EGI Data Federations

Lukasz Dutka (CYFRONET),
Tiziana Ferrari (EGI.eu),
Christian Bernardt (DESY),
Fabrizio Furano, Alejandro Alvarez (CERN),
Helmut Heller (LRZ)

EGI VT-FedData
Use cases and data flow

- UC1 Brain Scan Registration and Delivery
- UC2 Remote interactive multiresolution visualization of large volumetric datasets
- UC3: Feature extraction and analysis of large volumetric datasets
- UC4: Publication and citation of data
- UC5: Management of Access Control Rights
Federating e-infrastructures and data

- Distributed, federated storage, HTC and cloud facilities
- Virtual Research Environments
- > 200 registered user research projects
- ~38,000 researchers

- ~350 resource centres in 52 countries
- 500,000 logical CPU cores
- >290 PB disk, 180 PB tape
- > 99.6% reliability
- Computational jobs ~1.5 Million per day
- Virtual machines deployed ~2,2000 per day
Do you need operational services and tools to run a distributed IT infrastructure for research?

**Technologies, processes and expertise:**
- To manage operations of heterogeneous distributed infrastructures
- To integrate resources from multiple independent providers with lightweight central coordination

**Added Value**
- A cost-efficient framework to manage operations within a federated environment, while retaining responsibility of local infrastructure.
- Allows efficient implementation of Best Management Practices for IT services.

**Relies on the best practices for IT service management (fitSM)**
- **Operations, Technology, and Security Coordination**
- Helpdesk support
- Specialised consultancy services
- Technical consultancy and support
- Operations tools
Federated Data Management Solutions

- dCache
- DynaFed
- iRODS
- Onedata
- FTS
- Globus Transfer
Testbed

CESGA – Spain
CERN – Switzerland
CNRS – France
CSIC-IFCA – Spain
CYFRONET – Poland
DESY – Germany
GWDG – Germany
GRNET – Greece
INFN-Bari – Italy

Others soon
Integrations with other solutions: EGI-EUDAT Collaborations

• **Goal**
  – Enable researchers seamlessly access e-Infrastructures services pairing data and high-throughput computing resources together
  – Permit a user of both e-infrastructures to instantiate a VM on the EGI Cloud Federation for the execution of a computational job consuming data stored onto EUDAT resources

• **Communities**
  – ELIXIR, EISCAT-3D, BBMRI, ICOS, ENVRI+
Dynafed FTS3

Fabrizio Furano
Alejandro Alvarez
• FTS stands for “File Transfer Service”, a workhorse of LHC data movements
• Used by ATLAS, CMS and LHCb. CERN production instance transferred ~10PB, just one instance, just one month (June ’15)
• FTS3 is designed to "fill up" and efficiently use its infrastructure
• Protection from overload of network and endpoints
  – auto-tune transfers and direct/manual configuration of limits
• Smart transfer retry mechanism based on error classification

• Modular protocol support
  – gsiftp, http(s), xrootd and srm
  – Cloud extensions: dropbox, CERNBox
• JSON-based config, REST-style interface for xfer submission and status retrieval
  – Standard clients and/or libraries can be used to perform these tasks

• We used them to design WebFTS, a production-grade Web based tool to transfer files between grid/cloud storages
• https://webfts.cern.ch

• Initial development funded by
### WebFTS (Beta version)

#### Job Table

<table>
<thead>
<tr>
<th>File ID</th>
<th>Transfer Host</th>
<th>Source URL</th>
<th>Dest. URL</th>
<th>File Size (Bytes)</th>
<th>Throughput (MB/s)</th>
<th>Start Time</th>
<th>End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>27546292-3362-11e4-9809-02163e008cfa</td>
<td>gisftp://lxfsra10a01.cern.ch</td>
<td>gisftp://lxfsra10a01.cern.ch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0245cd88-3362-11e4-ad7b-02163e008cfa</td>
<td>gisftp://lxfsra10a01.cern.ch</td>
<td>dropbox://www.dropbox.com</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Submit Time

- a31397f6-350e-11e4-9dc8-02163e008cfa: 2014-09-05T15:09:30
- 27546292-3362-11e4-9809-02163e008cfa: 2014-09-03T12:02:18
- 0245cd88-3362-11e4-ad7b-02163e008cfa: 2014-09-03T12:01:16

#### No delegation detected
Dynamic Federations (Dynafed)

- An interactively browsable system able to dynamically discover its metadata content and transparently present it to the clients
  - **Makes a bigger system out of distributed ones.**
- Supports HTTP(s), WebDAV and S3, replicas, hierarchical listings and writing into the federation

- Makes it possible to browse and access a huge repository made of many sites without requiring a static index and **keeping sites’ independence**
  - No “registration”, no maintenance of additional file catalogues.
  - If catalogues are needed, can talk to more than one at the same time or just coexist

- Redirect intelligently clients asking for replicas
  - **Automatically detect and avoid sites that go offline or unreachable**
- Efficient with algorithmic name translations
  - Can also accommodate non algorithmic ones
- Accommodate client-geography-based redirection choices

- Dynamic partial namespace caching: fast and scalable
  - Top constant speed is in the order of 7-10K client transactions per sec per frontend machine
On the fly friendly visualization
Full WebDAV access
Redirection-based
Robust against failures
Fully scalable

With 2 replicas

Scalability, speed
Seamless presentation

Geography-based
Client-aware redirections

Flexible authentication/authorization, friendly with identity federations

Realtime detection of sites' up-ness

Makes S3 storage easy to use, scales it up and applies uniform security

Site A
(HTTP/WebDAV/S3)

.../dir1/file1
.../dir1/file2

Site B
(HTTP/WebDAV/S3)

.../dir1/file2
.../dir1/file3

Initial development funded by...
**HBP use cases - For more info**

- A **modular** set of road-tested and up-to-date solutions, demonstrated to be **integratable** across the LHC Grid over many years. These fulfill STORAGE – FILE XFER – DATA ACCESS – DATA FEDERATION

- **Everything is open source software** (licenses vary from Apache2 to LGPL)

- Focus on open-ness and flexibility with open data protocols, seamless integration with mainstream tools and mainstream Cloud storage techs
- Appropriate for integrating with investments in existing software solutions (e.g. custom metadata indexes)

**We see an easy integration path for HBP use cases, for moving and giving straightforward access to data**

- FTS and dCache have a long history of working well together, also the teams
- DynaFed is a powerful, lightweight scalable component that comes from this cooperation and from the experience in wide production scenarios

- Source code: [https://svnweb.cern.ch/trac/lcgdm/wiki/Dynafeds](https://svnweb.cern.ch/trac/lcgdm/wiki/Dynafeds)

- Source code: [https://svnweb.cern.ch/trac/fts3](https://svnweb.cern.ch/trac/fts3)
- Web FTS homepage: [https://webfts.cern.ch/](https://webfts.cern.ch/)
dCache cheat sheet

- dCache.org is an international collaboration, developing and distributing storage software (DESY, Fermilab, NEIC)

- dCache is in production in about 60 places around the world and stores (roughly) about 120 Pbytes in total for WLCG (about 50% of the total LHC data).

- dCache supports different storage media, like disk, SSD and tape and provides mechanisms for manual and automated internal and external replication and transitions.

- dCache storage can be accessed via standard protocols like WebDAV, NFS (POSIX), and gridFTP and proprietary protocols like dCap and xrootd.

- dCache supports ACLs with translation to unix permissions

- dCache supports a variety of authentication and mapping mechanisms, e.g. Kerberos, X509, User/Password, LDAP, NIS, NSSWITCH.

- Open source software
dCache - long-term funded storage expertise

• 15+ years of development deployment experience for laboratories around the globe

• 15+ years of funding

• Supporting multiple communities with diverse requirements
  • LHC
  • Fermilab Intensity Frontier
  • Belle II
  • IceCube
  • Photon Science

• Operational Experience at DESY, Fermilab, NDGF - tight loop between development and operations (quick propagation of requirements)

• Current development focus on data lifecycle policy with regard to Data Management Plans (DPM required for EU project proposals)
Possible Solution to HBP use cases

- Data Storage - disk short term, tape long term
- Mountable POSIX access: Parallel NFS + ACLs
- Currently developing (within Indigo-DataCloud)
  - Data lifecycle management
  - How securely a file is stored? (disk once, disk replicated, disk/tape, tape only, replicated across sites)
  - How quickly a file can be accessed? (SSD, Spinning disk, Tape)
  - Policy driven access rights

DynaFed

Common Namespace with replica detection, metadata in cooperation with the Metadata Center and FTS

Scalable network aware Transfer, Replication (with wrapper)
European Globus Community Forum
Globus Transfer

Dr Helmut Heller
Globus Data Transfer

• Based on GridFTP
  – works also with dCache

• Enhanced by a portal:
  – Globus Online => new name **Globus Transfer**
  – SaaS running in Amazon Cloud
  – using third party data transfer of GridFTP
  – [www.globusonline.eu](http://www.globusonline.eu)

• Basic transfer is and will always be free

• Premium services cost money
The Last Mile

• **Globus Connect Personal**
  - a little app for your Mac, Linux, Windows PC
  - basically an encapsulated GridFTP server
  - easy data transfer to your device
  - works also behind most firewalls
  - works with DHCP and on the road
  - single user

• **Globus Connect Server**
  - encapsulated GridFTP server
  - multi-user
Deployment

Globus Connect Personal

Globus Transfer
Globus Sharing
Globus Publish

LDAP

MyProxy CA

GridFTP

Brain Research Facility

Data Center

Supercomputing Center

GridFTP

GridFTP

GridFTP
Onedata

Lukasz Dutka
Cyfronet AGH
Onedata Spaces

Metadata change feed

Onedata Space Registry keeps control on
ACL to Data Spaces based on OpenID

User 1

User 2

User 3

Group 1

Each Data Space can be accessed (RW) by
multiple users according to ACLs

Data Space 1
is a kind of Virtual Directory

Data Space 2
is a kind of Virtual Directory
Onedata for HBP Usecases

- Fully Supported all use cases UC1 – UC5
- Solution for UC4 under development as a part of EGI-Engage Project
Onedata in a nutshell

• One solution for:
  – **Unified** data access
  – **Distributed** data management
  – High **throughout** data **migration** between sites
  – High speed **remote** data **access** on the fly
  – **Sharing** data between users
  – Flexible authorization and access control based on: **Tokens**, OpenID, Certs and others
  – **Open data** management
iRODS
Pre-RDA in Paris

Jerome Pansanel
and
the FG-iRODS team
• iRODS is an open source middleware designed for:
  – Managing large scale of distributed data collection
  – Data virtualisation
  – Automation of data operations
  – Data management policy enforcements
  – Data preservation and sharing
  – Being extensible (plugins)

• iRODS is based on a rule engine, that applies policies and rules and integrates a descriptive metadata system. iRODS comes with a basic set of rules that can be extended by user-defined rules and micro-services.

• The central component of an iRODS infrastructure is the iCAT-enabled resource. It is used for user authentication, authorization decisions and metadata storage. A group of iRODS servers with an iCAT-enabled resource server is called an iRODS zone. iRODS zones can be federated.

• iRODS can be used with many storage systems (Unix file system, HPSS, S3, HDS, Hadoop, FTP, HTTP, ...)

9/26/2015
iRODS use cases

• iRODS and several tools related to iRODS are released as open source software.
• It is used by world-renowned projects or institutes, like EUDAT, LSST, Wellcome Trust Sanger Institute or Novartis.
• Professional support (maintenance contracts, on-site support, ...) can be purchase from the iRODS consortium.
• iRODS homepage: http://irods.org
  • Documentation: http://irods.org/documentation
  • iRODS consortium members: http://irods.org/consortium/members
  • Source code: https://github.com/irods/irods
Metrics of Success

- simplicity of bringing up another active repository
- simplicity of the process transferring data to the active repository site (UC1)
- flexibility of accessing the same data by multiple scientists including intra groups access (UC3)
- decentralization of resource management. There are many collaborating groups but they remain independent. The system should be flexible to allow independently gain resources by those groups and still provide some integration level.
- simplicity of the access control (UC5)
Metrics of Success

• scalability of processing through load distribution/brokering, it should be easy (maybe automatic), instantiation of new processing units based on the traffic to active repositories. (UC2)

• how difficult distribution/upgrading of the application software onto the environment will be. Docker approach is under HBP investigation now.

• support for persistent identification (DOI and persistence life cycle approaches) (UC4)

• process of migration data between active repositories if needed, that might be limited
Thank you for your attention.

Questions?