Handling Software

- /opt/exp_software
- CVMFS
- Containerized CVMFS

Data vs Software

Software	Data
POSIX interface	put, get, seek, streaming
File dependencies	Independent files
$O(kB)$ per file	$O(GB)$ per file
Whole files	File chunks
Absolute paths	Relocatable

WORM ("write-once-read-many")
Billions of files
Versioned

Software is massive not in volume but in number of objects and meta-data query rates

The Old Style

❖ /opt/exp_software storage area made available in each site

❖ Available ReadOnly via POSIX in all WNs and UIs

❖ Special roles (sgm) can write on it

❖ Need manual deployment in all the sites of newer files/releases

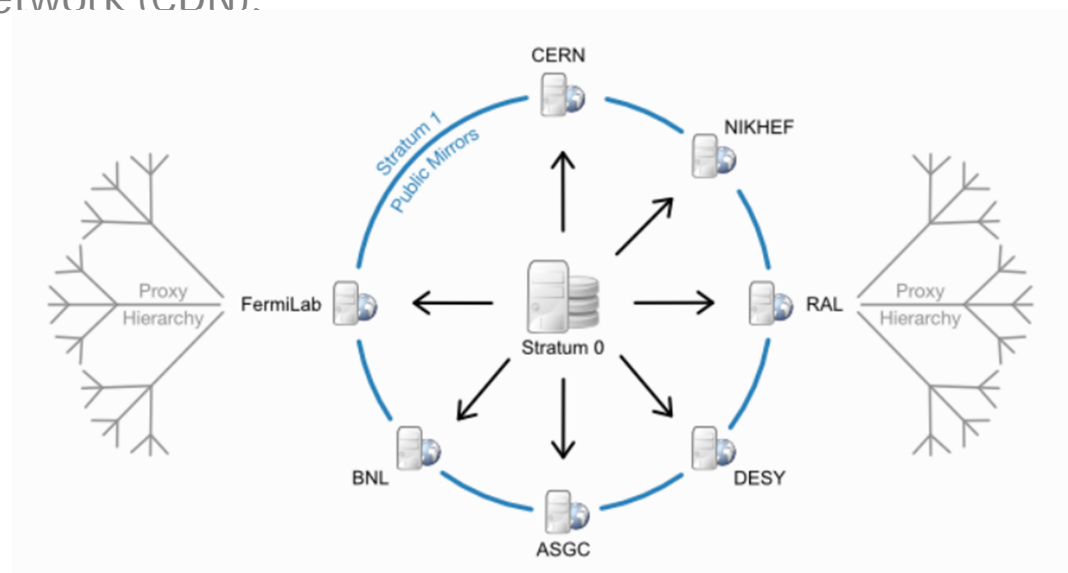❖ In general it is a file systems optimized for small files

CernVM-Fs

❖ CernVM-FS is a special-purpose virtual file system that provides a global shared software area for many scientific collaborations

❖ Asynchronous writing (publishing) key to meta-data scalability

❖ Provides a scalable, reliable and low- maintenance software distribution service.

❖ Implemented as a POSIX read-only file system in user space (a FUSE module).

❖ Files and directories are hosted on standard web servers and mounted in the universal namespace /cvmfs.

❖ Internally, CernVM-FS uses content-addressable storage and Merkle trees in order to maintain file data and meta-data

❖ Uses outgoing HTTP connections only
  ➢ Avoids most of the firewall issues of other network file systems.
  ➢ Transfers data and meta-data on demand and verifies data integrity by cryptographic hashes.

❖ Aggressive caching and reduction of latency, focuses specifically on the software use case

CernVM-Fs -Stratum 0

❖ CernVM-FS is a file system with a single source of (new) data
❖ This single source, the repository Stratum 0, is maintained by a dedicated release manager machine or publisher.
❖ A read-writable copy of the repository is accessible on the publisher.
❖ The CernVM-FS server tool kit is used to publish the current state of the repository on the release manager machine.
❖ Publishing is an atomic operation
❖ All data stored in CernVM-FS have to be converted into a CernVM-FS repository during the process of publishing.

CernVM-Fs -Stratum 1

❖ While a CernVM-FS Stratum 0 repository server is able to serve clients directly, a large number of clients is better be served by a set of Stratum 1 replica servers.

❖ Multiple Stratum 1 servers improve the reliability, reduce the load, and protect the Stratum 0 master copy of the repository from direct accesses.

❖ Stratum 0 server, Stratum 1 servers and the site-local proxy servers can be seen as a Content Distribution Network (CDN).

Scale of Deployment

Source / Stratum 0
Replica / Stratum 1
Site / Edge Cache

LHC infrastructure:
- > 1 billion files
- $\sim 100\,000$ nodes
- 5 replicas, 400 web caches

Picture by: jblomer<at>cern.ch