

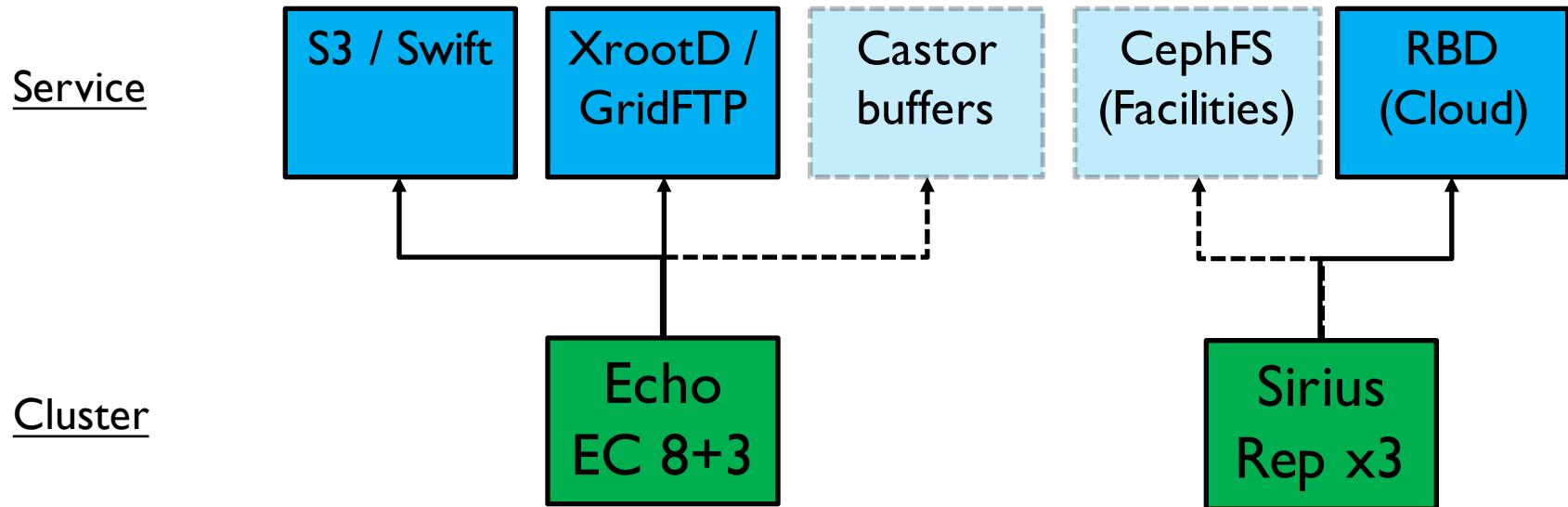
# Echo status and plans

Alastair Dewhurst, Alison Packer, Bruno Canning,  
George Vasilakakos, Ian Johnson, Tom Byrne



# Introduction

2



- GridFTP and XRootD now in production.
- Q2 upgrading Sirius to provide production quality service for Cloud.
- S3 / Swift to enter production July – Sept 17.
- No current requirement for “Castor on Ceph”
- Plenty of Castor only hardware can be used in Tape buffers.



# Echo cluster

- To early to make any insightful comments about running a large Ceph cluster.
- User community is much more active than Castor!
- We are leading the way with Erasure Coding.
  - Lots of interest in improving things for EC.
  - Stable version = Claimed stable version + 1

	Castor Currently / TB	Capacity to decommission / TB	New capacity added to Castor / TB	Echo allocation / TB	Capacity provided by Echo
<b>ALICE</b>	480	0	0	0	0%
<b>ATLAS</b>	5257	1693	980	3100	41%
<b>CMS</b>	2287	1337	653	2500	61%
<b>LHCb</b>	4906	1010	545	1500	25%



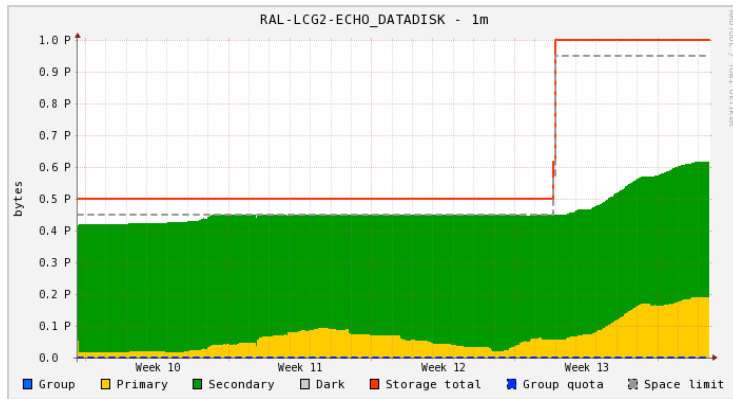
# GridFTP plugin

- GridFTP plugin was completed at start of October 2016.
- Average time to download single 2GB file: 40.7s.
  - On Castor 32.4s.
- ATLAS are using GridFTP for all transfers currently.
  - Until we are happy with XRootD.
- CMS Debug traffic and load tests also use GridFTP.
- Recently improvements have been made to:
  - Transfer timeouts.
  - Check-summing (to align with XRootD).
- Interest from US sites who have large Ceph clusters.

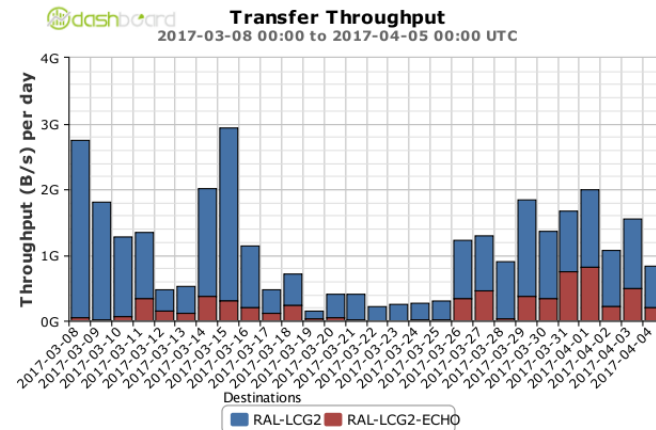


# ATLAS DDM

## Echo Space Usage



## Castor vs Echo Load

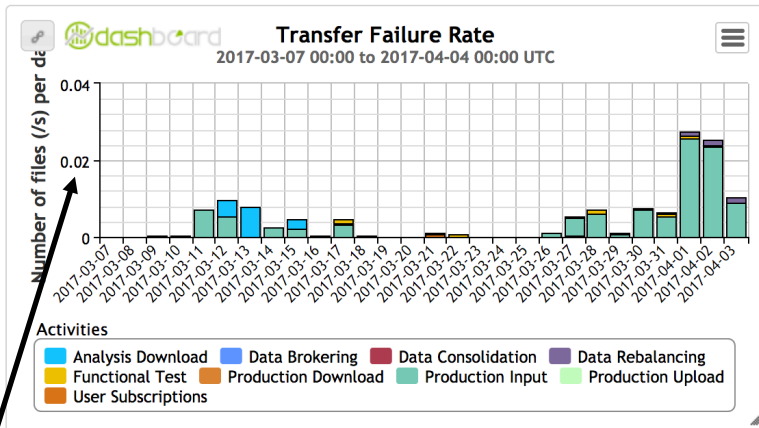
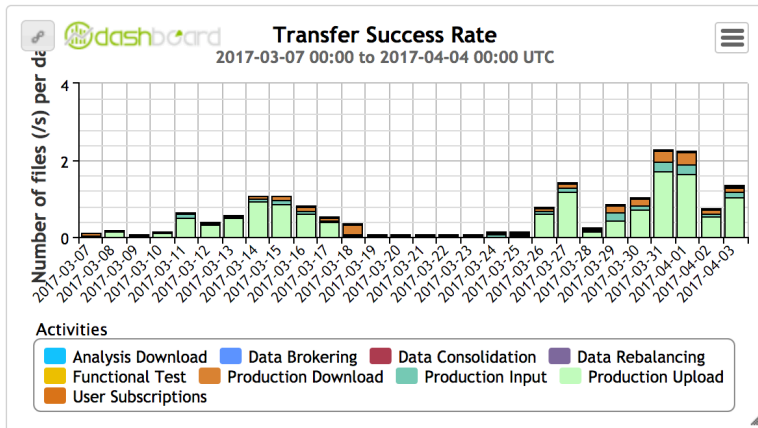
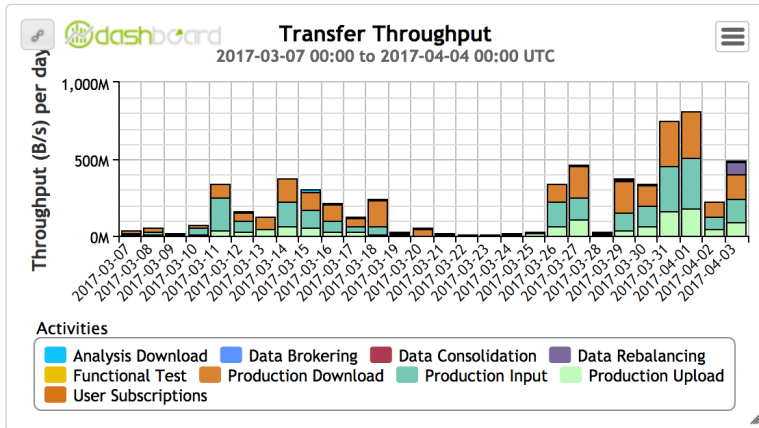
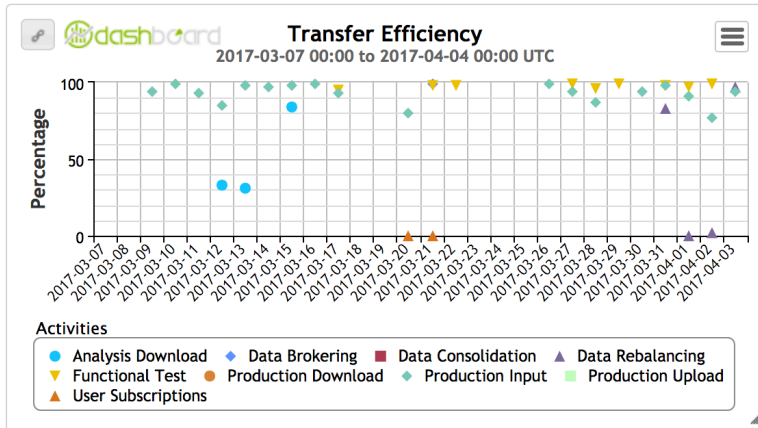


- Increasing ATLAS space on Echo by 500TB each week.
- Don't want to trigger large data migration.
- Can scale back if we hit load issues.



# ATLAS Transfers

## All ATLAS transfers to Echo in last month



Error rate < 1% and normally not Echo's fault

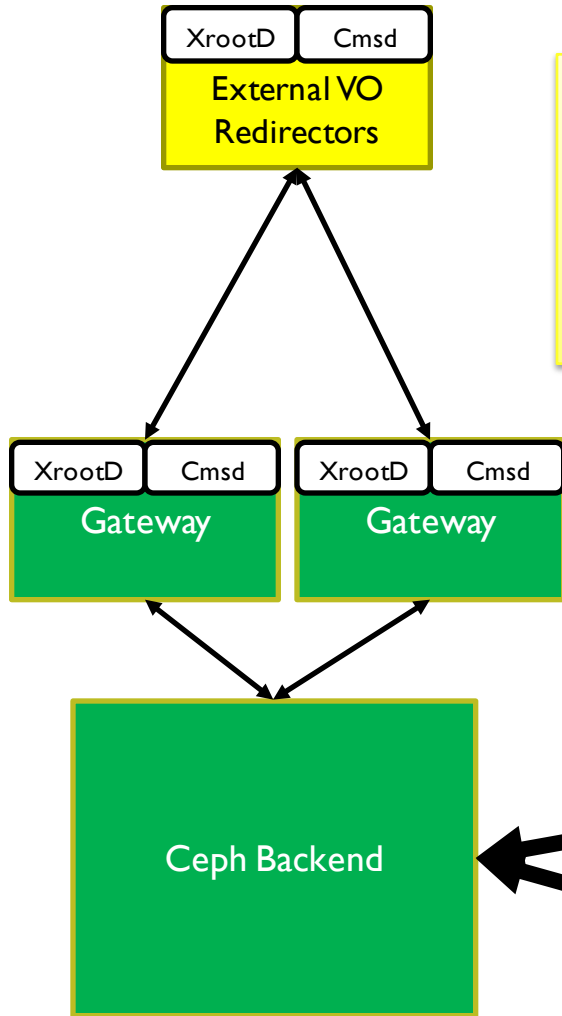


# XRootD plugin

- XRootD plugin was developed by CERN.
  - They have a very restricted use case.
  - They switched from EC back to using replication.
- We have a much more complex setup.
  - Optimizing direct I/O from Echo is biggest challenge.
- We have encountered many bugs:
  - Check-summing didn't work (Fixed).
  - Redirection didn't work (Fixed).
  - Caching proxy didn't work (Fixed).
  - File overwrite doesn't work.
  - N2N component doesn't work (Needed for CMS tests).



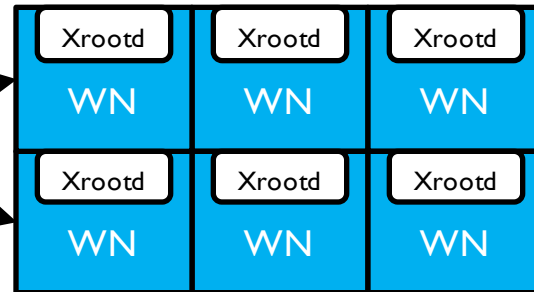
# XRootD Architecture



All XRootD gateways will have a caching proxy:

- On External gateways it will be large as AAA from other sites doesn't use LD.
- On WN gateways it will be small and used to protect against pathological jobs.

WN will be installed with an XrootD Gateway. This will allow direct connection to Echo. 25% of batch farm now using SL7 + containers.





# CMS Testing

- CMS have been testing XRootD access
  - Using Lazy-Download (fetches 128MB blocks at a time)
- Average file opening time: 0.01s
  - Castor has always struggled with this (50 times slower)

<b>Workflow</b>	<b>Echo with LD (Time / Event)</b>	<b>Echo no LD (Time / Event)</b>	<b>Castor with LD (Time / Event)</b>
<b>HIRun2015</b>	5.3	7.5	5.2
<b>PhaseIFall16DR82</b>	230	410	250
<b>PhaseIFall16GS82</b>	190	440	140
<b>PhaseIFall16GS</b>	37	91	29



# LHCb Plans

- Under the original migration plan, LHCb were going to move to Echo (primarily) next year.
  - And then the allocations got increased.
- Waiting for XRootD issues to be resolved.
- Will rely on copy-to-scratch model for jobs initially.
  - Some user jobs may suffer.
- LHCb can't use two separate storage elements as easily as CMS or ATLAS.
  - Clone each LHCb space token in turn and then switch over.



# New users on Echo



Alastair Dewhurst, 6<sup>th</sup> April 2017



# S3 / Swift

- We believe S3 / Swift are the industry standard protocols we should be supporting.

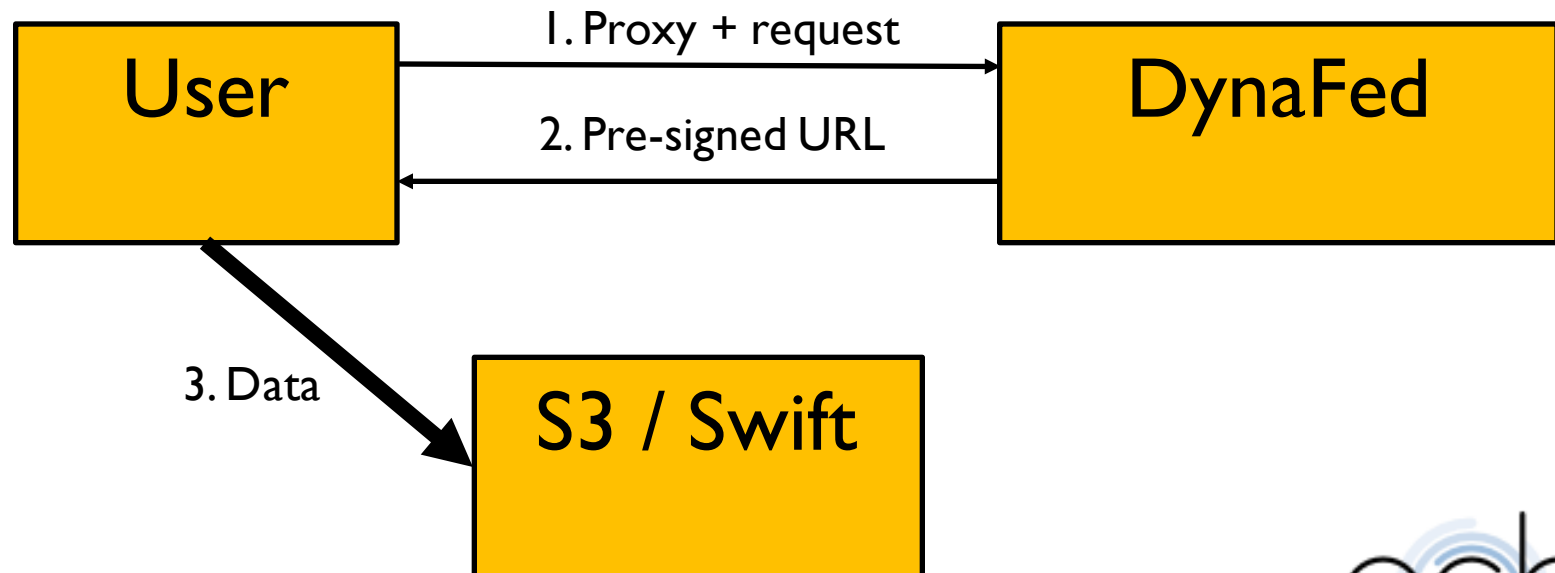
S3 / Swift API access to Echo will be the only service offered to new users wanting disk only storage at RAL.

- If users want to build their own software directly on top of S3 / Swift, that's fine:
  - Need to sign agreement to ensure credentials are looked after properly.
- We expect most new users will want help:
  - Currently developing basic storage service product that can quickly be used to work with data.
  - Focus on ease of use.

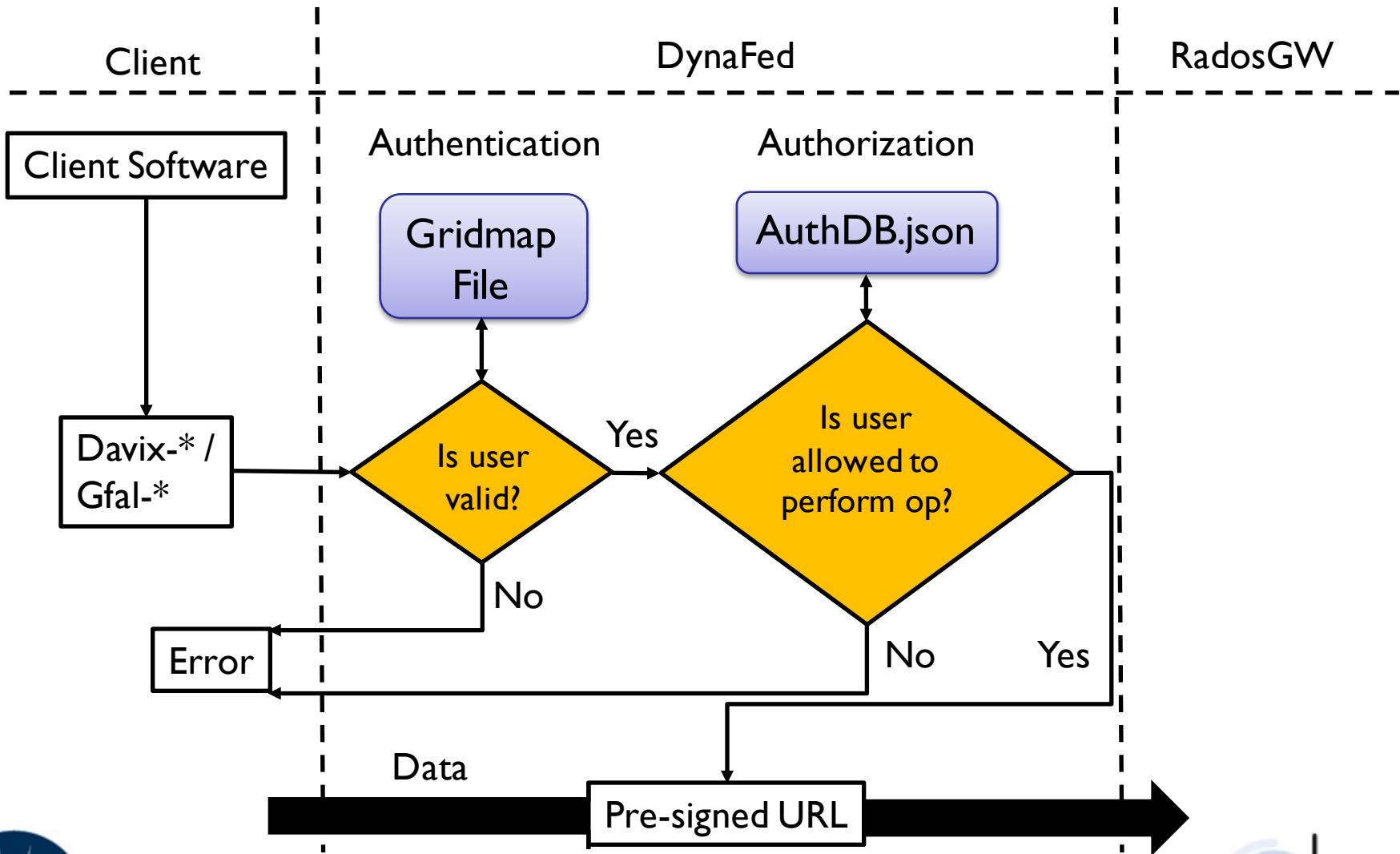


# DynaFed

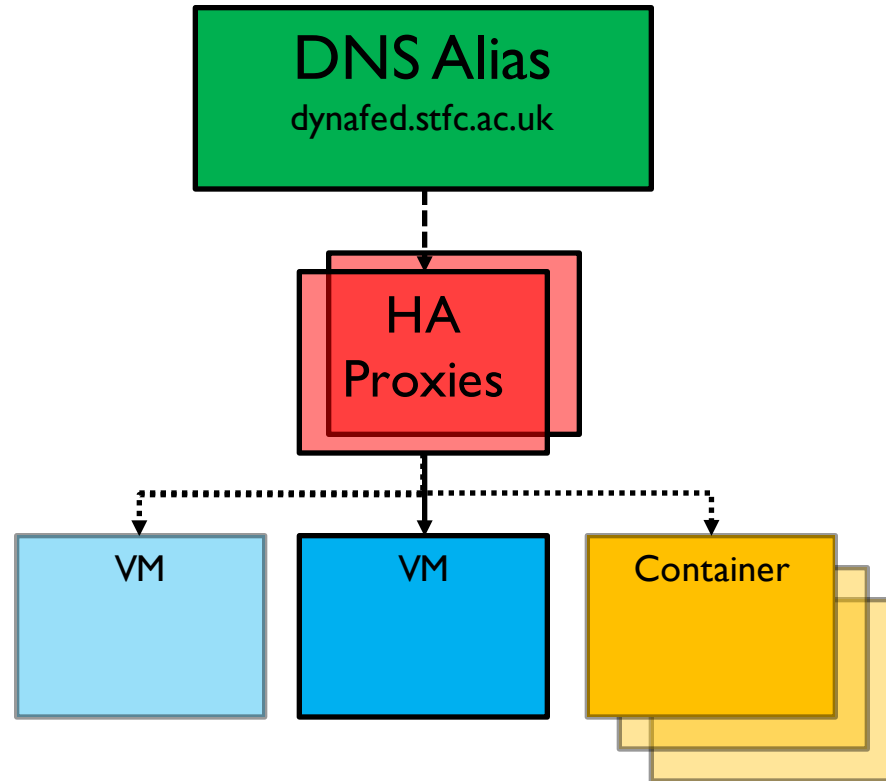
- We believe DynaFed is best tool to allow small VOs secure access.
- S3/Swift credentials stored on DynaFed Box.
- Users use certificate/proxy.
- Provides file system like structure.
- Good support for transfers to existing Grid storage.



# DynaFed AuthZ



# RAL Setup



- Service is behind HA proxy.
- Currently just one VM but easily scalable.



# Try it!!

Anyone with either atlas or dteam certificate has been granted access.

## CLI

```
# voms-proxy-init
# davix-ls -P grid davs://dynafed.stfc.ac.uk/gridpp/dteam-disk/
# davix-put -P grid testfile davs://dynafed.stfc.ac.uk/gridpp/dteam-disk/testfile
Or
# gfal-ls davs://dynafed.stfc.ac.uk/gridpp/echo/
# gfal-copy file:///home/tier1/dewhurst/testfile davs://dynafed.stfc.ac.uk/gridpp/dteam-disk/testfile2
```

# <https://dynafed.stfc.ac.uk/gridpp>

The screenshot shows a web interface for the 'gridpp' workspace. On the left, a sidebar lists the workspace 'DYNAFED.STFC.AC.UK' and the current directory 'gridpp'. Below this, a list of sub-directories is shown: atlas-eventservice, atlas-logs, azure, dteam-disk, lancaster-atlas-es, and lancaster-atlas-logs. The main area displays a table with columns for Metalink, Filename, Size, and Modified. The table is currently empty. Above the table, there are controls for 'New directory', 'Delete directory', 'Upload', and 'Download'. A search bar is also present with the text 'All Fields' and a 'Search...' button.

More commands (including testing 3<sup>rd</sup> party copy to/from your site) can be found here:  
<https://indico.cern.ch/event/505613/contributions/2230903/attachments/1346560/2040939/Oral-042.pdf>





# Summary & Plans

- Echo is now a production service providing some of the LHC VOs allocation of disk storage.
- ATLAS have started to move a significant fraction of their work on to Echo relying on GridFTP plugin.
- XRootD debugging and development ongoing.
- In the process of designing a new storage service:
  - Future facing but still want to be able to work with rest of GridPP sites.
  - VOs may need to evolve their storage models.

