



CernVM[-FS] Status Report

GDB October 2014

Gerardo Ganis, René Meusel

Agenda

1. Statistics of CERN-hosted Repositories
2. Successful Migration of CernVM-FS Stratum0s to 2.1.x
3. Alternative Storage Backends for Stratum 0 and 1
4. Simplified Client Configuration
5. Status of Garbage Collection in CernVM-FS
6. CernVM Status Update



CernVM-File System

CERN-hosted Repository Statistics

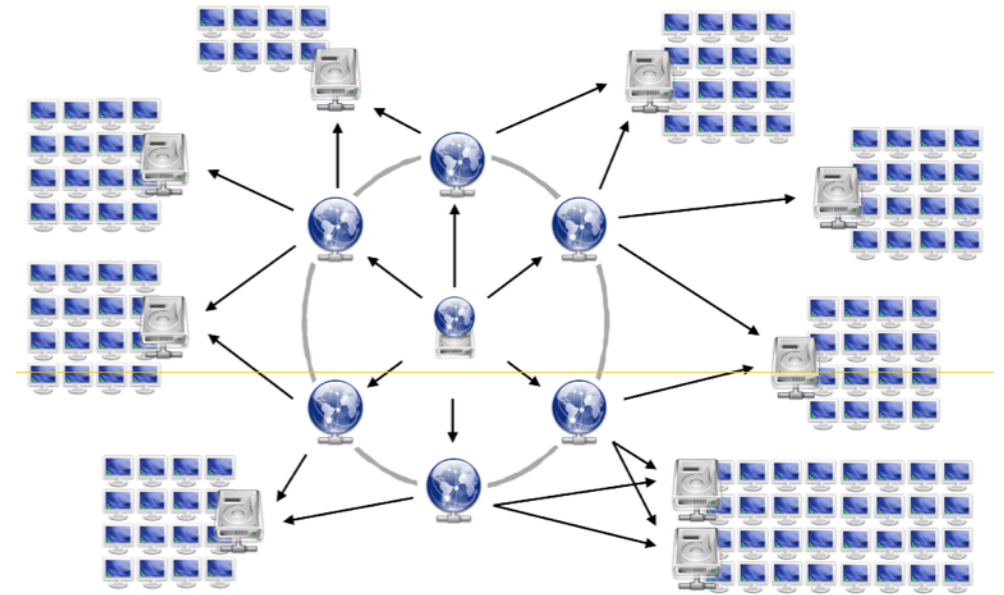
- LHC experiment software repositories doubled in size through the last 24 months

Repository	Files	Refer. Objects	Volume	avg. File Size	
atlas.cern.ch	34'500'000	3'700'000	2.1 TiB	66.2 kiB	Mainly Software
cms.cern.ch	30'600'000	4'800'000	0.9 TiB	33.1 kiB	
lhcb.cern.ch	13'600'000	4'600'000	0.5 TiB	41.9 kiB	
alice.cern.ch	5'900'000	240'000	0.5 TiB	90.7 kiB	
ams.cern.ch	2'900'000	1'900'000	1.9 TiB	0.7 MiB	Software + Conditions Data
alice-ocdb.cern.ch	700'000	700'000	0.1 TiB	0.2 MiB	Conditions Data
atlas-condb.cern.ch	8'000	7'800	0.5 TiB	60.8 MiB	

Effective: August 2014 (already presented at ACAT '14)

CernVM-FS Versions and Components

- Components
 - Installation Box / Release Manager Machine
 - Stratum 1 software (replication tools)
 - CernVM-FS client
- CernVM-FS 2.1.19 (released: end of May '14 - stable)
 - Consolidation release after CernVM-FS 2.1.17
- CernVM-FS 2.1.20 to be released before December '14
 - CVMFS_CONFIG_REPOSITORY
 - Experimental backend storage driver for S3
 - Experimental garbage collection on the server
 - Web API for Stratum 1 Servers
 - ...



New Features in CernVM-FS Server 2.1.x

Transactional Repository Updates

File System Snapshots

Snapshot History Database

Repository Rollbacks on Stratum 0

Parallel File Processing

Chunking of Large Files

Alternative Storage Backends




Multiple Repositories on one Installation Box

Aggregated Repository Statistics

Abandon 'Shadow Directory' on Installation Box

[...]

Migration Plan (*.cern.ch Repositories)

- Preconditions for server migration
 - All clients on CernVM-FS 2.1.x 
 - Stratum 1 replication servers on CernVM-FS 2.1.x 
 - Automatic repository migration available in CernVM-FS (First appeared in version 2.1.15 - fully stabilised in 2.1.20) 
- First migrated repository: **geant4.cern.ch** (April 11th)
 - lead to a couple of minor fixes in CernVM-FS 2.1.19
- Migrated other “small” repositories (April, May)
boss.cern.ch, **belle.cern.ch**, **grid.cern.ch**, **na49.cern.ch**, **na61.cern.ch**
- CernVM-FS 2.1.19 must be installed on all sites (by August 5th)
(decided in: WLCG Ops Meeting - June 5th [1])
- Migrated large repositories (August, September)
sft.cern.ch, **ams.cern.ch**, **atlas.cern.ch**, **alice.cern.ch**, **atlas-condb.cern.ch**, **cms.cern.ch**, **lhcb.cern.ch**

CernVM-FS Server Migration Status

- Overall smooth transition with only minor issues
 - Sporadic outages of some Tier 3 sites and individual users (still running CernVM-FS 2.0.x clients)
 - Test4Theory (LHC@Home 2.0) outage after migrating grid.cern.ch (master machine was running CernVM-FS 2.0.x)
- Minor adaptations needed for scripts on release manager machines (New: `cvmfs_server transaction`)
- One known issue:
 - Release manager machine on CernVM-FS 2.1.19 can get stuck in illegal mounting state after publish process interruption or reboot (Fix: <http://cernvm.cern.ch/portal/cvmfs/fix-failed-remount>)

CernVM-FS Server Migration Status

- Migration to CernVM-FS 2.1.x successfully completed on September 2nd for all CERN-hosted repositories

- Note: No more CernVM-FS 2.0.x components at CERN!

- Detailed schedule for reference:

April 11th

geant4.cern.ch

April 24th

boss.cern.ch

belle.cern.ch

May 5th

grid.cern.ch

na49.cern.ch

na61.cern.ch

August 5th

sft.cern.ch

ams.cern.ch

August 12th

atlas.cern.ch

alice.cern.ch

August 28th

atlas-condb.cern.ch

September 2nd

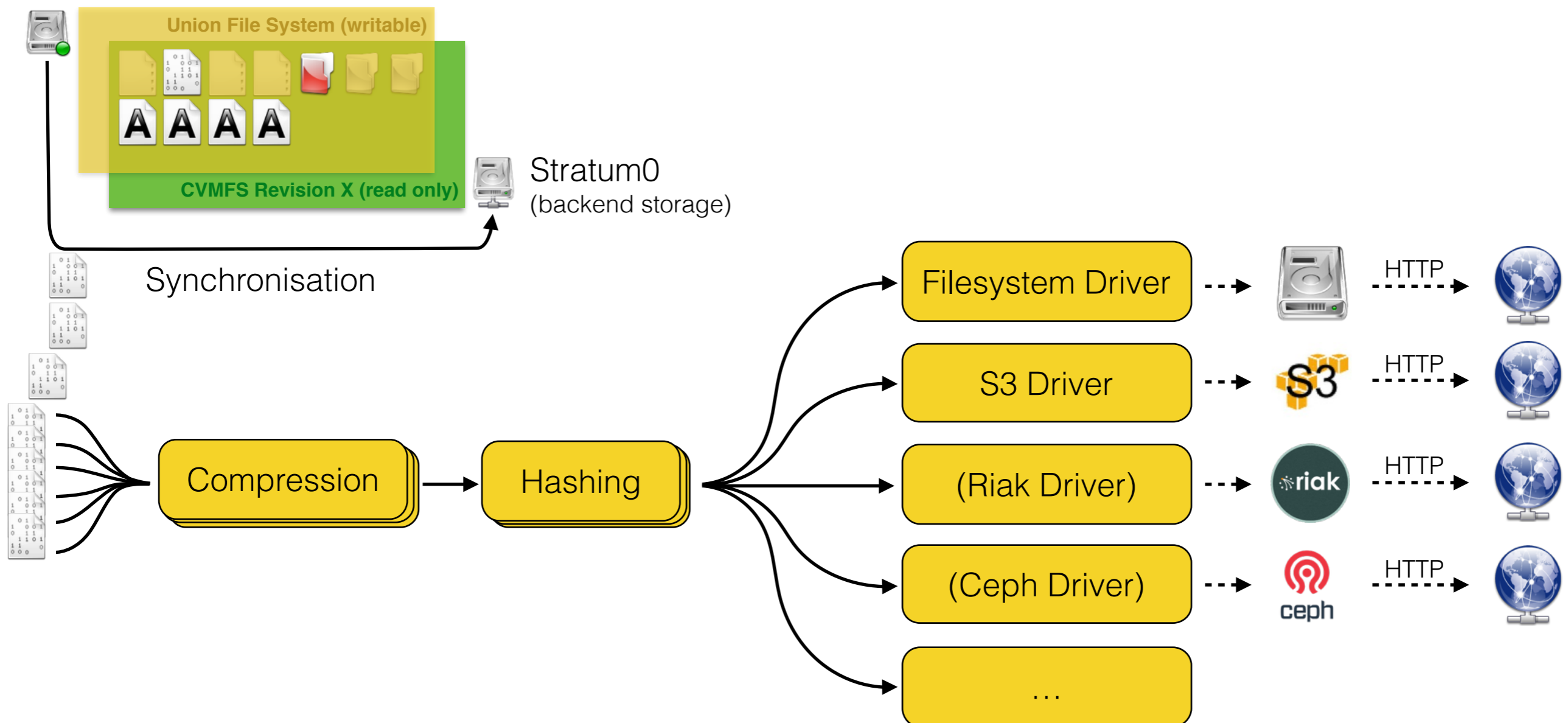
cms.cern.ch

lhcb.cern.ch



Alternative Storage Backends

- “Plug-in” Architecture since CernVM-FS Server 2.1.17
 - Potential for adding alternative storage drivers (S3, Ceph, Basho Riak, OpenStack Swift, ...)



CernVM-FS Server on S3 - Test Setup

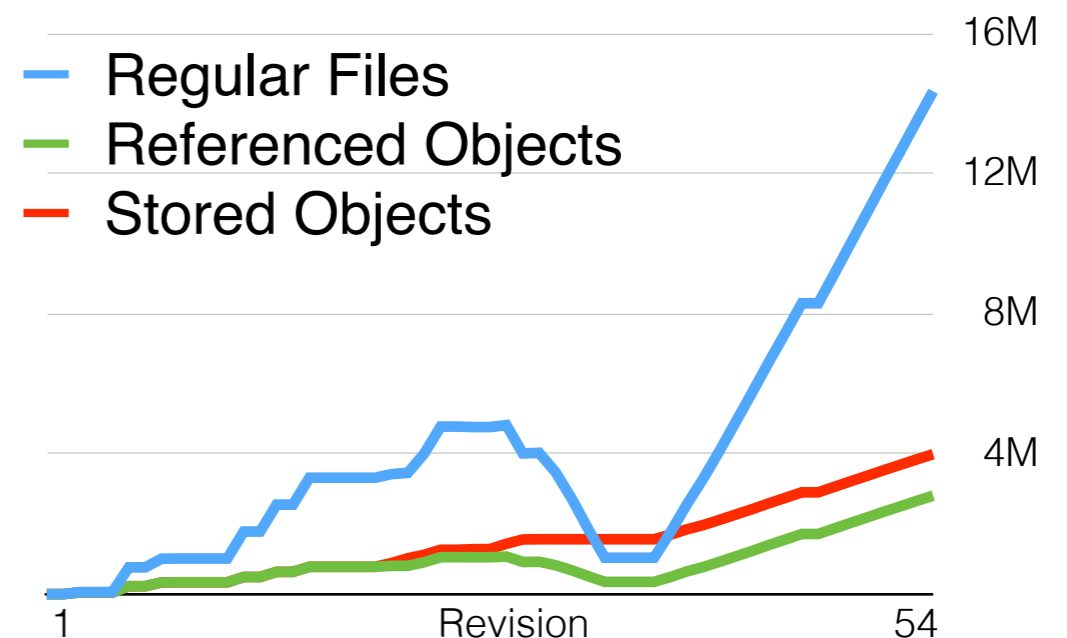
- S3 (contribution by Seppo Heikkila / CERN Openlab)
 - Field Test: Repository for LHCb Nightlies on S3
 - Experimental S3 setup hosted by Openlab
 - Publishes about 1'000'000 files and 50 GiB per night
 - Automatically replicated to an S3-based Stratum 1 (at CERN)
 - Available on lxplus through lhcbdev.cern.ch **[EXPERIMENTAL!]**
 - File publishing runs smoothly since about five weeks
 - Thanks to Ben Couturier (LHCb) for running the repository

Client Configuration Facilitation

- Just released: cvmfs-keys package version 1.5
 - Adds public keys for egi.eu and opensciencegrid.org
 - Monolithic cvmfs-keys package will be replaced by multiple cvmfs-config-[cern, osg, egi, ...] packages at some point
 - Disentangle CernVM-FS from CERN-specific configuration
- Support for CernVM-FS bootstrap repository (CernVM-FS 2.1.20)
 - Addresses tendency of “independent” CernVM-FS Stratum 0/1
 - Central place for client configuration and public keys

Garbage Collection for Stratum 0

- CernVM-FS initially designed as *insert-only* system
 - Historic snapshots stay reachable (long term preservation)
 - But: ever-growing backend storage volume
- Use-Case: Publishing of nightly integration build results
 - Requested by CMS and LHCb
 - Large amount of new files every day (f.e. LHCb: 1M files - 50 GiB)
 - Historic snapshots are of no interest
 - Garbage collection on revision level:
 - Sweep individual (old) snapshots
 - Sweep complete history



Other New and Upcoming Features

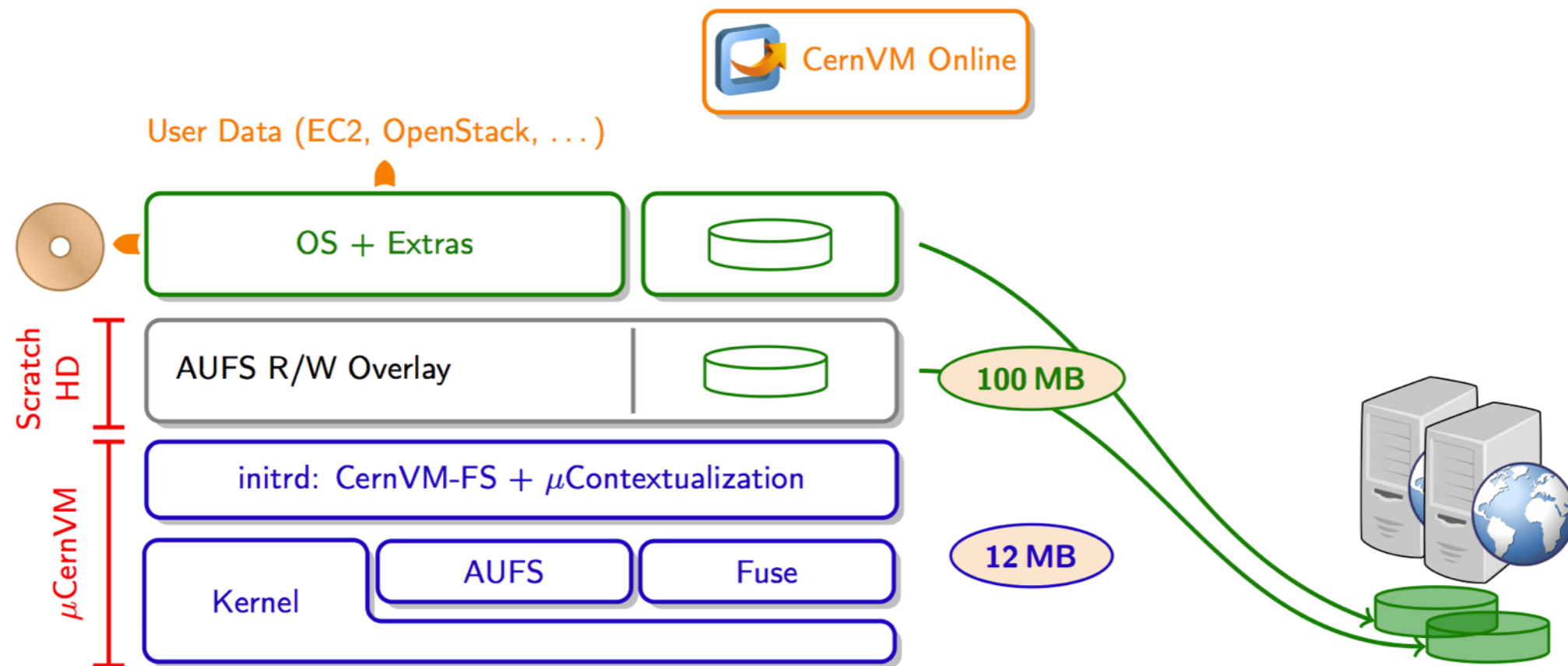
- CernVM-FS on Parrot
 - Using multiple repositories concurrently with Parrot is unstable
 - Improved switching of repositories in *libcvmfs* (CernVM-FS 2.1.20)
 - Adapted Parrot connector is submitted to *cctools* project
- Web API on Stratum 1 servers (CernVM-FS 2.1.20)
 - Automatic Stratum 1 ordering (contribution by Dave Dykstra)
 - Clients send list of configured Stratum 1 URLs to one Stratum 1
 - List is sent back ordered by geographic distance to requester
 - Based on GeoIP database (www.maxmind.com)
 - Basis for push replication of repositories (as requested by ALICE)



CernVM

CernVM Reminder

- CernVM 2 (SL5 + Conary + rPath)
 - No longer supported (End of Life: September 30th '14)
- CernVM 3
 - bootloader (μ CernVM) + SL6 (from CernVM-FS) + extras



Drastic reduction in size: 12 MB image + 100 MB cache

CernVM 3

- First production release (v3.1) on January 31st '14
- Current version 3.3 on May 27th '14
 - Based on SL 6.5, μ CernVM 1.18 (kernel 3.10.44-74)
 - Contextualisation: amiconfig, cloud-init
 - Web portal (CernVM-Online¹) with possibility to generate the user data file
 - Extras: HTCondor, ganglia, puppet, squid, xrootd, cloud clients
 - Integration with cloud-scheduler
 - cvm2ova tool to create custom OVA images
 - E.g. <http://cernvm.cern.ch/releases/ROOT6.ova> to run ROOT 6 on unsupported platforms

¹ <http://cernvm.cern.ch/portal/online>



- About
- Dashboard
- Marketplace
- Documentation
- Downloads
- Publications

Menu

- Dashboard
- Create Context
- Pair an instance
- Marketplace
- Create Cluster
- Logout

Recent context definition

- VAF Torino worker node v7
- ecsft
- CopilotVM
- ALICE Release Validation H...
- TutorialVM

Dashboard

Your context definitions

Name	Operations	WebAPI
Cvm3-LDT-1	Clone Publish	Launch now
Cvm3-OS	Clone Publish	
cvm3-test	Clone Publish	Launch now
ecsft	Clone Publish	Launch now
ecsft2	Clone Publish	Launch now
TutorialLP	Clone Publish	Launch now
TutorialVM	Clone Publish	Launch now

Get rendered context
Get raw user data

Create new context

CernVM Addressed Use-Cases

- Desktop development environment
- Image for IaaS clouds
- Volunteer computing
 - LHC@Home 2.0: T4T, LHCb, CMS, ...
- Long-term data preservation
 - Exploit time-machine features of CernVM-FS and flexibility of μ CernVM technology to recreate old environments

Hypervisor / Cloud Controller Support

Hypervisor / Cloud Controller	Status
VirtualBox	✓
VMware	✓
KVM	✓
Xen	✓
Microsoft Hyper-V	✓
Parallels	⚡ ¹
Openstack	✓
OpenNebula	✓
Amazon EC2	✓ ²
Google Compute Engine	✓ ³
Microsoft Azure	?
Docker	?

¹ Unclear license of the guest additions

² Only tested with ephemeral storage, not with EBS backed instances

³ Only amiconfig contextualisation

Long-term Data Preservation

Proved efficacy of new technology in two cases

1. ALEPH

- **Scientific Linux 4** compatible VM
- Full ALEPH software stack

2. CMS Open Data

- **Scientific Linux 5** compatible VM, complete development environment, frozen version of the CMS software framework
- Graphical environment, easy-to-install/use case (OVA bundle)

CernVM Now and Next

- Consolidation
 - Community feedback welcome and essential
- SL7-based version
- (Evaluation of) software containers (Docker) integration

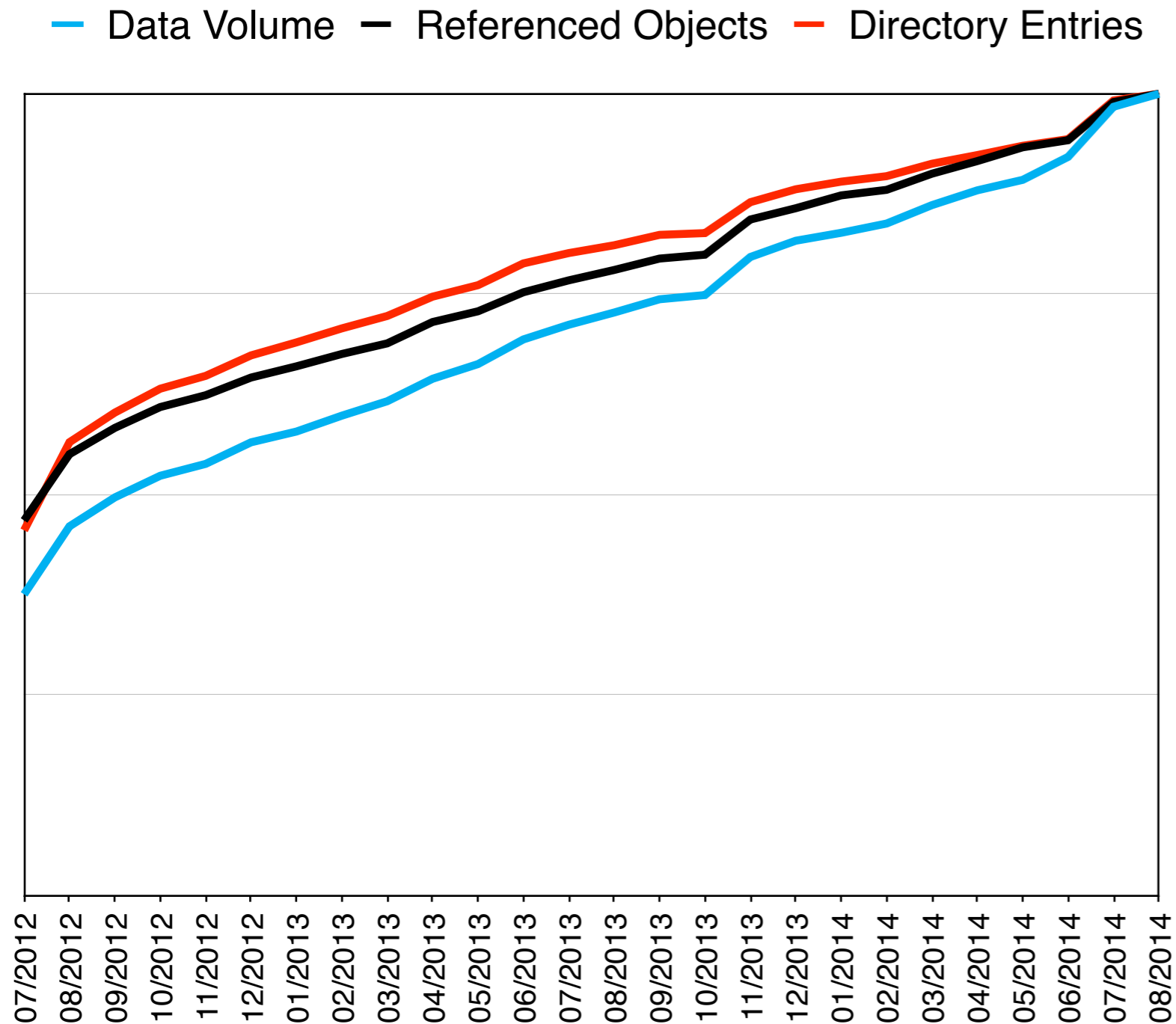


Kernel Deadlock Workaround

- Problem first encountered by ALICE in `alice-ocdb.cern.ch`
- Renaming certain files or directories on the Release Manager Machine causes a deadlock in the kernel (AUFS related)
- Reboot required, but no data loss

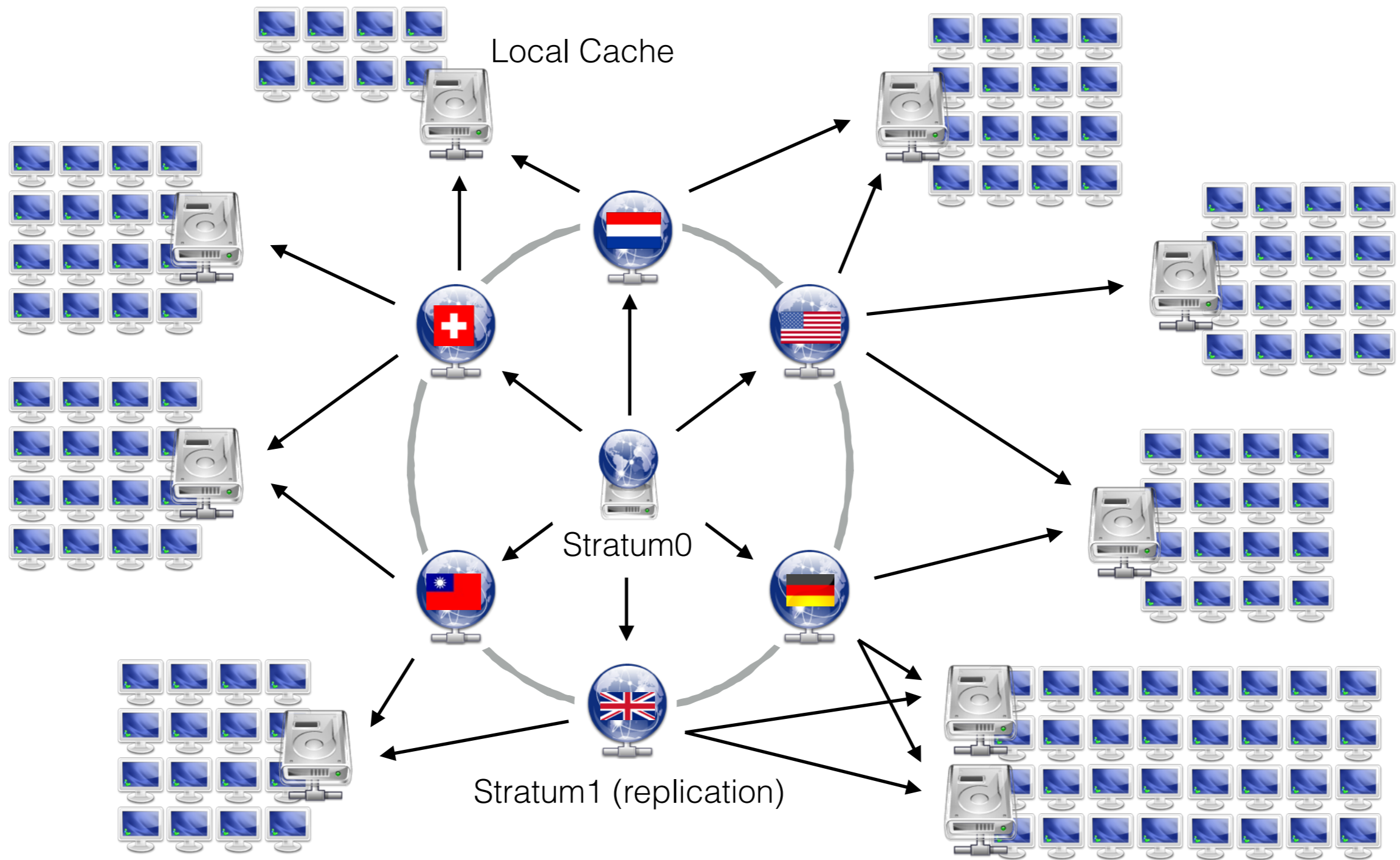
- Workaround by placing AUFS scratch space and CernVM-FS local client cache on separate file systems
(Details here: <http://cernvm.cern.ch/portal/cvmfs/workaround-krnl-deadlock>)
- Problem in AUFS is fixed as of kernel 3.10 (SL7)
- Kernel patch for SL6 based machines currently in testing
- CernVM-FS 2.1.20 detects vulnerable configurations

Growth Statistics for atlas.cern.ch

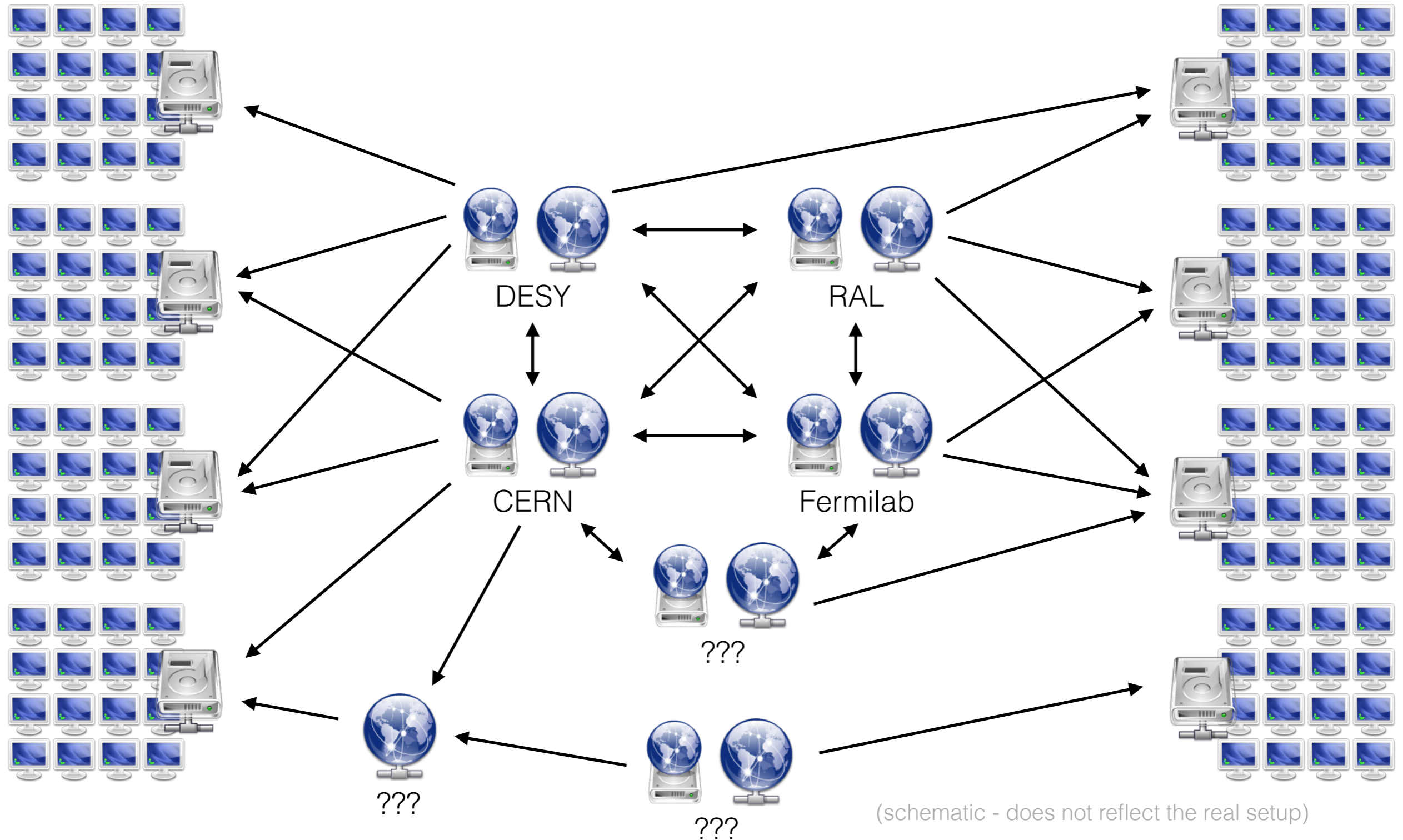


- Example Repository: **atlas.cern.ch**
- Size approximately doubled in two years
- Maximal values:
 - Data: 2.1 TiB
 - Entries: 48.0 M
 - Objects: ~3.8 M

Centralised CernVM-FS Structure



Current Mesh-like CernVM-FS Structure

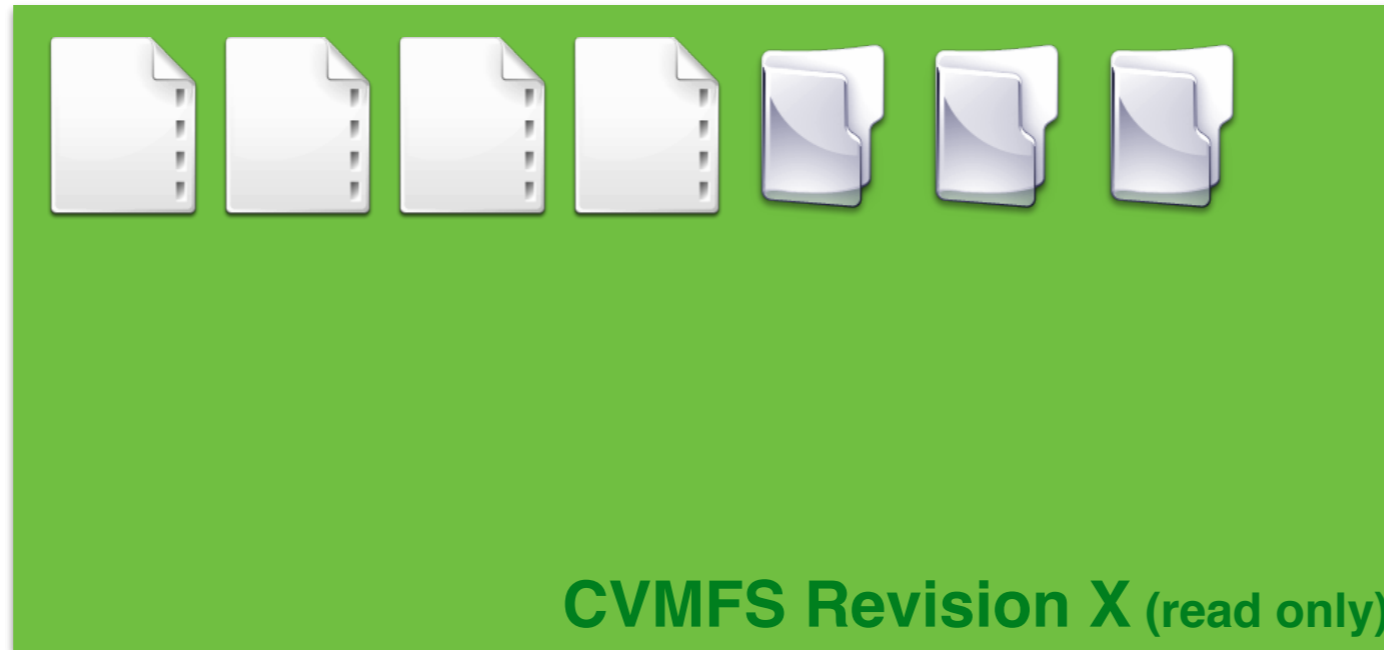


Transactional Repository Updates



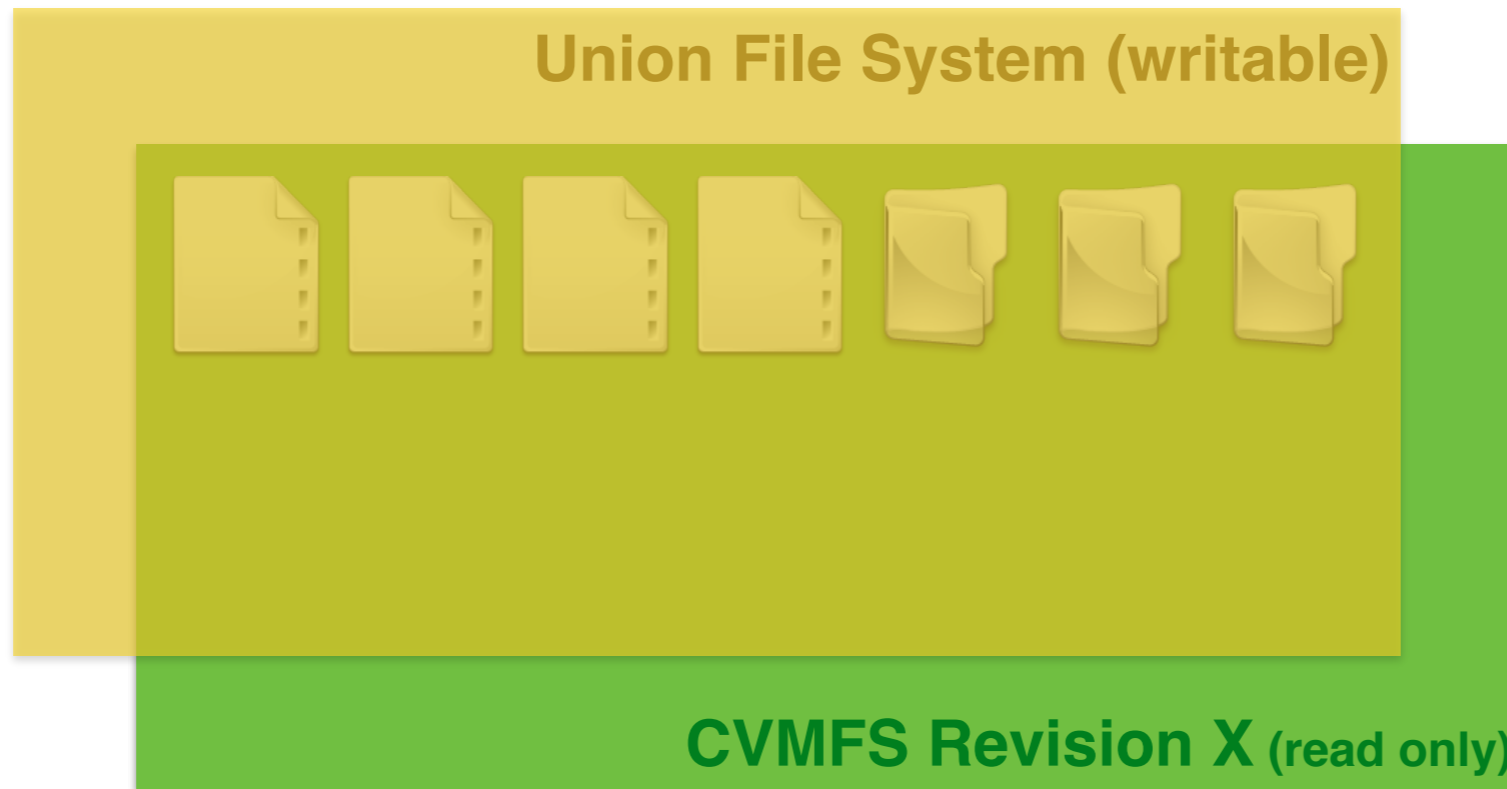
Stratum0
(backend storage)

Transactional Repository Updates



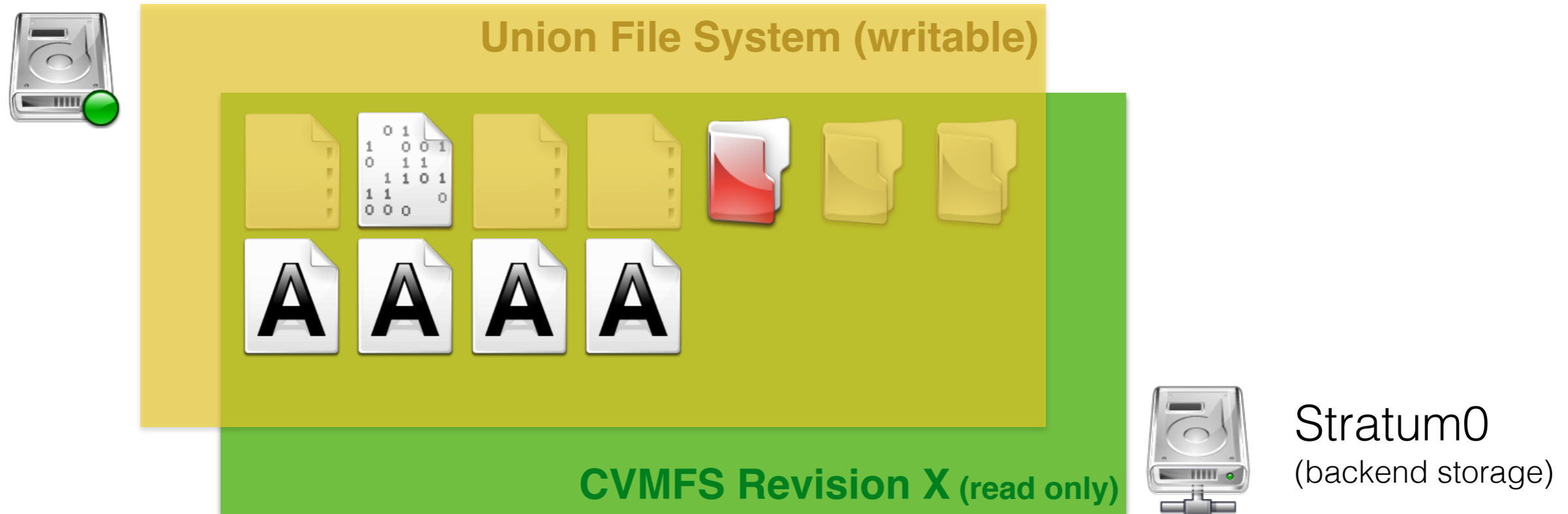
Stratum0
(backend storage)

Transactional Repository Updates

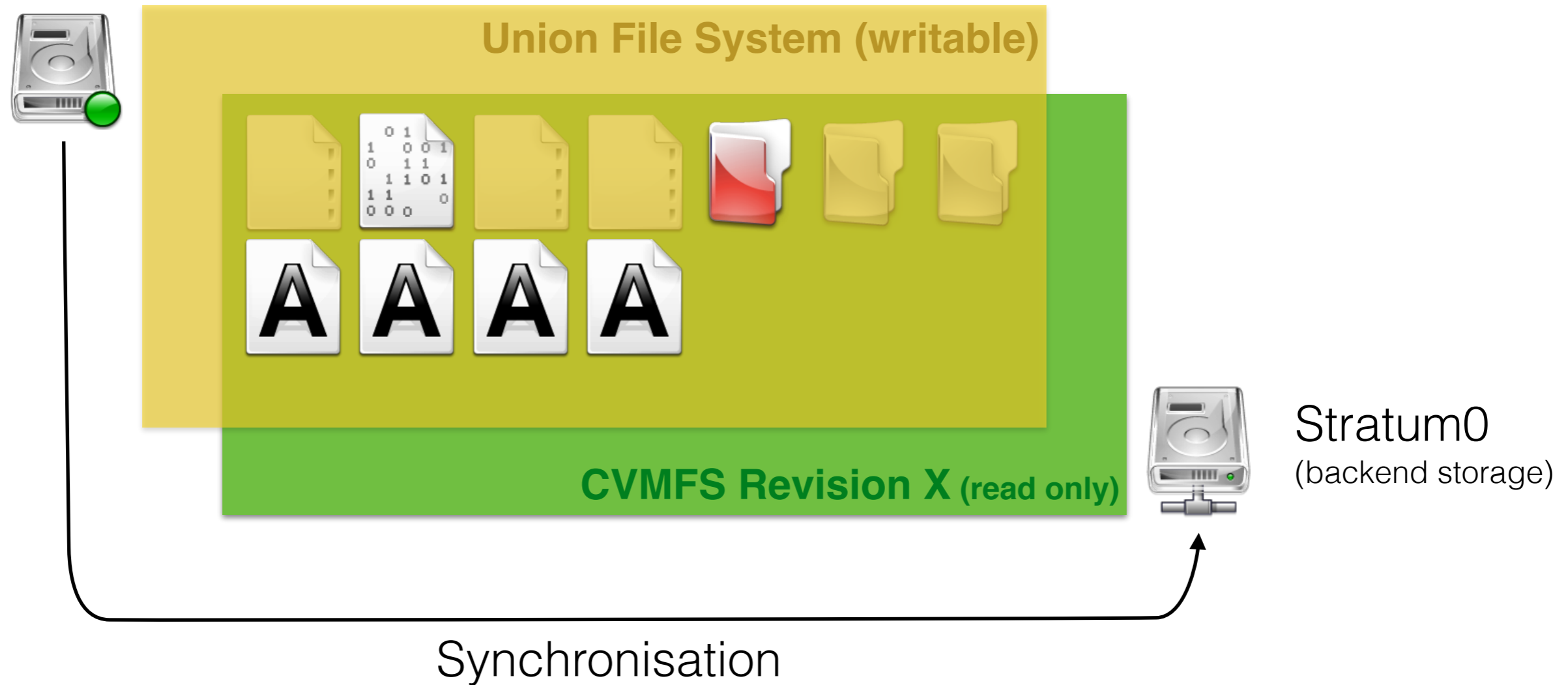


Stratum0
(backend storage)

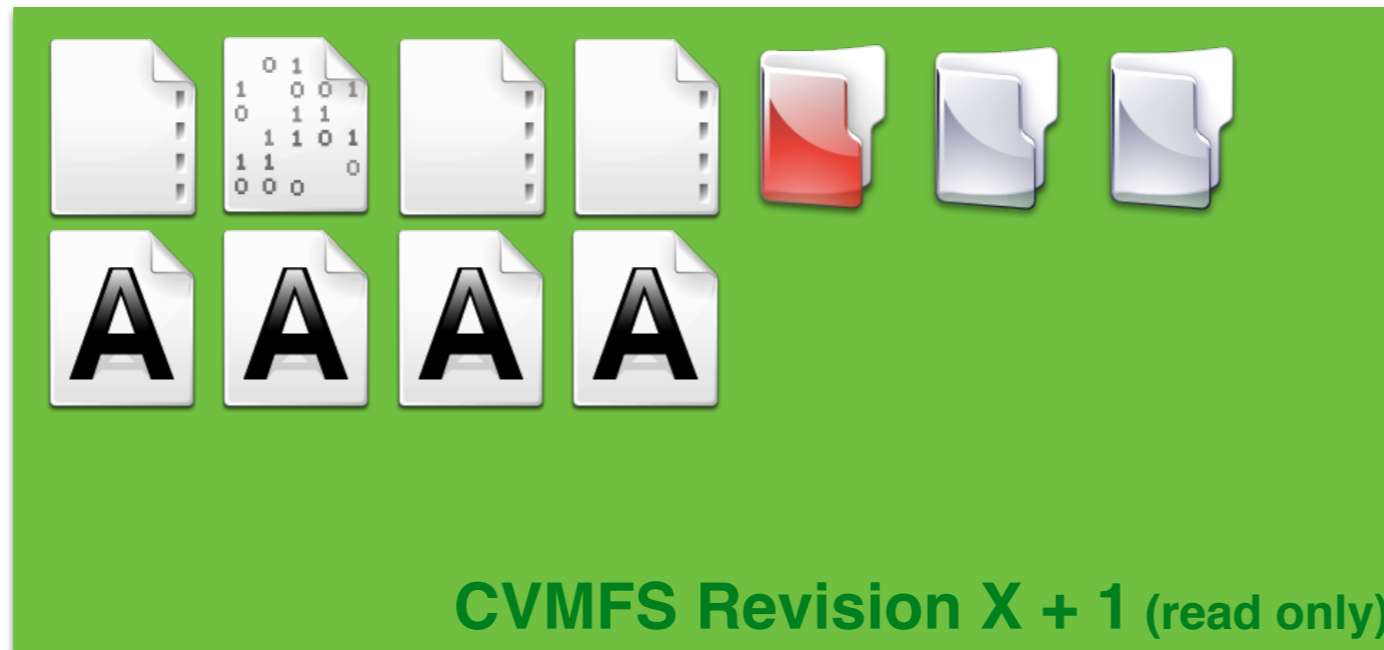
Transactional Repository Updates



Transactional Repository Updates

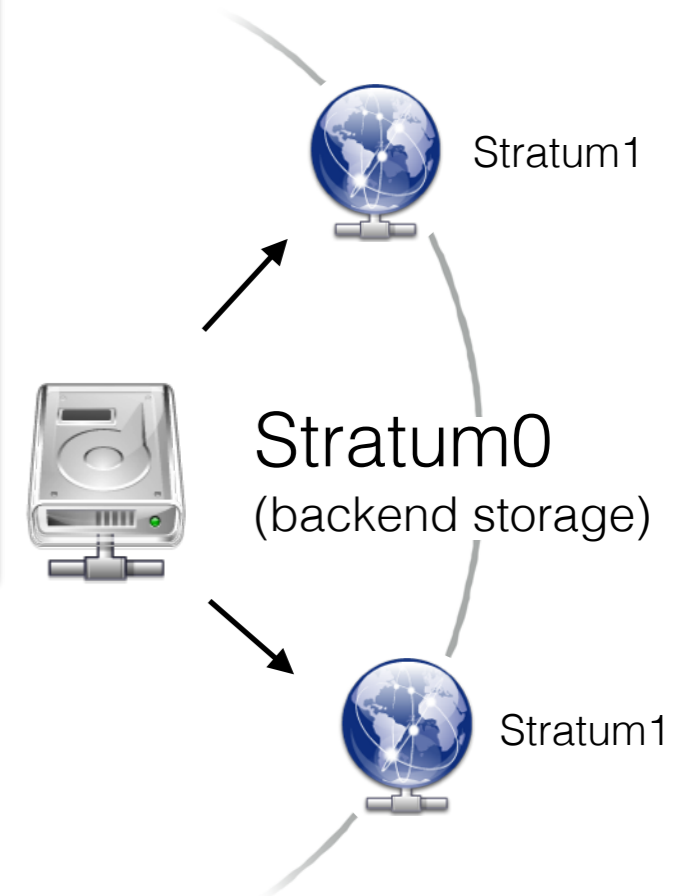
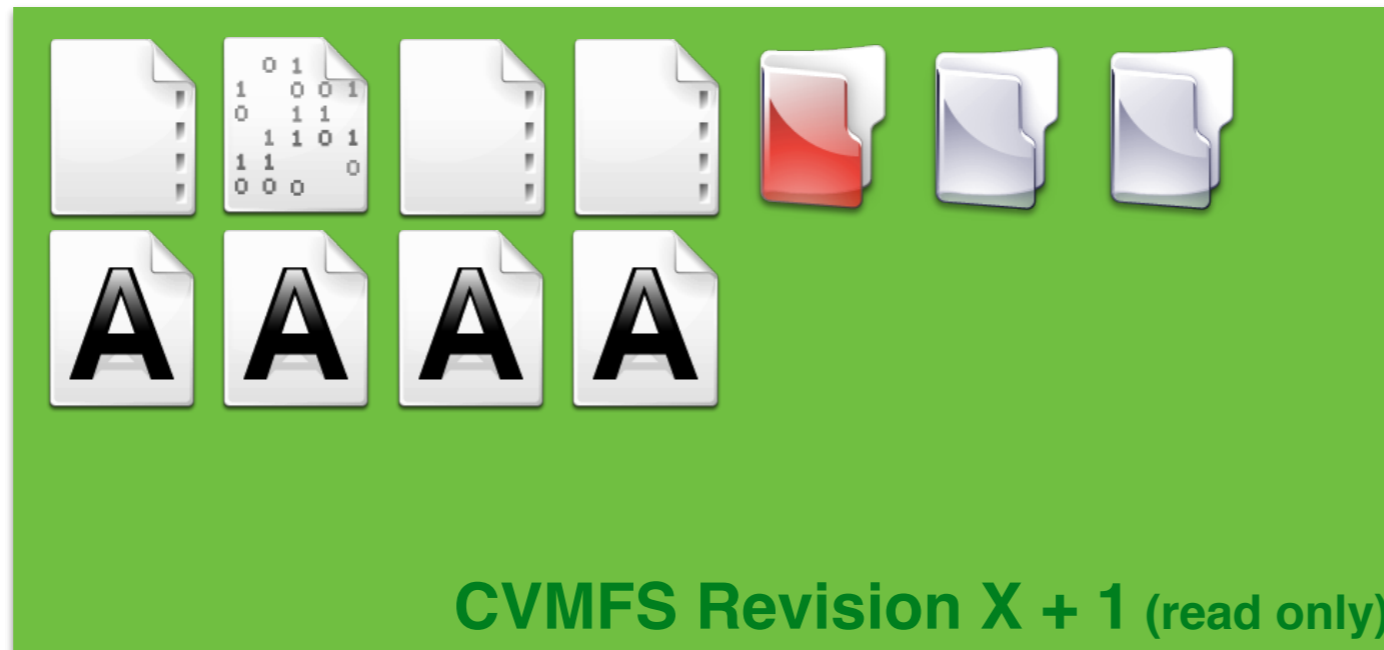


Transactional Repository Updates

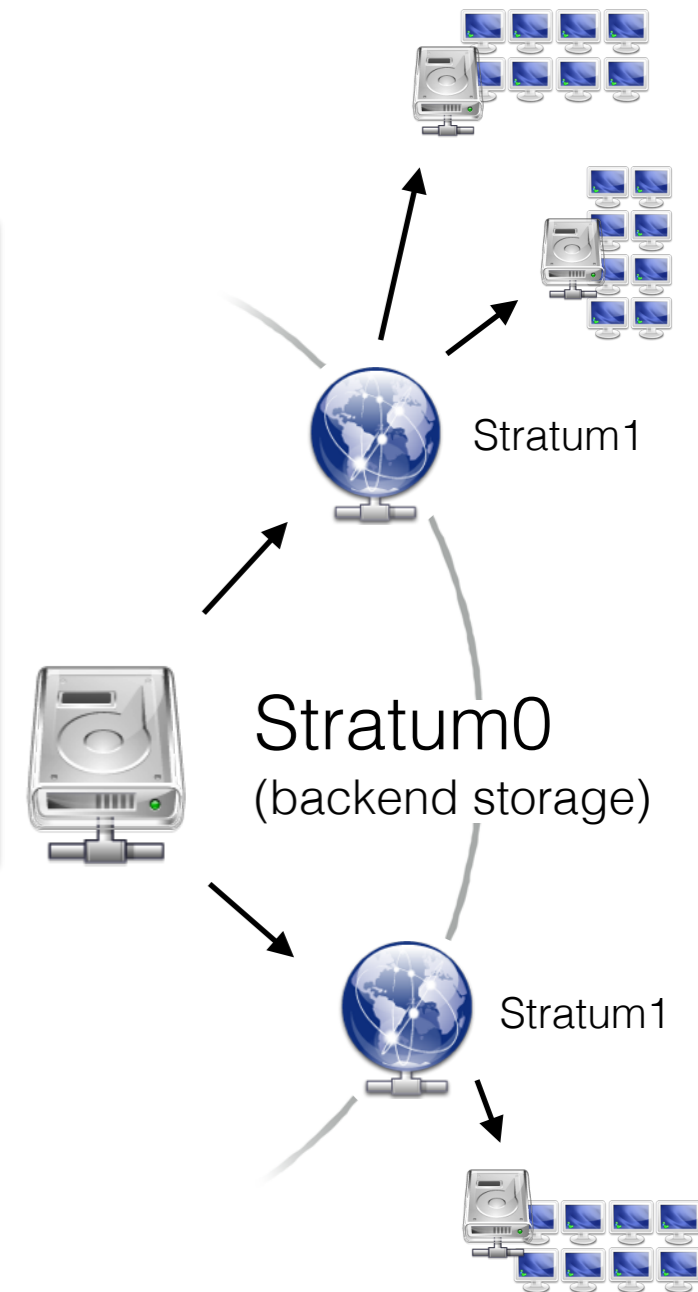
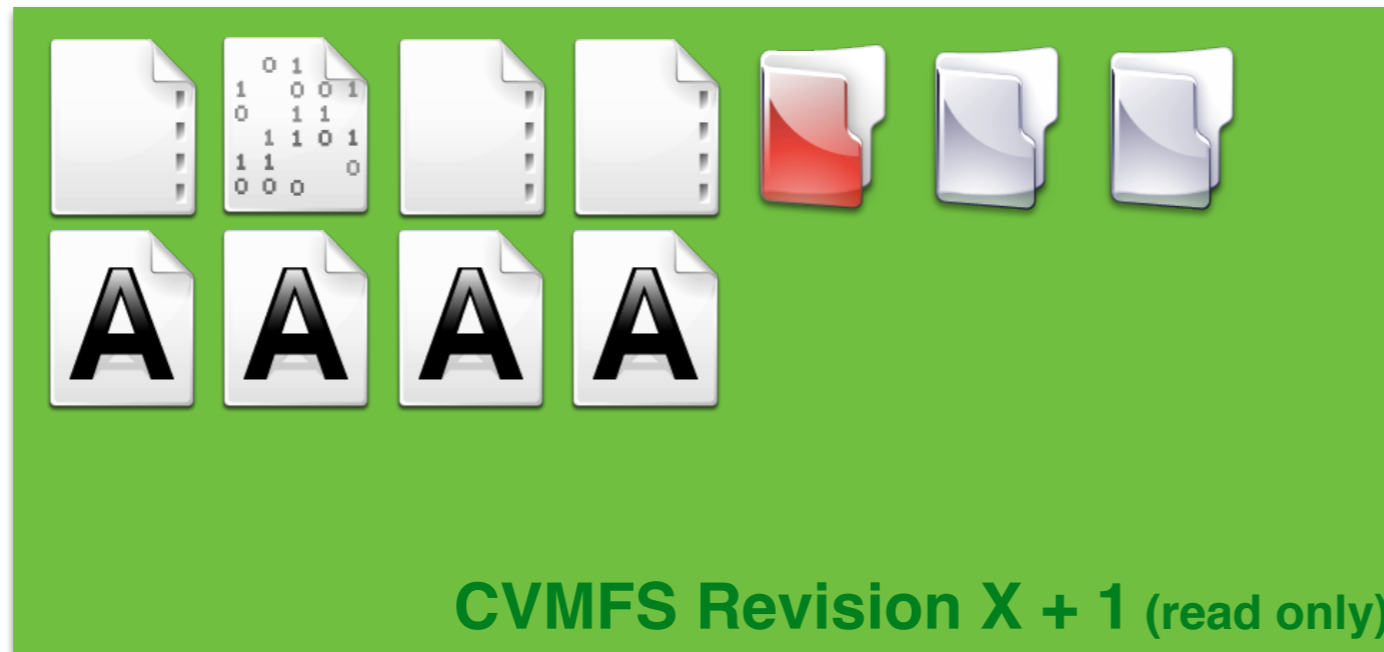


Stratum0
(backend storage)

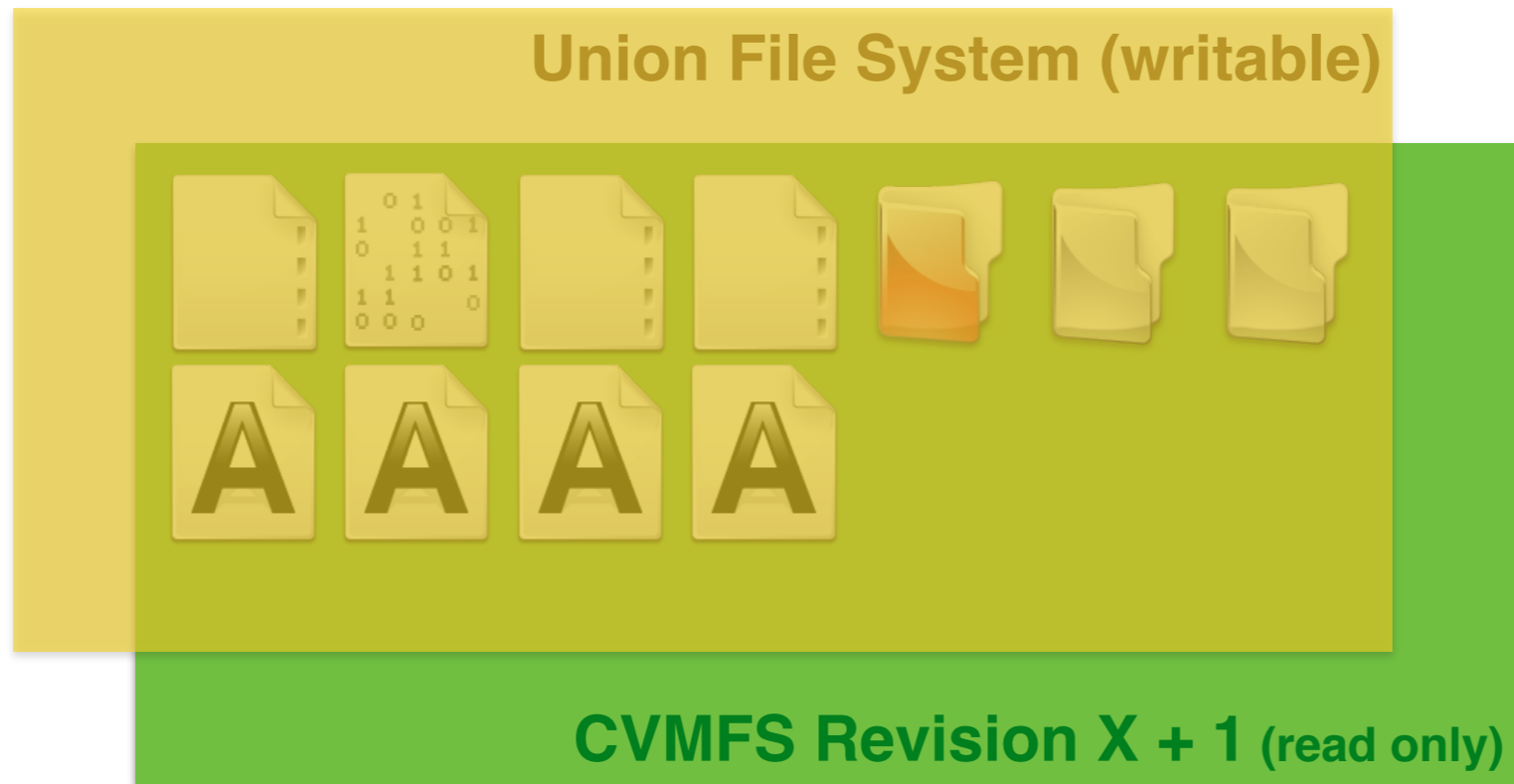
Transactional Repository Updates



Transactional Repository Updates

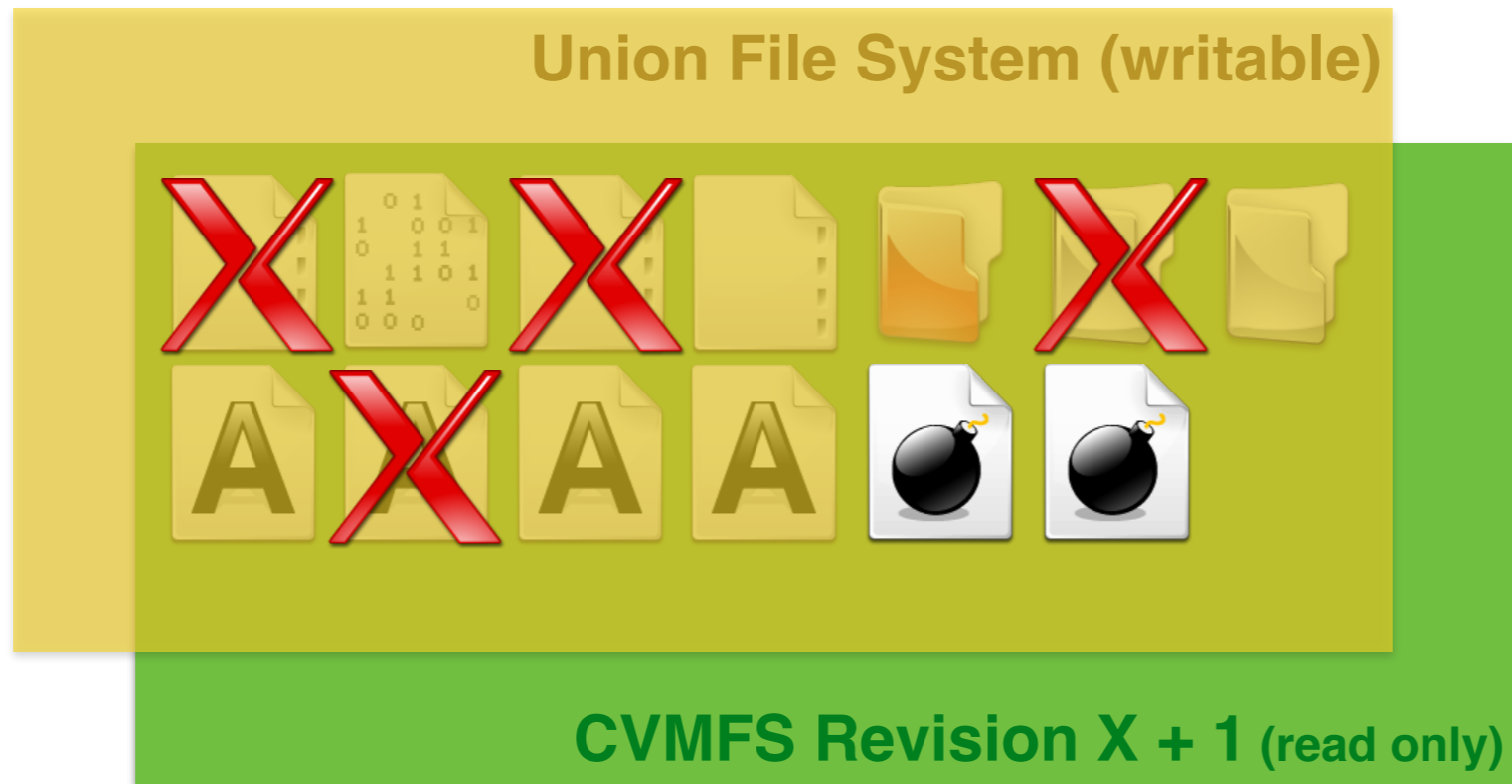


Transactional Repository Updates



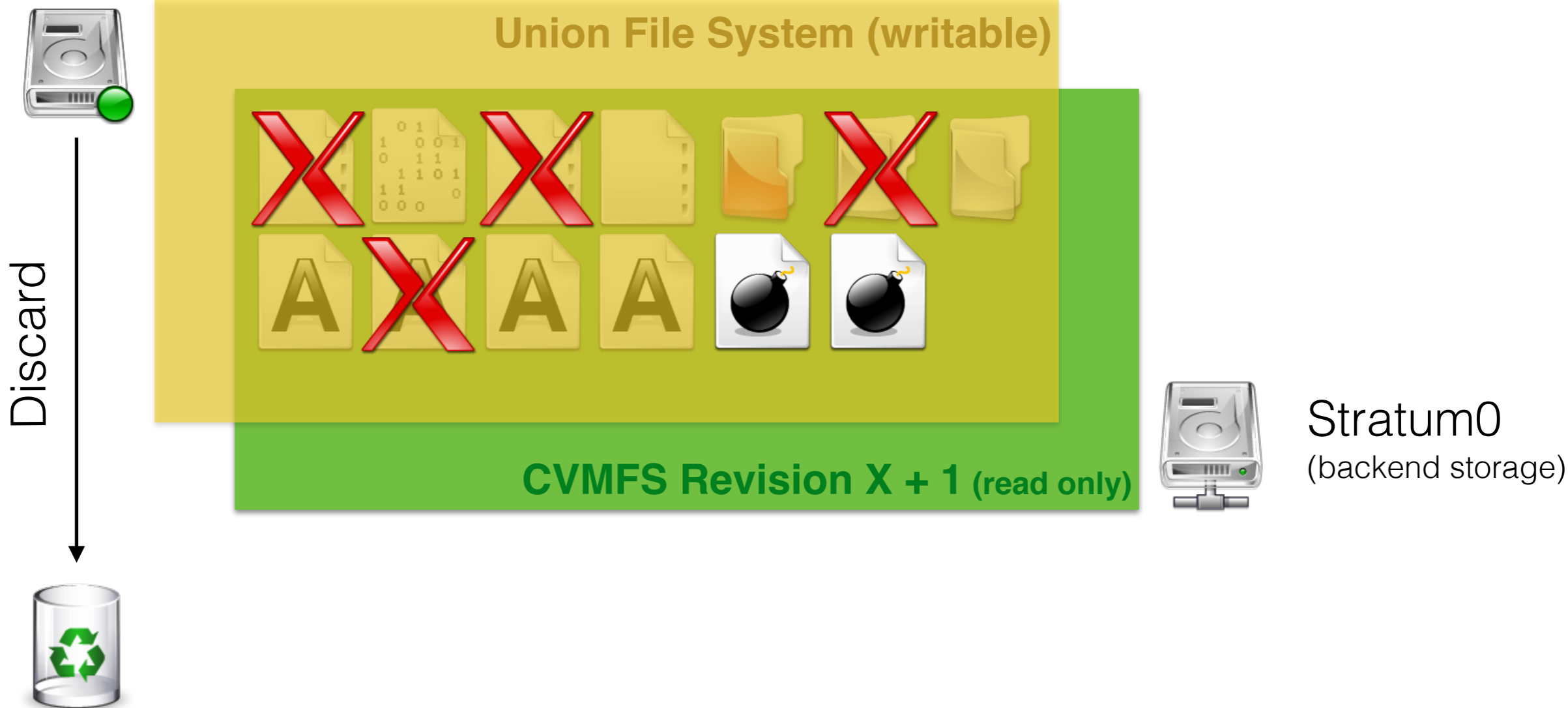
Stratum0
(backend storage)

Transactional Repository Updates

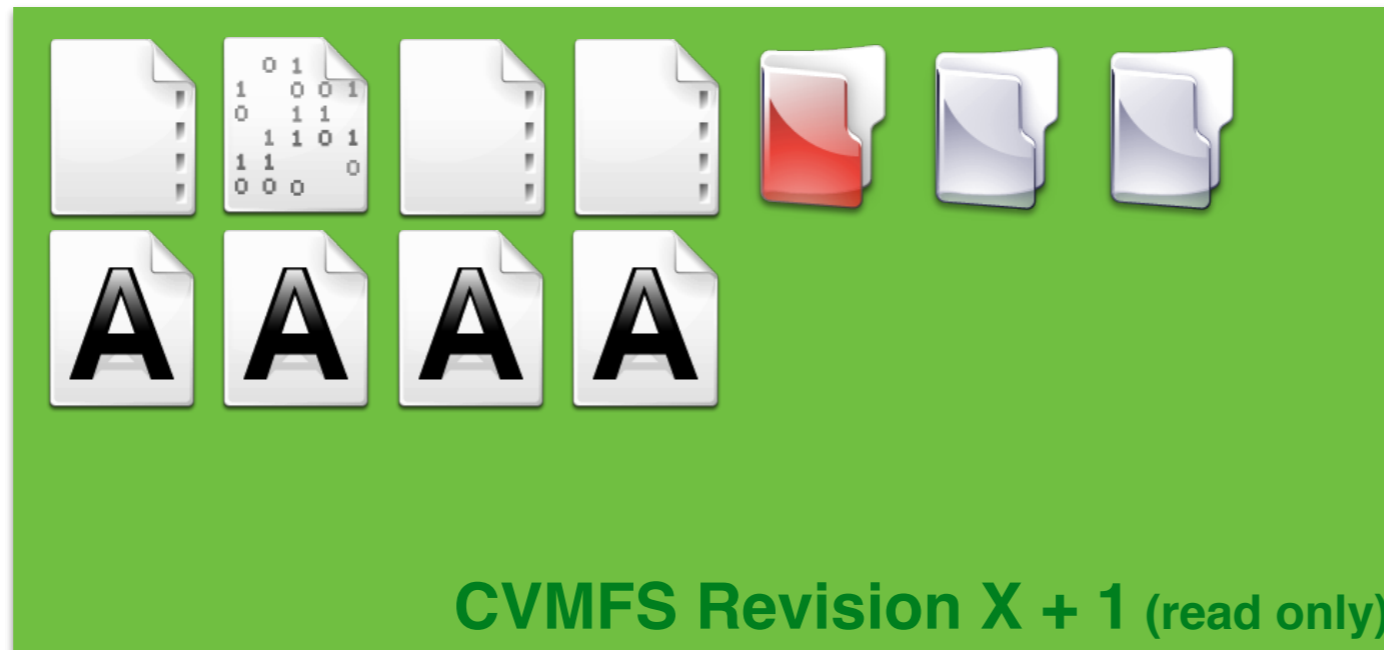


Stratum0
(backend storage)

Transactional Repository Updates



Transactional Repository Updates



Stratum0
(backend storage)

