btrfs: Introduction and Performance Evaluation

Douglas Fuller
Oak Ridge Leadership Computing Facility / ORNL
LUG 2011
btrfs: overview

- Implemented in Linux 2.6.29
- 64-bit
- Extent-based
- Copy on write
- Multiple devices
  - Object pools (RAID-0, RAID-1, and RAID-10 for now)
  - Dynamic rebalancing
- Subvolumes with snapshots
- Available transparent compression (zlib)
btrfs: still to come

- Erasure coding (RAID-5/RAID-6)
- fsck
- Dedup
- Encryption
Goals

• Determine baseline btrfs performance
• Evaluate pooling at the filesystem layer
  – Avoids expensive RAID controllers
  – Provides additional features
• Evaluate overhead from additional features
Test Platform

• Target: LSI Engenio (xbb2), 1TB fibre channel disks
  – DDR InfiniBand
  – Each disk on a separate backend channel
  – Controller write cache disabled
  – O_DIRECT
  – 1MB request sizes
  – I/O elevator noop

• Initiator
  – Nehalem with Linux 2.6.37
  – SRP
Baseline Performance

![Bar chart showing baseline performance across different groups with raw data for LUNs and Disks]
Sequential write

Sequential Write

- btrfs-multi
- btrfs-single
- ext3
Sequential read

Sequential Read

- btrfs-single
- btrfs-multi
- ext3
Random write

Random Write

- btrfs-single
- btrfs-multi
- ext3
Random Read

Random Read

- btrfs-single
- btrfs-multi
- ext3
Untar

untar

- btrfs-single
- btrfs-single-r
- btrfs-multipath
- btrfs-multipath-r
- ext3
Untar with compression
Random reads with compression

Random Read

- read-uncmp-s
- read-cmp-s
- read-uncmp-m
- read-cmp-m
Random write with compression

Random Write

- write-uncmp-s
- write-cmp-s
- write-uncmp-m
- write-cmp-m