Whamcloud

Lustre DNE Status Update

Lai Siyao

DDN® STORAGE

History

- Remote Directory in 2.4
- Striped Directory in 2.8
- Directory Migration in 2.10
- Directory QoS Allocation in 2.12
- Directory Restripe in 2.14

Remote Directory

Whamcloud

- Parent directory inode and its inode are located on different MDTs
- Dirent is stored in parent inode
- ‘lfs mkdir -i <MDT_index> …’

Striped Directory

Whamcloud

- Similar to file, directory is split into stripes
- Sub file dirents are stored in stripes by file name hash
- ‘lfs mkdir -c <stripe_count> -i <mdt_index> -H <mdt_hash> …’

Default Directory Layout

Whamcloud

- Similar to default file layout, used to create sub directories
- A major difference from default file layout: global default directory layout is not inherited, and it only affects subdirectory creation under ROOT
- MDT pool is not supported
- ‘lfs setdirstripe -D -c <stripe_count> -i <mdt_index> -H <mdt_hash> …’

Directory Striping Policy

Whamcloud

- Which MDTs are chosen in directory creation?
  1. User specified MDTs?
     'lfs mkdir -i <mdt_index> …'
  2. Parent directory has default directory layout, and default starting MDT index is not "-1"?
     Create directory on MDTs listed in default directory layout.
  3. Parent directory is plain directory, and user mkdir by 'lfs mkdir -i -1 …' or parent default dir layout starting MDT index is -1?
     QoS allocate objects on MDTs by space and inode usage.
  4. Create directory on MDT where its dirent is stored
  5. QoS allocate stripes on MDTs by space and inode usage if it's striped directory

QoS Allocation Policy

Whamcloud

- QoS allocate MDT inode for subdirectories
  - ‘lfs mkdir -i -1 …’
  - Parent directory is plain, and starting MDT index in default directory layout is “-1”, normally this only needs to be set on ROOT to balance MDT usage.
- QoS allocate stripes for striped directory by default.

Directory Migration

Whamcloud

- Migrate directory between MDTs
  - Source MDTs and target MDTs can overlap.
  - Almost all sub file inodes will be moved.
  - ‘lfs migrate -m …’, it can be rerun to resume failed migration.

Directory Restripe

Whamcloud

- Increase/Decrease directory stripe count based on current layout
- new directory hash type “crush”, a consistent hash algorithm
- minimum sub files need to be moved: delta/total stripe count
- ‘lfs setdirstripe -c <mdt_count> …’

Directory Auto Split

Whamcloud

- By default directory is created as plain directory
- When it becomes large, it will be split automatically
  - ‘lctl set_param mdt.*.enable_dir_auto_split=1’ to enable auto split
  - ‘lctl set_param mdt.*.dir_split_count=<count>’ to set split threshold
  - ‘lctl set_param mdt.*.dir_split_delta=<delta>’ to set split increase stripe count
  - ‘lctl set_param mdt.*.dir_restripe_nsonly=1’ to migrate dirents only
  - ‘lctl set_param lod.*.mdt_hash=crush’ to set default MDT hash type

DNE Namespace Inconsistency

Whamcloud

- Some directories or files are not accessible, or directory layout is broken
- LFSCK check namespace consistency:
  - ‘lctl lfsck_start -M <fsname>-MDT0000 -a -t namespace’ on MDT0 to start
  - ‘lctl lfsck_query -M <fsname>-MDT0000’ on MDT0 to query

Whamcloud

Thank You!

DDN® STORAGE