Cloud Management with Open Source Tools

Manish S. Thakare*, Akash T. Kuratkar
Dept.of computer sci. & engineering,
jawaharlal darda insti.of engineering & tech.
yavatmal-445001,India
manish.tahkare007@gmail.com*, akashkuratkar@gmail.com

*Abstract -*Cloud computing has gained a lot of hype in the current world of I.T. Cloud computing is said to be the next big thing in the computer world after the internet. Cloud computing is the use of the Internet for the tasks performed on the computer and it is visualized as the next-generation architecture of IT Enterprise. The 'Cloud' represents the internet. Cloud computing is related to several technologies and the convergence of various technologies has emerged to be called cloud computing.Open source systems refer to software systems whose source code is available, allowing for immediate incorporation of improvements and adaptations of the system by its users. This seminar reports on an evaluation of open source management tools for cloud computing. Cloud computing is not a magical solution, planning and management of resources and the cloud in itself is pivotal. In this report, we compare and evaluate a number of open source tools used in the industry to manage and automate the cloud. We evaluate provisioning tools, orchestration tools, configuration tools and monitoring tools.

Keyword: management of cloud with open source tools.

I. INTRODUCTION

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 utility (like the electricity grid) over a network (typically the Internet).

Cloud computing is an emerging computing technology that is rapidly consolidating itself as the next big step in the development and deployment of an increasing number of distributed applications.

Open source refers to any software system whose source code is made available for use or modification by third-party developers. Thus, unlike centralized, proprietary software development models, open source practical accessibility to the source code, allowing for immediate and concurrent incorporation of different approaches, and eventually, the branching of the system into customized variants.

Merging these two trends leads to a new breed of tools: Open source cloud computing tools (OSCCT). These tools provide a free, customizable infrastructure to deploy clouds for any type of application domain. Cloud computing is not a management solution it has to be managed well to yield success. Let's take a look at the various open source tools available to manage the cloud in the various management disciplines of the cloud.

II. WHAT IS CLOUD COMPUTING?

Cloud computing is a marketing term for technologies that provide computation, software, data access, and storage services that do not require end-user knowledge of the physical location and configuration of the system that delivers the services. A parallel to this concept can be drawn with the electricity grid, where in end-users consume power without needing to understand the component devices or infrastructure required to provide the service.

Cloud computing describes a new supplement, consumption, and delivery model for IT services based on

Internet protocols, and it typically involves provisioning of dynamically scalable and often virtualized resources. It is a byproduct and consequence of the ease-of-access to remote computing sites provided by the Internet. This may take the form of web-based tools or applications that users can access and use through a web browser as if the programs were installed locally on their own computers.

Cloud computing providers deliver applications via the internet, which are accessed from web browsers and desktop and mobile apps, while the business software and data are stored on servers at a remote location.

Cloud computing is built upon the broader concept of converged infrastructure and shared services. This type of data center environment allows enterprises to get their applications up and running faster, with easier manageability and less maintenance, and enables IT to more rapidly adjust IT resources (such as servers, storage, and networking) to meet unpredictable business demand. Most cloud computing infrastructures consist of services delivered through shared data-centers and appearing as a single point of access for consumer's computing needs. Commercial offerings may be required to meet service-level agreements (SLAs), but specific terms are less often negotiated by smaller companies.

A. *Layers:* Once an internet protocol connection is established among several computers, it is possible to share services within any one of the following layers.
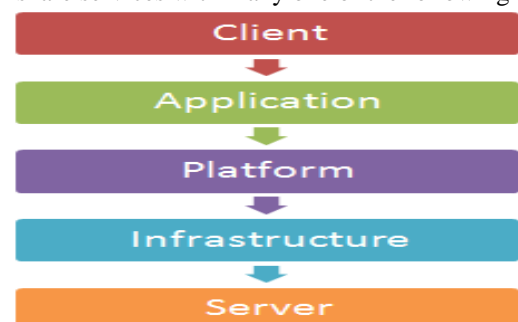

Figure 1: Cloud Computing Layers

B. Client: A cloud client consists of computer hardware and/or computer software that relies on cloud computing for application delivery and that is in essence useless without it. Examples include some computers, phones and other devices, operating systems, and browsers.

C. Application: Cloud application services or "Software as a Service (SaaS)" deliver software as a service over the Internet, eliminating the need to install and run the application on the customer's own computers and simplifying maintenance and support.

D. Platform: Cloud platform services, also known as platform as a service (PaaS), deliver a computing platform and/or solution stack as a service, often consuming cloud infrastructure and sustaining cloud applications. It facilitates deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers.

E. Infrastructure: Cloud infrastructure services, also known as "infrastructure as a service" (IaaS), deliver computer infrastructure typically a platform virtualization environment as a service, along with raw (block) storage and networking. Rather than purchasing servers, software, data-center space or network equipment, clients instead buy those resources as a fully outsourced service.

F. Server: The servers layer consists of computer hardware and/or computer software products that are specifically designed for the delivery of cloud services, including multi-core processors, cloud-specific operating systems and combined offerings.

III. NEED OF TOOLS IN CLOUD COMPUTING

A. Need of Cloud Management Tools:

a. A cloud requires a self-service capability, it must be designed to manage not just provisioning customer requests but also issues such as workload management, security, metering, monitoring, and billing services.

b. Many managers understand that for cloud services to be safe and effective, they must measure and monitor performance. In fact, performance monitoring will become increasingly important as companies rely more on third-party services. And, from all indications, a typical company may use more than one cloud services provider. For example, a company may use one cloud provider for a platform such as collaboration and a completely different provider for compute services. They may use another provider for storage.

Using tools to automate these various tasks are important. Having to perform these tasks manually is a tedious task and takes away from the cloud characteristics of speed, cost-efficiency and convenience.

B. Limitations :-

a. These tools are not open source so not developed as public collaboration and is not freely available

b. Can't freely share, modify and redistribute.

c. Source code is not available

d. Have to depend on particular software developer or vender due to proprietary restrictions thus it increases barrier to entry and exit

e. It has per-seat licensing cost and increases risk of cost growth due to licensing in situation where total number of users may not be known in advance

IV. OPEN SOURCE PHILOSOPHY

A. What is Open Source?:

Open source software is usually developed as a public collaboration and is often made freely available. Open source is indeed a certification mark owned by the open source initiative (OSI). It is a software that is intended to be freely shared, modified, and redistributed by others who may use the open source trademark, provided that the distribution terms conform to the OSI's open source definition.

The main elements in this definition are:

a. The software must be redistributed without restriction.

b. The source code must be made available.

c. The license can require improved versions of the software to carry a different name or version from the original software.

B. Advantages of Open Source Tools:

Here is a list of some of the aspects of open source that may represent a competitive advantage in the software market especially cloud computing.

a. The unrestricted ability to modify software source code enables the department to respond more rapidly to changing situations, missions, and future threats.

b. Reliance on a particular software developer or vendor due to proprietary restrictions may be reduced by the use of OSS, which can be operated and maintained by multiple vendors, thus reducing barriers to entry and exit.

c. Since OSS typically does not have a per-seat licensing cost, it can provide a cost advantage in situations where many copies of the software may be required, and can decreases risk of cost growth due to licensing in situations where the total number of users may not be known in advance.

Table 1: Comparison of Provisioning Tools

	Language	License	Targets
Cobbler	Python	GPL	RedHat, OpenSUSE, Fedora, Debian, Ubuntu
FAI	Perl	GPL	Debian
Kickstart	Python	GPL	Most Debian & Red Hat based Distros
Viper	Perl	GPL	Debian

C. State of Open Source Adoption in Cloud Computing:

Open source adoption in the cloud computing environment though has not reached the expected heights, is still highly relevant. It is observed that the adoption of OSS varies with the specializations and layers of cloud computing.

In the software as a service layer, open source has put up a disappointing performance. It was estimated every year that the next year SAAS will see more open source vendors.

CONFERENCE PAPER
"A National Level Conference on Recent Trends in Information Technology and Technical Symposium" On 09ᵗʰ March 2013
Organized by
Dept. of IT, Jawaharlal Darda Inst. Of Eng. & Tech., Yavatmal (MS), India

89

But, all of the major SAAS vendors such as google use closed source software. The more frustrating aspect to the open source community is that most of these services run on open source platforms such as LAMP.

Things are a little better in the platform as a service layer with a lot more involvement from the open source community and industry with services such as App Scale and RedHat's Open Shift project.

The layer in which you see maximum involvement from the open source industry is the infrastructure as a service layer. The leader in the public IAAS is Amazon's EC2 which is not open source, but it is in close competition with Rackspace's open source project CloudStack.

And in the on-premise or private IAAS, let's see total domination by open source technologies with Eucalyptus leading and other open source vendors such as Redhat, OpenNebula and Rackspace trailing close by.

V. OPEN SOURCE CLOUD MANAGEMENT TOOLS

Now that we have made a case for both cloud computing and open source we will list and evaluate the various disciplines, leading tools from each discipline.

A. Cloud Management Disciplines:

Cloud management is primarily classified into four disciplines:-

    a. Provisioning
    b. Configuration Management
    c. Automation
    d. Monitoring

B. Provisioning Tools:

In terms of networking, provisioning is the process of preparing and equipping a network to allow it to provide (new) services to its users. In the cloud, provisioning tools are mostly used for the installation Let us take a deeper look at some of the

Table 5.3: Comparison of Configuration Tools leading open source provisioning tools in detail:-

	Language	License	Support Organization
AutomateIT	Ruby	GPL	None
Capistrano	Ruby	MIT	None
Control Tier	Java	Apache	DTO Solutions
Func	Python	GPL	Fedora Project
RunDeck	Java	Apache	DTO Solutions

a. **Cobbler** is a linux provisioning server that centralizes and simplifies control of services including DHCP, TFTP, and DNS for the purpose of performing network-based operating system installations. It can be configured for PXE, reinstallations, and virtualized guests using Xen, KVM or VMware.

b. **FAI** is a non-interactive system to install, customize and manage linux systems and software configurations on computers as well as virtual machines from small networks to large-scale infrastructures.

It's a tool for unattended mass deployment of linux. You can take one or more new PC's, turn on the power, and after a few minutes, the systems are installed, and completely

configured to your exact needs, without any interaction necessary.

c. **Kickstart** installation method is used primarily (but not exclusively) by the Red Hat Enterprise Linux operating system to automatically perform unattended operating system installation and configuration. Redhat publishes cobbler as a tool to automate the kickstart configuration process.

d. **Viper** is a provisioning tool released under the GPLv3 license and can be used to install debian and some debian derived distributions. Viper is coded entirely in perl.

C. Configuration Management Tools:

Configuration management (CM) is a field of management that focuses on establishing and maintaining consistency of a system or product's performance and its functional and physical attributes with its requirements, design, and operational information throughout its life.

In the cloud, it is used mostly to set parameters for servers and start/stop various services.

Let us take a deeper look at some of the leading open source configuration management tools used in the cloud.

a. **Bcfg2** – This is a configuration management tool developed in the mathematics and computer science division of Argonne National Laboratory. Bcfg2 aids in the infrastructure management lifecycle conjuration analysis, service deployment, and conjuration auditing. It includes tools for visualizing configuration information, as well as reporting tools that help administrators understand configuration patterns in their environments.

b. **Cfengine** – This is a popular open source configuration management system, written by Mark Burgess. Its primary function is to provide automated configuration and maintenance of large-scale computer systems, including the unified management of servers, desktops, embedded networked devices, mobile smartphones, and tablet computers.

c. **Chef** – This is a configuration management tool

Table 2: Comparison of Automation Tools written in ruby. It uses a pure-ruby, domain-specific language (DSL) for writing system configuration "recipies". Chef was written by opscode and is released as open source under the Apache license 2.0.

	Year Started	Language	License
bcfg2	2003	Python	BSD
Cfengine	1993	C	Apache
Chef	2009	Ruby	Apache
Puppet	2004	Ruby	GPL

d. **Puppet** – This is a tool designed to manage the configuration of UNIX-like and Microsoft Windows systems declaratively. The user describes system resources and their state, either using puppet or ruby DSL (domain-specific language). This information is stored in files called "Puppet manifests". Puppet discovers the system information via a utility called facter, and compiles the Puppet manifests into a system-specific catalog containing resources and resource dependency, which are applied against the

target systems and any actions taken to remediate the system to the desired state will be reported

D. Automation/Orchestration Tools:

Orchestration describes the automated arrangement, coordination, and management of complex computer systems, middleware, and services. It is basically used in the cloud to automate tasks across systems.

Cloud service orchestration is the:-

- a. Composing of architecture, tools and processes by humans to deliver a defined service
- b. Stitching of software and hardware components together to deliver a defined service
- c. Connecting and automating of work flows when applicable to deliver a defined service

Table 5.4: Comparison of Monitoring Tools

	License	Language	Type of Monitoring	Collection Methods
Cacti	GPL	PHP	Performance	SNMP, syslog
Nagios	GPL	C/PHP	Availability	SNMP,TCP,ICMP, IPMI, syslog
Zabbix	GPL	C/PHP	Availability, Performance	SNMP,TCP/ICMPIPMI, Synthetic Trans- actions
Zenoss	GPL	Python	Availability, Performance, Event Management	SNMP, ICMP, SSH, Sys-log, WMI

Let us take a deeper look at some of the leading open source configuration management tools used in the cloud.

- a. **AutomateIT** is an open source tool for automating the setup and maintenance of servers, applications and their dependencies. It provides a surprisingly simple, yet powerful, way to manage files, packages, services, networks, accounts, roles, templates and more.
- b. **Capistrano** is an open source tool for running scripts on multiple servers; its main use is deploying web applications. It automates the process of making a new version of an application available on one or more web-servers, including supporting tasks such as changing databases.
- c. **Control Tier** is an open source, cross-platform build and deployment automation framework. Control Tier can help you to coordinate and scale service management and administration activities across multiple nodes and application tiers.
- d. **Func** is a two-way authenticated system developed under the Fedora project. You can build your own applications on top of it, and easily expand func by adding in additional modules, whether you want these to work through the func command line or by means of some other application.
- e. **Run Deck** is cross-platform open source software that helps you automate ad-hoc and routine procedures in data center or cloud environments. RunDeck allows you to run tasks on any number of nodes from a web-based or command-line interface. RunDeck also includes other features that make it easy to scale up your scripting efforts including: access control, workflow building, scheduling, logging, and integration with external sources for node and option data.

E. Monitoring Tools:

In terms of networking, monitoring describes the use of a system that constantly monitors a computer network for slow or failing components and that notifies the network administrator (via email, pager or other alarms) in case of outages.

Monitoring tools are used in the cloud to record errors and health of the IT infrastructure.

Let us take a deeper look at some of the leading open source configuration management tools used in the cloud.

- a. **Cacti** is an open source, web-based graphing tool designed as a frontend to RRD tool's data storage and graphing functionality. Cacti allow a user to poll services at predetermined intervals and graph the resulting data. It is generally used to graph time series data of metrics such as CPU load and network bandwidth utilization. A common usage is to monitor network traffic by polling a network switch or router interface via SNMP.
- b. **Nagios** is a popular open source computer system and network monitoring software application. It watches hosts and services, alerting users when things go wrong and again when they get better.
- c. **Zabbix** is designed to monitor and track the status of various network services, servers, and other network hardware.
- d. **Zenoss** is an open source application, server and network management platform based on the Zope application server. Released under the GNU General Public License (GPL) version 2, Zenoss Core provides a web interface that allows system administrators to monitor availability, inventory/configuration, performance and events.

VI. OPEN SOURCE CLOUD COMPUTING TOOLS USED IN WEATHER APPLICATION

Consider the development of an application for managing weather information. Such application feeds from a set of heterogeneous, distributed data sources including satellite images, GIS maps, and measurement stations for temperature, wind speed and atmospheric pressure. The application might need to collect historical data and present the user with aggregates from various levels of detail to summarize conditions during days, weeks, or months.

The application provides users with information about current weather conditions at various geographic locations. The application combines a Web component and Hadoop. The web component provides the user-interface and databases to manage common operations and keep system metadata. Hadoop provides the capabilities to run data analysis jobs on demand.

The web view of the application, which is based on the Google Maps API, is shown in Figure 6.1. The user can click on a location in the map to its view World coordinates and a link to see the details of the weather conditions at that location. User can save these locations into a local database and later retrieve them to compare data values.

A. Open Source Tools Used :-

- a. **Eucalyptus** – Provides a solution for IaaS. Eucalyptus was originally built as an open source cloud product and now supports enterprise-class private cloud as well

CONFERENCE PAPER
"A National Level Conference on Recent Trends in Information Technology and Technical Symposium" On 09th March 2013
Organized by
Dept. of IT, Jawaharlal Darda Inst. Of Eng. & Tech., Yavatmal (MS), India

91

as hybrid cloud computing. Application goal is to provide direct access to cloud from the outside world. Although Eucalyptus comes with Ubuntu Linux.
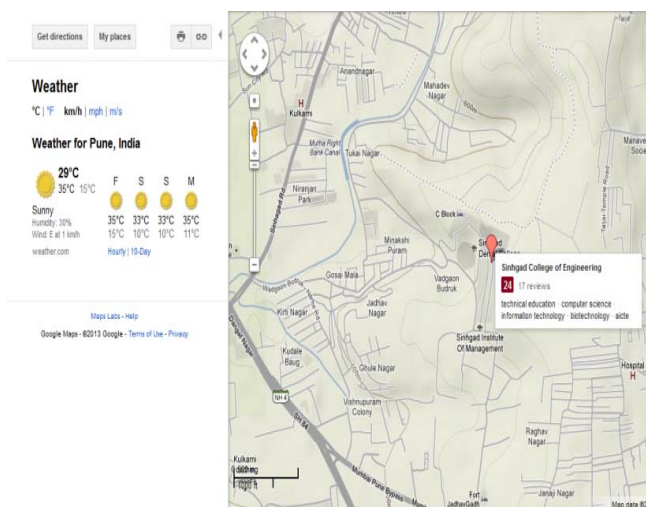

Figure 2: web view of the application based on Google   maps API

b. *Hadoop*– Provides an infrastructure for PaaS, specific for parallel processing. Hadoop provides the capabilities to run data analysis jobs on demand. Hive, a Hadoop-based data warehouse, provides storage for large data sets.

c. *Django-Python*– Provides an infrastructure for SaaS, specific for web application development.

VII.   CONCLUSION

The leading technical research agency Gartner reports that by 2015, 20% of the information technology based organizations would have completely adopted a cloud computing solution, and 98% of enterprises use open source tools to manage the cloud. Cloud computing and open source presents numerous opportunities to each other Open source software for the most part of their existence has been disrupters, be it the field of operating systems, databases or basic desktop and web programming. But, cloud computing provides the open source industry to be an innovator of the new rather than a disrupter of the old.

VIII.   ACKNOWLEDGEMENT

IX.   REFERENCES

[1].   Manuel Rodriguez-Martinez, Jaime   Seguel, ”Open Source Cloud Computing Tools: A Case Study with weather application”, IEEE Paper 2010

[2].   Judith Hurwitz, Robin Bloor, Marcia Kaufman, and Dr. Fern Halper, ”Cloud Computing for Dummies”

[3].   Mark R. Hinkle, ”Crash Course in Open Source Cloud Computing” in Proc. Infosec, 2011

[4].   Mark R. Hinkle, ”Build your Cloud: Simple, Cost Effective, Open Source” in Proc. OSCON 2011

[5].   Marten Mickos, ”Open Source and Cloud Computing: A Natural Fit or Mortal Enemies?” in Proc. OSCON 2010

[6].   Mark R. Hinkle, ”Open Source Tool Chains”, in Proc. OSCON 2010

[7].   Simon Wardley, ”Cloud Computing: Why it Matters?” in Proc. OSCON 2009 [8] Wikipedia.org

CONFERENCE PAPER
“A National Level Conference on Recent Trends in Information Technology and Technical Symposium” On 09th March 2013
Organized by
Dept. of IT, Jawaharlal Darda Inst. Of Eng. & Tech., Yavatmal (MS), India