CernVM: Ten Years After

Predrag Buncic (CERN/EP-SFT)

Agenda

- This is 10th anniversary of CernVM project and this keynote talk will be more about the project itself than about the project outcome
  - How this project came about and how it survived 10 years?
- This talk will be mostly about past and a little bit about the present.
- The discussion about the future I leave to the current development team and a growing community of our enthusiastic users ☺.

CernVM
Software Appliance

How did we get here?

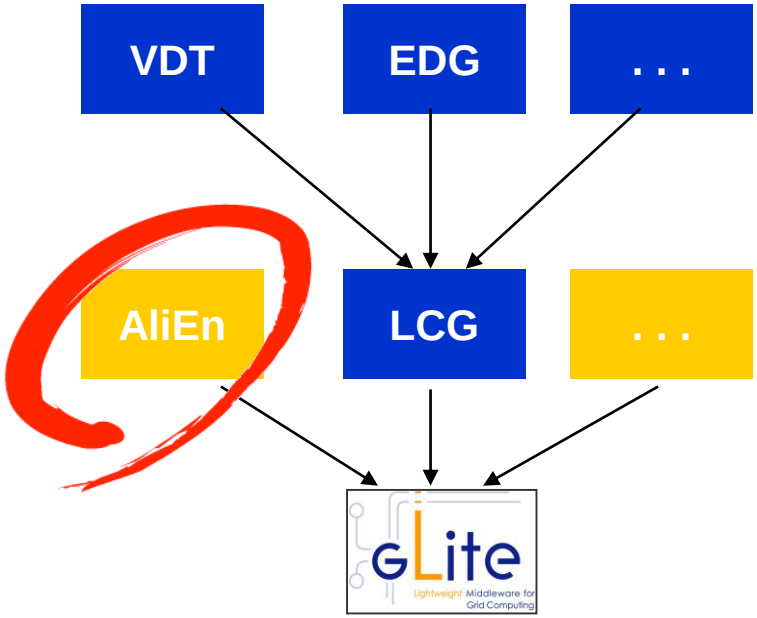
CernVM
Software Appliance

2004-2006

CernVM
Software Appliance

Grid old times...

EGEE Middleware

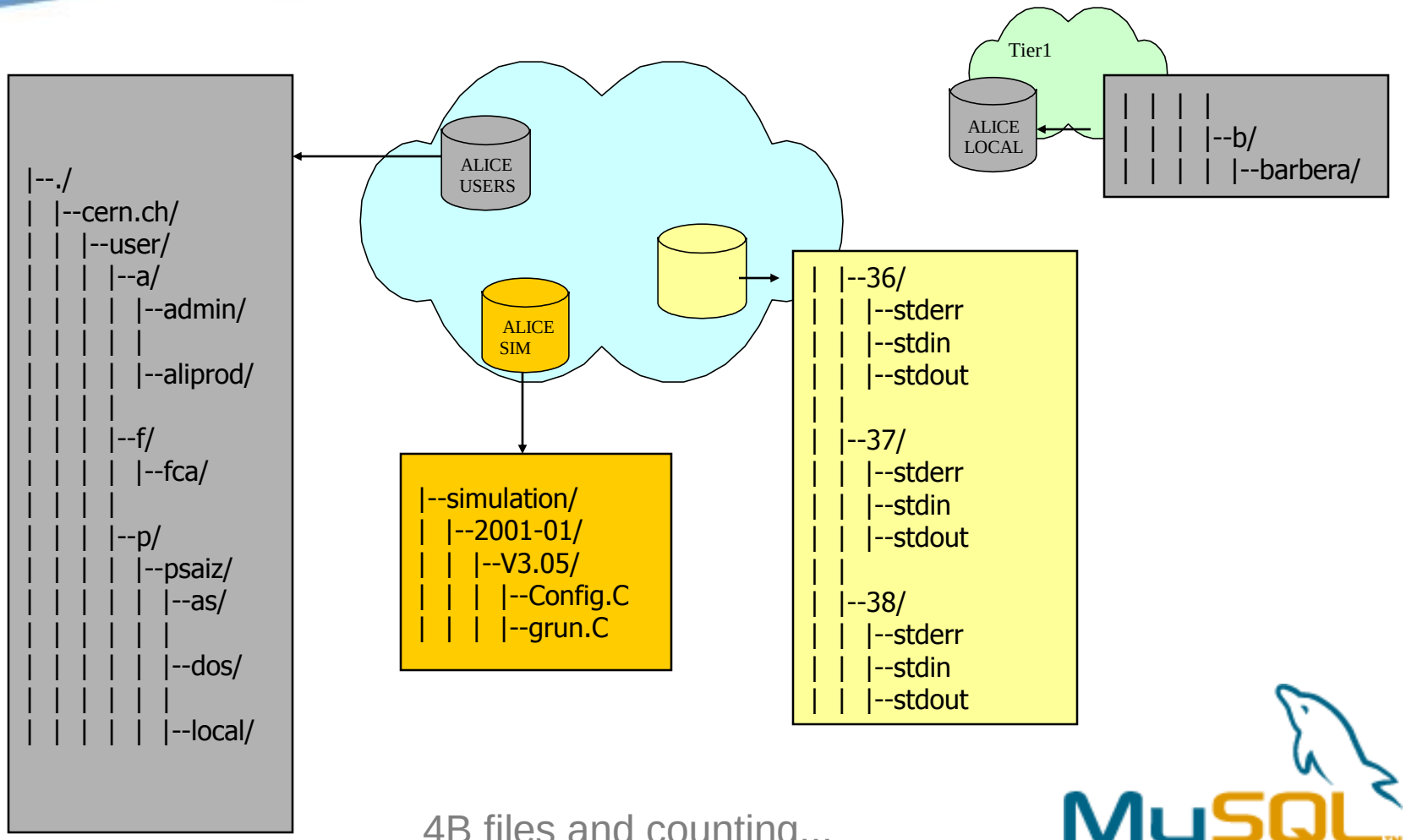


Goal:

Re-engineer and harden Grid middleware (AliEn, EDG, VDT and others)

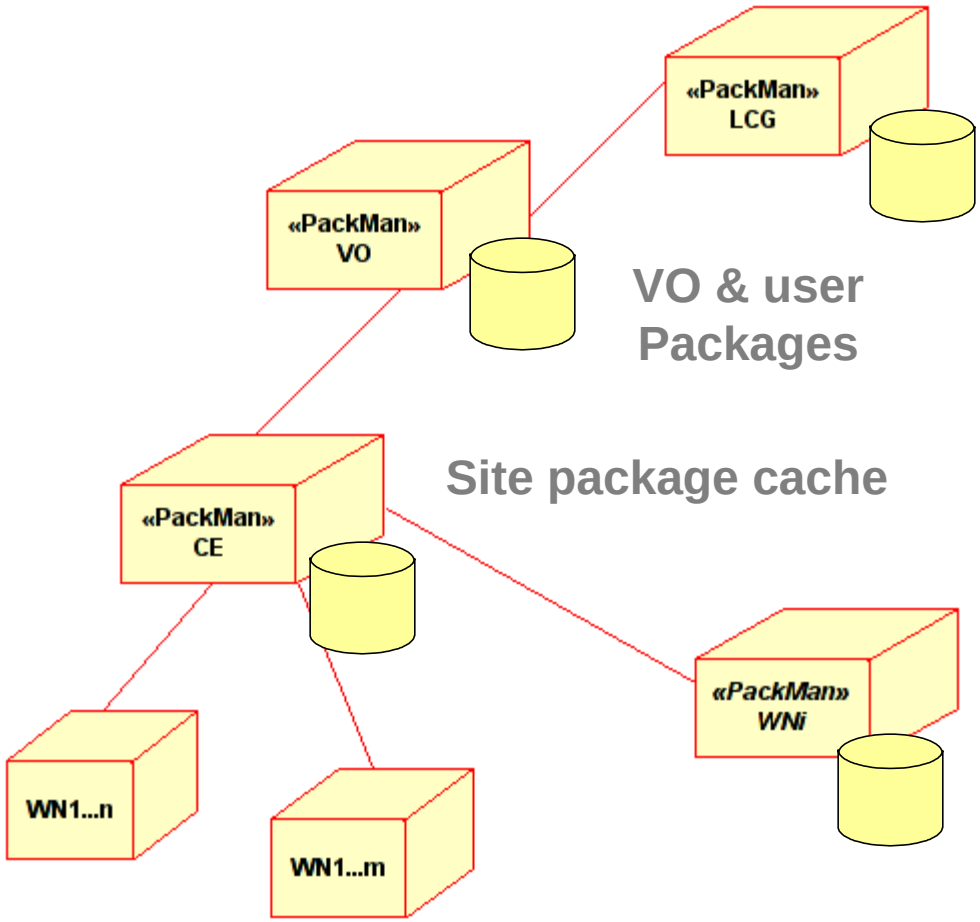
Provide production quality middleware

AliEn File Catalog



4B files and counting...

Package Manager



Common Packages (ROOT, POOL,..)

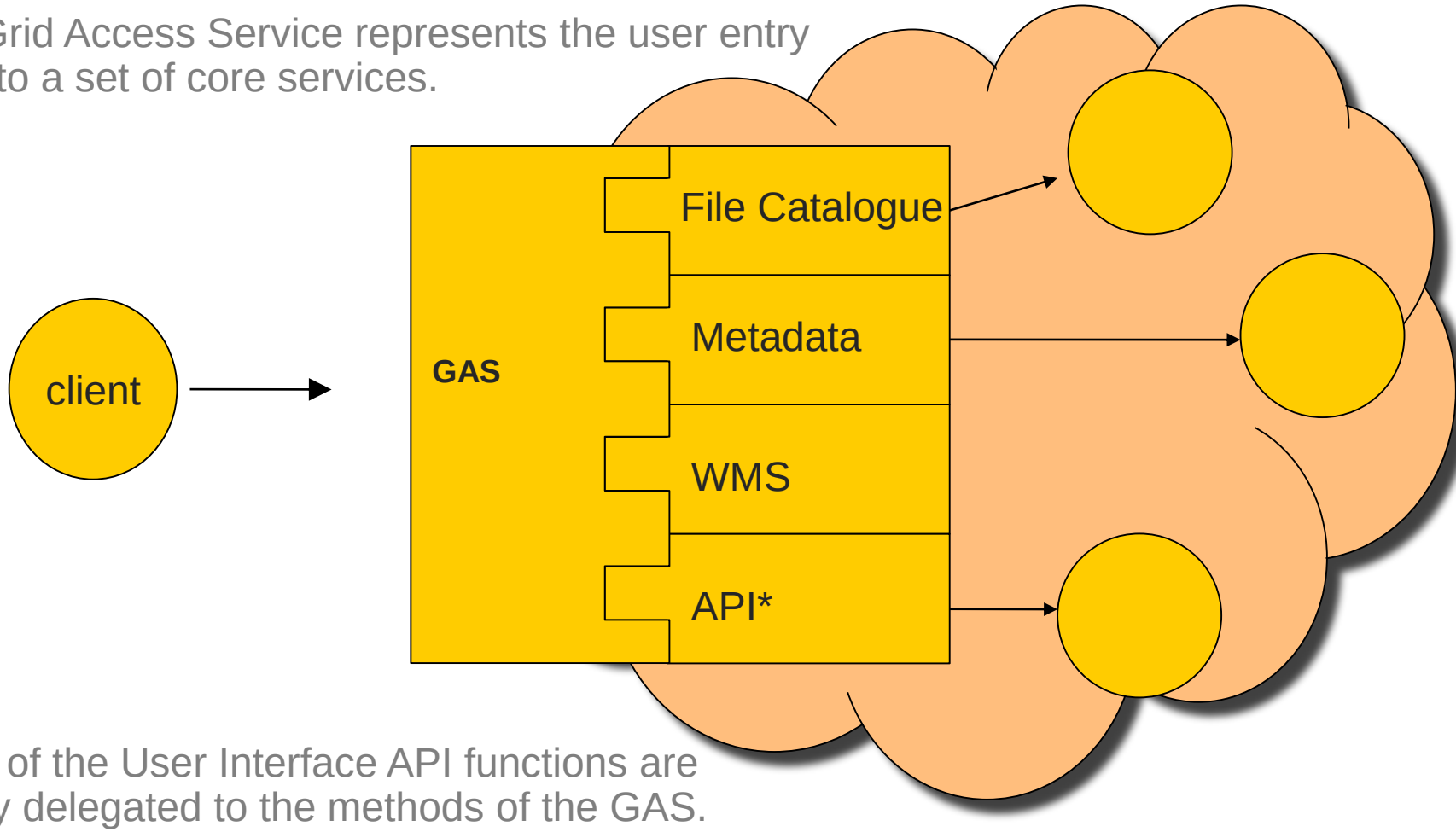
VO & user Packages

Site package cache

Worker node cache

Grid Access Service (GAS)

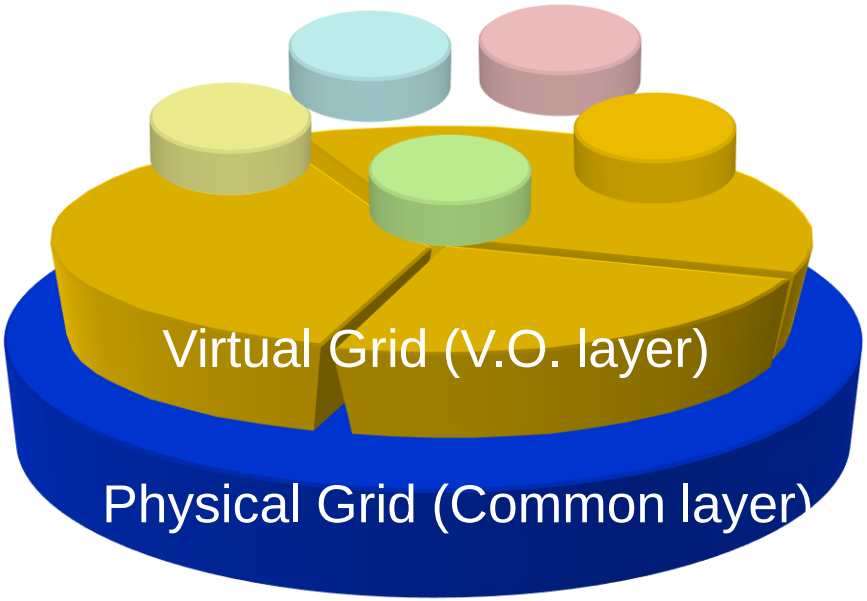
- The Grid Access Service represents the user entry point to a set of core services.



- Many of the User Interface API functions are simply delegated to the methods of the GAS. In turn, many of the GAS functions are delegated to the appropriate service.

The Big Picture

Virtual Cluster (User layer)



AliEn/PROOF demo at SC05

In 2006 moved from IT => PH/SFT

Virtual Machines

- Xen takes a novel approach by eliminating sensitive instructions directly in the guest OSs' original source code, which is called *para-virtualization*
  - These instructions are replaced with equivalent operations or emulated by replacing them with *hypercalls*, which call equivalent procedures in the VMM, or *hypervisor*
  - The hypervisor runs as the most privileged kernel, while guest OS kernels run less privileged on top of the hypervisor.
  - This method yields close-to-native performance

- *Once it is accepted that Job Agent can execute privileged commands, we are step closer to convincing the sites that they should let us run the Grid jobs within Virtual Machine*
  - *This can provide perfect process a file sandboxing*
  - *Software which is run inside a VM can not negatively affect the execution of another VM*

What's New | 2006

Announcing Amazon S3 - Simple Storage Service

Posted On: Mar 13, 2006

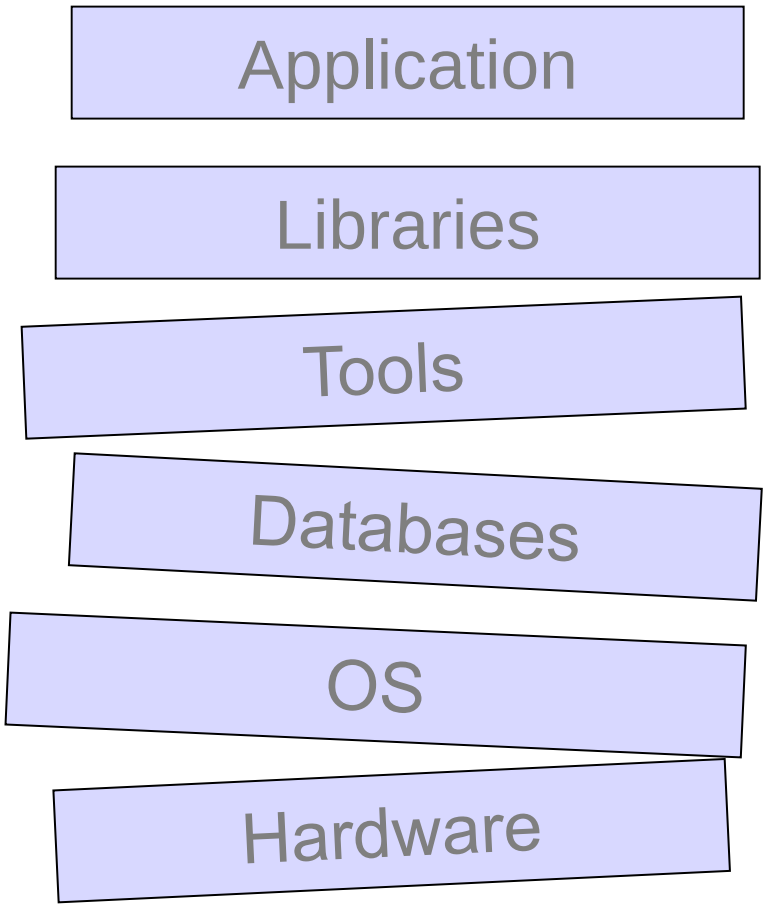
Amazon S3 is storage for the Internet. It is designed to make web-scale computing easier for developers. Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web. It gives any developer access to the same highly scalable, reliable, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites.

Announcing Amazon Elastic Compute Cloud (Amazon EC2) - beta

Posted On: Aug 24, 2006

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers. Just as Amazon Simple Storage Service (Amazon S3) enables storage in the cloud, Amazon EC2 enables "compute" in the cloud. Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use.

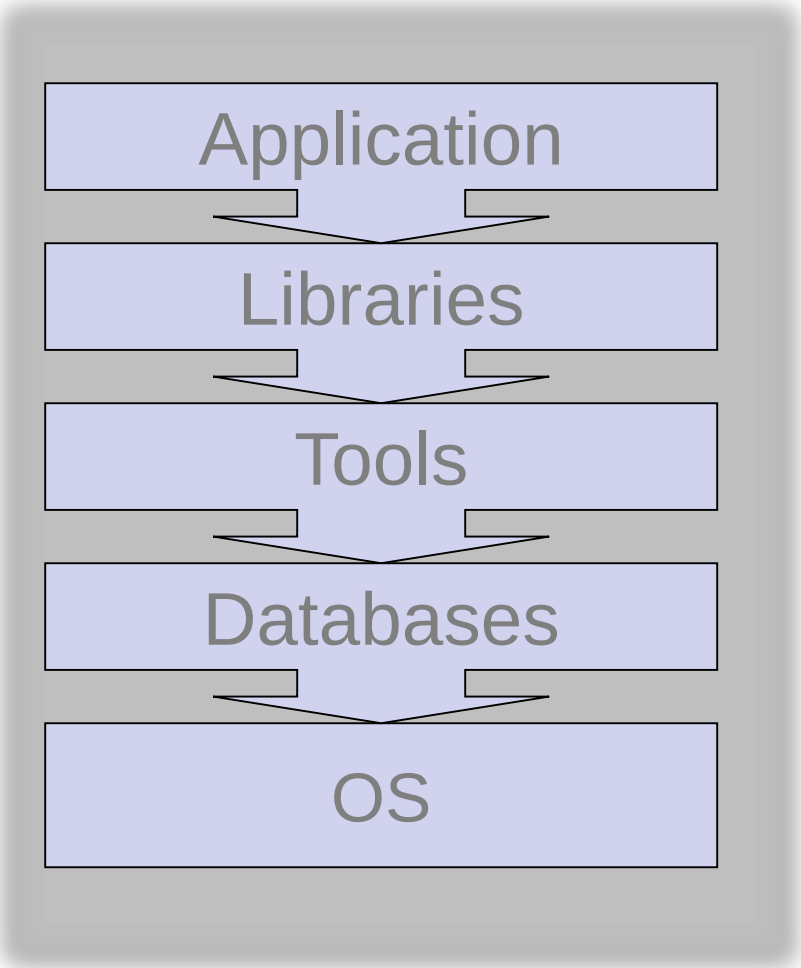
Software Integration



- Traditional model
  - driven by the platform (grid middleware)
  - Horizontal layers
  - Independently developed
  - Maintained by the different groups
  - Different lifecycle
- Application is deployed on top of the stack
  - Breaks if any layer changes
  - Needs to be certified every time when something changes
  - Results in deployment and support nightmare
- Difficult to do upgrades
  - Even worse to switch to new OS versions
  - Migration SL5 -> SL5 took several years and was completed in 2011
  - Now already obsolete and superseded by SL6

CernVM
Software Appliance

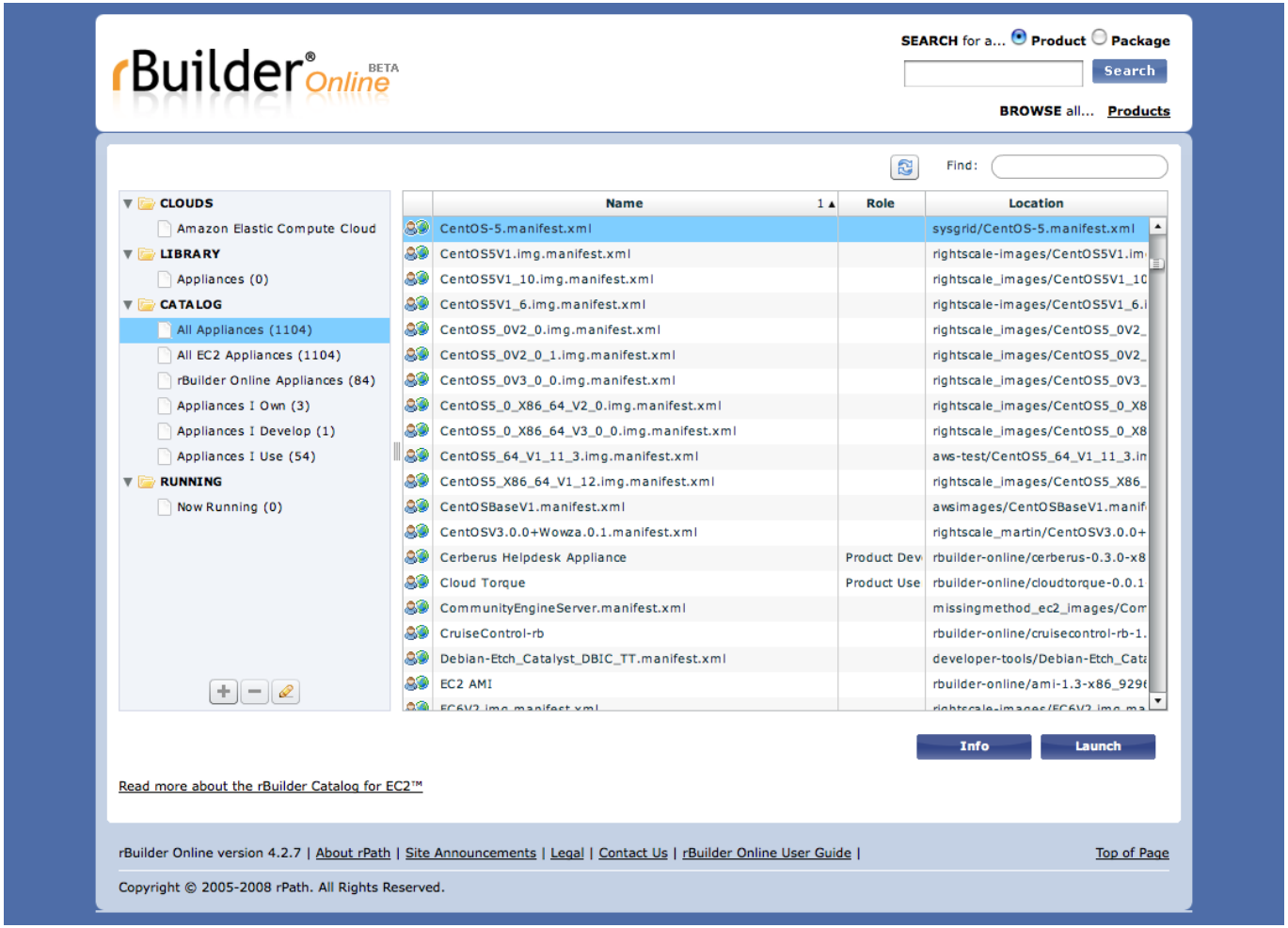
Possible Solution



- Vertical instead of horizontal software integration approach
- Application driven approach
  1. Start by analysing the application requirements and dependencies
  2. Add required tools and libraries
- Use virtualization to
  1. Build minimal OS
  2. Bundle all this into Virtual Machine image
- Problem solved?

CernVM Software Appliance

rBuilder from rPath



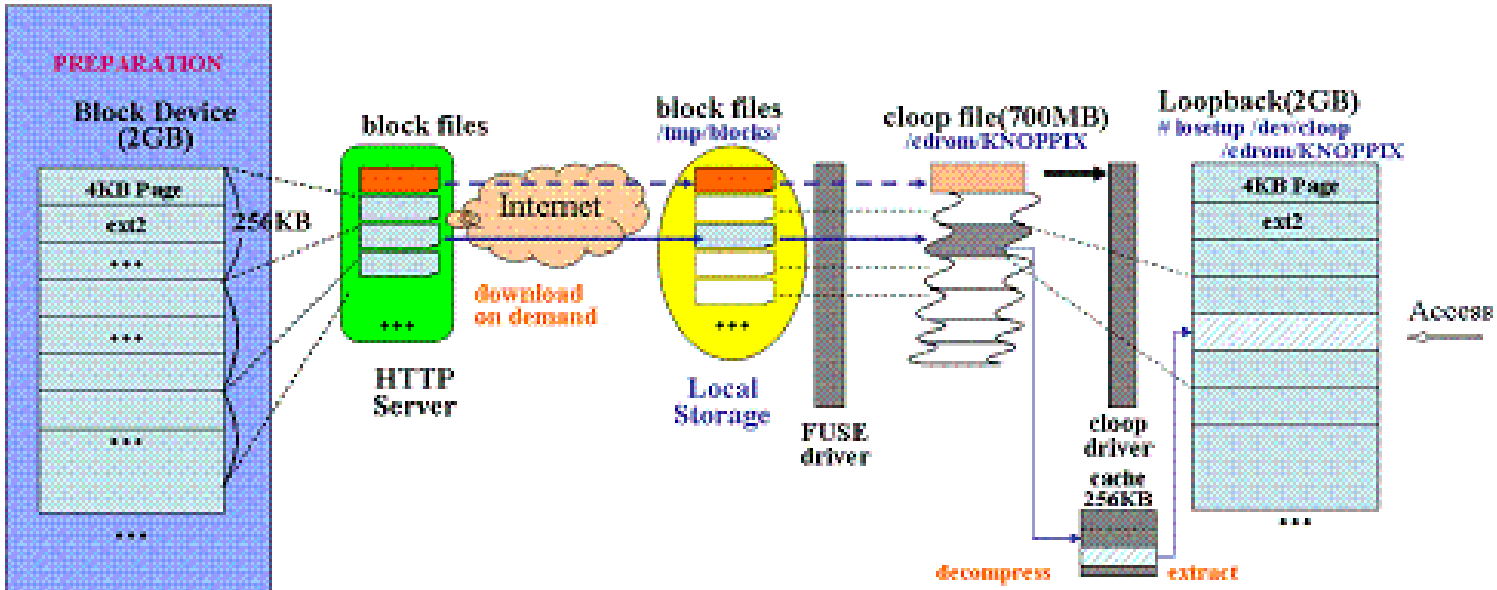
The Real Problem

- 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 flavor
    - The 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?

CernVM
Software Appliance

Virtual File Systems

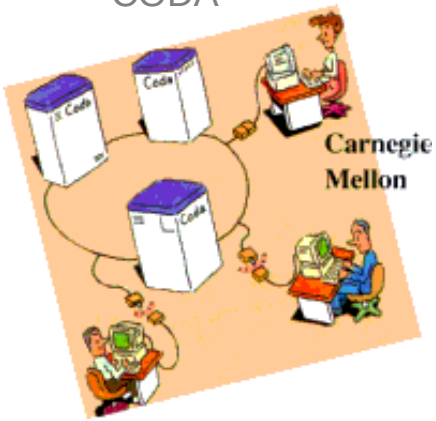
- Once we are allowed to run a Virtual Machine, whole new world of possibilities opens
  - We can (re)use a lot of code which was previously is system/kernel domain
  - We can build dedicated VMs with special kernel modules built in to support various fancy file systems
  - For example, HTTP-FUSE-CLOOP file system could be used for software distribution (Xennopix = Xen + Knoppix, boots from the 5Mb image!)



http://www.linux-kongress.org/2006/abstracts.html#4_2_2

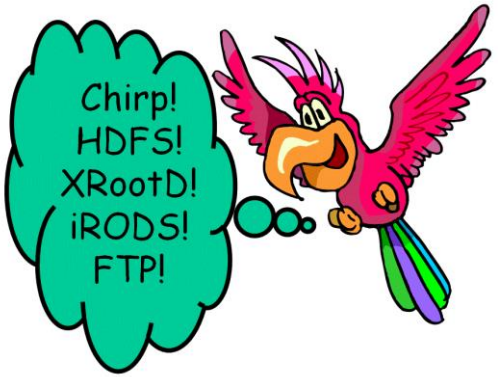
Virtual File Systems

CODA



http://www.coda.cs.cmu.edu/

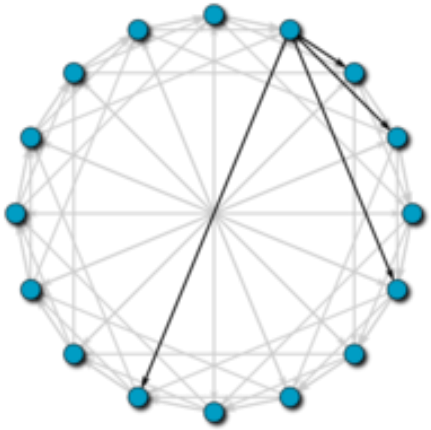
Chirp/Parrot/Grow-FS



https://ccl.cse.nd.edu/

https://www.openafs.org

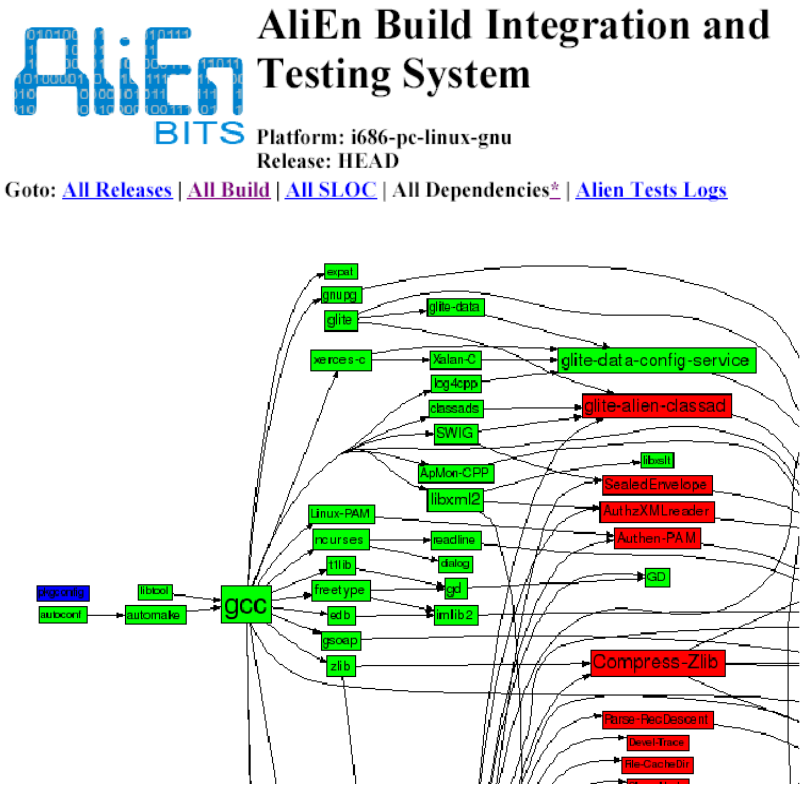
/

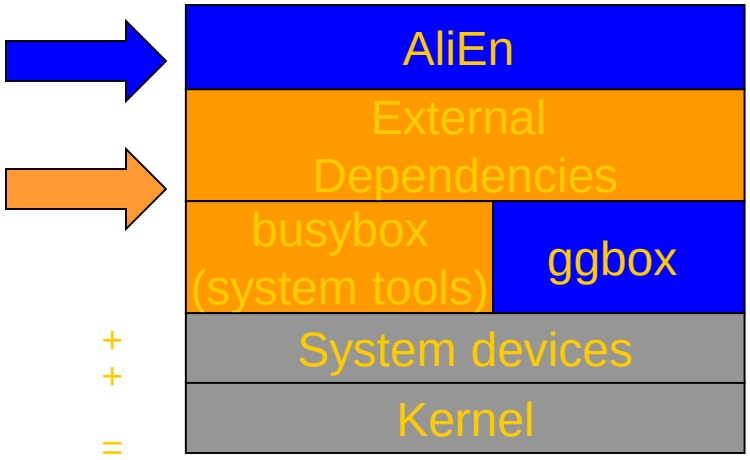


IgorFS

http://doi.ieeecomputersociety.org/10.1109/P2P.2008.19

AliEnX (2006)



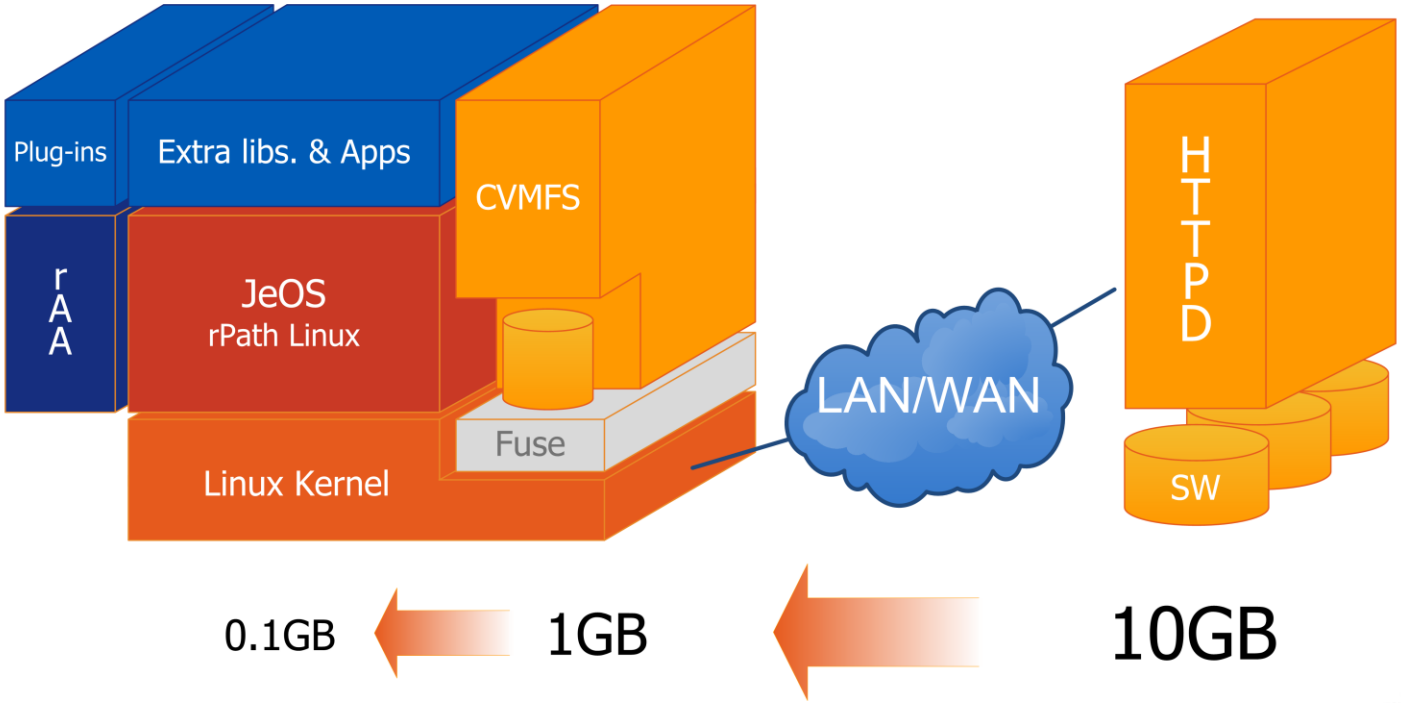


Grid Virtual Machine

From AliEnX to CernVM

- I realized that the old AliEn brand name has got some bad karma and decided to rename the project to CernVM.
- Tribute to IBM 3090 mainframe (CERNVM) that was switched off in 1996, marking the end of the mainframe era at CERN that had lasted 40 years.
- The name seemed to be perfectly adequate for the project and even DNS name cernvm.cern.ch was available.
- The brand name is IMPORTANT!

CernVM

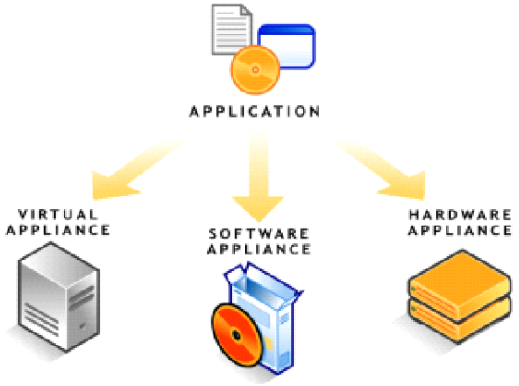


- Prototype CernVM File System (CVMFS) was derived from Parrot (http://www.cctools.org) and its GROW-FS code base and adapted to run as a FUSE kernel module adding extra features like:
  - possibility to use multiple file catalogs on the server side
  - transparent file compression under given size threshold
  - dynamical expansion of environment variables embedded in symbolic links

Leandro Franco

Key Building Blocks

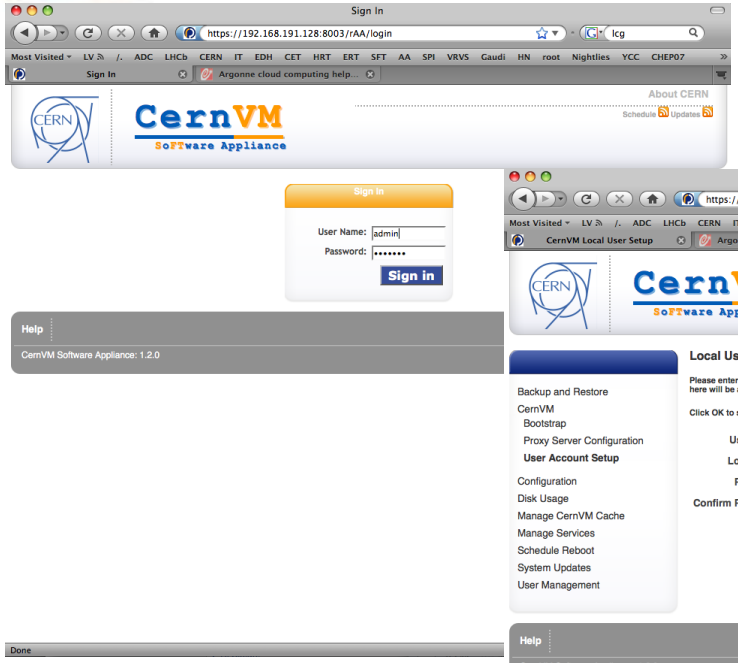
- rPath Linux 1 (www.rpath.org)
  - Slim Linux OS binary compatible with RH/SLC4
- rAA - rPath Linux Appliance Agent
  - Web user interface
  - XMLRPC API
- rBulder
  - A tool to build VM images for various virtualization platforms
- CVMFS - CernVM file system
  - Read only file system optimized for software distribution
    - Aggressive caching
  - Operational in offline mode
    - For as long as you stay within the cache



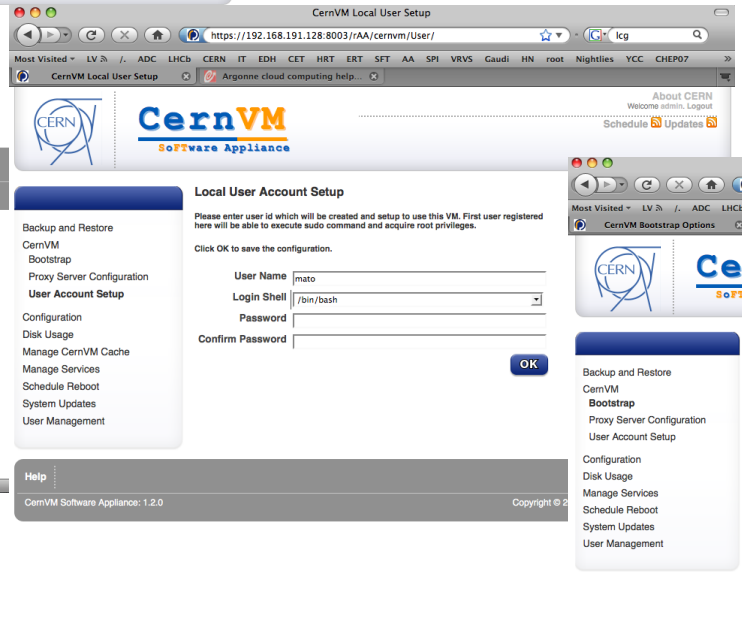
Build types

- 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 Apliance
- Xen Enterprise Virtual Appliance
- Virtual Iron Virtual Appliance
- Parallels Virtual Appliance
- Amazon Machine Image
- Update CD/DVD
- Appliance Installable ISO

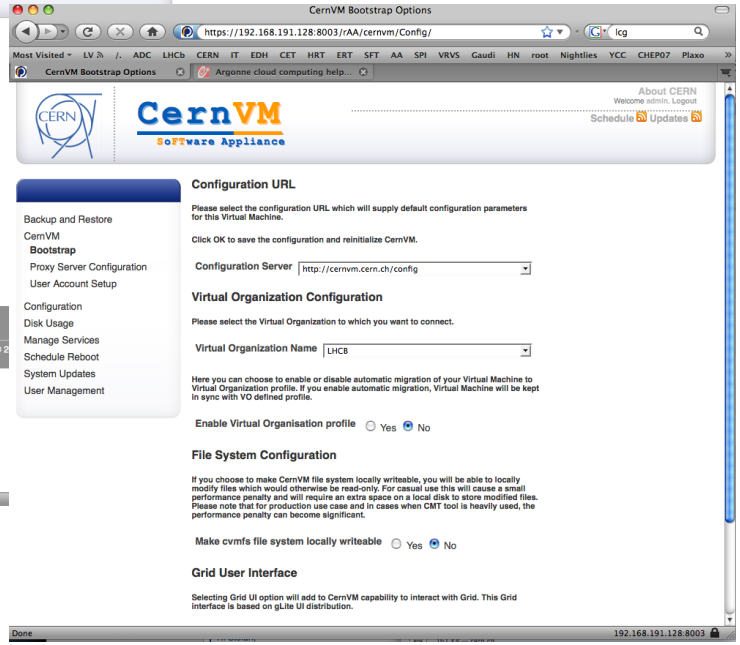
As easy as 1,2,3



1. Login to Web interface



2. Create user account



3. Select experiment, appliance flavor and preferences

White Paper R&D

- We were incredibly lucky to be ready to articulate and propose an interesting project exactly in the moment when the call for proposals was made
- The R&D framework gave us the opportunity to think out of the box
- Even the modest funding allowed us discover and hire some truly excellent people
- This also allowed us to speak to the experiments (and CERN/IT) and be actually heard
- In short, that was a great time

Project Goals and Deliverables

- Evaluation of the available virtualization technologies
  - Understand and validate technologies by checking their performance, usability and platform constraints
  - Evaluation of the tools to build and manage 'virtual appliances'
- Deployment of a read-only distributed file system with aggressive caching schema
  - Essential to avoid any pre-installation of layered software
  - Validate performance, scalability and usability
- Collect requirements from experiments and confront them with available technologies
  - Suggest optimal choice for given use case
- Provide prototypes of data analysis virtual appliances for at least 2 experiments
  - Asses their suitability for providing portable and easy to install data analysis environments
  - Assist experiments in adapting their software process to this platform

CernVM Team

Carlos Aguado

Service Infrastructure Development and Support

System Development Engineer at Amazon Web Services

Artem Harutyunyan

Co-Pilot Architect and Developer

Director of Engineering at Mesosphere

Jakob Blomer

CVMFS developer

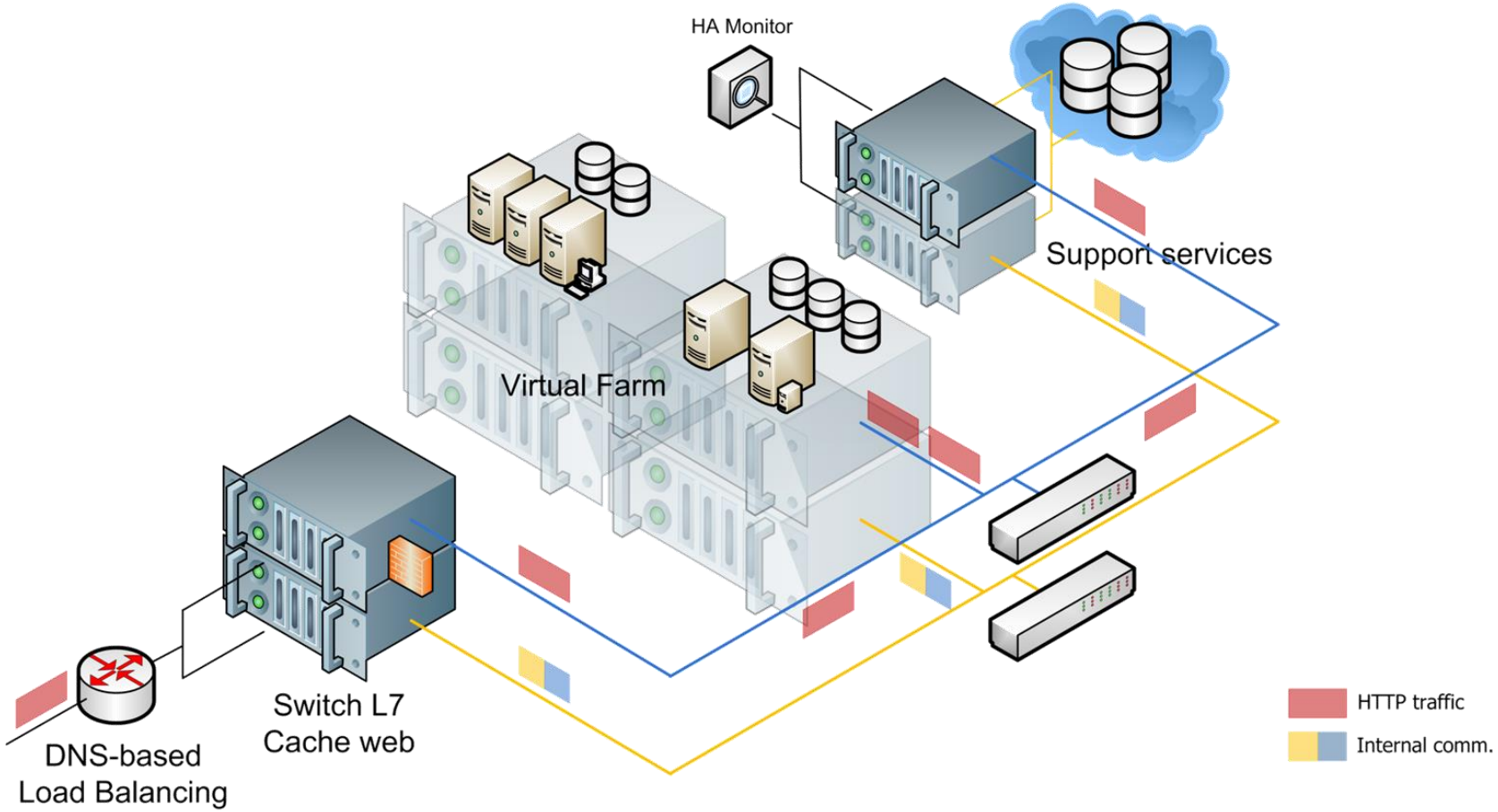
Now CERN staff in EP-SFT

Ioannis Charalampidis

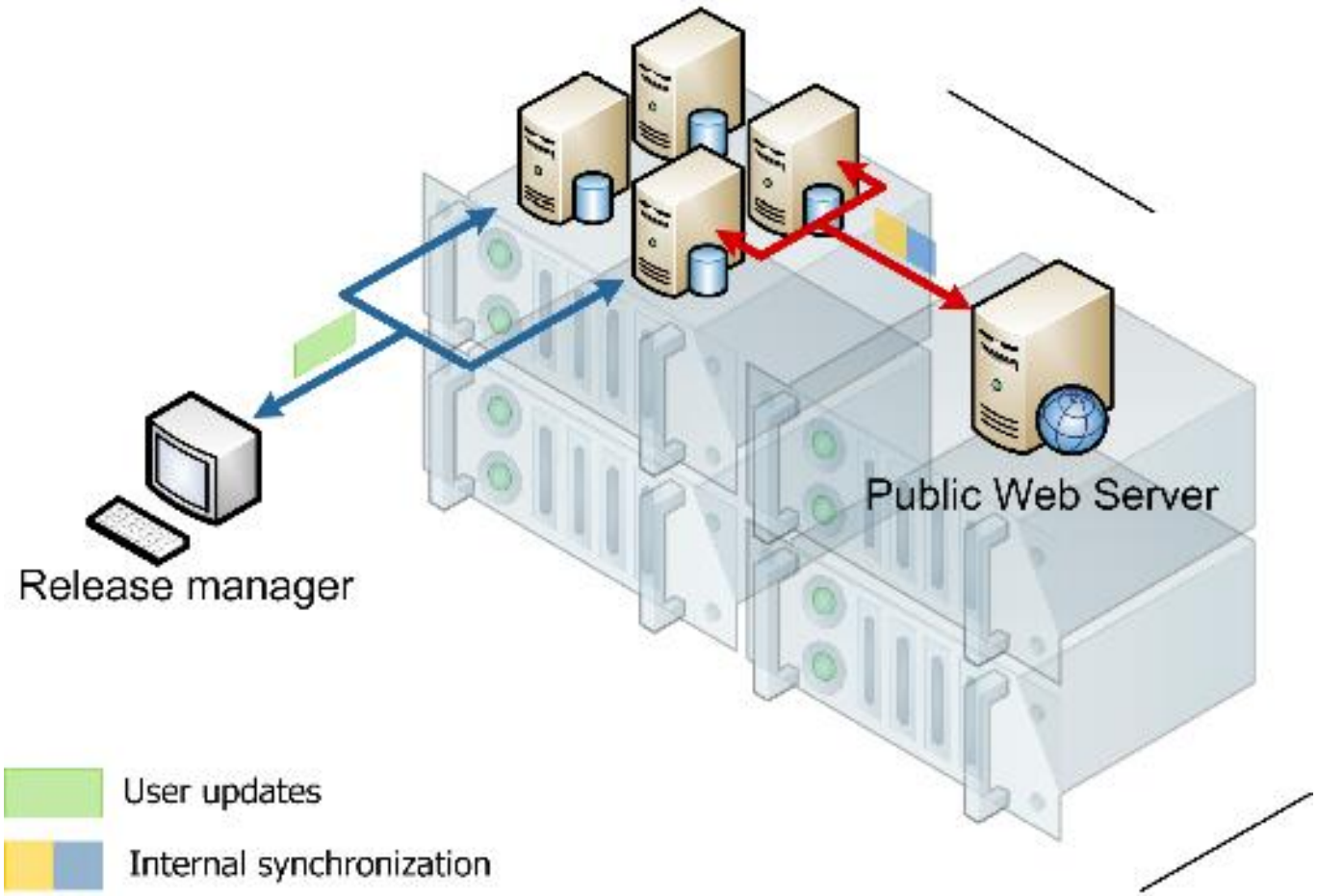
CernVM Online, micro CernVM, Test4Theory..

Software Engineer at Mesosphere

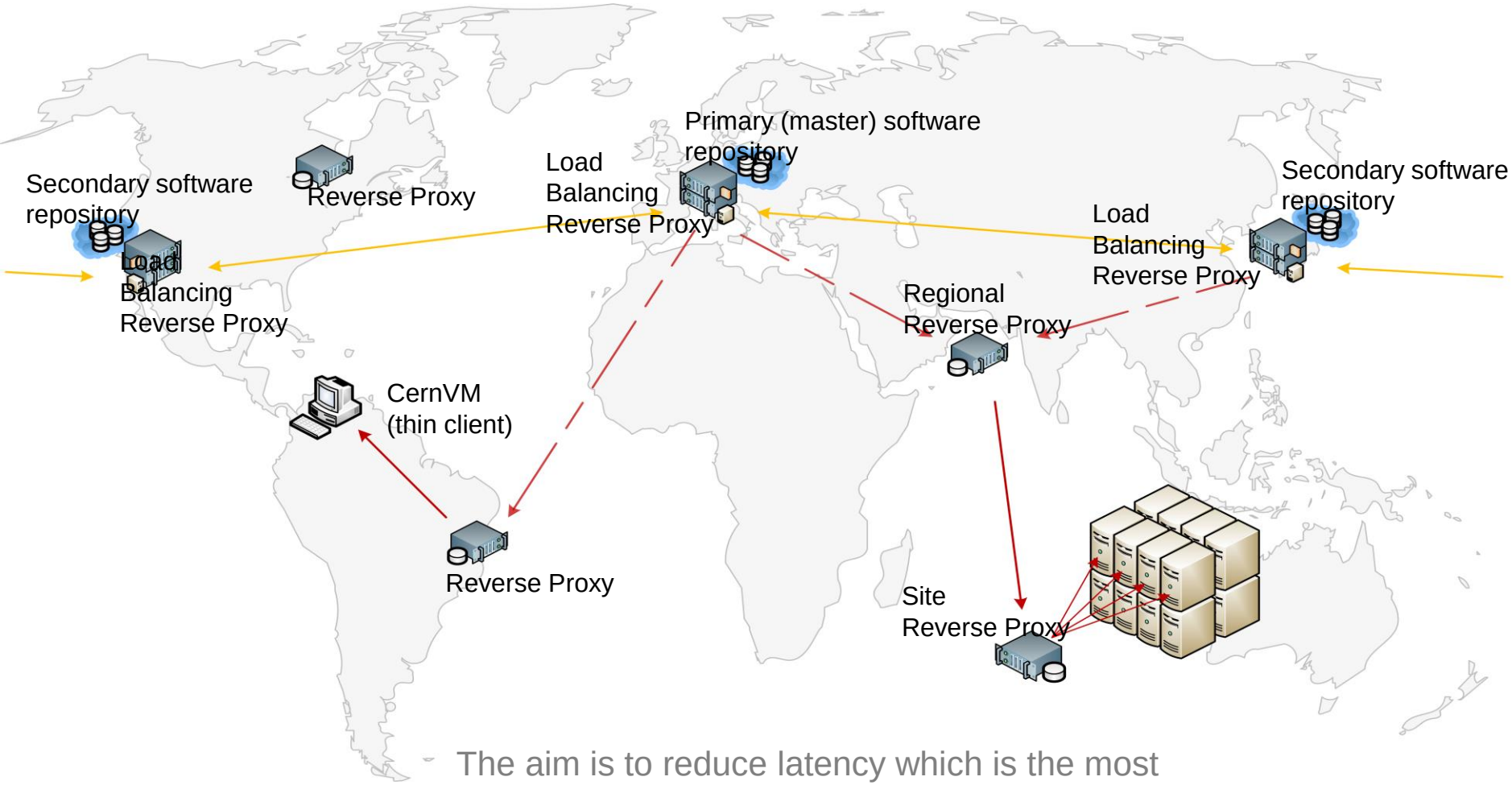
Scalable infrastructure



Publishing releases



Deployment picture



The aim is to reduce latency which is the most important issue for distributed network file systems

Sales pitch

- If you believe that it is time to jump on bandwagon and expolore directioin in which major software industry players are going and if you have one of these problems:
  - You work for LHC experiment and its software is not compatible with your favorite hardware or s/w platform running on your laptop/desktop
  - You do not want to spend time to manually keep software up to date
  - Your want to profit from the latest developments in CPU technology and use your multi/many core CPU to its maximal potential without modifying your application
  - You want to share spare CPU cores/cycles with others
  - You want to run your software on voluntary resources beyond the current Grid

…then CernVM might be what you are looking for.

- Version 1.0 available at http://rbuilder.cern.ch/project/cernvm/releases

CernVM
Software Appliance

CernVM Users

~6000 unique IP addresses

Expanding use cases

- How can profit from modern (multi and many core) CPU power of the user workstations and use them at least as a development platform?
- How to decouple application and infrastructure lifecycles and assure a homogeneous job execution environment compatible with one in which application was developed?
- How to preserve experiment software and keep it usable and accountable over many years?
- How to use unmodified HEP software and harvest CPU cycles in volunteer computing environment?

Buncic, P., Aguado Sánchez, C., Blomer, J. et al. Eur. Phys. J. Plus (2011) 126: 13.
https://doi.org/10.1140/epjp/i2011-11013-1

CernVM/FS

Ian Collier

- By the end of 4 year R&D project we were still struggling to get clear support from at least two LHC experiments
  - ATLAS, LHCb and CMS were interested and using CVMFS
  - We had many users from across all 4 collaborations using CernVM but only LHCb would speak loudly about that

- Ian Collier from RAL approached us suggesting to factor out CVMFS from CernVM project
  - It was a hard pill to swallow but we accepted it under condition that it remains officially called CernVM File System
  - He did an excellent work on convincing the HEPIX community that CVMFS meets the security requirements and put a convincing arguments backed up by operational evidence that it solves many problems of the large sites.
  - This is what finally cut it, CVMFS was accepted across the board and CernVM just won't die ☺

CernVM
Software Appliance

CernVM Today

It’s all about the Team...

Rene Meusel

Now in Mesosphere

Nikola Hardi

In the audience ☺

Radu Popescu

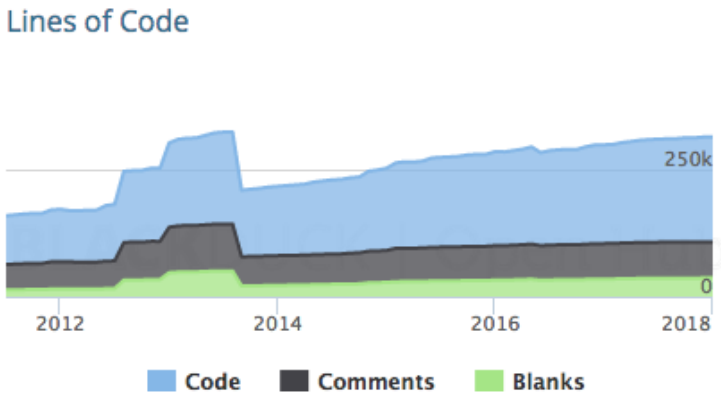
Jakob Blomer

Gerardo Ganis

- This project always had a great team of core developers and the tradition continues...

Enthusiastic community...

- **In a Nutshell, cvmfs...**
- ... has had 12,836 commits made by 31 contributors
  representing 207,047 lines of code
- ... is mostly written in C++
  with an average number of source code comments
- ... has a well established, mature codebase
  maintained by a large development team
  with stable Y-O-Y commits



30 Day Summary
Dec 14 2017 — Jan 13 2018
134 Commits
3 Contributors

12 Month Summary
Jan 13 2017 — Jan 13 2018
1855 Commits
Down -565 (23%) from previous 12 months
8 Contributors
Down -2 (20%) from previous 12 months

- https://www.openhub.net/p/cvmfs
- .

CernVM
Software Appliance

...and many happy Users!

Conclusions

- CernVM did not happen completely by chance
  - It was a premeditated success but we were incredibly lucky to have all ideas in the pocket and be ready to propose the R&D project when a rare opportunity appeared.
  - The formal R&D status assured necessary visibility and reach out to the users and the modest funding allowed us to grow and build an excellent team.
  - R&D should be a norm and not exception in our environment.
- While the most of development work is already done, there will always some new buzzwords to catch up with, security threats or a brand new OS release...
  - Change is the only constant thing in this world... and that bring the loads of support work which is not always fun for developers.
- The challenge will remain to keep the team in place
  - Help and cooperation from the extended community will be crucial for a long-term project survival.