



# Making the most of cloud storage - a toolkit for exploitation by WLCG experiments.

Alejandro Alvarez Ayllon, Maria Arsuaga Rios, Georgios Bitzes, Fabrizio Furano, Oliver Keeble (presenter), Andrea Manzi

# Object stores

- Advantages
  - Scalability and performance achieved through relaxing or abandoning many aspects of posix
  - Applications must be aware or adapted
- How can such resources be plugged into existing WLCG workflows?
  - Can apply to public or private cloud
    - NB ceph at sites

# Integration challenges

- How do you get data in and out?



- FTS
- DPM [ & dCache]

- How do you access the data using familiar tools?
- ... at an acceptable performance?



- Davix
- Gfal2
- [ROOT]

- How do you deal with the authentication and authorisation system?
- How do you deal with the lack of a namespace?



- Dynafed

# Transfer

# davix

- `davix-put /etc/services  
https://objbkt1.s3.amazonaws.com/file01 --s3secretkey  
<secret> --s3accesskey <access>`
- `davix-cp -P grid  
davs://dpm.cern.ch/dpm/cern.ch/home/dteam/file01  
s3s://objbkt1.s3.amazonaws.com/file01 --s3secretkey  
<secret> --s3accesskey <access>`



# gfal2/davix

- `gfal-copy file:///etc/services  
s3://objbkt1.s3.amazonaws.com/file01`
- `gfal-copy  
davs://dpm.cern.ch/dpm/cern.ch/home/dteam/fil  
e01 s3://objbkt1.s3.amazonaws.com/file01`



# FTS: Pre-signed URL

```
fts-transfer-submit --strict-  
copy -s
```

```
https://fts3.cern.ch:8446
```

```
https://dpm.cern.ch/dpm/cern  
.ch/home/dteam/file01
```

```
'https://objbkt1.s3.amazonaws  
s.com/tf_04?Signature=eFAy  
XMWISY%2BWEVcqfvGvux  
ZF6ZQ%3D&Expires=21057  
74242&AWSAccessKeyId=A  
KIAJZZQ2TYSEBKNVWKA'
```

First Previous 1 Next Last

File ID	File State	File Size	Throughput	Remaining	Start Time	Finish Time	Staging Start	Staging End
+ 357604505	FINISHED	0 bytes	0.00 MB/s	-	2016-09-23T14:2	2016-09-23T14:2		

🏠 <https://dpmhead-trunk.cern.ch/dpm/cern.ch/home/dteam/1.txt>

📄 [https://objbkt1.s3.amazonaws.com/tf\\_04?Signature=6qe6joRXpoSFYdAI8Hm9Bjno4%2B8%3D&Expires=1474643908&AWSAccessKeyId=AKIAJZZQ2TYSEBKNVWKA](https://objbkt1.s3.amazonaws.com/tf_04?Signature=6qe6joRXpoSFYdAI8Hm9Bjno4%2B8%3D&Expires=1474643908&AWSAccessKeyId=AKIAJZZQ2TYSEBKNVWKA)

- Transfer host: fts106.cern.ch
- Staging host:
- PID: 9601
- Hash: 20EA
- Activity: default
- Selection strategy: auto
- Attempts: 0
- Duration: 0.379 seconds
- Checksum:
- User specified size: 0
- Configuration:
- Parameters: nostreams:1,timeout:0,bufferize:0
- Job finished: 2016-09-23T14:21:14
- Finished time: 2016-09-23T14:21:14
- Error reason:
- Log file:  
[https://fts106.cern.ch:8449/var/log/fts3/transfers/2016-09-23/dpmhead-trunk.cern.ch\\_objbkt1.s3.amazonaws.com/2016-09-23-1421\\_dpmhead-trunk.cern.ch\\_objbkt1.s3.amazonaws.com\\_357604505\\_fb01b1fe-8198-11e6-8a3f-02163e00a39b](https://fts106.cern.ch:8449/var/log/fts3/transfers/2016-09-23/dpmhead-trunk.cern.ch_objbkt1.s3.amazonaws.com/2016-09-23-1421_dpmhead-trunk.cern.ch_objbkt1.s3.amazonaws.com_357604505_fb01b1fe-8198-11e6-8a3f-02163e00a39b)
- Metadata:  
null



# FTS: key management

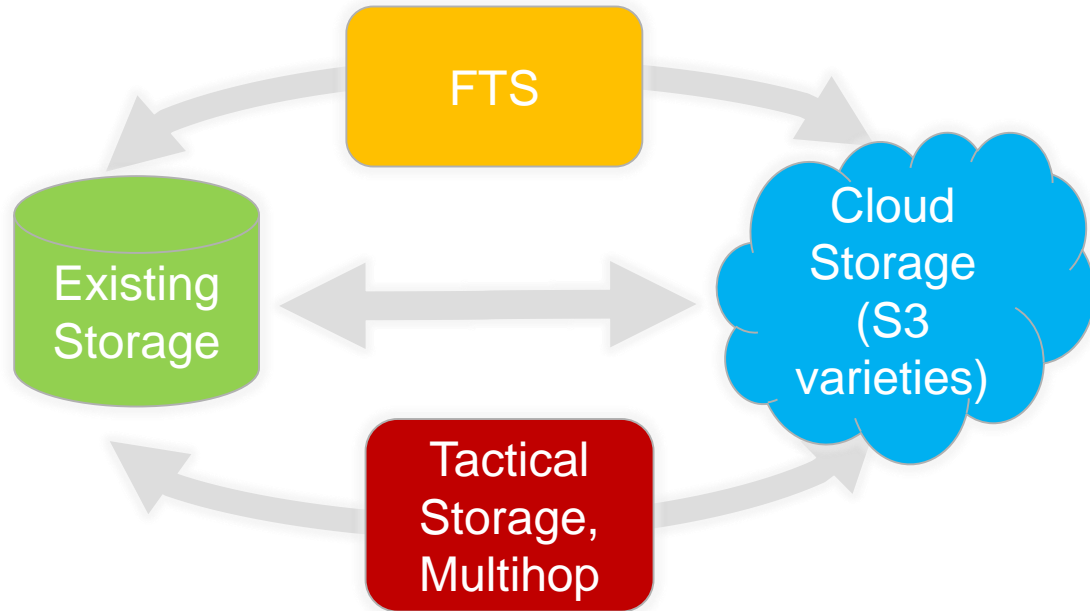
You can also allow FTS to hold the keys to your cloud storage

```
$curl[...] https://fts3devel01.cern.ch:8446/config/cloud_storage -H "Content-Type: application/json" -X POST -d '{"storage_name":"S3:s3.domain.com"}'
```

```
$curl[...] "https://fts3devel01.cern.ch:8446/config/cloud_storage/S3:s3.domain.com" -H "Content-Type: application/json" -X POST -d "${config}"
```

```
{  
  "vo_name": "dteam",  
  "access_key": "ACCESS_KEY",  
  "secret_key": "SECRET_KEY"  
}
```

# FTS: transport



- Solutions for import to and export from clouds
  - Several S3 variants supported
- Various architectures possible
  - FTS gateway
    - SRM<->S3
  - 3<sup>rd</sup> party transfer
  - Multi-hop with tactical storage

# Authentication

# Dynafed: authentication

- The authentication problem
  - Expiry of pre-signed URLs
  - Reuse of such URLs
    - ...there are some in this talk...
- Dynafed can hold your S3 keys and present a “grid standard” X509 (+ HTTP) interface to clients
  - VOMS support

```
davix-put -P grid /etc/services  
https://federation.desy.de/mys3fed/file01
```

# DynaFed: namespace

- DynaFed can also simulate a namespace in front of your S3 storage
  - Including aggregating multiple independent S3 backends

```
$ davix-ls -l dav://federation.desy.de/myfed/azure-s3-cephs3-clouds-together/dir01/
```

```
-rwxr-xr-x 0 426454 2014-09-05 04:04:17 file01  
drwxr-xr-x 0 0 1970-01-01 01:00:00 dir02
```

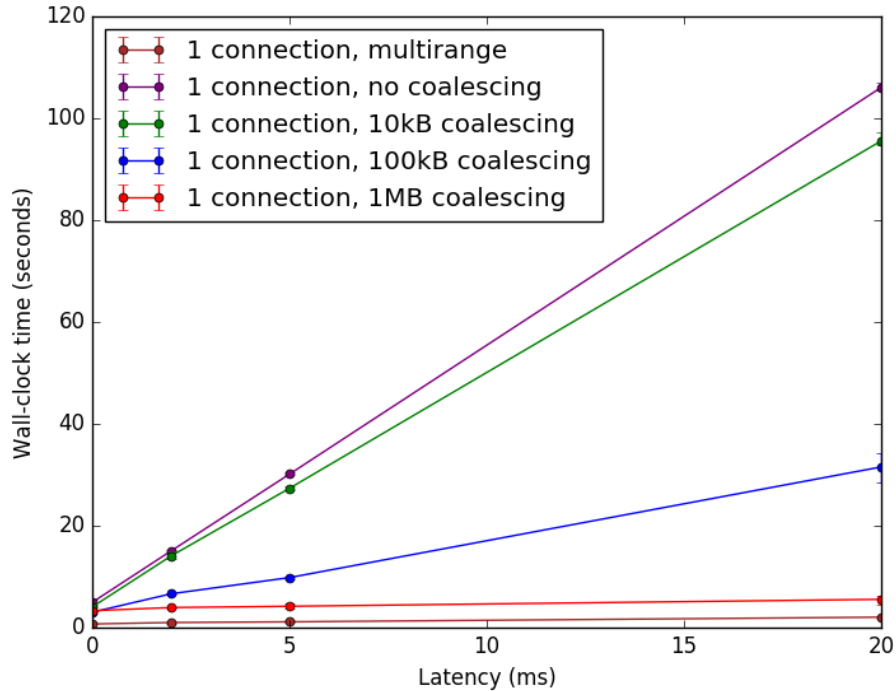
# Access

# Access Performance

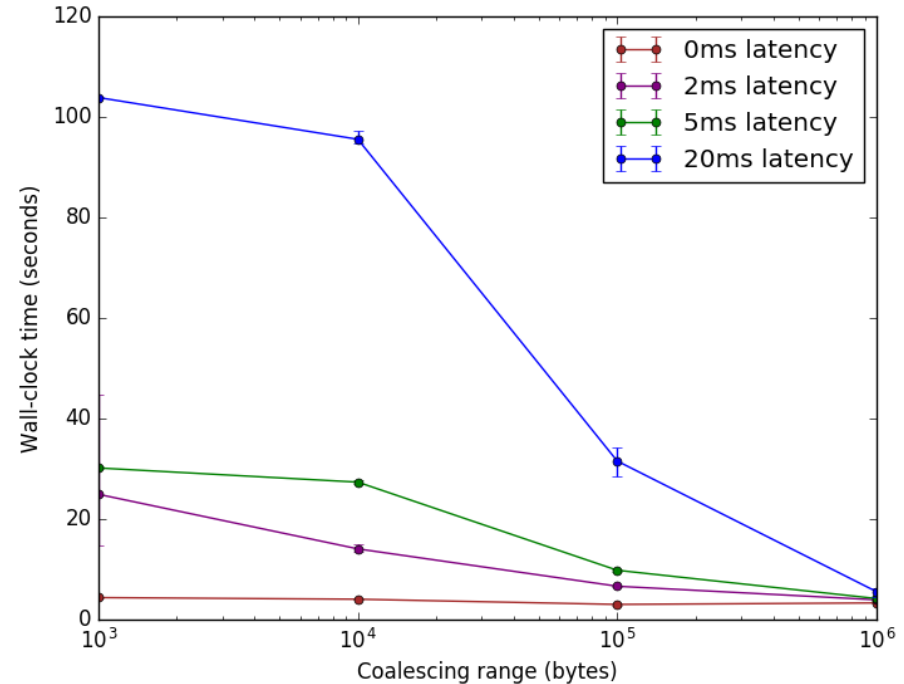
- The stack
  - ROOT + TDavixFile -> Amazon S3, Ceph, Azure, Swift, gcloud (partial support)
- The problem
  - Most S3 implementations do not allow a “vector read”
- The solutions
  - Davix range coalescing
  - Davix concurrency
- Plots
  - ROOT analysis, 267MB file
  - Multi-range behaviour: 5MB read in total, in 30 vectors of length ~180.
  - Tests run 4 times

# Davix range coalescing

Impact of latency on runtime



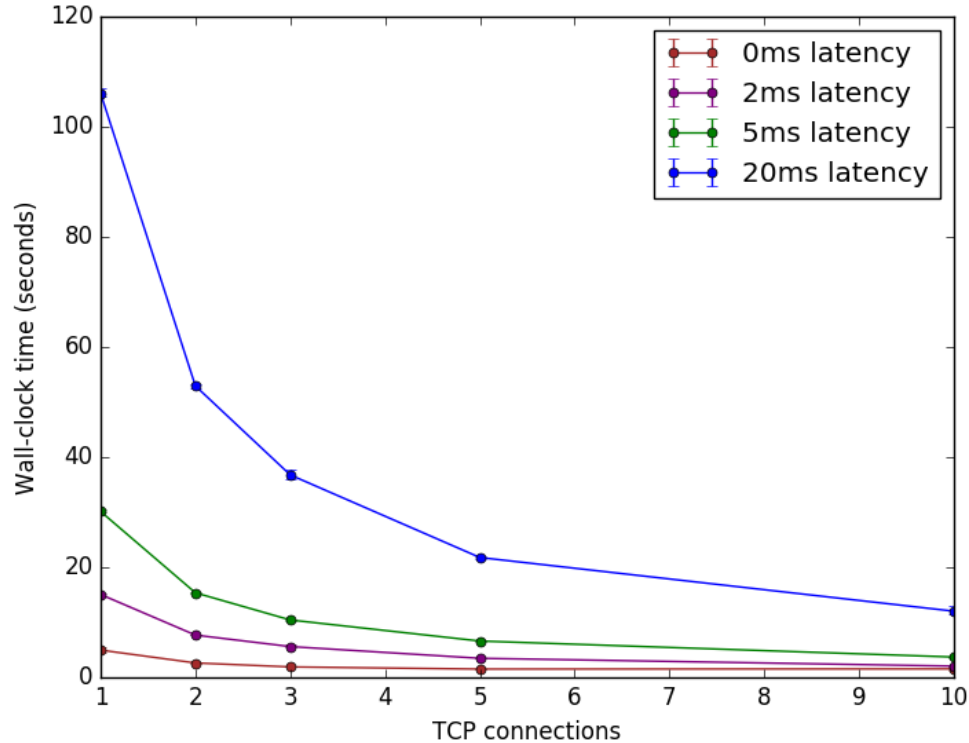
Improving wall-clock time through coalescing





# Davix concurrency

Improving wall-clock time through parallel connections



# Summary

A suite of mature and complementary tools exists which allow exploitation of cloud objects stores with minimal modification of existing clients and frameworks

```
fts-transfer-submit -s https://fts3.cern.ch:8446 https://dpmhead-  
trunk.cern.ch/dpm/cern.ch/home/dteam/file01  
s3://objbkt1.s3.amazonaws.com/file01
```

```
TFile *f=TFile::Open("davs://federation.desy.de/myfed/s3-  
federation/group.test.hc.NTUP_SMWZ.root");
```

# References

- <http://fts3-service.web.cern.ch/>
- <http://lcgdm.web.cern.ch/dpm>
- <http://dmc.web.cern.ch/>
- <https://svnweb.cern.ch/trac/lcgdm/wiki/Dynafeds>
- Try it!
  - <https://fts3.cern.ch:8446>

# Reserve Slides



# FTS: Transport

- 3<sup>rd</sup> party copy
  - DPM & dCache versions
  - Possible Combinations:
    - WebDAV endpoint for most DPM and dCache storages, to S3 (use `davs+3rd:// -> s3://`)
    - WebDAV endpoint for DPM, with version at least 0.17, allows copies from and to S3 (use `davs+3rd:// -> s3://`)
    - `gfal2 >= 2.10`
- Protocol Translation (e.g. SRM <-> S3)
  - Data routed through FTS
    - Demonstrated with CMS Asynchronous Stage Out
  - For limited data, just use `fts3.cern.ch`
    - For better performance, co-locate an FTS in the cloud
      - They are easy to set up!

# 3<sup>rd</sup> party copy with HTTP

```
COPY /srv/dpm/volume/dteam/2015-12-02/tf03.140064.0?token=relztCkNPQb%2FQJg%2FrmXUohGf9r0%3D%401475048508%400&dav_sfn=%2Fdpm%2Fcern.ch%2Fhome%2Fdteam%2Ftf03&dav_user=%2FDC%3Dch%2FDC%3Dcern%2FOU%3DOrganic%20Units%2FOU%3DUsers%2FCN%3Dokeeble%2FCN%3D609355%2FCN%3Doliver%20Keeble&AWSAccessKeyId=AKIAJZZQ2TYSEBKNVWKA&Signature=p8%2Bi28LH7uBBWs2f3MLEC5Sx7U%3D&Expires=1475051109&copyRedirected=1&AWSAccessKeyId=AKIAJZZQ2TYSEBKNVWKA&Signature=FWhqbKsihVhaiIESkeiOC4gFuHk%3D&Expires=1475051110 HTTP/1.1
```

User-Agent: libdavix/0.6.4 neon/0.0.2

Host: dpmdisk-rc.cern.ch

Destination:

http://objbkt1.s3.amazonaws.com/tf05?AWSAccessKeyId=AKIAJZZQ2TYSEBKNVWKA&Signature=bi2oXo5GPtly8m3TFISUk0lnWlg%3D&Expires=1475051109

X-Number-Of-Streams: 1

Secure-Redirection: 1

Copy-Flags: NoHead