OpenStack Use Case at GREE

GREE, Inc.

2014/2/14
OpenStack Days Tokyo 2014

Self Introduction

Koichi Watanabe (29)

Infrastructure Dept., GREE
Infrastructure Design and Development

Favorites:
富士山, 富士宮, 日馬富士.. sth like Fuji

Yohei Matsuhashi (27)

Infrastructure Dept., GREE
Infrastructure Design and Development

Favorites:
ものづくり (Robotics, etc)

Agenda

GREE

- Introduction
- Infrastructure Overview (before OpenStack)
- Why Virtualization ?
- System Overview
- Implementation
- Issues from testing
- Issues from operation
- Recent Work
- Conclusion

About Us

Company	GREE, Inc.
Est.	Dec 7th, 2004
Location	Roppongi, Tokyo
Business	Social Gaming Business
	Social Media Business
	Platform Business
	Advertising & Ad Network Business
	Licensing & Merchandising Business
	Venture Capital Business

GREE

Social Games

Many kinds of games

踊り子クリノッペ

釣り★スタ

探検ドリランド

聖戦ケルベロス

Infrastructure Overview (before OpenStack)

Infrastructure Overview

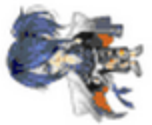
Over 1,250+ products

In-House Mobile Social Games

3rd Party Mobile Social Games

Mobile Social app Platform

Avatars



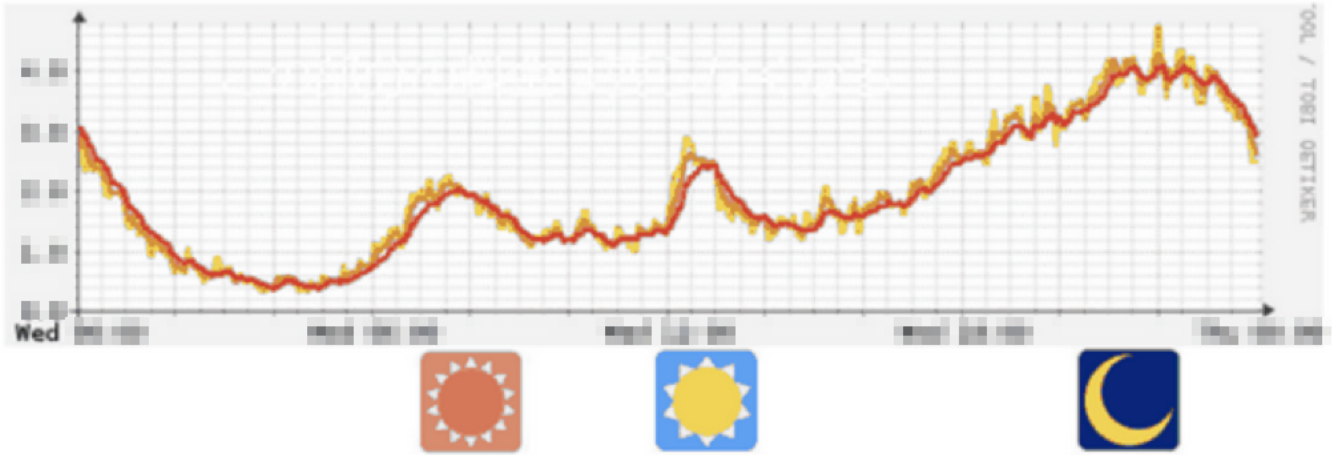
Social Networking Service

GREE

GREE's Server Farm

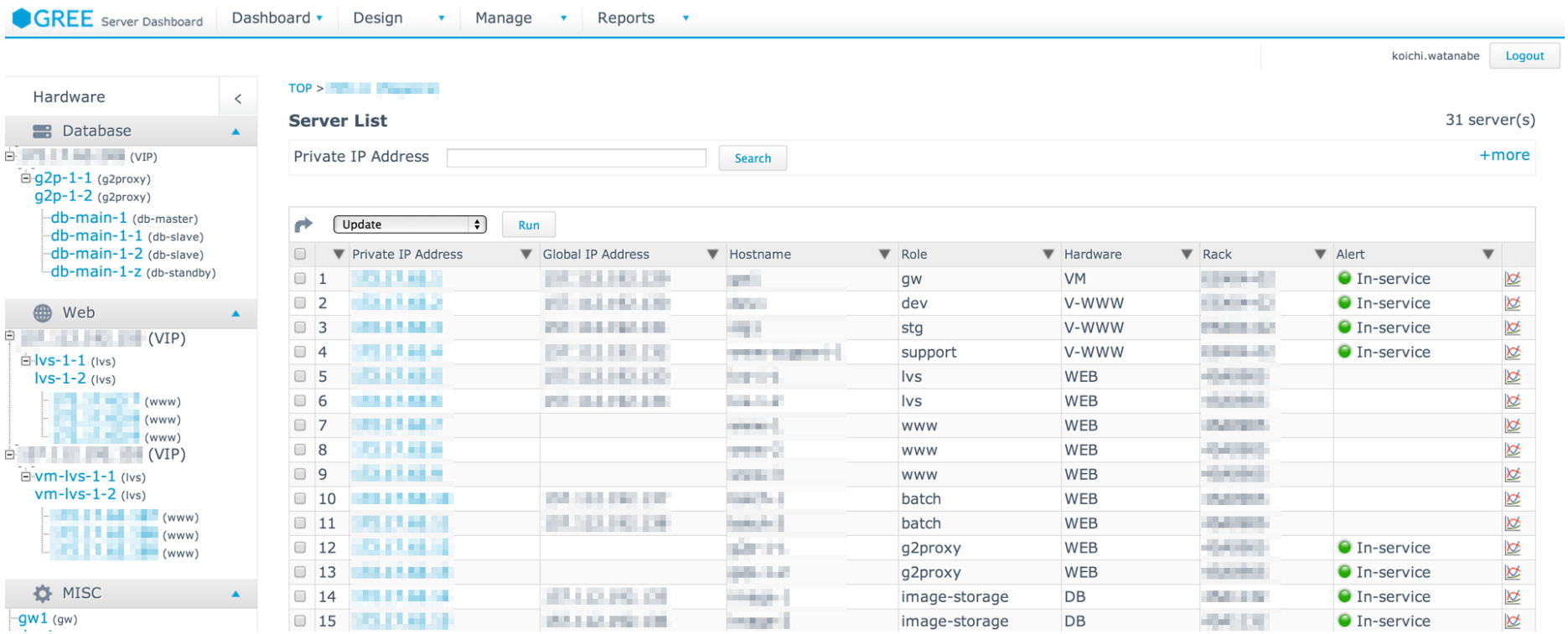
- On-premise
  - hundreds of servers per service/game

- Recurrent Peak traffic
  - several times per day



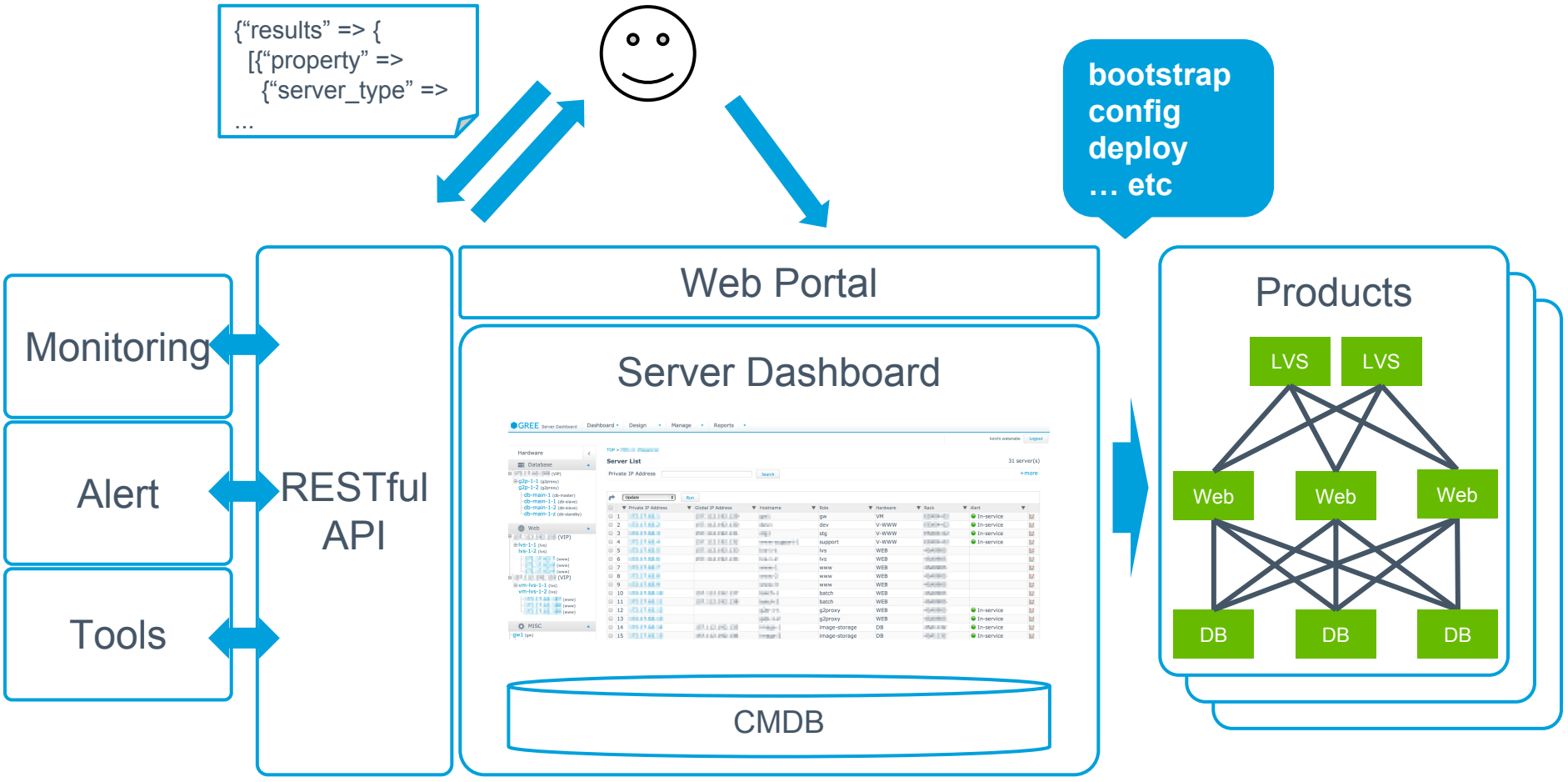
Server Dashboard

Server Management Portal, with Auto-Configuration features



Infrastructure Overview

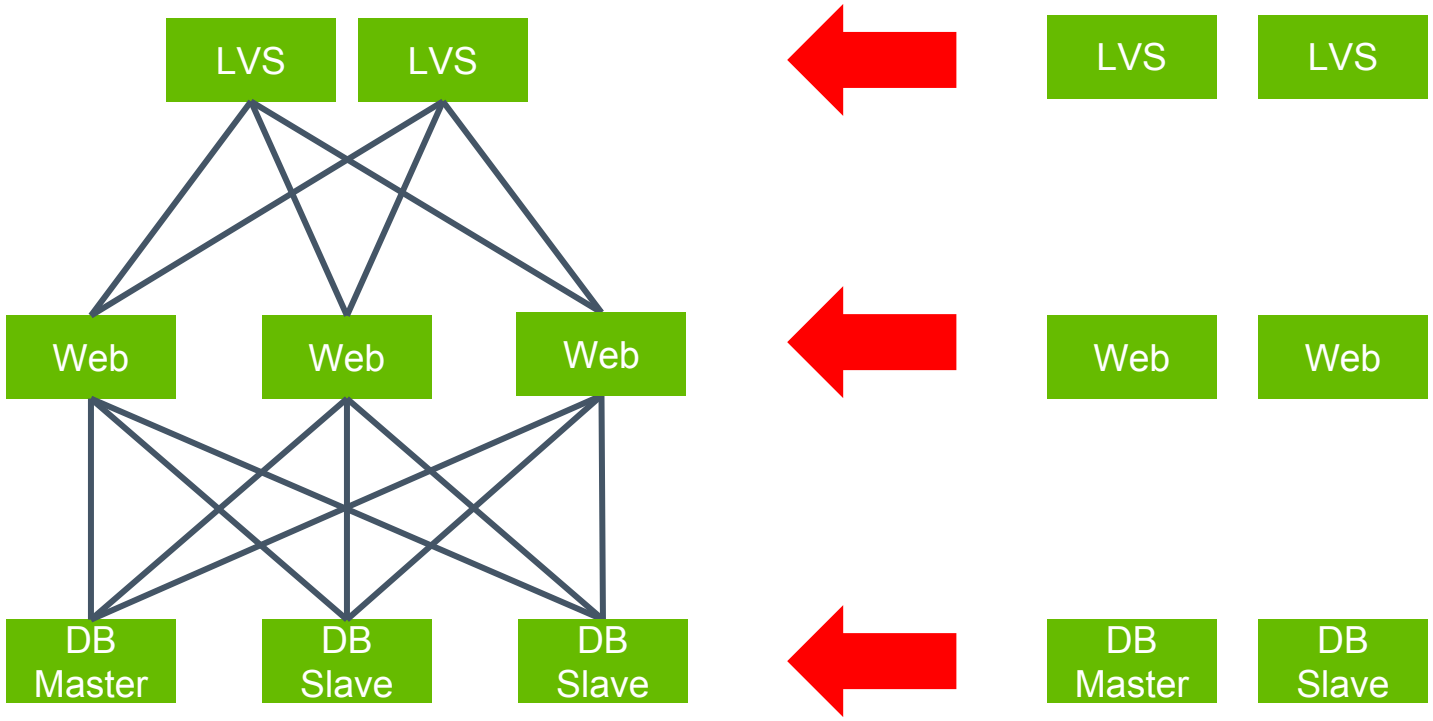
Controlled by RESTful API



GREE

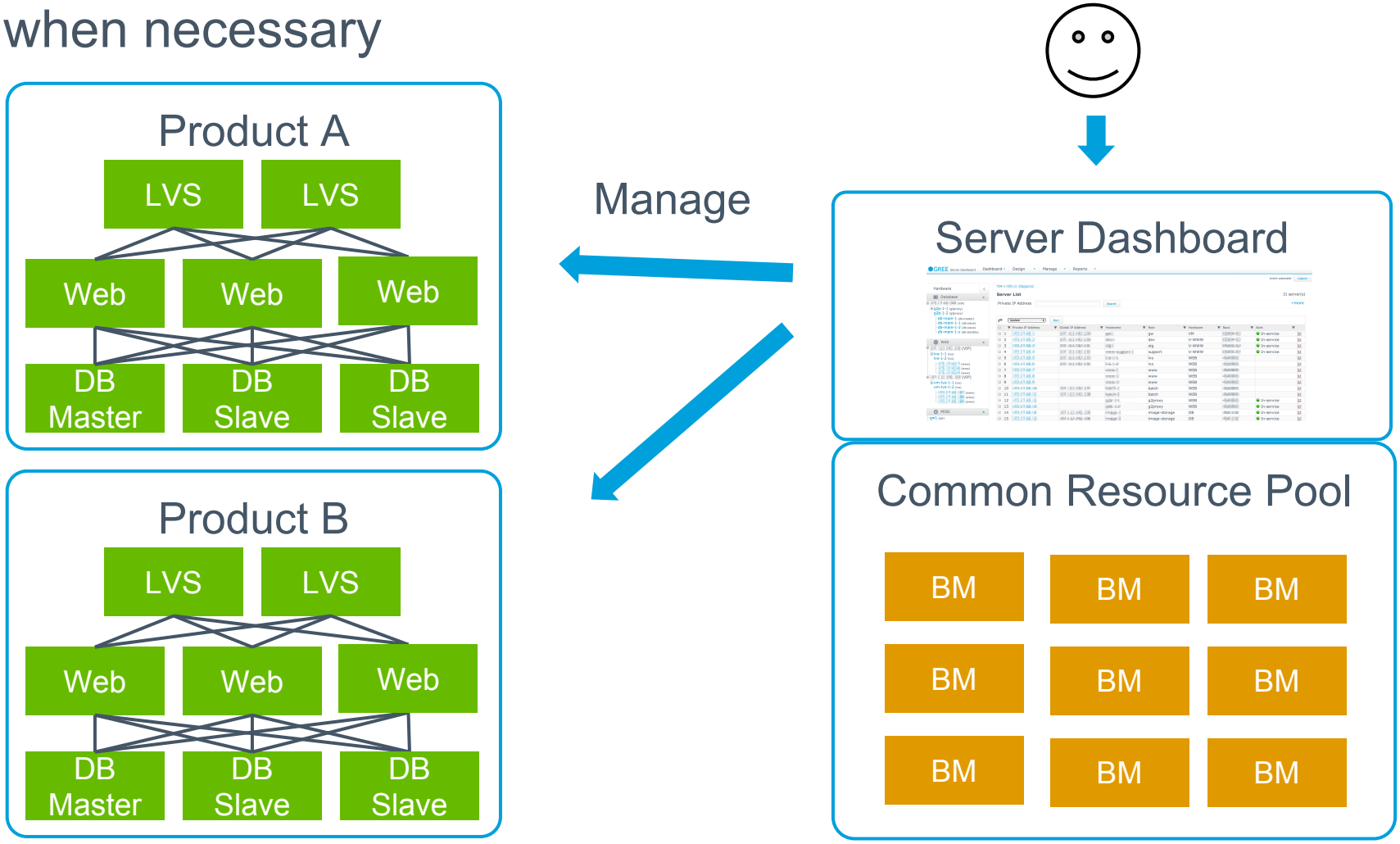
Elasticity

- Able to scale-in/scale-out all server components
  - Controlled with Server Dashboard



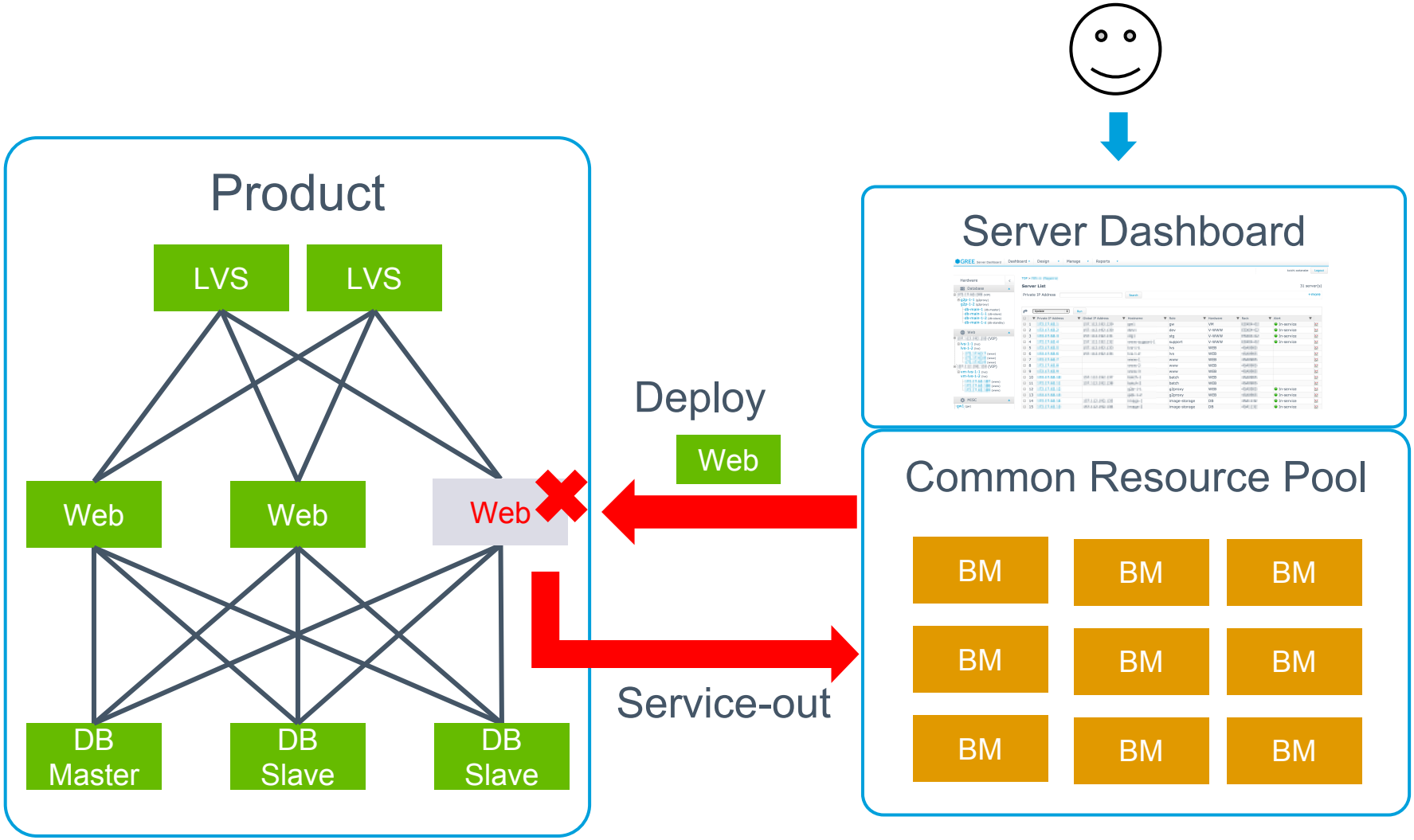
Server Resource Pool

Activate and use servers from common resource pool as and when necessary



Operation Cycle

Spare servers are used to deal with failures



Why Virtualization ?

GREE

Use case

- Improve server resource usage
  - we have various services and roles
  - some of them could be run on smaller servers
    - eg. gateway, batch servers

- Improve automation
  - Automate machine and software tests
    - boot, configure, test and destroy

- Reduce operation costs
  - HA/FT
  - Live migration
  - Floating IP

GREE

Points to consider

Reason	Comments
Reduced Performance	Approx 5-50% of Hypervisor resource overhead
Reduced Dependability	Noisy Neighbour problem
Decreased Visibility	More complex architecture
More Dependencies	More external packages, such as OpenStack

GREE

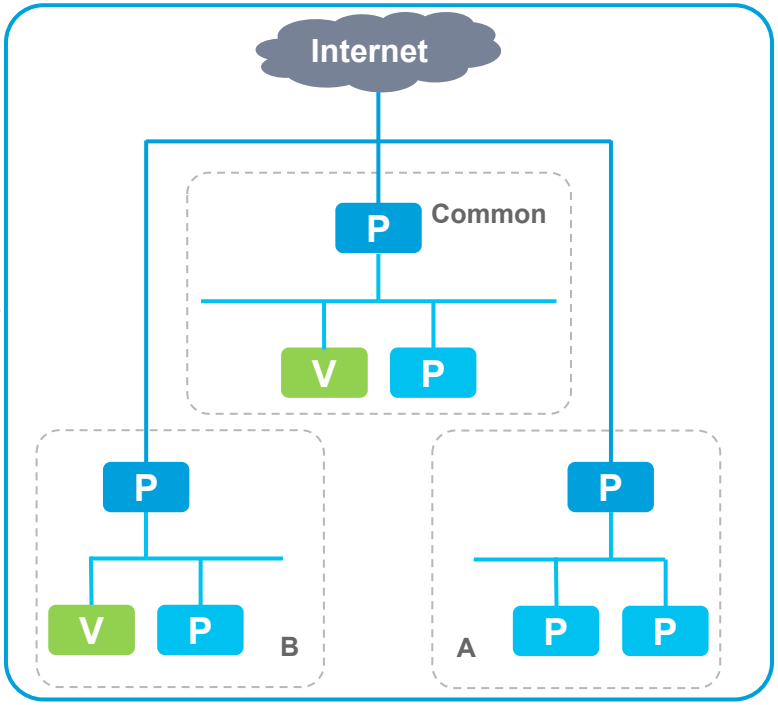
Evaluation Summary

Reason	Comments
Reduced administration effort	Automation tools are mandatory
Reduced fixed costs	Reduced overall server resource usage
Improved scalability	Machine provisioning is faster than physical machines
Increased flexibility	Multi-tenancy, better network and resource management
Increased application availability	HyperVisor overhead is still a drawback
Reduced development effort	Test automation would be helpful

GREE

OpenStack x GREE

- Major evaluation axis
  - OSS
  - All API
  - All Python
  - Scalability
  - Loose coupling in each elementals
  - Multi-tenancy
  - Compatible with Swift
  - Familiar with our architecture
  - Possible to use current servers
  - Very active contributors and community



openstack™

GREE

Current Production Environment (May 2013)

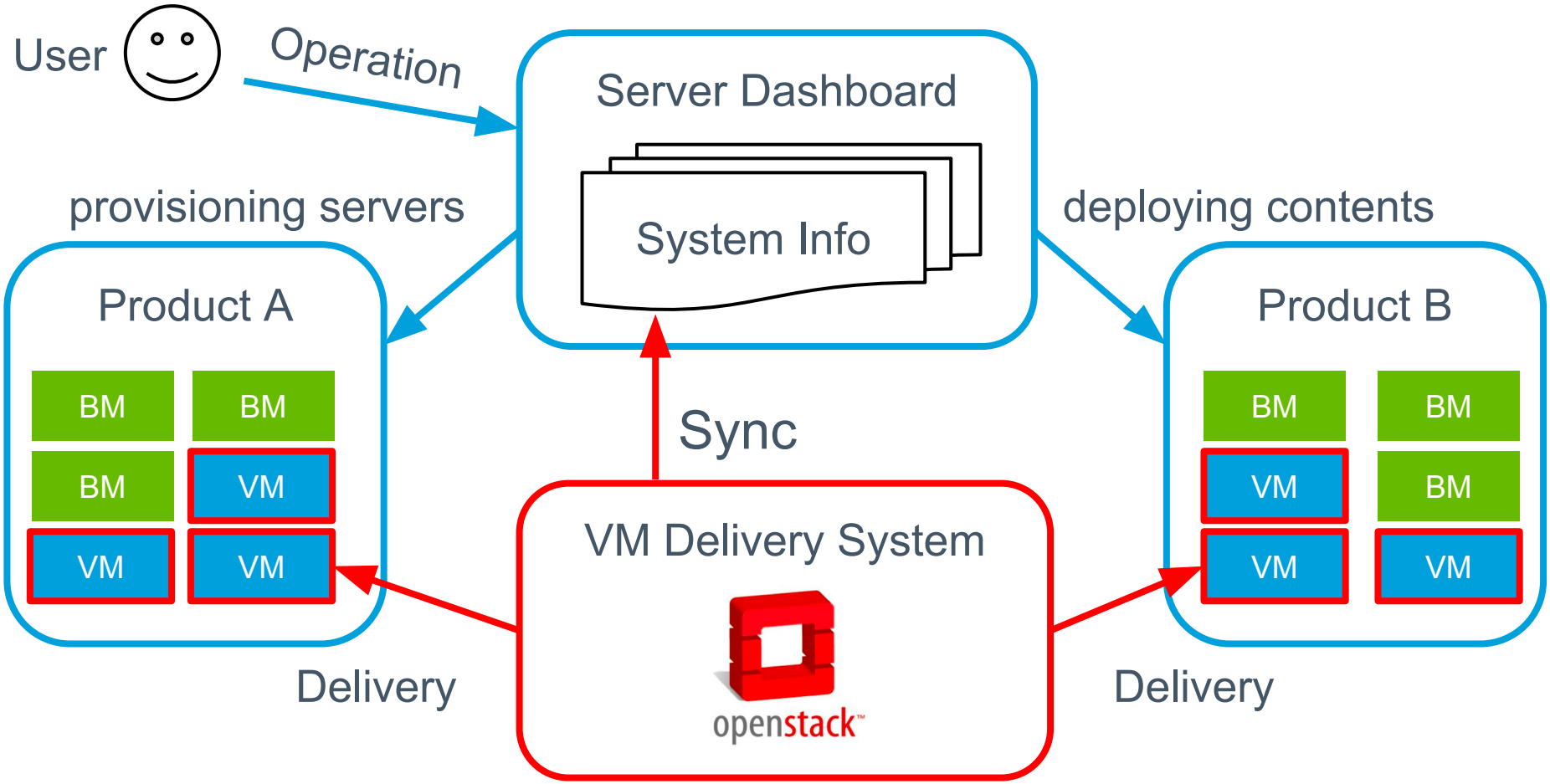
Category	Comments
OpenStack	Folsom
Hypervisor	KVM
Host OS	Ubuntu
Networking	Open vSwitch
Deployment Tool	Chef
Storage	LVM

System Overview

GREE

OpenStack integration to our existing system

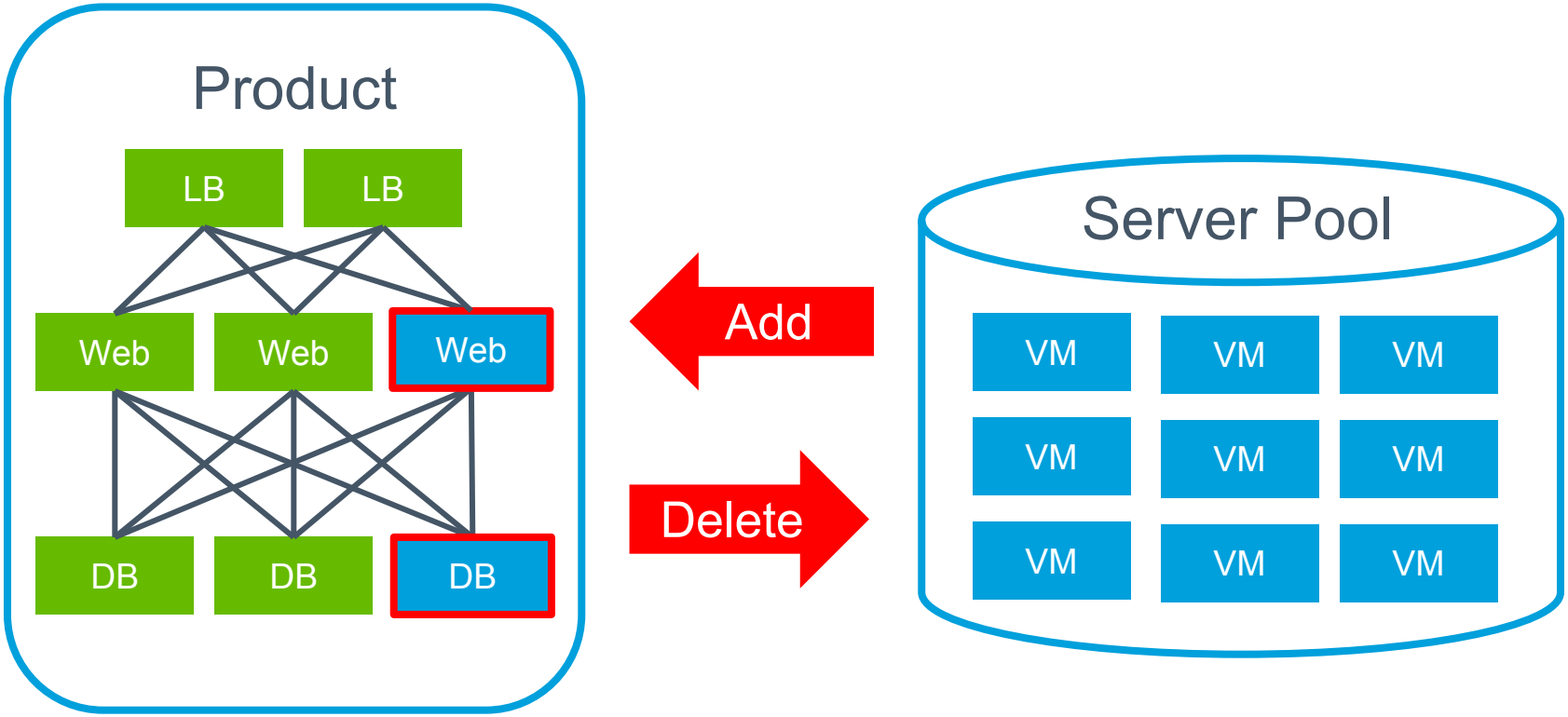
- Manage VMs and BMs on the same interface
  - Sync VM info with Server Dashboard



GREE

Rapid Scale In / Out

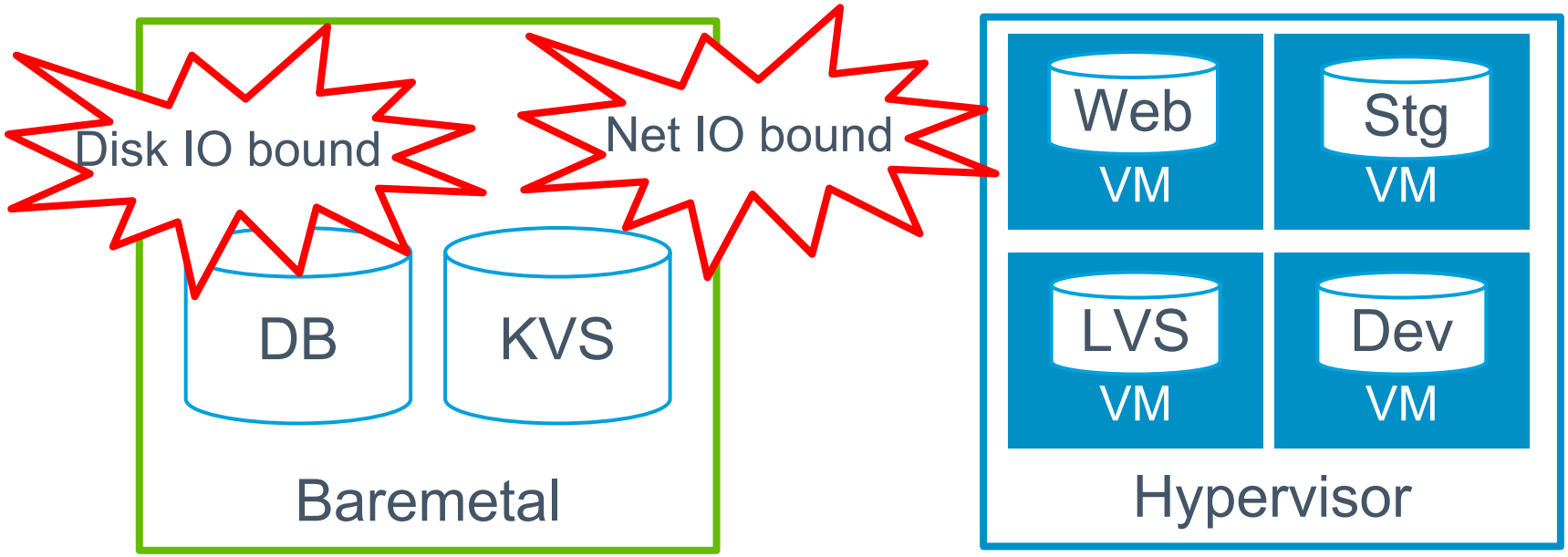
- On-Demand VM delivery
  - Add VMs from common server pool in OpenStack
  - Improved server deployment time



GREE

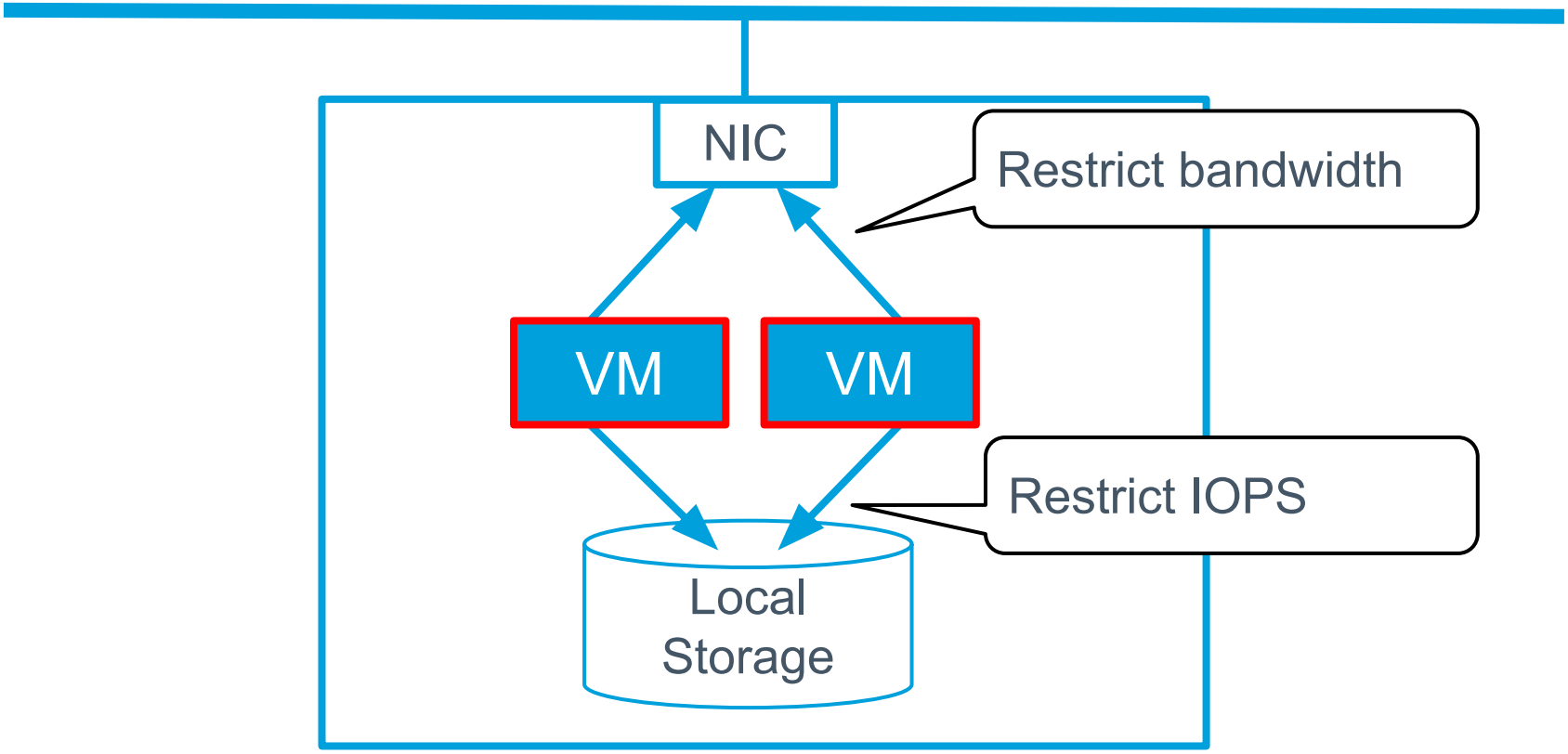
Hybrid BMs and VMs

- Able to choose BM or VM depending on app's workload
  - Baremetal : high I/O
  - VM : low resource usage



QoS

- Guaranteed VMs' performances
  - Network bandwidth
  - Disk IOPS

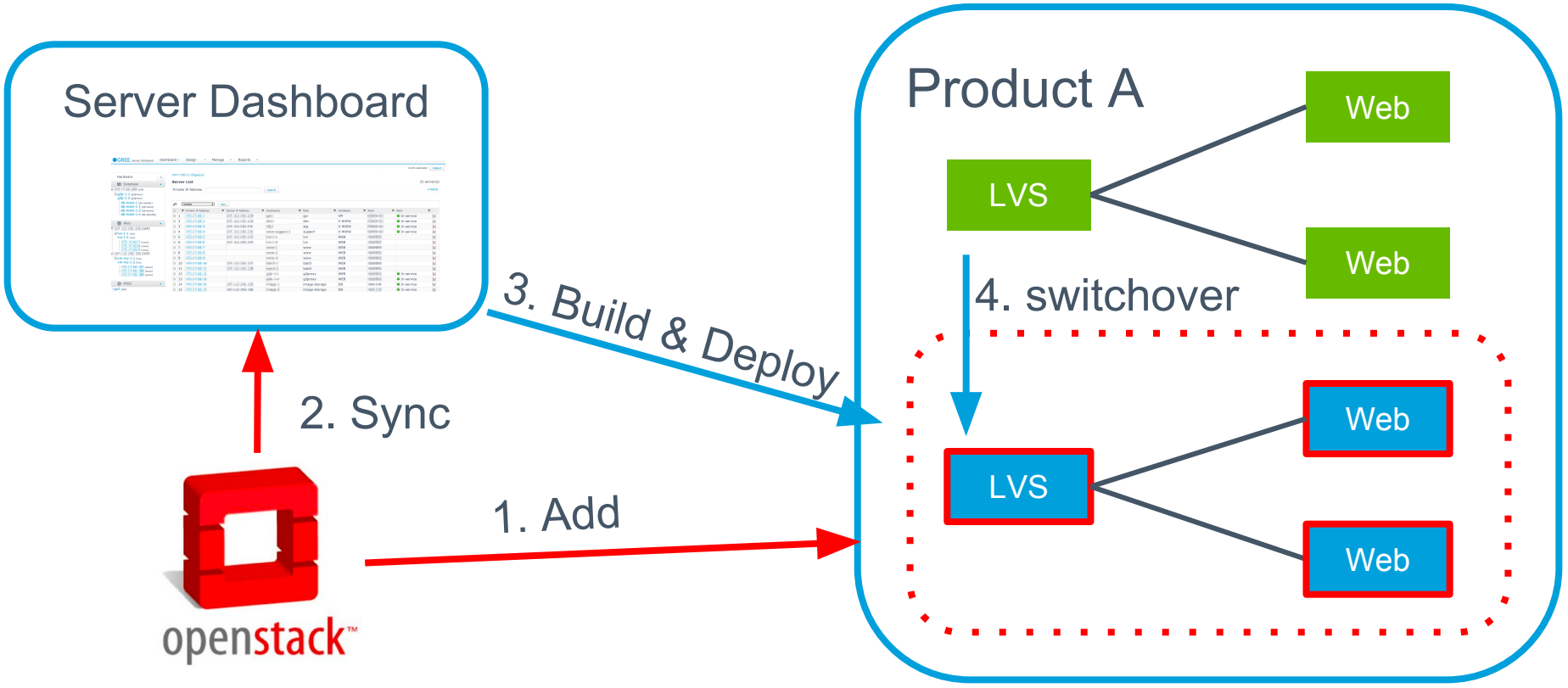


GREE

Zero-downtime migration

- Migrate between VM ↔ BM using automation tools

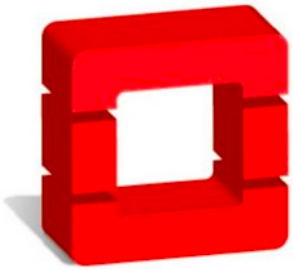
example :



Implementation

GREE

GREE is completely built on OSS



(not in production)

… and more!!

GREE

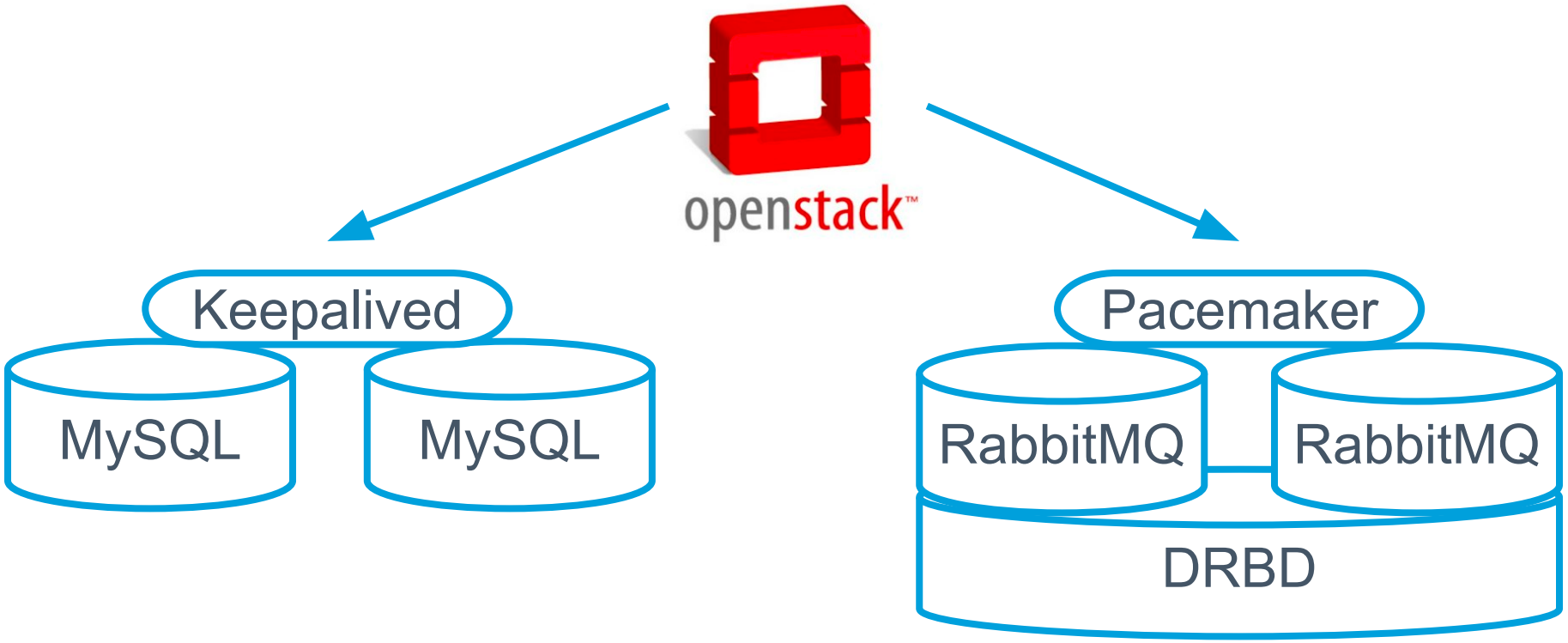
Deploying OpenStack

- Auto-provisioning with Chef
  - Middleware
  - Controller
  - Compute
- Provisioning flow
  - Install packages
  - Performance tuning
  - Apply patches
  - etc.

GREE

Middleware

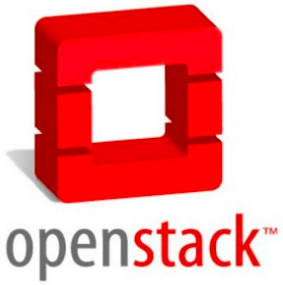
Component	Middleware	Redundant
Database	MySQL	Keepalived
Messaging queue	RabbitMQ	Pacemaker + DRBD

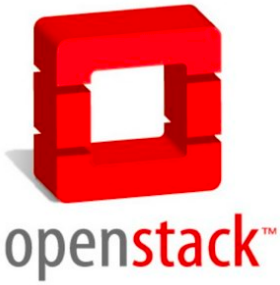


GREE

OpenStack Controller

- OpenStack APIs
  - Keystone :: Authentication
  - Nova :: Compute
  - Quantum/Neutron :: Network
  - Glance :: Image
  - Cinder :: Volume
- Redundant tool
  - Keepalived





GREE

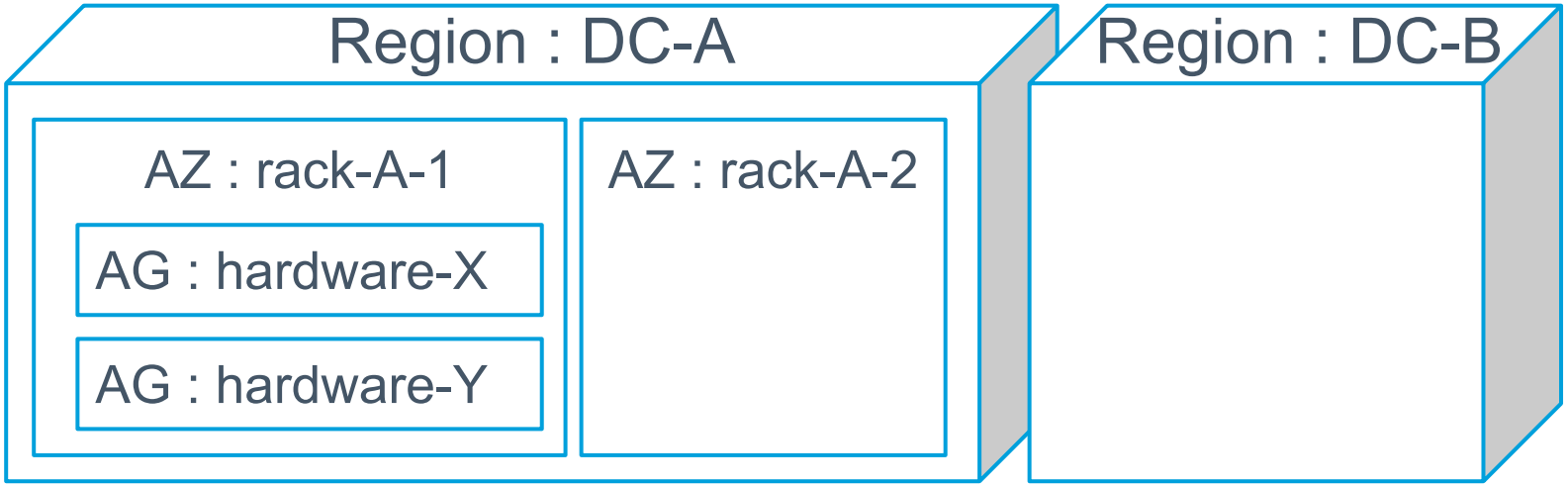
Compute Node

- Improve KVM performance
  - virtio :: storage
  - vhost :: network
  - hugepages :: memory
- Able to apply QoS
  - cgroup :: disk io
  - tc :: traffic control
- How
  - Chef configuration deployment
  - OpenStack patching

GREE

Region / Availability Zone / Aggregate

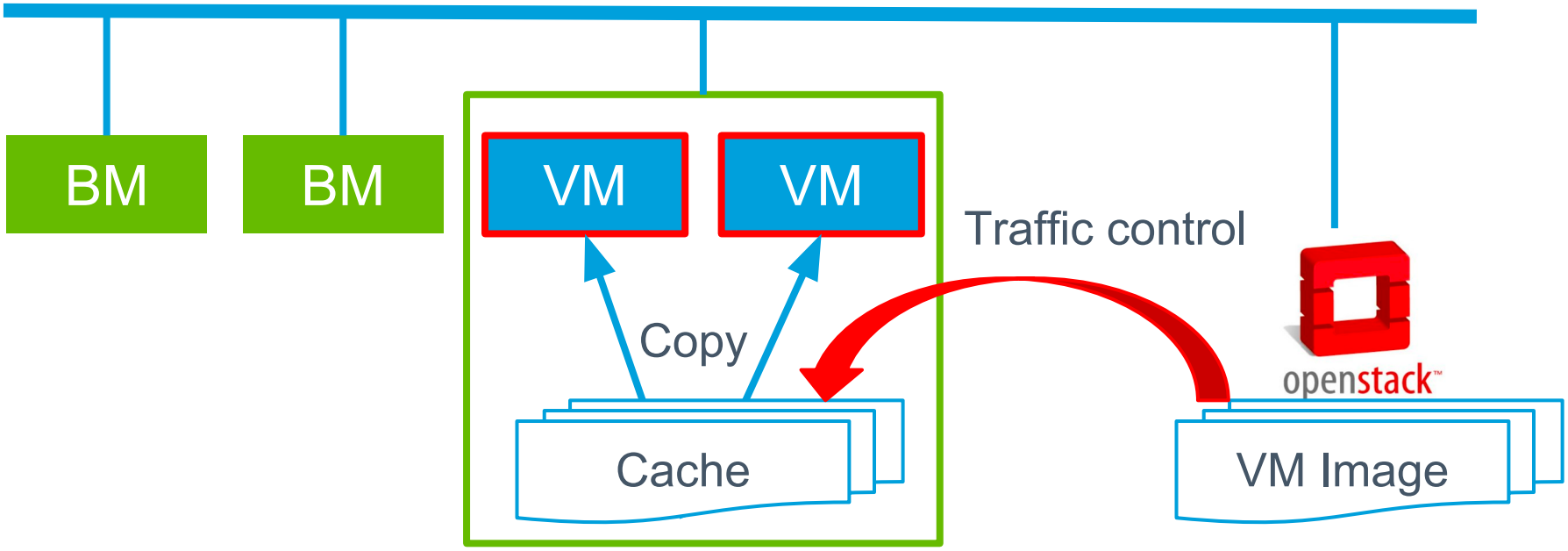
Concept	Apply to
Region	Datacenter
Availability Zone (AZ)	Rack
Aggregate (AG)	Hardware type



GREE

Resource management during VM deployment

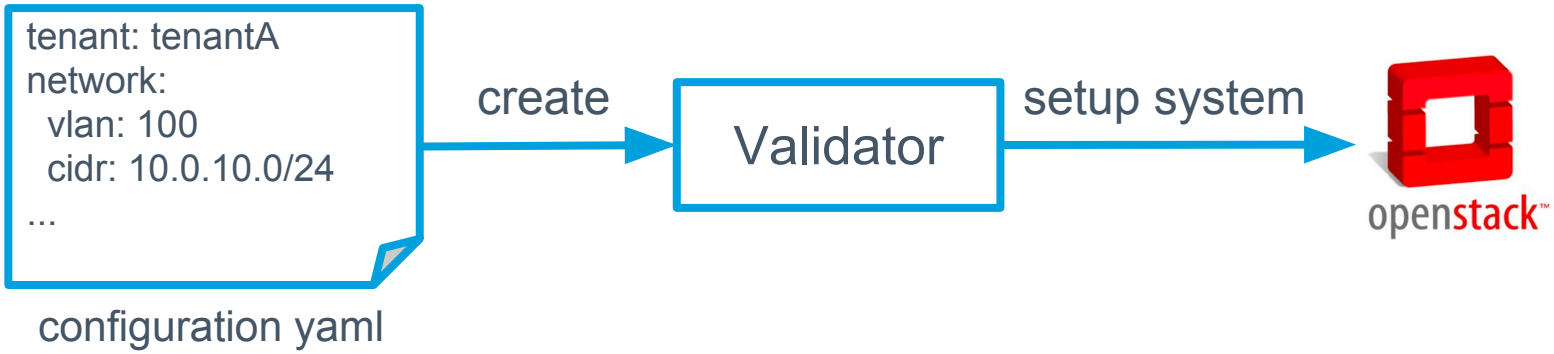
- Manage network IO by traffic control
- Prioritize disk IO with ionice
- Use local VM image cache in the hypervisor



GREE

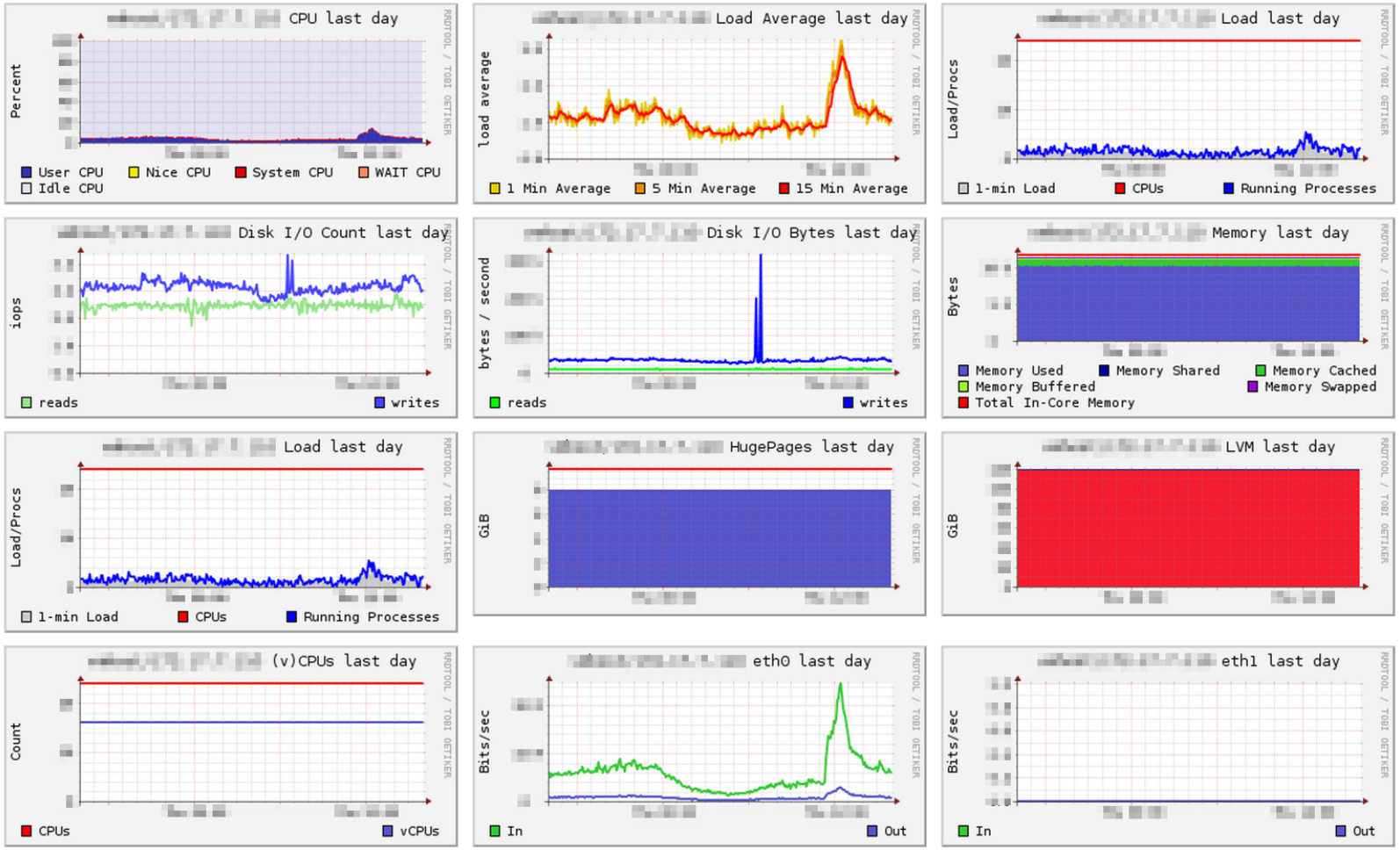
Original Operation Tools

- OpenStack command wrappers
  - manage tenant/user/network
  - include failsafe check mechanism
  - customize VM's parameters
- VM placement scheduler
  - Selects an appropriate place depending on the service and the server role



Gathering 100+ RRD metrics

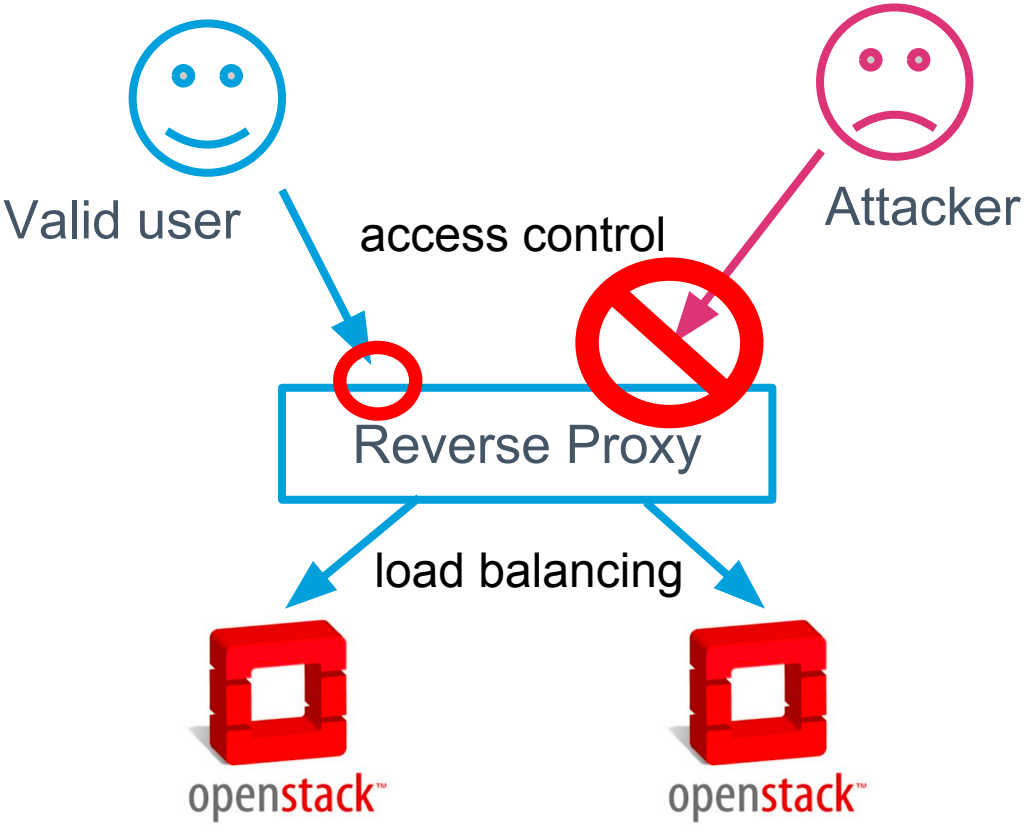
- Load measuring
  - ex. memory usage, disk usage, assigned CPU cores



GREE

Managing OpenStack API

- Use reverse proxy in front of OpenStack APIs
  - Apply Access Control List and SSL/TLS
  - API load balancing

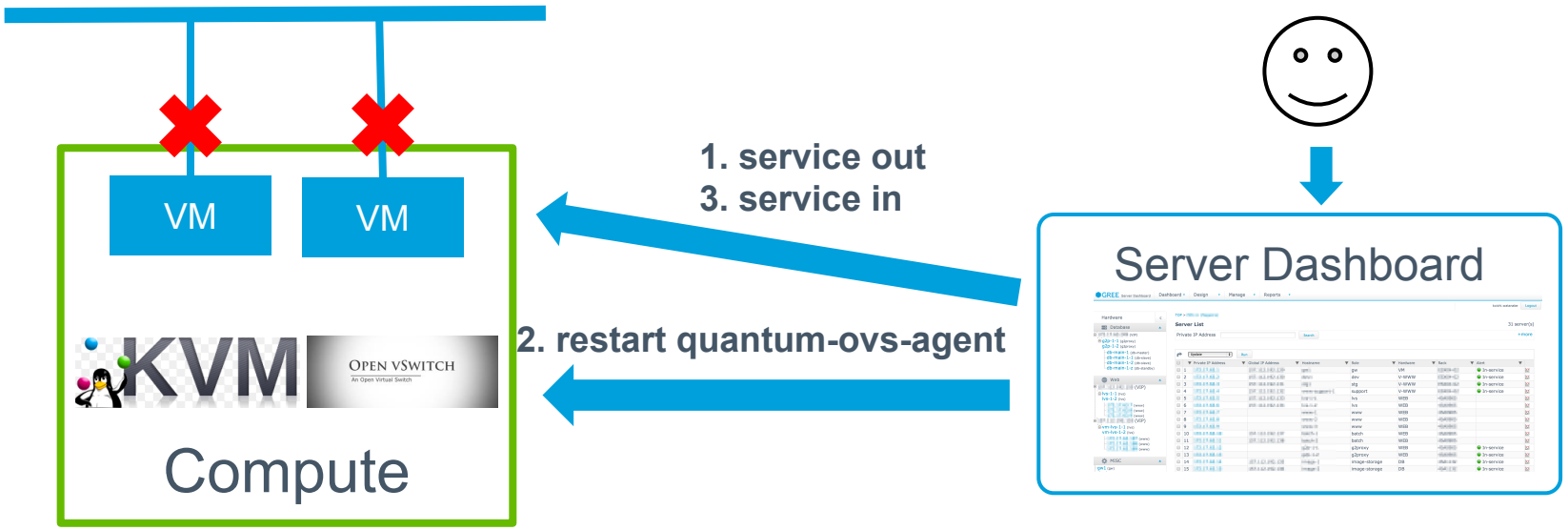


Issues from testing

GREE

Connection temporary fails when ovs-agent is restarted

- Issue
  - ovs flow entry is re-initialized upon restart
- Solution
  - Base policy of not restarting agent
  - Service out before restarting the agent

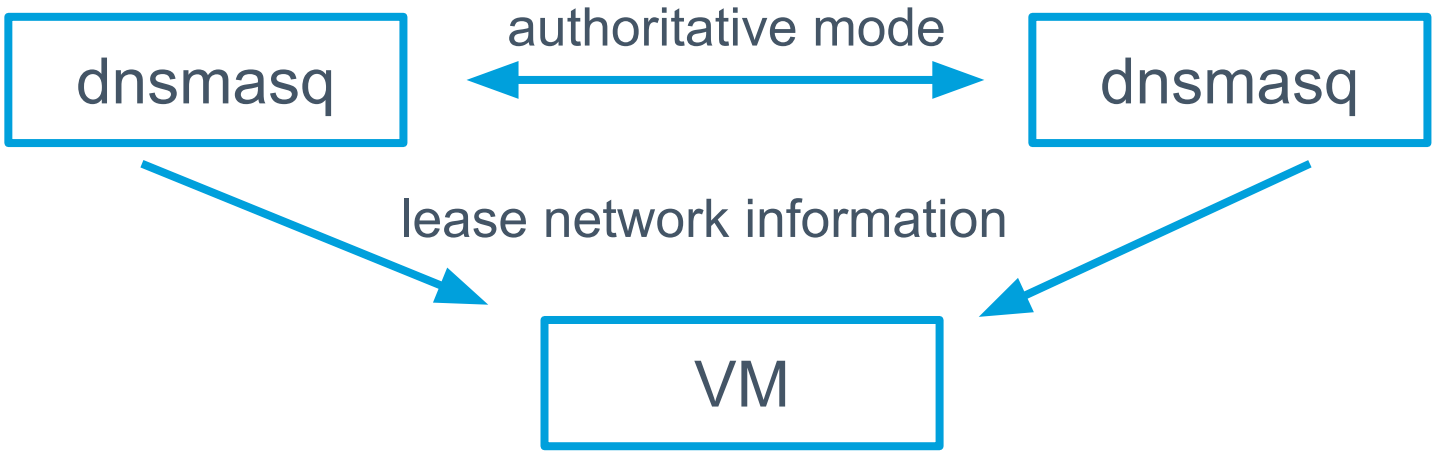


GREE

Invalid redundancy of DHCP servers

- Issue
  - dhcp-agent couldn't failover leased IP addresses
    - Depends on DHCP protocol specification
- Solution
  - Use DHCP authoritative mode

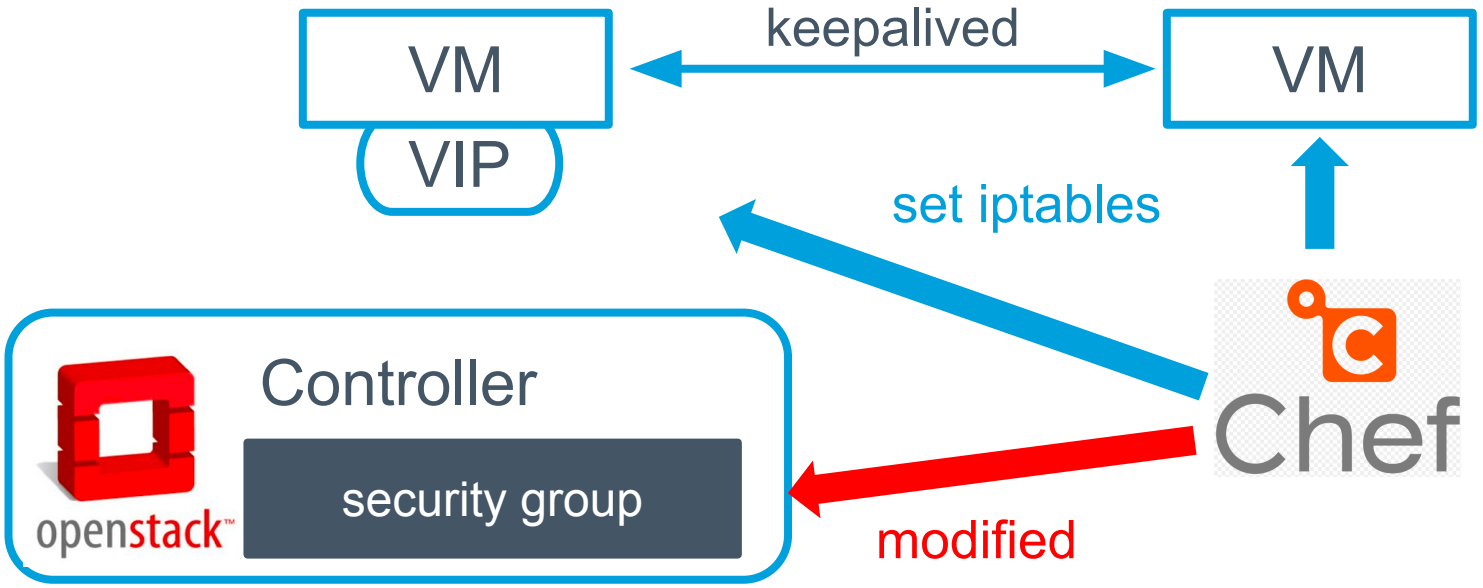
    or
  - Use static IP address



GREE

Unable to set VIP for LVS

- Issue
  - Denied VIP communication by security group
- Solution
  - Modified security group rules
    - Apply patches by chef



GREE

Extra Issues

Issue	Cause	Solution
KVM suddenly fails	memory overcommit	apply failsafe mechanism in the placement scheduler monitoring overcommit status

… and more!!

[Caution]

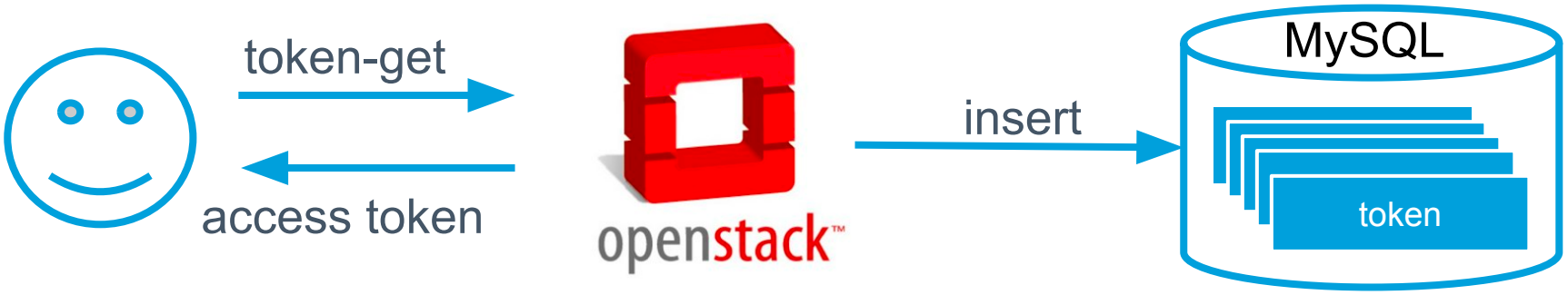
- Some of issues we experienced while testing OpenStack folsom have already been fixed in the latest release
- Solutions we introduced may not be the best ways (as you may know…)

Issues from operation

GREE

DB record consumption by keystone token

- Issue
  - Accumulating keystone tokens continuously
- Solution
  - Set cron to delete expired tokens in database
  - (Otherwise, migrate keystone backend to memcache)

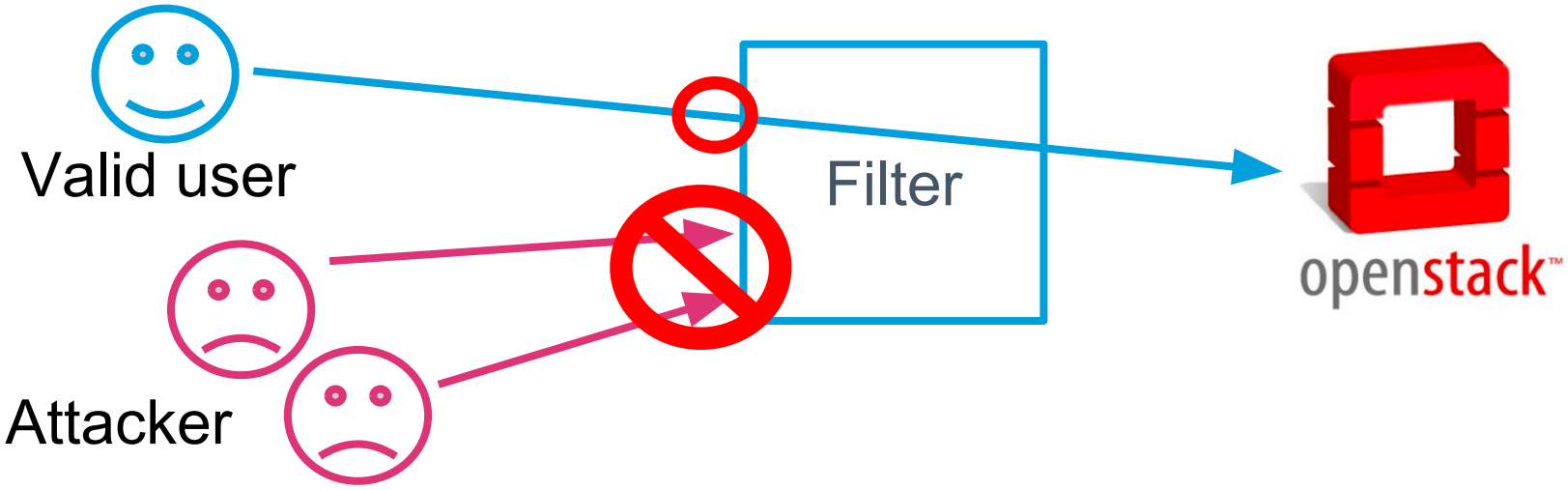


GREE

Taking measures to DDoS attack

- Issue
  - OpenStack controller could have many global IPs
  - ssh brute force attack, etc

- Solution
  - Set filter rules by iptables
  - (Otherwise, turn off DHCP server)



GREE

nova-compute process down after deleting flavor

- Issue
  - Nova-compute process crash when a running VM flavor is deleted
- Solution
  - Apply failsafe check mechanism by our tools

GREE

Extra Issues

Issue	Cause	Solution
VMs for DB	Overhead of the KVM	Use bare metal servers (Otherwise, replace to SSD/FusionIO)
KVM bugs	bugs in para-virt module. ex. kvm-clock	apply workarounds

… and more!!

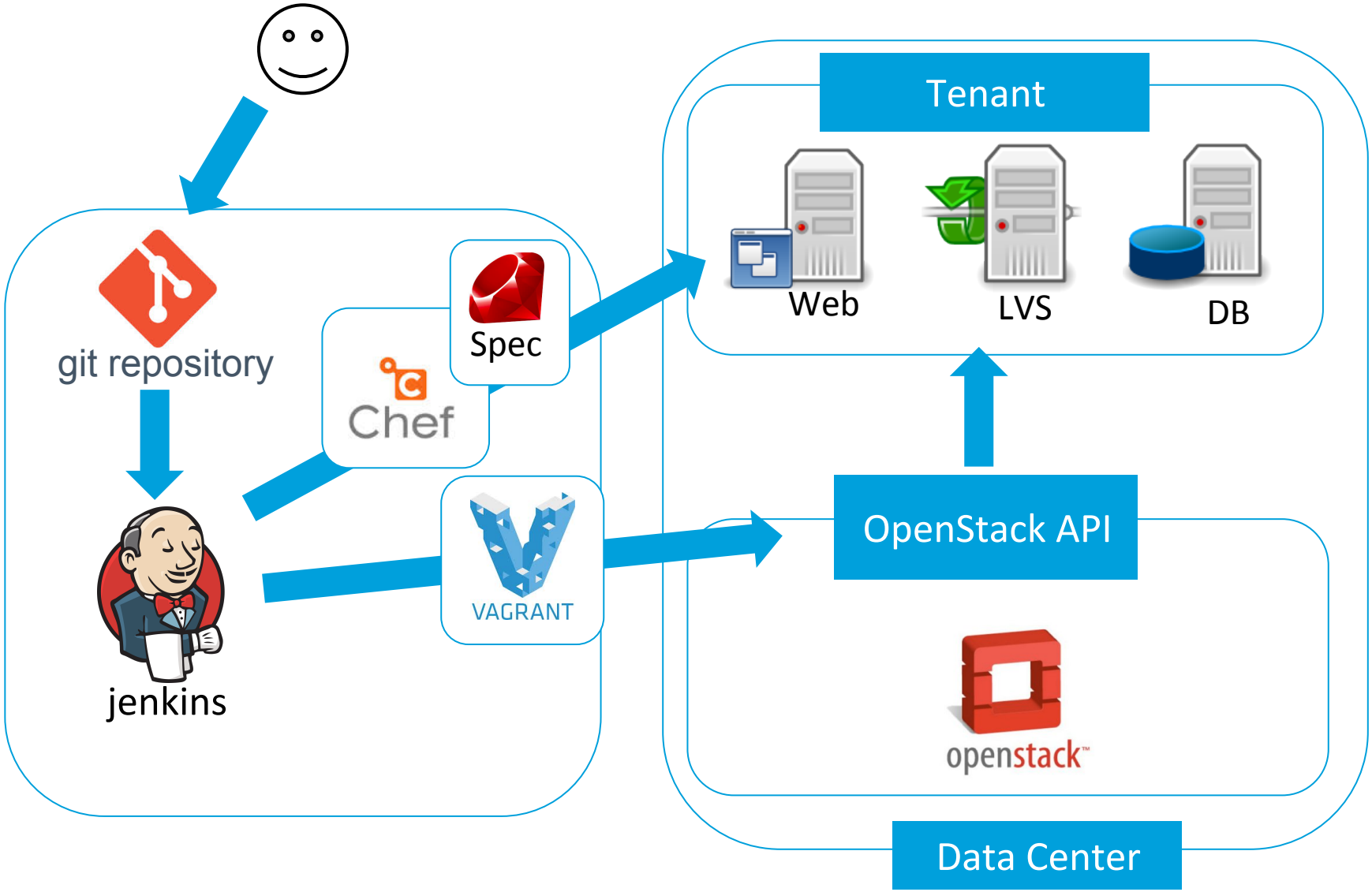
[Caution]

- Some of issues we experienced while testing OpenStack folsom have already been fixed in the latest release
- Solutions we introduced may not be the best ways (as you may know…)

Recent Work

GREE

OpenStack x Vagrant for Automated Integration Test

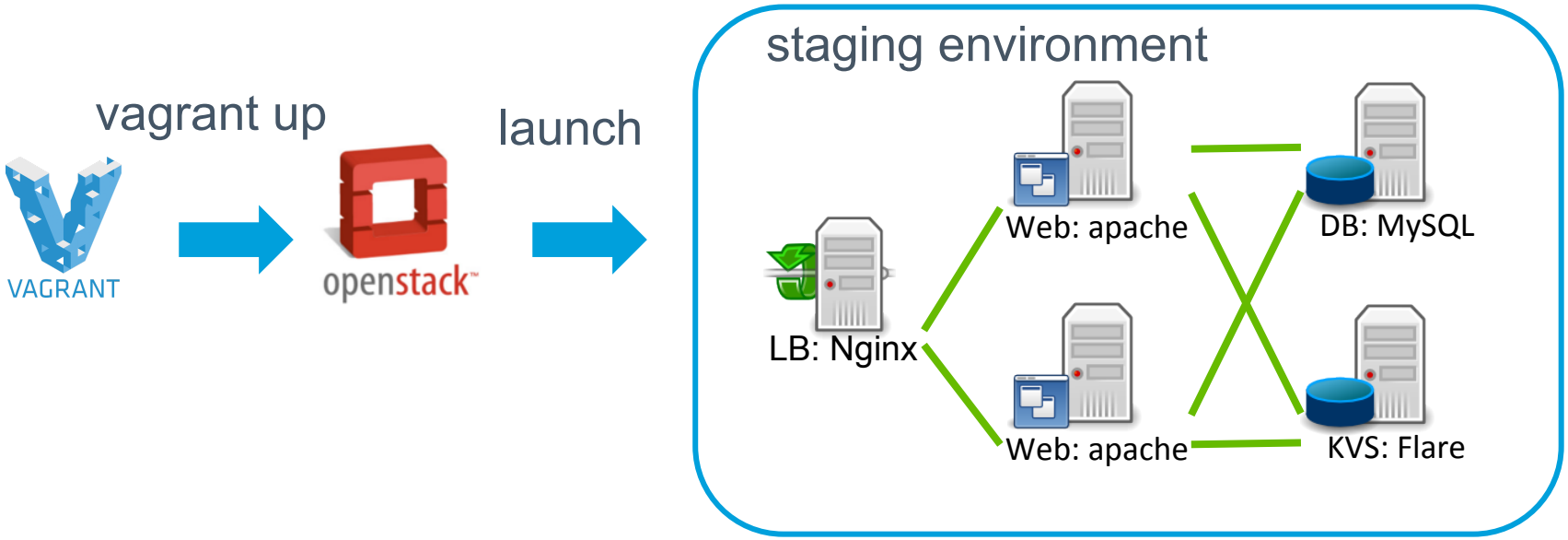


GREE

OpenStack x Vagrant for Automated Integration Test

- Automated Integration Test
  - Hook ➜ Delivery VMs ➜ Configuration ➜ Deploy Codes ➜ Test (➜ Destroy)

- Create staging environments instantly with latest codes



GREE

Step By Step…

- Improve UI/UX and tools
  - Adopt Domain Driven Development
- Networking
  - Edge-to-Edge
  - L2 Overlay
  - Taking a look into OpenStack L3 Agent...
- Linux Container
  - Docker
- Experiments in On-premise infrastructure
  - Log-Manageable immutable infrastructure
  - Blue-Green deployment

Conclusion

GREE

Impressions of OpenStack

- Should design on own workload and requirements
  - We have achieved High-Availability on application layer
    - No SPoF in all systems including VM's application service
    - Adopted local disk storage for VMs
- Test, test, and test...
  - Understanding how it works
  - Say "No" to extra features
    - Many new projects on OpenStack
    - Right person in the right place
    - Some features worked only in devstack...
- Loose coupling in each components, but tight coupling as OpenStack cluster
  - Upgrading OpenStack is a living hell

GREE

Expectation for OpenStack

- Free to choose, free to design
  - Respect for the culture as Open-Source
- Core infrastructure improvement
  - More than PaaS or other “as a Service”
  - Rolling Update
  - HA/FT
  - PCI Path Through
  - etc..
- Significant storage solution for Cinder