

A few thoughts on storage

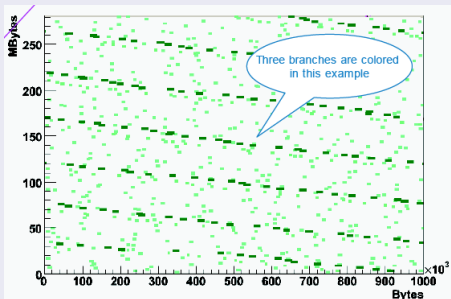
Greig A. Cowan

University of Edinburgh

GridPP21 Swansea

- If VOs use POSIX I/O protocols (rpio, dcap, xroot) to open and read data files, then you will see **LOTS** of random I/O on your disks.
- This is completely different to the sequential writes that you see when data is being transferred in by FTS.
- Being able to open 1 file using rpio is **NOT** the same as opening 1000.
 - You will see weird problems where things work for you, but not for users.
 - For rpio, first stop should always be to check certificates!

From a ROOT presentation at CHEP 07



- Sequentially processing events in a files does not mean sequential access to the file.

- RHUL has ~ 300 TB of disk. By December, Glasgow will have ~ 500 TB!
 - How will the storage middleware perform? dCache has been shown to operate at this level, for DPM, it's uncharted territory.
 - Do sites have the tools to help manage this amount of disk?
- We know what works well for WAN transfers:
 - Disk servers sitting in front of RAID'ed disk
 - XFS and Areca cards.
 - Spread disk over a sufficient number of servers (20TB/box?).
 - Bring new disk online at the same time.
- Thinking about chaotic user analysis. . .
 - Spread disk over a sufficient number of servers
 - You don't want 1000 jobs trying to access a dataset that's on a single disk.
 - Networking. You probably need >2 Gb/s links to these storage boxes.
- If you are thinking about using your WN disk, think very very carefully.

- RHUL has ~ 300 TB of disk. By December, Glasgow will have ~ 500 TB!
 - How will the storage middleware perform? dCache has been shown to operate at this level, for DPM, it's uncharted territory.
 - Do sites have the tools to help manage this amount of disk?
- We know what works well for WAN transfers:
 - Disk servers sitting in front of RAID'ed disk
 - XFS and Areca cards.
 - Spread disk over a sufficient number of servers (20TB/box?).
 - Bring new disk online at the same time.
- Thinking about chaotic user analysis. . .
 - Spread disk over a sufficient number of servers
 - You don't want 1000 jobs trying to access a dataset that's on a single disk.
 - Networking. You probably need >2 Gb/s links to these storage boxes.
- If you are thinking about using your WN disk, think very very carefully.

- Please talk to us!
- Mailing list, meeting, blog, IM are all there to be used.
- gridpp-storage list has a lot of good people on it who know a lot of stuff.
- Some sites very good at getting in touch.
- Other sites, you never hear anything from.