

# cvmfs at Fermilab: Overview

with notes on the OSG / OASIS implementation

Revised 11/27/2013 S. Fuess

## Document Set

This document is part of a set of documents describing how cvmfs is implemented and used at Fermilab. The document set consists of:

- cvmfs at Fermilab : Overview <<< this document
- cvmfs at Fermilab : User's Guide
- cvmfs at Fermilab : Release Manager's Guide
- cvmfs at Fermilab : System Manager's Guide

This document set reflects version 2.1.14 of the cvmfs product.

*Note: cvmfs is implemented elsewhere in different modes. Where useful, the differences in other implementations are noted.*

## References

The CERN VM File System (CernVM-FS, or cvmfs) was, as the name implies, originally developed at CERN for the function of distributing virtual machine (VM) images, and is a project within the larger CernVM toolkit. CernVM-FS was recognized as providing features as a stand-alone network file system. The home page is available at <http://cernvm.cern.ch/portal/filesystem>.

The OSG / OASIS implementation is documented at <https://twiki.opensciencegrid.org/bin/view/Documentation/Release3/InstallCvmfs> and <https://www.opensciencegrid.org/bin/view/ReleaseDocumentation/OasisUpdateMethod>.

## Overview

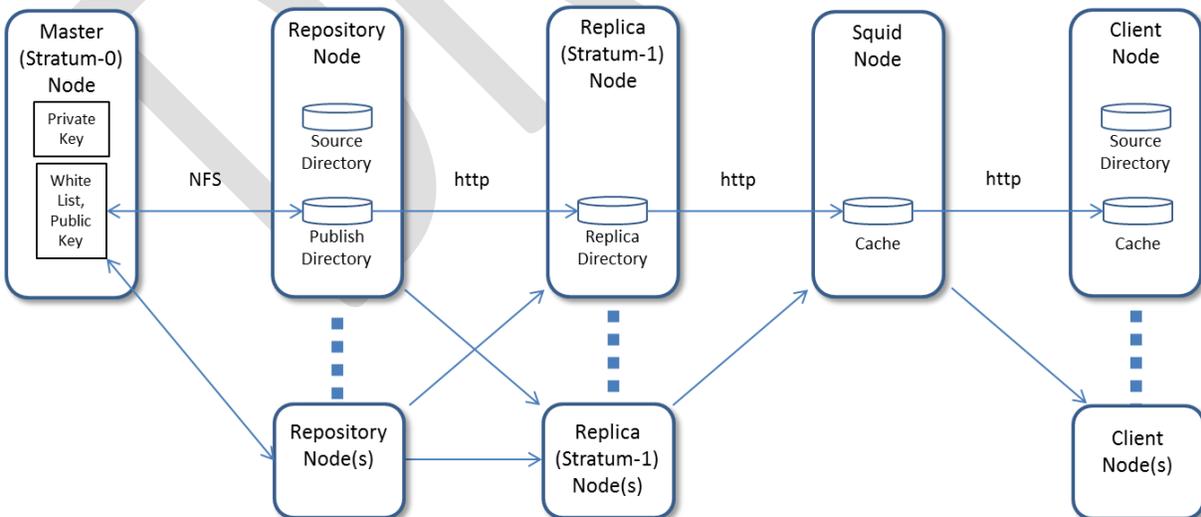


Figure 1 Elements of cvmfs file system

Figure 1 shows the elements within cvmfs and the path from the definitive source directory to the distributed copies of the source directory on the client nodes. Each of these elements is discussed in more detail in the following.

Note: The OSG configuration combines the *Master* and *Repository* functions within a single node.

## Master (Stratum-0)

- **Function**

The *Master node* at Fermilab is the root of the distribution of the cvmfs key signing mechanism. The OSG Oasis implementation also combines the functions of the *Repositories* within the *Master node*.
- **Master Nodes**

The Fermilab *Master node* and the corresponding *Master name*:

  - cfsmaster01.fnal.gov                      [fnal.gov](http://fnal.gov)

The OSG *Master node* and the corresponding *Master name*:

  - oasis.opensciencegrid.org              [oasis.opensciencegrid.org](http://oasis.opensciencegrid.org)

The above [names](#) are noted as <master> in this document.
- **RPMs required**
  - cvmfs-server
  - cvmfs-keys                                      (only needed to set directory structure)
- **File Systems / Directories**

[/srv/cvmfs/<repo>/pub](#)                                      (NFS mounted from *Repository* node)

The *Repository's published* directory, to which the *Repository* whitelist is written and signed with the *Master* key. In the Oasis configuration this is a local file system also used for the *Repository* function.
- **Configuration**

[/etc/cvmfs/keys/<master>.masterkey](#)  
The principal secret key underlying the cvmfs implementation.

[/etc/cvmfs/keys/<master>.pub](#)  
The public keys used to interpret the whitelist. This key is distributed out-of-band to *Clients*.

[/etc/cvmfs/keys/<repo>.crt](#)  
Certificates generated at the *Repository* creation time, managed out-of-band and replicated on the *Master*, to be signed by this *Master (Stratum-0)* server as part of the *Repository* whitelist.
- **Access**

Access is limited to root.
- **Notes**
  - The *Master (Stratum-0)* [name](#) is needed throughout the cvmfs distribution chain to identify the signing key. The *Master (Stratum-0)* [node](#) is not needed at any further point in the cvmfs chain (for the Fermilab implementation; for Oasis this node is also the *Repository* location).

## Repository

- **Function**

The *Repository* is the origin of the cvmfs distribution chain. The experiment/group release manager deposits files in a specific Source (*Shadow*) directory within the *Repository*. A cvmfs daemon watching the *Shadow* directory notes any changes in a change log file. Upon execution of an explicit "cvmfs publish" command, the change log is parsed to create/update a condensed and encrypted version of

the *Shadow* directory in a *Publish* directory. The cvmfs system propagates the *Publish* directory to clients, which a fuse file system layer translates back into a read-only mirror of the original *Shadow* directory.

- **Repositories (Nodes)**

The current Fermilab *Repository* [names](#), equivalent (by convention, not required) to the FQN of the *Repository* [node](#):

- [novacfs.fnal.gov](http://novacfs.fnal.gov)
- [gm2cfs.fnal.gov](http://gm2cfs.fnal.gov)
- [d0cfs.fnal.gov](http://d0cfs.fnal.gov)
- [ssidistcfs.fnal.gov](http://ssidistcfs.fnal.gov)
- [uboonecfs.fnal.gov](http://uboonecfs.fnal.gov)
- [lbnecfs.fnal.gov](http://lbnecfs.fnal.gov)

The OSG *Repository* [name](#) is:

- [oasis.opensciencegrid.org](http://oasis.opensciencegrid.org)

The OASIS *Repository* server also acts as the *Master / Stratum-0* server.

The above *Repository* [names](#) are noted as <repo> in this document.

- **RPMs required**

- cvmfs-server
- cvmfsflt
- cvmfs-keys (only needed to set directory structure)
- kmod-cvmfsflt
- kmod-redirfs (not included in base SLF distribution)
- redirfs (not included in base SLF distribution)
- httpd (Apache server)

- **File Systems / Directories**

[/srv/cvmfs/<repo>/staging](#)

User controlled directory for importing files, typically as a tar file containing a distribution. The Fermilab repositories mount this from a distinct volume for efficiency purposes.

[/srv/cvmfs/<repo>/shadow](#), bind mounted as [/cvmfs/<repo>](#)

The *Shadow* directory into which the release manager places files for *distribution*, typically as the result of unpacking a tar file from the *Staging* area.

[/srv/cvmfs/<repo>/ctrl](#)

Directory populated by the cvmfs daemon upon changes to the *Shadow* directory.

[/srv/cvmfs/<repo>/pub](#)

Directory populated by a cvmfs *Publish* command. In the Fermilab configuration, this directory is NFS exported to the *Master (Stratum-0)* node in order to receive a signed whitelist. This *Publish* directory is then further distributed via http through the cvmfs server chain.

- **Configuration**

[/etc/cvmfs/server.local](#)

Defines the *Repository* [name](#) and the directories noted above.

[/etc/cvmfs/keys/<repo>.crt](#)

[/etc/cvmfs/keys/<repo>.key](#)

Private key and self-signed certificate used to sign the *Repository*. (Note that this is different from the master key belonging to the *Master (Stratum-0)* [node](#) used to sign the whitelist.) In the Fermilab configuration, the certificate is distributed out-of-band to the *Master* [node](#).

[/etc/httpd/conf.d/cvmfs-stratum.conf](#)

Sets rewrite rules and access restrictions to allow only *Replica (Stratum-1)* access.

- **Access**

The Fermilab configuration allows login access to the *Repository node*. There is only a single user, cvmfs. The .k5login for that user contains the principals of the release managers controlling the specific *Repository*.

The OSG configuration utilizes a common login node, oasis-login.opensciencegrid.org. Logins are enabled via gsissh. The Oasis login node NFS mounts the *Shadow* directories from the joint *Master/Repository node*.

- **Notes**

- The *Repository* and *Master (Stratum-0)* functions are usually combined in a single node. The Fermilab installation chooses to separate the two, which allows experiments to fully/uniquely control the *Repository* but completes the distribution utilizing a single private/public key associated with the *Master (Stratum-0)* name.
- The *Repository* name is needed throughout the cvmfs distribution chain. The *Repository node* is needed only at the *Master (Stratum-0)* node for the NFS mount.

## Replica (Stratum-1)

- **Function**

The *Replica node* maintains a *Replica* directory that is a periodically updated snapshot of a *Published* directory from the *Repository node*. A *Replica node* may act on behalf of multiple *Repositories*, by means of generating multiple *Replica* directories from different *Published* origins. The intention is that there are one or more *Replica* servers per physical site, providing near-local access of the *Published* directories to *Clients*.

- **Replica Servers (Nodes)**

The current Fermilab *Replica nodes*, configured to distribute the Fermilab *Repositories*:

- cfs01.fnal.gov
- cfs02.fnal.gov

The OSG *Replica* server operated at the GOC, configured to distribute OSG supported *Repositories* is:

- oasis-replica.opensciencegrid.org

Fermilab also operates a *Replica node* that serves both OSG/OASIS and CMS:

- cvmfs.fnal.gov

- **RPMs required**

- cvmfs-replica
- cvmfs-keys (only needed to set directory structure)

- **File Systems / Directories**

[/srv/cvmfs/<repo>/pub](#)

A local copy of the directory originally populated by a “cvmfs publish” command on the *Repository node*. This directory is populated by web access from the *Repository node* as initiated by a cron job (every 30 minutes), then further distributed via a web server.

- **Configuration**

[/etc/cvmfs/replica.repositories](#)

Sets the access URLs for the *repositories* distributed by this *replica* server. The file contains information for each repository of:

- *Repository* name

- URLs of the *Repository* node
- Local location of the public master key
- Local location for the *Replica* of the *Publish* directory
- Parameters controlling the operation of the *Replica*
  - Number of parallel connections
  - Timeout period
  - Number of retries allowed

[/etc/cvmfs/keys/<master>.pub](#)

The public master key belonging to the *Master (Stratum-0)* and used to sign the whitelist.

[/etc/cron.d/cvmfs\\_snapshot](#)

The cron file to execute a `cvmfs_snapshot` command, which performs a web access from the served *Repository's* nodes.

- **Access**

Access is limited to root.

## Squid Proxy

- **Function**

Although not specific to `cvmfs`, a *Squid Proxy* is typically used as a cacheing layer between the *Clients* and the *Replica (Stratum-1)* servers. The *Replica (Stratum-1)* servers are typically sized to handle only moderate network traffic to the *Squid Proxy* servers, which are configured to serve multiple *Clients* potentially accessing multiple *Replica (Stratum-1)* servers.

## Client

- **Function**

The *Client* is the ultimate user of the `cvmfs` distribution. The *Repository* source directory appears as an auto-mounted read-only file system.

- **Clients (Nodes)**

These *Clients* are configured to use the Fermilab `cvmfs` distribution:

- GPCF interactive nodes (limited to the *Repository* associated with the VM)
- GPCF local batch nodes (all local Fermilab *Repositories*)
- GP Grid worker nodes

These *Clients* are additionally configured to use the OSG Oasis `cvmfs` distribution:

- GPCF interactive nodes
- GP Grid worker nodes

- **RPMs required**

- `autofs`
- `cvmfs`
- `cvmfs-keys` (only needed to set directory structure)
- `fuse`
- `oasis-config` (see the OASIS documentation [here](#))

- **File Systems / Directories**

[/var/cache/cvmfs2/<repo>](#)

Cache area used by the `cvmfs` daemon; not accessed by the user. This will match the *Publish* directory of the *Repository*.

[/cvmfs/<repo>](#)

The read-only *Distribution* directory, propagated by *cvmfs* from the *Repository*. This is the directory ultimately accessed by applications on the *Client*.

- **Configuration**

[/etc/cvmfs/default.local](#)

This includes

- a list of *Repository* [names](#) accessed by the *Client*
- the *Proxy* server through which the *Repositories* are accessed

[/etc/cvmfs/config.d/<repo>.conf](#)

A configuration file is required for each *Repository* to be accessed. Each includes:

- a list of URLs of *Replicas* that serve the distribution
- the location of the public master key for the distribution

[/etc/cvmfs/domain.d/<master>.conf](#)

[/etc/cvmfs/domain.d/<master>.local](#)

Configurations associated with the accessed *Master (Stratum-0)* sites.

[/etc/cvmfs/keys/<master>.pub](#)

The public master key belonging to the *Master (Stratum-0)* [name](#) and used to sign the whitelist.

This key is distributed out-of-band.

[/etc/auto.master](#) and [/etc/auto.cvmfs](#)

Configure the automounter to mount *cvmfs* distributions at */cvmfs*

- **Access**

The *Client* is configured to allow general access.

- **Notes**

- A *Client* may access multiple *Repositories* from multiple *cvmfs* paths.
- Versions of the *cvmfs Client* code prior to the 2.1.x (alpha release at the date of this document) manage the local cache independently for each served *Repository*. If the net sum of allocated cache space is oversubscribed, there is a possibility that the cache space will be filled. In this case the *cvmfs* fuse module panics, causing a system crash. In later versions of the *cvmfs Client* code the cache space is collectively managed across all served *Repositories*.