Rocks Virtual Clusters, Extended clusters in to Amazon EC2 w/Condor

Philip Papadopoulos, Ph.D
University of California, San Diego
San Diego Supercomputer Center
California Institute for Telecommunications and Information Technology (Calit2)

UC Cloud Summit, UCLA, 19 APR 2011

© UC Regents 2010

Outline

- Short Background on Rocks
- Virtual Clusters
- Practically extending a local cluster using “Hardware” in EC2 and Condor

© UC Regents 2010

Rocks – http:// www.rocksclusters.org

- Technology transfer of commodity clustering to application scientists
- Rocks is a cluster/System Configuration on a CD
  - Clustering software (PBS, SGE, Ganglia, Condor, …)
  - Highly programmatic software configuration management
  - Put CDs in Raw Hardware, Drink Coffee, Have Cluster.
- Extensible using “Rolls”
- Large user community
  - Over 1PFlop of known clusters
  - Active user / support list of 2000+ users
- Active Development
  - 2 software releases per year
  - Code Development at SDSC
  - Other Developers (UCSD, Univ of Tromso, External Rolls
- Supports Redhat Linux, Scientific Linux, Centos and Solaris
- Can build Real, Virtual, and Hybrid Combinations (2 – 1000s)

Rocks Core Development NSF award #OCI-0721623

© UC Regents 2010

A Mid-Sized Cluster Resource

Includes : Computing, Database, Storage, Virtual Clusters, Login, Management Appliances

TRITON RESOURCE

Large Memory PSDAF
- 256 GB & 512 GB Nodes (32 core)
- 8TB Total
- 128 GB/sec
- ~ 9TF

Shared Resource Cluster
- 16 GB/Node
- 4 - 8TB Total
- 256 GB/sec
- ~ 20 TF

ROCKS

Large Scale Storage (Delivery by Mid May)
- 2 PB (384 TB Today)
- ~60 GB/sec (7 GB/s)
- ~ 2600 (384 Disks Now)

UCSD Research Labs

Campus Research Network

http://tritonresource.sdsc.edu

© UC Regents 2010

The “Downloadable” Data Center

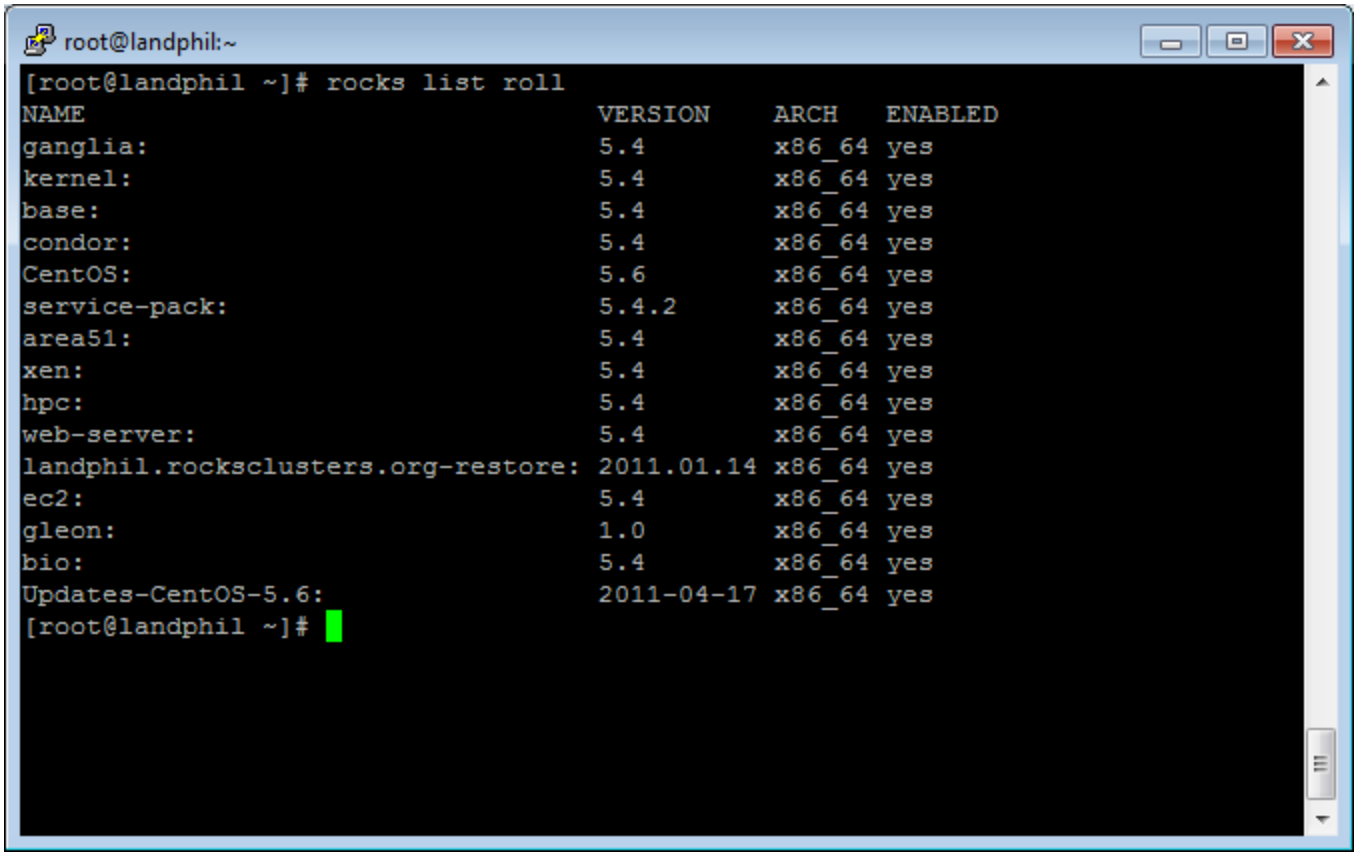
© UC Regents 2010

Key Rocks Concepts

- Define components of clusters as **Logical Appliances** (Compute, Web, Mgmt, Login DB, PFS Metadata, PFS Data, …)
  - Share common configuration among appliances
  - Graph decomposition of the full cluster SW and Config
  - Rolls are the building blocks: reusable components (Package + Config + Subgraph)
- Use installer's (Redhat Anaconda, Solaris Jumpstart) **text format to describe** an appliance configuration
  - Walk the Rocks graph to compile this definition
- Heterogeneous Hardware (Real and Virtual HW) with no additional effort

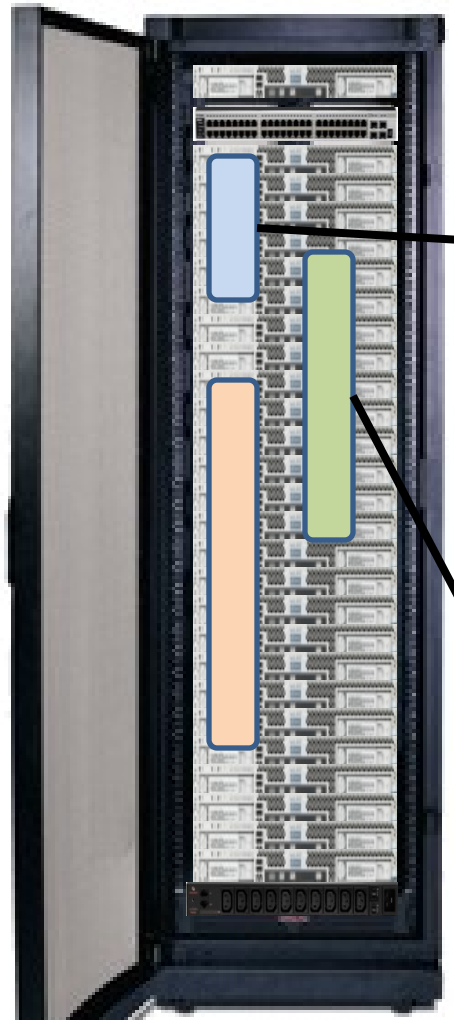
© UC Regents 2010

What's in YOUR cluster?

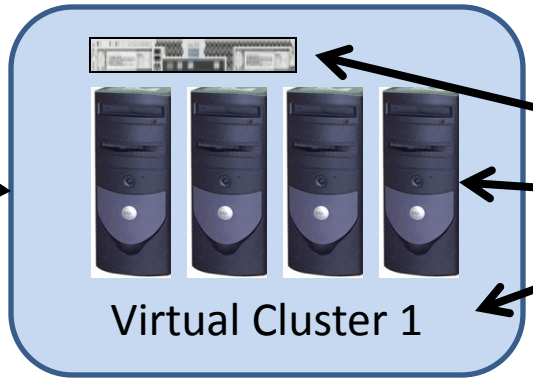


© UC Regents 2010

Virtual Clusters in Rocks Today



Physical Hosting Cluster
"Cloud Provider"



Require:
1. Virtual Frontend
2. Nodes w/disk
3. Private Network
4. Power

Virtual Clusters:
- May overlap one another on physical HW
- Need network isolation
- May be larger or smaller than physical hosting cluster

A single rocks command can allocate a Virtual Cluster

© U

How Rocks Treats Virtual Hardware

- **It's just another piece of HW**.
  - If RedHat supports it, so does Rocks
- Allows mixture of real and virtual hardware in the same cluster
  - Because Rocks supports heterogeneous HW clusters
- Re-use of all of the software configuration mechanics
  - E.g., a compute appliance is compute appliance, regardless of "Hardware"

Virtual HW must meet minimum HW Specs

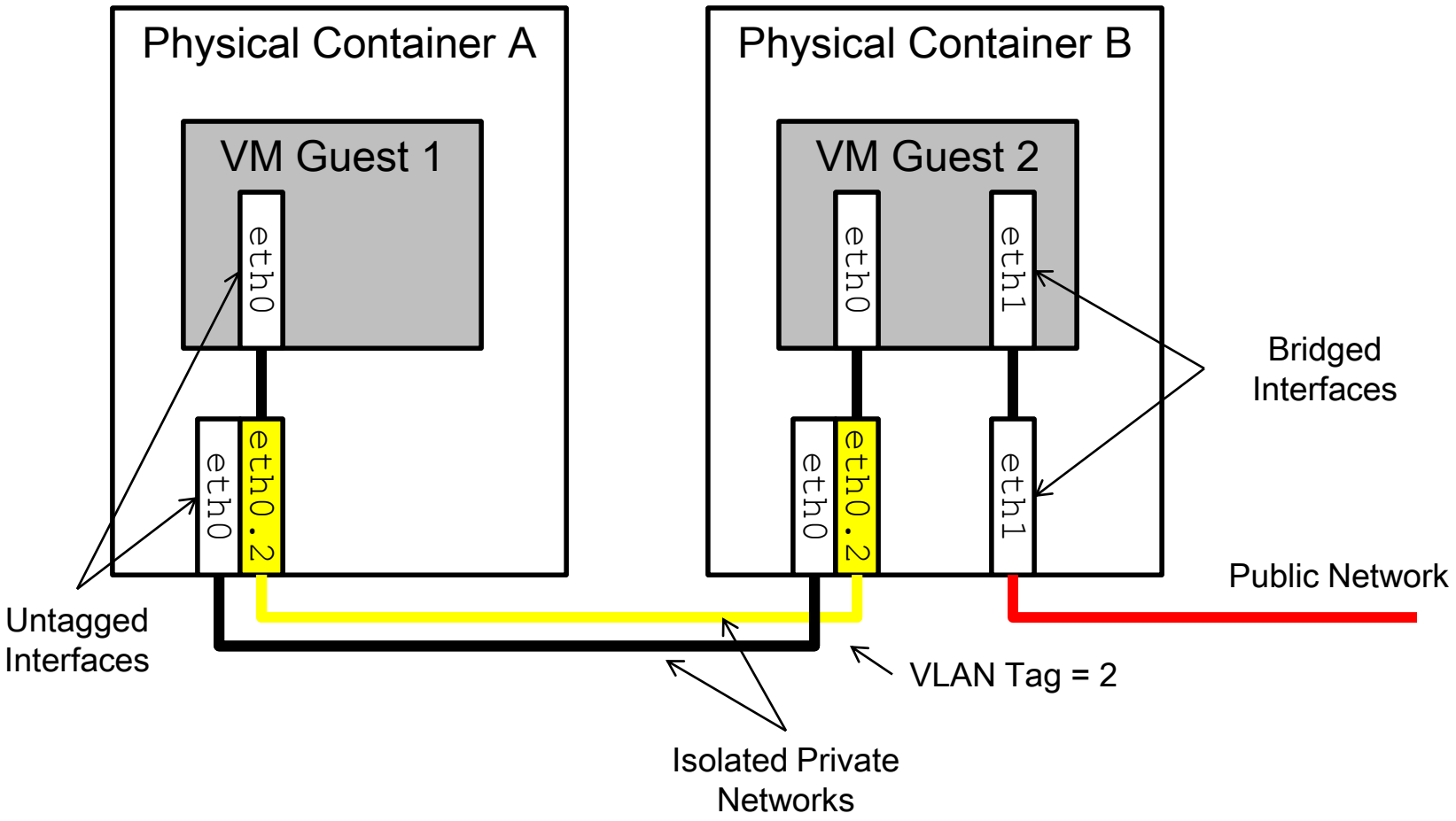
- 1GB memory
- 36GB Disk space*
- Private-network Ethernet
- + Public Network on Frontend

* Not strict – EC2 images are 10GB

© UC Regents 2010

Rocks Xen Roll Supports a Broad Spectrum of (Virtual) Networks

Bridged, VLAN Isolated, Multiple Interfaces/Node (Real or Virtual)



© UC Regents 2010

Rocks and EC2

We can build physical hosting clusters ,multiple isolated virtual clusters, and hybrid mixtures of virtual and physical nodes:

- Can I use Rocks to author “images” compatible with EC2? (We use Xen, They use Xen)
- Can I automatically integrate EC2 Virtual Machines into my local cluster (cluster extension)
  - Submit locally
  - My own private cloud

© UC Regents 2010

Some Challenges in EC2

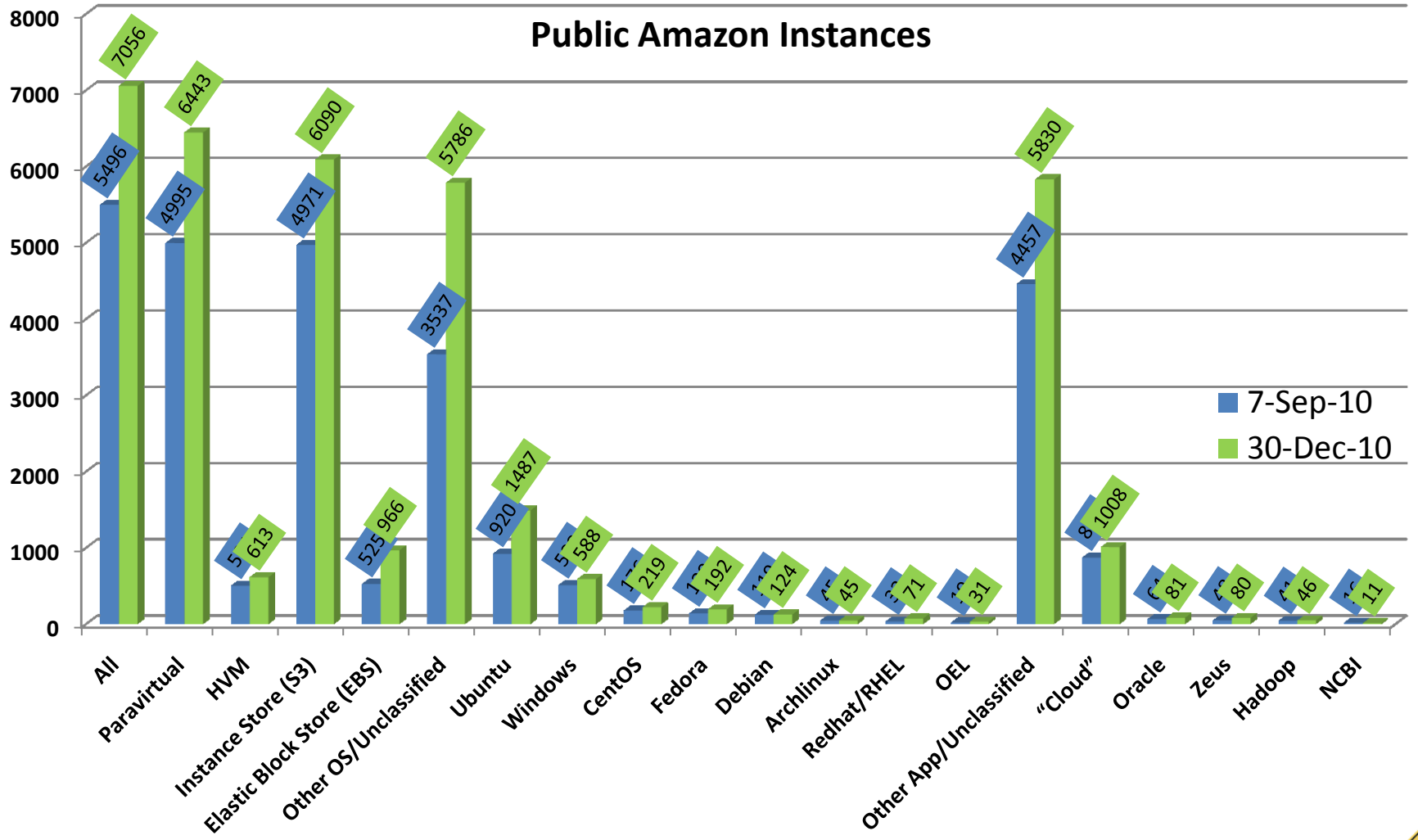
1. Defining the contents of your Virtual Machine (Software Stack) (Contents of AMI)
2. Understanding limitations and execution model
3. Debugging when something goes wrong
4. Remembering to turn off your VM
   - Smallest 64-bit VM is ~$250/month running 7x24

© UC Regents 2010

Why do we even care how a (IAAS) Cloud image is made?

© UC Regents 2010

A: Too MANY pre-existing AMIs. No Systematic (Scientific) Reproducibility



© UC Regents 2010

Condor Roll

- Condor 7.4.4
- Integration with Rocks command line to do basic Condor configuration customization
- To build a Condor Cluster with Rocks
  - Base, OS, Kernel, Condor Roll
  - Gives you local collector, scheduler
- Basic, Working Configuration that can be customized as required.

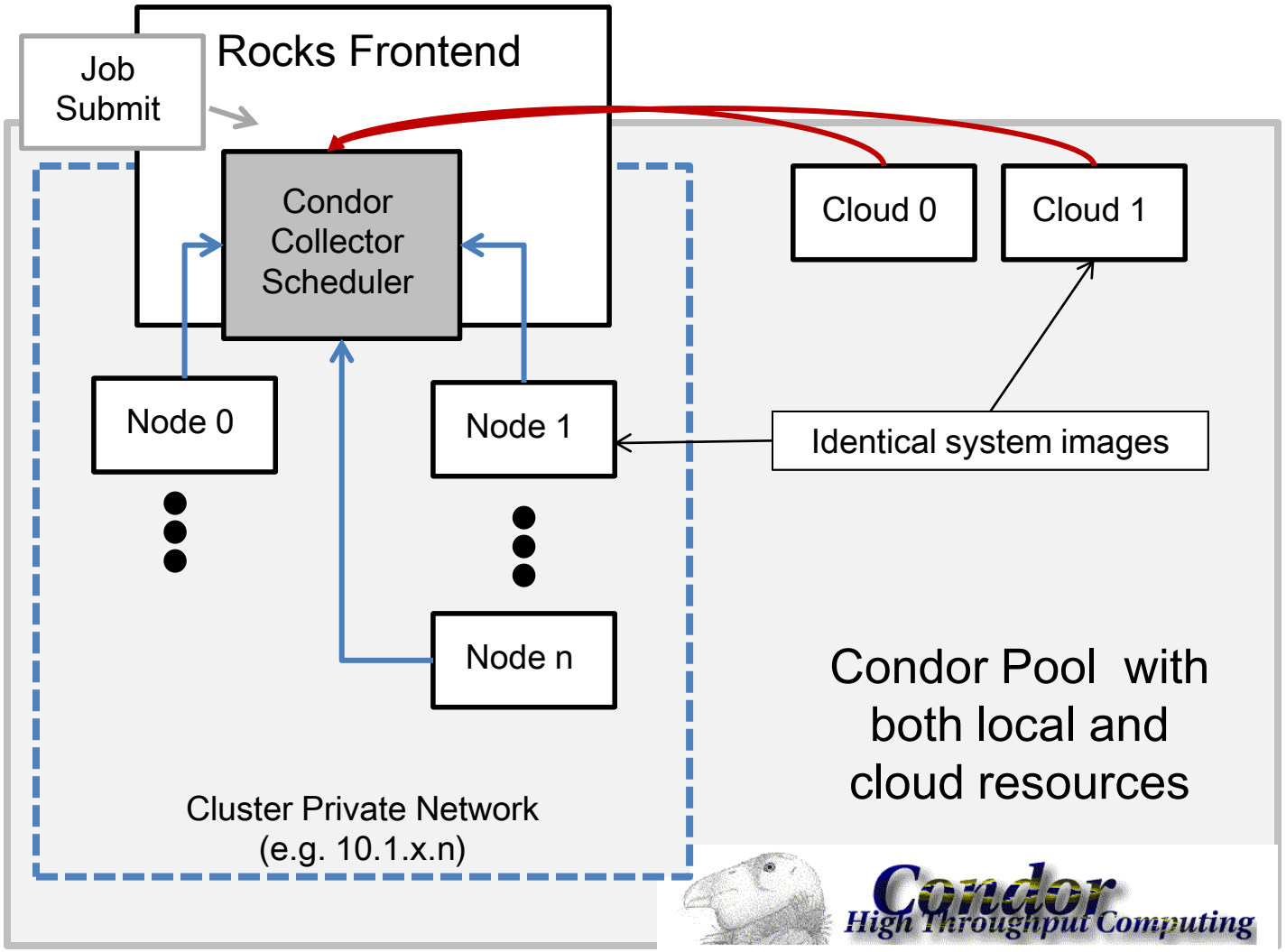
© UC Regents 2010

The EC2 Roll

- Take a Rocks appliance and make it compatible with EC2:
  - 10GB disk partition (single)
  - DHCP for network
  - ssh key management
  - Other small adjustments
- Create an AMI bundle on local cluster
  - `rocks create ec2 bundle`
- Upload a bundled image into EC2
  - `rocks upload ec2 bundle`
- Mini-tutorial on getting started with EC2 and Rocks
  - Register image and go.
- Experimental: automated tunneling setup

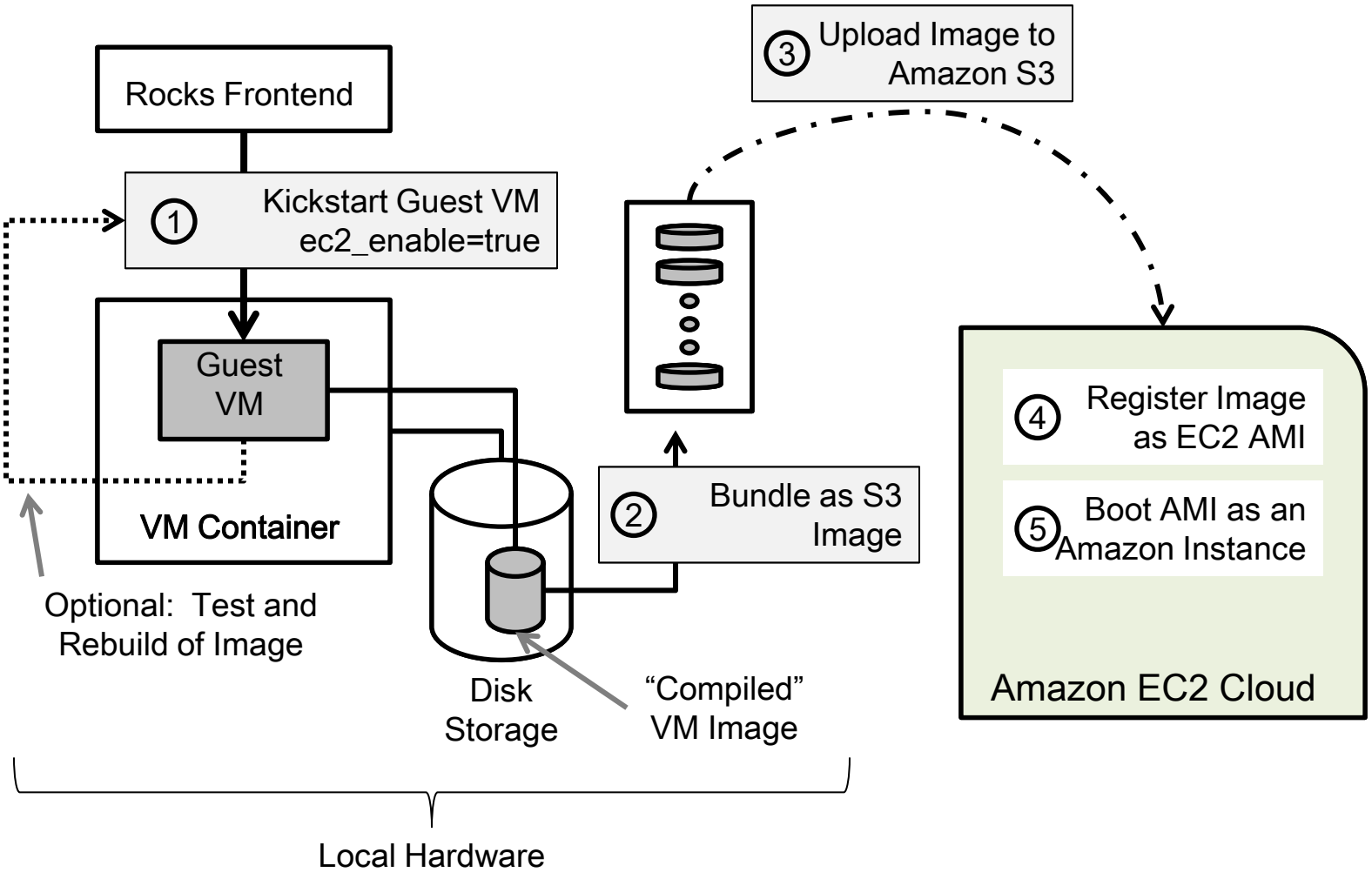
© UC Regents 2010

All Together: Extended Condor Pool with Consistent System Definitions



© UC Regents 2010

Complete Recipe



© UC Regents 2010

At the Command Line: provided by the EC2 Roll/Xen Rolls

1. `rocks set host boot action=install compute-0-0`
2. `rocks set host attr compute-0-0 ec2_enable true`
3. `rocks start host vm compute-0-0`
   - After reboot inspect, then shut down
4. `rocks create ec2 bundle compute-0-0`
5. `rocks upload ec2 bundle compute-0-0 <s3bucket>`
6. `ec2-register <s3bucket>/image.manifest.xml`
7. `ec2-run instances <ami>`

© UC Regents 2010

Summary

- Easily Extend your Condor pool into EC2
  - Others can do this as well
  - Condor supports the public/private network duality of EC2
- Have your software on both local cluster and remote VM in EC2
- Mix and match
  - Local Physical, Local Virtual, Remote Virtual
- Familiar tools and paradigms for cloud-hosted VMs.

© UC Regents 2010