Evaluation and Integration of OCP Servers from Software Perspective

Internet Initiative Japan Inc.
Takashi Sogabe
Who am I?

• Takashi Sogabe
  – @rev4t
  – Software Engineer, Internet Initiative Japan Inc.
  – Focusing
    • OpenStack
  – Involving
    • OpenContrail, mruby, Riak, etc.
Who is IIJ?

• Network operator
  – Provision of Internet connectivity and WAN service
    • Dedicated line
    • Mobile
  – SEIL
    • Next generation router developed by IIJ
      – Customer Premises Equipment
      – SEIL has a *SEIL Management Framework (SMF)* function for central management of various network functions

• Cloud operator
  – IIJ GIO
    • IaaS, PaaS
Ready-Made Data Centers

- co-IZmo/I
  - Container-Based Data Centers
Turn-Key Solutions for Hybrid Cloud

- Facility
  - co-Izmo/I
  - Conventional Data Center
- Racks, Servers, Switches
  - OCP
- CloudOS
  - OpenStack
- DCIM
  - (TBD)
Components of the System

- Each component is replaceable
- Customers can also buy a part of the system
PoC Environments

- Server
  - Winterfell, etc.
- JBODs
  - Knox, etc.
- Object Storage
  - Kinetic
- Networking
  - Cumulus, ARISTA, BROCADE
OCP Servers

• Simple
  – Minimum amount of peripherals
    • No VGA port
    • No Optical drive
    • 1 NIC with a Mezz slot
    • IPMI/DCMI
  – Easy to replace parts
Simple is better (1)

• Simple system leads to reduce time to boot
  – Winterfell
    • About 30sec
  – Conventional Servers
    • Over 120sec
Simple is better (2)

- OpenStack Integration
  - No need to buy unnecessary peripherals
  - Improve the degree of freedom in a combination of servers such as Compute nodes and Storage nodes
CloudOS

- OpenStack
  - Best way to deploy AWS-like IaaS

- Deployment Issue
  - hard to deploy OpenStack components
  - Commercial Distributions ease the difficulty
    - Metacloud, Mirantis, Piston, RedHat, etc.
  - TripleO (OpenStack on OpenStack)
  - Juju/MAAS
Components need to be Clustered

1. Horizon
   - Neutron

2. Glance
   - VM
   - Provide Image
   - Provision VMs
   - Save Images

3. Nova
   - VM
   - Provide Volume
   - Provide Volume

4. Cinder
   - Backup Volumes

5. Swift

Authenticate

Keystone
PoC Components

Management Server
Juju  MAAS  Nagios

neutron-gateway

The Internet

Clos Fabric

Pacemaker, Corosync, HAproxy

glance
dashboard
cloud-controller
cinder

nova-compute
keystone
swift-proxy
swift-storage

ceph
rabbitmq-server
mysql-cluster
Bare Metal Provisioning

• MAAS
  – Metal as a Service
  – Ideal for Ubuntu

• Cobbler
  – Suitable for other OS (RedHat, CentOS, etc.)
  – Used in OpenStack Distro (Mirantis, etc.)
Juju

• Deployment tool
  – Similar to Chef, Puppet, Ansible, etc.
  – Works well with MAAS
  – App
    • OpenStack
    • Hadoop
    • Etc.
Networking (1)

• Flat Network
  – Simple
  – Hard to scale out

• CLOS Topology
  – Scalable
    • ECMP forwarding balances flows
  – Need to overlay network
    • GRE, VXLAN, etc.
Networking (2)

• Single Pont of Management
  – Networking should incorporate the way of DevOps
    • Chef/Puppet/Ansible, CI
  – Automate everything!
  – Disaggregating hardware from software
    • We can use genuine Linux Distro
      – DevOps friendly 😊
Networking (3)

• Server Switch
  – FBOSS and Wedge
  – Pluribus Networks F64 and E68-M
    • Server with Switch chip

• Eliminate the barriers between servers and switches
  – RTT-sensitive apps can be deployed in server switches rather than conventional servers
Storage (1)

• Knox
  – High Density JBOD
  – 30 HDDs in a 2U chassis
Storage(2)

• Seagate Kinetic
  – Disaggregating HDDs from servers
    • Key-Value Store API
    • Protobuf with Ethernet
  – Benefits
    • Flexibility
    • Scalability
    • Efficiency
Storage (3)

• Swift with Kinetic
  – https://github.com/swiftstack/kinetic-swift
  – Works well with OpenStack

$ swift-ring-builder kinetic.builder
  kinetic.builder, build version 31
  1024 partitions, 3.000000 replicas, 1 regions, 1 zones, 4 devices, 0.00 balance
  The minimum number of hours before a partition can be reassigned is 1

  Devices:   id  region  zone  ip address  port  replication ip  replication port  name

<table>
<thead>
<tr>
<th>Devices</th>
<th>id</th>
<th>region</th>
<th>zone</th>
<th>ip address</th>
<th>port</th>
<th>replication ip</th>
<th>replication port</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.174.251.101:8123</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>127.0.0.1</td>
<td>6010</td>
<td>127.0.0.1</td>
<td>6010</td>
<td></td>
</tr>
<tr>
<td>10.174.251.102:8123</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>127.0.0.1</td>
<td>6020</td>
<td>127.0.0.1</td>
<td>6020</td>
<td></td>
</tr>
<tr>
<td>10.174.251.103:8123</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>127.0.0.1</td>
<td>6030</td>
<td>127.0.0.1</td>
<td>6030</td>
<td></td>
</tr>
<tr>
<td>10.174.251.104:8123</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>127.0.0.1</td>
<td>6040</td>
<td>127.0.0.1</td>
<td>6040</td>
<td></td>
</tr>
</tbody>
</table>

$ swift upload mycontainer test
test
$ swift download mycontainer test -o -awesome

$ for i in {1..4}; do kcmd -H 10.174.251.10$i list objects; done
objects.202f47d75a808c212d43c6dda051f39c.1407379571.83189.data.89a32569-fbbf-4ad3-8e8a-f46e1d632cbb
objects.202f47d75a808c212d43c6dda051f39c.1407379571.83189.data.dd7a79e0-3c92-43fb-937a-10fd11a28c32
objects.202f47d75a808c212d43c6dda051f39c.1407379571.83189.data.df9f0a7f-d2da-4d52-aa06-55e96a1f0dcd7
TIPS FOR OCP DEPLOYMENTS
IPMI Issues

• Need driver support for IPMI device on OCPv2 Windmill
  – https://bugs.launchpad.net/opencompute/+bug/1156667

• Workaround
  – Additional kernel options
    
    maas root tags new name='winterfell' comment='winterfell' ¥
    definition='//node[@class="system"]/vendor = "Wistron"' ¥
    kernel_opts='console=ttyS4 mei.blacklist=yes mei_me.blacklist=yes'
In-band Management

• Useful Information

– https://wiki.ubuntu.com/OpenCompute

```bash
$ sudo apt-get -y install build-essential debhelper dkms bzr libssl-dev
$ bzr branch lp:opencompute/mei
$ bzr branch lp:opencompute/dcmitool
$ bzr branch lp:opencompute/dcmi
$ cd mei
$ fakeroot dpkg-buildpackage -us -uc
$ cd ../dcmitool
$ fakeroot dpkg-buildpackage -us -uc
$ cd ~/dcmi
$ fakeroot dpkg-buildpackage -us -uc
$ cd ~
$ sudo dpkg -i mei-dkms_7.1.21.4.S_all.deb dcmi-dkms_2.1.6.28.MEI_all.deb
$ cd dcmitool
$ sudo dpkg -i dcmitool_1.8.10_amd64.deb
$ sudo bash
# echo "dcmi" >> /etc/modules
```
Summary

• We are developing container modules for IT users all over the world
• Users can have benefits by incorporating OCP design
• Disaggregation is a key factor to improve DevOps