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?

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

• How do you deal with the authentication and authorisation system?

• How do you deal with the lack of a namespace?

  • FTS
  • DPM [ & dCache]

  • Davix
  • Gfal2
  • [ROOT]

  • 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>

3rd party copy!!
gfal2/davix

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

3rd party copy!!
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.com/tf_04?Signature=eFAyXMWISY%2BWEVcqfvGvuxZF6ZQ%3D&Expires=2105774242&AWSAccessKeyId=A
KIAJZZQ2TYSEBKNVWKA'
```
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
  - 3rd 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/
-rwrx-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

Wall-clock time (seconds)

TCP connections

- 0ms latency
- 2ms latency
- 5ms latency
- 20ms latency
Summary

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

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

- 3rd 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!
COPY /srv/dpm/volume/dteam/2015-12-02/tf03.140064.0?token=relztCkNPQb%2FQJg%2FrmXUohGf9r0%3D%401475048508%400&display_sfn=%2Fdpm%2Fcern.ch%2Fhome%2Fdteam%2Ftf03&dav_user=%2FDch%2FDc%2Dcouncil%2DOrganic%2DUnits%2FO%3DUsers%2FCN%3Doliver%2Dkeeble&AWSAccessKeyId=AKIAJZZQ2TYSEBKNWKA&Signature=p8%2Bi28LH7uBBBWs23MLEC55x7U%3D&Expires=1475051109&copyRedirected=1&AWSAccessKeyId=AKIAJZZQ2TYSEBKNWKA&Signature=FWHqhKhVhiaeIESkeiOC4gFuHk%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=AKIAJZZQ2TYSEBKNWKA&Signature=bi2oXo5GPtly8m3TFISUk0lnW1g%3D&Expires=1475051109
X-Number-Of-Streams: 1
Secure-Redirection: 1
Copy-Flags: NoHead