ONEDATA PLATFORM FOR TRANSPARENT AND UNIFIED DATA ACCESS IN HYBRID CLOUDS

Presented by: Bartosz Kryza
DATA IN HYBRID CLOUD ENVIRONMENTS
ARCHITECTURE OVERVIEW
ONEDATA ARCHITECTURE

Onezone

External IdP

External IdP

ONE WORLD

FUSE Client

Oneclient

HTTP GUI

REST / CDMI

P2P

Data Space 1

Oneprovider

DS. 3

Data Space 2
MAIN FEATURES
PROBLEM: GLOBAL NAMESPACE AND TRANSPARENT ACCESS TO DATA IN MULTI-CLOUD ENVIRONMENTS

Thanks to Onedata now you can:

• Transparently access existing and create new data in multi-cloud environments
• Access data from anywhere
• Use many protocols to access the same data
• Use global namespace for globally distributed data collections
BUT WE WANT POSIX

- Support for most of the POSIX operations on virtual file system
- Global namespace automatically mapped to a virtual file system mountable on VM, Grid nodes, Docker containers, desktops
[...] WITH THE SAME PERMISSIONS

- LUMA - flexible mapping of POSIX access rights among different storages
- Extensible – easy to develop integration with custom AAI systems (e.g. LDAP)
PROBLEM: HOW TO ACCESS LOCKED-IN DATA

Thanks to Onedata now you can:

• Current data collections stored on locally available POSIX file system can be easily exposed
• Access data in public Cloud on-the-fly, without pre-staging, with smart caching
• Remotely created data can be replicated back to the private Cloud local storage automatically
SYNCHRONIZATION OF PRIVATE STORAGE
PROBLEM: HIGH THROUGHPUT DATA TRANSFERS

Thanks to Onedata now you can:

• Replicate files on demand and on the fly without any additional effort
• Migrate data between sites on demand with simple API interface
• Easily monitor transfers and current data distribution through Web GUI or REST API
REPLICA MANAGEMENT SIMPLIFIED

- Manage files not replicas
- Files distribution between locations is hidden below the virtual filesystem structure
- Replica management on a block basis
- Missing blocks delivered on the fly
- API for replica management for pre-staging and implementing external data policy management
HIGH THROUGHPUT DATA MIGRATION IN P2P MESH

- Data Transfer Mesh
- 3 Oneproviders connected by 20+Gbit/s links
- Simultaneous data transfers between all of them
- Single VM Node per Provider

Data Migration at combined Throughput 56 Gbit/s
Thanks to Onedata now you can:

• You can work with different type of storages: Ceph, S3, POSIX, SWIFT, GlusterFS, WebDAV (new)
• DirectIO architecture allows for high throughput data processing from all machines running Oneclient with only single Oneprovider deployment
MULTIPLE STORAGE BACKENDS

Parallel Processing Nodes using POSIX Oneclient, CDMI or REST

Protocols CDMI
Protocols S3
Protocols POSIX VFS

Control, Remote Data Access CDMI API

Direct Access if possible

Storage Access

Ceph S3 SWIFT NFS GlusterFS

Oneprovider
PROBLEM: DATA DISCOVERY AND REPLICATION BASED ON DISCOVERY RESULTS

Thanks to Onedata now you can:

• Work with data and metadata in one system – avoiding problems of consistency
• Monitor metadata data changes through API in order to integrate with other services
• You can discover data within data collections using highly efficient indices built on top of Map-Reduce
• New interface for originating replicas based on data discovery results
INTEGRATED METADATA MANAGEMENT

- All files and directories can have custom metadata
- API for metadata management
- API for data discovery based on metadata
- Virtual Folders based on metadata tags
Thanks to Onedata now you can:

• Control and Monitor local storage usage
• Monitor local file popularity
• Automatically evict locally least popular replicated data
• Expose interface for external logic to cache eviction
ONEDATA FILE POPULARITY AND SMART CACHING
ONEDATA IN HNSC
Onedata has been tested thoroughly using applications from:

- **EMBL**
  - Large genome files processed in public Cloud on the fly from storage on private Cloud
- **INFN**
  - Large number of files processed in a sequence of random read operations with very small block size (~1KB) in public Cloud on the fly from storage on private Cloud with automatic prefetch
- **DESY**
  - Complex deployment including LUMA for UID/GID mapping from private Cloud storage to public Cloud and automatic replication of results back to the private Cloud storage with correct access rights
ONEDATA HIGH THROUGHPUT DATA PROCESSING ON HNSC

Onedata Transparent POSIX File System
Processing transparently cached data - 37GBytes/sec
LESSONS LEARNED

• Implementing transparent hybrid cloud virtual filesystem with pure POSIX interface for real life applications is challenging – but possible!

• Seemingly simple issues such as authentication or uid/gid mapping can be in fact very challenging and time consuming to get right

• High-performance backend storage and low latency network are critical to achieve high IOPS

• Performance issues are hard to track instantly in hybrid Cloud deployments, as they can depend on multiple factors (network issues, storage issues, high system loads from other users, etc.)
• Spin-off company is being established, which will be in charge of evolution of commercial features of the project
• Several support plans will be provided with different reaction times and support levels
• Pricing will be based on the number of Oneprovider instances deployed and total storage size managed by them
• Tentative pricing, flat rate per 1PB/Oneprovider/month:
  • Silver – 48h reaction time
  • Gold – 24h reaction time
  • Platinum – 7h reaction time
We would like to thank everyone in the HNSC for this opportunity in particular:

- All Buyers Group members who have participated in Onedata evaluation for their patience and vigilance in pointing out missing features and bugs
- Both T-Systems and RHEA consortia for trusting us and helping us with evaluation on production Cloud infrastructures
QUESTIONS?

Please visit:
www.onedata.org
BACKUP
EXAMPLE - EMBL-EBI APPLICATION DEPLOYMENT

Cloud - OTC

300 VMs

Local processing cluster 300 VMs
Each VM process 4 files at the same time
Connected to POSIX like, Globally Shared Virtual File System via Oneclient

High Troughput ~10GB/s
Low Latency Access ~500µs

Local Cache
Ceph on 3 VMs
Capacity 20TB

Private EMBL Cloud - Hinxton

NFS Test Collection 26 TB
Full Collection size 1.5 PB

1 VM
Oneprovider

10G Link

1 VM
Oneprovider

10G Link

GEANT

LOCAL processing cluster 300 VMs
Each VM process 4 files at the same time
Connected to POSIX like, Globally Shared Virtual File System via Oneclient

High Troughput ~10GB/s
Low Latency Access ~500µs

Local Cache
Ceph on 3 VMs
Capacity 20TB

Private EMBL Cloud - Hinxton

NFS Test Collection 26 TB
Full Collection size 1.5 PB

1 VM
Oneprovider
## CDMI HTTP ACCESS

<table>
<thead>
<tr>
<th>Operations</th>
<th>Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic object GET PUT DELETE</td>
<td>cdmi_dataobjects, cdmi_read_value, cdmi_modify_value, cdmi_delete_dataobject</td>
</tr>
<tr>
<td>Basic container GET PUT DELETE</td>
<td>cdmi_list_children, cdmi_create_container, cdmi_delete_container</td>
</tr>
<tr>
<td>Metadata (container&amp;dataobject)</td>
<td>cdmi_read_metadata, cdmi_modify_metadata, cdmi_size, cdmi_(atime</td>
</tr>
<tr>
<td>Access control lists (rwx)</td>
<td>cdmi_acl</td>
</tr>
<tr>
<td>Big folders</td>
<td>cdmi_list_children_range</td>
</tr>
<tr>
<td>File System Export (FUSE client)</td>
<td>-</td>
</tr>
<tr>
<td>Move and copy</td>
<td>cdmi_(move</td>
</tr>
<tr>
<td>Big files</td>
<td>cdmi_read_value_range, cdmi_modify_value_range</td>
</tr>
<tr>
<td>Access by ObjectID</td>
<td>cdmi_object_access_by_ID</td>
</tr>
</tbody>
</table>
REPLICATION JOBS AND MONITORING
Thanks to Onedata now you can:

- You can hook your customized decision logic into the system
- Onedata offers a long polling HTTP API for monitoring changes in the system
- The API is eventually consistent
- The customized logic can be used for instance for:
  - automatic replication to certain locations
  - metadata ingestion