Hyper Converged PLCloud with CEPH

RongFan Wang
PLCloud Tech Director, PowerLeader

Shanghai Ceph Day
2015.10.18

About PLCloud

宝德云 plcloud.com

宝 德 PowerLeader

PowerLeader Science & Technology Group Co., Ltd, (8236.HK)

A diversified IT Group enterprise, is one of the China's biggest server providers with many subsidiaries based on the core business of server, cloud computing and big data.

PowerLeader Cloud —— PLCloud

- Started using OpenStack in 2013
- Started using Ceph in 2014
- Upgraded to OpenStack Icehouse, and provided Public Cloud Hosting Services in 2014
- Bringing OpenStack / Docker / Ceph to traditional enterprises combining with hardware

Agenda

宝德云 plcloud.com

1. Hyper Converged Architecture in PLCloud
2. UseCase: Ceph RadosGW in Media Storage
3. Hyper Converged: OpenStack/Docker/Ceph

Hyper Converged Architecture

宝德云 plcloud.com

From: Converged Infrastructure

The **hardware-focused**, building-block approach of VCE (a joint venture of EMC, Cisco, and VMware), simply known as converged infrastructure;

To: Hyper Converged Infrastructure

The **software defined** approach of Nutanix, VMware, and others called hyper-converged infrastructure.

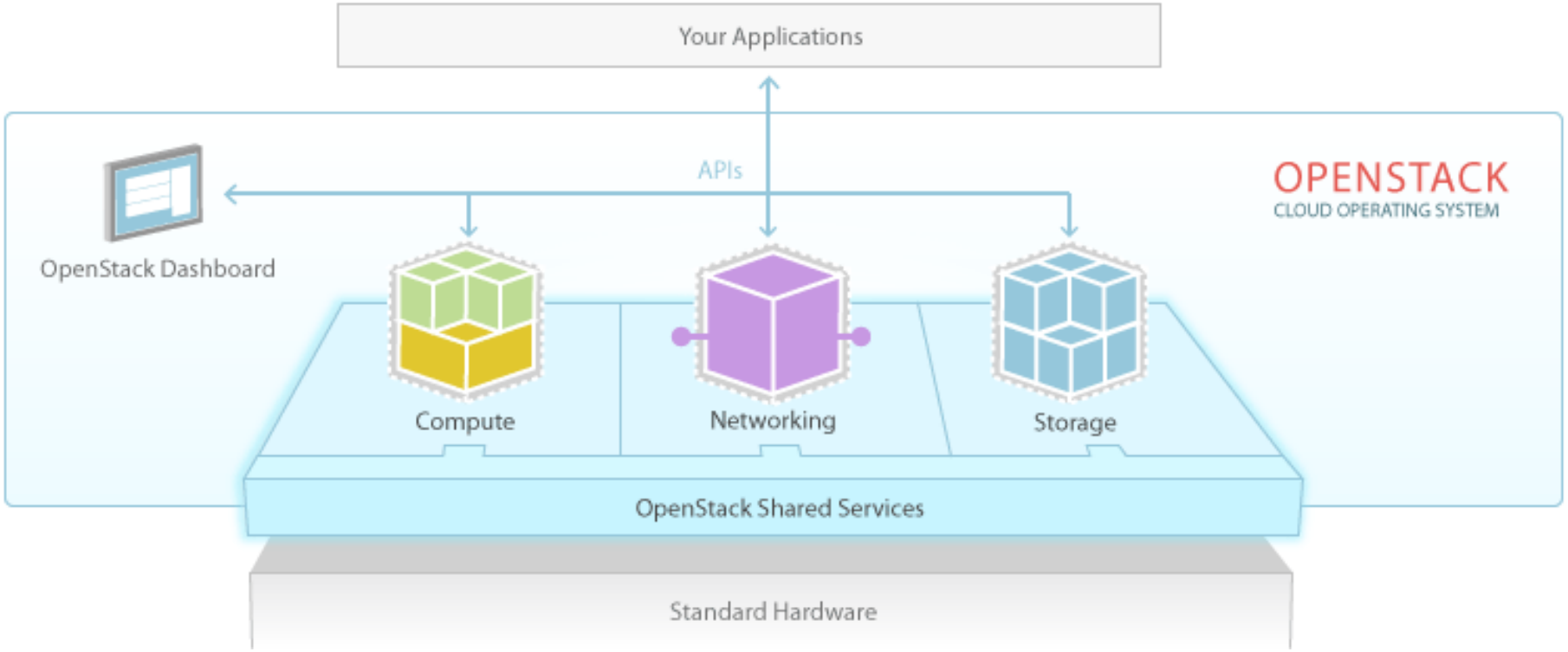
Hyper-convergence is a type of infrastructure system with a **software-centric architecture** that tightly integrates **compute, storage, networking and virtualization resources** and other technologies from scratch in a commodity hardware box supported by a single vendor.

Most important difference between Converged and Hyper-Converged

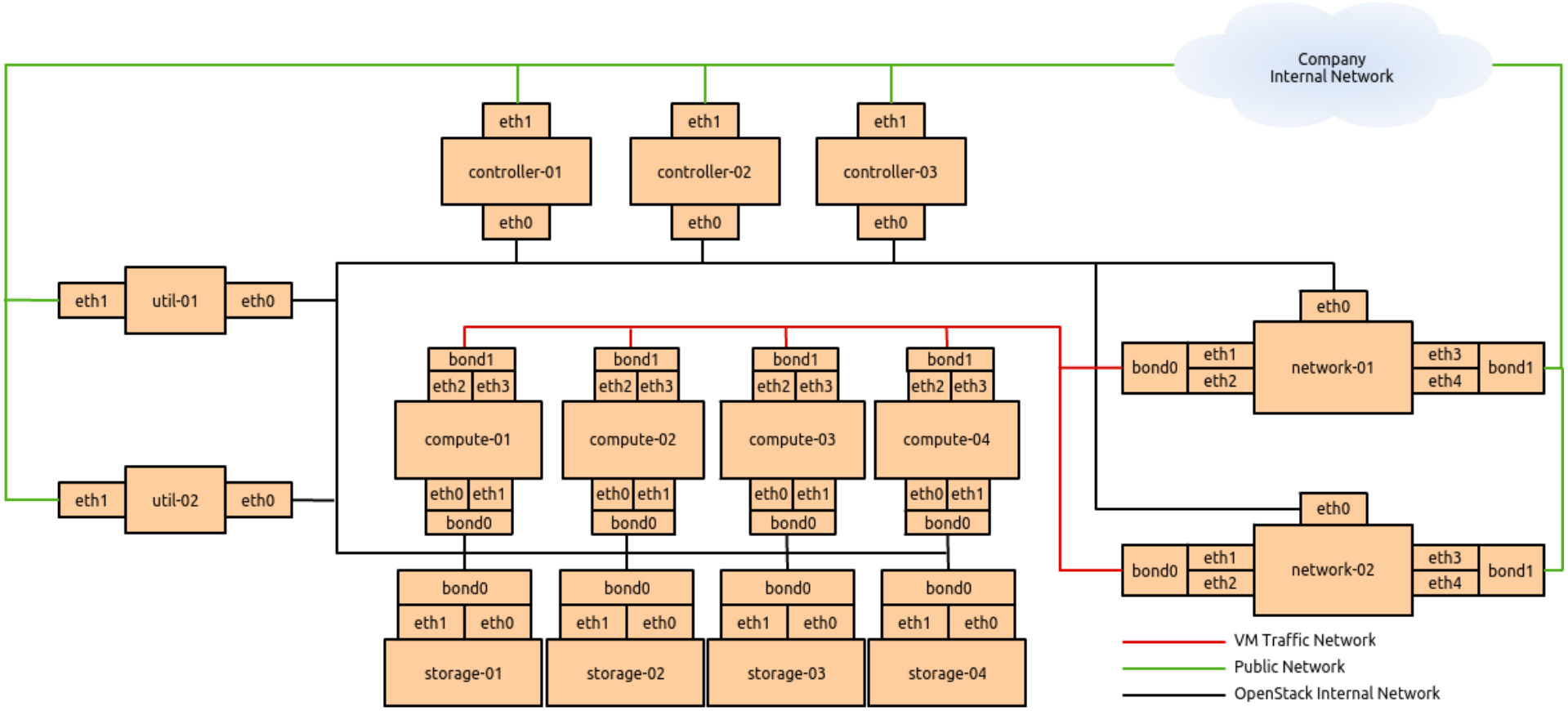
- The technologies in a converged infrastructure can be separated and used independently.

- The technologies in a hyper-converged infrastructure, however, are so integrated that they can not be broken down into separate components.

Hyper Converged OpenStack

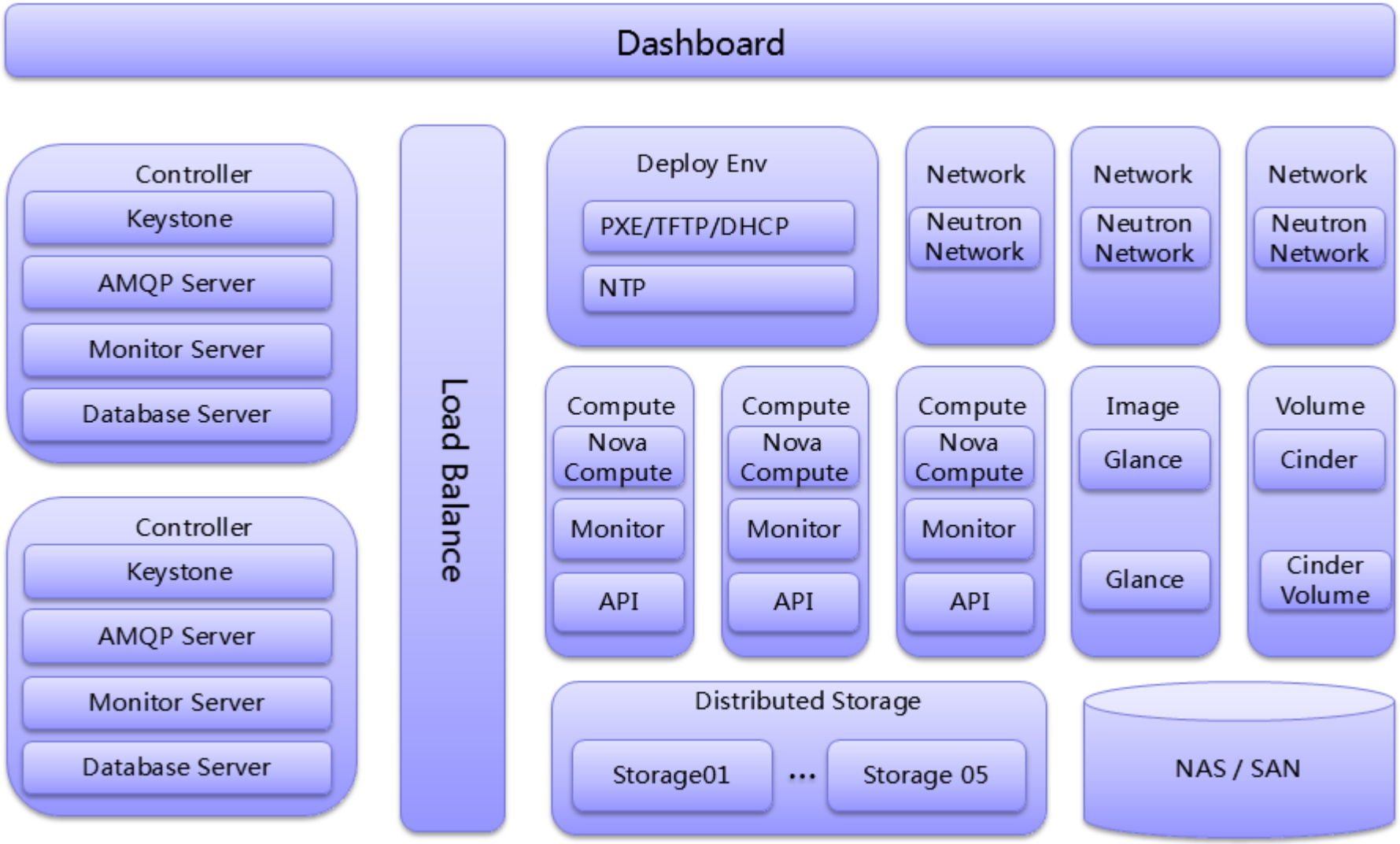


OpenStack Node Deployment

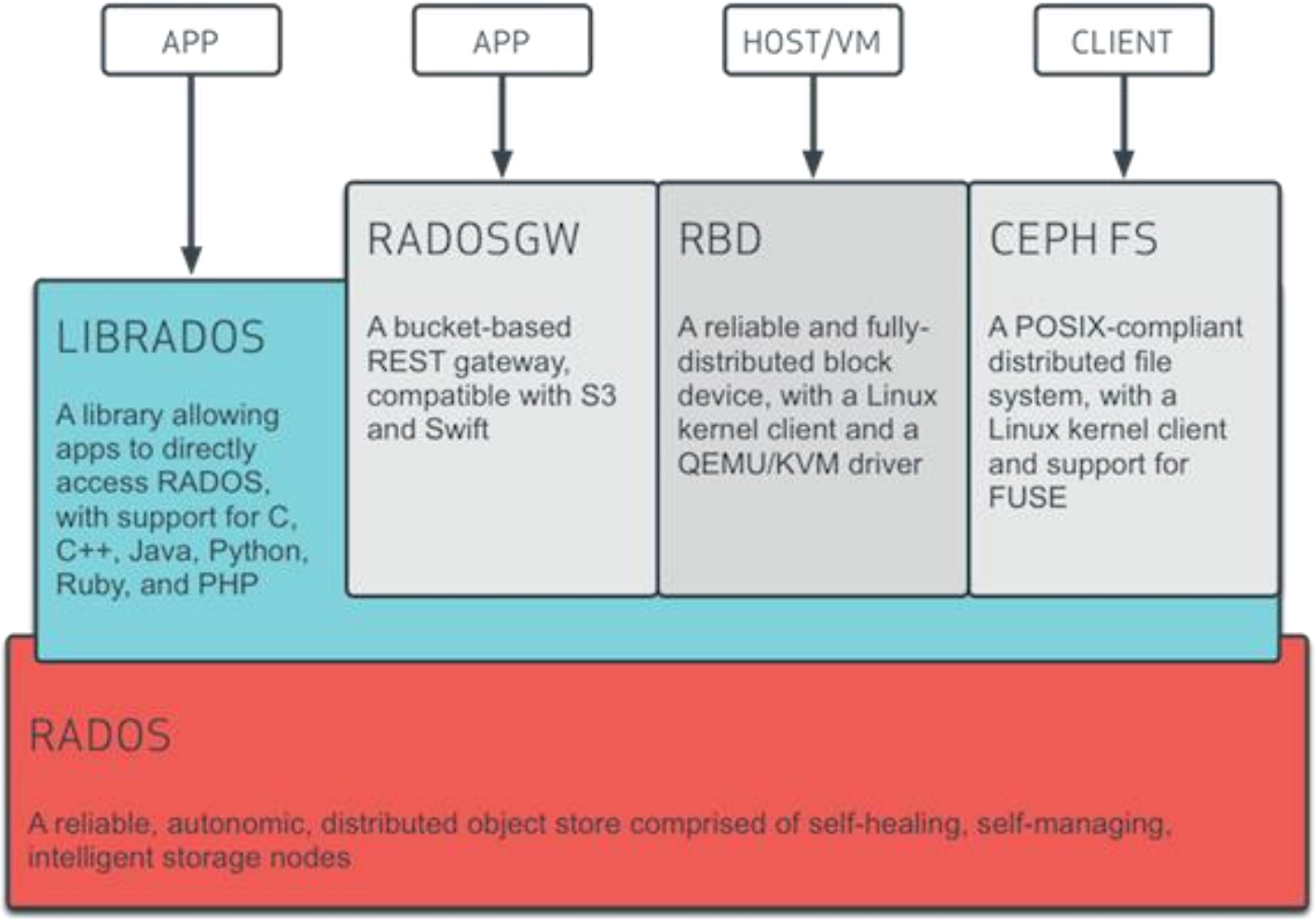


From: http://docs.openstack.org/openstack-ops/content/index.html

Hyper Converged PLCloud



Hyper Converged Ceph



Agenda

宝德云 plcloud.com

1 Hyper Converged Architecture in PLCloud

2 UseCase: Ceph RadosGW in Media Storage

3 Hyper Converged: OpenStack/Docker/Ceph

Ceph RadosGW in Media Storage

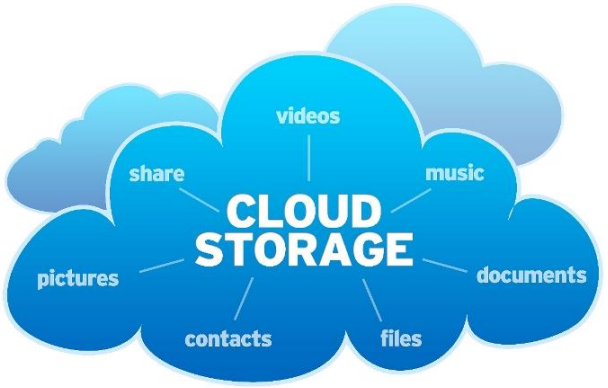
宝德云 plcloud.com

Enterprise Media Storage User Requirements:

- Large Capacity
- Rapid Growth
- Access Security
- Cost Effective

Enterprise Media Storage System Requirements:

- Scale out, not Scale UP
- High Concurrency, High Availability



Choosing an Appropriate Storage Solution

- SAN or NoSAN?
- Block or File or Object?
- Commercial or OpenSource?
- Public or Private?

宝德云 plcloud.com

Why do we choose Ceph?

- Pure Software
- Open Source, Commercial Support
- Unified Storage: RBD, RGW, CephFS
- Scale Out
- Self Healing
- Replication and Erasure Coding
- Integrate well with OpenStack

Ceph RadosGW in Media Storage

宝德云 plcloud.com

Use Case: iermu Media Service on PLCloud

- Run media web/app/db vm over OpenStack and Ceph RBD
- Use Ceph RGW as media resource storage
- Put video TransportStream/jpg file via c-language programme
- Manage resource via python-swiftclient
- 400+KB per video ts file
- Reserved video ts/jpg file 7 days or 30 days
- Allow media server temporary access to objects
- Provide media service for Internet and Intranet User

Used Ceph RadosGW APIs

- Authentication
- Container Ops: Create, List, Delete
- Object Ops: Create, List, Delete, Get, Copy
- Temp URL Ops: POST Temp-URL Keys, GET Temp-URL Objects
- Admin Ops: User Manage, Capability Manage, Quota Manage

Ceph RadosGW in Media Storage

宝德云 plcloud.com

Recommended Server Specifications(CIB)

	PL-4020	PL-4020Q	PL-4040	PL-4040Q	PL-4060	PL-4060Q
Form Factor	2U/1Node	2U/4Node	2U/1Node	2U/4Node	2U/1Node	2U/4Node
Nodes	4	4	4	4	4	4
Platform Vendor	PowerLeader	PowerLeader	PowerLeader	PowerLeader	PowerLeader	PowerLeader
CPU	E5-2630v3 8C	E5-2630v3 8C	E5-2650v3 10C	E5-2650v3 10C	E5-2670v3 12C	E5-2670v3 12C
Memory	64GB per node	64GB per node	128GB per node	128GB per node	256GB per node	256GB per node
HDD	900GB SAS	900GB SAS	900GB SAS	900GB SAS	-	-
SSD	-	-	INTEL DC S3710 400GB SSD	INTEL DC S3710 400GB SSD	INTEL DC S3710 800GB SSD	INTEL DC S3710 800GB SSD
NIC	2 * AOC-STG-I2T 10G(SPF+)	2 * AOC-STG-I2T 10G(SPF+)	2 * AOC-STG-I2T 10G(SPF+)	2 * AOC-STG-I2T 10G(SPF+)	2 * AOC-STG-I2T 10G(SPF+)	2 * AOC-STG-I2T 10G(SPF+)

Agenda

宝德云 plcloud.com

1 Hyper Converged Architecture in PLCloud

2 UseCase: Ceph RadosGW in Media Storage

3 Hyper Converged: OpenStack/Docker/Ceph

OpenStack Storage

- Ephemeral Storage
- Persistent Storage

OpenStack Object Storage

- Swift Object Storage
- Glance Image/Snapshot

OpenStack Block Storage

- Glance Image/Snapshot
- Nova VM Root Disk
- Cinder Volume
- Cinder Snapshot
- Cinder Backup

宝德云 plcloud.com

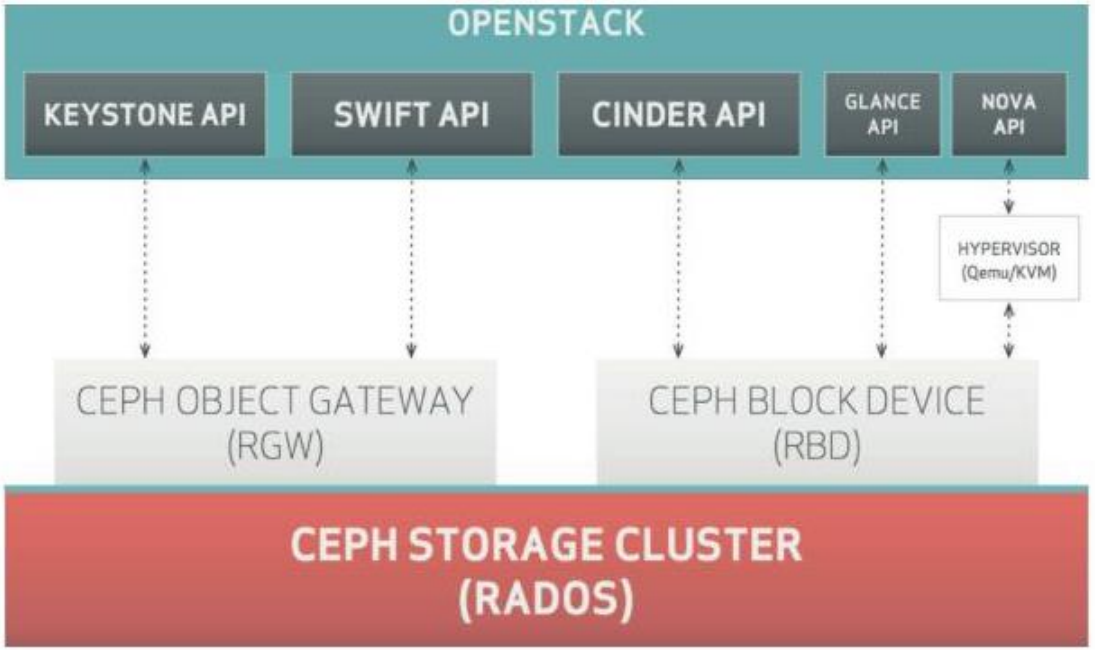
OpenStack Storage

	Ephemeral storage	Block storage	Object storage
Used to…	Run operating system and scratch space	Add additional persistent storage to a virtual machine (VM)	Store data, including VM images
Accessed through…	A file system	A block device that can be partitioned, formatted, and mounted (such as, /dev/vdc)	The REST API
Accessible from…	Within a VM	Within a VM	Anywhere
Managed by…	OpenStack Compute (nova)	OpenStack Block Storage (cinder)	OpenStack Object Storage (swift)
Persists until…	VM is terminated	Deleted by user	Deleted by user
Sizing determined by…	Administrator configuration of size settings, known as flavors	User specification in initial request	Amount of available physical storage
Encryption set by…	Parameter in nova.conf	Admin establishing encrypted volume type, then user selecting encrypted volume	Not yet available
Example of typical usage…	10 GB first disk, 30 GB second disk	1 TB disk	10s of TBs of dataset storage

http://docs.openstack.org/openstack-ops/content/storage_decision.html

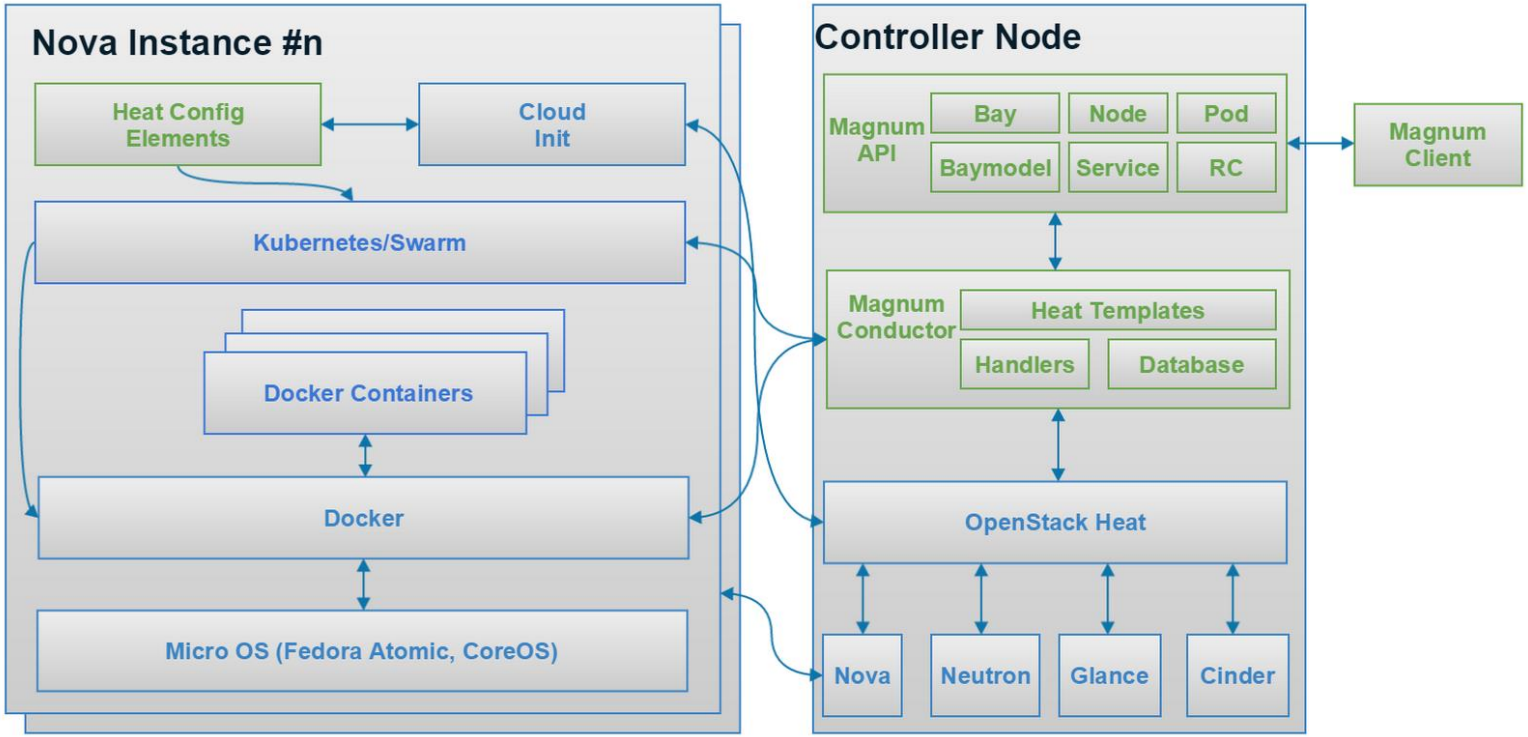
Ceph & OpenStack

- Glance Image/Snapshot
- Nova VM Root Disk: Boot From Image (Copy on Write)
- Cinder Volume: Boot From Volume / Attached Volume
- Cinder Snapshot: Flatten / Clone / Rollback
- Cinder Backup: ^backup\.([a-z0-9\-]+?)\.snap\.(.+)$
- Swift Storage: RadosGW Swift compatible



Docker & OpenStack

- Docker on Nova VM
- Nova Docker Driver
- Heat Docker Plugin
- Magnum: Kubernetes, Docker Swarm, Mesos



Storage in Docker

- Docker Registry Storage
- Docker Data Volumes
- Docker Data Volume Containers

Ceph & Docker

- Ceph Rados Gateway driver for Docker Registry
- Map RBD device inside Docker Container
- CephFS as Data Volume
- CephFS as NAS Storage
- Run Ceph in Containers

宝德云 plcloud.com

Hyper Converged in Private Cloud Case

- OpenStack + Docker + Ceph
- All OpenStack Storage on Ceph
- 18*(5 OSD+1SSD) / Ceph RBD / CephFS
- 785VM / 4vCPU32GB per VM
- Ubuntu14.04 / Docker1.6.1 / 150+ Containers per VM
- All VM Mount Ceph FS
- Mount VM Directory as Container’s Data Volume
- Boot 1 VM < 5s
- Boot 1 Container < 1s
- Boot 150+Containers < 120s

Giving Thanks to Linux and Open Source

@宝德云 in WeChat