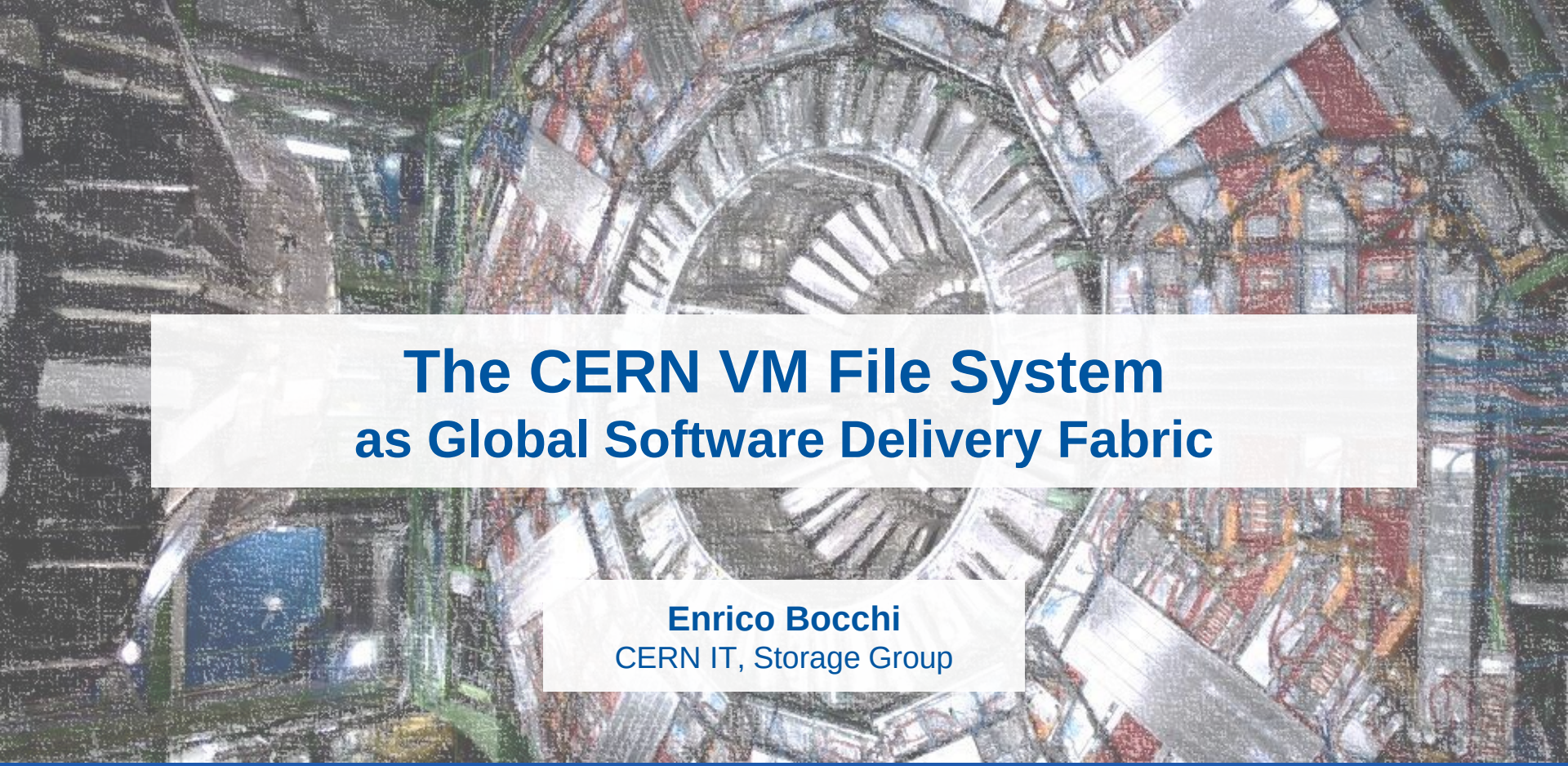
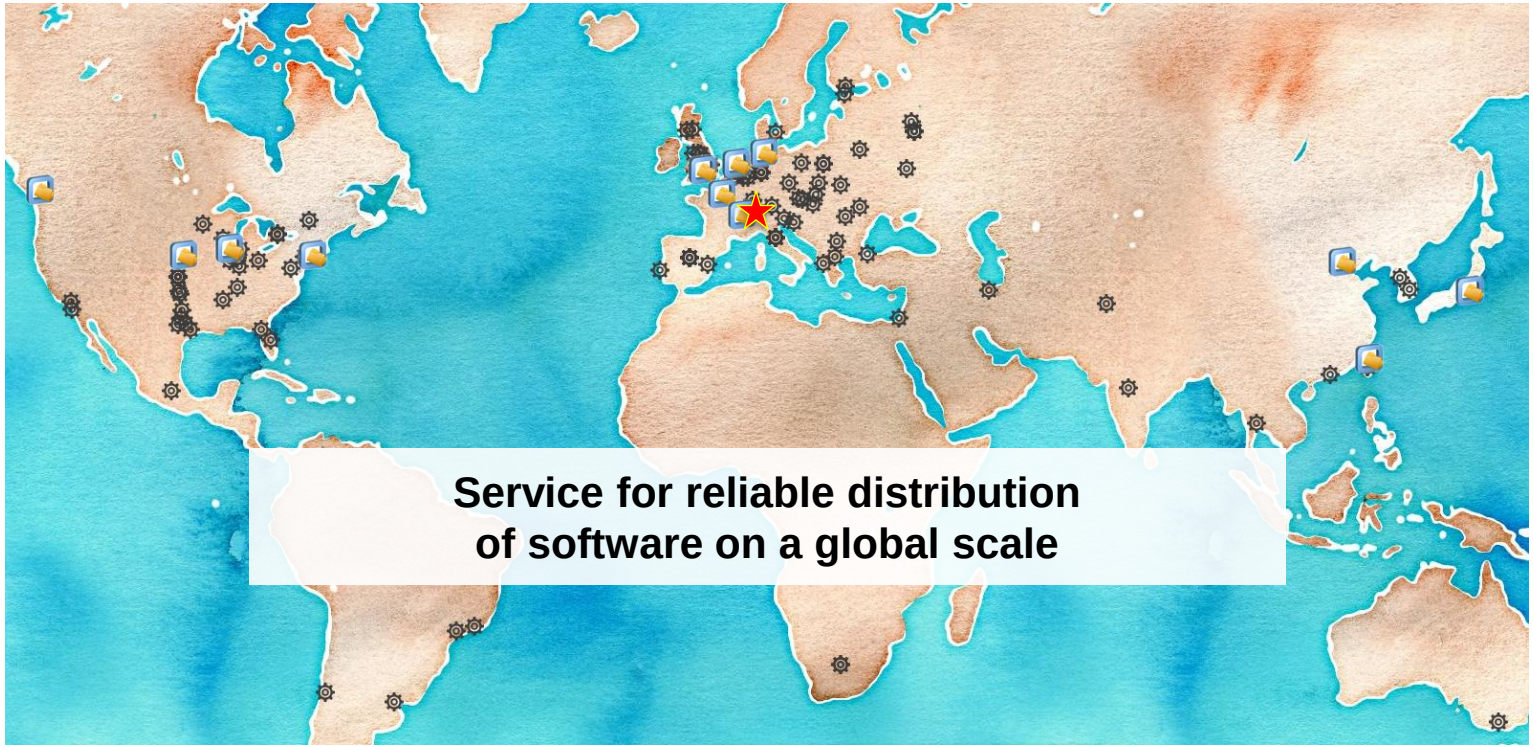


The CERN VM File System
as Global Software Delivery Fabric

Enrico Bocchi
CERN IT, Storage Group

CVMFS for Global Software Delivery

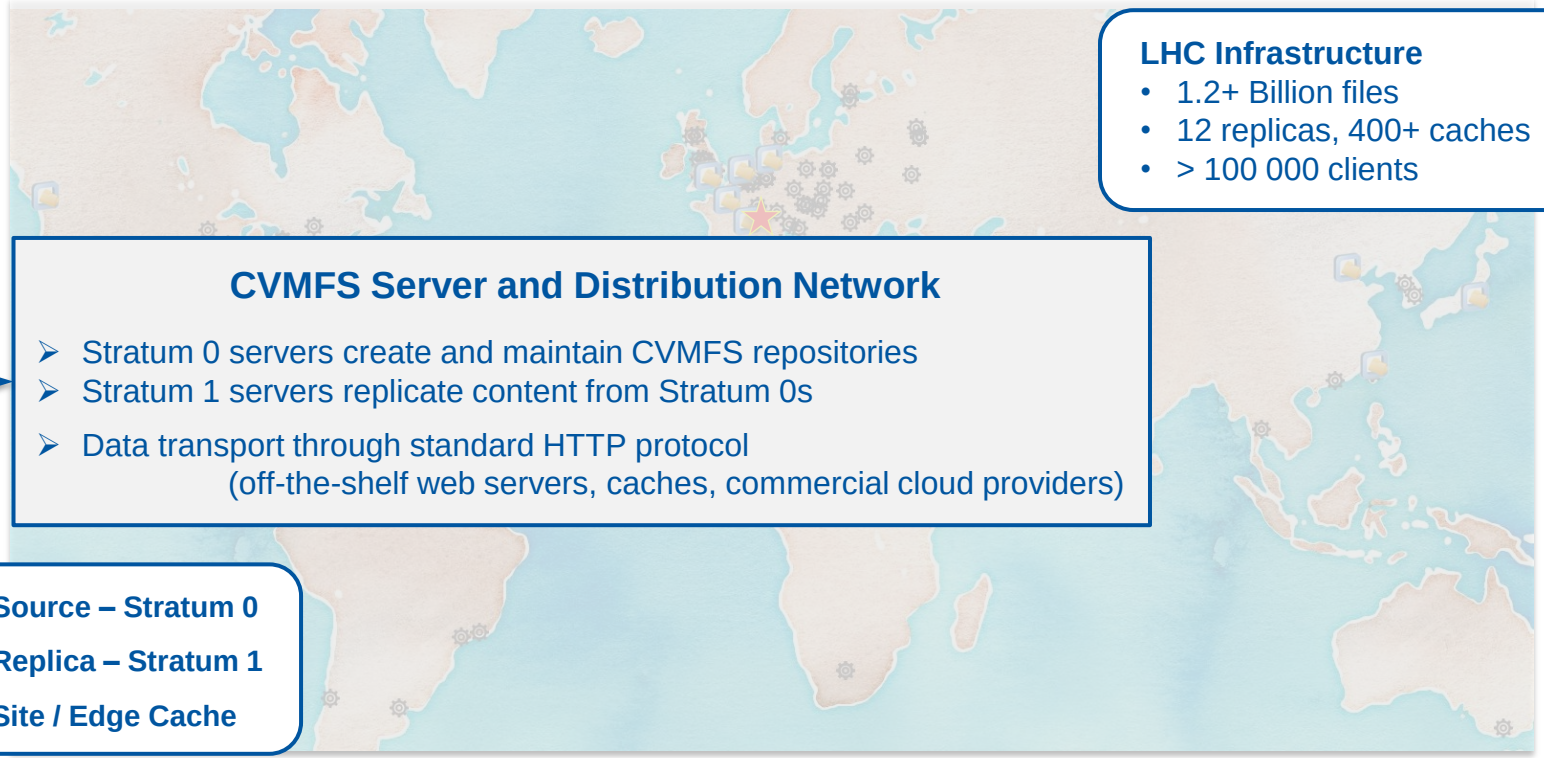


CVMFS for Global Software Delivery

LHC Infrastructure
- 1.2+ Billion files
- 12 replicas, 400+ caches
- > 100 000 clients

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

CVMFS for Global Software Delivery



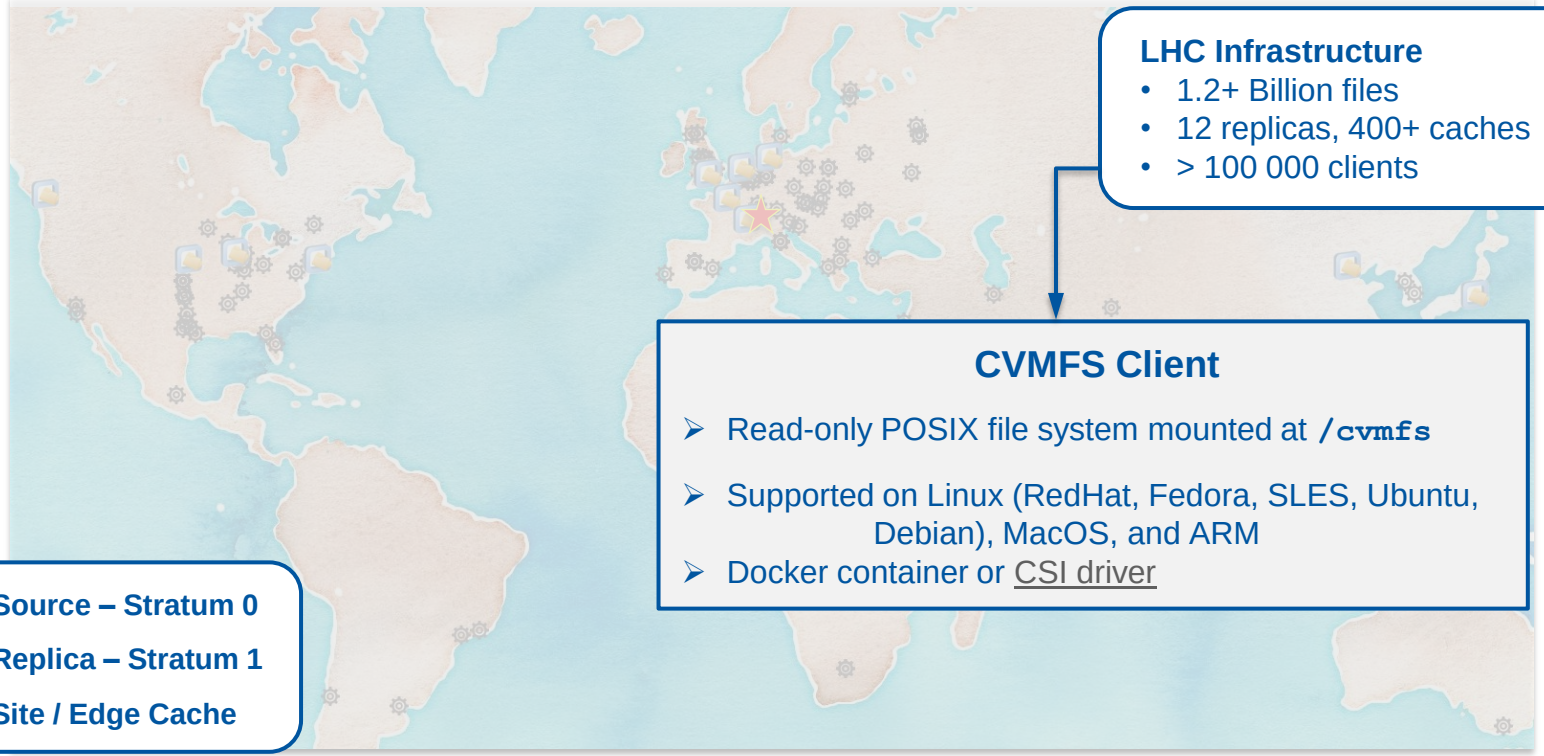
LHC Infrastructure
- 1.2+ Billion files
- 12 replicas, 400+ caches
- > 100 000 clients

CVMFS Server and Distribution Network

- Stratum 0 servers create and maintain CVMFS repositories
- Stratum 1 servers replicate content from Stratum 0s
- Data transport through standard HTTP protocol (off-the-shelf web servers, caches, commercial cloud providers)

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

CVMFS for Global Software Delivery



LHC Infrastructure
- 1.2+ Billion files
- 12 replicas, 400+ caches
- > 100 000 clients

CVMFS Client

- Read-only POSIX file system mounted at **`/cvmfs`**
- Supported on Linux (RedHat, Fedora, SLES, Ubuntu, Debian), MacOS, and ARM
- Docker container or CSI driver

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

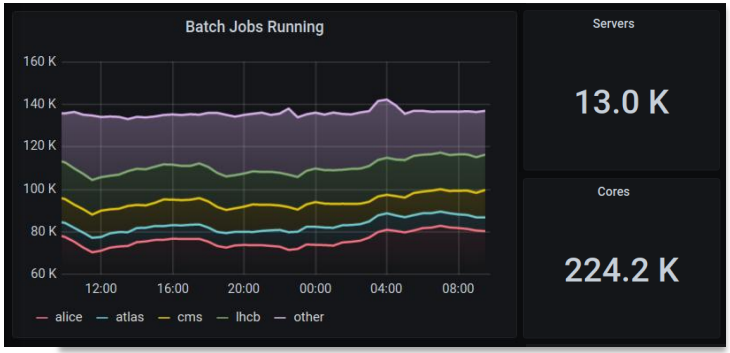
Use Cases

- Software Delivery for High Energy Physics
- “Small” Datasets Distribution
- Container Layers Ingestion and Distribution

Software Delivery at Scale

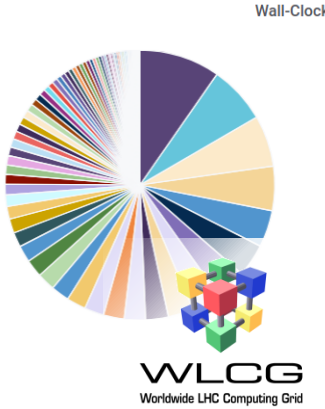
- Primary use case at CERN: High Energy Physics Software
  - 55 repositories, 70+ TB, 1.2B files
  - 30,000+ clients served via 23 caches (~9 TB)

- Ubiquitous CVMFS client
  - Batch Jobs
  - Software environments on lxplus
  - Hadoop clusters
  - Experiments' online farms
  - SWAN Jupyter Notebooks
  - Scientists' laptops



Software Delivery at Scale

- Worldwide LHC Computing Grid
  - 170+ computing centers, 40 countries



	total	percentage
CH-CERN	243.5 Mil	10%
US-FNAL-CMS	174.7 Mil	7%
US-T1-BNL	153.8 Mil	6%
US-MWT2	133.9 Mil	5%
T2_US_Wisconsin	110.5 Mil	4%
T2_US_Nebraska	98.9 Mil	4%
CA-TRIUMF	93.0 Mil	4%
T2_US_MIT	87.2 Mil	3%
RU-JINR-T1	75.9 Mil	3%
T2_US_Caltech	68.5 Mil	3%
T2_US_Florida	63.7 Mil	3%
US-NET2	62.3 Mil	2%

- Euclid Consortium, Astrophysics
  - Operate a decoupled CVMFS infrastructure across ~10 sites

- Large Synoptic Survey Telescope, Astronomical Survey

Distribution of “Small” Datasets

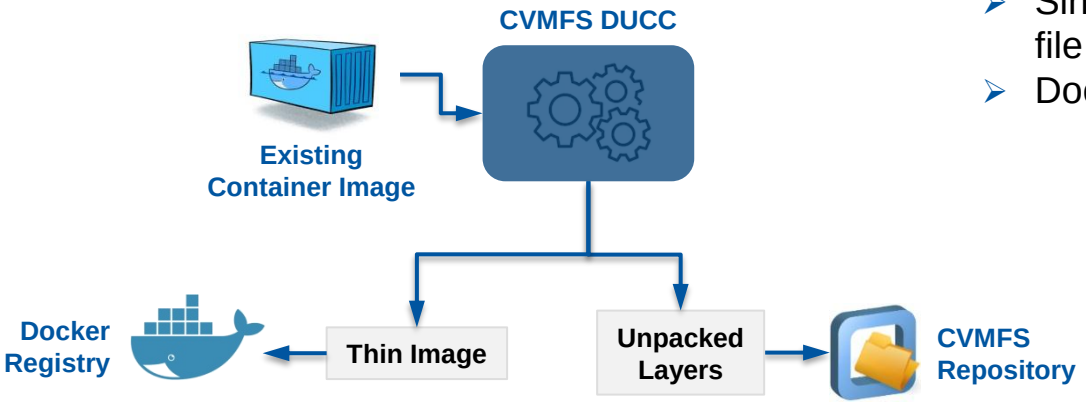
- Conditions data for major LHC experiments
  - 6 repositories, 1.39 TB, 1.69M files

- Physics Data for Gravitational Waves detectors
  - Virgo, Ligo, Kagra detectors produce ~5TB / year
  - CVMFS to export filesystem-like namespace
  - Actual payload is read from StashCache
  - Added authentication through x.509 plugin

S. Bagnasco, “*VIRGO and Gravitational Waves computing in Europe*”
https://indico.cern.ch/event/773049/contributions/3473801/

CVMFS for Container Images

- Server: Ingestion via DUCC
  - Publishes container images in their extracted form on CVMFS
  - Generates and uploads the *Thin Image* on Docker registries

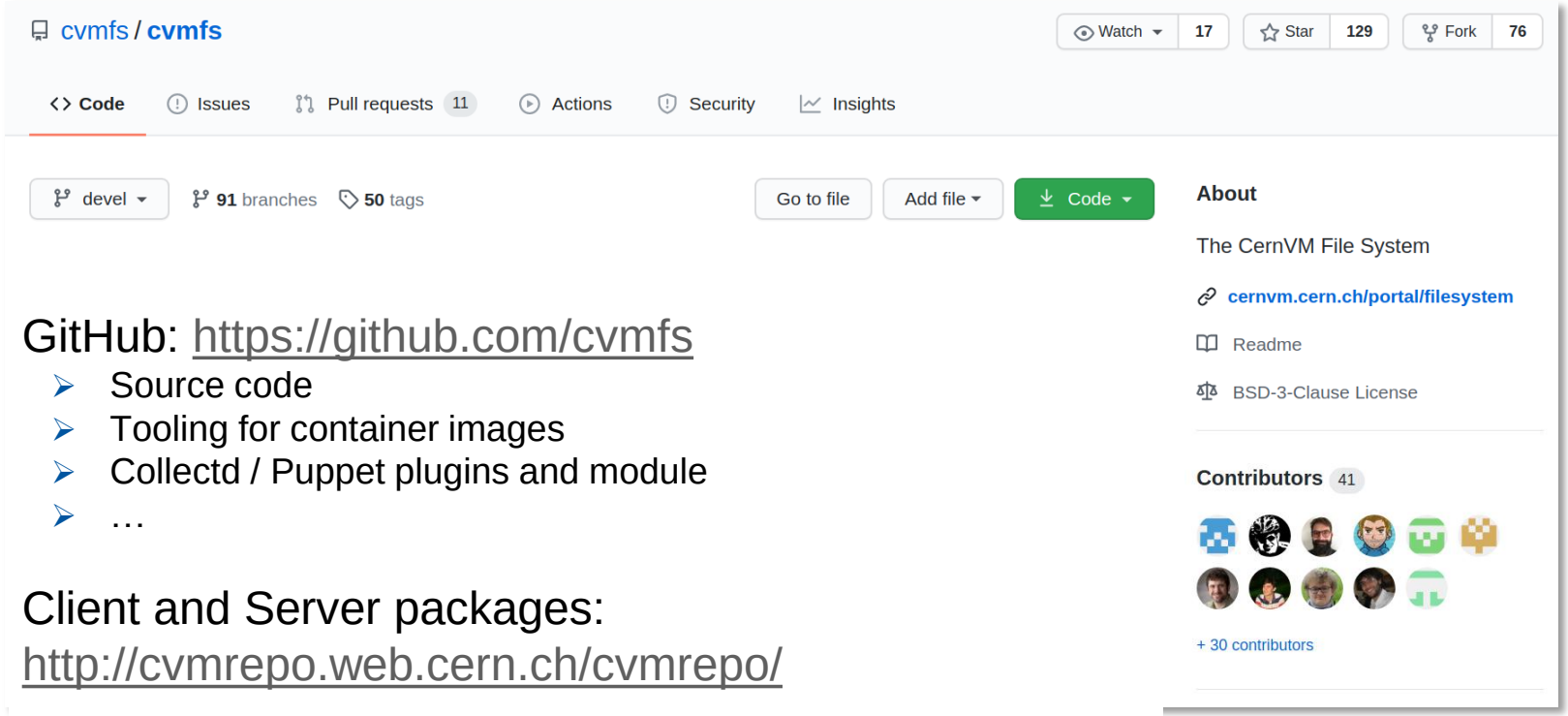


- Client
  - No download+extraction of images locally
  - Benefit from on-demand fetching of required files only
  - Singularity can use flat root file system directly from CVMFS
  - Docker requires *Graph Driver*

2

Where to Find CVMFS

Where to Find CVMFS



- GitHub: https://github.com/cvmfs
  - Source code
  - Tooling for container images
  - Collectd / Puppet plugins and module
  - …
- Client and Server packages: http://cvmrepo.web.cern.ch/cvmrepo/

Where to Find CVMFS

- Documentation:
  - http://cvmfs.readthedocs.io/en/stable/
  - http://cernvm.cern.ch/portal/filesystem

- Community mailing list:
  - cvmfs-talk@cern.ch

- CVMFS Workshop – Save the date!
  - CernVM Workshop 2021
    1-3 February 2021
    NIKHEF (Amsterdam, NL)

  - Last Workshop, June 2019: https://indico.cern.ch/event/757415/

The CERN VM File System
as Global Software Delivery Fabric

Thank you!

Enrico Bocchi
enrico.bocchi@cern.ch

CernVM
File system


CERN

Backup and Further Details

The CernVM File System

https://github.com/cvmfs/cvmfs

Write

- A publish-subscribe file system tuned for maximum dissemination

```
$ cvmfs_server transaction myrepo.cern.ch
$ cvmfs_server publish myrepo.cern.ch
```

- Publisher node is the single source of (new) data: read-write permissions
- Install applications once on the publisher, access from anywhere

Read

- POSIX file system access to globally available directory `/cvmfs`

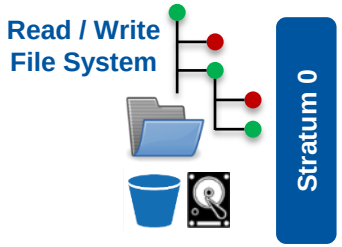
```
$ ls /cvmfs/myrepo.cern.ch
myFOLDER  myREADME.md
```

- HTTP-based read-only access
- RedHat, Debian, Ubuntu, macOS, …
- Clusters, cloud, supercomputers, end-user laptop

Infrastructural Components

- Tiered Servers: *Stratum 0s and 1s*
- Site Caches and CVMFS Clients

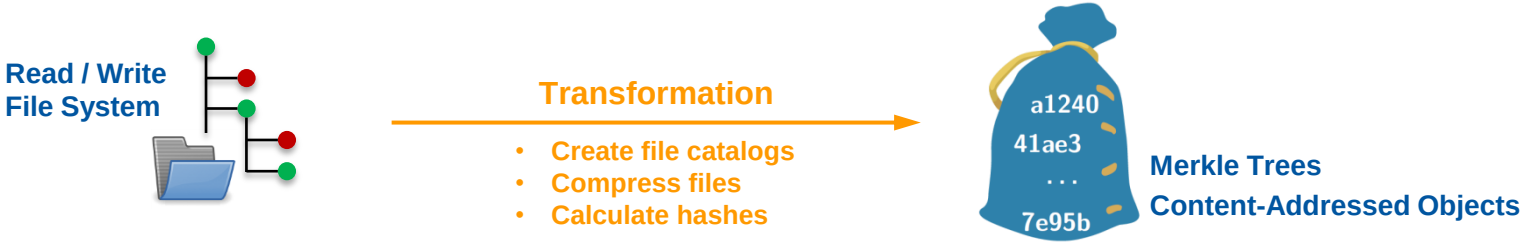
CVMFS Stratum 0s



- `cvmfs_server` package for repository administration
  - CLI to manage transactions and publish changes to the repository

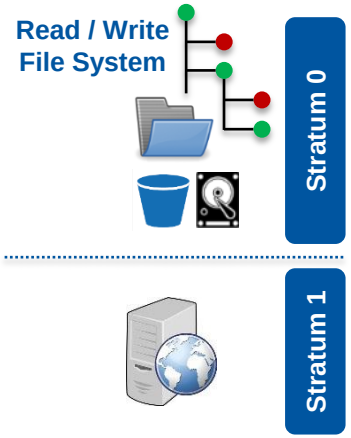
```
# cvmfs_server transaction myrepo.cern.ch
# tar xvf myarchive.tar.gz -C /cvmfs/myrepo.cern.ch
# cvmfs_server publish myrepo.cern.ch
```

  - Repository becomes writable on Stratum 0
  - Files are transformed into content-addressable objects
  - Authoritative storage can be local storage or S3 backend



CVMFS Stratum 1s

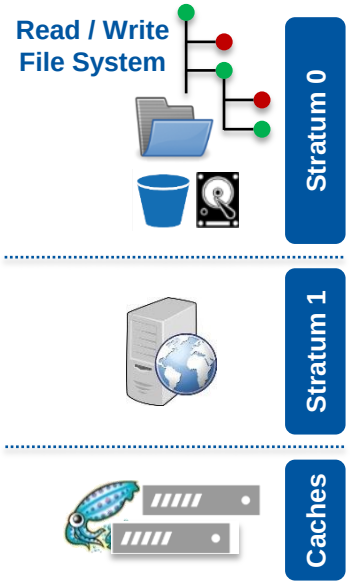
- Stratum 1 servers in Europe, US, Asia
  - Reduced RTT to caches and clients
  - Improved availability in case of Stratum 0 failure
- RESTful CVMFS GeoAPI service
  - Clients submit request with desired resource and Stratum 1s list
  - Stratum 1 returns sorted list of Stratum 1s
  - Based on MaxMind IP database




```
HTTP GET
http://s1.cs3.org/cvmfs/<desired_resource>/api/v1.0/geo/<list_of_known_stratum1s>
```

Site Caches

- Off-the-shelf HTTP caching software
- Squid-cache as forward proxy
  - Recommended for clusters of clients
  - Reduce latency to clients and load on Stratum 1s
- Take advantage of cloud based CDNs
  - OpenHTC on CloudFlare
  - Helix Nebula Cloud
    (RHEA, T-Systems, IBM Cloud)



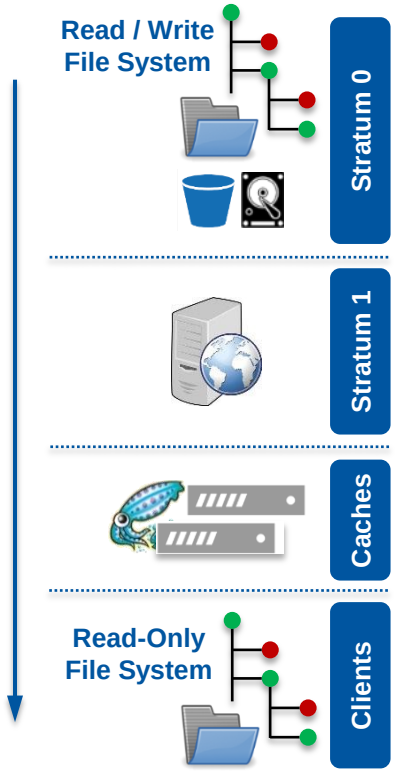
OpenHTC.io

CLOUDFLARE

HELIX NEBULA
THESCIENCECLOUD

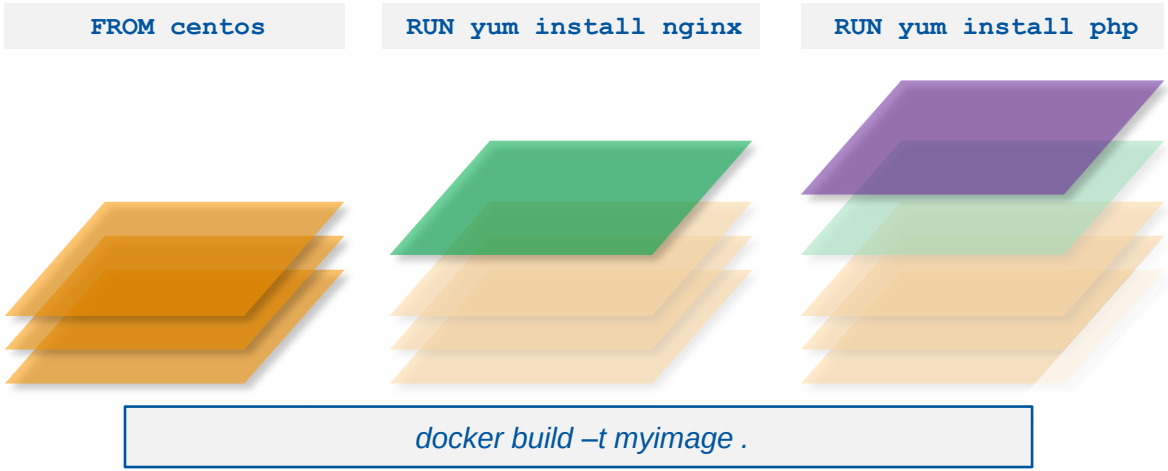
CVMFS Clients

- Client package for **`/cvmfs`** access
  - Runs on HPC, Supercomputers, end-users laptops
  - Kubernetes support with in-container mount or CSI driver
  - Dynamic mounting of repositories with autofs
- Local caching
  - Local file system (soft limit enforced)
  - Tiered: In-memory and disk
  - Alien: Cluster and Network file systems
- Embedded tools for troubleshooting and FS verification



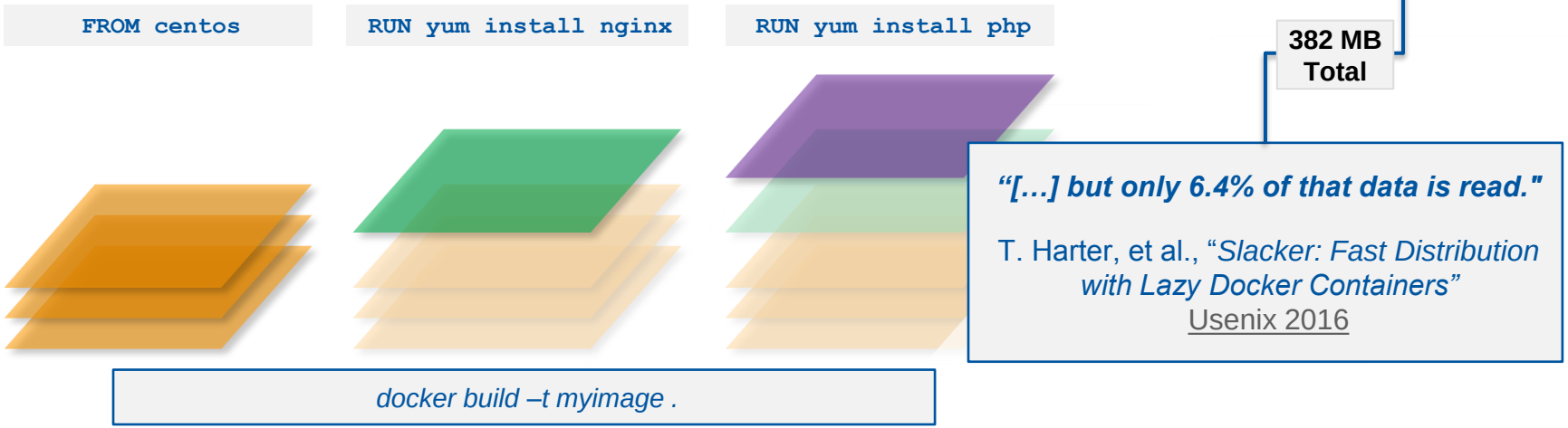
CVMFS for Docker Layers

- Docker images are the product of several layers
  - Layers are tarfiles
  - Need to be downloaded and extracted locally



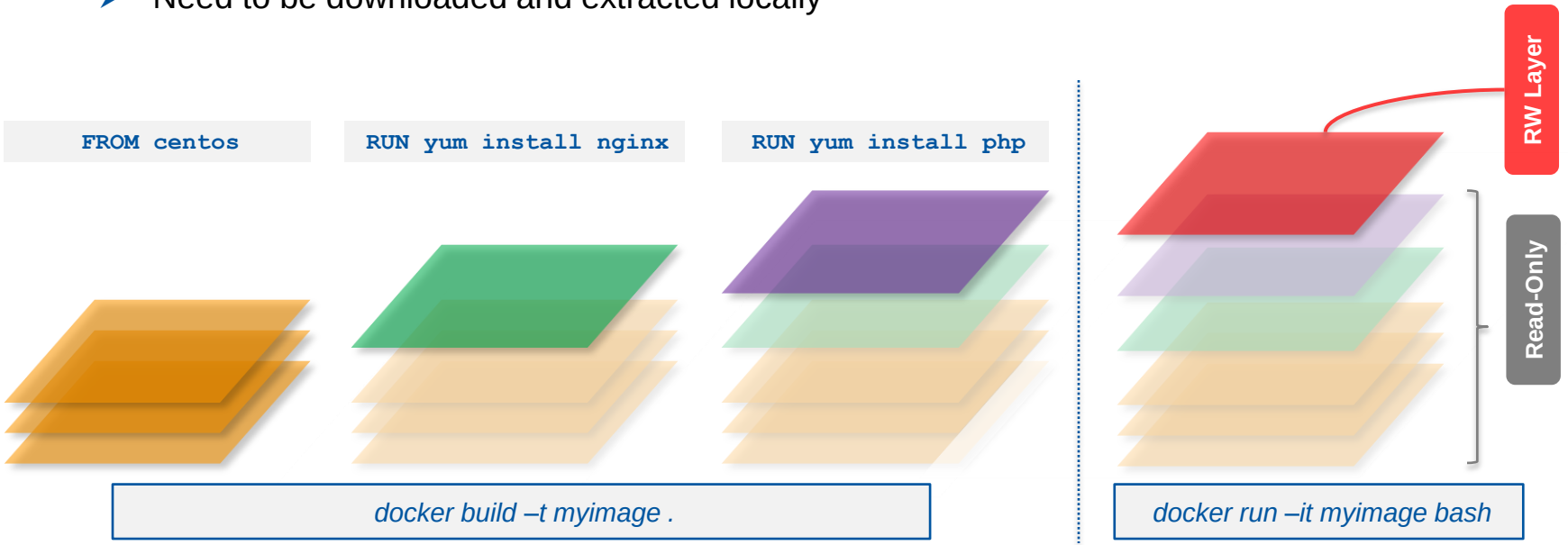
CVMFS for Docker Layers

```
[root@ThinkPad-X1]# docker history myimage
IMAGE               CREATED             CREATED BY                                      SIZE
75cc2375258a        4 seconds ago       /bin/sh -c yum -y install php                   66.9MB
e779b8a4024f        9 seconds ago       /bin/sh -c yum -y install nginx                 77.8MB
470671670cac        4 days ago          /bin/sh -c #(nop)  CMD ["/bin/bash"]            0B
<missing>           4 days ago          /bin/sh -c #(nop)  LABEL org.label-schema.sc…   0B
<missing>           7 days ago          /bin/sh -c #(nop) ADD file:aa54047c80ba30064…   237MB
```



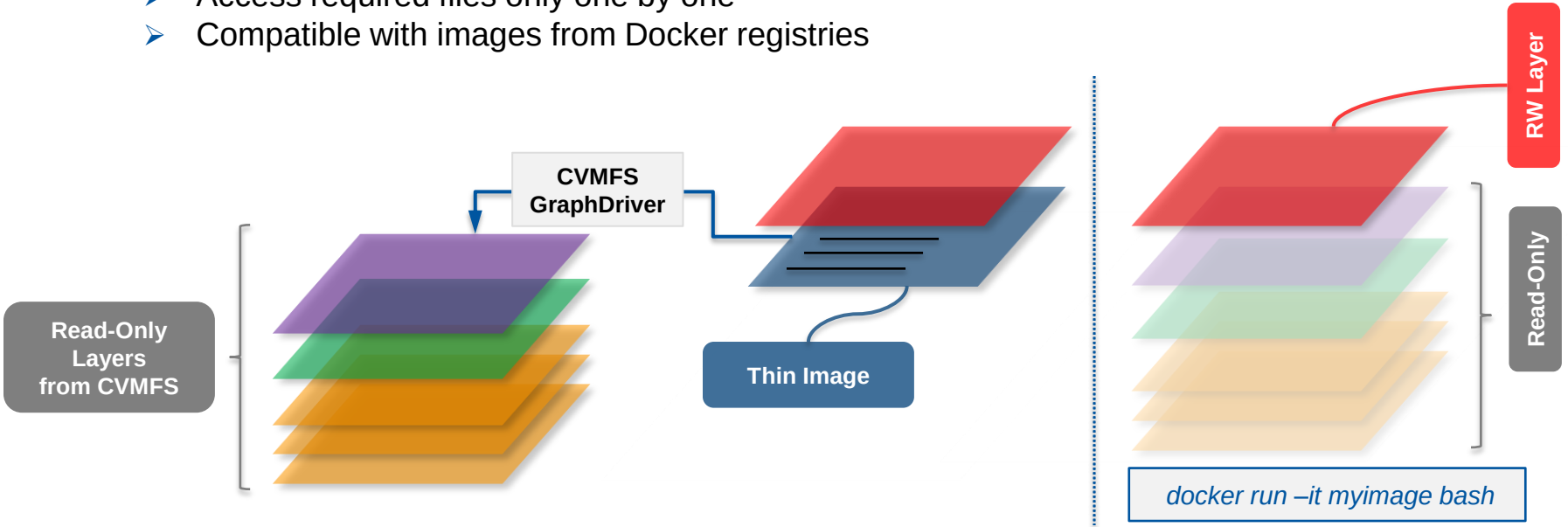
CVMFS for Docker Layers

- Docker images are the product of several layers
  - Layers are tarfiles
  - Need to be downloaded and extracted locally



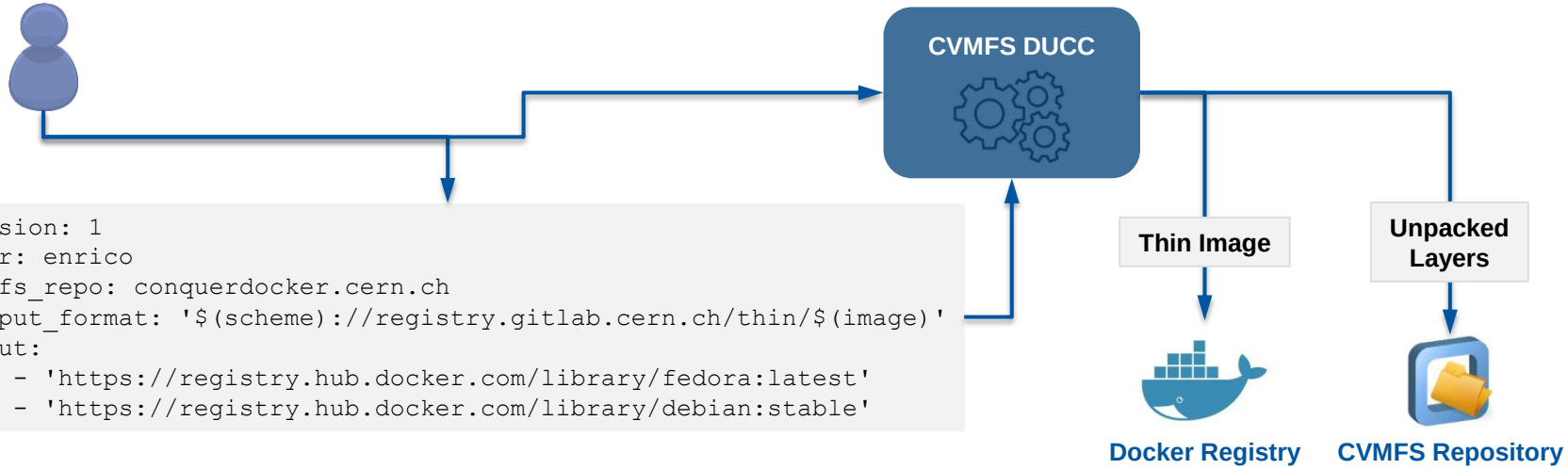
CVMFS serving Docker Layers

- CVMFS GraphDriver allows Docker daemon to use images from CVMFS
  - Images are stored in unpacked form
  - Access required files only one by one
  - Compatible with images from Docker registries



CVMFS ingesting Docker Layers

- DUCC: Daemon to convert and publish unpacked layers
  - Based on wishlist of Docker images to be ingested
  - Automatic generation and publication of thin image and unpacked layers




```
version: 1
user: enrico
cvmfs_repo: conquerdocker.cern.ch
output_format: '$(scheme)://registry.gitlab.cern.ch/thin/$(image)'
input:
    - 'https://registry.hub.docker.com/library/fedora:latest'
    - 'https://registry.hub.docker.com/library/debian:stable'
```

CVMFS Users' Contrib