Sysadmin Tools for Tackling the Cloud

MARK HINKLE

Mark Hinkle is the Senior Director, Open Source Solutions, at Citrix. He joined Citrix as a result of their July 2011 acquisition of Cloud.com, where he was Vice President of Community. He is currently responsible for Citrix Open Source Business Office and the Citrix efforts around Apache CloudStack, Open Daylight, Xen Project, and XenServer. Previously, Hinkle was the VP of Community at Zenoss Inc., a producer of the open source application, server, and network management software. He also is a longtime open source expert and author, having served as editor in chief for both *LinuxWorld Magazine* and *Enterprise Open Source Magazine*. Mark Hinkle wrote the book *Windows to Linux Business Desktop Migration* (Thomson, 2006). He is a contributor to NetworkWorld's Open Source Subnet, and his personal blog on open source, technology, and new media can be found at www.socializedsoftware.com. Follow him on Twitter @mrhinkle. mrhinkle@socializesoftware.com

If you read much in the way of the tech media you are probably numb from the cloud computing hype. You have real problems servers to run, backups to execute, and networks to configure. These things exist in two worlds for most sysadmins: your data center and an increasing number of third-party cloud services. Adding the following tools to your sysadmin toolbox will allow you to take advantage of the cloud without missing a beat.

Command and Control Stacks

Most system administrators who do things at scale are already utilizing configuration management tools such as Cfengine, Puppet and/or Chef. All three have considerable capabilities for automation and configuration; however as you move to the cloud, considering tools that utilize access methods that will likely exist natively on your cloud infrastructure as well as your legacy metal is prudent.

Ansible [1] is a simple open source orchestration stack that allows you to communicate with servers over SSH. In turn, this allows you to communicate with machines via a protocol that's likely already available on your machine to execute commands over SSH defined in YAML to call programs written in virtually any language, including Python, Ruby, or even Perl.

SaltStack [2] is another open source configuration management and execution framework along the same lines as Ansible, but it differs by using ZeroMQ as a message bus to execute changes across a network in parallel.

Cloud Controllers

You likely already have a bunch of infrastructure running on your metal and, as that infrastructure goes out of service, you probably are looking to move some of these workloads to the cloud. The decision you may struggle with is determining which cloud. Until you move those services, you won't know how your applications perform and the nuances of each cloud. If you are of a devops mentality, you probably have or will design systems that are easily replicated infrastructure across different architectures, cloud or otherwise. These tools will help you achieve that goal.

If you made your choice already, you may be instrumenting your shop to a certain API. Otherwise, one strategy is to instrument to a single cloud controller that has API compatibility with multiple clouds. Jclouds [3], an Apache incubator project (however, it is very mature), is a perfect example of this type of technology. Jclouds is a library that furnishes a single source to develop tools against—but still broker calls to—multiple clouds through the use of portable cloud abstractions. Jclouds users can take advantage of Java or Clojure as the domain-specific language (DSL). Python experts can take advantage of similar functionality in Apache Libcloud [4], and Ruby enthusiasts can take advantage of their existing skills using Fog [5].

Sysadmin Tools for Tackling the Cloud

Storage

As you start to utilize cloud services, you will quickly realize the advantages and challenges of deploying infrastructure in a much more geographically diverse landscape. Often the challenge is to provide data in a distributed environment with varying levels of utility. For example, you may have data that needs to be stored with varying degrees of availability and integrity. Google has done a considerable amount of research in this area [6] and, if you want to geek out on the considerations for globally distributed data, their findings are informative. Whereas Google focuses on the why, I would direct you to the how.

Gluster [7] is a network/cluster file system written in user space and uses Filesystem in User Space (FUSE) to hook itself with virtual file system (VFS) layer. Gluster works with common concrete file systems, such as ext3, ext4, and xfs. In terms of random file access, the more servers you add the better this scales. Common use cases are for content that is replicated and served behind caching services, such as images or music files delivered by Pandora, a rumored Gluster user.

Ceph [8] is similar in that it is an open source file system and distributed file store that can provide storage, much like Amazon's S3 or Block storage, through their RADOS block device for KVM and additional hypervisors, which will be added soon.

Built on top of Basho's Riak NoSQL database, Riak CS [9] is yet another open source object store. Riak provides a highly available, fault-tolerant storage system that includes compatibility with Amazon's S3 API. Riak CS can provide storage for images, documents, VM backups, archives of information, and other large objects on utility hardware, providing a foundation for storage that compliments your cloud at a much more attractive price point than legacy enterprise storage solutions.

Summary

There are lots of reasons to move to the cloud, and plenty of reasons to continue managing your own infrastructure. The solutions described in this article augment either strategy by providing tools to help you automate your increasingly distributed infrastructure, which lets you keep your options open as you explore new cloud services or look for affordable storage solutions that are uniquely suited to the cloud.

References

[1] Ansible: http://www.ansibleworks.com/

[2] SaltStack: http://saltstack.com/community.html

[3] Jclouds: http://jclouds.incubator.apache.org/

[4] Apache Libcloud: http://libcloud.apache.org/

[5] Fog: http://fog.io/

[6] Ford, D., Labelle, F., Popovici, F., Stokely, M., Truong, V., Barroso, L., Grimes, C., and Quinlan, S., 2010, Availability in globally distributed storage systems, Proceedings of the 9th USENIX Symposium on Operating Systems Design and Implementation, USENIX (2010), http://www.usenix.org/events/osdi10/tech/full_papers/Ford.pdf

[7] Gluster: http://www.glusterfs.org/

[8] Ceph: http://ceph.com/

[9] Riak CS: http://basho.com/riak-cloud-storage/