Quality of Service in Storage by
INDIGO-DataCloud

Anupam Ashish
Where did it come from?

- Amazon
  - S3 : online
  - Glacier : nearline

- Google
  - Standard
  - Durable Reduces Availability (DRA)
  - Nearline

- IBM (HPSS, GPFS)
  - Storage classes (user defined)

- dCache
  - Tape
  - Disk (spinning or SSD)
  - Resilient Management (‘n’ copies)
First Ideas

Platform as a Service

Canonical Storage Property Information System

IaaS

StoRM

dCache

GUI

REST API

Canonical Storage

Property Information

System

3
The INDIGO – DataCloud Approach

- Built a common (agreed) vocabulary for Data Management
- Map agreed vocabulary to a QoS protocol spec
- Provide a reference Implementation
CDMI Cloud Data Management Interface

- SNIA Cloud Data Management Interface (CDMI)
  - ISO/IEC standard
  - Interoperability for Data stored in the cloud
    - Cloud solution vendors

- More than 20 products that meet the CDMI
Storage Networking Industry Association

• Non-Profit Organization
  • Industrial and Scientific members from information technology

• Mission
  • Promoting vendor-neutral architectures
  • Standards and Educational Services

• Facilitate
  • Efficient Management of Data
  • Movement of Data
  • Security of Data

May 2017
Anupam Ashish – 11th International dCache Workshop
CDMI v1.1.1

- Discover capabilities available with storage provider
- Manage containers and the data that is placed in them
- RESTful principles in the interface design
- Providers can support a subset of CDMI
  - Must expose the limitations in the capabilities reported
CDMI v1.1.1

- **Object Model**
  - Data Objects → Files
    - Store Values
  - Containers → Directories
- **Capability Object**
  - Functionality provided by the storage system
  - Each data object or container has one or more capability objects
    - `cdmi_retention_period`

- **QoS Class**
  - Collection of capability objects
• Some Quality of Storage Attributes available
  • Data Redundancy cdmi_data_redundancy
  • Geo-Location cdmi_geographic_placement
  • Latency cdmi_latency
  • Retention cdmi_retention_period

• Data System can express an actual value to the exposed attributes
  • Geo-Location DE
  • Latency 3000

• User can request a specific value during data creation or update
  • Geo-Location FR, DE, JP
CDMI Considerations

• CDMI is an industry standard.
  • Allows dCache to express its QoS for other systems

• Not widely Used

• Doesn’t cover our use cases
  • Transition from one QoS to other
    • No clear way to determine if
      • the transition is allowed or
      • the system is capable of providing time

• Lifetime of QoS
  • Pinning in dCache
  • Lock QoS on a data object
CDMI Considerations

• Pros
  • CDMI is an industry standard.
  • Allows dCache to express its QoS for other systems

• Cons
  • Not widely Used
  • Transition from one QoS to other
    • No clear way to determine if
      • the transition is allowed or
      • the system is capable of providing time
  • Can’t express Lifetime of QoS
    • Pinning in dCache
    • Lock QoS on a data object
  • No way to manipulate the set of capabilities
INDIGO CDMI Extension

• Default attributes based on storage system configuration
• Assign QoS Class on creation and update of objects
• Changes to QoS restricted
  • Discover allowed transitions from one QoS Class to another
  • To the ones permitted by the storage system
  • Changes to individual capability of a QoS class not allowed
• Attribute for Capability Lifetime
• Additional attribute for monitoring QoS transitions
INDIGO QoS Architecture

Generic CDMI Web Service

Java Service Provider Interface (SPI)

HPSS Plug-in  CEPH/StoRM  dCache Plug-in

REST <-> HPSS API

High Performance Storage System

dCache QoS Controller

SRM
Current Deployment

- KIT (master server)
- KIT (GPFS, HPSS: Tape, Disk)
- CNAF (StoRM: disk for now)
- DESY (dCache: Tape, Disk)
- Poznan (CEPH: disk only)
## Infrastructure View

![Infrastructure View](image)

<table>
<thead>
<tr>
<th>Directory</th>
<th>Date/Time</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private</td>
<td>4/5/2017, 6:01:12 AM</td>
<td></td>
</tr>
<tr>
<td>public-file</td>
<td>4/5/2017, 6:05:00 AM</td>
<td>177 Bytes</td>
</tr>
<tr>
<td>private-file</td>
<td>4/5/2017, 6:06:01 AM</td>
<td>148 Bytes</td>
</tr>
</tbody>
</table>
## Available Qualities of Storage

<table>
<thead>
<tr>
<th>Name</th>
<th>Access Latency [ms]</th>
<th>Number of Copies</th>
<th>Storage Lifetime</th>
<th>Location</th>
<th>Storage type</th>
<th>Available Transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>DataobjectProfile1</td>
<td>3000</td>
<td>1</td>
<td>20 years</td>
<td>PL</td>
<td>Archival</td>
<td></td>
</tr>
<tr>
<td>DataobjectProfile2</td>
<td>2000</td>
<td>2</td>
<td>20 years</td>
<td>PL, UK</td>
<td>Archival</td>
<td></td>
</tr>
<tr>
<td>DataobjectProfile3</td>
<td>500</td>
<td>3</td>
<td>20 years</td>
<td>NL, ES, PL</td>
<td>Archival</td>
<td></td>
</tr>
<tr>
<td>disk</td>
<td>100</td>
<td>1</td>
<td></td>
<td>DE</td>
<td>Process</td>
<td></td>
</tr>
<tr>
<td>disk+tape</td>
<td>100</td>
<td>2</td>
<td></td>
<td>DE</td>
<td>Process</td>
<td></td>
</tr>
<tr>
<td>DiskAndTape</td>
<td>50</td>
<td>3</td>
<td>20 years</td>
<td>DE</td>
<td>Process</td>
<td></td>
</tr>
<tr>
<td>DiskAndTape</td>
<td>50</td>
<td>2</td>
<td></td>
<td>IT</td>
<td>Process</td>
<td></td>
</tr>
<tr>
<td>DiskOnly</td>
<td>50</td>
<td>3</td>
<td>20 years</td>
<td>DE</td>
<td>Process</td>
<td></td>
</tr>
<tr>
<td>DiskOnly</td>
<td>50</td>
<td>1</td>
<td></td>
<td>IT</td>
<td>Process</td>
<td></td>
</tr>
</tbody>
</table>

**Collected Capabilities**

- Access Latency
- Number of Copies
- Storage Lifetime
- Location
- Available Transition
Back to Infrastructure View

KIT / DiskOnly

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data storage lifetime</td>
<td>20 years</td>
</tr>
<tr>
<td>Latency</td>
<td>10</td>
</tr>
<tr>
<td>Throughput</td>
<td>4194304</td>
</tr>
<tr>
<td>Capability lifetime action</td>
<td>migrate-to:/cdmi_capabilities/dataobject/DiskAndTape</td>
</tr>
<tr>
<td>Capabilities allowed</td>
<td></td>
</tr>
<tr>
<td>Capability lifetime</td>
<td>0:30:00</td>
</tr>
<tr>
<td>Geographic placement</td>
<td>• DE</td>
</tr>
<tr>
<td>Data redundancy</td>
<td>3</td>
</tr>
</tbody>
</table>
The End

https://www.indigo-datacloud.eu
Better Software for Better Science.
The End

https://www.indigo-datacloud.eu
Better Software for Better Science.
Conclusion

• Apologies for not doing it the “right way”
• But we had to provide an implementation within 30 months.
• However, we can prove that we are serious.
• Process with SNIA is painful but helps to understand the difficulties, to map our ideas to a real protocol.
• Implementing the protocol helps to understand the issues with the different storage systems.
• We even now support limited transitions.
  • Dangerous !!! (Tape is a tricky beast 😊 )
• Brokering IG and Brokering WG

• Vocabulary groups

• NEW: data preservation:

• Overlap with data management plans.