Whamcloud

Lustre Development Discussion

Andreas Dilger

New ldiskfs Improvements

Whamcloud

- Major `ldiskfs` features merged into upstream `ext4/e2fsprogs`
  - Large xattrs (`ea_inode`), directories over 10M entries/2GB (`large_dir`), project quota
- One more Lustre-specific feature remains to be merged
  - Extended data in directory (`dirdata`) - needs unit test interface before merge
- Existing `ext4` features available that could be used by Lustre on `ldiskfs`
  - Tiny files (1-600ish bytes) could be stored directly in the MDT 1KB inode (`inline_data`)
  - Metadata integrity checksums (`metadata_csum`)
  - Efficient allocation for large OSTs (`bigalloc`)
- New `ext4` features currently under development
  - Verity – data checksums stored
  - Directory shrink – reduce space as files deleted

Foreign Layout Integration with HSM

Whamcloud

- Current HSM state stored as a separate `trusted.hsm` xattr
- PCC is also a form of HSM, using client-local NVMe storage as "archive"
- Archive copy is a mirror of the file, handled similar to FLR component
- New LOV *foreign* layout type added for linking to DAOS containers
  - Foreign type (4 bytes) + arbitrary xattr to identify content/objects
  - Can be used as a whole or partial-file component, similar to DoM or RAID-0
- Use foreign type for HSM state stored in a component
  - Allow multiple HSM archives per file
  - S3, HPSS, TSM, remote Lustre, …
- Allow partial file archive/restore
- Use for ldiskfs container images?