Scaling Xen within Rackspace Cloud Servers

Chris Behrens
chris.behrens@rackspace.com

XenSummit 2012, San Diego, CA USA
August 28, 2012

rackspace®
the open cloud company

Agenda

- Intro to Rackspace Cloud Servers
- Why Xen?
- Xen to XenServer
- OpenStack Basics
- OpenStack at Rackspace
- Scaling OpenStack Deployment
- Future

Intro to Cloud Servers

Intro to Cloud Servers

Overview

- Public Cloud product under Rackspace's Open Cloud portfolio
- First Gen
  - Code originated from Slicehost acquisition
  - Initially Xen 3.x, but migrated to XenServer
- Next Gen
  - Public Launch 8/1/2012
  - Powered by OpenStack
- XenServer 6
- Linux, Windows, and FreeBSD VMs

Intro to Cloud Servers

Statistics

- 180,000+ total Rackspace customers (not just Cloud Servers)
- Tens of thousands of hosts
- Hundreds of thousands of VMs
- Millions of snapshots

Why Xen?

Why Xen?

- Thin hypervisor layer
- Open source
- Excellent driver support due to Linux dom0
- Good API
- Good performance
- Along with Linux, Windows VMs are a must
  - Same virtualization technology desired for Windows
  - Microsoft will support
  - Good PV driver support

Xen to XenServer

Xen to XenServer

Cloud Servers use of Xen

- LVM
  - Logical volumes per VM
- Base images stored as tar files
- Kernel and ramdisk outside of images
- Shell out to run ‘xm’ commands
- Snapshots w/ LVM
  - Mounted and contents rsynced to another host
- No agent
  - Modifications inside images done via mounting powered off VM

Xen to XenServer

Xen to XenServer changes

- VHDs for virtual disk images
  - Tarred and gzipped
- Kernel and ramdisk inside images
- Snapshots via XenAPI
  - Whole VDI chain wrapped in OVF container for upload to glance
- Agent inside VM images
  - Communication via Xenstore
  - File injection, password resets, networking setup

Xen to XenServer

Migrating

- Pick a XenServer host
- Create VDI
- Attach VDI to dom0
- Partition, mkfs, and mount the new virtual disk
- LVM Snapshot original Xen VM and mount it
- rsync snapshot contents to mounted VDI on new host
- Power off original Xen VM and mount its virtual disk
- rsync again to new host to catch changes since snapshotting
- Extract a tar file into mounted VDI containing kernel, ramdisk, agent
- Unmount and detach VDI from dom0
- Create new VM record and attach VDI
- Start the new VM

OpenStack

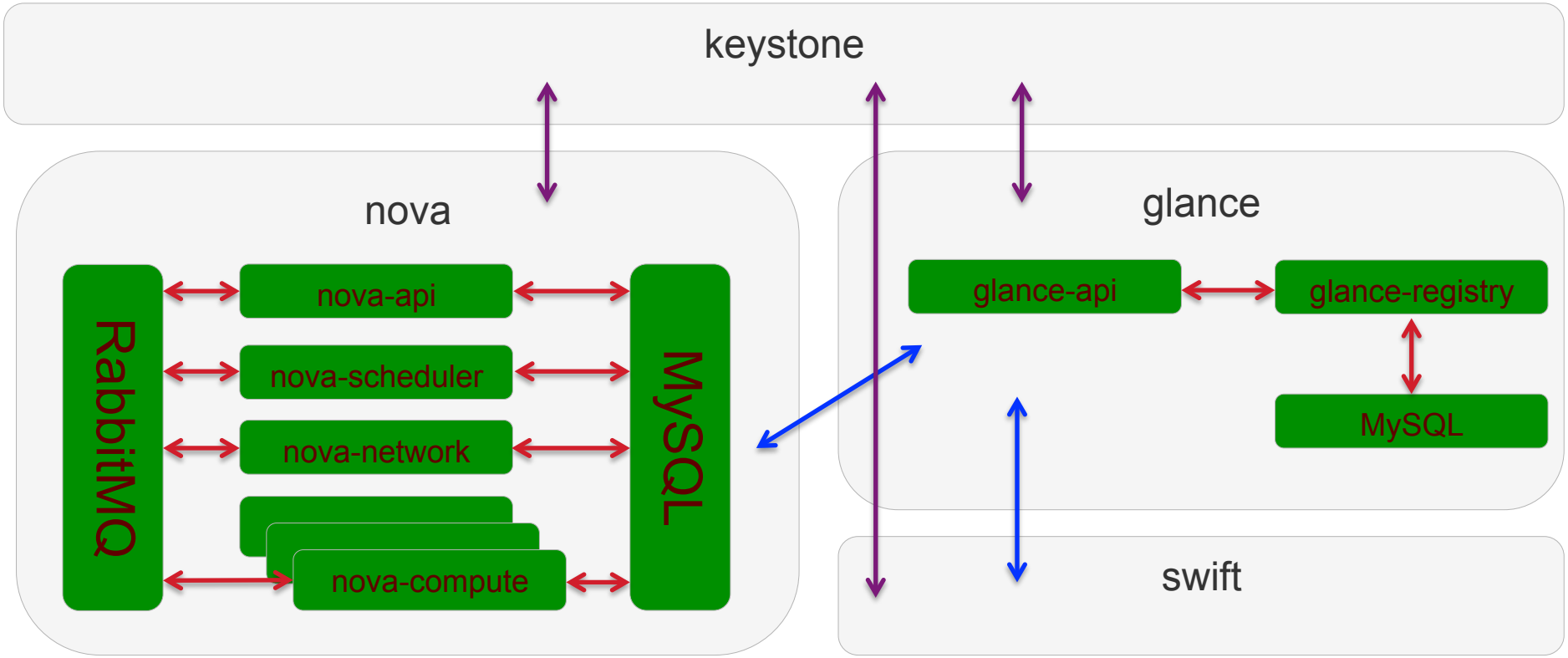
The Basics

OpenStack
Components

- Nova
  - Compute controller with various hypervisor drivers
- Glance
  - VM Image registry
- Swift
  - Object storage
- Quantum/Melange
  - Networking
- Cinder
  - Volumes/Block storage
- Keystone
  - Authentication
  - Authorization
  - API Endpoint discovery

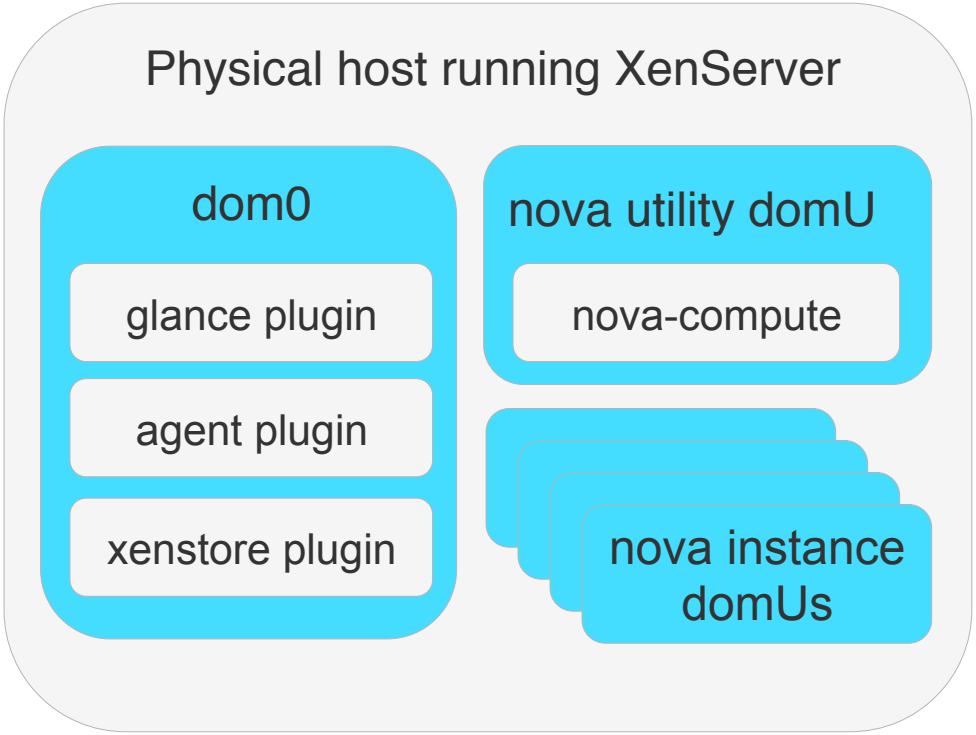
OpenStack

Typical OpenStack Deployment



OpenStack
nova-compute w/ XenAPI

- Runs in a utility domU on each host
- Images downloaded in dom0
  - glance XenAPI plugin
- Images attached to nova-compute domU
  - Partitioning changes
  - file-system resizing
- Monitors VMs' power_state
- Snapshots/backups
- VM resize/migration
  - rsync to new host
- Communicates w/ agent inside VMs
  - Via xenstore



OpenStack
XenAPI VM Calls

- VM.start / VM.start_on
- VM.destroy
- VM.clean_reboot
- VM.hard_reboot
- VM.clean_shutdown
- VM.hard_shutdown
- VM.pause
- VM.unpause
- VM.suspend
- VM.resume
- VM.snapshot
- VM.add_to_VCPUs_params
- VM.get_VBDs
- VM.get_record
- VM.get_by_uuid
- VM.get_by_name_label
- VM.set_name_label
- VM.add_to_xenstore_data
- VM.remove_from_xenstore_data
- VM.assert_can_migrate
- VM.migrate_send
- VM.pool_migrate

OpenStack

Rackspace Specifics

OpenStack Rackspace Specifics
Source Code

- Rackspace tracks trunk
  - Deployed code generally less than 2 weeks behind trunk
- Some custom patches on top of trunk
  - Features specific to Rackspace
  - Custom scheduling to meet Rackspace needs
  - Scaling with OpenStack Compute Cells

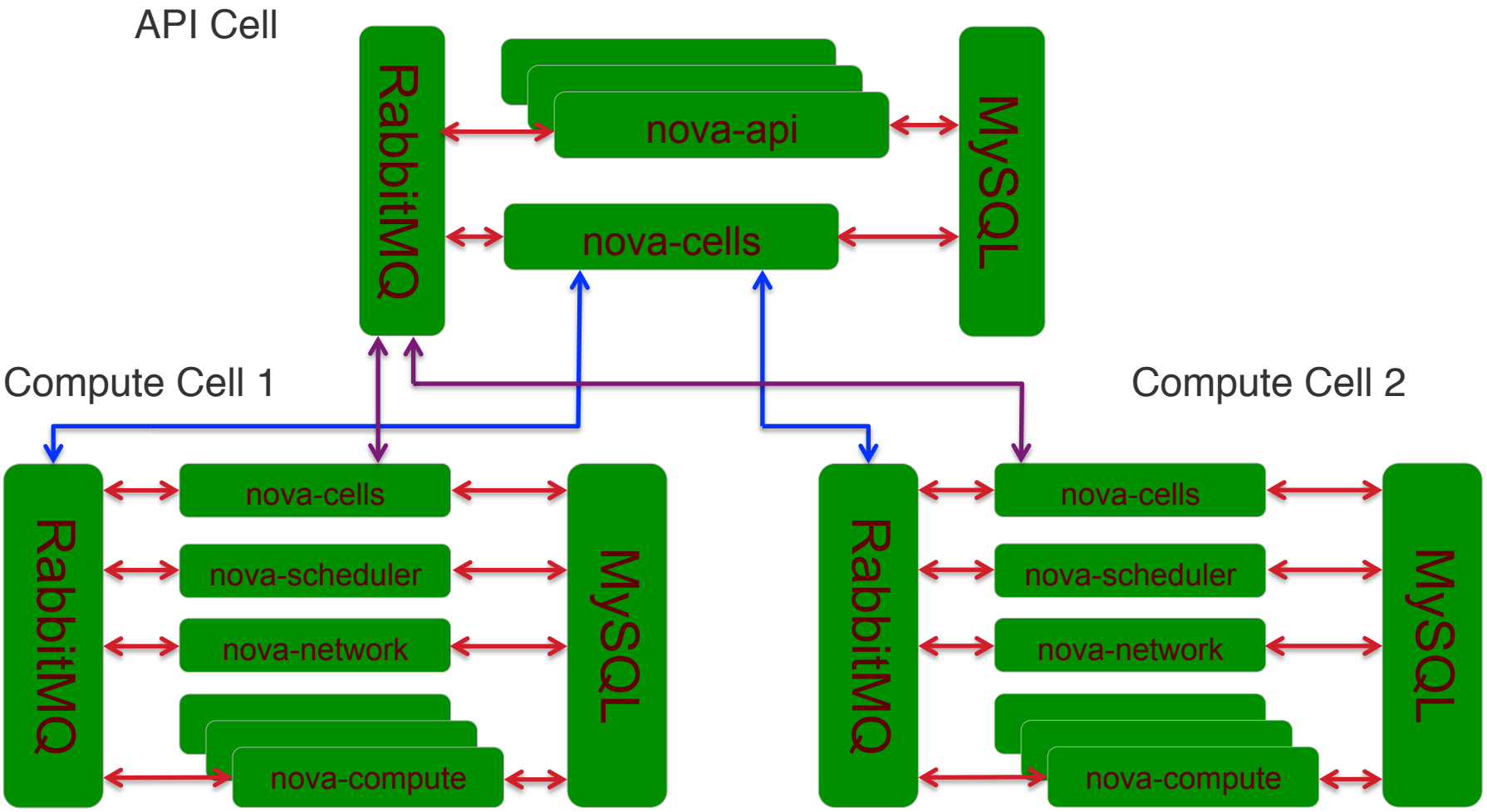
OpenStack Rackspace Specifics

OpenStack Compute Cells

- Scaling beyond small nova deployments
  - Effectively ties multiple nova deployments together
- Developed by Rackspace
- Targeted for Grizzly OpenStack release
- Hierarchical tree
  - Multiple parents supported
- API cell(s) at top of the tree
- Compute cell(s) underneath
- DB and RabbitMQ per cell
- Scalability
  - Splits connections to DB and RabbitMQ
  - Less work to schedule new VM placement

OpenStack Rackspace Specifics

OpenStack Nova Multi-cell Diagram

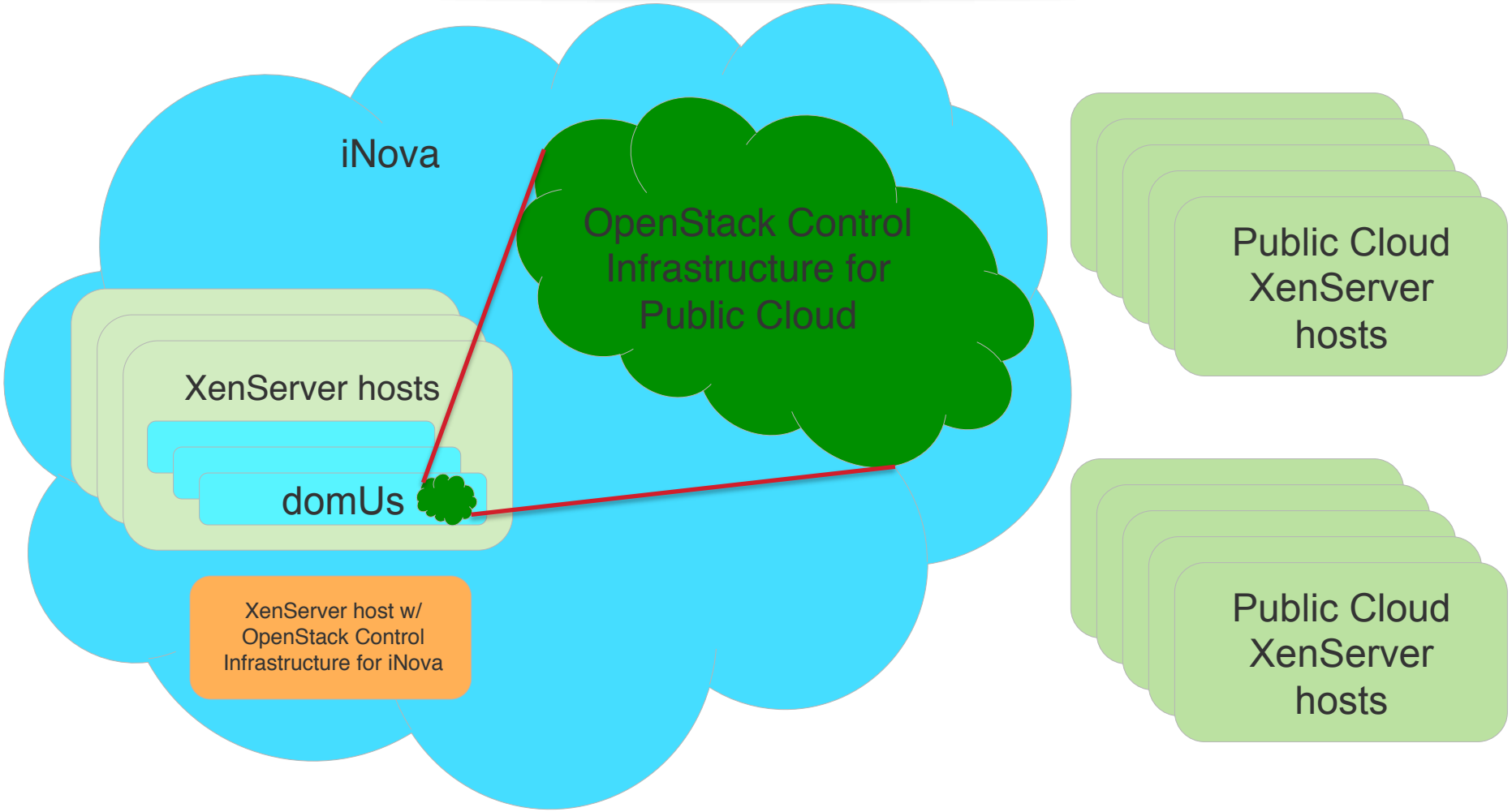


Deploying OpenStack

Deploying OpenStack
'Inception'

- Private internal cloud
  - Based on Openstack, called iNova
  - OpenStack services for public cloud run virtualized under iNova
- Easy to spin up additional services for load balancing
  - Enables quick reaction to load spikes
- Easier deployments of new compute cells
- Automated reaction to downed hosts
- Enables new strategies for deployments
  - Replace vs Upgrade

Deploying OpenStack
iNova Diagram



Deploying OpenStack
Continuous Deployment

- Trunk and custom branches merged multiple times daily
  - Unit Tested and Packaged
- Configuration managed with puppet
  - Follows same QE and Continuous Deployment rules as code
- QE Environment
  - OpenStack control infrastructure deployed on iNova
  - nova-compute upgraded on 200 hosts
  - Smoke tests
- Staging Deploy
  - Uses the same packages and puppet manifests

Future

Future

Improvements desired from Xen

- Better VM resource isolation
  - I/O scheduling weights per virtual disk
- Built in live migration without shared storage
- More information about VM resource usage
  - Disk usage
  - Help us improve OpenStack code
  - Help customers react to bad situations
- XenAPI call to write to live xenstore
- Improve XenServer upgrade experience

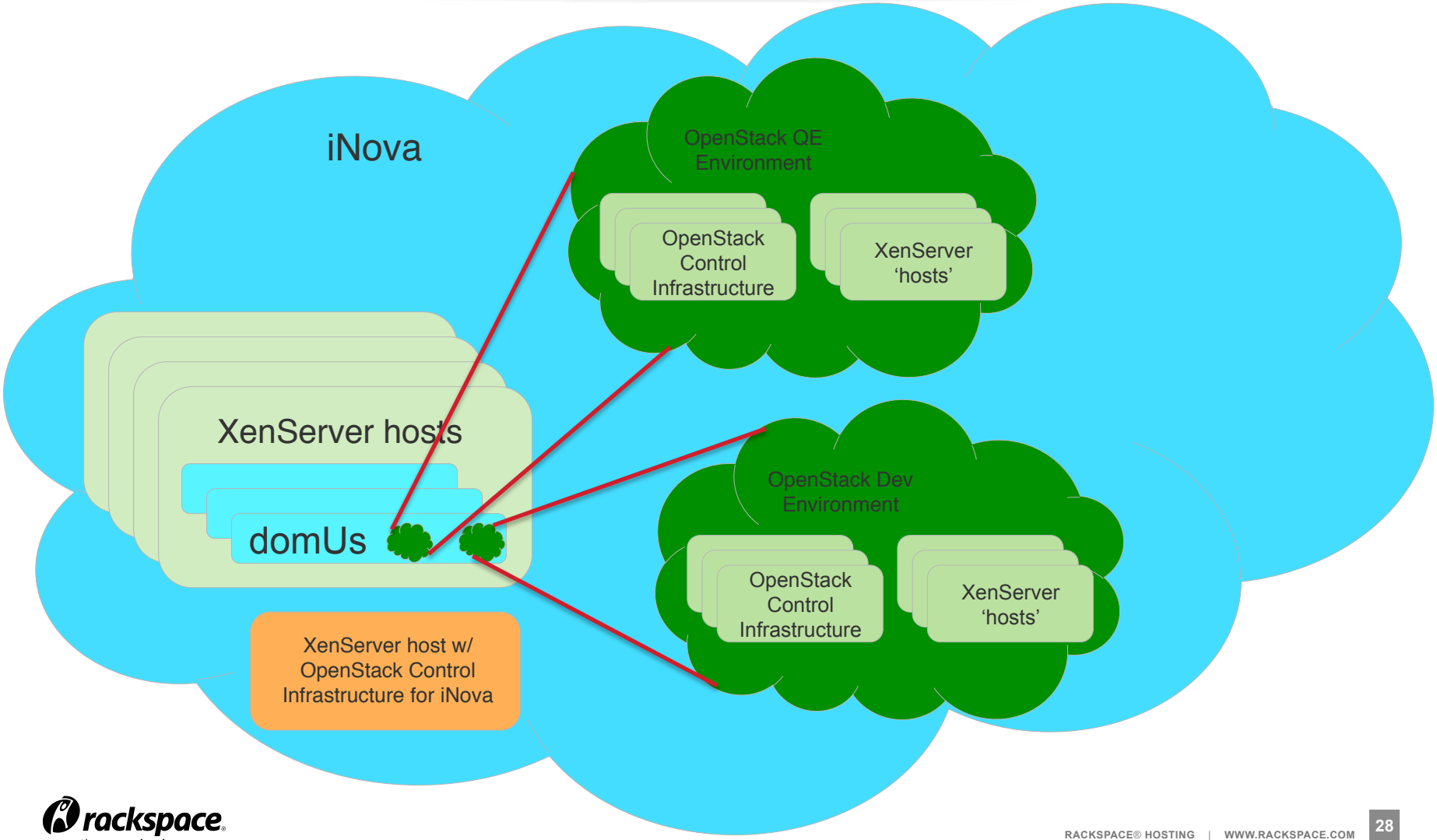
Future

Rackspace's Other Interests

- Better VDI introduction for new VMs
  - Look at VDI.introduce vs mv'ing VDIs and doing an sr-scan
- Incremental backups
  - Delta VHDs instead of all VHDs
- Updated plugins using python >= 2.6
  - Cleans up workarounds for no built-in 'uuid' module
  - Allows use of newer python features like context managers
- Better continuous deployment and testing
  - Virtualize all of OpenStack
    - Requires spinning XenServer on XenServer
    - Wipe out compute nodes for QE environment
    - Enable virtual labs for OpenStack developers
  - Puppet used to configure XenServer

Future

iNova with Dev/QE Environments



THANK YOU

Chris Behrens
chris.behrens@rackspace.com

RACKSPACE® HOSTING | 5000 WALZEM ROAD | SAN ANTONIO, TX 78218
US SALES: 1-800-961-2888 | **US SUPPORT:** 1-800-961-4454 | WWW.**RACKSPACE**.COM