



Application of the virtualisation technology

Predrag Buncic

CERN

CERN School of Computing 2009



Problem

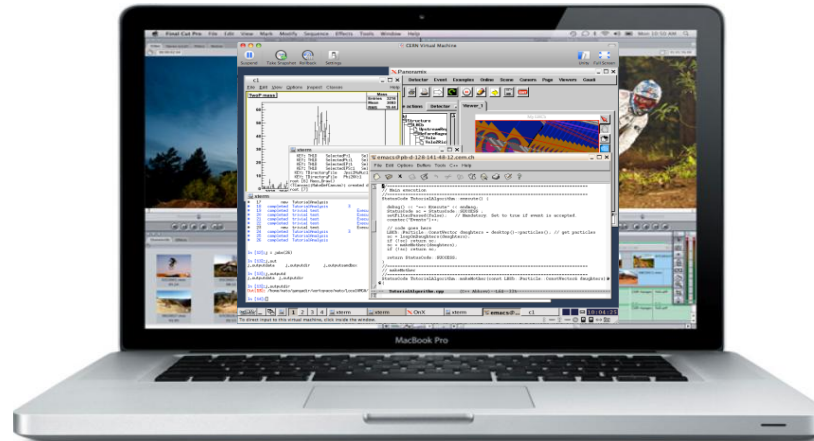
- **Software @ LHC**
 - Millions of lines of code
 - Different packaging and software distribution models
 - Complicated software installation/update/configuration procedure
 - Long and slow validation and certification process
 - Very difficult to roll out major OS upgrade (SLC4 -> SLC5)
 - Additional constraints imposed by the grid middleware development
 - Effectively locked on one Linux flavour
 - Whole process is focused on middleware and not on applications
- **How to effectively harvest multi and many core CPU power of user laptops/desktops if LHC applications cannot run in such environment?**
- **Good news: We are not the only one with such problems...**

Rethinking Application Delivery



<http://www.youtube.com/watch?v=idm16trjKPM>

Solution



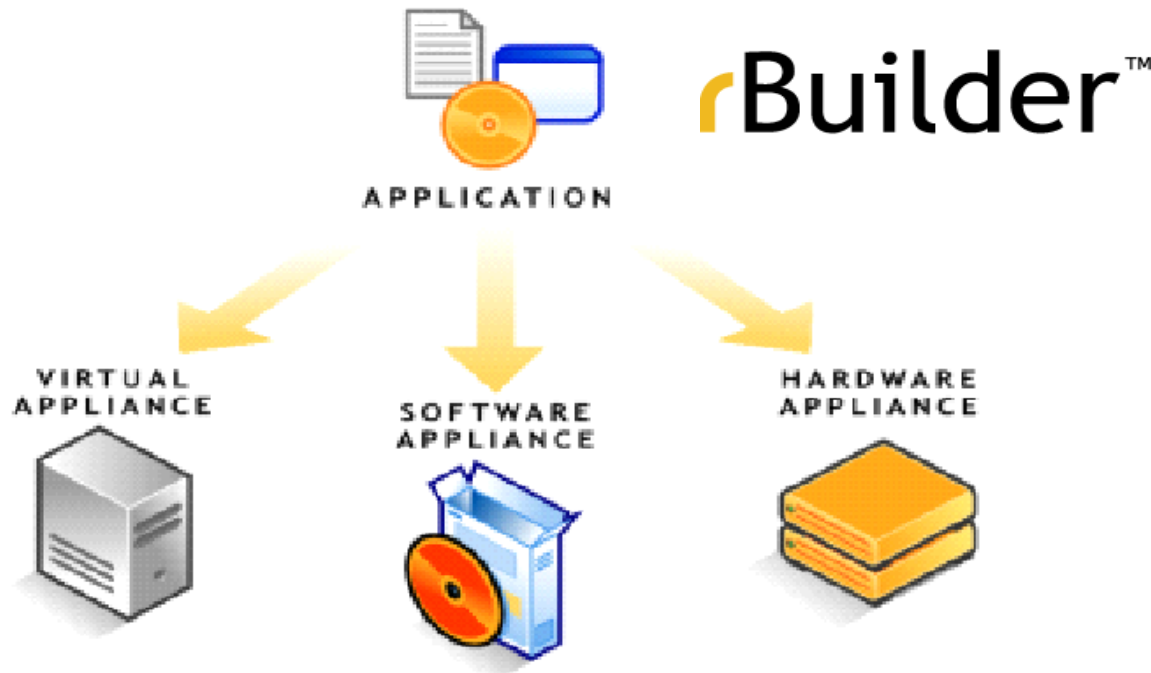
- **A complete, portable and easy to configure user environment for developing and running LHC data analysis locally and on the Grid**
 - Independent of physical software and hardware platform (Linux, Windows, MacOS)
 - Code check-out, edition, compilation, local small test, debugging, ...
 - Grid submission, data access...
 - Event displays, interactive data analysis
 - No user installation required
 - Suspend/Resume capability

Virtual Software Appliance

- **Virtual Software Appliance is a lightweight Virtual Machine image that combines**
 - minimal operating environment
 - specialized application functionality
- **These appliances are designed to run under one or more of the various virtualization technologies, such as**
 - VMware , Xen, Parallels, Microsoft Virtual PC, QEMU, User mode Linux, CoLinux, Virtual Iron...
- **Virtual Software Appliances also aim to eliminate the issues related to deployment in a traditional server environment**
 - Simplify configuration procedure
 - Ease maintenance effort

Virtualizing LHC applications

Starting from experiment software...



...ending with custom Linux specialised for a given task

Build types

- Installable CD/DVD
- Stub Image
- Raw Filesystem Image
- Netboot Image
- Compressed Tar File
- Demo CD/DVD (Live CD/DVD)
- Raw Hard Disk Image
- Vmware® Virtual Appliance
- Vmware® ESX Server Virtual Appliance
- Microsoft® VHD Virtual Appliance
- Xen Enterprise Virtual Appliance
- Virtual Iron Virtual Appliance
- Parallels Virtual Appliance
- Amazon Machine Image
- Update CD/DVD
- Appliance Installable ISO

Conary Package Manager

```
class Root(CPackageRecipe):  
    name='root'  
    version='5.19.02'  
  
    buildRequires = ['libpng:devel',  
                    'libpng:devellib', 'krb5:devel',  
                    'libstdc++:devel', 'libxml2:devel',  
                    'openssl:devel', 'python:devel',  
                    'xorg-x11:devel', 'zlib:devel',  
                    'perl:devel', 'perl:runtime']  
  
    def setup(r):  
        r.addArchive('ftp://root.cern.ch/root/%(name)s_v%(version)s.source.tar.gz')  
        r.Environment('ROOTSYS',%(builddir)s')  
        r.ManualConfigure('--prefix=/opt/root ')  
        r.Make()  
        r.MakeInstall()
```



1. Find what you need

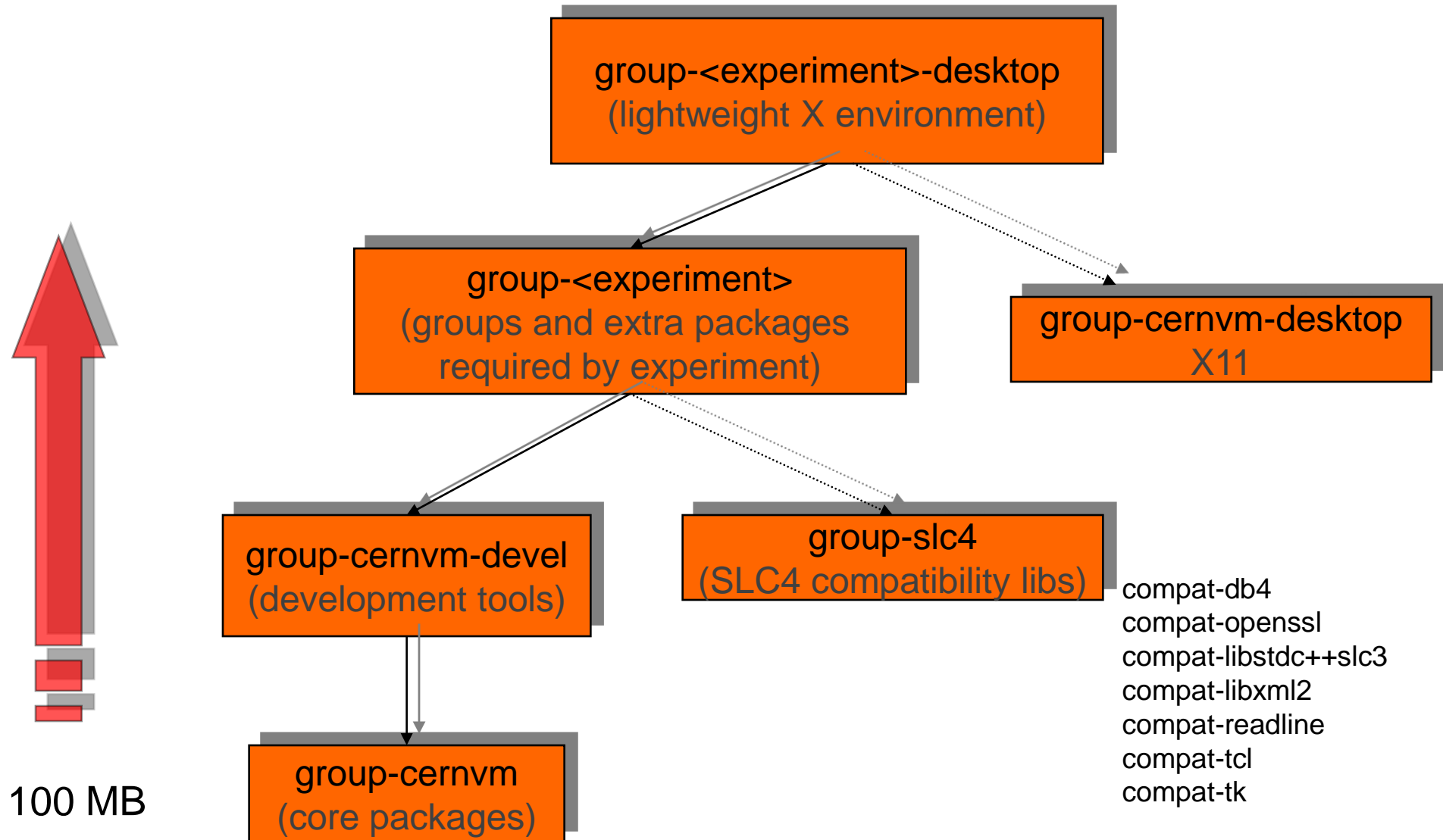


2. Build your recipe

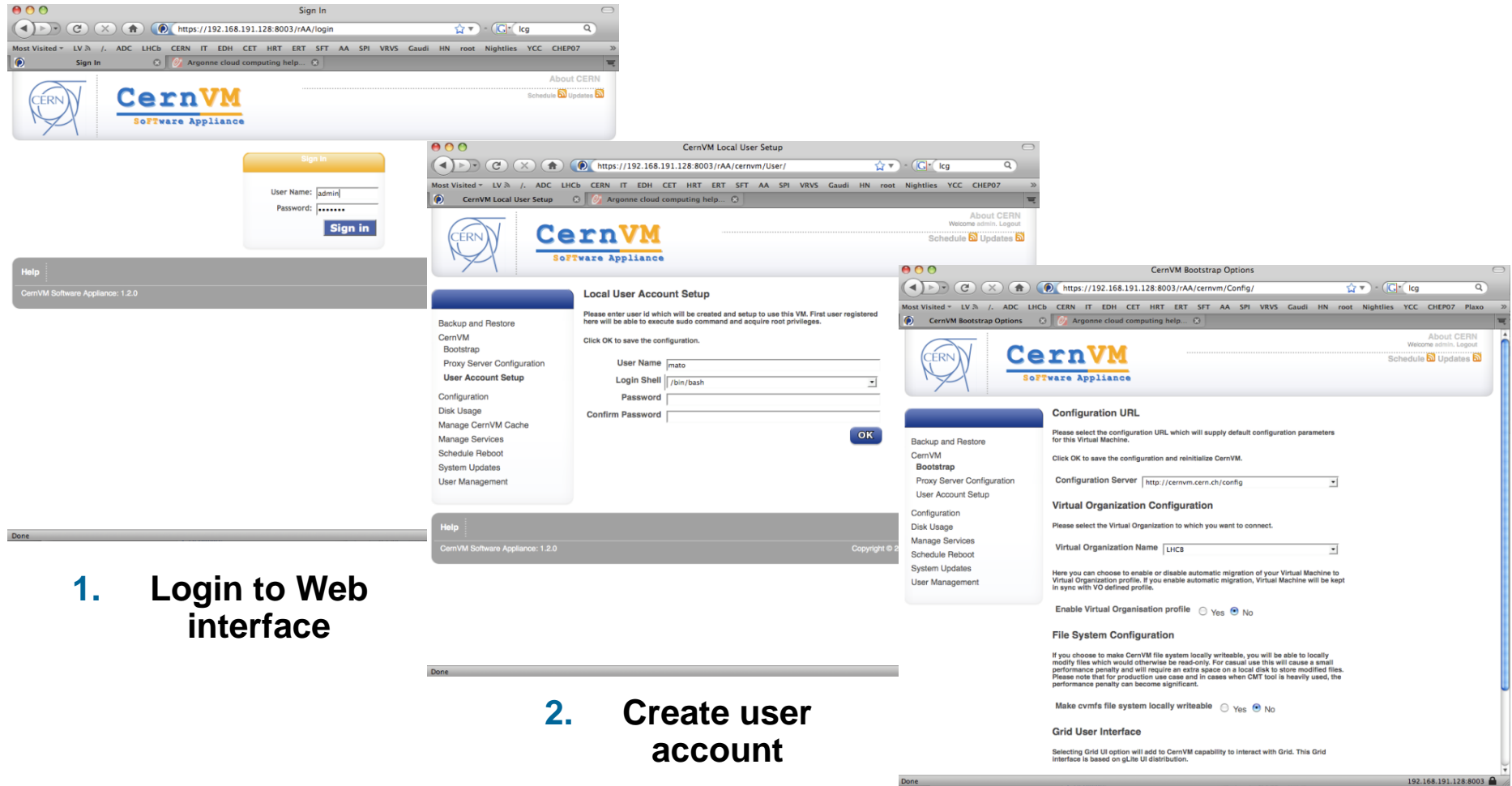


3. Cook it!

Package Groups



As easy as 1,2,3



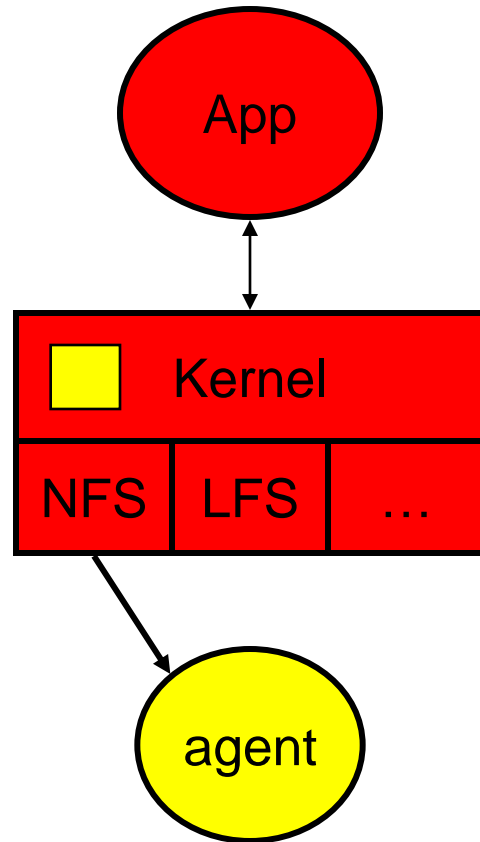
1. Login to Web interface

2. Create user account

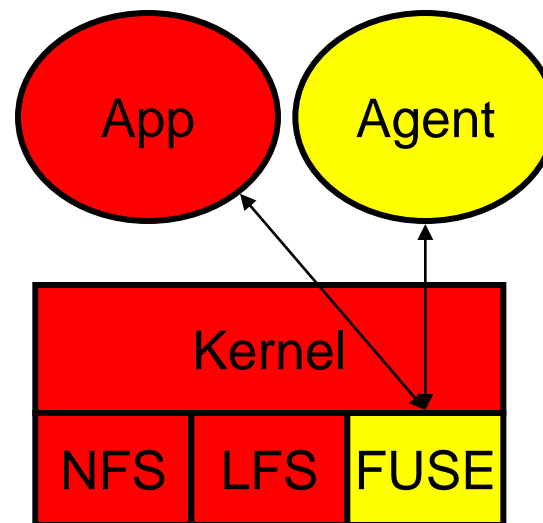
3. Select experiment, appliance flavor and preferences

Options for File System virtualization

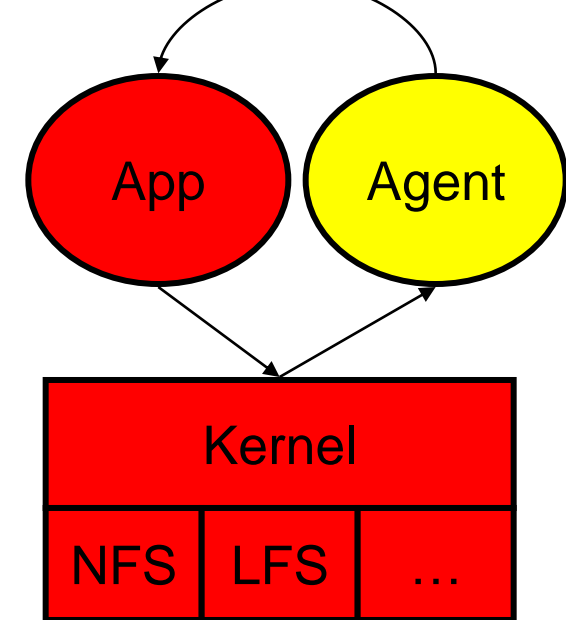
Kernel File System (NFS)



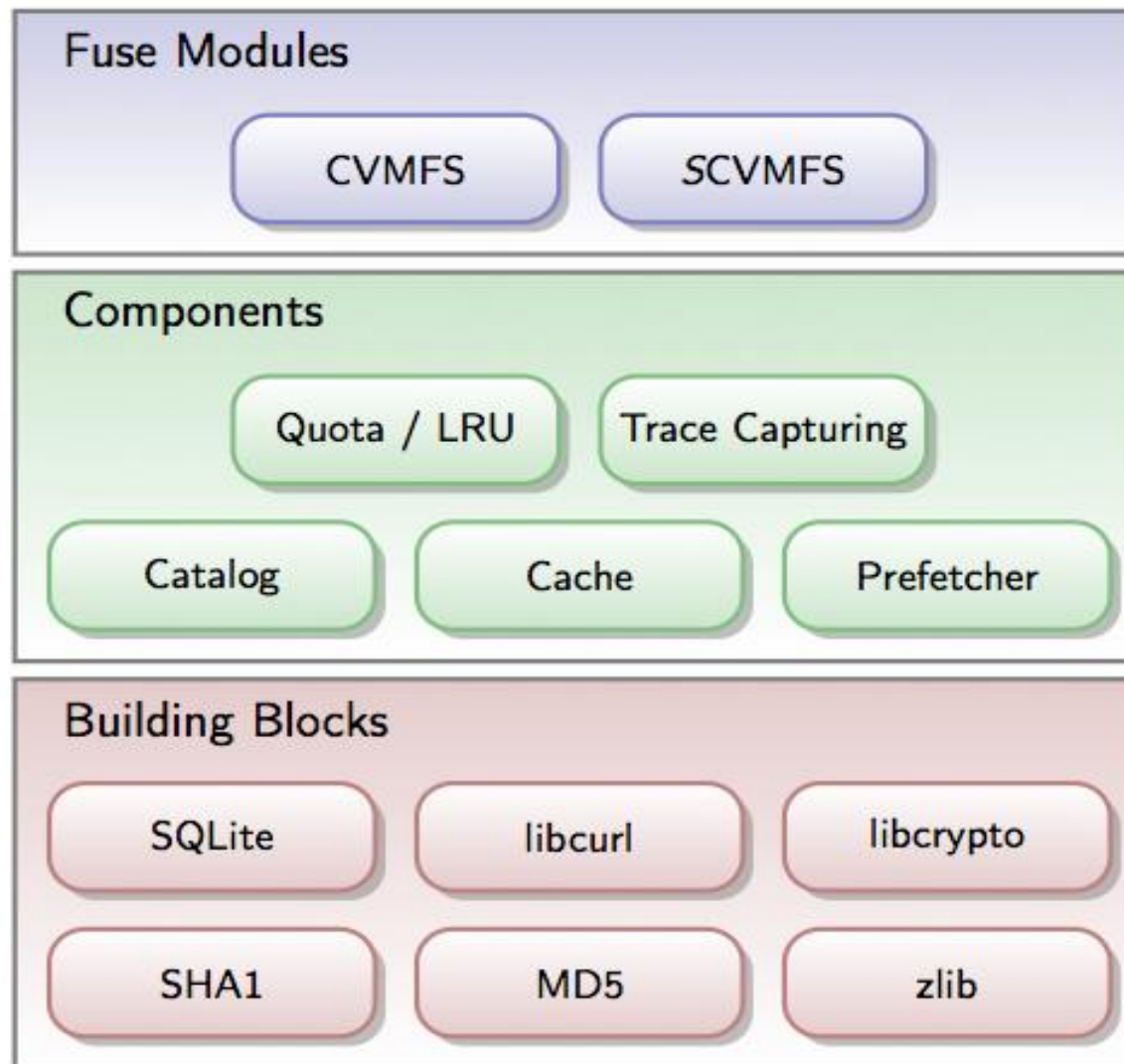
Kernel Callout (Fuse)



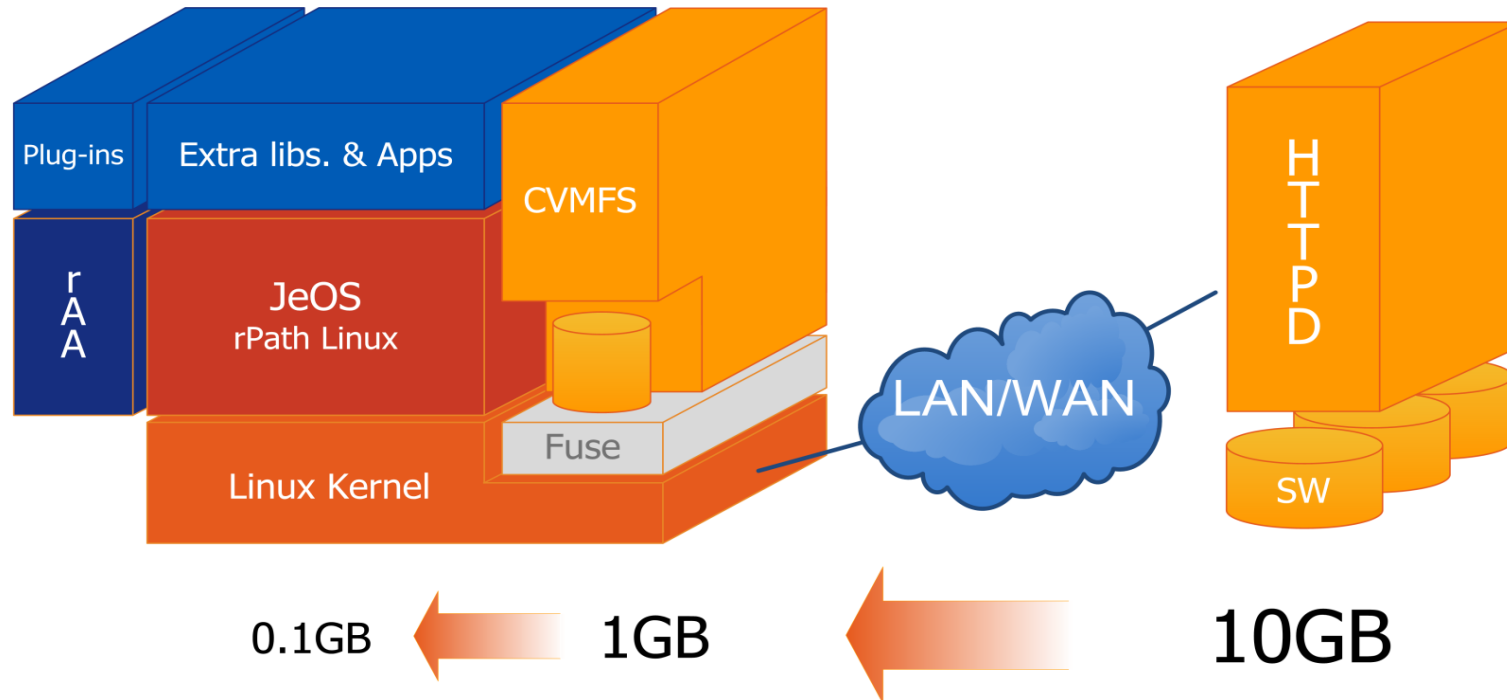
Debugger Trap (Parrot)



CVMFS



“Thin” Virtual Machine

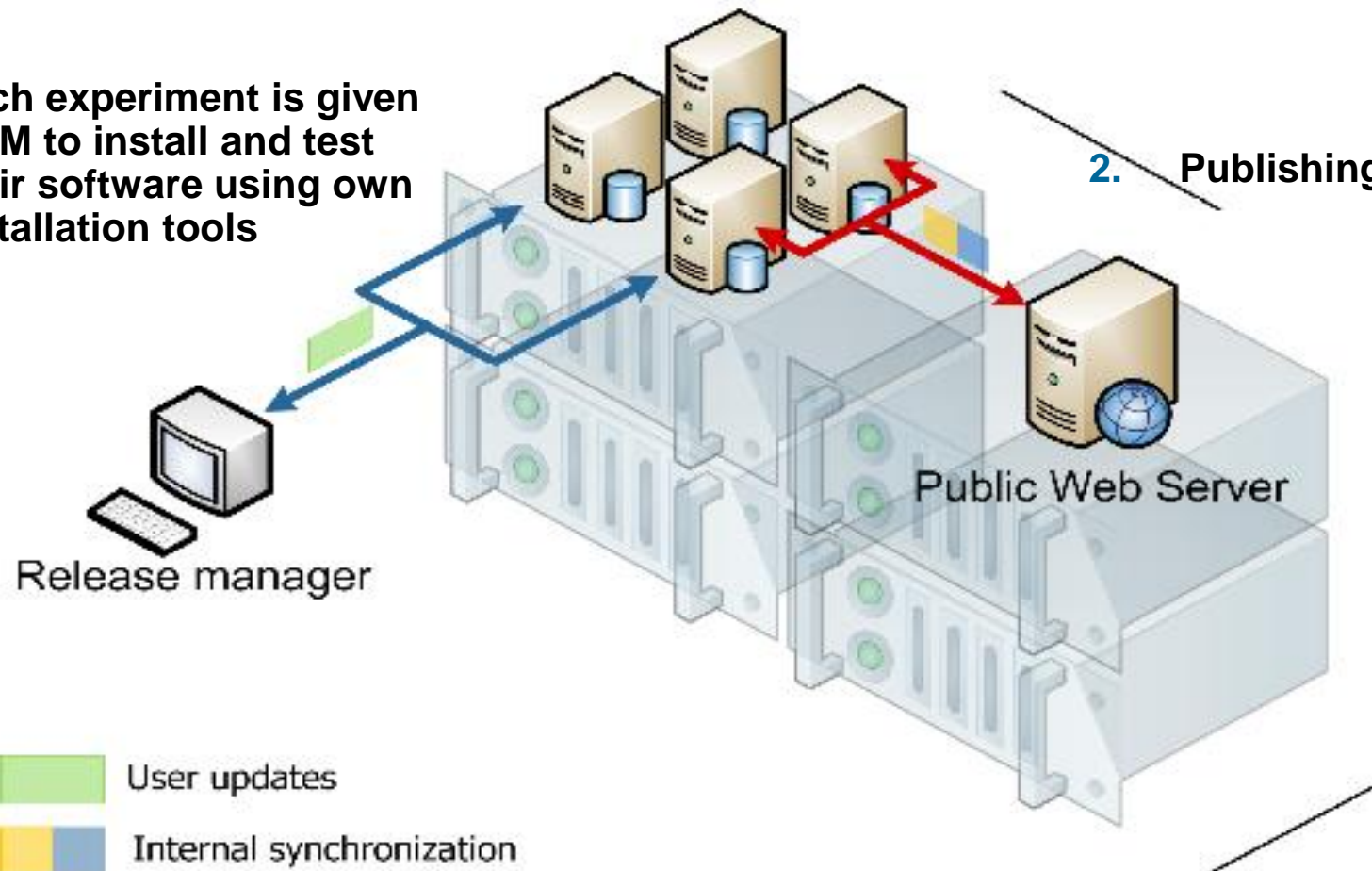


- **The experiment are packaging a lot of code**
 - but really use only fraction of it at runtime
- **CernVM downloads what is needed and puts it in the cache**
 - Does not require persistent network connection (offline mode)

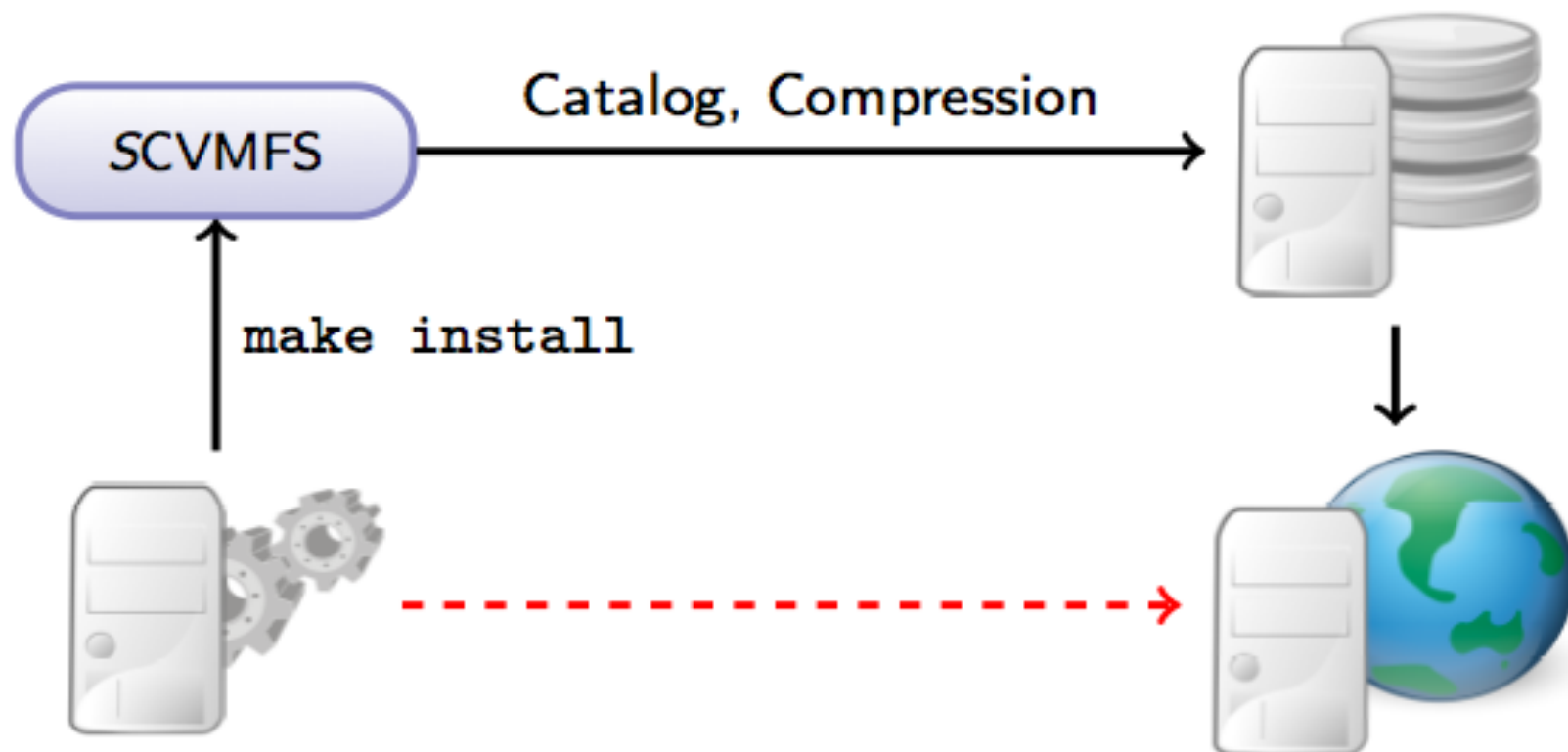
Publishing Software Releases

1. Each experiment is given a VM to install and test their software using own installation tools

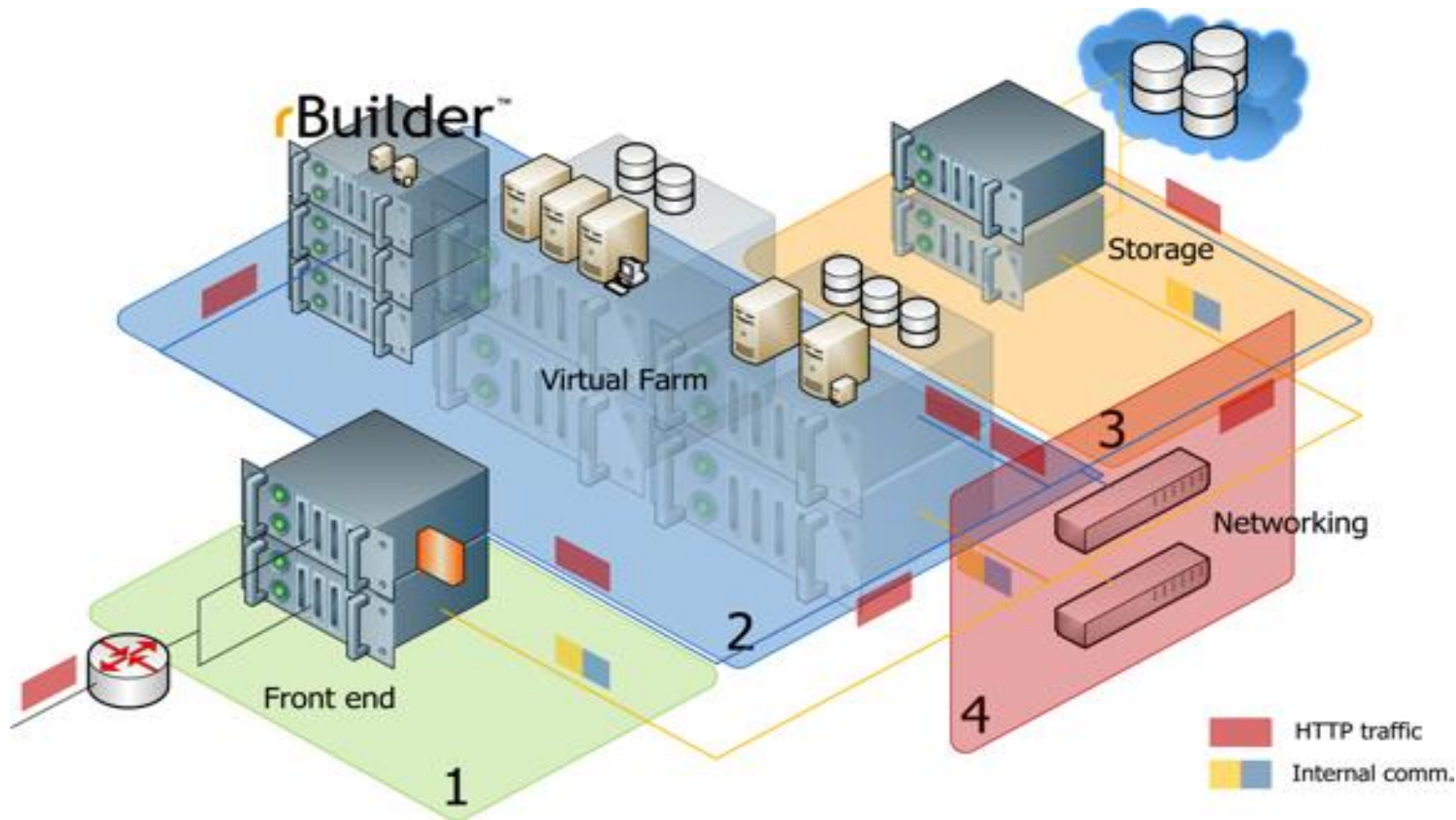
2. Publishing is an atomic operation



SCVMFS



Virtual Support Infrastructure



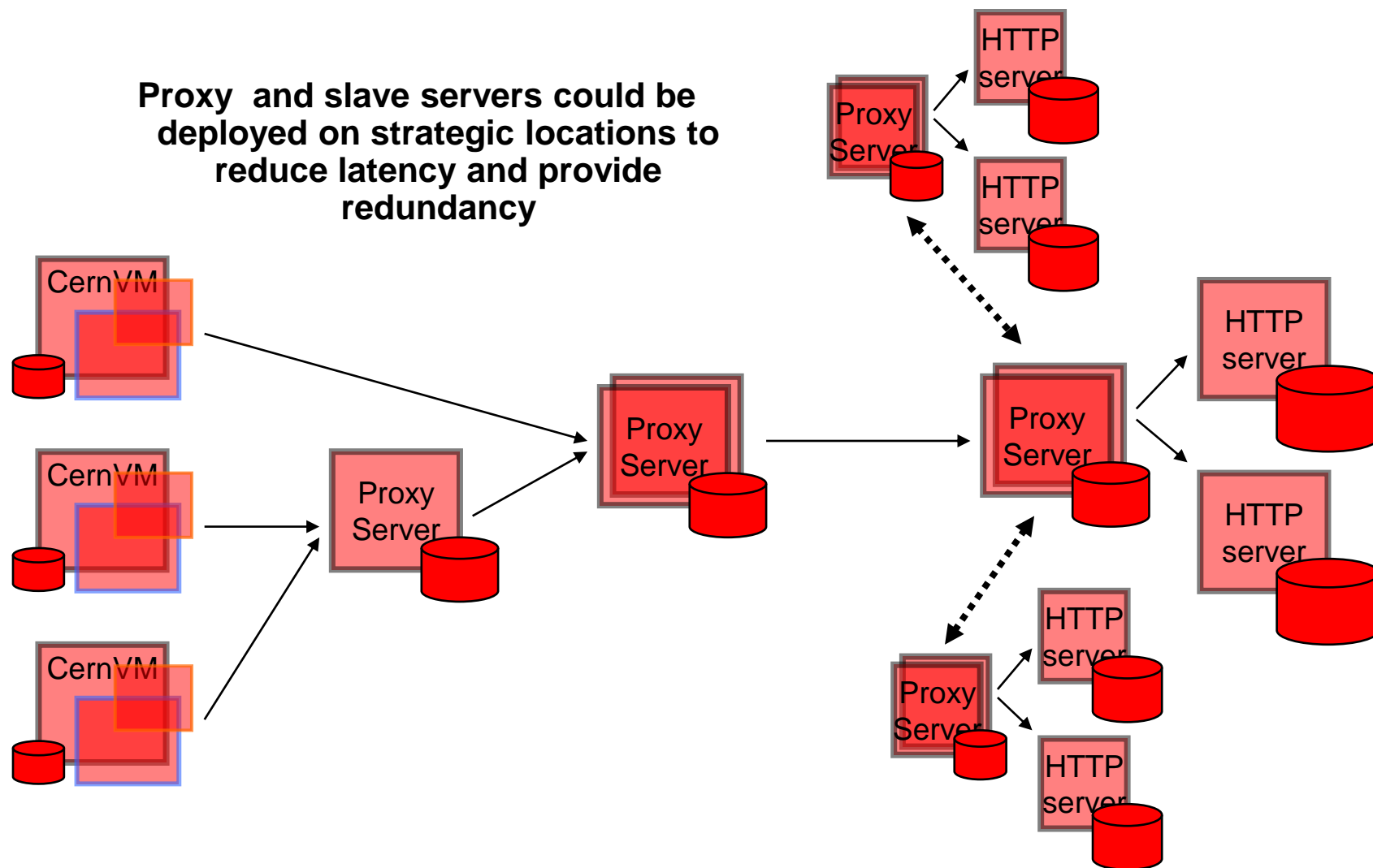
Where are our users?



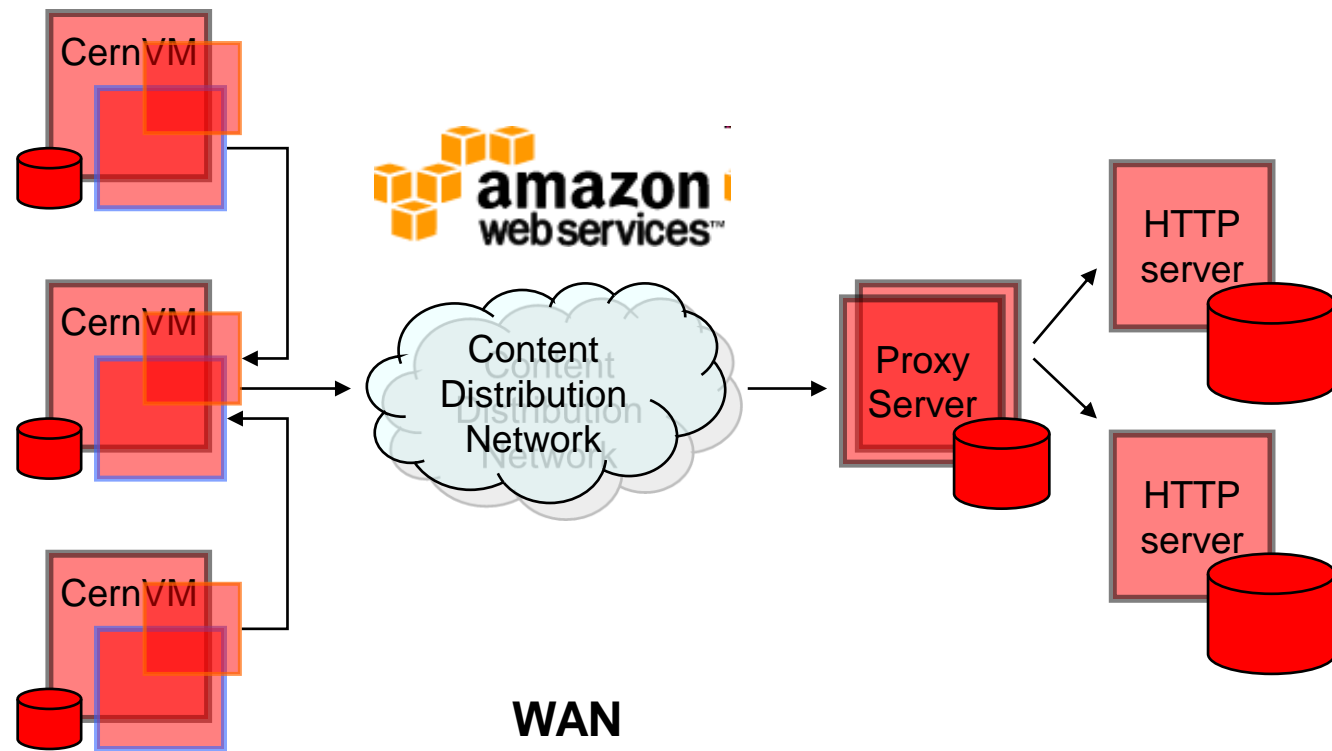
~800 different IP addresses

Scaling up...

Proxy and slave servers could be deployed on strategic locations to reduce latency and provide redundancy



LAN & WAN



LAN

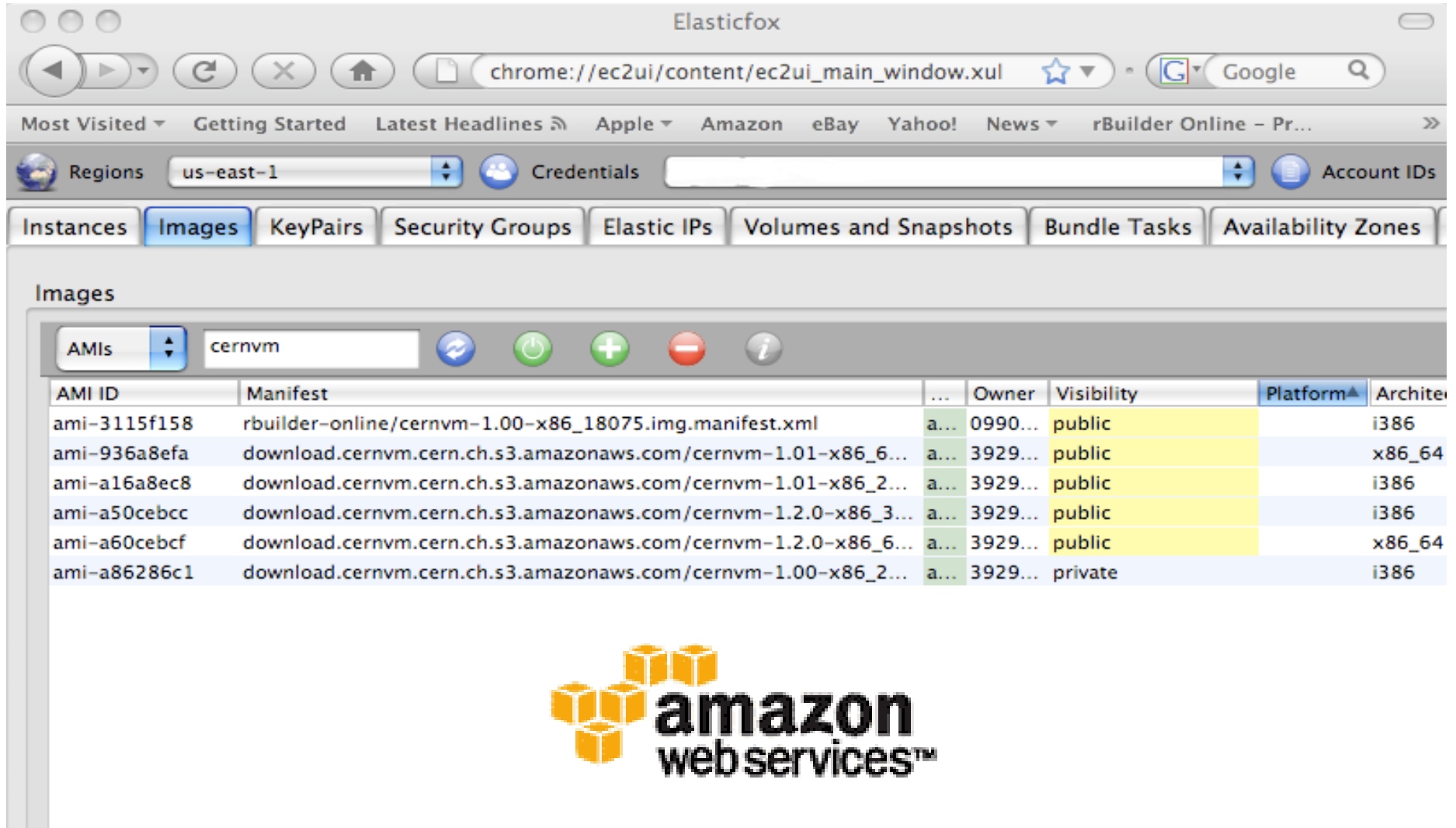
Use P2P like mechanism for discovery of nearby CernVMs and cache sharing between them. No need to manually setup proxy servers (but they could still be used where exist)

WAN

Use existing Content Delivery Networks to remove single point of failure

- Amazon CloudFront (<http://aws.amazon.com/cloudfront/>)
- Coral CDN (<http://www.coralcdn.org>)

Ready for Amazon EC2...

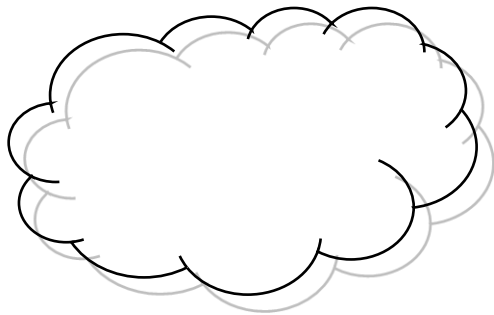


The screenshot shows the Elasticfox web interface, which is a wrapper for the Amazon EC2 console. The browser window is titled 'Elasticfox' and the address bar shows 'chrome://ec2ui/content/ec2ui_main_window.xul'. The interface includes a navigation bar with tabs for 'Instances', 'Images', 'KeyPairs', 'Security Groups', 'Elastic IPs', 'Volumes and Snapshots', 'Bundle Tasks', and 'Availability Zones'. The 'Images' tab is currently selected. Below the navigation bar, there is a search bar with 'cernvm' entered. A table of AMIs is displayed below the search bar.

AMI ID	Manifest	...	Owner	Visibility	Platform▲	Archite
ami-3115f158	rbuilder-online/cernvm-1.00-x86_18075.img.manifest.xml	a...	0990...	public		i386
ami-936a8efa	download.cernvm.cern.ch.s3.amazonaws.com/cernvm-1.01-x86_6...	a...	3929...	public		x86_64
ami-a16a8ec8	download.cernvm.cern.ch.s3.amazonaws.com/cernvm-1.01-x86_2...	a...	3929...	public		i386
ami-a50cebcb	download.cernvm.cern.ch.s3.amazonaws.com/cernvm-1.2.0-x86_3...	a...	3929...	public		i386
ami-a60cebcf	download.cernvm.cern.ch.s3.amazonaws.com/cernvm-1.2.0-x86_6...	a...	3929...	public		x86_64
ami-a86286c1	download.cernvm.cern.ch.s3.amazonaws.com/cernvm-1.00-x86_2...	a...	3929...	private		i386

Below the table, the Amazon Web Services logo is visible.

CernVM as job hosting environment



- Ideally, users would like run their applications on the Grid (or Cloud) infrastructure in exactly the same conditions in which they were developed
- CernVM already provides development environment and can be deployed on cloud (EC2)
 - Easily extensible to other communities beyond LHC experiments

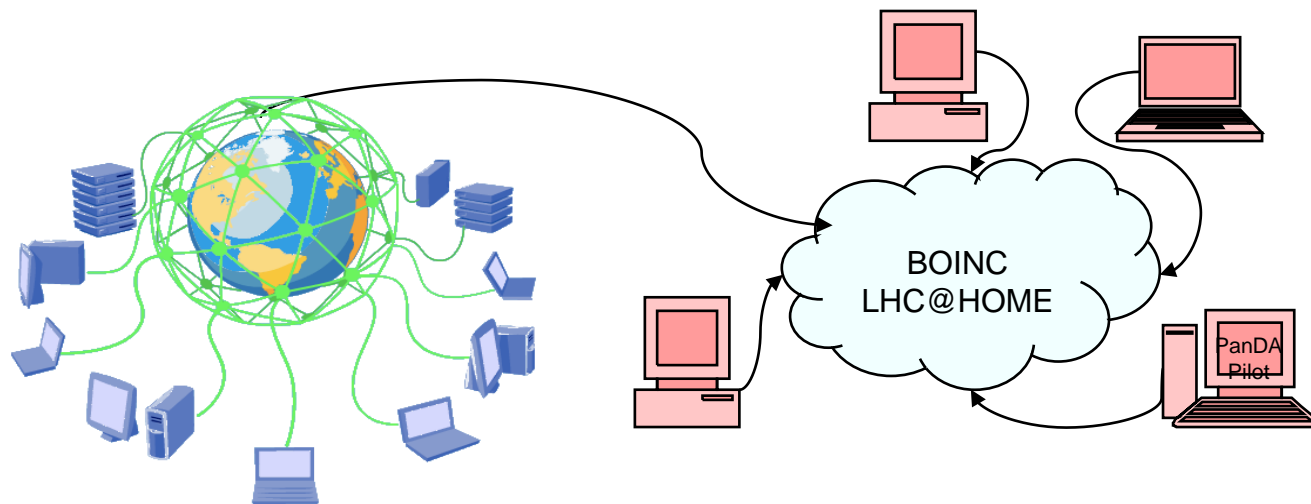
Advantages

- **Exactly the same environment for development and job execution**
- **Software can be efficiently installed using CVMFS**
 - HTTP proxy assures very fast access to software even if VM cache is cleared
- **Can accommodate multi-core jobs**
- **Deployment on EC2 or alternative clusters**
 - Nimbus, Elastic

Bridging Grids & Clouds

■ BOINC

- Open-source software for volunteer computing and grid computing
- <http://boinc.berkeley.edu/>
- Ongoing development to use VirtualBox running CernVM as a job container
 - <http://boinc.berkeley.edu/trac/wiki/VirtualBox>
- Adds possibility to run unmodified user applications
- Better security due to guest OS isolation



Infrastructure as a Service (IaaS)

- **Nimbus (former Globus Workspace Service)**
 - Nimbus is a set of open source tools that together provide an "Infrastructure-as-a-Service" (IaaS) cloud computing solution
 - <http://workspace.globus.org/>
 - Successfully created virtual AliEn site for ALICE with one command



Conclusions

- **Lots of interest from LHC experiments and huge momentum in industry**
 - ATLAS, LHCb, CMS, ALICE, LCD
- **Hypervisors are nowadays available for free (Linux, Mac and Windows)**
 - But managing tools and support are not
- **CernVM approach solves the problem of efficient software distribution**
 - Using its own dedicated file system
 - Reducing deployment problem
- **Initially developed as user interface for laptop/desktop**
 - Already deployable on the cloud (EC2, Nimbus)
 - Can be deployed on managed (and unmanaged infrastructure) without necessarily compromising the site security
- **Deployment on the grid or in the computer centre environment requires changes to some of the current practices and thinking**
 - Utilizing private networks to avoid shortage of IP numbers and to hide VMs from public internet
 - Use proxy/caches wherever possible