International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

An Approach for On-Demand Scalability in Platform as a Service

Manodaya B. Gavali[1], N. D. Kale[2]

[1]SavitriBai Phule Pune University, PVPIT College of Engineering, Bavdhan, Pune, Maharashtra, India

[2]Professor, SavitriBai Phule Pune University, PVPIT College of Engineering, Bavdhan, Pune, Maharashtra, India

Abstract: ***Cloud computing is nothing but the delivery of computational power in the form of service and not a product, here shared software, resources and data are provided to computers and other devices as a metered service over a network. Platform as a Service is a one of the cloud based approach that provides companies all the functionalities for develop, deploy, and administer services, without any the burden of configure, install and manage the middle ware, operating system and hardware. The objective of this paper is "To provide a event driven framework to manage computing platform on demand basis". We propose a framework to manage middleware and virtual machine to provide High-availability clusters. In this we utilize Deployment diagrams to capture the idea of topology of a cluster, which is used in scaling the cluster***

Keywords: cloud, IaaS, PaaS, SaaS, High availability cluster, Scalability

1. Introduction

Cloud computing is a model for enabling, presenting a suitable, on-demand network access to a shared pool of configurable computing resources such as network, storage, server in form of compute, applications complying SOA, and services that can be frequently provisioned and released with less management effort or less interaction of service provider.

Applications which are deployed on cloud computing environment have shown lots of inherent advantages. Important one of them is the flexibility and High Availability. For example, in Storage cloud the files and data can be accessed by Staff as per there need even if they are working remotely and outside office. Consider, for example, how quickly an cloud enabled application can scale up/out as and when needed. Cloud computing is often cheaper and require no manual efforts. Rather than heavy investment at initial phases of project, cloud computing aims on sharing the resources, software is provided "on-rent" basis. Cloud model contains of below five necessary characteristic, four deployment models and three service models.

Essential Characteristics of cloud:

1) **On-demand self-service:** A consumer can independently provision computing capabilities such as network, storage and server when needed automatically without any human interaction of service provider.
2) **Broad network access:** The capabilities are provided through the network and accessed through standard a mechanism that is promoted use by heterogeneous thin or thick client platforms such as mobile phones, laptops, workstations and tablets.
3) **Resource pooling:** The computing resources are pooled so that it can serve multiple consumers using a multi-tenant model, in which different virtual and physical resources dynamically are assigned and reassigned to consumer as per his demand.
4) **Rapid elasticity:** Capabilities can be elastically provided and released automatically to scale up rapidly inward and outward corresponding with demand.
5) **Measured service:** Cloud systems automatically optimize and control resource use with advantages of a metering capability, at the same level of appropriate abstraction to the different types of service such as storage, processing, bandwidth and active user accounts.

Four types of cloud deployment:

- Private Cloud
- Public Cloud
- Community cloud
- Hybrid Cloud

Cloud computing services models can be classified as :

1) Software as a Service (SaaS). The capability provided to the consumer is to use the provider applications running on a cloud infrastructure. Various client devices can access the application from various client devices through either a thin client interface from a web browser or through a program interface.
2) Platform as a Service (PaaS). The capability provided to the consumer is to deploy onto the cloud infrastructure created or acquired applications created using programming language, services, libraries and tools supported by the provider.
3) Infrastructure as a Service (IaaS). Infrastructure as a Service (IaaS) abstracts hardware (server, storage, and network infrastructure) into a pool of computing, storage, and connectivity capabilities that are delivered as services for a usage-based cost. It is foundation for PaaS and SaaS provides a flexible, standard, and virtualized operating.

Our proposed implementation deals with PaaS service model and deployment diagrams of URL to capture the HA Cluster

grammar and aims to address all the four essential characteristics of cloud.

High-availability cluster is group of computers working in tandem to minimize chance of complete downtime. In this there is no single point of failure, if one component is failed then its services taken over by other redundant component until crashed server start working. HA cluster is used by critical database, file sharing system and web application. HA cluster contains many nodes performing same role and working in tandem. These node configuration categories into various models such as Active/active, Passive/active, N+1. These HA architectures can be represented by set of UML representation named as - deployment diagrams, it is used for physical deployment of physical entities on the node. It consists of node and their relationship with each other and used to describe the systems static deployment view and hardware component on which software components deployed. Efficient deployment diagram is important because it control performance, scalability, maintainability and portability. We capture the HA architecture with the UML diagrams with topology files, file format of topology file is discussed further.

2. Related Work

Scalability is one of the key benefits of using cloud computing. Scalability is the ability of the system to increase its capability to handle larger loads by adding resources or nodes at runtime. Depending on business needs cluster can be easily up scaled or downscaled. For example, banking applications are heavily used in day time and not used in evening and weekends, in such lax time the servers can be lend to organizations in other time zones, and on holidays they can be lend to travel organizations which are heavily used. When business needs are changed then cloud service providers can increase existing resources without any need of expensive changes existing IT systems. There are two types of scaling: Horizontal scaling and Vertical scaling. Horizontal scaling is used when it is not possible to change one resource type to other. If IaaS layer is unable to scale up a virtual machine at runtime due to host machine resource crunch or inability of hypervisor, PaaS has to choose horizontal scaling viz scale out and scale in such cases. In which nodes (i.e. servers) are added (scale out) or removed(scale in) to the system as per requirement with less processor and RAM. In vertical scaling we add (scale up) or remove (scale down) resources (processor and RAM) in the system. It is easy to implement but more costly than vertical scaling. Modern day applications requires support from middleware like database, web servers, application servers, mail server, file server, reverse proxy server, message queuing etc. To deploy and scale up, down, out the application at runtime, we have to manage runtime deployment of middleware, initializing it and adding the server to the application cluster. Lack of such framework makes the legacy applications to fail to get deployed over cloud environment.

Similar work for Service Oriented Architecture (SOA) for deploying SOA based solutions on PaaS and presented by Bao Rong [2nd REF]. But here we are collaborating the Deployment model and scaling out / in of platform for generic cluster applications.

One more approach is proposed which provide high availability and high scalability for the enterprise resource planning. The WebSphere cloudburst is solution provided by the IBM for resource monitoring, designed to speed creation and deployment of application to cloud and virtual environment. This solution does not take the load on the individual server into account to scaling out/scaling in the cluster. Solution proposed here has the capability so that virtual nodes in the cluster can report back the health and load the system is going thru, hence can be used for automated deployment.

3. Proposed System

In our proposed system provide framework for automated and on-demand scaling with event driven mechanism. Virtual node on the cluster report back regarding its health and load to the monitoring server. When load at the virtual node increases then the agent present at each node report back to the monitoring server about it, which then scale out the component of particular role. Similarly, if load at virtual node decreases then the agent present at each node report back to the monitoring server about it, so that it then scale in the component of particular role. Deployment diagram and scripts are fed to management server for configuration.

The system architecture is shown in fig.1;

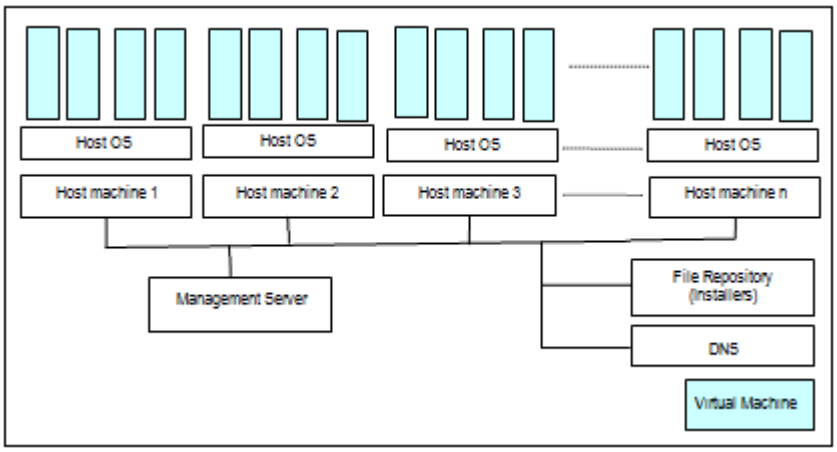

Figure 1: System architecture of proposed system.

4. Conclusion

This paper introduces the platform that provides an effective way to share resources on cloud and automated and on-demand scalability with event driven management. The resulting system could be more durable to load and stress over the resources, without need of manual intervention.

5. Acknowledgment

I like to acknowledge my vigorous thanks to Prof. Mr. N.D.Kale for providing and giving suggestions which helped me a lot in my research work and I also want to thanks our friends and classmates for helping me in this research work by giving me there timely suggestions and feedbacks on my research work.

References

[1] Hailong Sun, Xu Wang, Minzhi Yan, Yu Tang, Xudong Liu “Towards a Scalable PaaS for Service Oriented Software”, 2013 IEEE International Conference on Parallel and Distributed Systems
[2] Bao Rong Chang, Hsiu-Fen Tsai, Ju-Chun Cheng, Yun-Che Tsai High Availability and High Scalability to in-Cloud Enterprise Resource Planning System”, Intelligent Data analysis and its Applications, Volume II Advances in Intelligent Systems and Computing Volume 298, 2014.
[3] Bin Claudio A. Ardagna, Ernesto Damiani, Fulvio Frati, Davide Rebeccani, “Scalability Patterns for Platform-as-a-Service”, 2012 IEEE Fifth International Conference on Cloud Computing.
[4] X. Suo, Y. Zhu, and G. S. Owen, “CLOUD RESOURCE PROVISIONING AND BURSTING APPROACHES”, 2013 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing.
[5] ZENG Shu-Qing “The Improvement of PaaS Platform” 2010 First International Conference on Networking and Distributed Computing
[6] Srijith K. Nair, Sakshi Porwal, Theo Dimitrakos, Ana Juan Ferrer, Johan Tordsson, Tabassum Sharif, Craig Sheridan, Muttukrishnan Rajarajan and Afnan Ullah Khan., “Towards Secure Cloud Bursting, Brokerage and Aggregation”, 2010 Eighth IEEE European Conference on Web Services
[7] Nishant Agnihotri1,Aman Kumar Sharma “EVALUATING PAAS SCALABILITY AND IMPROVING PERFORMANCE USING SCALABILITY IMPROVEMENT SYSTEMS”, IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163
[8] I. Llorente, OpenNebula Project. http://www.opennebula.org/
[9] Amazon Elastic Compute Cloud (EC2),
[10] http://www.amazon.com/ec2/
[11] Google App Engine, http://appengine.google.com
[12] Microsoft Live Mesh, http://www.mesh.com
[13] Sun network.com (Sun Grid), http://www.network.com
[14] The NIST Definition of Cloud Computing [15]http://csrc.nist.gov/publications/nistpubs/800-145/SP800 .pdf
[15] Where are we at with Cloud Computing?: A Descriptive [18]Literature Review – Yang, Tate e, Inc.