dCache:
storage for advanced scientific use-cases and beyond
Tigran Mkrtchyan for dCache team
CHEP 2018, Sofia
Scientific Data challenges
“This Jen, is the INTERNET”

- 04 Oct. 1957 USSR launches “ΠС-1” (Sputnik-1)
- Feb. 1958 creation of ARPA
- Feb. 1966 project ARPANET
- Dec. 1973 TCPv4 (rfc675)
- 1990 – the first web browser
Scientific data challenges

- Volume
- Fast ingest
- Chaotic Access
- Sharing
- Access Control
- Persistence & Long term archival
- Immutability
Storage point of view

- Object store
  - HA, moderate access rate
  - Reduced probability of loss
- Fast data ingest
  - Low latency, high IOPS
- Data Delivery Service
  - Large number of chaotic clients
  - Some data more popular than other
- Time-series data and volatile space
  - Time based data eviction
- Who is who?
  - Authentication and Authorization
Data management & workflow control
(Rucio, Kafka, SSE)

High Speed Data Ingest

Interactive analysis & Sharing

Fast Analysis
NFS 4.1/pNFS

Wide Area Transfers (Globus Online, FTS) by GridFTP, HTTP
dCache design goals

- Single-rooted namespace, distributed data
- Client talks to namespace for metadata operations only
- Bandwidth and performance grow with number of Pool nodes
- Standard clients (OS native or experiment framework)
- Same data can be provided by any access protocol and security flavor
Tertiary storage support

- Native to dCache
  - essential part of original design
- Write-back/read-through -like behavior
  - Transparent for end-users
- Used with a wide variety of HSMs including S3.
- Supports multiple HSMs on the same instance
- Provides full functionality with/without HSM
  - tape and disk-only files can be mixed on the same data server
dCache around the world

- HERA
- Tevatron
- WLCG
- Belle II
- LOFAR
- CTA
- IceCUBE
- EU-XFEL
- Petra3
- DUNE
- And much more ...
Access protocols

- WebDAV
  - username+password
  - x509 certificates
  - SPNEGO
  - Macaroons
- FTP
  - user name+password
  - GSS-API (krb5, gsi)
- NFSv4.1/pNFS
  - RPCSEC_GSS (krb5, krb5i, krb5p)
- DCAP
- XrootD
ALS Project MinE

Searching for the genes that cause ALS (motor neurone disease)

https://www.projectmine.com/

WebDAV
- Authentications
  - X509
  - Username/password
- Can use port 443, bypassing firewall misery
- Redirects (to HTTP)
  - On: load balancing, but unencrypted
  - Off: TLS data encryption
- webdav.grid.surfsara.nl DNS round robin

Shameless stolen from Onno Zweers presentation.
Data life cycle

Beamline

Detector

PC-Layer

Online Storage

Scratch

Online Cluster

Long term storage

Offline Storage

ScratchHome

Offline Cluster

PC-Layer

Online Storage

Scratch

Online Cluster

Offline Cluster

storage for advanced scientific use-cases | Tigran Mkrtchyan | 12
dCache for Photon Science

- Storage for Off-line data processing
- Ex/Import to/from remote site
- Preserves Access control preservation on archival
- Data workflow control
Workflow control

See Paul's presentation on Thursday afternoon
dCache for EU-XFEL LINAC DAQ

- Fast ingest of accelerator telemetry data
  - 500 MB/s, 24x7
- Stored for 30 days
- Automatically removed
  - old data removed including namespace entry
dCache for WLCG

- Used by Tier-1 and Tier-2
  - ~50% of LHC data
- WAN access
  - HTTP
  - GridFTP
- LAN Access
  - DCAP
  - NFSv4.1/pNFS
  - Xrootd
  - HTTP (davix)
- Management
  - SRM
  - CDMI
  - REST-API
Data sharing challenges

- Authentication
- Authorization
- Access control
- Delegation
Multiple identities problem

- x509 (grid)
  
  /C=DE/O=GermanGrid/OU=DESY/CN=Tigran Mkrtchyan

- Kerberos
  
  tigran@DESY.DE

- LDAP
  
  - uid=tigran,ou=people,ou=rgy,o=desy,c=de

- Unix ID (uid)
  
  - 3750
plugable authn

• Pam -like system
• Allows to combine multiple plugins
  • specify plugin wiring
• Supports many standard and custom authentication plugins
  • from ActiveDirectory to gridmap file
Example config

```bash
# authenticate with username+password, or certificate
auth  sufficient ldap
auth  optional x509
auth  optional voms

# get uid, gids from ldap
map  optional vorolemap
map  sufficient ldap
map  sufficient authzdb

# get home directory from ldap
session sufficient ldap
session optional authzdb
```

If user comes with password
Or x509 certificate and VOMS

If there is a mapping for DN+VOMS to “user name”
Take it into account

Try local auth-db file
Macaroons: the other cookies

- Contextual Caveats
- Decentralized Authorization
Macaroons: 101

- HTTP Bare token
- No special knowledge on the client side
- Derive new macaroon by adding a caveat
- HMAC based chain of caveats
- All caveats must be fulfilled to authorize request

*Can be validated by issuer of initial macaroon.*
Example:

Storage System

Data portal

storage for advanced scientific use-cases | Tigran Mkrtchyan | 23
Example:

```
Storage System

Data portal

{READ, WRITE}

{+path:/data}

{+read-only}

{+path:/data/img.jpg,ip:1.2.3.4}

Storage System
```

storage for advanced scientific use-cases | Tigran Mkrtchyan | 24
Non scientific data

- Storage back-end for nextCloud
- Exposed as NFS server
- High metadata rate, low IO requirements
- High availability

nextCloud instance @ DESY
HA-nextCloud instance @ DESY
Summary

- dCache stores and delivers data for many (scientific) communities.
- Provides uniform authN and authZ independent from access protocol.
- Let experiments to manage data, not storage.
Thank You!