Practical Lessons from Building a Highly Available Openstack Private Cloud

Sebastian Kachel, Florian Haas

Fall 2013, OpenStack Summit Hong Kong, November 5 2013

hastexo!

pixelpark

Who are we?

Sebastian Kachel

pixelpark

IT-Cloud-Manager & IT-Operator

- Unix / Linux & SysAdmin / DevOps guy
- involved in OpenStack in 2012
- Based in Berlin, Germany
- One of the organizer from OpenStack-Meetup-Berlin

Florian Haas

hastexo!

CEO & Principal Consultant

- HA/Storage/Cloud guy, consultant, instructor
- hastexo co-founder & CEO
- Based in seat 10C, Economy Class
- Occasionally returns to home base near Vienna, Austria

What was our challenge to solve?

Ensure high availability for all services that we want use for our customers in a private cloud

What’s this about?

Pixelpark AG

Full service agency for multimedia communications & e-business solutions

departments: concepts, project management, editorial, design, development & hosting

Solution from pixelpark’s high availability private cloud

Why OpenStack?

Why OpenStack?

- Benefits of cloud computing like on-demand, scalable & elastic
- Fixed, time based release cycles
- Open source
- Support
- Rapid development
- Cloud software that goes beyond IaaS

make it **highly available**

Why High Availability?

We provide service level agreements with high availability up to 99,99%.

How did we do it?

Storage

HA Storage

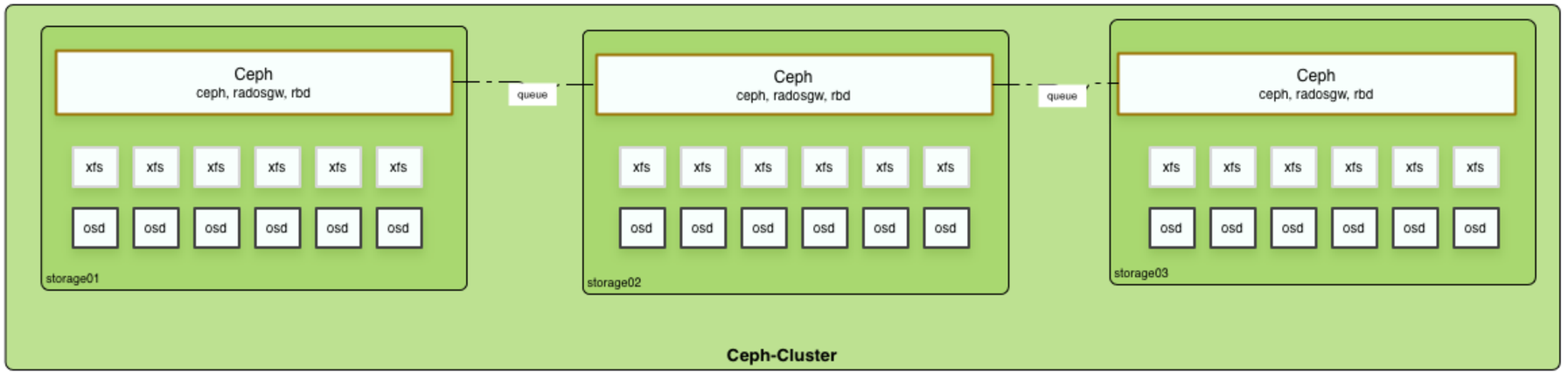
- Highly available storage as base data store to make it scale

Why Ceph ?

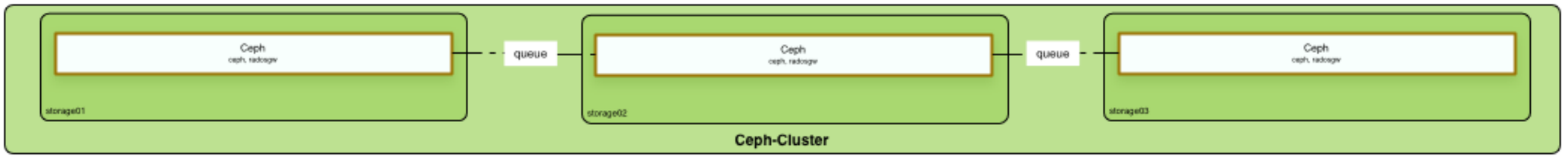
- Distributed storage platform designed to provide excellent performance, reliability and scalability
- Guarantees reliable storage with no data loss
- Stores: Cinder volumes, Glance images, static data (S3) over radosgw & instances

Ceph is an excellent, reliable basis for cloud storage

How did we build our Ceph store?



- Working with 3 copies
- 1 Disk per OSD
- XFS filesystem
- Journaling on separate SSD
- Every storage node has 8x Gbit ports in trunk mode

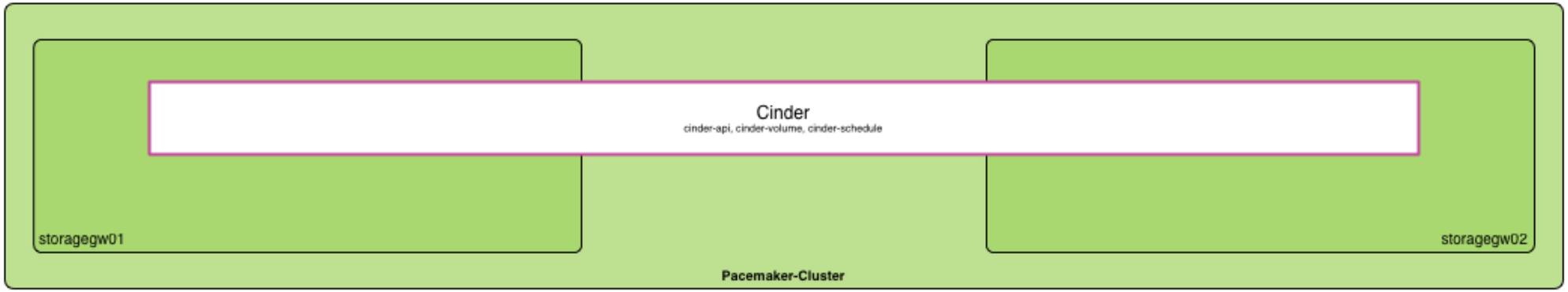

Ceph
ceph, radosgw
storage01
queue
Ceph
ceph, radosgw
storage02
queue
Ceph
ceph, radosgw
storage03
Ceph-Cluster

OpenStack Block Storage

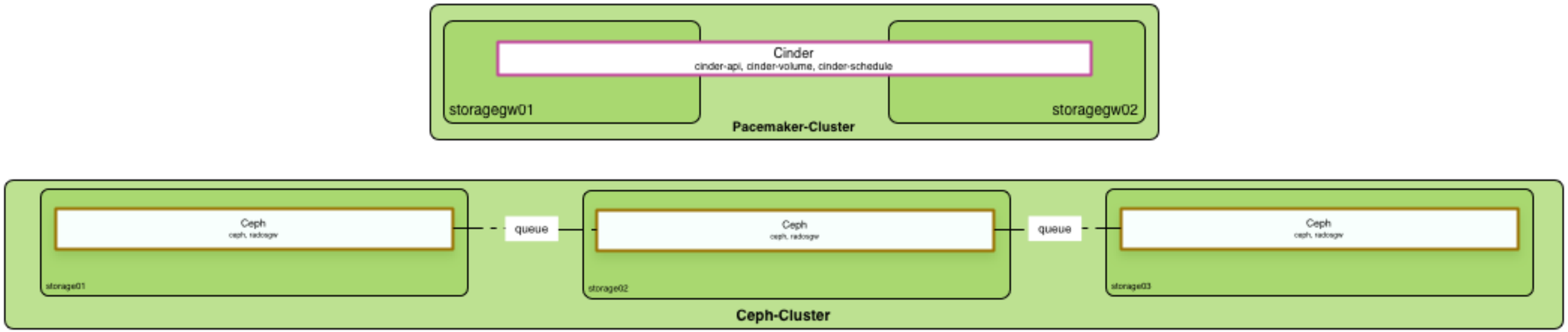
HA Block Storage

cinder-volume, cinder-api & cinder-schedule over two nodes

Pacemaker to monitor & control services



- cinder services in active/backup mode

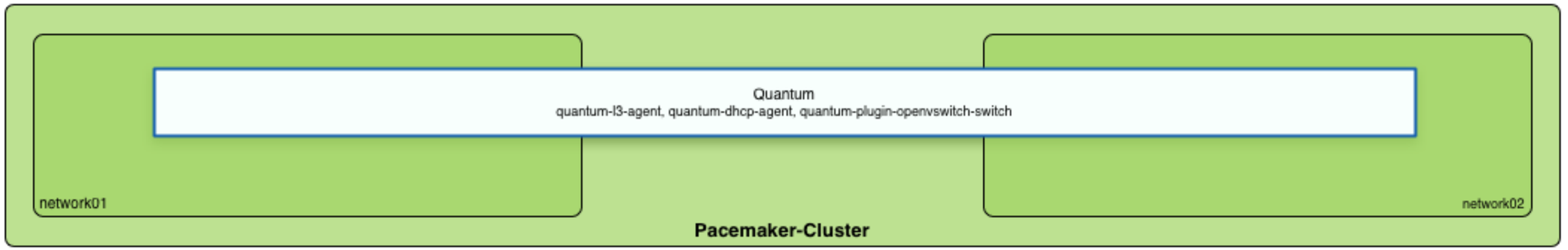

Cinder
cinder-api, cinder-volume, cinder-schedule
storagegw01
storagegw02
Pacemaker-Cluster
Ceph
ceph, radosgw
queue
Ceph
ceph, radosgw
queue
Ceph
ceph, radosgw
storage01
storage02
storage03
Ceph-Cluster

Network

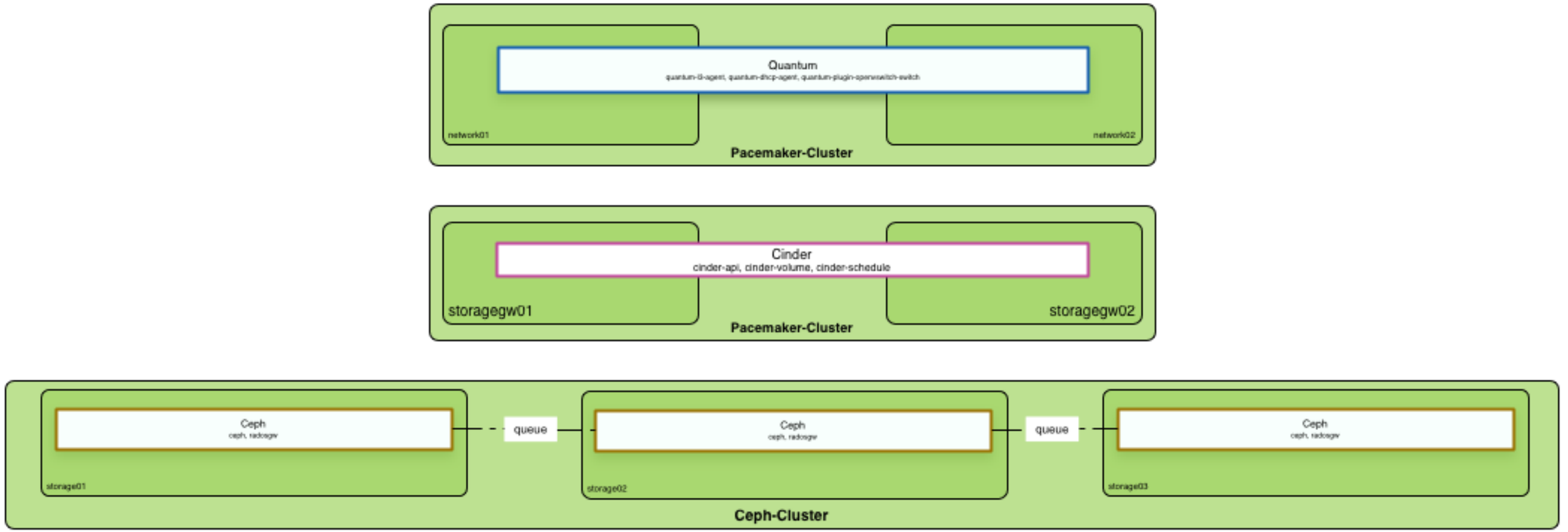
HA Network

quantum-dhcp-agent & quantum-l3-agent scalable over two nodes

Pacemaker to monitor & control network services

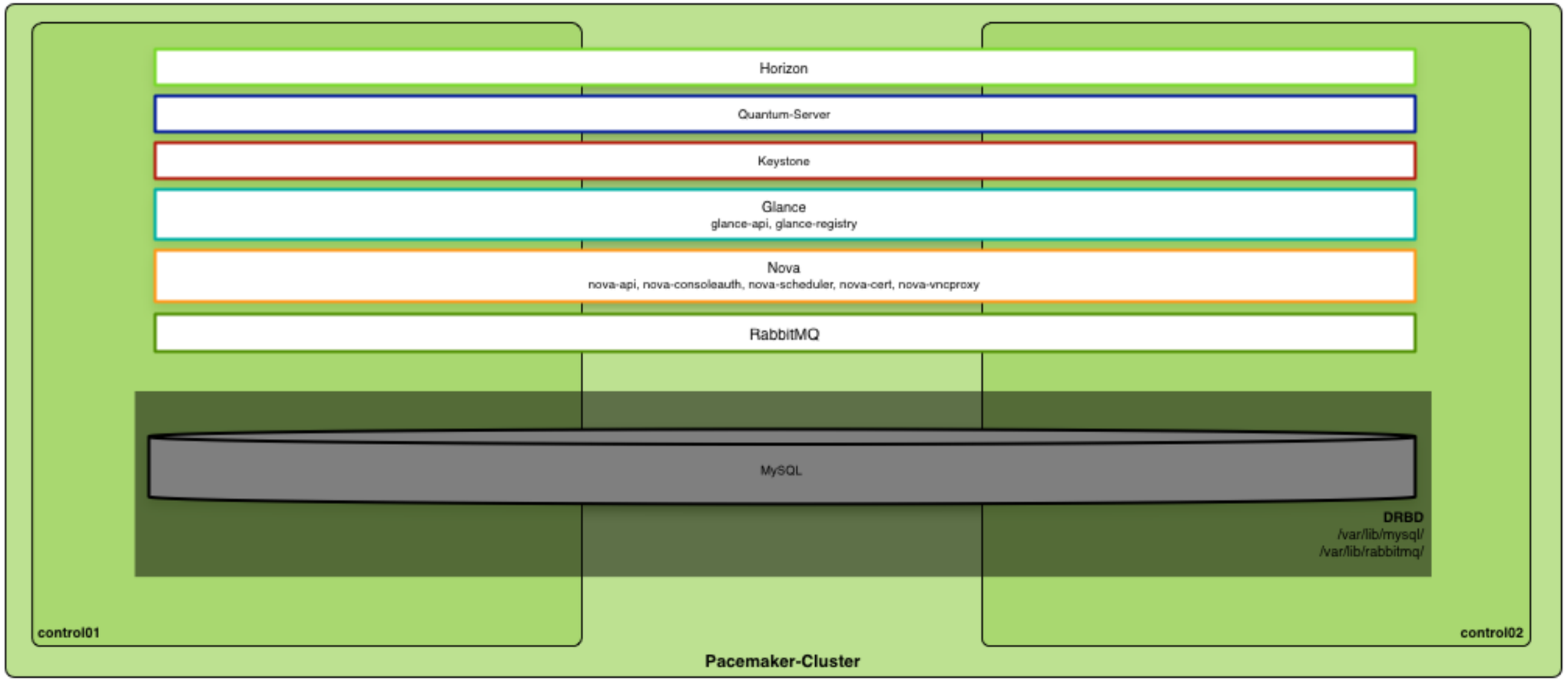


- quantum-dhcp-agents **active/active** over two nodes
- quantum-l3-agent **active/backup** distributed over two nodes
- quantum-plugin-openvswitch-agent **active/active** over two nodes


Quantum
quantum-l3-agent, quantum-dhcp-agent, quantum-plugin-openvswitch-switch
network01
network02
Pacemaker-Cluster
Cinder
cinder-api, cinder-volume, cinder-schedule
storagegw01
storagegw02
Pacemaker-Cluster
Ceph
ceph, radosgw
queue
Ceph
ceph, radosgw
queue
Ceph
ceph, radosgw
storage01
storage02
storage03
Ceph-Cluster

Services & APIs

HA OpenStack services and APIs

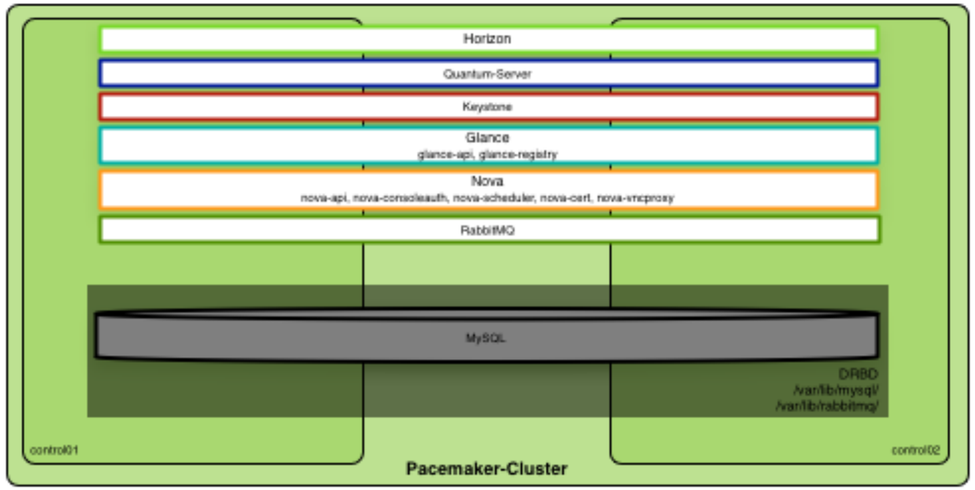


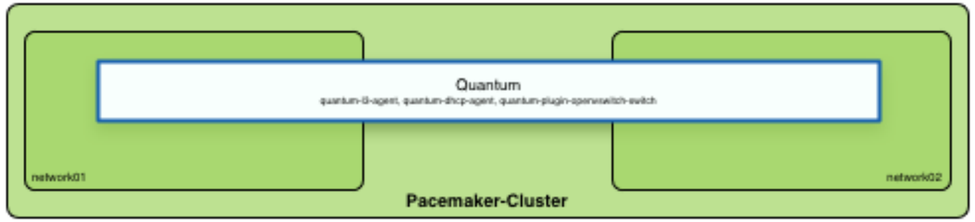
HA OpenStack services and APIs

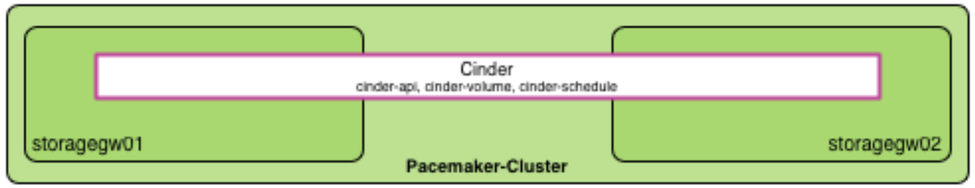
Pacemaker cluster with two controller nodes to keep

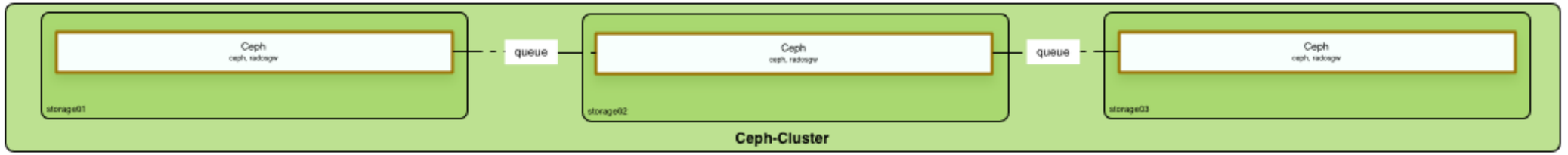
horizon,
keystone,
glance,
nova,
RabbitMQ,
quantum-server & mysql database

always on


Horizon
Quantum-Server
Keystone
Glance
glance-api, glance-registry
Nova
nova-api, nova-consoleauth, nova-scheduler, nova-cert, nova-vncproxy
RabbitMQ
MySQL
DRBD
/var/lib/mysql/
/var/lib/rabbitmq/
control01
control02
Pacemaker-Cluster

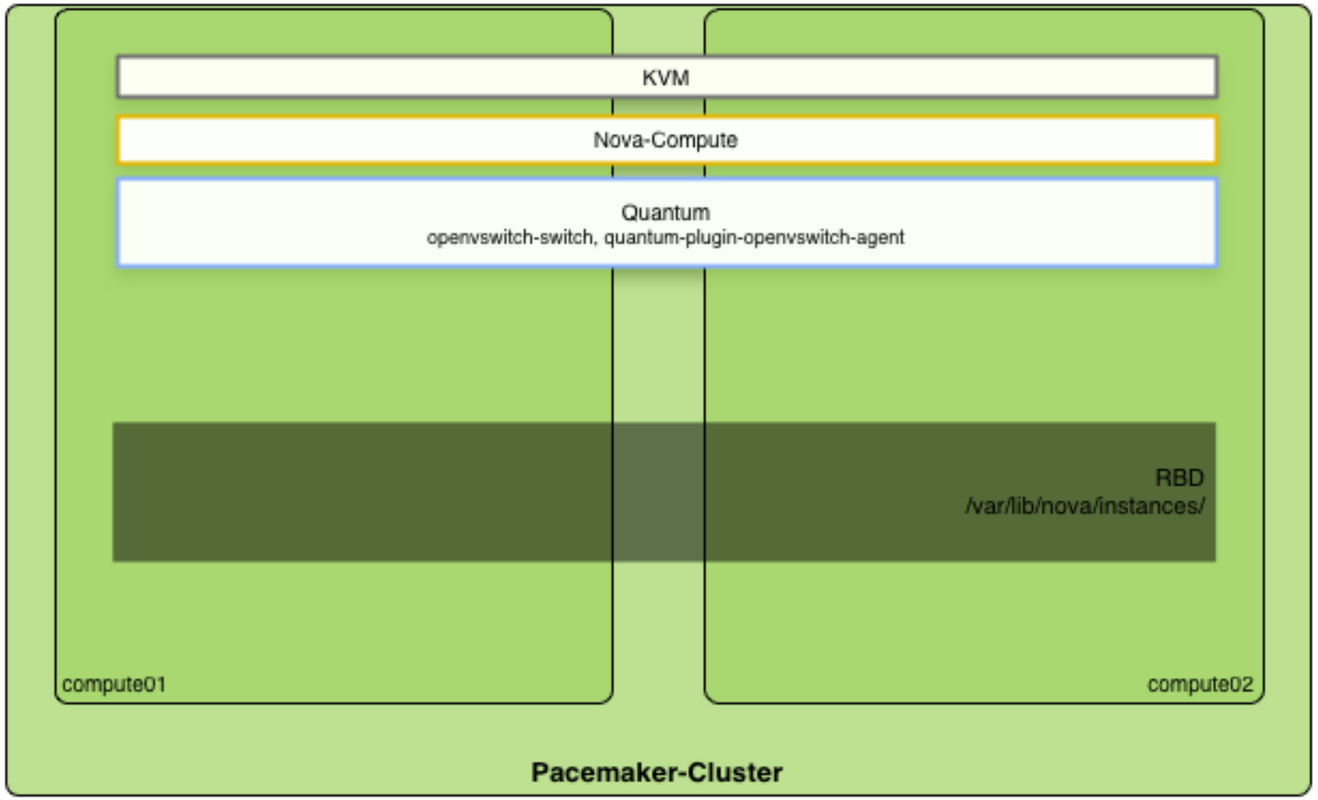

Quantum
quantum-l3-agent, quantum-dhcp-agent, quantum-plugin-openvswitch-switch
network01
network02
Pacemaker-Cluster


Cinder
cinder-api, cinder-volume, cinder-schedule
storagegw01
storagegw02
Pacemaker-Cluster

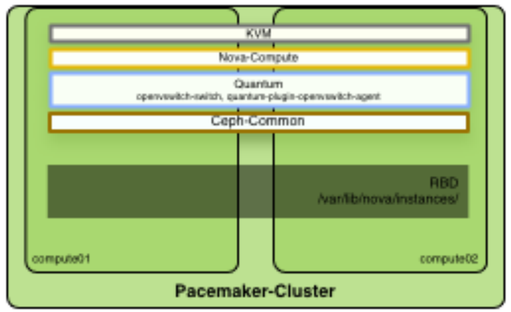

Ceph
ceph, radosgw
queue
Ceph
ceph, radosgw
queue
Ceph
ceph, radosgw
storage01
storage02
storage03
Ceph-Cluster

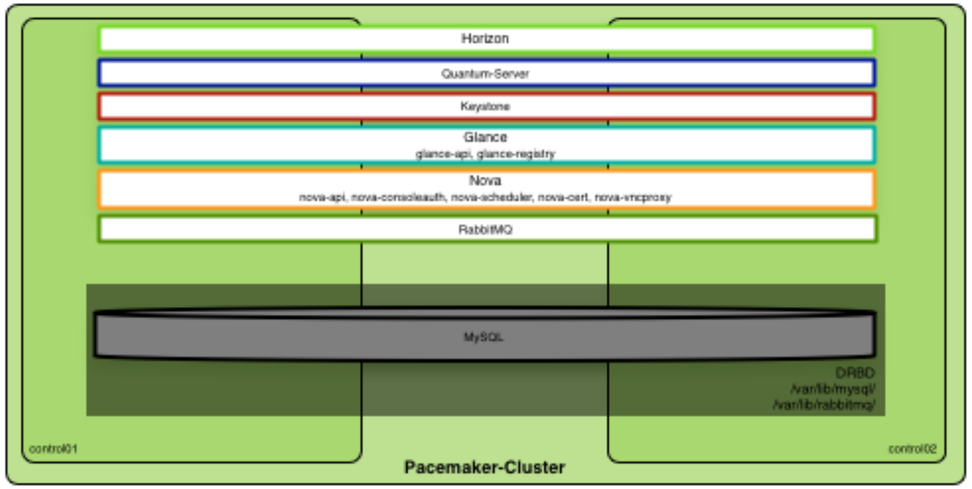
Compute

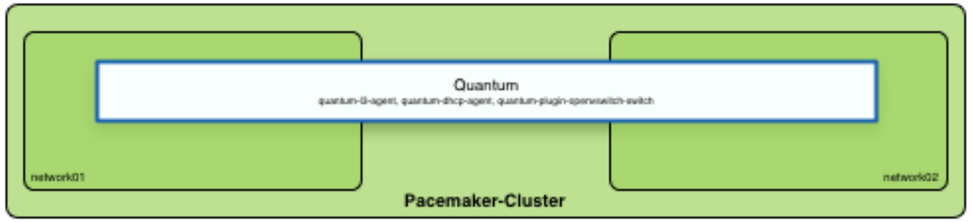
HA Nova guests

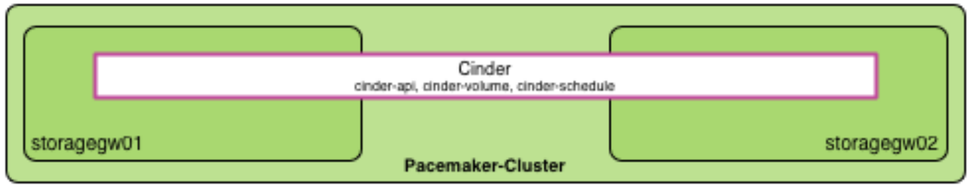


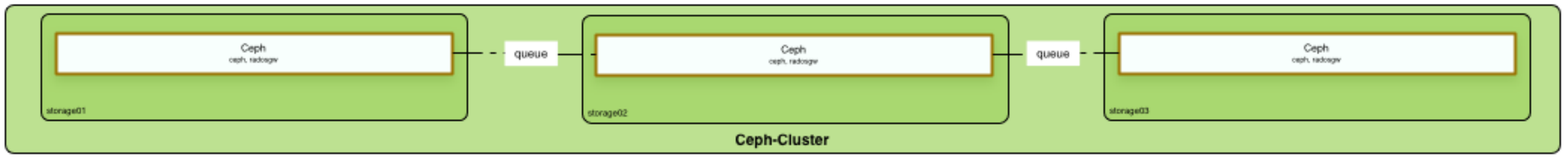
- rbd mount under /var/lib/nova/instances
- own pool in ceph-cluster
- 6 x 1GBit/s Ports on every compute


KVM
Nova-Compute
Quantum
openvswitch-switch, quantum-plugin-openvswitch-agent
Ceph-Common
RBD
/var/lib/nova/instances/
compute01
compute02
Pacemaker-Cluster


Horizon
Quantum-Server
Keystone
Glance
glance-api, glance-registry
Nova
nova-api, nova-consoleauth, nova-scheduler, nova-cert, nova-vncproxy
RabbitMQ
MySQL
DRBD
/var/lib/mysql/
/var/lib/rabbitmq/
control01
control02
Pacemaker-Cluster


Quantum
quantum-l3-agent, quantum-dhcp-agent, quantum-plugin-openvswitch-switch
network01
network02
Pacemaker-Cluster


Cinder
cinder-api, cinder-volume, cinder-schedule
storagegw01
storagegw02
Pacemaker-Cluster


Ceph
ceph, radosgw
queue
Ceph
ceph, radosgw
queue
Ceph
ceph, radosgw
storage01
storage02
storage03
Ceph-Cluster

How did this affect our organization?

Lessons learned

- Implementing an OpenStack environment is a **challenge**
- Getting **training and support** is a good idea
- Ensure **quality**, work **efficiently**
- **programmable infrastructure**
- Create a basis for **further innovation**
- Be ready for up & coming **technology**
- Sponsor an **OpenStack Meetup group**

Have fun!

Get in touch!

Sebastian Kachel
Pixelpark AG
www.pixelpark.com
sebastian.kachel@pixelpark.com

Florian Haas
hastexo
www.hastexo.com
florian.haas@hastexo.com


Thank you !
hastexo!
pixelpark