Lustre 2.12 in production

Stéphane Thiell
Stanford Research Computing Center (SRCC)
sthieill@stanford.edu
The wrong perception of “testing”...

In testing for 6+ months (Lustre 2.12.0 beta/RC on Fir)

▸ Limited number of 2.12 clients on Sherlock at that time
▸ Almost everything was working great during small scale testing (< 10 clients)
  ▸ even with DNE + DoM + PFL

In production since February 5, 2019

▸ servers: Fir (2MDS 72TB/8OSS 3PB)
▸ clients: Sherlock (~1,400-node cluster)
▸ unstable for ~3 months, finally stable by early May 2019
▸ HUGE impact on users and their jobs, and their experience
Critical problems seen in production with Lustre 2.12.0

DNE
- Locking issue with rename to remote directories
  - LU-12037 Possible DNE issue leading to hung filesystem
    lctl set_param mdt.*.enable_remote_rename=0
    (reinstate -EXDEV, option available in 2.12.1)

DoM
- MDS deadlock due to locking issues
Lustre 2.12.0: patching needed

Our #1 patch for DoM: async_discard

- **LU-11359** racer test 1 times out with client hung in dir_create.sh, ls, ... and MDS in ldlm_completion_ast()
- Special thanks to Mikhail Pershin (Whamcloud)
- Hopefully will make it into 2.12.2!
- Requires ALL Lustre clients to be patched too!
- ```bash
[[ $(lctl get_param mdc.*.import) =~ async_discard ]] && echo ok
```
Lustre 2.12: conclusion

Lustre 2.12 is finally stable on Sherlock, just in time for LUG’19 after problem mitigations and patching

Lustre 2.12.1 doesn’t include important DoM fixes

More tests at scale needed for new features in Lustre

Many thanks to Peter Jones, Andreas Dilger, Patrick Farrell, Mikhail Pershin, Lai Siyao, Oleg Drokin, Alex Zhuravlev, Amir Shehata, Joe Gmitter, and the rest of the Whamcloud team!
Extra slides
Fir uses DNE (Distributed Namespace)

2 MDS, 4 MDTs of 18 TB each (SSDs)
DNEv2 only used to stripe new user and group directories

```bash
$ lfs getdirstripe /fir/groups

lmv_stripe_count: 4  lmv_stripe_offset: 1  lmv_hash_type: fnv_1a_64
mdtidx  FID[seq:oid:ver]
  1 [0x2400000400:0x8:0x0]
  2 [0x2c0000401:0x6:0x0]
  3 [0x280000403:0x2:0x0]
  0 [0x200000401:0x6:0x0]

$ lfs getdirstripe /fir/groups/aaiken

lmv_stripe_count: 0  lmv_stripe_offset: 1  lmv_hash_type: none

$ lfs getdirstripe /fir/groups/aauclert

lmv_stripe_count: 0  lmv_stripe_offset: 3  lmv_hash_type: none
```
Fir uses DoM (Data-on-MDT)

Our DoM sizing process was described at LAD’18

TL;DR:
- DoM stripe size set to 128KB
- 75% of scratch files expected to be stored on MDTs only...
Fir uses PFL (Progressive File Layout)

<table>
<thead>
<tr>
<th>File size range</th>
<th>PFL striping</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-128KB</td>
<td>1 MDT stripe</td>
</tr>
<tr>
<td>128KB-16MB</td>
<td>1 OST stripe</td>
</tr>
<tr>
<td>16MB-1GB</td>
<td>2 OST stripes</td>
</tr>
<tr>
<td>1GB-32GB</td>
<td>4 OST stripes</td>
</tr>
<tr>
<td>32GB-256GB</td>
<td>8 OST stripes</td>
</tr>
<tr>
<td>256GB+</td>
<td>16 OST stripes</td>
</tr>
</tbody>
</table>
Lustre 2.12.0: problem mitigation examples

LU-11976 req wrong generation leading to I/O errors on 2.12 clients

- Disabled idling connection feature:
  ```bash
  lctl set_param osc.*.idle_timeout=0
  ```

LU-11888 + LU-11937 Inet.service randomly load tcp NIDs

- Added to `/etc/systemd/system/Inet.service.d/deps.conf` on all clients
  ```ini
  [Unit]
  After=dkms.service network.service

  [Service]
  # we don't want tcp nids
  ExecStartPost=-/usr/sbin/lnetctl net del --net tcp
  ```
Lustre 2.12.0: mitigation examples cont’d

LU-11989 Global filesystem hangs in 2.12
  ▶ Reduced ptlrpc timeout:
    options ptlrpc ld1m_enqueue_min=30
    options ptlrpc at_max=20

LU-12037 Possible DNE issue leading to hung filesystem
  ▶ lctl set_param mdt.*.enable_remote_rename=0
    (reinstate -EXDEV, option available in 2.12.1)
Lustre 2.12.0: our current patchset

The following patches have finally been working for us since end of April 2019

- LU-10777 dom: disable read-on-open with resend
- LU-11285 ldlm: reprocess whole waiting queue for IBITS ⚠ not in 2.12.1!
- LU-12018 quota: do not start a thread under memory pressure
- LU-11359 mdt: fix mdt_dom_discard_data() timeouts ⚠ not in 2.12.1!
- LU-11964 mdc: prevent glimpse lock count grow
- LU-12037 mdt: add option for cross-MDT rename
- LU-12065 Ind: increase CQ entries
- LU-12037 mdt: call mdt_dom_discard_data() after rename unlock
LU ticket list (for reference)

**LU-9341** PFL: append should not instantiate full layout
**LU-11761** blocked MDT mount and high cpu usage from lodXXXX_recYYYY threads
**LU-11964** Heavy load and soft lockups on MDS with DOM
**LU-11976** req wrong generation leading to I/O errors on 2.12 clients
**LU-11888** Unreachable client NID confusing Lustre 2.12
**LU-11937** Inet.service randomly load tcp NIDs
**LU-11939** ASSERTION(tgd->tgd_tot_granted >= ted->ted_grant) on OSS
**LU-11967** MDS LBUG ASSERTION(o->opo_reserved == 0) failed
**LU-11989** Global filesystem hangs in 2.12
**LU-12037** Possible DNE issue leading to hung filesystem
**LU-12096** ldlm_run_ast_work call traces and network
**LU-12178** MDS deadlock with 2.12.0 (quotas?)
**LU-12215** OSS network errors 2.12
**LU-12250** MDS with high rate of ldlm_cancel and Prolong DOM lock