Death of Data – actually {/,/g/}data* filesystems

Robert C. Bell | CSIRO IMT Scientific Computing

29 May 2018

Outline

- Problem of shared filesystems
- Some issues
- Some solutions
  - Dataflows, not data stores
  - Use scratch more – scalable flushing
  - Use /data less
  - Provide utilities to manage data flow

  - See paper on Storage Magic

Problem of shared filesystems

- Problem of shared filesystems for users
  - Home: too small
  - Scratch: too dangerous
  - Data: too full
  - HSM: too hard
  - Cloud: too new
- Problem of shared filesystems for system managers
  - Home: backup requirements
  - Scratch: too many files to flush
  - Data: too stagnant
  - HSM: too hard? Takes careful management
  - Cloud: too hungry

Some example issues

- NCI
  - /short not flushed for 5 years
    - can’t get enough campaign space (412 M files on 22-05-2018)
  - /g/data areas – allocations locked up
  - Massdata – too small, perceived to be hard to use:

    'mdss put ...' of two or more files will fail if the average file size is smaller than 1024Kb.
- Pawsey
  - struggling to flush /scratch
    - large number of files (787 M files on 22-05-2018)
  - limited access to HSM storage without extensive project proposals
- CSIRO SC
  - overload on shared areas led to stalls
  - no longer giving access to /data area for new users

/data and project area issues

- Risk – little protection from loss – no backup or HSM
  - Leads to complacency
  - No user perception of consequences of loss
- Hard to manage usage and get space for new projects
  - used to do “name and shame” – peer pressure => bullying
- Hard to recover allocations, except project end or user cessation
  - difficult even then to get either action to copy the data to safe keeping, or someone to take responsibility to delete data in the era of “keep everything”

Some solutions

- Dataflows, not data stores
  - data is a journey, not a destination
- Use scratch more, with big quotas
  - scalable flushing makes this feasible
- Use /data less: use scratch and/or HSM-enabled
  - no support for un-managed static data
- Collections
  - Enables scaling
  - CSIRO was adding 5 M files per day to science data stores!
- Utilities to manage data flow

Utilities to manage data flow

- Allow data collections to be maintained on scratch storage
- Allow protection from loss by maintaining reference copy in HSM
  - Consolidation to get scalability
- Allow optimised re-syncing in the face of flushing
- Allow updating
- Replace the use of {/,/g/}data* filesystems
- Can also be used for dealing with departed users, or users who have filled Home

Capabilities from prototype utility

- In a working area in /flush* or /short

```
mirror.sh help
mirror.sh create
mirror.sh list
mirror.sh sync
mirror.sh update
mirror.sh moveto new_area
mirror.sh flush
mirror.sh release
mirror.sh recall
…  9 more operations
```

- Available of SC systems bracewell, pearcey and ruby, and NCI raijin
  - man ~bel107/bin/man/mirror.1 man ~rcb599/bin/man/mirror.1
- Work with a mirror in $STOREDIR or /g/data* & mdss (at NCI)

Examples from prototype utility

- 11) To undertake a move of an existing area, say in SC $DATADIR/mydata to a $FLUSH* area, but not need to have the data ready now. This can be quite quick, since no user data is moved from the old to the new location.
- cd $DATADIR/mydata
- mirror.sh create (if not done already).
- mirror.sh sync
- mirror.sh flush
- mirror.sh moveto $FLUSHDIR/mydata
- cd $FLUSHDIR/mydata
- mirror.sh list

Examples from prototype utility

- 12) To squirrel away part of the $HOME area on pearcey, bracewell or NCI raijin because quota limits have been hit.
- cd $HOME/dormant
- mirror.sh create
- mirror.sh list
- mirror.sh flush
- mirror.sh release

Conclusion

- With scratch under control:

Death of {/,/g/}data!

- Hopelessly unmanageable storage!
- Users can use either scratch, or HSM-managed, or Bowen Research Cloud
- Prototype utility to mirror scratch areas into persistent storage, and have ability to re-create after flushing

Thank you

CSIRO IMT Scientific Computing
Robert C. Bell
CSIRO HPC National Partnerships

t +61 428 108 333

e Robert.Bell@csiro.au

w www.csiro.au