship between users and providers, which contains three parties (i.e., Cloud service user, Cloud service provider/Cloud user, Cloud provider). Multi located data storage and services (i.e., applications) in the Cloud make privacy issues even worse. Hence, adapting released acts for new scenarios in the Cloud, it will result in more users to step into Cloud. We claim that the prosperity in Cloud Computing literature is to be coming after those security and privacy issues having be resolved[4].

There is a critical need to securely store, manage, share and analyze massive amounts of complex (e.g., semi-structured and unstructured) data to determine patterns and trends in order to improve the quality of healthcare, better safeguard the nation and explore alternative energy. Because of the critical nature of the applications, it is important that clouds be secure. The major security challenge with clouds is that the owner of the data may not have control of where the data is placed. This is because if one wants to exploit the benefits of using cloud computing, one must also utilize the resource allocation and scheduling provided by clouds. Therefore, we need to safeguard the data in the midst of untrusted processes[5].

Cloud computing, a rapidly developing information technology, has aroused the concern of the whole world. Cloud computing is Internet-based computing, whereby shared resources, software and information, are provided to computers and devices on-demand, like the electricity grid . Cloud computing is the product of the fusion of traditional computing technology and network technology like grid computing, distributed computing parallel computing and so on. It aims to construct a perfect system with powerful computing capability through a large number of relatively low-cost computing entity, and using the advanced business models like SaaS (Software as a Service) [6].With the development of parallel computing, distributed computing, grid computing, a new computing model appeared. The concept of computing comes from grid, public computing and SaaS. It is a new method that shares basic framework. The basic principles of cloud computing is to make the computing be assigned in a great number of distributed computers, rather then local computer or remoter server. The running of the enterprise's data center is just like Internet. This makes the enterprise use the resource in the application that is needed, and access computer and storage system according to the requirement. This article introduces the background and principle of cloud computing, the character, style and actuality. This article also introduces the application field the merit of cloud computing, such as, it do not need user's high level equipment, so it reduces the user's cost. It provides secure and dependable data storage center, so user needn't do the awful things such storing data and killing virus, this kind of task can be done by professionals. It can realize

data share through different equipments. It analyses some questions and hidden troubles, and puts forward some solutions, and discusses the future of cloud computing. Cloud computing is a computing style that provide power referenced with IT as a service. Users can enjoy the service even he knows nothing about the technology of cloud computing and the professional knowledge in this field and the power to control it[7].

Existing systems –

A. The Force.com

Introduction –

In 2008, The Force.com provides the first Multitenant architecture for SaaS. The focus of this system is multi tenancy, a fundamental design approach that can dramatically help improve the manageability of SaaS applications. This figure 1 defines multi tenancy, explains the benefits of multi tenancy, and demonstrates why metadata driven architectures are the premier choice for implementing multi tenancy. The world's first PaaS, which delivers turnkey multi tenancy for Internet-scale applications. The system details Force.com's patented metadata-driven architecture components to provide an understanding of the features used to deliver reliable, secure, and scalable multitenant applications.

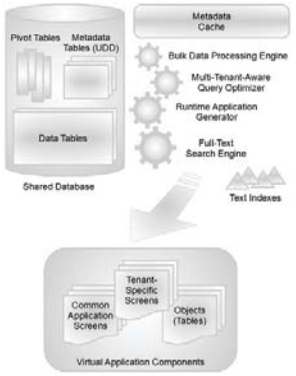


Fig. 1 : Force.com Platform Architecture.

Forms, reports, work flows, user access privileges, tenant-specific customizations and business logic, even the definitions of underlying data tables and indexes, are all abstract constructs that exist merely as metadata in Force.com's Universal Data Dictionary (UDD). For example, when a developer is building a new custom application and defines a custom table, lays out a form, or writes some procedural code, Force.com does not create an "actual" table in a database or compile any code. Instead, Force.com simply stores metadata that the platform's engine can use to generate the "virtual" application components at runtime. When someone wants to modify or customize something about the application, all that's required is a simple non-blocking update to the corresponding metadata.

Because metadata is a key ingredient of Force.com applications, the platform's runtime engine must optimize access to metadata; otherwise, frequent metadata access would prevent the platform from scaling. With this potential bottleneck in mind, Force.com uses metadata caches to maintain the most recently used metadata in memory, avoid performance sapping disk I/O and code recompilations, and improve application response times. Force.com stores the application data for all virtual tables in a few large database tables that serve as heap storage. The platform's engine then materializes virtual table data at runtime by considering corresponding metadata.

To optimize access to data in the system's large tables, Force.com's engine relies on a set of specialized pivot tables that maintain denormalized data for various purposes such as indexing, uniqueness, relationships, etc. Force.com's data processing engine helps streamline the overhead of large data loads and online transaction processing applications by transparently performing data modification operations in bulk. The engine has built-in fault recovery mechanisms that automatically retry bulk save operations after factoring out records that cause errors[8].

System Strength :

1. This is first SaaS multi tenant system.
2. System focused on most of the database related terms.

System Lack **:**

1. Any detailed information is not provided about event log.
2. No information given about the user profiles.
3. No mapping is provided.
4. System again lacking on tenant management.
5. No any specific algorithm is given for Customer account security.

B. LABS hp

Introduction –

On 21 Feb 2009, LABS^{hp} provides Multi-tenancy in Cloud – based Collaboration Services.

In this system it is mentioned that increased outsourcing of non-core competencies will drive the demand for a new generation of multi-tenanted cloud-based platforms that address the needs of content-

centered collaboration between organizations. This system is based on the FRACTAL conceptual prototype which has allowed us to evaluate the suitability of current enterprise content management (ECM) technologies for this type of platform.

Fractal Conceptual Prototype –

Goals of the Prototype

We had three distinct goals for the prototype: first, we wanted a functioning system that would help us to better envision FRACTAL from an end user perspective; second, we wanted to clarify requirements for the underlying platform; and third, we wanted to understand limitations of current ECM technologies for realizing multi-tenanted cloud-based applications.

Key Features –

Content Spaces : hosted spaces that bring together people, content, collaborative tools, and customizable active behaviors.

Active behaviors : a way for end users to define functional extensions operating within the context of a content space involving content, metadata, automated processing services and tasks carried out by other users. An active behavior may be manually invoked as needed, or it may be automatically triggered by a change to a content space or the passing of time. An invocation may involve a single content object or many objects in parallel. Their complexity ranges from automatically creating up to date PDF versions of documents as they are modified, to running workflows to automatically collate information from several collaborating organizations into a single document.

Agile configuration : must be light-weight, low-touch and customizable by end users without IT involvement.

Open and extensible by third parties : an Internet platform with open APIs, where third parties are motivated to develop customizations/extensions that can then be published through a marketplace and easily discovered by end users[9].

System Strengths :

1. System explain prototype for Multi-tenant.
2. System properly handled the SaaS multi-tenant concept.

System Lack :

1. System lack on users security.
2. System lack on event log.
3. System lack on mapping.
4. System lack on Load distribution.

C. EMC²

Introduction-

In June 2009, EMC² provides powerful capabilities for creating Software as a Service(SaaS). This system offers SaaS built on top of the Documentum 6.5 platform stack. The primary focus is on multi-tenancy as one of the major areas of complexity.

This system provides multi tenancy with following repository

1. Shared repository – model isolation is achieved via partitioning the data within single repository. The main reason for choosing a shared repository model is to support efficiently a large amount of small customers with a very load. If an application serves customers with one to 10 users it is probably the best model. Typically, a unique Customer ID is associated with the data that must be partitioned. All queries must be qualified by Customer ID to guarantee that customer will never have access to others customers data.

2. Dedicated repository – In this case, each customer will be guaranteed isolation from all possible perspectives. Using the dedicated model minimize the impact on application design and implementation. The only requirement is a mapping of the Customer ID to a repository. In shared repository data isolations are also provided.

3. Content Isolation - The simplest way to isolate the content is by creating a base custom type and making all content-enabled types derived from it.

User Isolation- In shared repository take care that each username does not collide with the names of other uses. This is important since the username is actually a UID and is used by other objects to refer to a user.

3.1 Schema Isolation - Two types of schema elements are most common : Custom attribute and Custom values. In Custom attribute are often required to model extension to document type such as Contacts or purchase orders. Each customer could have a unique set of metadata necessary for each type of document. In custom values one would need to create a special type and manage all the customer-defined values separately[10].

System Strengths :

1. This system focused on the user profile.
2. By using the data isolations the customer have the facilities to work more efficiently.
3. Customers account is safe because of the above mentioned repository.

System Lacks :

1. System more focused on customer instead the SaaS multi tenancy.
2. No method given through that user get more details about applications.
3. No work schedule is mentioned.
4. The mapping is not provided to the customer.
5. No security norms are mentioned.
   - ✓ Security
   - ✓ Privacy
   - ✓ Reliability
   - ✓ Compliance
   - ✓ Freedom

III. PROPOSED SOLUTION

As on the based on the above mentioned 'Literature Survey' we try to give proposed solution which will covered probably all lacunas of the above described systems. Multi-tenancy refers to the ability to run multiple customers on a single software instance installed on multiple servers. This is done to increase resource utilization by allowing load balancing among tenants, and to reduce operational complexity and cost in managing the software to deliver the service.

Three Tier Architecture of Web Services

As shown in following fig. 2, the developed solutions are divided into following layers.

Presentation Layer

The presentation layer provides an interface for clients to access the portal application. This layer consists of the following elements:

- ✓ Web Forms: The primary web form is the Default.aspx. This page is the entry point to the portal. It is responsible for loading the other elements of the presentation layer.
- ✓ Skins: The Default.aspx web form loads the skin for the page based on the settings for each page or portal.
- ✓ Containers: The Default.aspx web form also loads the containers for the modules based on the settings for each module, page, and portal.
- ✓ Module User Controls: Modules will have at least a single user control that is the user interface for the module.

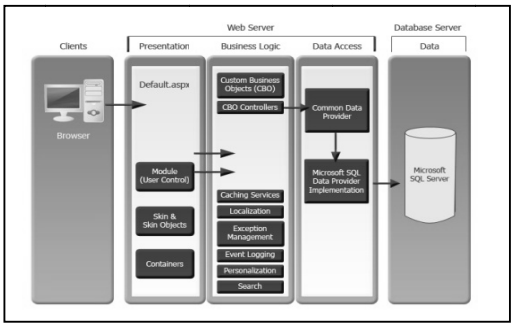


Fig. 2 : Three Tier architecture of Web Services

Business Logic Layer

The business logic layer provides the business logic for all core portal activity. This layer exposes many services to core modules. These services include

- ✓ Localization
- ✓ Caching
- ✓ Exception Management
- ✓ Event Logging
- ✓ Personalization
- ✓ Search
- ✓ Installation & Upgrades
- ✓ Security

Data Access Layer

The data access layer consists of two elements:

- ✓ Data Provider API: This is an abstract base class that establishes the contract that the implementation of the API must fulfil.
- ✓ Implementation of Data Provider API: This class inherits from the Data Provider API class and fulfills the contract by overriding the necessary members and methods. The core SaaS Multi-Tenant release provides a Microsoft SQL Server implementation of the Data Provider API.

Data Layer

The data layer provides data to the data access layer. The data store used in the data layer must be supported by the implementation of the Data Provider API to fulfil the data requests.

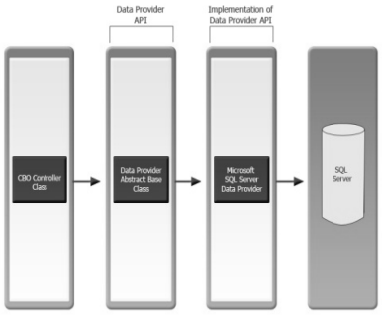


Fig. 3 : Data layers

Securities

To secure the data in cloud computing, we try to proposed Digital Signature with RSA encryption algorithm[2].

So, we are proposed to do the following.

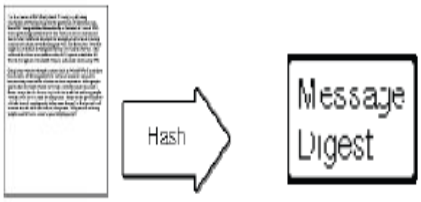


Fig. 4 : Document crunch into message Digest

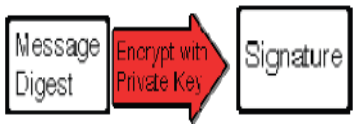


Fig. 5 : Encryption of message digest signature

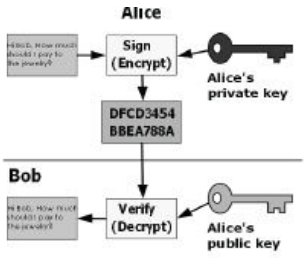


Fig. 6 : Encryption of Digital Signature into cipher text

With the increase of the amount of data and users in information system, the requirements of data integrity in system need to be improved as well. To handle the multiple users in cloud computing again proposed MD5 algorithm[11] which is shown in the figure 6.

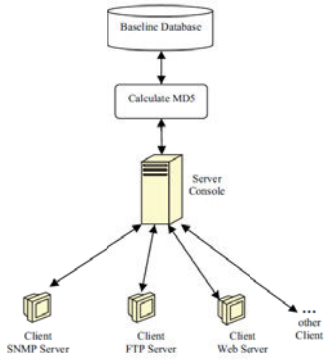


Fig. 7 : System architecture of MD5

IV. CONCLUSIONS

Cloud computing is still struggling in its accuracy , with positive and negative comments made on its possible implementation for a large-sized enterprise. IT technicians are spearheading the challenges. Several groups have recently been formed, with the goal of exploring the possibilities offered by cloud computing and to establish a common language among different providers. In this cloud computing is facing several issues in gaining recognition for its merits. Its security deficiencies and benefits need to be carefully handle before making a decision to implement it.

After study of various IT companie's SaaS Multi-tenant system , one thing is clear that, it is very fast growing technology now a days. But, each SaaS multi-tenant have some remedies which make it less powerful for working. This proposed system will overcome on most of the remedies in previous systems.

AKNOWLEDGEMENT

At the time of making a survey on this area many peoples were helped me. I specially thanks Prof. R. B. Wagh and Prof.G.K. Patnaik and for his valuable guidance on this area.

REFERENCES

[1] John W. Rittinghouse, James F. Ransome, "Introduction", Cloud Computing-Implementation, management & security, Ed. Taylor & Francis Group, LLC ,2010, pp.27-29.

[2] Cor-Paul Bezemer, Andy Zaidman, "Multi-Tenant SaaS Applications: Maintenance Dream or Nightmare?", Delft University of Technology Software Engineering Research Group Technical Report Series, 2010, http:// www.se.ewi.tidelft.nl/techreports, pp 1-5.

[3] Uma Somani, Kanika Lakhani, Manish Mundra, "Implementing Digital Signature with RSA Encryption Algorithm to Enhance the Data Security of Cloud in Cloud Computing", 2010 1st International Conference on Parallel, Distributed and Grid Computing (PDGC - 2010), pp.211-216.

[4] Minqi Zhou, Rong Zhang, Wei Xie, Weining Qian, Aoying Zhou, "Security and Privacy in Cloud Computing: A Survey", 2010 Sixth International Conference on Semantics, Knowledge and Grids, pp. 105-112.

[5] Kevin Hamlen, Murat Kantarcioglu, Latifur Khan, Bhavani Thuraisingham, "Security Issues for Cloud Computing", International Journal of Information Security and Privacy, 4(2), April-June 2010, pp 39-51.

[6] Jianfeng Yang, Zhibin Chen, "Cloud Computing Research and Security Issues", International Conference on Computational Intelligence and Software Engineering(CiSE), Wuhan, 10-12 Dec. 2010, pp 1-3.

[7] Shuai Zhang, Shufen Zhang, "Cloud Computing Research and Development Trend", 2010 Second International Conference on Future Networks,ICFN'10 , 2010, pp 93-97.

[8] The Force.com Multitenant Architecture: Understanding the Design of Salesforce.com's Internet Application Development Platform. http://www.salesforce.com/au/assets/Force.com_Multitenancy_WP_101508.pdf.

[9] David Banks, John S. Erickson, Michael Rhodes, "Multi-tenancy in Cloud-based Collaboration Services", Hewlett-Packard Development Company, L.P., February 21, 2009 http://www.hpl.hp.com/techreports/2009/HPL-2009-17.pdf.

[10] "Implementing SaaS Multi-tenancy with EMC Documentum 6.5- Best Pratices Planning", EMC2, http://www.emc.com /collateral /software /white-papers/h4701-oem-multitenancy-wp.pdf.

[11] Danyang Cao, Bingru Yang, "Design and implementation for MD5-based data integrity checking system", International Conference on Information management and Engineering (ICIME), 2010, pp.608-611.

◈◈◈

Improving Accessibility of Mobile Cloud Computing Using Multidimensional Data Structure

Rinku V. Rajankar & R.W. Jasutkar

Computer Science & Engineering, G.H. Raisoni College of Engineering, (G.H.R.C.E.), Nagpur, India

Abstract - Mobile cloud computing provides a solution to meet the increasing functionality demands of end-users, as all application logic is executed on distant servers and only user interface functionalities reside on the mobile device. As a result there is increasing demand of online personal data storage which should be used anywhere to provide a high performance, easy to use universal data access service thru various computing resources. In this paper we introduce multi-dimensional data structure to improve the accessibility of cloud .In the developing process of various servers proposed work make use of Microsoft's latest windows Azure cloud computing platform.

Keywords - *Cloud Computing, Azure framework.*

I. INTRODUCTION

Cloud computing [1] is one of the emerging technologies that will lead to the next generation of Internet. It provides optimized and efficient computing through enhanced collaboration, agility, scalability, and availability that reduces hardware and software investment cost. The essential cloud characteristics are on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. Computing capabilities, such as server time and network storage, can be unilaterally provisioned or de-provisioned as needed automatically. They are available over the Internet and accessible through heterogeneous client platforms, such as laptops and mobile phones. The computing resources are pooled and dynamically assigned and reassigned to serve multiple consumers. The capabilities appear to be unlimited, as they can be rapidly and elastically provisioned to quickly scale out and rapidly released to quickly scale in.

So Cloud computing is a paradigm in which data, applications or software are accessed over a network. This network of servers is called as "Cloud". A cloud application leverages the cloud in software architecture, often eliminating the need to install and run the application on the customer's own computer, thus alleviating the burden of software maintenance, ongoing operation, and support. Commercial cloud computing has three distinct offerings.

Software as a service (SaaS): SaaS delivers a single application through the browser to thousands of customers using a multitenant architecture. On the customer side, it means no upfront investment in servers or software licensing; on the provider side, with just one application to maintain, costs are low compared to conventional hosting.

Infrastructure as a Service (IaaS): IaaS is the delivery of computer infrastructure (typically a platform virtualization environment) as a service. Raw infrastructure, such as servers and storage, is provided from the vendor premises directly as an on-demand service.

Platform as a service (PaaS): Development platforms and middleware systems hosted by the vendor, allowing developers to simplify code and deploy without directly interacting with underlying infrastructure. That means it is possible to build own applications that run on the provider's infrastructure and are delivered to users via the Internet from the provider's servers.

The remainder of the paper is organized as follows, in section II we present key technology of cloud computing. In section III we reported related work and finally conclusion in section IV .

II. KEY TECHNIQUES OF CLOUD COMPUTING

In this section, we would take Google's cloud computing techniques [2] as an example, summed up some key techniques, such as data storage technology

(Google File System), data management technology (BigTable).

A. Google File System (GFS)

Google File System (GFS)[3] is a proprietary distributed file system developed by Google Inc. for its own use. It is designed to provide efficient, reliable access to data using large clusters of commodity hardware. GFS is optimized for Google's core data storage and usage needs (primarily the search engine), which can generate enormous amounts of data that needs to be retained. Files are divided into chunks of 64 megabytes, which are only extremely rarely overwritten, or shrunk; files are usually appended to or read. It is also designed and optimized to run on Google's computing clusters, the nodes of which consist of cheap, "commodity" computers, which means precautions must be taken against the high failure rate of individual nodes and the subsequent data loss. Other design decisions select for high data throughputs, even when it comes at the cost of latency.

The nodes are divided into two types: one Master node and a large number of Chunkservers. Chunkservers store the data files, with each individual file broken up into fixed size chunks (hence the name) of about 64 megabytes, similar to clusters or sectors in regular file systems. The Master server doesn't usually store the actual chunks, but rather all the metadata associated with the chunks, such as the tables mapping the 64-bit labels to chunk locations and the files they make up, the locations of the copies of the chunks, what processes are reading or writing to a particular chunk, or taking a "snapshot" of the chunk pursuant to replicating it (usually at the instigation of the Master server, when, due to node failures, the number of copies of a chunk has fallen beneath the set number). All this metadata is kept current by the Master server periodically receiving updates from each chunk server ("Heart-beat messages").

B. BigTable

A Bigtable[4] is a sparse, distributed, persistent multidimensional sorted map. The map is indexed by a row key, column key, and a timestamp; each value in the map is an uninterpreted array of bytes. BigTable is now used by a number of Google applications, such as MapReduce, which is often used for generating and modifying data stored in BigTable, Google Reader, Google Maps, Google Book Search, "My Search History", Google Earth, Blogger.com, Google Code hosting, Orkut, YouTube, and Gmail. Google's reasons for developing its own database include scalability, and better control of performance characteristics.

III. RELATED WORK

The aim of this paper is to combine cloud computing technologies with multidimensional data structure so that cooperating organizations can share vast amounts of data with improved accessibility. Recently an increasing number of commercial cloud platform has established to offer flexible services for end user around world. Amazon simple storage services (S3)[5] aim to provide storage as a low-cost ,highly available service via an HTTP-like interface. Generic operations such as get, put, delete and list are supported so that other services can be developed based on S3.

A. Microsoft Azure Platform

The Microsoft Azure Platform [6] is a cloud computing platform that offers a set of cloud computing services similar to the Amazon Web Services. Windows Azure Compute allows the users to lease Windows virtual machine instances. Azure compute follows a platform as a service approach and offer the net runtime as the platform. Users can deploy their programs as an Azure development package through a web application. Platform-as-a-service infrastructures have a greater capability to offer quality of service and automated management services than infrastructure-as-a-service offerings. Azure offers a limited set of instances on a linear price and feature scale.

B. Virtualization

Resource virtualization is at the heart of most cloud architectures. The concept of Virtualization allows an abstract, logical view on the physical resources and includes servers, data stores, networks, and software. The basic idea is to pool physical resources and manage them as a whole. Individual requests can then be served as required from these resource pools. For example, it is possible to dynamically generate a certain platform for a specific application at the very moment when it is needed. Instead of a real machine, a virtual machine is used. Proposed work create virtual cloud environment by using Azure Framework.

C. Web Application

The Web Application is the only interface provided system to the user to access the cloud infrastructure. We provide different functions based on the permissions assigned to a user. Our system is based on the client-server[7] approach. The client tools provide files and directories operations for users except for the management of users. It takes care of the registration of users and communities. It is designed to present users a friendly interface to simplify the management of users and communities.

D. Data access service.

In the windows Azure platform, a platform called Windows Azure Storage is specifically designed to build file storage service. Windows Azure Storage allows programmers to store any data they want. In accordance with "cloud computing" concept, the data once stored in the "cloud" will never be lost, programmers can gain access to any size of data at any time, from any terminal, anywhere. Different types of data storage available on windows azure are

- Blob storage: It is for long-term data. Blobs are binary objects together with <name, value> pair metadata. Each blob can be up to 50 GB and blobs are grouped into logical containers. Blobs are replicated three times in the data center for reliability purposes and they can be accessed from any server or by a URL over the Internet.
- Table storage: Another type of persistent storage. A table can be very large (millions of rows and columns) and is partitioned by rows and distributed over the storage nodes in Windows Azure. It is also triply replicated. Tables are not full SQL tables because there is no join operator..
- Queue: Asynchronous messaging service.

Table storage is used in proposed module as it work with multi-dimensional data structure

IV. CONCLUSION

Now, mobile computing user are looking for more effective ways to store and access their large amount of personal data. In our proposed system is providing more flexibility to access the data using multi-dimensional data structure. Also that data may not be precise .In future proposed system may include fuzzy concept to improve accessibility.

REFERENCES

[1] Irena Bojanova & Augustine Samba "Analysis of cloud computing delivery arch. Model". 2011 Workshops of International Conference on Advanced Information Networking and Applications

[2] Xu Wang, Beizhan Wang & Jing Huang ."Cloud computing and its key techniques". 2011 IEEE

[3] Ghemawat S, Gobioff H, Leung ST, "The Google file system' In:Proc. Of the 19th ACM Symp. On Operating Principles. New York: ACM Press. 2003. 29-43.

[4] Chang F, Dean J, Ghemawat S, Hsieh WC, Wallach DA, Burrows M, Chandra T, Fikes A, Gruber RE. "Bigtable: A distributed storage system for structure data" In:Proc. Of the 7th USENIX Syrup. On Operating Systems Design and Implementations. Berkeley: USENIX Association, 2006. 205-218.

[5] Keith R. Jackson, Lavanya Ramakrishnan &Krishna Muriki "Performance Analysis of High Performance Computing Applications on the Amazon Web Services Cloud" 2nd IEEE International Conference on Cloud Computing Technology and Science

[6] Wei Lu, Jared Jackson, Jaliya Ekanayake,et al "Performing Large Science Experiments on Azure: Pitfalls and Solutions" 2nd IEEE International Conference on Cloud Computing Technology and Science

[7] Gwan- Hwan Hwang**. "**Supporting Cloud Computing in Thin- client/server Computing model " 2010 IEEE International Symposium on Parallel and Distributed Processing with Applications

◈◈◈

Packet Monitoring Approach to Prevent DDoS Attack in Cloud Computing

Vikas Chouhan & Sateesh Kumar Peddoju

Electronics & Computer Engineering Department,Indian Institute of Technology Roorkee, Roorkee-247667, India

Abstract **-** In cloud environment, cloud servers providing requested cloud services, sometimes may crash after receiving huge amount of requests. This is exactly what happens in a denial of service (DoS) attack. It prevents the authentic clients from getting service. DoS attack is accompanied by IP Spoofing so as to hide the source of flooding and to make every request look different.

In this paper, we present an approach for packet monitoring in Cloud Environment to prevent DDoS attacks. This new approach of Hop Count Filtering provides a network independent and readily available solution to prevent DoS attack in Cloud environment. Also, this method decreases the unavailability of cloud services to legitimate clients, reduces number of updates and saves computation time. The presented approach is simulated in CloudSim toolkit environment and corresponding results are then produced.

Keywords **-** *Cloud Computing, TTL, IP, Hop Count, Denial-of-Service.*

I. INTRODUCTION

Cloud computing can be defined as a new style of computing in which dynamically scalable and often virtualized resources are provided as a services over the Internet. Advantages of the cloud computing technology include cost savings, high availability, and easy scalability [1].

DoS attacks do not wish to modify data or gain illegal access, but instead they target to crash the servers and whole networks, disrupting legitimate users' communication. DoS attacks can be launched from either a single source or multiple sources. Multiple-source DoS attacks are called distributed denial-of-service (DDoS) attacks [2].

When the operating system notices the high workload on the flooded service, it will start to provide more computational power to cope with the additional workload. The attacker can flood a single, system based address in order to perform a full loss of availability on the intended service [3, 4].

These attacks are a type of Flooding Attack [2, 5], which basically consist of an attacker sending a large number of nonsense requests to a certain service, which is providing various services under cloud. As each of these requests has to be handled by the service implementation in order to determine its invalidity, this causes a certain amount of workload per attack request, which in the case of a flood of requests usually would cause a Denial of Service to the server hardware [2].

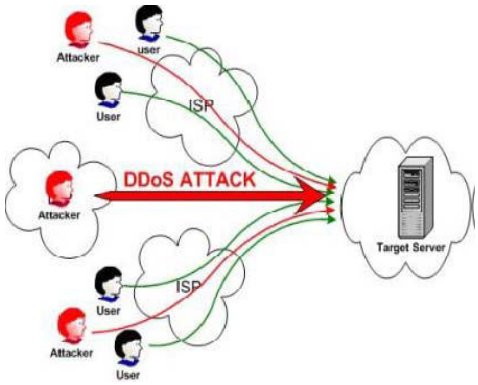


Fig. 1 : DDoS attack [2]

II. HOP-COUNT COMPUTATION

Since hop-count information is not directly stored in the IP header, one has to compute it based on the Time-to-live (TTL) field. TTL is an 8-bit field in the IP header, originally introduced to specify the maximum lifetime of each packet in the Internet. Each intermediate router decrements the TTL value of an in-transit IP packet by one before forwarding it to the next-hop [6, 7].

A. Extract final value of TTL

When a Packet reaches its destination and extracting its TTL field value, this value is known as final TTL. The challenge in hop-count computation is that a destination only sees the final TTL value. It would have been simple had all operating systems (OSs) used the same initial TTL value, but in practice, there is no

consensus on the initial TTL value. Furthermore, since the OS for a given IP address may change with time, we cannot assume a single static initial TTL value for each IP address [6].

B. Investigate the initial value of TTL

According to [6], most modern OSs uses only a few selected initial TTL values, 30, 32, 60, 64, 128, and 255. Only a few Internet hosts are apart by more than 30 hops, thus one can determine the initial TTL value of a packet by selecting the smallest initial value in the set that is larger than its final TTL. For example, if the final TTL value is 112, the initial TTL value is 128, the smallest of the two possible initial values, 128 and 255. Thus, given the final TTL value one can find the initial TTL value. Initial TTL values can be calculated as follows [8]:

Initial TTL=32 if final TTL <=32

Initial TTL =64 if 32<final TTL<=64

Initial TTL =128 if 64<final TTL <=128

Initial TTL =255 if 128<final TTL <=255

C. IP2HC Table

The IP2HC table [8] is a mapping between Source IP Address of a packets and stored hop count for that IP Address. It is a structure with Source IP address serving as index to match the hop count information.

III. PROPOSED ALGORITHM TO PREVENT DOS ATTACK

The proposed algorithm, uses the hop count filtering mechanism, and provides a clear idea of implementation so that it can be used in Cloud environment to prevent DoS attacks.

The algorithm requires continuous monitoring of packets travelling over the network in the Cloud, and thus, we extract SYN flag, TTL and source IP information from these monitored TCP/IP packets. The algorithm recognises four cases for each captured packet in the whole operation.

i. If SYN flag is set and source IP address exist (Syn=1 and Src=1) in IP2HC table then calculate hop-count by using TTL value of IP packet. Now check if the hop-count matches with the stored hop-count, if not, then update source hop-count field of table for that source IP address.

ii. If SYN flag is set and source IP address does not exists (Syn=1 and Src=0) in the IP2HC table then calculate hop-count and add a new entry for the Source IP address with the corresponding hop count in the IP2HC table.

iii. If SYN flag is not set and source IP address exists (Syn=0 and Src=1) in IP2HC table then calculate hop-count and if this hop count does not matches the stored hop count entry in the IP2HC table for the corresponding source IP address, then packet is spoofed, else the packet is legitimate.

iv. If SYN flag is not set and source IP address does not exists (Syn=0 and Src=0) in IP2HC table then it means that the packet is spoofed, because every legitimate IP address having a valid TCP connection will have its entry in the IP2HC table.

The inspection algorithm extracts the source IP address and the final TTL value from each IP packet [6]. The algorithm infers the initial TTL value and subtracts the final TTL value from it to obtain the hop-count. The source IP address serves as the index into the table to retrieve the correct hop-count for this IP address. If the computed hop-count matches the stored hop-count, the packet has been "authenticated" otherwise; the packet is likely spoofed [6].

ALGORITHM -1

Consider the following notations:

synflag = Syn bit of TCP packet.

mcount =malicious packet counter.

Tf= final value of TTL.

Ti=initial value of TTL.

```
Initialize mcount=0;
 For each packet
 Set TTL = ExtractFinalValueOfTTL( );
       //get time-to-leave field of IP packet
 Set srcIp = ExtractSourceIP( );
       //get source IP address from IP packet
 Set synflag = ExtractSynBit( );
      //get Syn flag value from TCP packet

 If (synflag is set)
 {
   If (establish_tcp_connection( ))
         //true when connection established
  {
```

```
        If ( srcIp is exist in IP2HC table )
        {
          ComputePacket ( srcIp , TTL , synflag);
            // function call which filter the spoofed
packet
        }
        else //new connection packet
        {
          Hc=ComputeHopCount( TTL );
            //get hop-count value
          NewEntryInTable(srcIp,Hc);
            //Add entry into IP2HC table
        }
    }
    else
    {
          // ignore packet
    }
 }
 else //synflag is not set
 {
    If ( srcIp exist in IP2HC Table)
    {
        ComputePacket ( srcIp , TTL, synflag );
          // function call which filter the spoofed
             packet
    }
    else
    {
        'drop the packet' //Packet is spoofed
        mcount++;     // increment in malicious
                         packet by 1

    }
 }
 ComputePacket ( string srcIp , int Tf , boolean
  synflag)
 {
    Hc=ComputeHopCount( Tf ); //get hop-count
                                  value
    Hs=RetreiveStoredHopCount(srcIp);
        //get stored hop-count value
    If ( Hc != Hs )
    {
        if( synflag is set)
        {
          UpdateTable ( srcIp , Hc);
            //update hop-count value in IP2HC
              table
        }
        else
        {
          'drop the packet' //Packet is spoofed
           mcount++;
            // increment in malicious packet by 1
        }
    }
    else
    {
         'allow the packet' // packet is legitimate
    }
}

int ComputeHopCount( int Tf )
{
     Set Ti= InvestigateInitialTTL(Tf);
     return Ti - Tf; //return hop-count value
}
```

IV. SIMULATION RESULTS

We simulated our algorithm on CloudSim toolkit having an arrival rate of 1000 packets per/sec at cloud server. Experimental results are shown in Table I. Various conditions discussed in Table I include pair of SYN flag (Syn) and source IP Address (Src) to provide information of packet. A value of Syn=0 represent SYN

flag not set and Syn=1 represents SYN flag is set. Similarly, Src indicate the presence of source IP Address in IP2HC table. A value of Src=0 represent entry does not exist and Src=1 represents entry exists. First experiment consists of 580 (337+243, see Table I) malicious packets, and 173 new entries and only 83 entries are updated. In contrast, the packets which require an update in the table are 130 (Syn=1 & Src=1). So, the effective (in fact, reduced) number of updates are 47 (130-83). The total reduction in the number of updates in the table is 30.15% (total allowed packets/ total packets), which is a considerable amount of improvement over the conventional method.

TABLE I : RECEIVED NUMBER OF PACKETS

Experi -ment No.	Syn=0 & Src=0	Syn=0 & Src=1			Syn=1 & Src=0	Syn=1 & Src=1			Improvement
	Malicious packets	Total packets	Malicious packets	Allowed packets	New Entry in IP2HC table	Total packets (tp)	Updated packets (up)	Allowed packets (ap)	Percentage of reduction in the no. of updates(ap/tp)
1	337	360	243	117	173	130	83	47	36.15
2	374	332	223	109	164	130	92	38	29.23
3	340	377	257	120	155	128	86	42	32.81
4	333	383	275	108	161	123	86	37	30.08
5	345	372	267	105	156	127	95	32	25.19
6	360	364	259	105	154	122	82	40	32.78
7	331	382	272	110	165	122	92	30	24.59

For computation time simulation the sample inputs are taken as arrival rate 'A' and various results has been analyzed and presented in Table II.

TABLE III : SAMPLE INPUTS

Sample	Sample input (Arrival rate in packets/sec)	Computation time (in ms)
1	1000	20
2	2000	38
3	4000	26
4	6000	25
5	8000	44
6	9000	62
7	10000	68

The graph in (Figure. 2) show how our proposed approach saves the possible computation time. In case of samples 2, 3 and 4 there is variation. Sample 2 needs more time then sample 3 and 4 because it depends on receiving field of packets. Computation time is relevant factor for performance measurement of cloud network and it improves processing power of cloud server and minimizes loss of available resources.

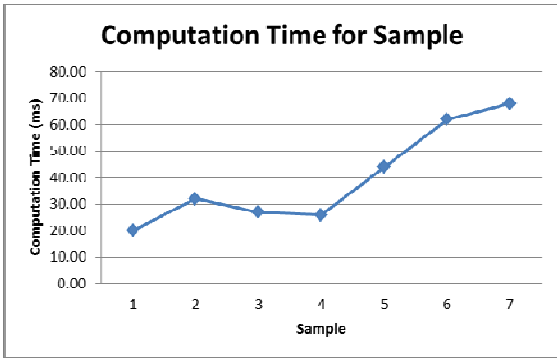


Fig. 2 : Graph showing computation time

IV. CONCLUSIONS

Cloud Computing is gaining popularity, but with the widespread usage of cloud the issue of cloud security is also surfacing. One of the major threats to Cloud security is Distributed Denial of Service Attack (DDoS) or simply Denial of service attack (DoS). To improve resource availability of resources, it is essential to provide a mechanism to prevent DDoS attacks. One of the methods for prevention is Hop Count Filtering method (HCF). This paper presented a version of Hop Count filtering method which not only detects malicious packets but also includes update of IP to Hop count Table (IP2HC) with a mechanism that reduces the number of updates and thus saves computation time by analyzing SYN flag of TCP protocol.

REFRENCES

[1] B. Furht and A. Escalante, Handbook of Cloud Computing: Springer, 2010, pp. 3-11.

[2] D. GARG, "DDOS Mitigation Techniques-A Survey," in International Conference on Advanced Computing, Communication and Networks, 2011.UACEE '11, pp. 1302-1309.

[3] P. A. R. Kumar and S. Selvakumar, "Distributed Denial-of-Service (DDoS) Threat in Collaborative Environment - A Survey on DDoS Attack Tools and Traceback Mechanisms," in Advance Computing Conference, 2009. IACC 2009. IEEE International, 2009, pp. 1275-1280.

[4] P. S. Mann and D. Kumar, "A Reactive Defense Mechanism based on an Analytical Approach to Mitigate DDoS Attacks and Improve Network Performance," International Journal of Computer Applications, vol. 12-No.12, pp. 43-46, January 2011.

[5] S. T. a. K. Levitt, "Detecting spoofed packets," in Proceedings of The Third DARPA Information Survivability Conference and Exposition (DISCEX III) '2003, Washington, D.C., 2003.

[6] W. Haining, et al., "Defense Against Spoofed IP Traffic Using Hop-Count Filtering," Networking, IEEE/ACM Transactions on, vol. 15, pp. 40-53, 2007.

[7] I. B. Mopari, et al., "Detection and defense against DDoS attack with IP spoofing," in Computing, Communication and Networking, 2008. ICCCn 2008. International Conference on, 2008, pp. 1-5.

[8] N. Venkatesu, et al., "An Effective Defense Against Distributed Denial of Service in GRID," in Emerging Trends in Engineering and Technology, 2008. ICETET '08. First International Conference on, 2008, pp. 373-378.

◈◈◈

Challenges in Cloud Environment

Himanshu Shah, Paresh Wankhede & Anup Borkar

Accenture Services Pvt. Ltd., Mumbai, India

***Abstract* -** Organizations have been skeptical about moving from traditional data-centre onto cloud. Along with well known concerns like security, loss of control over infrastructure, data theft, lack of standards, etc; cloud does pose issues like portability, reliability, maintainability, ease-of-use, and etcetera.

This whitepaper talks about these concerns around system quality attributes using Amazon Web Services (AWS) and Azure cloud as reference. The whitepaper encompasses the recent challenges faced and probable solutions for the same. It also covers one specific issue related to RHEL (Red Hat Enterprise Linux) [Ref 3] migration on AWS in detail.

This whitepaper also discusses and recommends cloud vendor(s) and certain management tools based on the parameters related to system quality attributes such as portability, reliability, maintainability, etc.

I. INTRODUCTION

A. Portability

It is the ease with which an environment and an application can be ported on and off the cloud or to other public cloud:

Description

There are no common standards for cloud vendors to adhere to. Current development efforts do not suffice the purpose. And they do not force the cloud service providers to prioritize and focus on the issue of interoperability. Some examples of it are:

1) *Microsoft's Azure only supports Windows OS compared to other vendors like AWS which support various flavors of Unix/Linux.*
2) *A java application hosted in Google Apps Engine (GAE) is bound to the DataStore which isn't exactly an RDBMS.*
3) *AWS PaaS services like SQS [Ref 4], RDS [Ref 5] creates a vendor lock-in.*

All the cloud vendors have the liberty to implement their services the way they deem beneficial. AWS provides a bunch of loose coupled services which can be used either in conjunction or independently. AWS, as IaaS, does not force the user to change the architecture of an application to make it compatible to host on AWS cloud. So at any given point of time user has a flexibility to move out of AWS cloud to any other cloud or traditional data-centre. Though there are PaaS services provided by AWS, if used, would require change in application architecture and that in turn would mean vendor lock-in.

Recommendation

There are no open standards. Cloud users have to keep this constraint in mind while designing the application, so that they don't get locked-in with the vendor. AWS provides support for application portability as long as PaaS services are not used.

B. Recoverability

It's the ease with which the infrastructure and application could be recovered in case of a disaster and/or threat to Business continuity:

Description

Services provided by cloud vendors are unique to their implementation. The compute node management, back-up procedures, firewall implementation, is different for all the vendors. Considering these facts application recoverability could be a time consuming.

It is quintessential in today's fast growing business environment to automate IT provisioning. Managing and configuring the IT infrastructure is one of the most time consuming and error prone task. The desire to implement something that would make the paradigm shift has given birth to a concept which treats Infrastructure management as Code. This transition, from traditional server maintenance to automation, would make the building and maintaining a modern infrastructure environment look more like maintaining a software project. AWS provides a service called CloudFormation [Ref 6] using which a user can rebuild the whole of infrastructure right from the scratch without much of human intervention.

Treating the Infrastructure as Code has been widely acknowledged. There are certain service

providers/technologies which offer Infrastructure management. They are:

1) *Chef [Ref 7]*
2) *Puppet [Ref 8]*
3) *RightScale [Ref 9]*
4) *Eucalyptus [Ref 10]*

Recommendation

It is recommended that Chef or Puppet should be used be it any infrastructure environment. These tools come handy for the purpose of recoverability and maintainability. AWS and Azure along with cloud management tool provides a unique way to manage cloud servers. The pros and cons of each of these tools have been discussed in details in coming sections.

C. Maintainability

It is the ease with which the application environment could be maintained.

Description

One of the biggest motivation factors for organizations to move onto cloud from traditional data-centre is maintainability. Users need not to worry about the maintenance of infrastructure when on cloud. Though, users have to keep a tab on application's health and status. This can be done with the help of cloud management tools available. Chef and Puppet are specially built considering this requirement. These tools help maintain the required status of an application. All the leading cloud service providers support cloud management through APIs, CLIs and web consoles. For example, AWS provides Java APIs while Azure supports .Net APIs.

To discuss the maintainability an experience of RHEL migration on AWS cloud has been discussed in detail in subsequent section.

Recommendation

Though there are bunch of APIs and CLIs available with each of the cloud vendor; it is easy to manage and maintain the infrastructure on cloud with the help of cloud management tools.

D. Transparency and Control

It is the extent to which the cloud vendor allows its user to control the infrastructure.

Description

Transparency in cloud's perspective is the capability to have a view at day-to-day activity of cloud infrastructure or point-in-time status and health of the servers. Users would like to have full control over the servers which are running under their account and hosting their applications. AWS implementing IaaS gives the user right amount of control over the servers to suffice the purpose. Azure on the other hand being PaaS does not provide that level of transparency and control at the hardware level.

Recommendation

If transparency and control over infrastructure is too major a deciding factor, data-centre is a way to go. AWS provides some control over the infrastructure which should suffice in most of the cases.

E. Security

It is the extent to which the cloud vendor supports security and reliability for cloud environment.

Description

One of the deciding factors for any organization while choosing cloud environment is its security and reliability. There are numerous regulations appertain to the storage and use of data, including Payment Card Industry Data Security Standard (PCI DSS), ISO, the Health Insurance Portability and Accountability Act (HIPAA), the Sarbanes-Oxley Act, to name a few.

Beside this there are privacy issues that arise from virtualization. AWS runs the server on virtualized environment and this brings in the issue that could arise from multi-tenancy. Organizations are skeptical about sharing the same hardware with multiple users. Though there are few options such as virtual private cloud and virtual private network that helps resolve the data security and network security to some level.

Recommendation

There are provisions that could help resolve the security issues to some extent, but if the nature of data in question is very sensitive it is better to keep it on on-premise servers.

F. OS support

It is the number and verity of operating systems that are supported.

Description

AWS supports almost all the types of leading OS [REF 1] ranging from enterprise level Unix systems to Microsoft server OS.

Recommendation

AWS has various flavors of operating systems to choose from. That makes a viable contender for application hosting.

II. RHEL MIGRATION ON AWS CLOUD – AN EXPERIENCE

AWS along with Red Hat offers enterprise level operating system on cloud. AWS-Red Hat collaboration allows users to rent Red Hat Enterprise Linux instead of purchasing a license. With this combined effort users were able to:

1) *Purchase RHEL by paying hourly fees.*
2) *Use RHEL OS which was provided on Amazon EC2 [Ref 1] cloud*
3) *Get the supported from Red Hat*

To sum it all up; the images were provided by AWS EC2 cloud, supported by Red Hat and managed by end user (Ref Figure 1 below).

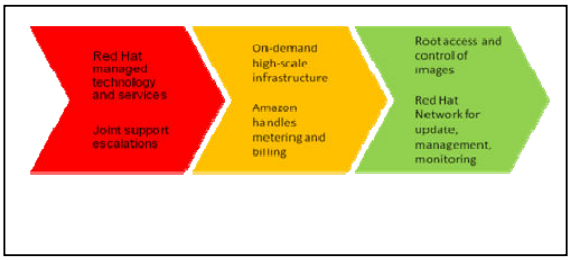


Image source: AWS

Fig. 1 : RHEL on AWS

The EC2 images provided by Red Hat were available in RHEL version 5.x. There were following ways to subscribe to that service:

1) *On-demand pay-as-you-go*
2) *RHEL premium subscription*

This was the beta offering which has now been discontinued to introduce hourly on demand offering. Subsequent section talks about the underlying architecture of AWS on top of which RHEL is offered.

A. *Amazon as-is architecture*

All the services provided by AWS are based on virtualization and its computing unit EC2 is no exception. Following subsections discuss in detail the kernel level architecture of AWS with respect to RHEL.

1) *Amazon Amazon Kernel Image (AKI) and Amazon RAM Image (ARI) concept*

Amazon gives the flexibility to use the kernels of user's own choice, other than the default EC2 kernels (AKI). To complement these kernel images AWS also provides ARI. The point to be noted here is that the architecture of the selected AMI, AKI and ARI should match and be compatible [Ref 2]. Even if they match, there is no guarantee that the combination would work and server would instantiate. It could so happen that the combination in place is not meant or designed to work in conjunction.

2) *Para-Virtual GRUB Loader (PVGRUB)*

AWS EC2 has a provision to load a Para-virtual Linux kernel within an Amazon Machine Image (AMI). The option is provided to create images that contain a kernel and initrd (initial RAM disk), and behave in a manner that is closer to traditional virtual or physical Linux installations. This feature allows seamless upgrade of kernel on Amazon EBS-backed instances. It enables the user to load any kernels of their choice. This PVGRUB acts as a mini-OS that runs at the boot and selects the kernel to boot from by reading /boot/grub/menu.lst from an AMI. It loads the specified kernel and eventually shuts itself down to free the resources.

B. *Migration from DevPay to hourly on demand*

Initially, RHEL (beta image) was made available on EC2 via DevPay subscription model. Users could simply subscribe to the offering, electronically, and have access to the RHEL AMIs. The subscription came with monthly fees on top of pay-per-use (hourly fee) model of AWS.

Red Hat has recently made an announcement to perform a major AMI update. This update will remove all the existing DevPay AMIs and replace them with the new set up AMIs. This move has been introduced to change the subscription model from DevPay to hourly on demand. This new subscription model does not include any monthly fees.

There is no migration tool or policy made available by RHEL or AWS. Considering the fact that DevPay AMIs would not work post the retirement, all the data from the configured AMIs had to be moved onto the new offering images called hourly on demand. There are two ways to get this done:

1) *Doing the whole configuration right from the scratch.*
2) *Utilizing available cloud management tools.*

Subsequent section details out the theory of utilizing cloud management tools, which is the recommended way to manage the infrastructure. These management tools are based on a concept arisen from treating the Infrastructure as Code.

C. Infrastructure as Code

It is quintessential in today's fast growing business environment to automate IT provisioning. Managing and configuring the IT infrastructure is one of the most time consuming and error prone task. The desire to implement something that would make the paradigm shift has given birth to a concept which treats Infrastructure management as Code. This transition, from traditional server maintenance to automation, would make the building and maintaining a modern infrastructure environment look more like maintaining a software project.

Using cloud management tools gives an edge over traditional methods in a sense that they are agile, modular and customizable. This need of managing the infrastructure and automating the provisioning process has made the users to look for a solution which could be perceived as a software module. Cloud management tools such as Chef and Puppet are open source scripting based management tools. These scripts are idempotent, cross platform and modular which makes them a reliable to use.

Various cloud management tools are compared and discussed in following sections [Ref Table I]. Recommendations based on predefined parameters are also provided.

III. RECOMMENDATIONS

A. Manageability

1) Script the image configuration process

Scripting the instance configuration process gives an edge over manual configuration process in an ever changing environment (refer the RHEL case discussed above). It's easier to modify a piece of code than repeating the manual process right from the scratch.

B. Configurability

1) Assigning a role to an instance on the fly

It's always good to use the base image for instance creation and configure that instance, by assigning it a role, on the fly. This reduces the burden of maintaining the bunch of images. Any changes that are to be made would have to be incorporated in the base image.

C. Maintainability

1) Usage of cloud management tools

Puppet and Chef are two such cloud management tools which offers infrastructure automation. They provide a framework to automate system admin tasks. There are certain benefits of using scripting tools to set up the infrastructure against traditional, manual method. The advantages are as follows:

a) Easier to maintain.

b) No scope for errors inflicted by human intervention.

c) Same scripts could be used in traditional datacenter.

d) No vendors lock-in as far as cloud vendors are concerned.

e) Scripts are idempotent, cross platform and modular

There are some negatives associated with this. They are as follows:

a) Big learning curve.

b) Requires deep knowledge and understanding of how system would behave if code were to change even a bit.

AWS provides a service called CloudFormation which actually works as cloud management tool to recreate whole of the infrastructure with a single click. This is based on a concept of Infrastructure as Code. Azure does not provide any such functionality right out of the box.

2) Third party cloud management tool

Using a third party management tool, like RightScale, gives a flexibility to move the application from one cloud service provider to another with ease. But there is a possibility of getting tied up to RightScale services. RightScale has its own terminology and features like ServerTemplate, RightScripts which are very handy and useful, but are RightScale patented.

D. Portability

1) Avoid vendor lock-in

An application to be ported on cloud should be designed such that it becomes cloud agnostic. Services that could make the user get locked-in with vendor should be avoided to make the application portable in future. Such applications could make use of cloud management to its full extend. Services provided by Azure are more tightly coupled as compared to services provided by AWS. There are high chances of vendor lock-in with both the service providers if specific services like queuing service, Content Delivery Network (CDN), etcetera, are used.

Following table (Table I) lists out the recommendation for different cloud management tools based on system quality attributes:

.

TABLE I : COMPARISON AND RECOMMENDATION ON CLOUD MANAGEMENT TOOLS

Capabilities	RightScale	CloudFormation	Chef/Puppet	Eucalyptus	Recommendation
Cloud vendor Portability	Supports various clouds, but not all of them.	Specific to AWS cloud	Scripts works on any cloud and Data-centre.	Supports AWS and on-premise cloud.	Chef/Puppet.
Recoverability	It is helpful	Built for recoverability	Supports recoverability	Supports recoverability	CloudFormation.
Usability	Small learning curve considering the fact that it has a web console	Small learning curve. Based on JSON	Big learning curve	It has a learning curve.	CloudFormation.
Maintainability	Easy to maintain	Easy to maintain	Somewhat difficult	Easy to maintain	RightScale.
Lock-in	User gets tied up to RightScale	Only usable with AWS	No lock-in	User gets tied up to Eucalyptus	Chef/Puppet
User group	Big user group. Components are sharable.	No formal platform to share components	Big user group. Components are sharable.	Small user community.	RightScale or Chef/Puppet
Safety	Have to share cloud credentials.	No need to share credentials	No need to share credentials	Have to share cloud credentials.	Chef/Puppet

IV. CONCLUSION

System quality attributes are the key parameter on which maturity of a cloud service can be evaluated. Cloud computing still being in a nascent state strives to provide quality service over this attributes. Evaluation and recommendation in this whitepaper is based on current state and maturity of various clouds and cloud management tools. This is subjected to change down the line, considering the fact that cloud technologies are in evolving state.

Treating Infrastructure as Code for cloud environment would reduce an overhead to configure and manage the infrastructure manually. Same scripts, with little or no modifications, could be used to setup an infrastructure in a datacenter. These scripts/tools are not vendor specific and hence provide a great flexibility and agility. Scripts are idempotent, cross platform and modular which makes them a reliable to use.

RHEL migration case discussed in this whitepaper is an apt example on maintainability issue that one could face on cloud and it also highlights the importance of treating the Infrastructure as Code.

REFERENCES AND FURTHER READING

[1] http://aws.amazon.com/ec2/

[2] http://aws.amazon.com/articles/1345

[3] http://aws.amazon.com/rhel/

[4] http://aws.amazon.com/sqs/

[5] http://aws.amazon.com/rds/

[6] http://aws.amazon.com/cloudformation/

[7] http://www.opscode.com/chef/

[8] http://puppetlabs.com/

[9] http://www.rightscale.com/

[10] http://www.eucalyptus.com/

◈◈◈

QoS Based Scheduling of Workflows in Cloud Computing

Jayadivya S K & S. Mary Saira Bhanu
Department of Computer Science & Engineering, National Institute of Technology, Tiruchirappalli, Tiruchirappalli – 620 015, India

Abstract **-** Cloud computing is a collection of virtualized computers that are probed on-demand to service applications. These applications are represented as workflows which are a set of tasks processed in a specific order based on their required services. Scheduling these workflows to get better success rate becomes a challenging issue in cloud computing as there are many workflows with different QoS (Quality of Service) parameters. In this paper, we introduce a strategy, QoS based Workflow Scheduling (QWS) to schedule many workflows based on user given QoS parameters like Deadline, Reliability, Cost etc. The strategy is to schedule the tasks based on QoS negotiation between user requirements and the services provided by Computation and Storage servers. The scheduler does the QoS negotiation based on the surplus information. The experiments were conducted in a simulated cloud environment by generating services and workflows randomly. The results show that our strategy gives the effective success rate as reliability of the service is considered during QoS negotiation.

Keywords- *Cloud computing; Scheduling; Workflows; User defined QoS parameters;*

I. INTRODUCTION

Cloud computing [1][2] has emerged as a global – infrastructure for applications by providing large scale services through the cloud servers. The services can be either storage service or computation service. Hardware and software resources can be utilized by users as services. These services can be configured dynamically by making use of virtualization.

Cloud computing provides a computing environment for the applications which can be represented by the workflow. Workflow is a sequence of tasks processed in a specific order based on dependency of services between these tasks. A workflow has a set of QoS parameters and it is represented as a Directed Acyclic Graph (DAG) [3] in which the nodes represent individual application tasks and directed arcs stand for precedence relationship among the tasks. Mapping between these tasks and services depends on the scheduling algorithm which is an NP complete problem [4].

Scheduling of workflows is a challenging one when many workflows are considered with many QoS parameters. There are many scheduling algorithms [5][6][7] developed for QoS parameters which consider either Execution time or Budget constraints or both. Along with these parameters, reliability is also an important factor to be considered in various applications. For example in some real-time applications like medical surgery, banking, etc., require urgent execution of workflows. So workflow has to exhibit high levels of reliability because applications process may get delay due to workflow failures.

In this paper, we have considered deadline, reliability and cost as QoS parameters for scheduling. Each task should have some parameters to satisfy QoS requirements of a workflow, but in most of the cases, user will mention the QoS parameters for whole workflow [7]. So in the proposed strategy, QoS based Workflow Scheduling (QWS), calculates the surplus information to achieve QoS negotiation for a workflow by using the distribution of parameters among tasks. QWS accepts multiple workflows and multiple QoS parameters from the users at any time and it reduces makespan and cost by considering reliability factor, which increases success rate of scheduling.

Rest of this paper is structured as follows: Section 2 will explain the related work. Section 3 discusses about the strategy. Section 4 describes about experimental results and section 5 gives the conclusion and future work.

II. RELATED WORK

In this section, few scheduling algorithms are described which use workflows as input. As for cloud workflow systems, similar to many other grid and distributed workflow systems, scheduling is a very important component which determines the performance of a whole system.

According to [5], there are two major types of workflow scheduling algorithms: Best effort based and QoS constraint based algorithms.

Best effort based scheduling algorithms attempts to minimize the makespan of a workflow. Some examples are Heterogeneous Earliest Finish Time algorithm [6], Min – Min algorithm [8] used by GrADS, Throughput maximization strategy used by SwinDeW-G [9] and SwinDeW-C [10]. But these algorithms neither have any provision to specify their QoS parameters nor any specific support to satisfy them.

Zhifeng Yu and Weisong Shi [11] proposed an algorithm for multiple workflows. In this algorithm, the ready tasks are scheduled based on ranking of tasks but this algorithm does not consider any user defined QoS parameters.

However, the scheduling algorithms which consider QoS parameters like deadline, cost, reliability, etc., were designed for transaction intensive workflows not for multiple. workflows. The transaction intensive workflows are multiple instances of one workflow whereas multiple workflows are different types of workflows with different requirements.

QoS constraint based scheduling algorithms, consider the user requested parameters while scheduling. These parameters can be deadline, reliability, cost, availability etc. Cost-based Scheduling of Scientific Workflow Applications on Utility Grids was proposed by Jia Yu, Rajkumar Buyya and Chen Khong Tham [7], which takes care of deadline as one QoS parameter and it minimizes the cost of a workflow. Meng Xu, Lizhen Cui, Haiyang Wang, Yanbing Bi [12] proposed a scheduling algorithm for multiple workflows with deadline and cost, but other parameters like reliability, availability, etc., were not included.

Although the above mentioned algorithms have their benefits for which they have designed for, none of them discusses about multiple workflows with many QoS parameter like Deadline, Cost, Reliability, Availability etc. Thus we have included Deadline, Cost and Reliability parameters with multiple workflows in QWS. The objective is to reduce the makespan and cost by considering the reliability factor.

III. QWS ARCHITECTURE AND METHODOLOGY

A. System Model

Fig. 1 shows the architecture of the QWS in a cloud environment. There are two major types of servers in cloud which are storage server and computational server. Storage server provides the service related to data storage and modification which does not require any mapping of services. Computational server provides the service related to computing resources which requires mapping of services based on QoS parameters required by a task.

QWS process is designed by using three modules which are Preprocessor module (PM), Scheduler module (SM) and Executor module (EM) with rescheduling if required (Backfilling). The control flow diagram of these modules is shown in Fig. 2.

Users first submit their workflows with QoS parameters in Abstract data structure format to preprocessor module. PM discovers the services required for those tasks and generates the DAG [3] based on the dependencies between them and divides the tasks based on computation services and storage services. SM allocates the particular services to these tasks based on the requirement of QoS parameters and attributes of the services in cloud environment. After mapping, EM sends the tasks to the mapped servers and checks for the result of these tasks. If EM gets the successful result then it activates all the tasks which are dependent on these results. If it fails, then SM re-schedules those tasks.

B. Problem Definition

Workflow ω_i is represented by a set of four tuples which are $<T_{i,j}, D_i, R_i, C_i>$. $T_{i,j}$ is a set of finite tasks { $T_{i,1}$, $T_{i,2}$, $T_{i,3}$,… $T_{i,j}$ }. Each task $T_{i,j}$ has a set of attributes like task-id, deadline, execution time, datasets and services needed, size, etc.

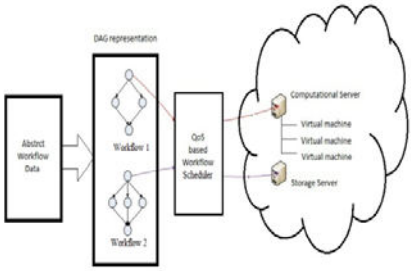


Figure 1: QWS Architecture

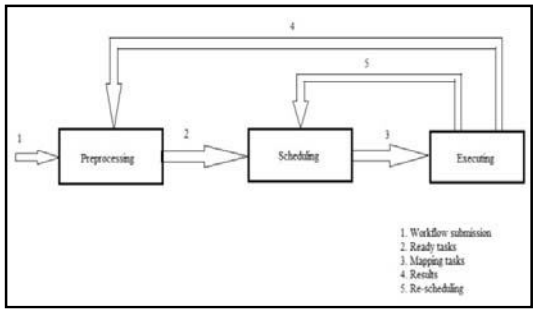


Figure 2: Control flow diagram

Deadline of each task is calculated by distributing the deadline of workflow among tasks in a critical path. Deadline distribution algorithm is described in Methodology. Most of the times the execution time of

the task depends on the performance of the machine in which job has to be executed [13]. For this reason, execution time is calculated when it is assigned to a node based on its MIPS rate. D_i is the Deadline of the workflow ω_i before that the workflow has to be executed. R_i is the minimum reliability of workflow ω_i, which should be maintained for executing the workflow. C_i is the cost of the workflow ω_i required to execute it.

Let *m* be the number of services available in cloud. Let s_k be the set of services which are capable of executing the task $T_{i,j}$. The QWS will schedule a set of tasks by mapping each task to suitable s_k by achieving reduced makespan and budget and maximum reliability with the help of the following equations from (1) to (3).

$$makespan = \max_{Tj \in \tau_i} executiontime(T_j) \quad (1)$$

$$budget = \sum_{j=1}^{n_i} \cos t(T_j) \quad (2)$$

$$reliability = \prod_{j=1}^{\tau_i} reliability(T_j) \quad (3)$$

$$where\ \tau_i\ is\ the\ critical\ paths\ in\ the\ \omega_i$$

C. Methodology

1) *Preprocessing Module:* This module accepts the workflow data from the user in the form of abstract data structure as shown in the following example.

No. of workflows : 1

No. of tasks : 4

task1 : {Datasets:d1,d2,d3, Services : s1,s2}

task2 : {Datasets:d2,d4,d5, Services : s3,s4}

task3 : {datasets:d3,d6,d1, services : s10,s5}

task4 : {datasets:d5,d6,d10, services : s7,s12}

Deadline : Sun Dec 4 12 : 53 : 18 2011

Cost : 150

Reliability : 9.356

From the above information, it generates the DAG by finding the dependencies between the tasks of a workflow using (4). DAG is a directed set of arcs of the form (Ti,Tj) where Ti is called the parent task and Tj is called the child task of Ti. Child task cannot be executed until all its parent tasks complete its execution.

$$Dep(T_i, T_j) => datasets(T_i) \cap datasets(T_j) == \phi\ ?\ 0 : 1; \quad (4)$$

The QoS parameter, deadline, will be mentioned for the whole workflow [7]. But each task should also have a deadline to meet users' deadline. So the deadline of the whole workflow is distributed in to sub-deadlines among the tasks in a critical path based on their size. Backtracking algorithm as shwon below is used to find the critical paths in a workflow.

```
void backtrack( )
{
    get DAG of workflow
    critical _ path(DAG)
}

  void critical _ path(DAG)
  {
      /* start -> root node
         end -> ter minating node in DAG */
      for(each set of start and end tasks in ω_i)
          allpaths(DAG, start, end);
  }
void allpaths(DAG, current, end)
{
    /* current -> node from which
                  path has to be found */
    push(current);
    if(current == end)
        store each task in critical path set τ of ω_i;
    for(i from 0 to number of neibours of i)
        allpaths(DAG, i, end);
}
```

It finds ready tasks using (5) in all workflows which are tasks whose predecessor tasks executed successfully. Then it sends these tasks to the scheduler module by placing them in to a ready queue. Initially the tasks are the root nodes of all DAGs.

$$\sum_{k=0}^{n} \sum_{l=0}^{k} DAG_i(T_k, T_l) == 0 \quad then\ T_k\ is\ ready\ task \quad (5)$$

where n is total number tasks in a workflow

2. *Scheduling module:* This module sorts all tasks in the ready queue based on:

- Instructions_time_ratio
- Number of services

Instructions_time_ratio is the ratio between the number of instructions in the task and the deadline of

the task. The tasks with less instructions_time_ratio are scheduled first.

The number of services can be accessed in cloud is limited. Many times, the number of tasks waiting to be executed will be more than the total number of services available. So task with fewer services should be scheduled first.

After sorting the tasks in ready queue, it checks whether the task requires computation service or storage service. If the task is related to storage service, then it maps the task to the storage server. If it is related to the computational service, then it checks for the availability of services in registry.

If available services can service the requirements of the task then it maps the task to that particular service in a data center. If the services are available in datacenters and they are busy then it checks for other datacenters and assigns to the next free server. If all are busy then it will be mapped to the wait queue of the data center which has lesser load compare to others. The algorithm for mapping and scheduling tasks on services is shown below:

```
void schedule(t, ω_i)
{
    /* t -> task
       s -> service */
    while ( ready_queue not empty )
    {
        t ← first task in ready_queue
        s ← getservice(t)
        send task to the data center
                    by placing in run queue
        dequeue(ready_queue)
        make s as busy
    }
}
```

```
int getservice(t)
{
 /* PTC -> cost required to execute previous tasks
    PTR -> Reliability obtained by previous tasks */

        select service S such that
            executiontime(t) < deadline(t) &&
            cost(t on s) + PTC < budget(ω_i) &&
            rel(s) * PTR < reliability(ω_i)
        return S
}
```

All available data centers will be registered in the registry with the attributes MIPS rate, available memory, services it can provide, reliability factor of those services, etc. Reliability of the service is maintained dynamically [14] and it is defined as the probability that the server returns a correct result within a time period which indicates it is not a binary property. A server can return correct result or wrong result based on the circumstances. The reliability of the server may also change with time due to fluctuating load, malicious node behavior, outage, etc. So the reliability has to be monitored in regular intervals which can be done by using (6). Here we assumed that if the number of tasks failing to meet the QoS parameters on that service increased to more than 20% then the reliability of the service is changed accordingly.

$$r(S) = \frac{n_S(t) + 1}{N_S(t) + 2}$$

where $r(S)$ *is the reliability of the service* S

$n_S(t)$ *is the total number of tasks executed successfully on* S

$N_S(t)$ *is the total number of tasks executed on* S

(6)

3. *Executor module:* EM sends all mapped tasks to the repective data centers and also waits for the acceptance and reply from the data center. The data center can accept the task or reject the task based on surplus information as shown in (7).

$$T(t_{ready\_queue} + t_{wait\_queue} + t_{new} + resolve(t_{new})) <= Deadline(t_{new})$$

$$\text{Re}liability(t_{new}) <= \text{Re}liability(service\ provided\ by\ data\ center)$$

$$Execution\_cost(t_{new}) + Commnication\_cost(t_{new}) <= Budget(t_{new})$$

If the datacenter accepts the task, then EM sends the task to the data center and waits for the result. Once EM gets the result, it checks for the QoS parameters of the completed task. If the completed task meets the QoS requirements, then it sends the result to preprocessor which masks the dependencies of all its child tasks. If the completed task does not meet the QoS requirements, then it sends an error signal to preprocessor which displays message to the user.

If the datacenter rejects the task, then it sends the task back to the scheduler i.e. backfilling, then the scheduler re-schedules that task.

We assumed that if the task continuously failing for more than 50 times, then it displays error message to the user.

The complete process of the QWS is shown in Fig. 3. This shows how the tasks can move from initiation state to completion state.

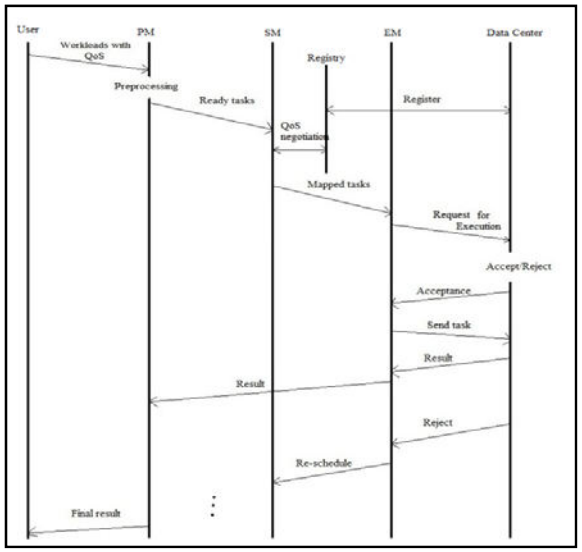


Figure 3. QWS Process

IV. EXPERIMENTAL RESULTS

In this section, the experimental results of the proposed QWS model are discussed. The Cloud environment is simulated using VMware virtual machines. Communication between them is achieved by using Ssh programming.

There are around 25 services in cloud environment which are scattered in different computational data centers and storage servers. To evaluate the performance multiple workflows ranging from 5 to 50 with a minimum of 8 tasks in each workflow with QoS parameters are generated randomly using random generation algorithm with uniform distribution.

The most common method for generating the random sequence $\{r_1, r_2, r_3, \ldots\ldots, r_k\}$ over the [n, m] is known as the linear congruent method. This method is, multiplying the previous random number r_{i-1} by the constant 'n' and adding with constant 'c' to it, then the modulus of the result is taken by dividing it by 'm' which gives r_i as shown in (8). This helps in distributing the values over [n, m] uniformly.

$$r_i = (nr_{i-1} + c) \bmod m \qquad \ldots\ldots \quad (8)$$

The services are chosen for each task randomly in the set $\{S_1, S_2, S_3, \ldots\ldots, S_{25}\}$. Then this information is sent to PM which starts the process of QWS as explained in section III.

The proposed QWS algorithm was run to evaluate its performance for various test cases with different number of workflows and different set of QoS parameters for each workflow.

The graphs for comparison between user given QoS parameter values and the obtained values of deadline, reliability and cost are shown in Fig. 4, Fig. 5 and Fig. 6 respectively. From these graphs we found that, QWS is meeting the QoS requirements of the workflow by achieving minimum makespan and cost and maximum reliability when compared to user required values.

QWS is compared with the MQMW algorithm [12] which is considering the QoS parameters Deadline and Cost but not the Reliability.

From the analysis we found that, the success rate of the scheduling is better than MQMW even after considering Reliability parameter. The graph of success rate is shown in Fig. 7.

V. CONCLUSION AND FUTURE WORK

The workflows in cloud computing platform have different QoS requirements. The main goal is to schedule many workflows by considering its QoS requirements. Many existing systems have addressed either for deadline or cost or both but not for reliability.

The proposed algorithm, QoS based workflow scheduling (QWS), allows users to execute their workflows by satisfying their QoS requirements like deadline, cost and reliability.

Experiments were conducted to test QWS algorithm with random generation of workflows in a simulated cloud computing environment. The results of these experiments were compared with the results of MQMW algorithm. This showed that QWS algorithm produced good success rate of scheduling even after considering reliability along with deadline and cost.

In our future work, we will include some more QoS parameter like availability, etc., and data grouping and movement before scheduling the tasks.

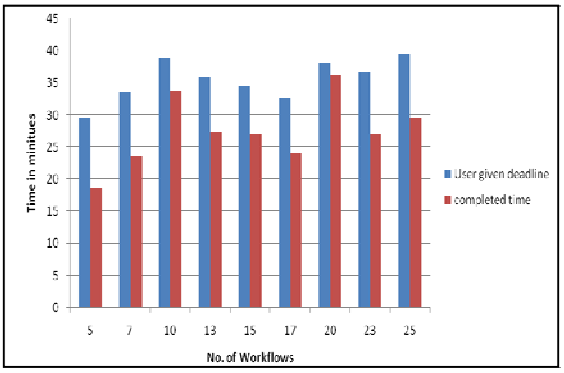


Figure 4. User defined Deadline versus Obtained Deadline

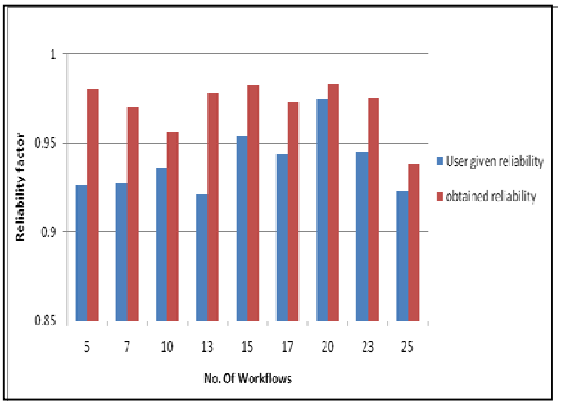


Figure 5. User defined Reliability versus Obtained Reliability

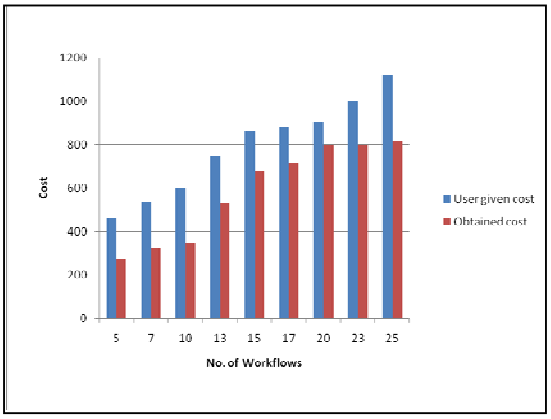


Figure 6. User defined Cost versus Obtained Cost

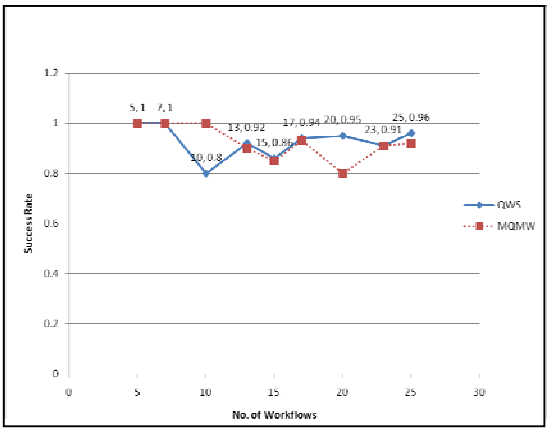


Figure 7. Success Rate of Scheduling

REFERENCES

[1] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg, and Ivona Breandic, "Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility", Future Generation Computer Systems, Elsevier Science, Amsterdam, June 2009, Volume 25, Number 6, pp. 599-616.

[2] C. Germain-Renaud and O.Rana, "The Convergence of Clouds, Grids and Autonomics" IEEE Internet Computing, p. 9, 2009

[3] Z. Shi and J.J.Donagarra, "Scheduling workflow applications on processors with different capabilities," Future Gen. Comupter systems 22 (2006) 665-675

[4] M.R. Garey and D.S. Johnson, "A Guide to the Theory of NP-Completeness", Computers and Intractability, New York, Freeman, 1979.

[5] J. Yu and R. Buyya, "Workflow Scheduling Algorithms for Grid Computing", Metaheuristics for Scheduling in Distributed Computing Environments, F. X. a. A. Abraham, ed., Springer, 2008.

[6] H. Topcuouglu, S. Hariri and M. Wu, "Performance-effective and low-complexity task scheduling for heterogeneous computing," IEEE Transactions on Parallel and Distribution Systems, vol. 13, no. 3, pp. 260-274, 2002.

[7] Jia Yu, Rajkumar Buyya and Chen Khong Tham, "Cost-based Scheduling of Scientific Workflows Applications on Utility Grids", In 1st IEEE International Conference on e-Science and Grid Computing, Melbourne, Australia, Dec. 5-8,2005.

[8] K. Cooper et al., "New Grid Scheduling and Rescheduling Methods in the GrADS Project", In NSF Next Generation Software Workshop, International Parallel and Distributed Processing Symposium, Santa Fe, IEEE CS Press, Los Alamitos, CA, USA, April 2004.

[9] Ke Liu, Jinjun Chen, Yun Yang and Hai Jin, "A throughtput maximization strategy for scheduling transaction-intensive workflows on SwinDeW-G", Concurrency and Computation: Practice and Experience, 20(15):1807-1820, wiley, Oct. 2008.

[10] Yun Yang, Ke Liu, Jinjun Chen, Xiao Liu, Dong Yuan, Hai Jin, "An Algorithm in SwinDeW-C for Scheduling Transaction-Intensive Cost-Constrained Cloud Workflows", Proc. Of the Fourth IEEE International Conference on e-Science and Grid Computing (e-Science08), pp. 374-375, Indianapolis,USA, Dec. 2008.

[11] Zhifing Yu and Weisong Shi, "A Planner-Guided Scheduling Strategy for Multiple Workflow Applications," icppw, pp.1-8, International Conference on Parallel Processing – Workshops, 2008.

[12] Meng Xu, Lizhen Cui, Haiyang Wang, Yanbin Bi, "A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows for Cloud Computing", IEEE International Symposium on Parallel and Distributed Processing with Applications 2009.

[13] D Dutta, R C Joshi, "A Genetic – Algorithm Approach to Cost – Based Multi-QoS Job Scheduling in Cloud Computing Environment", International Conference and Workshop on Emerging Trends in Technology (ICWET 2011)-TCET, Mumbai,India.

[14] Jason Sonnek, Abhishek Chandra, Jon B. Weissman, "Adaptive Reputation-Based Scheduling on Unreliable Distributed Infrastructure", IEEE Transactions on Parallel and Distributed Systems, Vol. 18, No. 11, November 2007

◈◈◈

Stepping into the Cloud

Venubabu Kunamneni

Bellingham MA - USA - 02019

Abstract **-** This article gives the basic concept, defines the terms used in the industry, and outlines the general Architecture, SaaS, PaaS and IaaS of Cloud computing. It gives a summary of Cloud Computing and provides a good foundation for understanding. Cloud computing is a better way to run your business. Instead of running your apps on your data center, they run on a shared data center.

I. INTRODUCTION

A Cloud is a set of machines and web services that implement cloud computing. Users of a cloud request access from a set of web services that manage a pool of computing resources (i.e., machines, network, storage, operating systems, etc). When granted, some fraction of the resources from the cloud pool is dedicated to the requesting user until same user releases them. It is called Cloud computing. Cloud computing is offered as a service based on demand. It eliminates the need for organizations to build and maintain expensive data centers plus software costs, too.

It enables organizations to stand up new systems quickly and easily. It provides elastic resources that allow applications to scale as needed in response to market demands. Its "pay as you go" rental model allows organizations to defer costs. It increases business continuity by providing inexpensive disaster-recovery options. It will also reduces the need for organizations to maintain a large IT staff.

Cloud computing is a general term for anything that involves delivering hosted services over the Internet. These services are broadly divided into three categories: Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). Again they are categorized into Private, Public and Hybrid based on the nature of access and control with respect to use and provisioning of virtual and physical resources.

Cloud computing is an evolution in which IT consumption and delivery are made available in a self fashion via the internet or internal network, with a flexible pay-as-you-go business model.

But moving to the cloud turns out to be more challenging than it first appears. The cloud-computing business model is still in development stages for large organizations to use cloud tech for high critical applications, and quite a few issues remain for venders to work out.

II. BENEFITS FOR ORGANIZATION IF THEY CAN MOVE TO CLOUD ENVIRONMENT

- Fast start-up : Cloud computing is good for any start-up because it allows us to test your business plan very quickly for little money. We can also scale as we grow.
- Scalability : Like electricity, gas and water, cloud computing services allow businesses to only pay for what business use. And as your business grows, you can accommodate by adding more resources.
- Faster product development : We can build and deliver applications 4 times faster, at about ½ the cost of traditional software platforms. We can deliver a complete platform with a simplified programming model so anyone can use it to install their applications.
- No capital expenditures : Cloud computing reduces paperwork, lowers transaction costs, minimizes the capital expenditure on hardware and the resources. Moving your business to 'the cloud' also reduces the need for an IT staff.
- Lower computer costs : We don't need to maintain a high-powered and high-priced computer to run cloud computing's web-based applications. Because applications run in the cloud, not on the desktop PC. When we are using web-based applications, our PC can be less expensive, with a smaller hard disk, less memory, more efficient processor, we don't even need a CD or DVD drive, we don't have to install any software's, we access all the software from clod
- Instant software updates : When the application is web-based, updates happens automatically and are

available the next time when log into the cloud. When you access a web-based application, you will get the latest version.

- Unlimited storage capacity : Cloud computing offers limitless storage. Your computer's current 100 gigabytes hard drive is peanuts when compared to the hundreds of terabytes
- Increased data reliability : For desktops hard disk crash can destroy all your valuable data, but computer crashing in the cloud shouldn't affect the storage of your data.
- 24x7 universal document accesses : With cloud computing, we don't have to take our documents with us, they stay in the cloud, and we can access them whenever we have a computer and an Internet connection. All our documents are instantly available from wherever you are.
- Latest version availability : If company moves to cloud computing, every one can access same latest version from cloud, which also helps to migrate all users to latest version with one install
- Easier group collaboration : Multiple users can collaborate easily on documents and projects, because the documents are hosted in the cloud, not on individual PC's, all you need is a computer with an Internet connection

III. WHAT ARE THE CONCERNS WITH CURRENT CLOUD TECH FOR LARGE ORGANIZATIONS?

- App performance can suffer : Generally we don't see performance issues if application are running on private cloud, but performance issues can me be more on public cloud, so high rated application required to be on private clod.
- Losing Internet Connection : We should ask our self whether the risk of losing internet connection and therefore access to your database will affect. It is a risk some companies are willing to take, but something to consider.
- Data might not be secure : With cloud computing, all your data is stored on the cloud may not be a secure. How do we know how secure is our data? We know that all cloud computing venders are saying that data is secure, but we don't know that for sure.
- Security : Security is still big concern to adopt venders cloud computing, venders still need to develop more secure layers and also required to convince their clients.
- Need Constant Internet : It is impossible to access applications that are running in the cloud if we can't connect to the Internet. Since we all use the Internet to connect to applications, if we don't have an Internet connection it means we can't access anything. No Internet connection means no work.
- Stored data can be lost : Data stored in the cloud is unusually safe, replicates across multiple nodes. Our data can be missing, we have no local backup. (Unless we download all our cloud data to our own desktop). Relying on the cloud puts us at risk if the cloud goes down.
- Doesn't work with low-speed connections : A low-speed Internet connection, like dial-up services, makes cloud computing very painful, sometimes we can not even use application. As we know web-based apps require a lot of bandwidth to download, as do large documents.
- Application access can be slow : Even you have fast Internet connection, web-based applications can sometimes be slower than using application on your desktop. It all depends on load on the application and bandwidth of connection.

IV. TYPES OF CLOUD COMPUTING

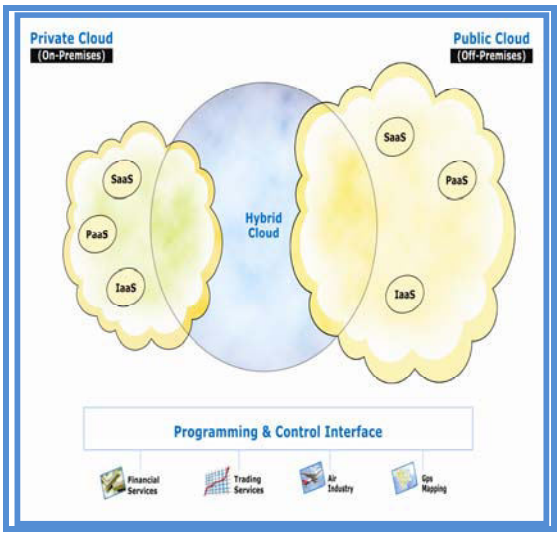


Fig. 1 : Types of Clouds

1. Public cloud : Public clouds Environments will be hosted by the cloud providers in their premises and provide access for the general public over the Internet. Customers are allowed to self-provision resources typically via a web service interface but they don't have any control over where and how the cloud computing infrastructure is run by the provider. Customer's rent access to resources as

needed on a pay-as-you-go basis. Public clouds offer access to large pools of scalable resources on a temporary basis without the need for capital investment in data center infrastructure. See Figure#1 for more details.

2. Private cloud : Private clouds give users immediate access to computing resources hosted within an organization's infrastructure. Users self-provision and scale resources drawn from the private cloud via web service interface, just as with a public cloud. But it is deployed within the organization's existing data center and behind the organization's firewall, a private cloud is subject to the organization's physical, electronic, and procedural security measures and thus offers a higher degree of security over public cloud. Externally hosted clouds are cheaper than on premise private clouds, but required more security. Private clouds are more expensive and secure than public clouds See Figure#1 for more details.
3. Hybrid cloud Organizations may host critical applications on private clouds and applications with relatively less security concerns on public cloud. The hybrid cloud combines resources drawn from one or more public clouds and one or more private clouds. The usage of both private and public clouds together is called hybrid cloud. In Hybrid cloud, an organization uses their own computing infrastructure for normal usage, but accesses the public cloud for high and peak load requirements. This ensures that a sudden demand in computing requirement is handled gracefully. See Figure#1 for more details.
4. Community cloud : involves the sharing of computing infrastructure between organizations of the same community. For example all Government organizations within the state of MA may share computing infrastructure on the cloud to manage data related to citizens residing in MA.

V. CLOUD SERVICES DELIVERY MODELS

5.1 Software as a Service (SaaS)

Software as a Service (SaaS) is a software distribution model in which applications are hosted and executed on the provider's infrastructure and front-end is made available to customers over a network, typically the Internet.

SaaS has become a common delivery model for most business applications, including accounting, collaboration, customer relationship management, enterprise resource planning, human resource management, content management, and help desk management. SaaS has been incorporated into the main strategy of all leading enterprise software companies with cloud. Refer to Figure#2 for more details

Benefits of the SaaS model include:

- Easier administration
- Automatic updates and patch management
- All users will have the same version of software
- Easier collaboration, for the same reason
- Global accessibility
- Less Admin costs

5.2 Platform-as-a-Service (PaaS)

Platform as a Service (PaaS) provides access to a programming or runtime environment with scalable compute and data structures embedded in it. With PaaS, users develop and execute their own applications within an environment offered by the service provider. PaaS enables you to commission applications quickly, without the cost and complexity of buying and managing the underlying software/hardware. Refer to Figure#2 for more details

Benefits with PasS

- Time to Market
- Requires no up-front investments
- Minimize operational costs
- Centralized information management
- Enhanced productivity
- Access to information anywhere, anytime
- Easy collaboration

5.3 Infrastructure-as-a-Service (IaaS)

Infrastructure as a Service is a provision model in which vendor provides access to virtualized computer hardware resources, including machines, network, and storage. The client typically pays on a per-use basis. Refer Figure#2 for more details. Infrastructure as a Service is sometimes referred to as Hardware as a Service (HaaS).

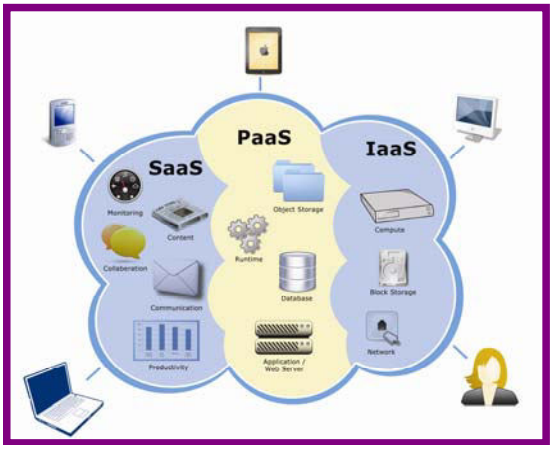


Fig. 2 : Cloud Delivery Models

VI. MAJOR CLOUD COMPETITORS IN TODAY'S MARKET

Fig. 3 : Major Cloud Competitors in today's Market

Gmail : A web mail from Google Inc. Webmail is an email client implemented as a web application accessed via a web browser. It is an oldest and popular SaaS provided by players like Microsoft, Yahoo, Google, etc. Most of them are ad-revenue driven and free for public.

Go Daddy : They are into internet domain registration, web hosting, email hosting, etc.

Net Suite : They are into cloud based integrated business management software delivered in SaaS model

Amazon : Amazon Elastic Computing Cloud (EC2) is a well known service of Amazon Web Services (AWS). EC2 allows users to rent virtual computers on which to run their own computer applications. EC2 allows scalable deployment of applications through a web service

Rack Space : Rack space is into web hosting, application hosting, email hosting, cloud servers, cloud storage, etc. Their products are implemented on different type of cloud delivery models ranging from PaaS to IaaS

GoToMeeting : is a Web-hosted application created and marketed by Citrix Online, a division of Citrix Systems. It is a remote meeting and desktop sharing software that enables the user to meet with other computer users, customers, clients or colleagues via the Internet in real-time. This will comes under SaaS model.

MS Office 365 : is a commercial application containing services which offer SaaS. Microsoft Office 365 for professionals and small businesses is a subscription service that combines the familiar Microsoft Office Web Apps with a set of web-enabled tools that are easy to learn and use, that work with your existing hardware,

Dropbox, Asigra, 3X systems, Jungle disk : These are the cloud storage vendors for web based files hosting services, that used for cloud storage to enable users to store and share files and folders with other the internet.

VII. LARGE ENTERPRISES CONCERNS REGARDING CURRENT CLOUD PROVIDERS

- How can Cloud providers secure our products or data?

  Cloud security remains a top concern for enterprise cloud deployments, protecting information is everyone's responsibility, as per current security tools or layers organizations can only put the lowest-risk data and applications into the cloud. Cloud vendor required to justify and prove how they will secure critical products and customer data, they will also required to show all the security layers are in place

- Can they utilize our manpower to support our product?

  The question in large organizations before they move their products or data into vender cloud infrastructure, since our current internal team has lot of experience and Knowledge on our products, would same vender take our manpower to support our products instead of requiting new manpower.

- Do they use onshore team to support infrastructure?

  Most Lot of high risk application can not support by offshore teams due to Federal Laws, in those cases vender has to provide support with local teams only. So organizations required to confirm same with provider before they move to vender

cloud Infrastructure. Condition like those will increase operational costs to Providers.

- Where will Cloud providers place Cloud servers?

  Organizations are concerned about Cloud infrastructure physical location where provider will install our application and store the sensitive data., because as per Federal laws high rated applications and data required to be local, so this will increase operations costs.

- Can cloud venders provide Disaster Recovery strategy?

  Disaster recovery from an IT perspective is making sure we can recover our systems and our data. Disaster recovery in the cloud is a relatively new concept, but recovery plan is the main aim of the entire IT disaster recovery planning project. For these plans that cloud Vender will set out the detailed steps needed to recover your IT systems to a state in which they can support the business after a disaster, those plans required to test before organizations adopts Cloud concept.

- Can we have our own cloud?

  This is still an ongoing effort. All the organizations are in the process determining same, for what services can go for vender, which we still required to keep in-house cloud. I personally believe, high rated application ad data still required to be in-house cloud.

- Should vender cloud infrastructure be used for all applications or only those with certain criticality ratings?

  This is one of the main concerns for large organizations, since security is not in place for 100%. I believe its good idea to have vender cloud for less critical application.

- How can we address latency?

  Latency means the time taken for data to be transmitted over the Internet between the provider and the customer, higher latency values mean longer response times. This is the one of the main issue for large organizations and cloud providers, many cloud providers are still failing to get good grips with latency issues in the cloud, many organizations may locate their office in a different country from their datacenter, datacenter is one side of the world and their customer from other side of the world, latency issue can be any reason, so cloud providers should consider this before they setup cloud Infrastructure for any organizations, good idea to have servers in multi location based on their customers.

- How can we address external network issues?

  No one can control or predict external network issues, so cloud providers should have at least few net providers, if one provider servers goes down or performance issues, backup one can address same.

VIII. CLOUD SECURITY

As we all know, there are a number of security issues and concerns associated with cloud computing but these issues falling into two mainly categories: Security issues faced by cloud providers who are providing SaaS, PaaS and IaaS via the cloud and security issues faced by their customers. In most cases, the cloud provider must ensure that their infrastructure is secure and their client's data and applications are protected. This enables the client to be able to communicate to their customer that the cloud provider has taken the proper security measures to protect their information.

Sometimes Cloud Computing Security is an evolving sub-domain of computer security, network security, and, more broadly, information security.

Cloud providers need to ensure that data is secure, it cannot be accessed by any unauthorized users or simply lost, and that data privacy is maintained. Cloud providers will need to justify the following areas:

8.1 Protecting the Data

To be considered protected, data from one customer must be properly segregated from another. It must be stored securely when "at rest" and it must be able to move securely from one location to another. Cloud providers should have systems in place to prevent data leaks or access by third parties. Proper separation of duties should ensure that auditing and monitoring cannot be defeated.

8.2 Identity management

All enterprise will have its own identity management system to control access to information systems and computing resources. Cloud providers either integrate the customer's identity management system into their own infrastructure, using SSO technology, or provide an identity management solution of their own.

8.3 Physical and personnel security

Providers must ensure that physical machines are adequately secure.

8.4 Availability

Cloud providers must ensure that applications available as a service via the cloud are secure by implementing application security measures and application-level firewalls in the production environment.

8.5 Application security

Cloud providers must ensure that applications available as a service via the cloud are secure by implementing application security measures and application-level firewalls to be in place in the production environment.

8.6 Privacy

Cloud providers must ensure that all critical data are masked and that only authorized users have access to that data in its entirety.

IX. COMPARISON BETWEEN GRID AND CLOUD

9.1 Grid

- Grid is a form of distributed computing whereby a super virtual computer is composed of many networked and loosely coupled utility computers acting together to perform very large tasks
- The programs are executed in a special type of parallel computing model that relies on a complete computers connected to the network by conventional network interface.
- Since various processors and local storage areas in a grid do not have high-speed connections, this arrangement is thus well-suited to applications in which multiple parallel computations can take place independently, without the need to communicate intermediate results between processors (non-interactive workloads)
- For a job to be suited to grid computing, the code needs to be parallelized. The source code should be structured to create separate tasks out of the program. Then the controlling unit assigns each task to an available node. These tasks need not be non-interactive, however messages sent between tasks increase the time factor
- As the tasks complete on various computing nodes, the results are sent back to the controlling unit, which then collates them forming a cohesive output. Refer to Figure#4 for more details

  Eg: BOINC, ROCKS, gLite, Sun Grid engine, etc

9.2 Virtualization

- Virtualization refers to abstraction of computing resources like processor cycles, ram, storage, network, etc from a physical computer (host) and present them to more than one virtual computers (virtual machines) that acts like a real computer with an operating system in a transparent way
- Piece of software that is responsible for this abstraction of physical resources is referred to as Hypervisor
- Software executed on these virtual machines is separated from the underlying hardware resources
- There are numerous advantages of this technique such as optimal usage of resources, ease in manageability and configuration, scalability, etc. Eg: KVM, Xen, ESXi, Hyper-V, etc. Refer to Figure#4 for more details

9.3 Cloud

- Cloud computing is the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a metered service over a network
- Users of a cloud request this access from a set of web services that manage a pool of computing resources (i.e., machines, network, storage, operating systems, application development environments, application programs)
- When granted, a fraction of the resources in the pool is dedicated to the requesting user until he or she releases them
- It is called “cloud computing” because the user cannot actually see or specify the physical location and organization of the equipment hosting the resources they are ultimately allowed to use. That is, the resources are drawn from a “cloud” of resources when they are granted to a user and returned to the cloud when they are released
- The decoupling of the VM from the underlying physical host allows the same VM to be started on different host. Likewise abstraction of resources from a host allows them to be allocated to some other VM running on another host. Thus virtualization is seen as an integral part and enabler for cloud computing, allowing the cloud provider the necessary flexibility to move and allocate the computing resources requested by the user wherever the physical resources are available.

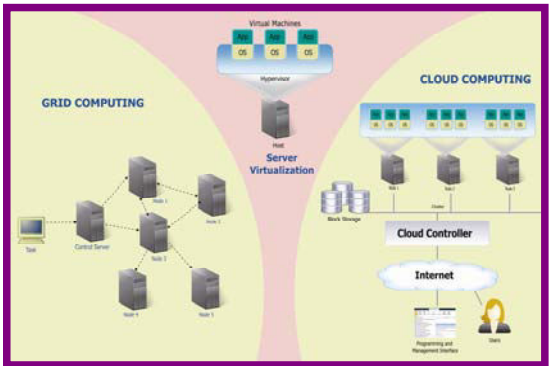


Fig. 4 : Comparison between Grid, Virtualization and Cloud

Each request for resource allocation in a Grid generally calls for more than the available resources and only a few of these allocations can be serviced at a time and others need to be scheduled for when resources are released. This results in sophisticated batch job scheduling algorithms of parallel computations. But in a cloud, the allocations need to be real-time and in fact there is no provision for queuing. This is a completely different resource allocation paradigm, a completely different usage pattern, and all this results in completely different method of using computing resources. In a nutshell, grid is great if you have an app that needs a lot of combined compute cycles and Virtualization/cloud is great if you have a lot of apps that need little compute cycles each. Refer to Fig. 4 for more details

X. MAJOR CLOUD PLATFORM VENDORS

OpenNebula is an open-source cloud computing toolkit for managing heterogeneous distributed data center infrastructures. The OpenNebula toolkit manages a data center's virtual infrastructure to build private, public and hybrid IaaS (Infrastructure as a Service) clouds.

Nebula toolkit will combine both data center resources and remote cloud resources. The toolkit includes features for integration, management, scalability, security and accounting. It also emphasizes standardization, interoperability and portability, providing cloud users and administrators with a choice of several cloud interfaces

Nimbula was founded by Chris Pinkham and Willem Van Biljon who led the team that created Amazon's Elastic Compute Cloud (Amazon EC2). Nimbula Director is a Cloud Operating System that allows users to implement IaaS (Infrastructure as a Service) style private, public and hybrid clouds.

The software is aimed at both enterprise customers and service providers. It can manage both on- and off-premise infrastructure through a Web UI, an API or a command line interface.

Enomaly Inc founded in 2004, initially founded as an open source consulting company by Reuven Cohen, George Bazos and Lars Forsberg, the company quickly grew from an open source consultancy and system integrator into one of the first focused on the emerging cloud computing space. The company was among the first to provide a self-service Infrastructure-as-a-Service (IaaS) platform with the first version launched in 2005 under an open source platform called Enomalism, later renamed ECP and made available commercially as closed sourced.

Enomaly's current software, called Elastic Computing Platform, Service Provider Edition (ECP/SPE) was released in July, 2009 and allows web hosts and service providers to offer public facing IaaS and cloud services to their customers in a means similar to that of Amazon Ec2.

In November 2010 Enomaly launched SpotCloud.com, described as the first commodity style Clearinghouse & Marketplace for unused cloud computing capacity. According the SpotCloud site the service is "Built on Google App Engine and the Enomaly ECP platform SpotCloud is an easy to use, structured cloud capacity marketplace where service providers can sell their excess computing capacity to a wide array of buyers and resellers.

OpenStack is an IaaS cloud computing project by Rackspace Cloud and NASA. Currently more than 120 companies have joined the project among which are Citrix Systems, Dell, AMD, Intel, Canonical, SUSE Linux, HP, and Cisco. It is free open source software released under the terms of the Apache License.

OpenStack integrates code from NASA's Nebula platform as well as Rackspace's Cloud Files platform. In July 2010, Rackspace Hosting and NASA jointly launched a new open source cloud initiative known as OpenStack. The mission of the OpenStack project is to enable any organization to create and offer cloud computing services running on standard hardware.

Eucalyptus implements IaaS (Infrastructure as a Service) style. Eucalyptus can use a variety of virtualization technologies including VMware, Xen and KVM hypervisors to implement the cloud abstractions it supports.

Eucalyptus enables the creation of on-premise private clouds out of organizations existing data centre. Eucalyptus implements IaaS private cloud that is accessible via an API compatible with Amazon EC2 and Amazon S3. This compatibility allows any Eucalyptus cloud to be turned into a hybrid cloud, capable of drawing compute resources from public cloud.

Cloud.com implements IaaS (Infrastructure as a Service) style private, public and hybrid clouds. Their software, CloudStack, is designed to make it easier for Service Providers and Enterprises to build, manage and deploy IaaS offerings similar to Amazon EC2 and S3.

CloudStack is available in three editions: the Enterprise Edition, the Service Provider Edition and the open-source Community Edition.

AppLogic, flagship product of 3Tera, is a turn-key cloud computing platform for scalable applications and web services. AppLogic is the first cloud computing platform that is designed for distributed applications.

It uses advanced virtualization technologies to ensure complete compatibility with existing operating systems, middleware and applications. As a result, AppLogic makes it easy to move existing web applications into the cloud without modifications.

You'll be able to deploy and scale existing applications without changing code and architecture. You'll integrate, monitor and scale applications and infrastructure using just a GUI, and you can add or remove resources and storage without disrupting users.

Quickly offer new services such as SaaS, PaaS, IaaS

Scale applications without changing code and architecture

Easily replicate services for other departments or customers

We can Build a flexible, efficient datacenter with VMware vSphere. Run business critical applications with confidence and respond faster to your business needs with VMware vSphere, the industry-leading virtualization platform for building cloud infrastructures.

vSphere accelerates the shift to cloud computing for existing datacenters, With over 250,000 customers worldwide and the support of over 2500 applications from more than 1400 ISV partners, VMware vSphere is the trusted platform for any application. Discover for yourself why vSphere is the #1 virtualization platform in the industry.

Windows Azure is an open and flexible cloud platform that enables you to quickly build, deploy and manage applications across a global network of Microsoft-managed datacenters. You can build applications using any language, tool or framework. And you can integrate your public cloud applications with your existing IT environment.

Windows Azure enables you to easily scale your applications to any size. It is a fully automated self-

service platform that allows you to provision resources within minutes. Elastically grow or shrink your resource usage based on your needs. You only pay for the resources your application uses. Windows Azure is available in multiple datacenters around the world, enabling you to deploy your applications close to your customers.

AppScale allows users to upload multiple App Engine applications to a cloud. It supports multiple distributed backends such as HBase, Hypertable, Apache Cassandra, MySQL Cluster, and Redis.

AppScale is an open-source framework for running Google App Engine applications. It is an implementation of a cloud computing platform (Platform-as-a-Service), supporting Xen, KVM, Amazon EC2 and Eucalyptus

Cloud Foundry is an open source cloud computing platform as a service (PaaS) software developed by VMware released under the terms of the Apache License 2.0. It is primarily written in Ruby. Cloud Foundry is an open platform as a service, providing a choice of clouds, developer frameworks and application services.

Cloud Foundry makes it faster and easier to build, test, deploy and scale applications.

XI. CONCLUSION

As we all know, there are several advantages of cloud computing. Here's a summary.

Remote Access : With cloud computing, your business is not restricted to be in a particular location. This applies to individuals also. You can access the services from anywhere. All you need is your ID and password. In some cases, there may be extra security requirements, but once in place you can easily access your cloud services from any part of the world.

Optimal usage of resources (and easy of management in case of public cloud) : You are aggregating all the resources of all physical computers in a cluster to make them available to all of the virtual computers running on it. You will monitor resource usage in real time, get alerts for peak resource consumption, allocate resources based on policies, etc. You find any surplus resources in a cluster, you move in a virtual computer running on some other cluster running low on resources, or you will shutdown one or few physical computers. All this mechanism ensures optimal usage of resources and that means lower power and air conditioning requirements.

In case you are on a public cloud, you can do away with all the hassles of managing and administering computers, networks, storage, operating systems, applications, etc. and it implies fewer technical head count and lesser expense.

Highly scalable environment (and much lower CAPEX in case of public clouds): Cloud is very scalable. You can add more physical nodes should you require more resources in any cluster and the beauty is that the newly added resources are portable and can be assigned to any virtual machine running on the cluster.

In case you opt for public cloud, you can add more resources quickly whether you are a startup or in need of more computing resources for a short time without any upfront capital expenditure. In most cases, the entire process is automated so the expansion takes just a few minutes. The same is applicable if you wish to use fewer resources. One of the best advantages of cloud computing is easy re-allocation of resources.

Data security in cloud : Despite most people's doubts that cloud computing is not secure, the reality is that clouds are as secure as your data centre is and it all depends on how you perceive security. As regards to outage is concerned, clouds offer on-the-fly load balance, backup and disaster recovery which results in data safety and cheaper business continuity options.

With respect to threats originating from external or internal sources, clouds are susceptible to damage based on the fact that numerous systems will be sharing the same infrastructure, if the operator's infrastructure's security is breached, it will affect all the systems sharing that infrastructure. So, we are expected to be more cautious while planning perimeter security of the infrastructure, patching of applications on a regular basis and isolation of instances to the possible extent.

XII. PREDICTIONS ABOUT HOW THE CLOUD WILL EVOLVE IN THE FUTURE

- Faster proliferation of clouds : Recently, market research firm Ovum published the findings of a survey aimed at large, multi-national corporations on cloud computing. The report found that cloud adoption is up 61 percent from 2010 and well known benefits, such as scalability and cost-reduction were among the reasons respondents

chose to shift IT to the cloud. The findings of the Ovum survey coincide with other market analytics within the field. According to a report by Advanced Micro Devices, 70 percent of businesses worldwide will make the switch to the cloud within the next couple of years.

- System or infrastructure outages will decrease with cloud adoption : As we know cloud environment is cheaper than having our internal datacenter, so we can have more redundancy sites than before.
- Cloud mega-data centers : In future cloud providers required to maintain mega data centers, because more organizations are going to adopt cloud concept, so venders are required to have more resources in datacenter. As the connectivity options increase with decreased bandwidth costs, organizations tend to have more hybrid clouds using different services from different vendors on a robust network backbone
- Open source will dominate the future of the cloud : I personally believe open source is playing a dominant role in Cloud computing changing the platform and application dynamics. Open source is the foundation for cloud computing although ideas vary across the environment. Open source is also helping make the cloud more scalable
- Cloud Standards will emerge : With increased usage and dependence on cloud, we look forward to see emergence of some standards regarding interoperability, portability, security framework, etc either through intervention of Federal agencies or through vendor association bodies
- Government will adopt cloud : Government is a big consumer of IT in any nation. By looking at the benefits cloud is offering such as energy conservation, better manageability, security and fault tolerance, Government will itself embrace cloud technology in very near future.
- SaaS will lead the band : SaaS is oldest cloud delivery model we have seen and email web application is the most popular of them. More applications are delivered on SaaS model such as office suites, CRM, ERP, etc. It is widely discussed that almost all the applications will be delivered over browser in near future except a few like media development tools which rely more on local computing resources, thanks to the decreased ratio of price to bandwidth where dependable connectivity is no more a luxury and advent of latest web technologies like HTML 5 and CSS 3

  We are also looking at a gaining ground for web based operating systems like Google Chrome OS, eyeOS, Glide, JoliCloud, etc. We don't have any operating system installed on our computers except but a small firmware which will connect to the service provider's cloud and provide us the OS as well as applications based on our subscription. More corporates are even embracing Virtual Desktop Infrastructure (VDI) where personalized OS instance is delivered on to the client's browser including personal storage.

  Even some cloud management platforms are being offered on SaaS foot print. Eg. RightScale is an web based application enabling management of our cloud infrastructure from multiple providers including private and hybrid clouds. RightScale enables users to migrate workloads between their private clouds and public clouds operated by Amazon Web Services (AWS), Rackspace, etc
- Convergence of grid and cloud computing models : Some startup may come up with ideas such as building a commercial cloud offering based on grid of public computers with metering and monetizing of those contributions, similar as solar energy fed back to the national electricity grid, aka, buy computing resources from public on grid model and sell them on the cloud. GridGain (http://www.gridgain.com) is a step towards integrating both of these technologies for maximizing benefits of both worlds
- Computing will be traded on bourses like a commodity : Enomaly, a cloud solution provider has already offering a service called spotCloud where you can even sell excess computing capacity of your datacenter

REFERENCES

[1] The Cloud at Your Services. Jothy R and Arthur Mateos

[2] Cloud Security A Comprehensive Guide to Secure Cloud Computing by Ronald L. Krutz AND Russell Dean Vines.

[3] Clod Computing Prod and Cons for End Users by Michael Miller from http://www.informit.com/articles/article.aspx?p=1324280&seqNum=3

[4] Clod Computing over View. Retrieved September 15 2011 from Wikipedia http://en.wikipedia.org/wiki/Cloud_computing

[5] Clod Adoption up in large companies 6/7/2011 from http://www.navisite.com/news-and-events-cloud-computing.htm

[6] Types of Cloud Computing Retrieved September 9 2011 http://www.allcovered.com/ technology/ cloud-computing/types-of-cloud-computing.asp

[7] A web mail from Google Inc info Retrieved November 24, 2011 http://en.wikipedia.org/ wiki/Webmail

[8] Go Daddy business model with cloud Retrieved November 24, 2011 http://en.wikipedia.org/ wiki/web_hosting

[9] Net Suite business model Retrieved November 24, 2011 http://www.netsuite.com/ portal/home.shtml

[10] 3Terea services model Retrieved November 24, 2011 http://www.3tera.com/AppLogic/

[11] VMware services info Retrieved November 24, 2011 http://en.wikipedia.org/wiki/Vmware

[12] Eucalyptus cloud services models Retrieved November 24, 2011 http://open.eucalyptus.com/learn/what-is-eucalyptus and http://open.eucalyptus.com/ learn/what-is-eucalyptus

[13] Amazon Web services cloud computing services models Retrieved November 24, 2011 http://en.wikipedia.org/wiki/Amazon_EC2

[14] Rack Space business model in cloud computing info Retrieved from http://en.wikipedia.org/ wiki/Rackspace

[15] Right Scale business roles in cloud computing Retrieved form http://en.wikipedia.org/ wiki/RightScale

[16] IBM cloud business model info Retrieved November 24, 2011 http://www.ibm.com/cloud-computing/us/en/

[17] Yahoo cloud computing business info Retrieved November 24, 2011 http://labs.yahoo.com/Cloud_Computing

[18] Cloud Vendors details are reviewed from Wikipedia and providers websites

◈◈◈

Dynamic Load Balancing for the Cloud

Venubabu Kunamneni

Bellingham MA - USA – 02019

Abstract **-** This article on dynamic load balancing for a cloud environment will provide a basic over view on load balancing, static load balancing, software and hardware load balancing vendors in the market, and dynamic load balancing functionality for the cloud environment. The goal of a cloud-based architecture is to provide elasticity, and the ability to expand capacity on-demand. Dynamic load balancing is required for a cloud implementation. The load balancing feature effectively reduces the waiting time at each process step and the lead time of all products.

I. INTRODUCTION

Load balancing is the technology to distribute workload across multiple computers or a computer cluster, central processing units, disk drives, RAM, or other resources, to achieve optimal resource utilization, maximize throughput, minimize response time, avoid overload, and minimize application down time.

The load balancing services is usually provided by dedicated software or hardware.

Most of the times a single webserver is insufficient to handle the amount of traffic or load received. In this situation we have several options. If we have the ability to add new webservers into our setup the following load balancing options are available based on our environment and requirements.

- Purchase a piece of dedicated load-balancing hardware from vendor
- Use a software solution such as:
    - Round-robin DNS
    - Load balancing with a software

The hardware solution might be the best one, but if we don't have the funds for on dedicated hardware then a software-only solution might be our only option.

Round-robin DNS gives you the ability to setup a pair of two or more machines and have users "randomly" connect to a different host. This is simple and reasonably effective, however it doesn't give you much redundancy. If one machine fails then some users will still be sent to that host, and will receive errors. This approach is not ideal for high rated apps where zero impact to our customers is requirement.

There are a few advantages and disadvantages with DNS load balancing

Advantages

- Relatively simple configuration
- Effective distribution of traffic among multiple servers

Disadvantages

- No mechanism to deal with unavailable servers
- No means of ensuring that subsequent requests will be fulfilled by the same server
- Each server will require public IP addresses

II. WHY DO WE NEED LOAD BALANCING?

Consider a scenario where we only have one web server in operation to handle all incoming requests to your company website. When the business is being established, it may be possible to handle the volume of traffic your site receives with one web server. However, as the business grows, the one server will no longer be sufficient. If we don't add new web server instances our WebPages will load slowly and you will have users waiting till the server is free to process client requests. This is not a good thing as people are not very good at playing the waiting game. In these days where the market is very competitive customers service is hugely important or potential customers will move to the competition.

When you have multiple web servers in your server group, the incoming traffic can be evenly allocated among the different servers. This process is called static load balancing (see Figure2). However, it will only appear to the client as one server only rather than several. A case in point is the internet browser. The purpose include:

- To spread the load amongst a number of machines/locations

- To provide redundancy in case one machine/server fails
- To provide zero down time during patch installations on servers or updates to applications on server

III.COMPANIES THAT OFFER TECHNOLOGIES SUPPORTING LOAD BALANCING

Fig. 1: Companies that offer technologies supporting load balancing

IV. FEW LOAD BALANCING PRODUCTS FUNCTIONALITY

A10 Networks : This is a privately held company specializing in the manufacturing of application traffic delivery controllers with software and hardware. Founded in 2004, it originally serviced just the identity management market. In 2007, A10 Networks launched AX Series, a family of application delivery load balancing appliances, The AX Series application delivery controllers consists of 12 models, including nine 64-bit and three 32-bit models.

Array Networks : Founded in 2000 and headquartered located in Silicon Valley, California, Array Networks is a global technology company that addresses problems related to securely delivering enterprise applications to end users. Array takes a hardware-based load balanced approach to deliver enterprise applications, including SPX series SSL VPN enterprise platforms. It delivers applications for Enterprise Platforms, and Net Cert PKI Certificate Management Solutions.

Barracuda Networks : Barracuda Networks, Inc. is a privately held company providing security, networking and storage solutions based on appliances and cloud services. The company's security products include solutions for protection against email, web surfing, web hackers and instant messaging threats such as spam, spyware, trojans, and viruses. Barracuda Networks was established in 2003 and helps provide a company's networking and storage solutions including web filtering, load balancing, application delivery controllers, backup services and data protection. As of October 2009, Barracuda had over 85,000 customers. As of November, 2011, Barracuda had more than 130,000 customers.

Citrix Systems : Citrix is publicly held company and multinational corporation founded in 1989, that provides server and desktop virtualization, load balancing, software-as-a-service (SaaS), and cloud computing technologies, including Xen open source products. Citrix currently services around 230,000 organizations worldwide and is based in Fort Lauderdale, FL. Citrix load balance product is called as a NetScalet load balancer. The NetScaler load balancer continuously monitors the availability and health of not only the server hardware, but the state of back-end databases and applications. Network links, operating systems, and even individual application elements are also monitored by the NetScaler load balancer. Some load balancers simply provide basic ping.

F5 Networks : This publicly held company is a networking appliances company. It is headquartered in Seattle and, has development and marketing offices worldwide. It originally manufactured and sold some of the very first load balancing products. In 2010, F5 Networks was featured in Fortune's 100 Fastest-Growing Companies list. The F5 BIG-IP network appliance was originally a network load balancer but today also offers other functionality such as access control and application security. F5 offers products in various segments of the Application Delivery Controller market. According to Gartner, F5 has "a continued market-leading position in the Application Delivery Controller market.

Inlab : Inlab Software GmbH is an independent software vendor located in Grünwald, Germany.It develops and markets load balancing software, networking system software, and programming languages. Inlab's main product is a software TCP/IP load balancer for Linux and Solaris operating systems. Inlab's Balancing is used at many international commercial and academic customer sites.

Radware : Radware is a provider of integrated Application delivery, Network Security and Load balancing solutions based in Tel Aviv, Israel. Radware,

which is a member of the Rad Group of companies, is a public company and its shares are traded on NASDAQ. In April 2007, Radware acquired Covelight Systems, a provider of web application auditing and monitoring tools based Cary, North Carolina, for $16 USD million. In February 2009 Radware acquired Nortel's Application Delivery business.

Brocade : Brocade Communications Systems, Inc. is an American multinational corporation and a technology company specializing in data and storage networking products. The companys product portfolio spans across enterprise (LAN, WLAN) Switches, WAN (Internet) Routers, SAN Switches, Application Delivery Controllers, Network Security Appliances, Ethernet/Storage Network Adapters and PHY Transceivers. Founded in 1995, Brocade Communications is headquartered in San Jose, California, USA.

V. COMPARISON BETWEEN STATIC AND DYNAMIC LOAD BALANCE

Here are the most significant difference between static and dynamic load balancing:

- The most significant difference between static and dynamic LB is not having to allocate idle resources for dynamic Load balancing. Resources are allocated when needed and released when no longer required. This minimizes expensive allocation of idle resources.
- Adding new instances into static load balance configuration requires some time and testing, but with dynamic load balance everything will be automatic.
- Manual mistakes can happen in static load balance but are not present in dynamic load balancing
- Environment change approvals are required from management to add new instances into static environments but not for dynamic environments.
- The challenge in developing applications is making them state agnostic. By creating applications that allow its sessions to be transferred from one site to another allows for dynamic allocation. Hence the single largest benefit of dynamic vs. static allocation.

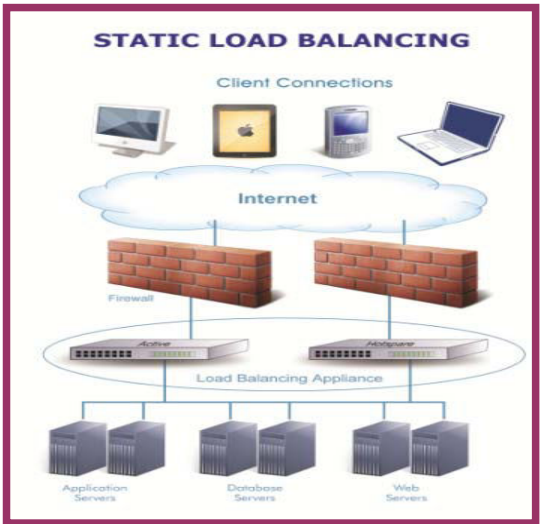


Fig. 2 : Static Load balancing overview

VI. WHY IS DYNAMIC LOAD BALANCING REQUIRED FOR A CLOUD ENVIRONMENT?

Dynamic load balancing is a major key for a successful implementation of cloud environments. The main goal of a cloud-based architecture is to provide elasticity, the ability to expand and contract capacity on-demand. Sometimes additional instances of an application will be required in order for the architecture to scale and meet demand. That means there is a need for a mechanism to balance requests between two or more instances of that application. The mechanism most likely to be successful in performing such a task is a load balancer. .

There's no other way to assume increased load other than adding new instances and distributing that load with software or hardware Similarly, when the additional instances of that application are de-provisioned, the changes to the network configuration need to be reversed, but software and hardware load balance is easy to scale up or scale down.

Obviously a manual process would be time consuming and inefficient, effectively erasing the benefits gained by introducing a cloud-based architecture in the first place. The below is an example of how dynamic load balancing can be implemented.

1. Let's assume that cloud management console, or a custom developed application or cloud tool kit, triggers an event that indicates a new instance is required to maintain availability. How it determines capacity limitations may be based on VM status via VMware APIs or data received from the load balancer, or a combination both(see Fig 2).

2. A new instance is launched in the cloud Environment for same application. This is accomplished via the cloud management console or cloud tool kit (see Fig 2).

3. The cloud management console or tool kit grabs the IP address of the newly launched instance and instructs the load balancer to add it to the configuration as new resources for same application. This is accomplished by the standards-based API which presents the configuration and management control plane of the load balancer to external consumers as services (see Fig 2).

4. The load balancer adds the new application instance to the appropriate configuration and as soon as it has confirmation that the instance is available and responding to requests, begins to direct traffic to that new instance without disturbing existing instances (see Fig 2).

This process should be easily reversed upon termination of an instance, load balancer should be able to release termination instance IP. Note: there may be other infrastructure components that are involved in this process that must also be considered on launch and decommission, but for this discussion we're just looking at the load balancing piece as it's critical to the concept of auto-scaling.

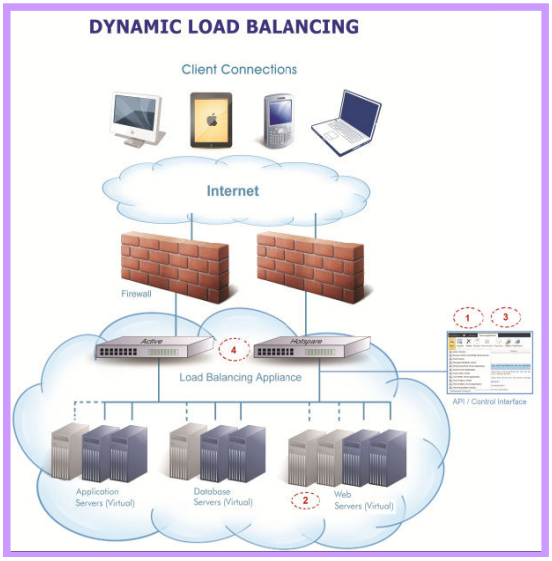


Fig. 3 : Dynamic load balancing overview

VII. PROS OF LOAD BALANCING

- Esnures that connections are not directed to a server that is down.
- Good for scaling out for multiple clusters on different segments.
- Is highly configurable, with rules allowing for client affinity, weighting, filtering, availability etc
- Works as a driver rather than as a service
- Allows for mixed-version clusters
- Manages resources efficiently
- Utilizes all the systems resources as efficiently as possible
- Improves the application response time by sending traffic round robin
- If we have two members in load balance pool, with priority function we can send all the traffic to one node and keep other node as a backup
- Helps with disaster recovery
- OS and application patching is made easier by routing traffic to different during change windows (less customer downtime)

VIII. MY THOUGHTS REGARDING THE FUTURE LOAD BALANCING TECHNOLOGY

The following does not exist with current Load Balancing technology but should be considered:

- Automatic alerting of any issues with their members to a support group
- Dynamic traffic routing (based on incoming load) in allocating resources in a cloud environment
- Monitoring of the server response time and stop sending new requests if that that server is taking longer than designed
- Report longer server response times to a system administrator
- The elimination of the middle man (such as cloud management console) that is used to add and remove application instances (sharing a new instance IP to the load balanced application to be added as a new member into configuration). Load Balanced applications should be able to track the load , add/remove instances , and configure that IP into the required configurations

IX. CONCLUSION

Cloud Computing is a vast concept and load balancing plays a very important role. As we all know, there are several advantages with load balancing for IT Environments, especially the dynamic implementation. It was designed to help organizations achieve their availability objectives. Dynamic load balancing helps

with comprehensive failover capabilities in case of server failures, distribution of traffic across multiple servers, and disaster recovery.

REFERENCES

[1] http://en.wikipedia.org/wiki/Load_balancing_(computing)

[2] Load balancing Vendors (listed in Fig 1) websites also reviewed

[3] http://www.loadbalancing.org/

[4] http://devcentral.f5.com

◈◈◈