SC18
Dallas, TX | hpc inspires.

Whamcloud

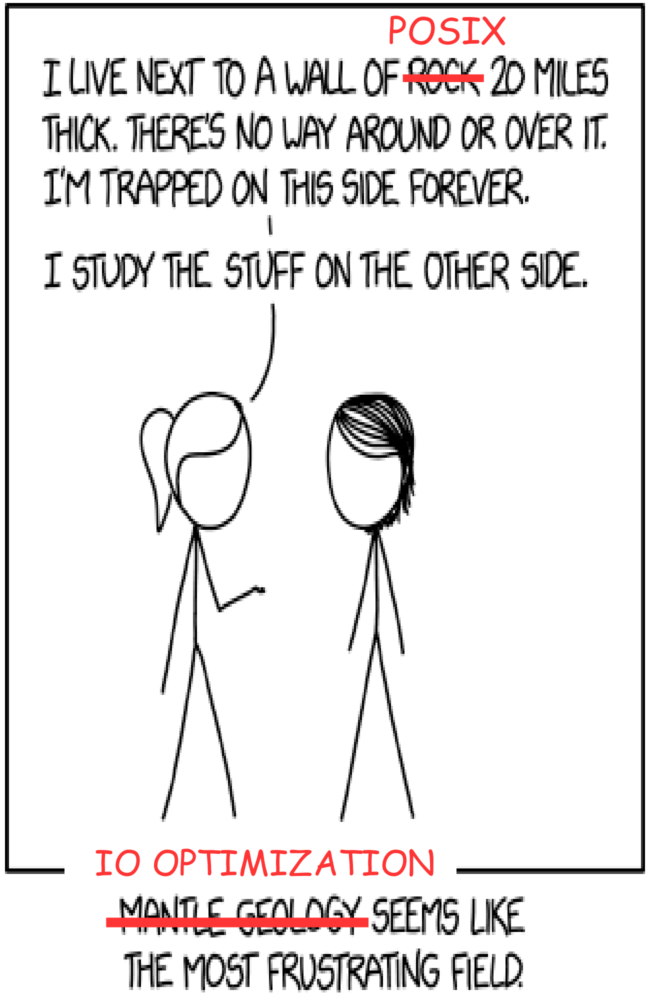
Analyzing Parallel I/O BOF - SC'18

Andreas Dilger, Lustre CTO and Principal Engineer

DDN® STORAGE

What Lessons Have I Learned So Far?

- Storage/IO is hard, and will continue to be so
  - Compute and communication is essentially stateless… *lucky!*
  - Storage has long-term behaviors – that's the whole point!
    - But… fragmentation, alignment, location, age, pattern, intermixing, …
- What are applications and storage *really* doing?
  - Mental model and reality often misaligned
  - Unknown IO *intent* at storage layer, have to guess desired behavior
  - Benchmark and compare to theoretical performance at each layer
- Label and track IO by application
  - Live monitoring/debug, post-run summary, understand IO patterns
  - Filesystem can use labels to improve scheduling, allocation/grouping
  - Like any resource, IO needs accounting - space, IOPS, peak bandwidth
- Always-on monitoring at some level
  - JobStats – MPI JobID sent from client to server with every Lustre RPC
  - Darshan – learn what application is doing, users often do not know


POSIX
I LIVE NEXT TO A WALL OF ~~ROCK~~ 20 MILES THICK. THERE'S NO WAY AROUND OR OVER IT. I'M TRAPPED ON THIS SIDE FOREVER.
I STUDY THE STUFF ON THE OTHER SIDE.
IO OPTIMIZATION
~~MANTLE GEOLOGY~~ SEEMS LIKE THE MOST FRUSTRATING FIELD.

https://xkcd.com/2058/

What Is Needed Next To Continue Improving?

Whamcloud

- Deeper integration of compute, comms, storage analysis
  - Facilitate understanding of global system analysis and behavior
  - Improve utilization of compute, network, and storage – jitter, bottlenecks
  - Single application optimization also has limitations – intra-job contention
- POSIX embrace and *ad-hoc* extend outside of existing applications
  - Can't tune all apps or remove POSIX (cf. FORTRAN), need bypass methods
- Better integration of IO libraries with apps and storage system
  - Concentrate knowledge/optimization efforts in **common** libraries
  - Communicate IO patterns for file/directory creation/access/lifecycle
  - Provide hints to IO library/storage to allow ***dynamic*** IO optimization
- Automated (client-side) analysis and tuning of IO workloads
  - Learn IO pattern for app/user from repeated run cycles
  - **Persistent storage** of IO patterns and optimization hints with **user runtime**

Whamcloud

Thanks. Questions?

adilger@whamcloud.com

DDN®
STORAGE