WBS 2.3.2 CVMFS Content Management Approach 2021-08-24 Meeting

Dates

24 Aug 2021

Agenda, Discussion, and Action Items

Agenda & Discussion

Goal(s)

Figure out how to securely manage the content published from our pilot cvmfs.xsede.org server.

Related Information

Background material

- https://easybuild.io/eum/002_eum21_cvmfs.pdf

Discussion

Pilot Server Setup (Jeff)

1. Host cvmfs.xsede.org provisioned on Jetstream with local user accounts that match XSEDE usernames
    a. We agreed that XSEDE matching usernames is the best approach
2. Logins via password or ssh keys
    a. We agreed to configure login MFA right away for the pilot as this is how it would work in production
3. Basic configuration managed using Ansible, but not checked in to a GitHub/BitBucket repo yet
4. Worked with OSG to publish under the path 'colorado.xsede.org'
5. colorado.xsede.org content is updated by the the local 'colorado' account based on configured CVMFS privileges and unix permissions
    a. Might want to rename 'colorado' to ''cvmfs_colorado' or something similar to separate it from the XSEDE username space

Since we would likely have multiple use cases and publishers ask OSG:

- What are the costs for having multiple top levels exports besides 'colorado.xsede.org'?
- Can a single server host and publish multiple top levels?

Use Cases

- The five likely CVMFS use cases identified by Lee are: RC-08, -09, -10, and SPI-11 and -12
    - The RC ones are the researcher ones (using content), the SPI ones are the service provider ones (hosting content)
- Publishing use cases
    - Could be published by XSEDE SPs, XSEDE itself, or XSEDE allocated projects
    - Publishers, such as service providers, could publish directly on their own with XSEDE involvement if they have the capability to do so
    - Publishers could use an XSEDE wide publishing server if they can't publish on their own, like XSEDE allocated projects or community software software owners
- Would be good for Lee, Jeff, Shava, and JP to review the use cases to see if they can be made more specific (narrows) or if anything in the use cases is just plain wrong
- Derek suggested we focus on use cases that CVMFS is designed for and not get sidetracked by other things that it could be used for, like distributing packaged software
- Community software areas owners who don't have the ability to run their own servers could use XSEDE's CVMFS
    - This would be a supplement to a local network file-system community software area typically used to share HPC software
- XSEDE SPs typically install HPC software on a local network file-system (like onto ceph using spac at SDSC)
    - Installing in CVMFS would make that software available across HPC systems

Managing Content

Steps on cvmfs.xsede.org for the 'colorado' user to publish:

1. Configure CVMFS with the account that owns and can modify the contents of 'colorado.xsede.org'
2. The owner start an update transaction with command "cvmfs_server transaction colorado.xsede.org"
    a. This mounts /cvmfs/colorado.xsede.org which the owner has write access to
3. The owner modifies contents

4. The owner closes transaction by running "cvmfs_server publish"

Security considerations:

- Server logins should follow standard login logging practices
- We need audit logs of who update files: either who used sudo or whatever cvmfs logs

Alternate content management approach:

- By default the owner updates the mounted file-system directly as shown above e
- Have a separate local file-system protected using standard permission and updated first
- Use a sudo wrapped command that initiates the transaction, rsync's to local file-system to the publishing file-system, and terminates the transaction
- Don't see any immediate value for in using GCS v5 to put content on the server
- Encourage SPs who have the ability to publish on their own to do so directly through OSG and not rely on XSEDE

Production platform

- Pilot is on Jetstream
- Production could be on Jetstream2 or AWS

Jim recommended a pilot report to hand over to ACCESS rather than implementing in production before the end of XSEDE.

Action Items

□ Jeff Makey will configure MFA on cvmfs.xsede.org

□ Jeff Makey will check-in the Ansible recipes into XSEDE's server configuration repository

☑ Jeff Makey will rename the 'colorado' account to 'cvmfs_colorado', which does not potentially collimate with XSEDE usernames

- usermod -l cvmfs_colorado -d /home/cvmfs_colorado -m colorado
- groupmod -n cvmfs_colorado colorado
- change *CVMFS_USER=colorado* to *CVMFS_USER=cvmfs_colorado* in /etc/cvmfs/repositories.d/colorado.xsede.org/server.conf

□ Jeff Makey will contact OSG about the cost of multiple top levels and publishing them from a single server, share results with JP and Shava

□ Jeff Makey will share our proposed "Alternate content management approach" with OSG folks for feedback

□ Lee LimingJeff MakeyJP NavarroShava Smallen will review and fine tune the use cases

□ Jeff Makey will implement the alternate content management approach if OSG considers it reasonable

□ Jeff Makey will confirm and share with this team what is logged when published files are changed

□ JP Navarro and Shava Smallen to explore with Colorado if they could publish on their own

□ JP NavarroLee Liming and Shava Smallen to explore potential XSEDE allocated projects and community software areas would prefer to use CVMFS

Attendees

- Jeff Makey ✓
- Jim Basney ✓
- Derek Simmel ✓
- Eric Blau ✓
- Lee Liming ✗
- JP Navarro ✓
- Shava Smallen ✗