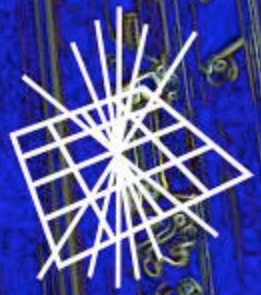


GridPP

UK Computing for Particle Physics

CernVM-FS Production Service Status

HEPiX Spring 2011, GSI

2-6 May

Ian Collier, STFC-RAL

ian.collier@stfc.ac.uk

Science & Technology Facilities Council

e-Science

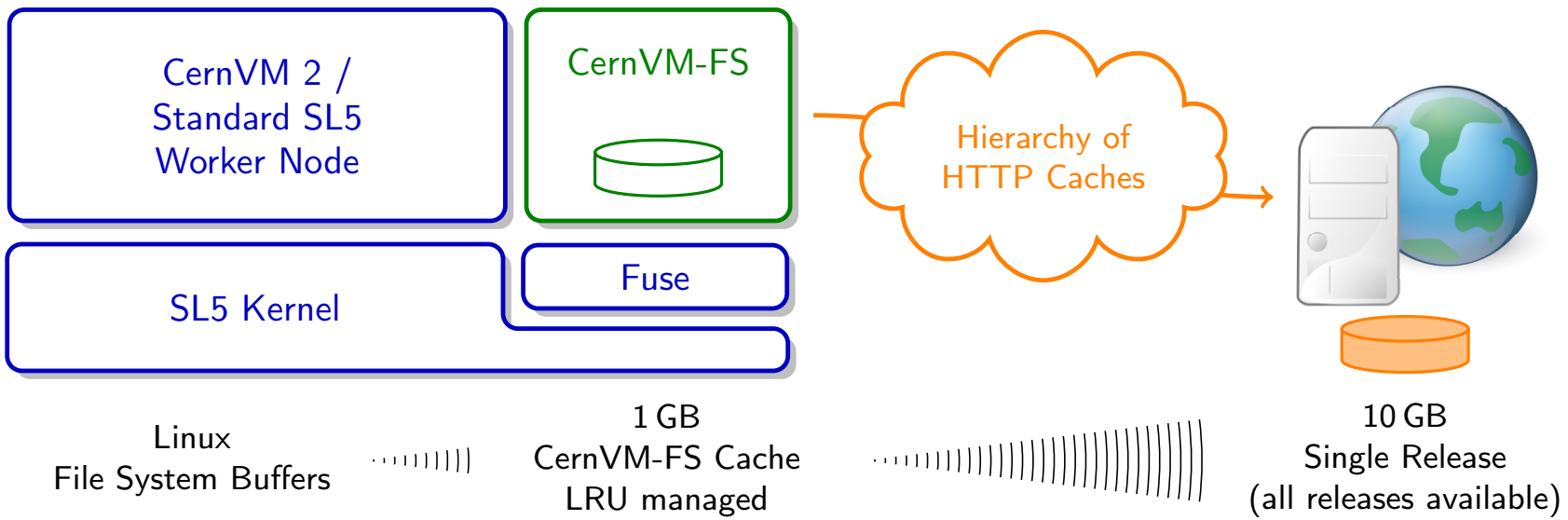
Overview

- CernVM-FS Introduction
  - Concepts
  - Performance
- CernVM-FS Service Status
- CernVM-FS Site deployment
  - Client Config
  - Squids...
- Summary

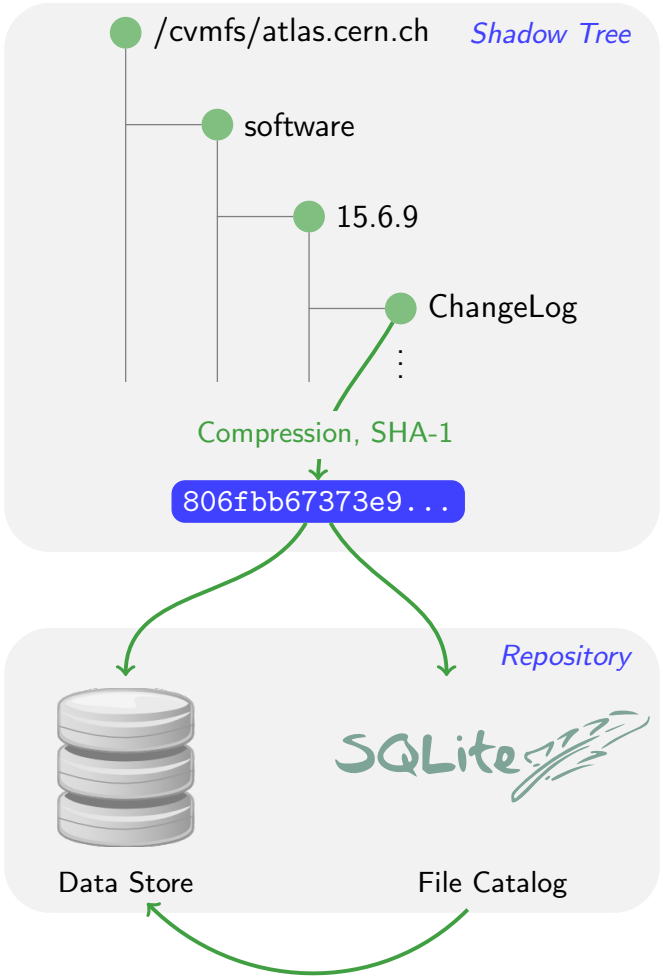
CernVM-FS - Introduction

CVMFS Principle:

Virtual software installation by means of an HTTP File System

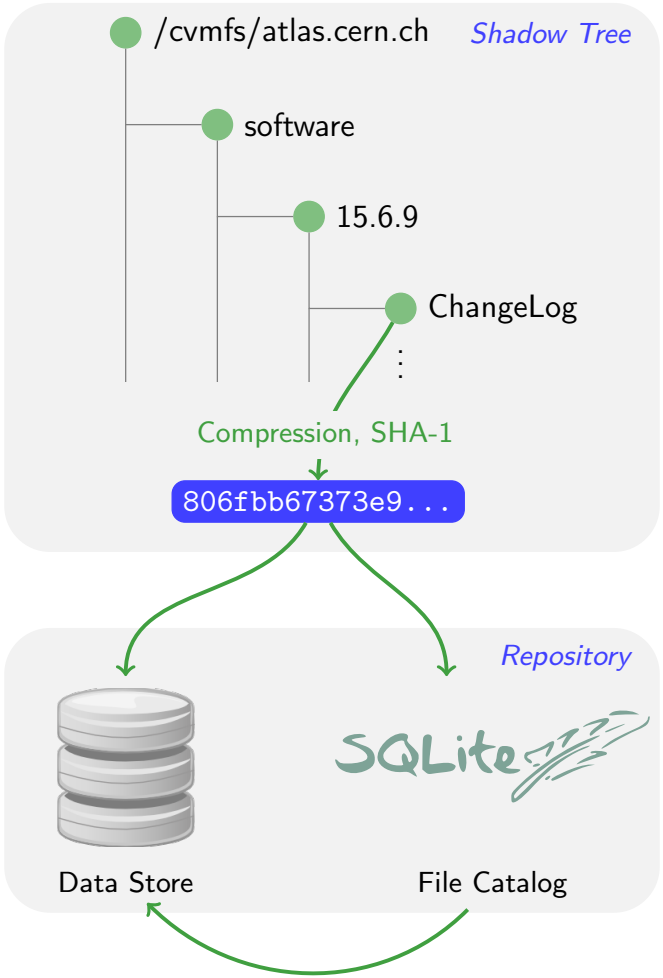


CernVM-FS - Introduction



- **Data Store**
  - Compressed Chunks (Files)
  - Eliminates Duplicates
  - Never Deletes
- **File Catalog**
  - Directory Structure
  - Symlinks
  - SHA1 of Regular Files
  - Digitally Signed
  - Time to Live
  - Nested Catalogs

CernVM-FS - Introduction



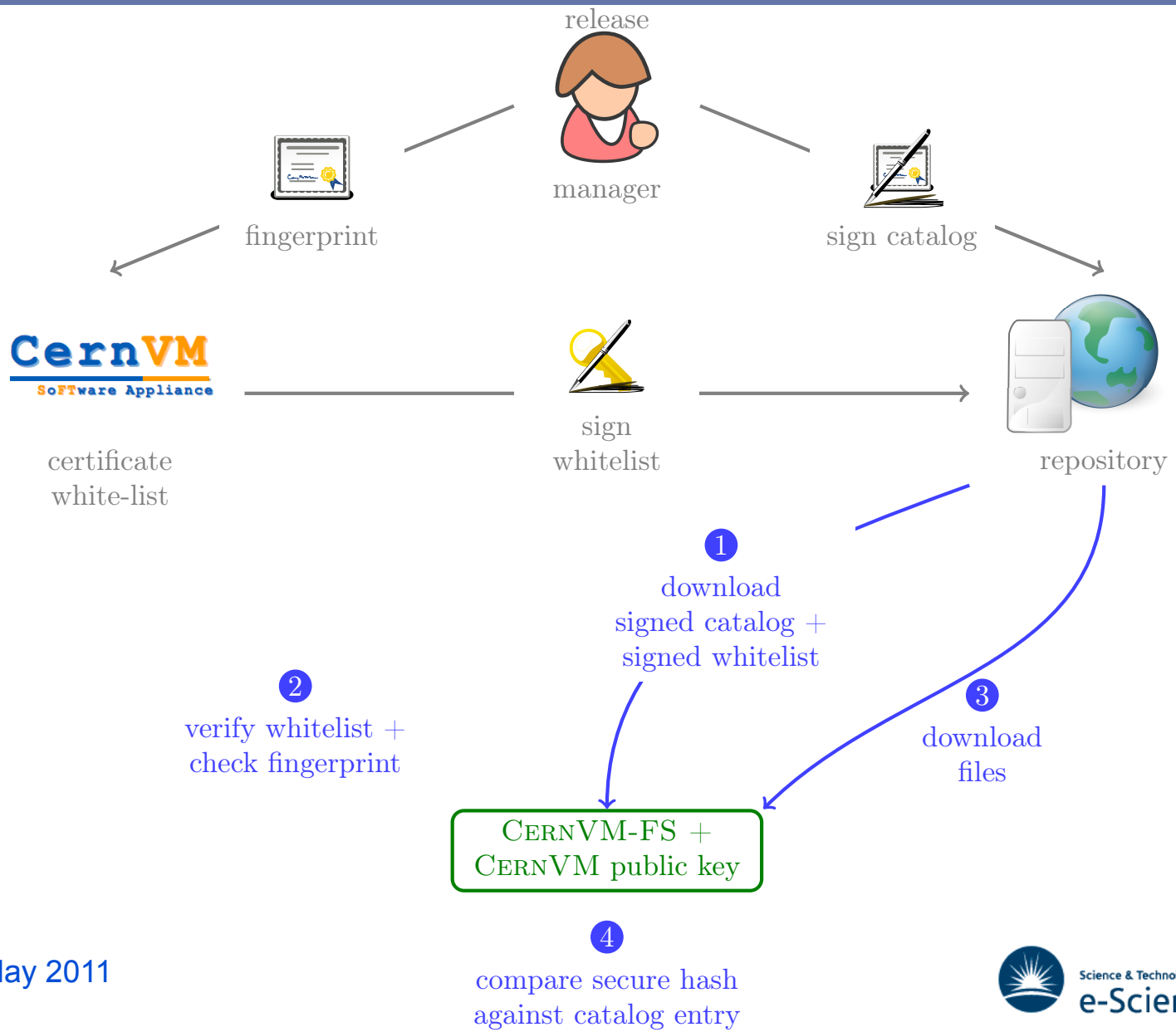
- **Data Store**
  - Compressed Chunks (Files)
  - Eliminates Duplicates
  - Never Deletes
- **File Catalog**
  - Directory Structure
  - Symlinks
  - SHA1 of Regular Files
  - Digitally Signed
  - Time to Live
  - Nested Catalogs

⇒ Immutable Files, trivial to check for corruption

CernVM-FS - Introduction

- CVMFS client:
  - Provides a **`/cvmfs/`** filesystem area.
  - Files served from a web-server.
  - File accesses are intercepted by fuse.
  - Metadata operations (**`ls`**, **`cd`**, ...) work on a signed sqlite database - downloaded from web-server.
  - File operations (**`cat`**, **`open`**, ...) trigger download of file from web server.
- Lot's of caching:
  - On batch workers between jobs.
  - On intermediate squid servers.
- Everything transparent to user

CernVM-FS - Release Process



CernVM-FS within WLCG

- CVMFS is set to replace current software installation methods.

- Current method:
  - Run <VO>sgm job via grid
  - Write files within job to some shared storage.
  - Validate software.
  - Publish tag in BDII.
  - Process has to be repeated at every site.
    - Process has to be debugged at every site.

CernVM-FS Advantages

- Install once (on stratum zero)
  - Files appear everywhere across WLCG.
- This can be many days faster.
- Hopefully less variation across sites.
  - Common path /cvmfs/...
  - Very few variables:
    - Cache size, Squid QOS.
- Same install bugs everywhere - fix once.
  - e.g LHCb have had problems with CMT usage on /cvmfs but at least it's everywhere.
- Some sites are struggling to provide scaled NFS/AFS.
  - squids scale very well - load seems to be small on squids

CernVM-FS Performance

- So, this all looks pretty good for the site admins, and not so bad for the VO release managers - surely there must be a downside - performance perhaps?

CernVM-FS Performance

- So, this all looks pretty good for the site admins, and not so bad for the VO release managers - surely there must be a downside - performance perhaps?

- Well, Victor Méndez at PIC tested just last month....

CernVM-FS Performance

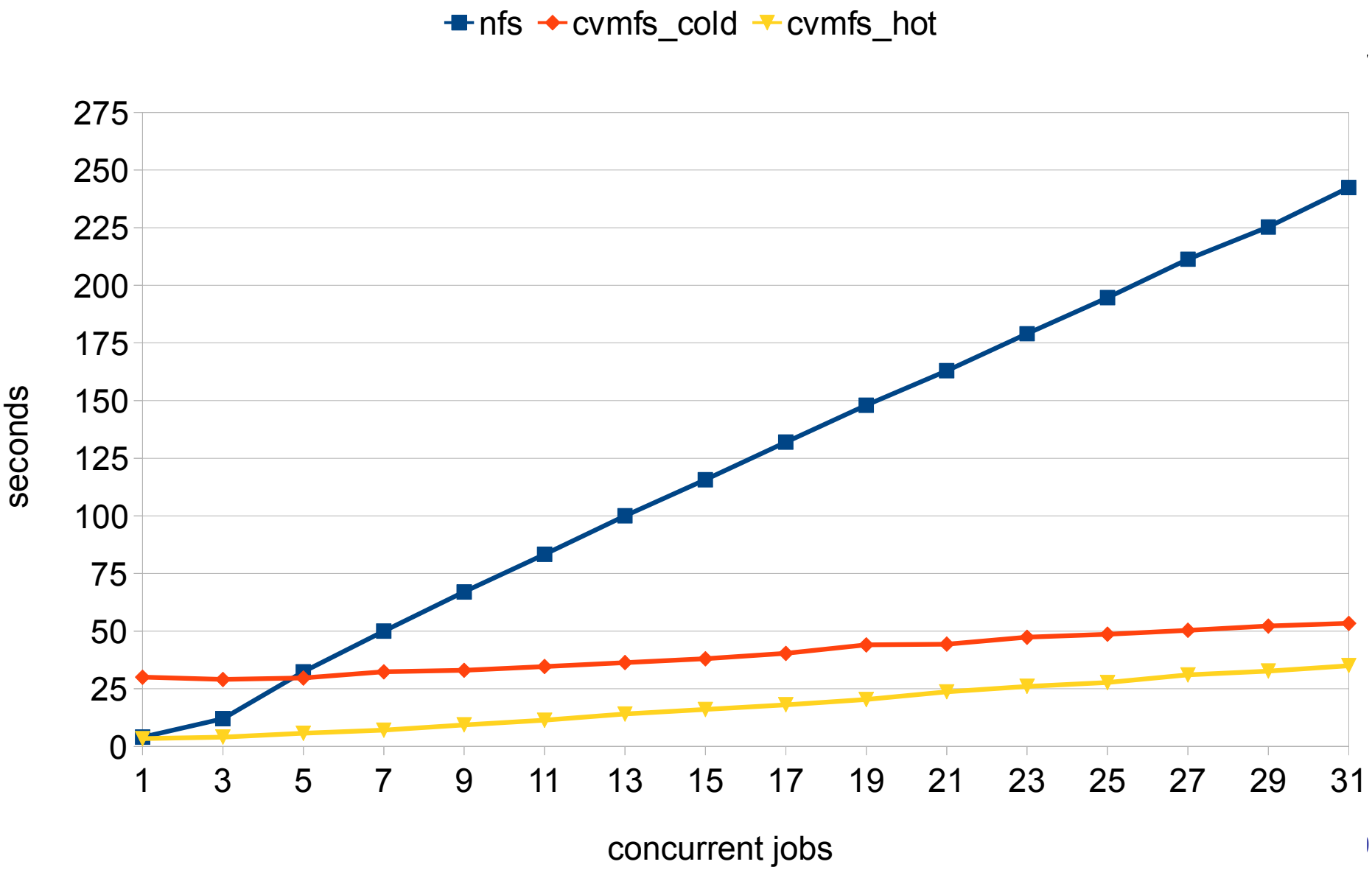
- So, this all looks pretty good for the site admins, and not so bad for the VO release managers - surely there must be a downside - performance perhaps?

- Well, Victor Méndez at PIC tested just last month....

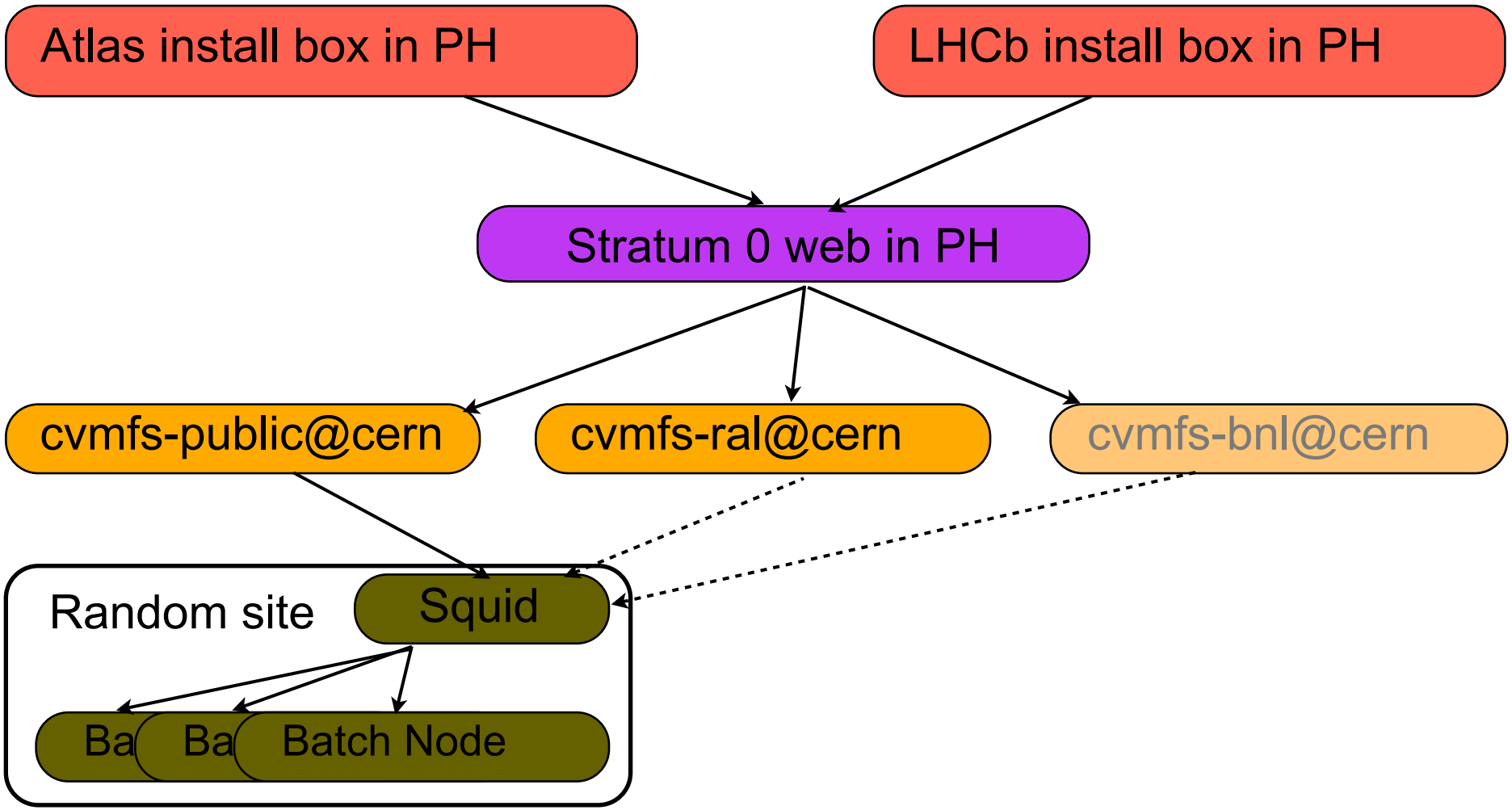
- Metrics measured:
  - Execution time for SetupProjec Gauss v38r9 - the most demanding phase of the job for the software area (huge amount of stat() and open () calls)
  - Dependence on the hot and cold local cache. Cache size 174MB (catalog 148MB), 1 job run: cold hit ratio = 0.54, another run with hot cache hit ratio = 0.99
  - Comparison with standard NFS shared area
  - Dependence on the number of concurrent jobs

GridPP
UK Computing for Particle Physics

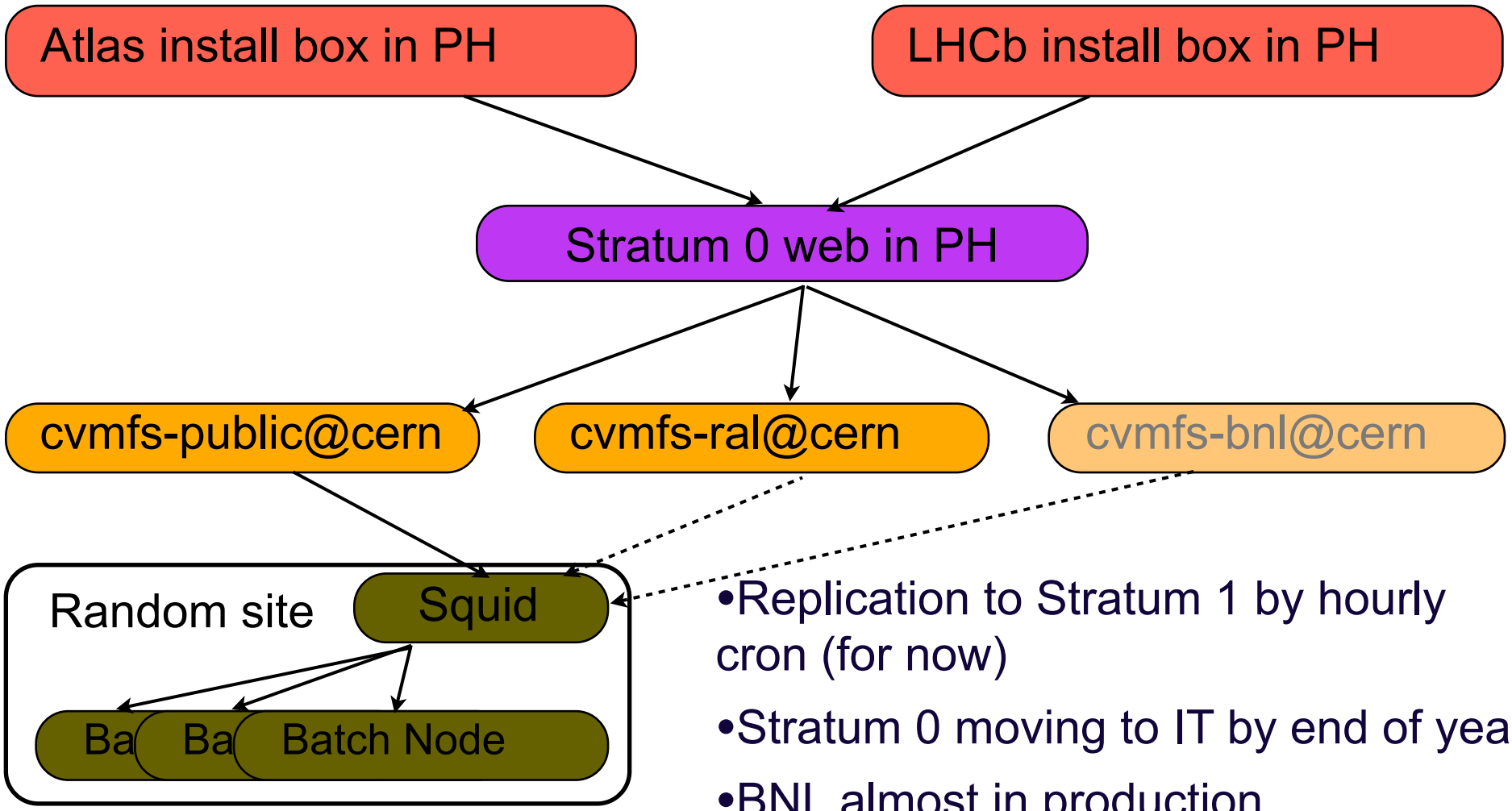
CernVM-FS Performance



CernVM-FS Service Status

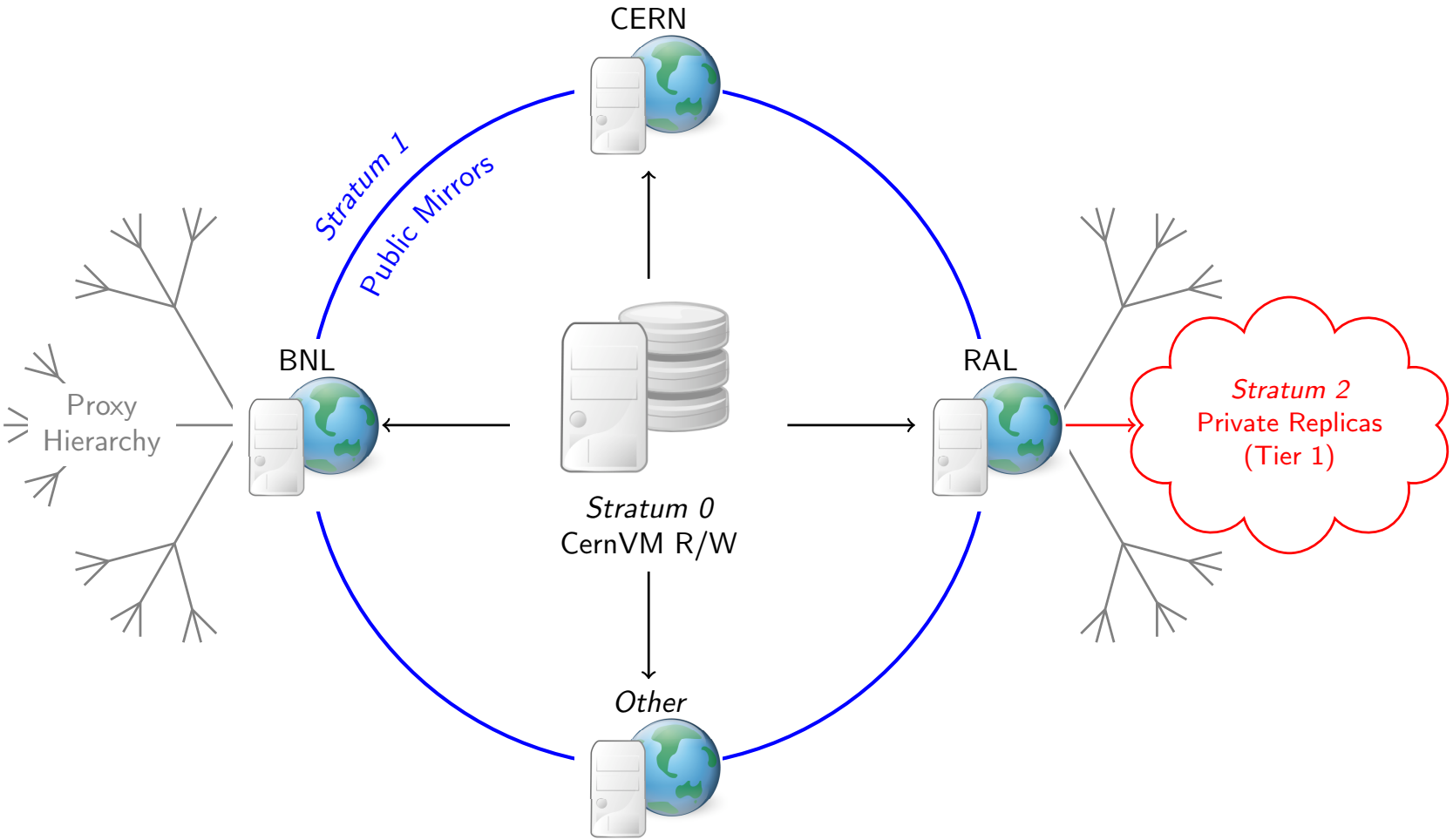


CernVM-FS Service Status



- Replication to Stratum 1 by hourly cron (for now)
- Stratum 0 moving to IT by end of year
- BNL almost in production

CernVM-FS Service Backend



Mirror servers: Web servers listening on port 80, 8000

Proxy servers: Local load-balanced Squid forward proxy (SL5 Squid)

CernVM-FS Site Deployment

You need:

- Some worker nodes
- Some rpms installed on your WNs:
  - `cvmfs` (currently 0.2.61-1)
  - `cvmfs-init-scripts` (currently version 1.0.8-2)
  - `cvmfs-keys` (currently version 1.1-1)
  - `cvmfs-auto-setup` (optional - aimed at T3s)
  - `fuse` (should be at latest version from OS)
  - `fuse-libs` (should be at latest version from OS)
  - `autofs` (should be at latest version from OS)

CernVM-FS Site Deployment

- You also should have:
  - A squid cache (ideally two or more for resilience rather than load)
  - Configured (at least) to accept traffic from your site to one or more cvmfs repository servers
  - You could use exisiting frontier-squids

CernVM-FS Site Deployment

What you need to configure

The repositories required at your site (atlas, cms, lhcb, etc.)

The source repository URL(s) to use

`http://cernvmfs.gridpp.rl.ac.uk/opt/@org@` `http://cvmfs-stratum-one.cern.ch/opt/@org@`

Ideally set one primary and at least one secondary - failover is built in to the client

The size of the cache(s) on your WNs

You need an entry in the autofs master map **`/etc/auto.master`** (**`/cvmfs /etc/auto.cvmfs`**) and also **`/etc/fuse.conf`** (**`user_allow_other`**)

`VO_<voname>_SW_DIR`

i.e: **`VO_LHCB_SW_DIR, VO_ATLAS_SW_DIR`** etc.

CernVM-FS Site Deployment

What you need to configure

There is also `/usr/bin/cvmfs_config` which can be run manually or be triggered by the `cvmfs-auto- config` rpm (aimed at T3s), since we carry out system config with Quattor I don't have direct experience.

```
cvmfs_config

Common configuration tasks for CernVM-FS

Usage: /usr/bin/cvmfs_config {setup [nouser]
[nocfgmod] [noservice] [nostart]| chksetup|
```

CernVM-FS Site Deployment

Configuration files

/etc/cvmfs

```
[cvmfs]# ls -R *
config.sh default.conf default.local

config.d:
alice.cern.ch.conf          cms.cern.ch.conf      lhcb.cern.ch.conf
atlas.cern.ch.conf          cms.cern.ch.local     lhcb.cern.ch.local
atlas.cern.ch.local         grid.cern.ch.conf     na61.cern.ch.conf
atlas-condb.cern.ch.conf    hepsoft.cern.ch.conf  sft.cern.ch.conf
atlas-condb.cern.ch.local   hone.cern.ch.conf
boss.cern.ch.conf           lcd.cern.ch.conf

domain.d:
cern.ch.conf cern.ch.local

keys:
cern.ch.pub
```

The ones in red are added by the site (RAL in this case)

Specific Changes

`/etc/domain.d/cern.ch.local`

```
CVMFS_SERVER_URL=http://
cernvmfs.gridpp.rl.ac.uk/opt/@org@;http://cvmfs-stratum-
one.cern.ch/opt/@org@
```

Here we tell cvmfs that for repositories hosted originally at CERN (ie all the WLCG experiment software, first look to the replica at RAL, and fail over to CERN. We can add BNL when it becomes available.

But as other cvmfs servers become available (for nightlies, non lhc VOs etc.) this mechanism will allow multiple sources to be accessed from the same WN (specified in `/etc/domain.d/gridpp.rl.ac.uk.conf` perhaps.

CernVM-FS Site Deployment

Specific Changes

```
/etc/cvmfs/config.d/atlas.cern.ch.local

CVMFS_QUOTA_LIMIT=10000
```

Here we tell cvmfs that for Atlas, the repository quota limit is 10GB instead of the 5GB in **`/etc/cvmfs/default.local`**, and so on for the other repositories used if required.

CernVM-FS Site Deployment

Tools

```
cvmfs_talk
cvmfs_config chksetup
cvmfs_config showconfig <repository>.cern.ch
service cvmfs status
service cvmfs probe
service cvmfs restartclean
service cvmfs restartautofs
```

cvmfs_talk allows us to ‘interrogate’ the caches cvmfs_config shows/verifies the configuration the service commands allow granular stopping/starting/ restarting/probing of components

CernVM-FS Site Deployment

Some Links	
Download	http://cernvm.cern.ch/portal/downloads
Yum	http://cvmrepo.web.cern.ch/cvmrepo/yum
News	http://twitter.com/cvmfs
Bug Tracker	https://savannah.cern.ch/bugs/?group=cernvm
Mailing list	cvmfs-talk@cern.ch

Summary - Since last time

- CernVM-FS is being used in production
  - LHCb at many Tier 1s (including RAL)
  - Atlas
    - Tier3s, RAL, PIC, NIKHEF, CERN, Wuppertal, QMUL, Munich, Lancester, Dortmund, JINR, . . .
  - Others of course, and some running their own servers
- Core service supported at CERN
- Replicas in place at CERN, RAL and BNL
  - for resilience not load
- In SL - right Troy?
- Set to transform WLCG VO software distribution

Questions