

Analysis Of Cloud Based Multitenant Load Scaling

Ankita Kaushik, Dr. Vikas Kumar, Dr. Amit Chaturvedi

Research Scholar, Department of Computer Science,
Bhagwant University, Ajmer
Email Id: ankita.home123@gmail.com
Head, Department Of Computer Science & Engineering,
Bhagwant University, Ajmer
Assistant Prof., Govt. Engg College, Ajmer

ABSTRACT : An internet based computing, through which resources are shared to customers .In multitenancy ,customers share same hardware and database .Many clients are working under one server and single instance of application is available to many customers . With minimum resource and cost , requirements can be fulfilled .With the help of scaling , resources can be allocated as per requirements. Increasing the potentiality of system, by adding /removing components or functionality or components of systems can be termed as scalability.

Keywords : Multitenancy, scaling, AGILE, elasticity

1. Introduction

Cloud computing which is internet based computing , through which resources and services are being provided to customers . Among many services , Software as a Service or SaaS is an important service where softwares are being provided to users on demand . A tenant is group of customers who have common access to resources . Multitenancy refers to application where customers (tenants) share same hardware resources ,one shared application and database instance . A technology that clouds use to share IT resources and securely among tenants , is called multitenancy .

One of the most important concepts for SaaS applications is multitenancy .SaaS serves both as business and for software purpose. SaaS applications usage in multitenant scenario , users nowadays login to software provider's web site and use software as and when required , but they don't need to install it .

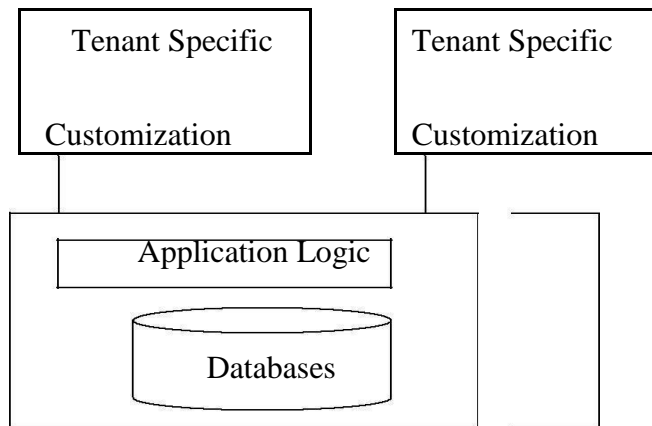


Fig :1 Multitenant Architecture Single , shared stack of S/w & H/w

Cloud computing , an upcoming technology , which helps in implementing services with on demand IT infrastructures , so that users can access services at any computer with the help of internet which is termed as Cloud Services . Three layers available are IaaS(Infrastructure as a service) ,PaaS(Platform as a service) and SaaS(Software as a service) .In order to take care of activities related to utilization of on demand virtual IT infrastructures , IaaS uses physical storage space and processing capabilities . PaaS layer main use is to provide middleware services to implement cloud services like security services or management services called multitenancy services. SaaS uses previous two layers to offer cloud services to end users .

Cloud scaling generally helps in allocating resources on demand , i.e. when needed . Its aim is to fulfill requirements of the applications with minimum resource and application cost . Service Level Objectives (SLO) should be met . Between a service provider and customer , a service level objective (SLO) is the most important factor . [1]SLO measure performance of service provider and used to avoid disputes . To scale up and down resource allocation at run time elasticity is major aspect in clouds . To maintain elasticity , allocation/ deallocation of virtual machine instances is done and here cost effectiveness is the major concern . VM level scaling and scaling at the resource level (CPU's, memory , I/O, etc) is needed .

Scalability ,in simple terms, is enhancing the potential of a system , network , or process to handle growing workload, where adding new functionality at minimum effort is functional scalability and in order to accommodate loads, components of system are added , removed called load scalability

2. Techniques used for resource scaling

1)Adding / removing VM's dynamically at run time subject extra overheads . Modifying VM's capacity can be done . A lightweight approach for cost effective scaling of cloud resources is being forwarded at IaaS cloud provider's side , provides for multi-tier applications implemented using multiple VM's . Virtual machine scaling and also resource level scaling is done at CPU,memory,I/O etc is being discussed in this paper.[2]

2)AGILE is elastic distributed resource scaling for infrastructure- as- a- Service . Resource management of the servers inside local VMs is continuously checked by AGILE , and future resource usage is predicted . [3] To provide scalability and increase ability , a Cloud Hosting Provider (CHP) which is a elastic web hosting provider is being presented by taking hold of outsourcing .

3) A model has been forwarded called a decision model to make the best use of costs , performance ,reliability , users and run time requirements considering SLA(Service Level Agreement) constraints like sufficient resource availability .[4]

4)PAC finds out significant patterns called signatures of different VMs. PAC efficiently finds out resource usage in applications and can minimize resource prediction errors . PAC handles load scaling in cloud and on – demand resource allocation . To continuously investigate VMs PAC : Pattern-driven Application Consolidation for Efficient Cloud Computing is being presented[5]

5)To increase efficiency of clouds infrastructure , an auto- scaling method with deadline and budget constraints is being proposed , so as to scale up and down VMs dynamically to handle load and deal with performance issues . Cloud applications are enabled to complete tasks within deadline by regulating work and reducing user cost . Two important factor taken into consideration are performance and budget . From performance point of view , with cloud scaling methods , cloud applications complete their tasks within given time limit by acquiring VM instances

.Taking cost into consideration ,by having instance types of VMs , cost effective working can be regulated in the clouds . This methodology has been implemented in Windows Azure Platform . [6]

6)In today’s cloud scenario , a new cost effective elastic provisioning approach called Kingfisher has been proposed to reduce time to transition to new configuration , cost can be minimized of virtual server resources . The computing infrastructure presents server configurations for rent and price . At run time, provisioning of resources , optimization is key factor . Optimization is considered as infrastructure cost for again configuring an application, to enhance capacity is considered . In public cloud , such as Amazon EC2 this approach can be implemented with the given case study .Using OpenNebula this approach has been implemented and evaluated successfully . Multiple mechanisms have been integrated , and cost effective configuration for enhancement in capacity and transitioning among different configurations is considered .[7]

8) To transfer smartphone applications to cloud , a framework ThinkAir , is presented . Smartphone are becoming more efficient and organized . ThinkAir needs few changes to an application’s source code , more experiments and evaluations are done . ThinkAir enhances elasticity and scalability of cloud and increases capacity of cloud computing using virtual machine (VM) . Smartphones ranging from Android to Xen have been

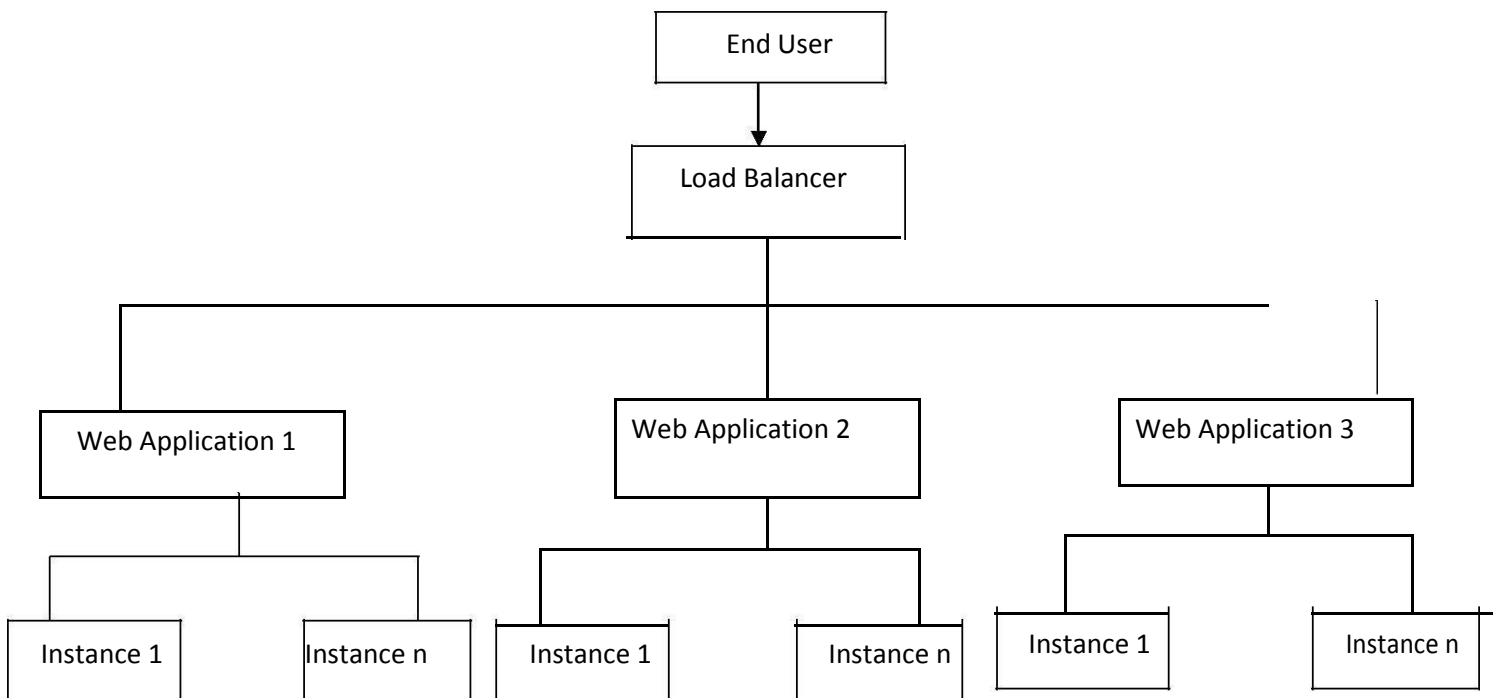
ported to be run on cloud applications , and we continue to work on applications . ThinkAir uses smartphone virtualization in cloud . Execution time and energy consumption decreases for N-queens puzzle.[9]

9)Ability of clouds to add or remove VMs and to predict future load an adaptive hybrid elasticity controller is being proposed [10]

3. Consideration regarding scalability : Low Speed of internet

Internet is nowadays a boon for upcoming industries , companies and business . Internet provides global information it is generally called mesh of networks , providing variety of information and communication facilities .In multi tenant environment low speed of internet is becoming a major issue . Cloud computing is based itself on internet and if there is low speed this can arise dissatisfaction among general users .

Figure : Multitenancy Load Scaling



4. Resource Scaling : Analysis

Low resource allocation

Multiple users consume resources of cloud . At any point in time , many users can access resources , many users are accessing resources of cloud applications . Many other applications

users , sometimes try to access same cloud resources. If load increases , there is lack of resources which is termed as low resource allocation .

High resource allocation

When more resources are allocated than the requirement , excess of resources are not utilized . As without having proper knowledge if supply is more than demand excess of resources get waste . Cloud Computing is based on pay as u go , as users have to pay only for usage . But without having prior knowledge of resource management , users who are in need don't get proper supply but users who are already having face wastage .

So there should be some proper scaling of resource allocation so as to dynamically allocate resources to users on run time .

Traditional methods of virtual machine scaling of cloud resources may use resources in excess so resource provisioning in cloud applications is being provided with VM level elasticity and at resource level itself

(CPU's, memory, I/O) [2]

To deal with load of cloud applications , number of virtual machines (VMs) assigned to cloud applications must be dynamically adjusted , but future knowledge is not available . So to deal with such issues , resource demand prediction can be done [3]

To deal with resource management limitations , an elastic web hosting provider is being presented which is according to SLA (Service level Agreement) . It provides solution to dynamic load received by applications and also provides scalability and high availability capabilities . [4]

To minimize resource wastage and also to enhance resource sharing in clouds a new Pattern Driven Application is being proposed which can discover significant patterns called signatures of different applications and hosts . Resource Prediction errors can be reduced by 50-90 % which is better than old methods[5]

A new scheme , Cloud auto scaling deals with business works does automatic scaling of resources according to workload change [6]

Replication and migration of cloud capacity is done by dynamically allocation of resources [7]

With the help of ThinkAir, smartphone applications migration to clouds and virtualization is now possible . Multiple VM's can be executed as per requirement . Many virtual machine

images can be used for parallelizing method execution to increase elasticity and scalability [8]

For running of cloud applications dynamically , number of VM's can be changed as per demand for that two adaptive hybrid controllers can be used . Hereover allocation of resources is reduced from 32% to 15% as compared to regression based elasticity controllers .[9]

Resources from three domains , compute , storage and network are acquired or released as per requirements , an integrated and autonomic cloud resource scaler or IACRS , scale resources integrate resources from three domains for effective scaling of resources [10]

5 . Imporatant factor : Cost Effectiveness

Without having proper knowledge of VMs , there are chances that cloud systems become over provisioned (costly) , so AGILE is elastic resource scaling method which reduces cost [2]

To maximize revenue of the providers of cloud computing applications an Elastic Cloud Hosting Provider is being introduced so that economic variables and functions can be studied and action should be taken to maximize its income .[3]

With signal processing techniques , using signatures , a feasible method is presented which little overhead[5]

Within budget stictness , automatic scaling does management of resources [6]

Cost effective system is being suggested that minimizes cost , here methods are present which helps in selection of a virtual server configuration . Infrastructure cost or transition cost can be reconfigured and upto 24 % cost can be reduced in total [7].

5. Performance Issues in scalability

Virtual machine cloning is being presented to reduce cloud applications startup time [2]

Back end server management needed for running a web application and how to replicate them.[3]

In virtualized cloud applications , when running large number of VM's performance can be improved by 50% .[5]

Workload change is taken into account and time constraints are considered as work is performed according to budget and deadline specified so as to achieve better performance in cloud auto-scaling methods [6]

The efficiency can be evaluated of kingfisher prototype , resource allocation provision can be dynamically provisioned . Using OpenNebula , approach can be implemented , efficiency can be evaluated on laboratory based Xen Cloud [7]

Execution time and energy consumption can be reduced for N-Queens problem , also Android to Xen have been experimented to run on clouds . Distributed feature of cloud can be used more efficiently by parallelizing feture of applications [8]

Traditional elasticity engines have 1.48 to 2.1 times SLA violation rates than adaptive hybrid elasticity controller ,This technique is also proves better for increasing loads [9]

Optimization of resource scaling is performed , here tenant can itself regulate what resources can be scaled according to condition so auto scaling feature . Network resources are generally not auto – scaled this limitation is reduced here ,also avoids fake scaling. [10]

6. Conclusion

Multitenancy allows single instance of software to serve many organizations by reconfiguring at same time according to different needs .Load scaling as we can say The ability by which the different resources can be expand and contract to adjust heavier or lighter loads or number of inputs and the ease by which components can be modified, added, or removed, so as to adjust changing load. We have studied a lightweight approach for cost effective scaling of cloud resources , AGILE technique: elastic distributed resource scaling for Infrastructure-as-a-Service, Cloud Hosting Provider (CHP) which is a elastic web hosting provider , a model called a decision model to make the best use of costs , performance ,reliability , users and run time requirements considering SLA, PAC: Pattern-driven Application Consolidation for Efficient Cloud Computing, Cloud Auto-scaling with Deadline and Budget Constraints, a new cost effective elastic provisioning approach called Kingfisher has been proposed to reduce time , a framework ThinkAir transfer smartphone applications to cloud , an Adaptive Hybrid Elasticity Controller for Cloud Infrastructures . We have analysed the performance , cost and resource scaling factors .

7. References

- 1)Jose M . Alcaraz Calero, Jose M. Alcaraz Calero HewlettPackard Labs and University of Murcia Nigel Edwards, Johannes Kirschnick, Lawrence Wilcock, and Mike Wray HewlettPackard Labs ,Toward a multi-tenancy, authorization system for cloud services , Copublished By The IEEE Computer And Reliability Societies , 1540-7993/10 , 2010 IEEE , November/December 2010
- 2)Rui Han, Li Guo, Moustafa M. Ghanem, Yike Guo , Lightweight Resource Scaling for Cloud Applications , CCGrid 2012 Department of Computing Imperial College London London, UK {r.han10, liguo, mmg, y.guo}
- 3)SLA-driven Elastic Cloud Hosting Provider J. Oriol Fito, Inigo Goiri and Jordi Guita Barcelona Supercomputing Center , Technical University of Catalonia Barcelona, Spain, Conference on Parallel, Distributed and Network-based Processing 1066-6192/10 , 2010 IEEE

DOI, 10.1109/PDP.2010.16

4) 2010 18th Annual IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication ,Decision Model for cloud computing under SLA constraints Arthur Andrzejak,Zuse Institute, Berlin (ZIB),Germany ,email : arturuni@gmail.com, Derrick Kondo,INRIA,France Email : derrick.kondo@inria.fr, Sangho Yi INRIA, France Email : sangho.yi@inria.fr

5) Computing Zhenhuan Gong, Xiaohui Gu ,PAC: Pattern-driven Application Consolidation for Efficient Cloud Department of Computer Science North Carolina State University {zgong}, 2010 18th Annual IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems.

6) Cloud Auto-scaling with Deadline and Budget Constraints Ming MaoJie LiMarty Humphrey Department of Computer Science University of Virginia Charlottesville, VA, USA 22904 {ming, j13yh, humphrey}@cs.virginia.edu 978-1-4244-9349-4/10/\$26.00 © 2010 IEEE 11th 41 IEEE/ACM International Conference on Grid Computing

7)2011 31st International Conference on Distributed Computing Systems A Cost-aware Elasticity Provisioning System for the Cloud Uendra Sharma, Prashant Shenoy University of Massachusetts {upendra,shenoy}@cs.umass.edu Sambit Sahu and Anees Shaikh IBM Watson Research Center {sambit,aashaikh}@us.ibm.com

8)Sokol Kosta, Andrius Aucinas, Pan Hui, Richard Mortier, Xinwen Zhang, ThinkAir: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading,978-1-4673-0775-8/12 ,2012 IEEE

9)Ahmed Ali-Eldin, Johan Tordsson and Erik Elmroth ,An Adaptive Hybrid Elasticity Controller for Cloud Infrastructures

Department of Computing Science, Umea University ° Umea, Sweden ° Email: {ahmeda, tordsson, elmroth}@cs.umu.se

10) Xiaoqiao Meng, Canturk Isci, Jeffrey Kephart, Li Zhang, Eric Bouillet, Dimitrios Pendarakis, Efficient Resource Provisioning in Compute Clouds via VM Multiplexing, *ICAC'10*, June 7–11, 2010, Washington, DC, USA. Copyright 2010 ACM 978-1-4503-0074-2/10/06