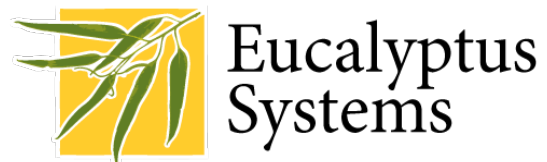


Cloud Computing and Open Source: IT Climatology is Born



Eucalyptus Systems, Inc.

© 2010

Clouds, Public and Private

Cloud computing has emerged as a new platform for implementing Information Technology (IT) functionality in a more flexible and cost effective way than has been possible previously. Originally, the concept of a computing cloud was that of a large-scale, publically accessible utility from which customers could draw resources on an as-needed basis, much as they draw networking capacity from the publically accessible Internet today. However, just as local networking is necessary in the enterprise in addition to public Internet access, there is also a need for an on-premise version of cloud computing in the private data center.

Cloud computing software in the data center layers a self-service utility capability over IT infrastructure (machines, network equipment, storage devices, etc.) in a way that permits end-users to manage resource provisioning and scale automatically, under their own control. IT manages the cloud platform itself and the infrastructure that implements it, but

the “nuts and bolts” associated with the allocation and release of data center resources are automated and “zero touch” by the IT administration, thereby streamlining the provisioning process.

In addition to the efficiency gains that on-premise or “private” clouds offer, another advantage is that they can be managed according to locally defined policies. For example, most data centers have well-defined security and privacy policies that the infrastructure must be able to audit and enforce. Moreover, as the mission of the organization changes, these policies may be updated or modified while the IT organization is responsible for ensuring the integrity of the overall system. In a public cloud, because the infrastructure is operated by a third party, it is often difficult to guarantee policy enforcement, auditing, and integrity particularly with respect to security and privacy. In contrast, because a private cloud uses the infrastructure already maintained and protected by the IT organization, it is

easier to implement and control policy.

Open-source Clouds

To achieve the degree of control that IT requires of its infrastructure, however, the software platform implementing the cloud paradigm needs to be transparent. That is, the private cloud platform cannot obscure any layer of the hardware or software infrastructure from the IT staff while, at the same time, it implements a secure and controlled execution environment for the cloud users. Two decades of experience with operating systems in the data center culminating in the emergence of Linux as the predominant choice indicates that the best way to achieve this transparency is through open source.

Open-source software platforms, such as Linux or Java, are developed publically in a way that exposes all aspects of their architecture and implementation as source code. Because the source code is freely available, “open source” is often considered synonymous with “free” by the public at large, and hence it has a great deal of appeal as a cost-saving measure. In fact, most IT organizations using open source software components today do not use them free of charge even when they could. Moreover, free versions of closed source, proprietary software is almost never used. Thus it is not because open source is free that it is widely used in a production IT

settings. Rather, it is the free and open development model that provides the benefits which ultimately lead to cost savings.

One such benefit is transparency. Because the community at large can review, comment on, and modify all aspects of the software platform, its behaviors (good and bad) are not mysterious to the IT personnel who are responsible for managing them. Often, the community maintains a set of publically available discussion venues where problems and their resolutions are shared. Thus when a problem occurs, IT management has immediate access to a large collection of resolution strategies without consulting with a (possibly unresponsive) third party vendor.

Somewhat counter-intuitively, this transparency is particularly important to the implementation and enforcement of local security policies. Originally, open source platforms were thought to be less secure than their proprietary counterparts because of the access that the public has to their internals. As open source has matured as a software development methodology, however, it has become clear that access to the source code itself does not create undue security liabilities while, at the same time, it does facilitate rapid remediation strategies when security problems do arise. In contrast, proprietary platforms today are some of the most vulnerable to attack and security breach. Summarizing,

obscured proprietary system code does not offer additional security while open source offers IT personnel both a way to see the liabilities “up front” and to counter their effects rapidly.

These lessons have manifested themselves with the ascendancy of Linux and Java as production computing software platforms in the data center. Today, cloud computing is a new but rapidly maturing paradigm for IT management and, as such, it will benefit from the experiences and best practices engendered by its predecessors. To do so, however, the cloud computing platform in the data center must be, and ultimately will be, an open-source platform.

Eucalyptus – the Open-source Cloud Computing Platform

Cognizant of modern software adoption trends, Eucalyptus has been developed as an open-source platform for implementing on-premise and private clouds. Its origins are as a research project in the Computer Science Department at the University of California, Santa Barbara where its authors were studying the use of open source to create new, highly scalable, and high performance distributed computing environments. Having gained wide acceptance in the research and open-source communities, it is now maintained by Eucalyptus Systems Inc. – a company

formed to provide enterprise-focused products as well as the professional development and support resources that are necessary to maintain and grow its user community.

Eucalyptus is both open source and packaged in a way to make it available via a number of different Linux distributions. This latter feature makes it easy and familiar to install in data centers where some version of Linux is already in use. In addition to being “agnostic” with respect to Linux version, Eucalyptus is also able to work with a number of open-source and proprietary hypervisors allowing virtualized data centers to continue to use the virtualization technology they already have in place, if desired.

Perhaps the most striking feature of Eucalyptus, however, is its choice of the Amazon AWS APIs as the API it currently supports. The progenitors of Eucalyptus experimented with a number of different cloud APIs (including several of their own definition) before it became clear that AWS had emerged as the de facto standard. The result is that it is possible to run an application on AWS and on Eucalyptus without modification. This support for “The Standard” as it is practiced today ensures that applications developed for Eucalyptus have access to the same ecosystem of technologies that those developed for AWS do. The AWS API is also well-tested and demonstrably scalable mitigating the risk that an API feature will “cripple”

the scalability of some future version of Eucalyptus.

The features of Eucalyptus that are making it the predominant choice for on-premise and private clouds can be summarized as:

- **Data center optimization.** Eucalyptus optimizes existing data center resources with consolidation through virtualization of all data center elements, including machines, storage and network. Eucalyptus is compatible with most widely used virtualization technologies, including Xen and KVM hypervisors.
- **Automated self-service.** Eucalyptus automates computer resource provisioning by allowing users to access their own flexible configurations of machines, storage, and networking devices as needed through standardized Web services.
- **Web services based.** Eucalyptus uses universally accepted Web service protocols internally, making its installation, operation, and maintenance similar to that for a high-quality e-commerce site.
- **Scalable data center infrastructure.** Eucalyptus clouds are highly scalable, which enables an organization to efficiently scale-up or scale-down data center resources according to the needs of the enterprise.
- **Elastic resource provisioning.** The elasticity of a Eucalyptus cloud allows users to flexibly reconfigure computing resources as requirements change. This flexibility helps the enterprise workforce remain adaptable to sudden changes in business needs.
- **Open source innovation.** Highly transparent and extensible, Eucalyptus' open source core architecture supports value-adding customizations and innovations provided by the open source development community. The Eucalyptus open source software core is available for free download at www.eucalyptus.com.
- **Hybrid cloud capability.** Engineered to emulate Amazon Web Services (AWS), Eucalyptus interacts seamlessly with Amazon public cloud services, including EC2 and S3, with no software modification required. This compatibility allows IT organizations to quickly "cloudburst" into the public cloud space without purchasing additional data

center hardware during very large spikes in data center resource demand.

As an open-source IT platform, one of Eucalyptus' key features is its ability to conform to local security and access policies. Eucalyptus can be configured to run within the bounds of electronic or physical security perimeters. It is fully self-contained and requires no access to resources beyond the resources on which it is installed to operate.

Also because it is agnostic with respect to Linux distribution, it implements cloud functionality without modification to local infrastructure. That is, it is able to "overlay" the existing hardware and software installation without requiring new or dedicated hardware, specific Linux versions, kernel modules, etc. At the same time, the system is highly tunable and configurable so that as the infrastructure evolves Eucalyptus can be customized to deliver the performance and stability characteristics of the resources under its control.

Its recent emergence as a popular cloud platform is analogous to that Linux enjoyed when it first gained acceptance in the data center. Eucalyptus implements as open source the same API that a large-scale, proprietary system does (AWS). In the case of Linux, the API was the Unix system call interface, with which

many data center operators were already familiar. However, the Eucalyptus-AWS relationship is different from that for Linux-Unix in that Eucalyptus on-premise complements AWS as a public cloud and vice versa. That is, Eucalyptus is not designed as a replacement technology for AWS but rather a way to get AWS functionality on-premise or in a private cloud setting.

As a final similarity, just as Linux supports a variety of applications (open source and proprietary), so does Eucalyptus support a growing ecosystem of cloud tools and applications (open source and proprietary). In particular, while the majority of its ecosystem is being generated by the external community, Eucalyptus Systems acts as the focal point for ensuring that this ecosystem is adequately supported. In the same vein, Eucalyptus Systems also offers "enhancements" to the base platform to support specific functionalities that are not properly part of the open-source platform. The participation by the commercial entity behind Eucalyptus in its ecosystem ensures that high-quality, enterprise grade software components specifically designed to work with Eucalyptus are available on a time scale that does not depend entirely on the Eucalyptus open-source community.

The Birth of IT Climatology

The self-service, SLA-based access to virtualized infrastructure that cloud



computing offers will continue to drive IT professionals to clouds as a platform. While public clouds offer tremendous scale at an attractive on-demand price point, on-premise and/or private clouds provide the infrastructure control and policy enforcement capabilities necessary in the mission-oriented data center. Experience with previous data center platforms (e.g. Linux and Java) shows that future platforms will be open

source, particularly because of the transparency and speed of problem remediation that open source platforms engender. Eucalyptus is the leading and most mature open-source cloud platform today, supporting the de facto standard APIs. With it, and the ecosystem that is developing around it, the future of on-premise cloud computing IT is becoming clearer and easier to forecast.

**Eucalyptus Systems, Inc.
130 Castilian Drive, Goleta, CA 93117 USA
1 (866) 456-3822 (EUCA)
www.eucalyptus.com**

**Copyright © 2010
Eucalyptus Systems, Inc. All rights reserved.
Eucalyptus is a registered trademark of Eucalyptus Systems, Inc.**