Cloud computing and storage with OpenStack

Discover the benefits of using the open source OpenStack IaaS cloud platform

M. Tim Jones

August 06, 2012

The Infrastructure as a Service (IaaS) cloud platform space is quite diverse, with well-known solutions like Nebula and Eucalyptus. But a relative newcomer to this space has shown considerable growth, not only in users but a large number of supporting companies. Get to know the open source platform OpenStack, and discover whether it's really the open source cloud operating system.

The idea behind clouds and public Infrastructure as a Service (IaaS) is certainly not new. In fact, Amazon Elastic Compute Cloud (Amazon EC2) will be six years old this year. What has changed is the focus on IaaS as a means of private cloud computing to satisfy enterprise computing with sensitive data. Private cloud computing applies the IaaS idea to private infrastructure. Although doing so lacks the economic advantages of public clouds (pay-as-you-go services), it exploits the core principles of cloud computing, with a scalable and elastic infrastructure within a corporate data center.

Let's begin with a quick introduction to IaaS and its architectures, and then jump into the leading open source solution: OpenStack.

IaaS and cloud architectures

Cloud computing architectures tend to focus on a common set of resources that are virtualized and exposed to a user on an on-demand basis. These resources include compute resources of varying capability, persistent storage resources, and configurable networking resources to tie them together in addition to conditionally exposing these resources to the Internet.

The architecture of an IaaS implementation (see Figure 1) follows this model, with the addition of other elements such as metering (to account for usage for billing purposes). The physical infrastructure is abstracted from the application and user through a virtualization layer implemented by a variety of technologies, including hypervisors (for platform virtualization), virtual networks, and storage.
Figure 1. High-level view of IaaS

Did NASA drop OpenStack?

In May 2012, NASA withdrew its participation of OpenStack, the solution it helped create with Rackspace. NASA didn't drop OpenStack but instead has changed its role to user rather than developer. OpenStack has a growing number of developers and supporters (from companies like AT&T, IBM, Dell, Cisco, Yahoo!, and Red Hat), so instead of competing with commercial developers, NASA will be a consumer of OpenStack. NASA also announced that it will cease development of Nebula.

Although OpenStack is the most popular open source cloud solution available today, it certainly wasn't the first. In fact, OpenStack is a combination of two older solutions developed in both the public and private sectors.

An earlier open source IaaS solution, Elastic Utility Computing Architecture for Linking Your Programs To Useful Systems (Eucalyptus), was developed as a research project at the University of California, Santa Barbara. Other solutions include OpenNebula (an open source cloud computing toolkit) and Nimbus (another open source toolkit for IaaS clouds). OpenStack integrated pieces of the U.S. National Aeronautics and Space Administration's (NASA) Nebula platform and the Rackspace Cloud Files project (cloud storage).

Cloud computing's newcomer: OpenStack

OpenStack is a relative newcomer to the IaaS space, its first release having come in late 2010. Despite the solution's presumed lack of maturity and given that it has been around for less than two years, OpenStack is now one of the most widely used cloud stacks. Rather than being a single solution, however, OpenStack is a growing suite of open source solutions (including core and newly incubated projects) that together form a powerful and mature IaaS stack.

As shown in Figure 2, OpenStack is built from a core of technologies (more than what is shown here, but these represent the key aspects). On the left side is the Horizon dashboard, which exposes a user interface for managing OpenStack services for both users and administrators. Nova provides a scalable compute platform, supporting the provisioning and management of large numbers of servers and virtual machines (VMs; in a hypervisor-agnostic manner). Swift implements a massively scalable object storage system with internal redundancy. At the bottom are Quantum and Melange, which implement network connectivity as a service. Finally, the Glance project implements a repository for virtual disk images (image as a service).
As shown in Figure 2, OpenStack is a collection of projects that as a whole provide a complete IaaS solution. Table 1 illustrates these projects with their contributing aspects.

**Table 1. OpenStack projects and components**

<table>
<thead>
<tr>
<th>Project</th>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizon</td>
<td>Dashboard</td>
<td>User and admin dashboard</td>
</tr>
<tr>
<td>Nova</td>
<td>Compute/block device</td>
<td>Virtual servers and volumes</td>
</tr>
<tr>
<td>Glance</td>
<td>Image service</td>
<td>VM disk images</td>
</tr>
<tr>
<td>Swift</td>
<td>Storage as a Service</td>
<td>Object storage</td>
</tr>
<tr>
<td>Quantum/Melange</td>
<td>Networks</td>
<td>Secure virtual networks</td>
</tr>
</tbody>
</table>

Other important aspects include Keystone, which implements an identity service that is crucial for enterprise private clouds (to manage access to compute servers, images in Glance, and the Swift object store).

**OpenStack architecture**

OpenStack is represented by three core open source projects (as shown in Figure 2): Nova (compute), Swift (object storage), and Glance (VM repository). Nova, or OpenStack Compute, provides the management of VM instances across a network of servers. Its application programming interfaces (APIs) provide compute orchestration for an approach that attempts to be agnostic not only of physical hardware but also of hypervisors. Note that Nova provides not only an OpenStack API for management but an Amazon EC2-compatible API for those comfortable with that interface. Nova supports proprietary hypervisors for organizations that use them, but more importantly, it supports hypervisors like Xen and Kernel Virtual Machine (KVM) as well as operating system virtualization such as Linux® Containers. For development purposes, you can also use emulation solutions like QEMU.

Swift, or OpenStack Object Storage, is a project that provides scalable and redundant storage clusters using standard servers with commodity hard disks. Swift does not represent a file system but instead implements a more traditional object storage system for long-term storage of primarily static data (one key usage model is static VM images). Swift has no centralized controller, which improves and overall scalability. It internally manages replication (without redundant array of independent disks) across the cluster to improve reliability.
Glance, or OpenStack Image Service, provides a repository for virtual disk images that Nova can use (with the option of being stored within Swift). Glance provides an API for the registration of disk images in addition to their discovery and delivery through a simple Representational State Transfer (REST) interface. Glance is largely agnostic of the virtual disk image format, supporting a large variety of standards, including VDI (VirtualBox), VHD (Microsoft® Hyper-V®), QCOW2 (QEMU/KVM), VMDK/OVF (VMware), and raw. Glance also provides disk image checksums for integrity, version control (and other metadata), as well as virtual disk verification and audit/debug logs.

The core OpenStack projects (Nova, Swift, and Glance) were developed in Python and are all available under the Apache License.

**Demonstrating OpenStack**

With a large number of independent projects that must be installed and configured to work in concert with one another, installing OpenStack can be a time-consuming task (see Related topics for more information on complete installations). But there are other options that can greatly simplify getting OpenStack up and running for the curious reader.

Anyone who's read some of my prior articles knows that I'm a fan of VM images for simplified use of Linux-based software. VMs allow you to easily create a new instance to try out or demonstrate software. The VM is a self-contained Linux instance (sometimes called a virtual appliance) that you can pre-install with the necessary software and preconfigure for your use. Provisioning software in this way greatly simplifies its use, allowing you to experiment with software that would otherwise be difficult or time-consuming to acquire. Check out Related topics for installation options that fit your particular hardware and base operating system needs.

For this demonstration, I decided to go with the latest Ubuntu release (12.04) and OpenStack's Essex release. Essex is available as an ISO using the uksysadmin's installation procedure (see Related topics). After a clean installation of OpenStack Essex on Ubuntu Precise, an external web browser should be able to view the OpenStack dashboard. Figure 3 shows the System Panel Images tab with the guest VM image in two container formats.

**Figure 3. OpenStack Dashboard view of the available guest images**
The image is used to create a demo instance, which, as Figure 4 shows, has been started. This instance is now available for use.

**Figure 4. OpenStack Dashboard view of the compute instances**

With a compute image now running in OpenStack, I can access it using its IP address (172.16.1.1) through a simple Secure Shell (SSH) session (see Listing 1, user input shown in bold).

**Listing 1. Accessing the OpenStack compute instance via SSH**

```
$ ssh -i Downloads/demo.pem ubuntu@172.16.1.1
The authenticity of host '172.16.1.1 (172.16.1.1)' can't be established.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added '172.16.1.1' (RSA) to the list of known hosts.
Welcome to Ubuntu 12.04 LTS (GNU/Linux 3.2.0-23-virtual x86_64)
* Documentation:  https://help.ubuntu.com/
System information disabled due to load higher than 1.0
0 packages can be updated.
0 updates are security updates.
Get cloud support with Ubuntu Advantage Cloud Guest
   http://www.ubuntu.com/business/services/cloud
The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.
ubuntu@demo1:~$
ubuntu@demo1:~$ hostname
demo1
ubuntu@demo1:~$ ps
   PID TTY          TIME CMD
835 pts/0    00:00:06 bash
948 pts/0    00:00:06 ps
```
With all of these layers running, it can be difficult to visualize what's happening. Figure 5 illustrates the entire stack and hopefully helps demystify it. In this demonstration, a Mac running Mac OS X provides the base platform. VirtualBox runs on Mac OS X, providing the platform for the execution of OpenStack (running on Ubuntu Linux). Note that VirtualBox is a type-2 hypervisor. Within the OpenStack Linux layer, QEMU is used as the guest hypervisor, which is ideal from a commodity hardware perspective but lacks the performance needed in true production settings.

**Figure 5. OpenStack demonstration stack running on commodity hardware**

Without support for nested virtualization (efficiently running a hypervisor on top of another hypervisor), I rely on QEMU for my guest hypervisor running in OpenStack. This allows me to run a guest VM on a guest hypervisor, running on a type-2 hypervisor. Although this setup can be slow, it fully demonstrates an IaaS stack running on a commodity computer system. Note that certain AMD processors provide an efficient way to support nested virtualization today.

Although using QEMU is not ideal from a performance perspective, it is largely compatible with KVM (Linux as a hypervisor), and therefore it is simple to migrate between the two hypervisors (in addition to the VM images being compatible between the two). What makes QEMU ideal in this case is that it can be executed on hardware that provides no virtualization support. Note that my platform in this example is virtualization capable, but because I'm running on VirtualBox (a hypervisor in its own right), the lack of nested virtualization forces me to use a guest hypervisor that has no reliance on virtualization extensions. In either case, I use libvirt to manage the VMs (starting, stopping, monitoring, and so on), so migrating to KVM on virtualization-capable hardware is as simple as a two-line modification in an OpenStack configuration file.

**Other ways to use OpenStack**

If you lack a cluster of your own, there are other options for enjoying the benefits of OpenStack. Rackspace, one of the creators of OpenStack, is offering what it hopes will be the Linux of the cloud. Rackspace's OpenStack cloud platform provides the benefits of OpenStack with the flexibility and scalability of public cloud infrastructure.

To simplify OpenStack installation for private clouds, numerous companies have focused on making it easy to use OpenStack within your private cluster. Companies like Piston Cloud Computing offer the Piston Enterprise OS, a private cloud operating system based on OpenStack. Mirantis provides professional services to enterprises to build out an OpenStack infrastructure.
What's next for OpenStack?

OpenStack continues to integrate new functionality, raising the bar on the definition of an IaaS solution. Numerous other projects under the OpenStack umbrella are available, still others are in the incubation process. The Keystone project provides an identity service that unifies authentication across OpenStack components while integrating with existing authentication systems. Community projects also exist for load balancing as a service (Atlas-LB); a cloud installation and maintenance system (Crowbar); a cloud-provisionable and scalable relational database (RedDwarf); a REST-based API for cloud orchestration (Heat); and a cloud management tool covering monitoring, billing, and more (Clanavi). Numerous other projects are in development within and outside of the OpenStack project, and this list grows each day as OpenStack builds on its momentum.

OpenStack is not without competition, as older projects continue to evolve and new projects appear. For example, CloudStack (first released in 2009) has several production installations but lacks the level of open source contributor support that can be found with OpenStack.

Going farther

Similar to the way Linux has evolved into the all-purpose operating system that fits all usage models, OpenStack is driving toward representing the operating system for the cloud. Instead of managing a limited set of cores and local resources, OpenStack manages a massive network of servers containing compute and storage resources along with the virtual network glue that ties them all together.

Since its first release in late 2010 (Austin), the OpenStack project has released four more versions, the last in April 2012 (Essex). With each release, OpenStack continues to drive new and improved functionality, raising the bar on other IaaS solutions. Now under the Apache umbrella, it's no surprise that OpenStack is the standard in cloud stacks.
Related topics

- **The OpenStack official website** is the unique source for information on the OpenStack family of projects, news on community projects, documentation, and everything else OpenStack.
- **Cloud computing with Linux** (M. Tim Jones, developerWorks, February 2009) is an introduction to cloud computing and its various themes (IaaS, Platform as a Service, Software as a Service), with an angle toward Linux-based options.
- **Anatomy of an open source cloud** (M. Tim Jones, developerWorks, March 2010) introduces cloud computing anatomy from the perspective of open source. This article introduces node architecture, cluster architecture, and the open source technologies that implement these requirements.
- **Anatomy of a cloud storage infrastructure** (M. Tim Jones, developerWorks, November 2010) explores the internals of a cloud storage infrastructure, including general architecture, manageability, performance, scalability, and availability. The article also explores cloud storage models, from private to public and hybrid.
- OpenStack isn’t a single project but an umbrella over a variety of projects that collectively implement a scalable and reliable cloud. Core projects in OpenStack include **Nova**, **Swift**, and **Glance**. Two projects currently in the incubator (soon to be core projects) include **Keystone** and **Horizon**. Finally, there are several community projects that extend or add functionality to OpenStack, including **Quantum**, **Melange**, **Atlas-LB**, **Crowbar**, **Heat**, and **Clanavi**.
- **Installing OpenStack Walk-through** provides a complete introduction to installing OpenStack for production uses.
- Several options exist for using OpenStack in the context of a VM (over the various OpenStack releases, including the bleeding edge). Have a look at **DevStack** (from Rackspace Cloud Builders), OpenStack's **Running OpenStack Compute (Nova) in a Virtual Environment**, and the System Administration and Architecture Blog’s **Screencast Video of an Install of OpenStack Essex on Ubuntu 12.04 under VirtualBox** (I used this example for the demonstration).
- If you need professional help with an OpenStack private cloud, several companies can provide this support. Two such companies include **Piston Cloud Computing** and **Mirantis**.
- **CloudStack** is a competitive stack to OpenStack. It has several of production installations.
- In the developerWorks **cloud developer resources**, discover and share knowledge and experience of application and services developers building their projects for cloud deployment.
- Follow **developerWorks on Twitter**. You can also follow this author on Twitter at **M. Tim Jones**.
- **Evaluate IBM products** in the way that suits you best: Download a product trial, try a product online, or use a product in a cloud environment.