S3 Storage with DynaFed for Dune

Alastair Dewhurst

Introduction

- RAL is currently committed to providing 1PB of storage to DUNE.

- This space is provided on our Echo storage service which provides an S3 / Swift API.

- RAL provides DynaFed which acts as the authentication and authorization layer.

  - Uses a Grid-mapfile.

- Rob Illingworth et al managed to resolve many problems and get transfers working.

  - Blocking issue hit when hard limit of 5GB on single file transfers into S3.

Davix

- WebDav is the extension of http(s) that allows clients to perform remote web content authoriing operations.

- Davix is the CERN implementation of WebDav.

  - It contains some optimizations for S3 endpoints.

  - It is the client that the gfal- commands use.

- Multi-part uploads were added in Davix 0.7.0 which was released on October 22nd 2018.

  - Should use Davix 0.7.1 which is now available in EPEL.

/gridpp/dune/test/

Mode	Links	UID	GID	Size	Modified	Name
-rwxr-xr-x	0	0	0	17	Fri, 21 Sep 2018 13:07:19 GMT	another_test.txt
-rwxr-xr-x	0	0	0	5.9G	Mon, 26 Nov 2018 12:23:49 GMT	deleteme1

Alastair Dewhurst, 26th November 2018

echo

DynaFed as an S3 Gateway
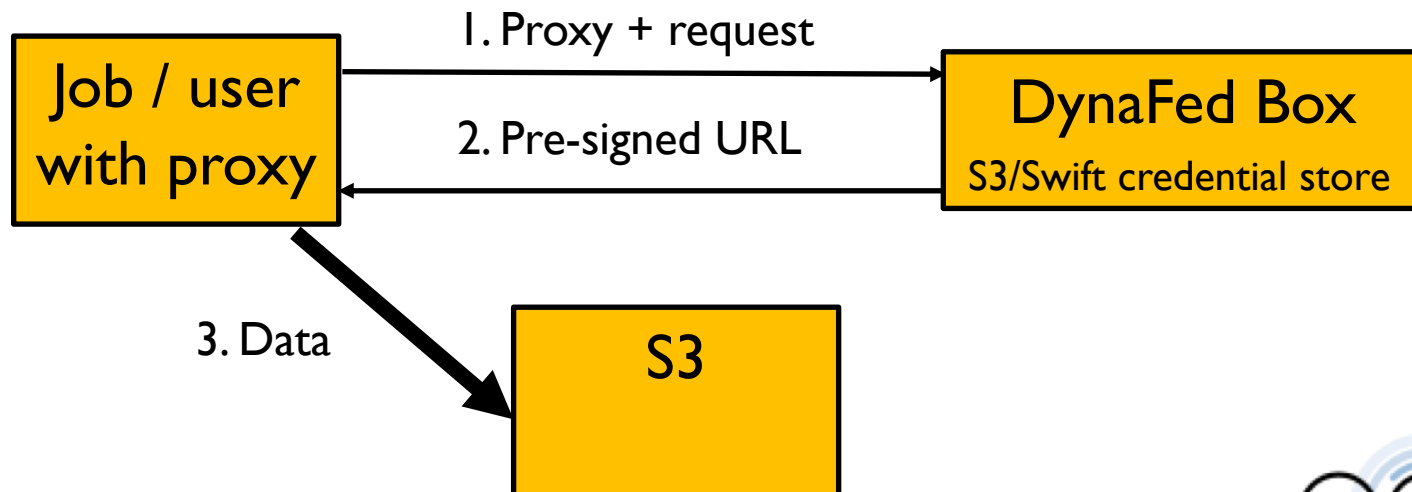
Directly to S3 endpoint:
davix-ls --s3alternate --s3secretkey XXXXX --s3accesskey YYYYY s3s://s3.echo.stfc.ac.uk/dune-test/
davix-put --s3alternate --s3secretkey XXXXX --s3accesskey YYYYY testfile s3s://s3.echo.stfc.ac.uk/dune-test/testfile

With Dynafed:
voms-proxy-init
davix-ls -P grid davs://dynafed.stfc.ac.uk/gridpp/dune/test/
davix-put -P grid testfile davs://dynafed.stfc.ac.uk/gridpp/dune/test/testfile
Or
gfal-ls davs://dynafed.stfc.ac.uk/gridpp/dune/test/
gfal-copy file:///home/tier1/dewhurst/testfile davs://dynafed.stfc.ac.uk/gridpp/dune/test/testfile2

Job / user with proxy

1. Proxy + request

2. Pre-signed URL

DynaFed Box
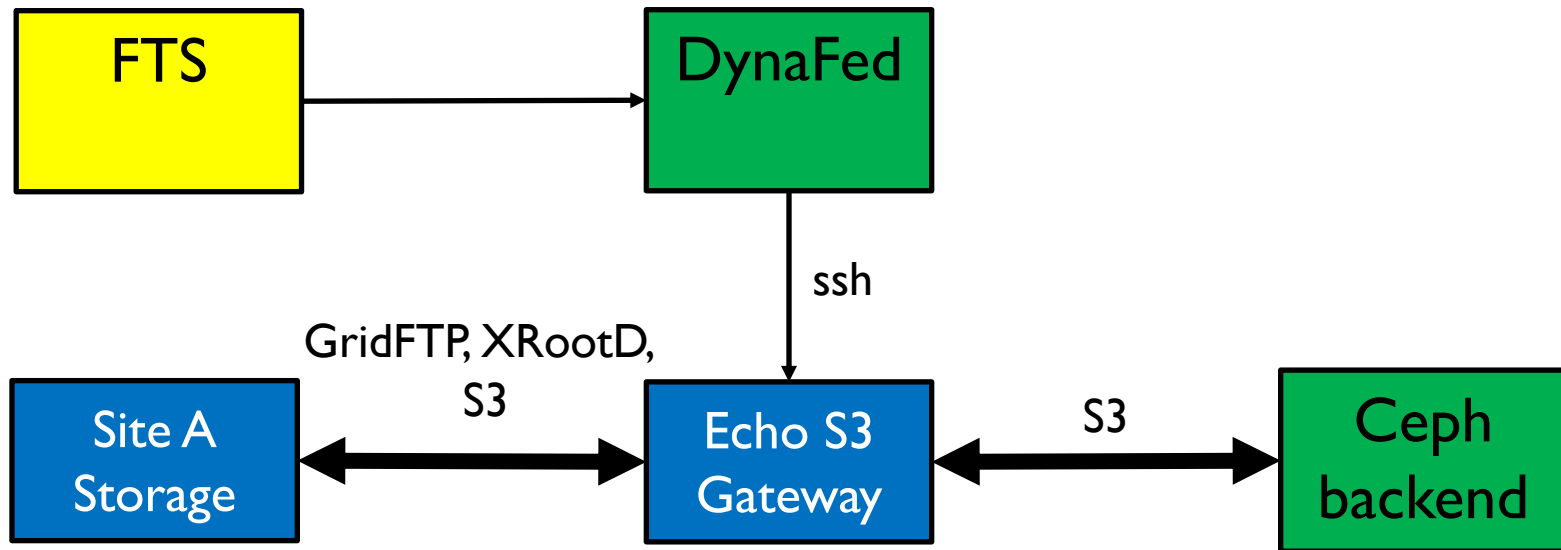S3/Swift credential store

3. Data

S3

Third Party Copies

- S3 can only be the passive partner in a TPC

- When performing a TPC the FTS Server will try:

  - Push (Source active)

  - Pull (Destination active)

  - Stream (Proxy transfers through FTS client)

- DynaFed 1.4 introduces significant improvements to the way DynaFed handles TPC.

  - Release Candidate was produced today!

  - Will be tested (and hopefully deployed) this week.

DynaFed Setup

```
┌──────────┐                    ┌──────────┐
│   FTS    │ ─────────────────► │ DynaFed  │
└──────────┘                    └──────────┘
                                     │
                                     │ ssh
    GridFTP, XRootD,                 ▼
         S3
┌──────────┐      ┌──────────┐  S3  ┌──────────┐
│ Site A   │◄────►│ Echo S3  │◄────►│  Ceph    │
│ Storage  │      │ Gateway  │      │ backend  │
└──────────┘      └──────────┘      └──────────┘
```

- If DynaFed is requested to mediate the transfers it will "stream" it by running a script:

  - https://svnweb.cern.ch/trac/lcgdm/browser/ugr/trunk/src/utils/ugrpullscript_gfal.sh

- We can modify this to ssh into an Echo gateway and stream the transfer through that.

- This should offer "full TPC support" for all protocols!

Alastair Dewhurst, 26th November 2018

Host Certificates

- RAL use QuoVadis commerical certificates on our S3 endpoints.

  - Web browsers will automatically trust it.

- Problems for Grid applications as the certificates are not IGTF approved.

  - Google, Amazon, Microsoft have similar problem.

  - FermiLab have already worked around this.

- Long term aim to change Grid Policy.

- Short term solution would be for RAL to provide an RPM which other sites could install and trust us.

Checksumming

- S3 does not currently store Checksums.

  - You can store arbitrary meta data with S3.

- Transfers via https will check file integrity which is where a significant amount of file corruption occurs.

- ATLAS (Frank Berghaus) are working on getting checksums to work with DynaFed and S3.

  - No ETA on when this will be ready.

  - Frank has produced work arounds for Rucio.

Conclusions

- Large file problem is solved for S3 endpoints.

- RAL will be testing and intends to upgrade the DynaFed endpoint this week.

- Big improvements in the range of transfers that should work to S3.

- Needs to keep Grid-mapfile up to date:

  - Who can view this:
    https://dynafed.stfc.ac.uk/gridpp/dune/