

14. Comparison of cloud management platforms

Kimmo Ahokas
Aalto University School of Science
kimmo.ahokas@aalto.fi

Abstract

Cloud computing allows fast and efficient resource provisioning within data centers. In large companies this can lead to significant savings, thus creating market for complete cloud platforms. In addition to commercial products, several open source cloud platforms exist. This paper compares four cloud management platforms and identifies the factors affecting future success of each of the platforms. We also estimate the future development of the cloud platform market.

KEYWORDS: cloud platform, IaaS, CloudStack, OpenStack, OpenNebula, Eucalyptus, VMware

1 Introduction

Cloud computing has rapidly changed the way in which resources in data centers can be provisioned. With virtualization and easy to use platforms the resource utilization can be optimized dynamically based on service load. This has led to situation where growing number of companies are setting up their own internal cloud computing platforms.

Building production-ready and scalable cloud platform is a technically challenging task. Resource management, authorization, backups and failure handling are all difficult subjects. That's why companies mostly rely on existing platforms. In addition to commercial cloud platforms, there are number of open source alternatives.

This paper studies various open source IaaS platforms and tries to identify the platform that will gain most success in the future. In the second section we introduce the general background for open source cloud computing platforms, third section compares various differences in the four open source IaaS platforms and in the fourth section we discuss how the differences might affect the future of each of platforms. In the final section we provide our conclusions.

2 Background

In this section we describe what cloud computation is, what cloud management platform is and briefly introduce the main open source cloud management platforms, including OpenStack, OpenNebula, CloudStack and Eucalyptus.

The NIST has defined cloud computing as "–a model for enabling ubiquitous, 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." In the defini-

tion cloud computing is divided into three different service models, namely Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). In this paper we are only interested in IaaS service model, which is defined as "The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls)." [10]

Cloud management platform is a software system that controls the allocation of physical resources on the data center. In the IaaS model users can launch virtual machines using the management console, which causes the platform to reserve needed resources, such as network capacity, storage, memory and CPU from physical machines. Cloud management platforms usually handle scaling, hardware failures and other problems automatically.

2.1 Main open source virtualization platforms

Currently there are four main open source virtualization platforms on the market: Apache CloudStack, OpenStack, OpenNebula and Eucalyptus.

OpenNebula started as a research project in 2005 and the first public release was in 2008. The latest stable version is 4.4.1. [11]

The Eucalyptus started as research project in University of California in 2007. In 2009 the project was commercialized through Eucalyptus open source company. During the same year Ubuntu launched "Ubuntu Enterprise Cloud (powered by Eucalyptus)", but later the company decided to move to OpenStack. The latest stable version of Eucalyptus is 3.4. [6]

The Apache CloudStack project began in 2008 by company called VMOps. The company was later sold to Citrix, but the code was released as open source. In 2012 Citrix relicensed the code under Apache License and was the project was adopted by the Apache Software Foundation. The latest stable version of CloudStack is 4.2.1. [4]

OpenStack was founded by Rackspace and NASA in 2010. Currently the stable version is OpenStack 2013.2 (Havana). [12].

3 Comparison

In this section we compare the four cloud platforms in both technical and non-technical terms. We aim to identify the important differences and how they might affect the future of each of the platforms.

3.1 Community

In open source projects the activity of community behind the project is one of the most important factors determining the future of the project. Without active community the development is likely to slow down and the project might die.

Jiang 2014 compares the community of four different open source cloud platforms. His analysis is based on public forum posts, mailing list conversations and commits in public version control databases. The collected data shows that currently the discussion concerning OpenStack and CloudStack consists of roughly 3000 messages per month while OpenNebula and Eucalyptus communities send barely 500 monthly messages. Furthermore, with OpenStack the number of monthly participants has been growing during the last year. Currently OpenStack project has about 800 monthly participants while CloudStack has roughly 200 monthly participants. Both OpenNebula and Eucalyptus have about 100 monthly participants. [8]

While the plain number of messages per month is significantly higher within OpenStack and CloudStack communities than within OpenNebula and Eucalyptus communities, the average number of messages per user varies more. OpenStack users send roughly 3.75 messages per month, CloudStack users 15 messages/month and both OpenNebula and Eucalyptus users send about 5 messages/month. This hints that OpenStack has really large community, but most members are not active. CloudStack on the other hand seems to have really active community. Both OpenNebula and Eucalyptus rank in between of the another two projects.

Jiang also notes that the community population of each project is steadily growing [8]. However, it seems that the active population stays relatively small, indicating that lots of users leave the community shortly after participating to discussion. For each of the projects, Jiang calculates percentage of active community members during last quarter over total community members. The percentage is 32.4% for OpenStack, 21.3% for CloudStack, 10.5% for OpenNebula, and 4.8% for Eucalyptus. This could indicate that OpenStack users choose to stay with the community, while most of the users of other platforms leave the community after a while. On the other hand it might mean that the other projects are easier to maintain so that users do not need support from community after their installation is up and running.

In a blog post from March 2014 Ignacio M. Llorente points out that public discussions in mailing lists and forums do not necessarily tell the whole truth about community activity [9]. For example some projects might organize parts of their communications in private channels.

Apart from public discussion the activity in source code management systems tells a lot about open source software projects. Each of the four compared virtualization platforms use public git repositories to manage the development. From

project	git commits	contributors	domains
OpenStack	2000	350	100
CloudStack	500	50	20
Eucalyptus	300	25	4
OpenNebula	200	10	3

Table 1: Rough estimate of monthly git activity in each of the projects based on figures by Jiang. The domains where commits come from are based on the email address in each commit.[8]

the commit logs it is possible to analyze current community.

The table 1 summarizes monthly git activity for each of the projects. The "git commits" column lists estimated number of monthly commits, "contributors" lists estimated monthly number of people sending commits and "domains" list estimated number of different email domains where the commits are sent from. Based on the plain numbers the OpenStack project seems to be most popular, CloudStack is in the middle and both Eucalyptus and OpenNebula fall behind.

The number of monthly commits in OpenStack seems to vary a lot, but the trend is still rising. Furthermore, the number of monthly contributors has been rising more steadily. However, during latest few months there has been a significant drop both in number of commits and in number of monthly contributors. This could signal some problems in community. In CloudStack the git activity seems to be increasing while in OpenNebula and Eucalyptus the git activity is mostly constant.

However, the plain number of commits can be fairly misleading figure. OpenStack project consists of many independent subprojects and a change in functionality may require commits to multiple subprojects. Furthermore, the commit size can vary significantly.

The number of contributors and domains where the contributions come from are more interesting figures. Clearly OpenStack code is contributed from wide range of different organizations while in Eucalyptus and OpenNebula the contributors are tied to small number of organizations. One possible explanation for this difference is that both Eucalyptus and OpenNebula could collect commits from community members and then hide the original author information. On the other hand organizations participating to OpenStack and CloudStack might encourage developers to send commits using their private email address rather than company email address.

In total the OpenStack project seems to have the largest and most active community both in terms of number of members and development activity. However, as the development of all the projects proceeds steadily, there must be some differences in the communities. It is possible that CloudStack, OpenNebula and Eucalyptus community members are more committed to the project and individual contributions are more significant than in OpenStack.

3.2 Organizational support

Large organizations using certain platform can hugely benefit the platform. If the core business of the organization depends on the platform, the organization is probably ready to dedicate vast amounts of resources to platform development. In most cases these efforts also contribute to the main project and benefit the whole community. Each of these cloud platforms have large organizations using and developing the platform.

OpenStack lists over 200 companies and organizations as members, sponsors and supporters of the OpenStack Foundation on their website¹. Among the most notable ones are the founders Rackspace and NASA. Other supporters include names such as AT&T, HP, IBM, Canonical, Red Hat and SUSE.

OpenNebula project does not report directly supporters, but instead the project provides a list of featured users². These include IBM, Dell, CentOS, Akamai and NASA.

Apache CloudStack is backed by Apache Software Foundation. The foundation does not list backers for any individual project, but the foundation sponsors³ include companies such as Yahoo, Citrix, Facebook, Microsoft and Google. According to Jiang at least Citrix is active in CloudStack development [8].

Eucalyptus project lists large number of companies as their partners⁴. These include companies such as Amazon Web Services, Dell, HP, Intel, Novell, Red Hat and Ubuntu.

Interestingly both OpenStack and Eucalyptus report Canonical/Ubuntu and Red Hat as project supporters. However, both Ubuntu⁵ and Red Hat⁶ advertise support for OpenStack. Furthermore, companies such as IBM, Dell and HP are associated with multiple cloud platform projects. Of course it is possible that large companies support multiple cloud platforms, but it seems logical that the primary efforts are targeted to specific platform.

For any virtualization platform the partnership with popular Linux distributions is really important as it makes deploying the platform easier for users. According to W3Techs Web Technology survey, the four most popular Linux distributions powering websites are Debian (used by 29.2% of internet sites in the world), Ubuntu (22.8%), CentOS (19.8%) and Red Hat (6.7%) [3]. As mentioned earlier, Ubuntu and Red Hat advertise support for OpenStack. Debian⁷ supports both OpenStack and OpenNebula. CentOS is derived from Red Hat thus supporting at least OpenStack. However, the project is looking for expanding and seems to support⁸ OpenNebula, OpenStack and CloudStack. Any of these distributions does not advertise support for Eucalyptus, but Eucalyptus installation instructions⁹ indicate support for both CentOS and Red Hat.

¹<http://www.openstack.org/foundation/companies/>

²<http://opennebula.org/community/users/>

³<http://www.apache.org/foundation/thanks.html>

⁴<https://www.eucalyptus.com/partners/all>

⁵<http://www.ubuntu.com/cloud>

⁶<http://www.redhat.com/products/cloud-computing/>

⁷<https://wiki.debian.org/Cloud>

⁸<http://wiki.centos.org/Cloud>

⁹https://www.eucalyptus.com/docs/eucalyptus/3.4/index.html#install-guide/installing_euca_release.html

OpenStack seems to have the widest range of Linux distributions supporting the platform. For end users this is the ideal case, as they can easily install the platform on operating system they are already familiar with. Furthermore, the support from distribution means that all the components are installable using distributions standard tools without configuring additional software repositories or building source files manually.

Other important range of supporters for virtualization platforms are cloud service providers such as Amazon Web Services or Rackspace. Company running their core business on open source platform signals that the platform is stable and the quality is high enough for serious use. Furthermore, large providers can push the limits of the platform enabling development towards even larger deployments.

The two largest IaaS service providers in the world, Amazon Web Services (AWS) and Rackspace [14] both advertise their support for open source virtualization platforms. Amazon is one of the key partners of Eucalyptus. Amazon does not directly use Eucalyptus in their platform, but the company supports Eucalyptus development and the APIs are compatible. However, Rackspace uses OpenStack to build their public and private cloud services. For OpenStack this is a huge benefit, as the platform is proven to work with large user space.

3.3 Commercial support

Building own cloud infrastructure for a company is very complicated task, which makes it important to have enterprise support available when needed. Each of these platforms have number of companies selling consulting and support services for the platform, but the service levels and locations vary.

3.4 License

The licensing of software has been significant issue within open source software community. The license determines how the software may be used, modified and distributed, strongly affecting the potential users. These four different cloud platforms are provided under two different licenses.

OpenStack, CloudStack and OpenNebula are provided under the Apache License, version 2.0. The Apache license allows users to use the software for any purpose, distribute it, modify it and distribute the modified version. It is also allowed to distribute modified parts under different license but the unmodified parts must be distributed under original license and any copyright notices must be preserved. [13]

Eucalyptus is distributed under the GNU General Public License Version 3 (GPLv3). Unlike the Apache License, GPLv3 requires that any modified parts of the code must be distributed using the original license. GPL license is often called copyleft license because of the requirement for using same license with derivative works. [7]

The Open Source Initiative approves the Apache License 2.0 and General Public License version 3 as open source licenses, e.g. that both of the licenses comply with the Open source definition [1]. However, the Apache license is more permissive than the GPL license.

The Apache License can be seen better for the projects, is it is more permissive. For example companies might want to distribute their modifications to software, but only with more restrictive terms. The required licensing might be a reason not to distribute modifications to community at all. However, under some circumstances if some parts of the code are licensed under different terms, it might affect the future usability of the software.

3.5 API

All the cloud platforms provide a browser-based graphical user interface for managing resources. In addition the platforms also have Application Programming interfaces (API) that allow developers to programmatically manage the resources. Good API's allow developers to create better tools for managing the cloud resources and compatibility with other vendors, such as Amazon EC2 can be seen as an advantage.

As companies have created internal tools to automate workflow with Amazon EC2, compatible APIs in open source virtualization platforms facilitate easier migration to these platforms. Moreover, compatible APIs reduce vendor lock-in.

One of the main selling points of Eucalyptus is the compatibility with AWS (Amazon Web Services). The platform is especially targeting hybrid cloud setups, where part of the computation is performed in own environment and part of computation is offloaded to public cloud such as AWS. Most parts of the Eucalyptus API are compatible with respective APIs from Amazon. For end users this offers a huge benefit, as Eucalyptus resources can be managed using same tools as with AWS. Moreover, it is possible to manage AWS resources directly from Eucalyptus control panel. [5]

OpenStack mainly relies on it's own API but additionally provides EC2 compatibility API. The compatibility API supports the most important actions, but is far from complete [2]. However, the OpenStack API has quite broad support, so in the future it might have similar status as EC2 API nowadays.

Similarly both CloudStack and OpenNebula mainly rely on their own APIs but provide AWS compatible APIs parallel to the native API.

3.6 Publicity

To gain users the platform should be familiar to potential users. Platform that is not familiar to potential users will have hard time gathering more users.

One of the easiest metrics for the amount of publicity each of the projects has, is the number of results the search for platform name yields in Google. On 19.03.2014 the results were following: CloudStack 5 510 000, OpenNebula 5 400 000, Eucalyptus 3 110 000 (used search was "Eucalyptus cloud" to avoid confusion to tree species) and OpenStack 2 480 000.

4 Discussion

In this chapter we discuss the future of these platforms and the whole cloud platform market.

5 Conclusion

References

- [1] Open source licenses.
- [2] Openstack wiki: Ec2 api comparison matrix.
- [3] Usage statistics and market share of linux for websites.
- [4] Apache Software Foundation. Cloudstack's history.
- [5] Eucalyptus Systems, Inc. Aws and eucalyptus compatibility.
- [6] Eucalyptus Systems, Inc. The eucalyptus story.
- [7] Free Software Foundation, Inc. GNU general public license, version 3, Jun 2007.
- [8] Q. Jiang. Open source IaaS community analysis, OpenStack vs OpenNebula vs Eucalyptus vs CloudStack, jan 2014.
- [9] I. M. Llorente. Balance between user base and community in OpenStack and OpenNebula.
- [10] P. Mell and T. Grance. The nist definition of cloud computing. *National Institute of Standards and Technology*, 53(6):50, 2009.
- [11] OpenNebula.org. About the opennebula project.
- [12] Rackspace. About rackspace.
- [13] The Apache Software Foundation. Apache license, version 2.0, jan 2004.
- [14] D. Toombs, L. Leong, B. Gill, T. Haynes, and G. Petri. Magic quadrant for managed hosting, north america. Technical report, Gartner, aug 2013.