

Understanding the Open Cloud

Open Cloud Projects Profiled

A Linux Foundation Publication

October 2013

The open cloud, like Linux, is fueling dramatic enterprise innovation and growth, which in turn is spurring a worldwide transformation of the technology landscape.

Introduction to the Open Cloud

This history of the technology industry has largely been defined by individual companies building software behind closed doors. Today, thanks in part to the path that Linux has blazed, software is increasingly built collaboratively. This is resulting in better software, more innovation and faster realization of new technologies. This is nowhere more evident today than in the cloud. This open approach to building software is advancing cloud computing at an accelerated pace and customers have the opportunity to invest in a strategic platform that is open from the ground up.

The cloud is emerging as one of the most important new technology platforms in decades, and it represents a unique inflection point, one that offers unprecedented choice for both developers and users. Developers are more engaged to build what they want, how they want, all by using the best software from a variety of projects. Users are getting involved in the development of these technologies unlike ever before, resulting in a change in culture that is referred to as DevOps. They're taking an active interest in how technology will be applied as they build it and are collaborating with other users of that technology. Consider Netflix, Amazon and Salesforce as just a few examples.

The open cloud, like Linux, is fueling dramatic enterprise innovation and growth, which in turn is spurring a worldwide transformation of the technology landscape.

The flurry of innovation is resulting in a variety of different open cloud technologies across a very tall stack of software, which can easily lead to confusion. This paper profiles some of the open source cloud projects at each level of the stack to provide a snapshot of their status and to help customers understand how they might leverage the projects to build, provision and maintain their own open cloud.

This paper is not intended to be a comprehensive survey of all potentially related projects; rather, it is a starting point from which users can begin to assess the building blocks for an open cloud. We expect to update the paper each year and welcome the community's input.

What is the Open Cloud?

The open cloud is one in which every component, from the software to the APIs used by application developers, is open to vendors, developers and customers alike. While proprietary cloud vendors promote their open APIs, an open cloud cannot be constructed from closed components. APIs that are open today, after all, may be closed tomorrow, unless the code is available.

Components of the open cloud today share a couple of key characteristics. Most obviously, they are released under an open-source license; this can be a reciprocal-style license such as the GPL used by the Linux kernel or an alternative like the Apache or MIT licenses. While the licenses carry with them different obligations, they all guarantee full and unrestricted access to the project's codebase.

While the code is critical, the community around the project is of equal long-term importance. To ensure the ongoing viability of a given project, it must have an ecosystem in terms of contributors and users that work overtime to support it. It is this community that is most likely to enable the initial experimentation and proof-of-concept build-outs that represent the first steps towards an open cloud.

Building upon an open cloud provides all the same benefits as the rest of the universe of open-source software, such as:

- improved code quality of the underlying cloud implementation;
- insurance against lock-in as well as the failure of your cloud provider;
- increased security, with the ability to fix vulnerabilities yourself; and
- freedom from mandatory licensing costs; among others.

The full stack comprising the open cloud starts at the operating system (OS) level with a kernel on bare metal, and it rises all the way through what cloud-based developers see and access when writing applications. In the case of Infrastructure as a Service (IaaS), that may mean the top layer of the open cloud is software providing APIs to administer virtualization software for setup, take-down and control of virtual machines. In the case of Platform as a Service (PaaS), the open cloud necessarily goes even farther up the stack, through the guest OS to the software that provides developer-accessible APIs.

Why is the Open Cloud Important?

The next-generation of the enterprise is being built now with open cloud technologies. It's important to know what are these open cloud projects, where they came from and what is the status of their communities. It's also important to get involved and start participating in these projects. Your contributions will accelerate technology innovation that benefits everyone, including you.

Profile Methodology

Building a cloud, whether on public or privately owned infrastructure, is a complicated endeavor involving many moving pieces, both in terms of personnel and of technology. The approach of this paper is to provide a general background on some of the fundamental building blocks of an open cloud, rather than serving as a comprehensive survey of all potentially related projects. Projects profiles are listed into five general categories: Hypervisors & Containers; IaaS; PaaS; Provisioning & Management; and Storage. We by no means mean to slight anyone who was not included and recognize there are many good projects that we simply couldn't or didn't include. We hope this can be a starting point for a discussion at our annual CloudOpen events and for future updates to this paper.

The projects below were selected according to three basic characteristics:

1. Their relevance to the open cloud
2. Their relative maturity
3. Their relative visibility

Specifically, this paper prioritizes projects that are building blocks of the open cloud, have demonstrated their relevance over an adequate time period, and are not already generally accepted mainstays of enterprise infrastructure.

Linux, as one example, in spite of its status as the de facto operating system of the cloud, is not included in the profiles below because its role and significance is already widely known. Likewise, Git, best known as the version control system behind GitHub and also used to manage the Linux kernel, is increasingly being used as the deployment mechanism for cloud platforms. Given Git's overall visibility, however, it was not deemed necessary to profile the project in detail here.

The following list is intended to serve as a primer on open cloud related technologies, and the omission of a given project should not be construed as a judgment on its quality, significance or general importance. As projects continue to evolve, so will this list – for example, as software-defined networking gains maturity, projects such as OpenDaylight or Open vSwitch may be added.

All profile data with the exception of lines of code was collected from public sources, including project websites and source code repositories. Estimates for lines of code are courtesy of Ohloh, and licensed under the Creative Commons Attribution 3.0 Unported license.

Open Cloud Profiles: HYPERVISOR & CONTAINER

KVM

- **Description:** KVM is a lightweight hypervisor that was accepted into the Linux kernel in February 2007.
- **History:** KVM was originally developed by Qumranet, a startup that was acquired by Red Hat in 2008.
- **Website:** <http://www.linux-kvm.org/>
- **Key Contributors:** HP, IBM, Intel, NetApp, Red Hat
- **Commercial Support:** HP, IBM, Red Hat
- **Project License:** GPL
- **Primary Programming Language:** C (95%)

XEN PROJECT

- **Description:** Xen is a cross-platform software hypervisor that runs on platforms such as BSD, Linux and Solaris.
- **History:** Xen was originally written at the University of Cambridge by a team led by Ian Pratt.
- **Website:** <http://www.xenproject.org>
- **Key Contributors:** AMD, Bromium, Calxeda, CA Technologies, Cisco, Citrix, Intel, Oracle, Samsung
- **Commercial Support:** Citrix, Oracle
- **Project License:** GPL
- **Primary Programming Language:** C (78%)
- **Lines of Code:** 507,000
- **Key Users:** Amazon, Rackspace, Verizon

Open Cloud Profiles: IAAS

APACHE CLOUDSTACK

- **Description:** CloudStack is an open source IaaS platform with Amazon Web Services (AWS) compatibility.
- **History:** CloudStack was originally created by Cloud.com (formerly known as VMops), a startup that was purchased by Citrix in 2011. In April of 2012, CloudStack was donated by Citrix to the Apache Software Foundation.
- **Website:** <http://cloudstack.apache.org/>
- **Key Contributors:** Citrix Basho, Citrix, Cloudera, Spotify, SunGard, WebMD
- **Commercial Support:** Citrix
- **Project License:** Apache 2
- **Primary Programming Language:** Java (67%)
- **Lines of Code:** 2.7M
- **Key Users:** Alcatel Lucent, BT, Datapipe, Edmunds.com, Gilt, IBM (Softlayer), Grid5000, NTT, Spotify, TATA Communications, Terramark, TomTom, Verizon, WebMD, Zynga

EUCALYPTUS

- **Description:** Eucalyptus is an open-source IaaS platform for building AWS-compatible private and hybrid clouds.
- **History:** Eucalyptus began as a research project at UC Santa Barbara and was commercialized in January 2009 under the name Eucalyptus Systems.
- **Website:** <http://www.eucalyptus.com/>
- **Key Contributors:** Eucalyptus Systems
- **Commercial Support:** Eucalyptus Systems
- **Project License:** GPLv3
- **Primary Programming Language:** Java (49%)
- **Lines of Code:** 1.4M
- **Key Users:** App Dynamics, FDA, NIST, Nokia Siemens Network, Puma, Riot Games

OPENNEBULA

- **Description:** OpenNebula is an open-source IaaS platform for building and managing virtualized enterprise data centers and private clouds.
- **History:** OpenNebula began as a research project in 2005 authored by Ignacio M. Llorente and Rubén S. Montero. Publicly released in 2008, development today is via the open source model.
- **Website:** <http://www.opennebula.org/>
- **Key Contributors:** C12G Labs
- **Commercial Support:** C12G Labs, inovex, Logica, Netways, Terradue
- **Project License:** Apache 2
- **Primary Programming Language:** C++ (32%)
- **Lines of Code:** 190,000
- **Key Users:** Akamai, BBC, Blackberry, CentOS, China Mobile, Deutsche Post E-Pos, European Space Agency, FermiLab, Prohuban - Santander Group, SARA Supercomputing Center

Open Cloud Profiles: IaaS (Cont.)

OPENSTACK

- **Description:** OpenStack is an open source IaaS platform, covering compute, storage and networking.
- **History:** In July of 2010, NASA and Rackspace joined forces to create the OpenStack project, with a goal of allowing any organization to build a public or private cloud using the same technology as top cloud providers.
- **Website:** <http://www.openstack.org/>
- **Key Contributors:** eNovance, HP, IBM, Intel, Mirantis, Rackspace, Red Hat, SUSE, VMware
- **Commercial Support:** Canonical, Cisco, Cloudscaling, Haxexo, HP, IBM, Mirantis, Rackspace, Red Hat, SUSE
- **Project License:** Apache 2
- **Primary Programming Language:** Python (63%)
- **Lines of Code:** 1.69M
- **Key Users:** Best Buy, CERN, Comcast, eBay, Deutsche Telekom, HP, MercadoLibre, PayPal, Rackspace, Sony, Wikimedia, Workday

Open Cloud Profiles: PaaS

CLOUD FOUNDRY

- **Description:** Cloud Foundry is an open Platform-as-a-Service, providing a choice of clouds, developer frameworks and application services.
- **History:** VMware announced Cloud Foundry in April 2011 and built a partner ecosystem.
- **Website:** <http://cloudfoundry.org/>
- **Key Contributors:** ActiveState, AppFog, IBM, Pivotal
- **Commercial Support:** ActiveState, AppFog, IBM, Pivotal
- **Project License:** Apache 2
- **Primary Programming Language:** Ruby (36%), JavaScript (33%)
- **Lines of Code:** 578,000
- **Key Users:** Baidu, GE, Rakuten

DOCKER.IO

- **Description:** Docker is an open-source engine for building, packing and running any application as a lightweight container and is built upon the LXC container mechanism included in the Linux kernel.
- **History:** Docker was written by dotCloud and released in 2013.
- **Website:** <http://www.docker.io>
- **Key Contributors:** dotCloud
- **Commercial Support:** dotCloud
- **Project License:** Apache 2.0
- **Primary Programming Language:** Go (63%)
- **Lines of Code:** 28,000
- **Key Users:** Cloudflare, eBay, Mozilla, Uber

OPENSIFT

- **Description:** OpenShift is Red Hat's Platform-as-a-Service offering. OpenShift is a cloud application platform where application developers and teams can build, test, deploy, and run their applications in a cloud environment.
- **History:** The OpenShift technology came from Red Hat's 2010 acquisition of start-up Makara (founded in May 2008). OpenShift was announced in May 2011 and open-sourced in April 2012.
- **Website:** <https://openshift.redhat.com/app/>
- **Key Contributors:** Red Hat
- **Commercial Support:** Red Hat
- **Project License:** Apache 2
- **Primary Programming Languages:** Ruby (52%)
- **Lines of Code:** 153,000
- **Key Users:** Accenture, Cloud9.io, PayPal, Roche

Open Cloud Profiles: Provisioning and Management

CHEF

- **Description:** Chef is a configuration-management tool, controlled using an extension of Ruby.
- **History:** Released by Opscode in January 2009.
- **Website:** <http://www.opscode.com/chef/>
- **Key Contributors:** Opscode
- **Commercial Support:** Opscode
- **Project License:** Apache 2
- **Primary Programming Language:** Ruby (95%)
- **Lines of Code:** 102,000
- **Key Users:** ancestry.com, IGN, LAN Airlines, University of Minnesota, Voxel

PUPPET

- **Description:** Puppet is IT automation software that helps system administrators manage infrastructure throughout its lifecycle.
- **History:** Founded by Luke Kanies in 2005.
- **Website:** <http://www.puppetlabs.com/>
- **Key Contributors:** Puppet Labs
- **Commercial Support:** Puppet Labs
- **Project License:** Apache 2
- **Primary Programming Language:** Ruby (98%), though PuppetDB is written in Clojure
- **Lines of Code:** 295,000
- **Key Users:** eBay, Google, JP Morgan Chase, Twitter, Viacom

SALT

- **Description:** Salt is a configuration management tool focused on speed and incorporating orchestration features.
- **History:** Salt was written by Thomas S Hatch and first released in 2011.
- **Website:** <http://www.saltstack.com/>
- **Key Contributors:** SaltStack
- **Commercial Support:** SaltStack
- **Project License:** Apache 2
- **Primary Programming Language:** Python (83%)
- **Lines of Code:** 82,000
- **Key Users:** Cars.com, HP Cloud, Hulu, LinkedIn

VAGRANT

- **Description:** Vagrant is an open source tool for building and managing development environments, often within virtual machines.
- **History:** Vagrant was written in 2010 by Mitchell Hashimoto and John Bender.
- **Website:** <http://www.vagrantup.com/>
- **Key Contributors:** Mitchell Hashimoto
- **Commercial Support:** HashiCorp
- **Project License:** MIT
- **Primary Programming Language:** Ruby (100%)
- **Lines of Code:** 19,000
- **Key Users:** BBC, DISQUS, Expedia, Mozilla, O'Reilly

Open Cloud Profiles: Provisioning and Management (Cont.)

JUJU

- **Description:** Juju is a service orchestration management tool.
- **History:** Juju was released by Canonical as Ensemble in 2011 and then renamed later that year.
- **Website:** <https://juju.ubuntu.com>
- **Key Contributors:** Canonical
- **Commercial Support:** Canonical
- **Project License:** AGPL
- **Primary Programming Language:** Go
- **Lines of Code:** 196,000
- **Key Users:** ScraperWiki

OVIRT

- **Description:** oVirt provides a feature-rich management system for virtualized servers with advanced capabilities for hosts and guests.
- **History:** Red Hat first announced oVirt as part of its emerging-technology initiative in 2008, then re-launched the project in late 2011 as part of the Open Virtualization Alliance.
- **Website:** <http://www.ovirt.org/>
- **Key Contributors:** Canonical, Cisco, IBM, Intel, NetApp, Red Hat, SUSE
- **Commercial Support:** IBM, Red Hat
- **Project License:** GPL
- **Primary Programming Language:** Python (69%)
- **Lines of Code:** 206,000
- **Key Users:** Primarily hosting companies, research institutions and universities, and small businesses. A case study with Alter Way was recently published on the oVirt website: http://www.ovirt.org/Alter_Way_case_study

Open Cloud Profiles: Storage

CEPH

- **Description:** Ceph is a distributed object store and file system.
- **History:** Ceph was originally created by Sage Weil for a doctoral dissertation. After Weil's graduation in 2007, he continued working on it full-time at DreamHost as the development team grew. In 2012, Weil and others formed Inktank to deliver professional services and support.
- **Website:** <http://ceph.com/>
- **Key Contributors:** Inktank
- **Commercial Support:** Inktank
- **Project License:** LGPL
- **Primary Programming Language:** C++ (81%)
- **Lines of Code:** 387,000
- **Key Users:** Bloomberg, Dreamhost, eBay

GLUSTER

- **Description:** GlusterFS is a scale-out, distributed file system. It is developed by the Gluster community, a global community of users, developers and other contributors.
- **History:** GlusterFS was originally developed by Gluster Inc., then acquired by Red Hat in October 2011.
- **Website:** <http://www.gluster.org/>
- **Key Contributors:** Citrix, DataLab, Harvard FAS Research Computing, Harvard University, Hortonworks, Intel, Open Source Lab at Oregon State University, Red Hat
- **Commercial Support:** Red Hat, Scalable Informatics
- **Project License:** Dual-licensed under GPL v2 and LGPL v3+
- **Primary Programming Language:** C (92%)
- **Lines of Code:** 588,000
- **Key Users:** Box.net, Intuit, Harvard FAS Research Computing, Pandora, Samsung

RIAK CS

- **Description:** Riak CS is an open source storage system built on top of the Riak key-value store.
- **History:** Riak CS was originally developed by Basho and launched in 2012, with the source subsequently released in 2013.
- **Website:** <http://basho.com/riak-cloud-storage/>
- **Key Contributors:** Basho
- **Commercial Support:** Basho
- **Project License:** Apache
- **Primary Programming Language:** Erlang (94%)
- **Lines of Code:** 24,000
- **Key Users:** Yahoo Japan

Building an Open Cloud

Composed of open-source components from the ground up, from kernel to the developer APIs, the open cloud helps to level the playing field. The availability of multiple open-source projects at each level of cloud architecture makes it possible to not only build an entirely open platform, but to have choices of components while doing so. And with multiple options available at nearly every level, the competitive dynamics between options is accelerating the ongoing functional improvements to the wider open cloud.

To start constructing your open cloud, first seek to identify those open-source components from each level of the stack that best meet your requirements. Whether selection is driven by internal programming language standards, internal resource skill sets, open source license preferences, existing commercial vendor relationships or a combination of all of the above, open-source projects will grant you the ability to build a resilient, scalable open cloud.

As capable as the open cloud is today, work is already underway to further improve it. For example, one common request is improving the ease of installation and administration, which includes increasing the integration between components at various levels. Another area where active improvements are being made is in the final few steps needed to provide an open cloud to others, such as the ability to meter use of memory, CPU, I/O, and so on, so that cloud providers have more flexibility in how they bill customers.

Conclusion

The lessons that Linux taught us are today informing how to build the open cloud. Collaborative development and open source software are defining the next generation of the enterprise and are offering developers and users unprecedented opportunity to innovate.

All of the open cloud projects included in this paper, as well as many others, will be represented at our annual CloudOpen conferences that take place in North America, Europe and Asia. Project leaders and committers from every one of these projects will be speaking or attending, and other projects will be sharing their latest work in the booth areas.



The Linux Foundation promotes, protects and standardizes Linux by providing unified resources and services needed for open source to successfully compete with closed platforms.

To learn more about The Linux Foundation or our other initiatives please visit us at www.linuxfoundation.org