OSD-Based pNFS

Benny Halevy bhalevy@panasas.com
Boaz Harrosh bharrosh@panasas.com

2008 Linux Storage & Filesystem Workshop
San Jose, CA
February 26, 2008
The Big Picture

pNFS client

<table>
<thead>
<tr>
<th>Layout cache</th>
<th>Device Cache</th>
</tr>
</thead>
</table>

Objects Layout Driver

I/O

Device

OSD Initiator Library

iSCSI / iSER / FC

Files/Blocks Layout Drivers

iSCSI Discovery Daemon
Status

• Panasas layout driver
  git://linux-nfs.org/~bhalevy/linux-pnfs.git

• OSD kernel patches
  git://git.open-osd.org/open-osd.git

• OSD Library
  git://git.open-osd.org/osd-lib.git
Requirements

- Bidirectional SCSI CDBs (in 2.6.25)
- Variable Length CDBs (2.6.26?)
- Long Sense Buf (patches sent for review)
- iSCSI Discovery Daemon
osd-lib design

- osd_initiator.c: low level request helpers
- osd_sec.c: security manager helpers
- osd_attributes.c: definitions for standard attrs

- Wish list: osd_target ...
OSD Command Overview

- CDB
  - Get/Set Attrs
  - Capability
  - Signature

- Data out
  - Set Attrs
  - Get Attrs
  - Signature

- Data in
  - Get Attrs
  - Signature
Performing an OSD command

- **osd_start_request**: allocate a request
- **osd_req_{format,create_partition,...}**: encode CDB
- **osd_add_{get,set}_attr**: add get/set attrs to CDB
  Both list or page modes supported
- **osd_finalize_request**: set final pointers and sign the request
- **osd_execute_request**: execute request via block layer
- **osd_req_decode_sense**
- **osd_req_decode_get_attr{,_list}**
- **osd_end_request**: destroys the request
Discussion

• DM
  - useful for layout driver stripe I/O?
    • Per-file layout, nested layouts.
    • Distributed error handling

• ULD
  - Need to register OSD type (possible via sg)
  - needed for representing objects as block devices?
    • Dynamic Object ID namespace => osdfs

• Open pnfs-obj back-end filesystem
  - Reference implementation
    • metadata and namespace on MDS local file system