CSIRO ASC
Data Protection Technologies

Robert C. Bell | Technical Services Manager
24 July2012

CSIRO IM&T ADVANCED SCIENTIFIC COMPUTING
www.csiro.au

With Jeroen van den Muyzenberg and Peter Edwards

CSIRO ASC

Advanced Scientific Computing

- Part of IM&T
- History of HPC – more than 20 years
- In-house systems, and access to partner facilities
- Systems, Storage, Software, Services, Support

Data Protection

- HPC led to need for storage facilities
- HSM for over 21 years
- CSIRO ASC Data Store
  - Primary storage system
  - Direct user access, plus SAMBA, HTTP, etc
  - Data on primary disc, secondary cache disc, MAID and tapes
  - Highly protected large-scale storage

Other (non-HSM) systems: need to do backup

ASC – Backup for non-HSM systems

- Techniques developed over period from 1998
  - **Needed to backup home areas on supercomputers, support systems, non-HSM file systems on Data Store host, etc.**
  - **Attracted to standard UNIX tools**
  - **Had access to an HSM system as a target**
  - **Needed efficient, reliable system, and scalable**
- Four key technologies
  1. **Use of rsync for transfer and storage of only the changed data at each cycle**
  2. **Use of hard-linking to reduce the storage demands**
  3. **Extended Tower of Hanoi scheme for management**
  4. **Integration with HSM**

Features

The techniques in use provide:

1. coverage in time matching the likelihood of restores
2. full backups every time, for the cost of incrementals
3. simple visibility of the backup holdings
4. simple recovery for individual files and complete file systems (with fast option at the cost of more storage)
5. no vendor dependency (open source and in-house)

Features

The techniques in use provide:

6. use of HSM as a target for backups to protect against ‘all the backup eggs in one basket’
7. centralised tape management
8. no scanning of the HSM needed
9. space saving of about a factor of five compared with conventional backups
10. strategies for superseded and retired backups
11. hosted on Linux with SGI’s DMF HSM:

    clients for Linux, Windows and partially for OS X

News

* Bad Moon Rises Over Cloud Perceptions *

It's a glorious day for those with backup and recovery services to pitch, not to mention for those who are nipping at the heels of Amazon Web Service's cloud empire. On the flip side, for those who just got word that their data is unrecoverable via a short letter, not to mention those who have been proclaiming cloud computing as the safe bet for mission critical applications--it's been quite a bleak morning.

Read More: http://www.hpcinthecloud.com/ct/uz5613109Biz11013583

HPC in the Cloud Weekly Update, May 2011

e-mail from Amazon

Hello,

A few days ago we sent you an email letting you know that we were working on recovering an inconsistent data snapshot of one or more of your Amazon EBS volumes. We are very sorry, but ultimately our efforts to manually recover your volume were unsuccessful. The hardware failed in such a way that we could not forensically restore the data.

What we were able to recover has been made available via a snapshot, although the data is in such a state that it may have little to no utility...

If you have no need for this snapshot, please delete it to avoid incurring storage charges.

We apologize for this volume loss and any impact to your business.

Sincerely,

Amazon Web Services, EBS Support

This message was produced and distributed by Amazon Web Services LLC, 410 Terry Avenue North, Seattle, Washington 98109-5210

HPC in the Cloud Weekly Update

Data protection is not easy!
Example: Google

http://gmailblog.blogspot.com.au/2011/02/gmail-back-soon-for-everyone.html

- Gmail back soon for everyone
- Posted by Ben Treynor, VP Engineering and Site Reliability Czar (24x7)

  Imagine the sinking feeling of logging in to your Gmail account and finding it empty. That's what happened to 0.02% of Gmail users yesterday, and we're very sorry. The good news is that email was never lost and we've restored access for many of those affected. Though it may take longer than we originally expected, we're making good progress and things should be back to normal for everyone soon.

Data protection is not easy!
Example: Google

I know what some of you are thinking: how could this happen if we have multiple copies of your data, in multiple data centers? Well, in some rare instances software bugs can affect several copies of the data. That's what happened here. Some copies of mail were deleted, and we've been hard at work over the last 30 hours getting it back for the people affected by this issue.

To protect your information from these unusual bugs, we also back it up to tape. Since the tapes are offline, they're protected from such software bugs. But restoring data from them also takes longer than transferring your requests to another data center, which is why it's taken us hours to get the email back instead of milliseconds.

…

Outline

- The need for backups
- Unitary versus discrete
- Disc target versus tape target (feature 6)
- Use of HSM as a target (features 6, 7 and 8)
- Production of backups: rsync with - -link-dest (features 2, 3, 4, 5, 9)
- Management of backups: Tower of Hanoi (features 1, 3 and 10)
- Scope:

  The backup of non-HSM file systems (source) to an HSM-managed file system (destination or target).

  Sources can be remote, local or the same host as the destination.

The need for backups

- Need to do risk assessment
  - What are the consequences if the data is lost?
  - How could the data be lost?
    - User errors (oops, wrong directory)
    - Systems staff errors (even bigger oops)
    - System software failures (RAID software)
    - Hardware failures (disc drives)
    - Intruder (physical or network)
    - External event
- Backups are snapshots of data holdings, made to allow recovery in some events.
  - Need latest, and going back in time.
- Backups are not archives!

University of Queensland

Unitary versus discrete

Dilemma: have many files. Make backup to:

1. a container (dump file or tape); or

   easy to manage, and good for recovery of the lot, but requires serial access (hence slow) for individual file recovery, and has no addressability. So, if you want to keep a current full dump, have to write the full information each time (or deal with fulls and incrementals/differentials).

2. a file system (equal numbers of files)

   not normally feasible, because of the number of files. Good for individual file recovery.

Disc versus tape target (features 6 & 7)

- Backups to disc
  - attractive – easy to manage, reliable
- Backups to tape
  - off-line, hence safer, but need drive on each system, or backup over network: less reliable
- Backups to disc then to HSM
  - best of both worlds.
  - with an HSM already dealing with tapes, no extra tape management for the backups. Get stacking for nothing!
    - just target to an HSM-managed filesystem.
    - bonus – get controlled protection.
- 6. use of HSM as a target for backups to protect against 'all the backup eggs in one basket'
- 7. Provides for centralised tape management

Rsync with --link-dest

- rsync: copies files, but only as necessary.

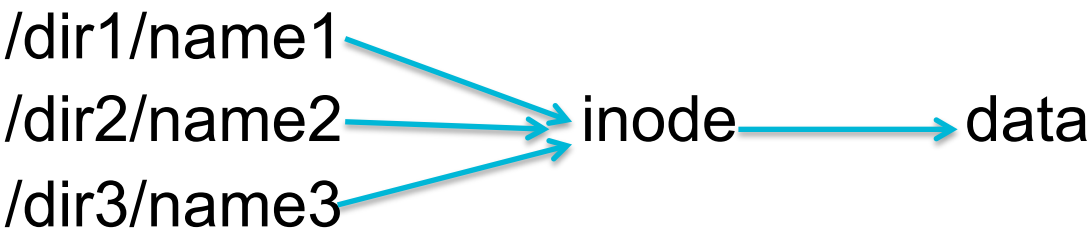
source -->	destination directory

- Colleague, Jeroen van den Muyzenberg discovered a feature of the rsync utility in 2007

	--link-dest= previous backup directory
source ->	destination directory

- If file is unchanged since last backup:
  - not copied from source, but hard-linked from previous directory.
- Hard-links:
  - same file, but aliased in multiple directories

Hard-links



- Unix directories contain file names and point to inodes
- A unix inode contains everything about a file except its name and the data – it points to the data
- Hard-links: multiple names for the same inode and hence data
- For backups, have:

  /dir_time1/filename
  /dir_time2/filename
  /dir_time3/filename

Rsync with --link-dest (features 2, 3)

- Each directory is a full backup, but only changes need to be transferred.
- 2. full backups every time, for the cost of incrementals
- Can use standard UNIX/Linux to look at backups (ls, etc).
- Can see a view back in time.
- 3. simple visibility of the backup holdings

Rsync with --link-dest (feature 3)

```
dmls –ail home.*/u1/file1
inode p    l size   date                          state directory
..199 -rw  1 6649 2007-11-26 16:24 (OFL) home.20080409/u1/file1
..762 -rw  1 9364 2009-08-10 11:11 (OFL) home.20090919/u1/file1
..593 -rw  1 5915 2010-05-26 10:47 (OFL) home.20100617/u1/file1
..509 -rw  1 6396 2010-09-30 17:14 (OFL) home.20101025/u1/file1
..670 -rw  2 6303 2010-12-10 15:27 (OFL) home.20101228/u1/file1
..670 -rw  2 6303 2010-12-10 15:27 (OFL) home.20110302/u1/file1
..354 -rw  5 6413 2011-03-09 13:54 (OFL) home.20110318/u1/file1
..354 -rw  5 6413 2011-03-09 13:54 (OFL) home.20110403/u1/file1
..354 -rw  5 6413 2011-03-09 13:54 (OFL) home.20110407/u1/file1
..354 -rw  5 6413 2011-03-09 13:54 (OFL) home.20110410/u1/file1
..354 -rw  5 6413 2011-03-09 13:54 (OFL) home.20110411/u1/file1
..718 -rw  3 6217 2011-04-12 16:11 (REG) home.20110412/u1/file1
..718 -rw  3 6217 2011-04-12 16:11 (REG) home.20110413/u1/file1
..718 -rw  3 6217 2011-04-12 16:11 (REG) home.20110414/u1/file1
```

Rsync with --link-dest (features 4, 5, 9)

- Can use standard UNIX/Linux cp to recover individual files, or rsync to recover the lot
  - A good recall from HSM first might help!
  - Option to keep the latest copy of all files on-line, for fast restores
- 4. simple recovery for individual files and complete file systems (with fast option at the cost of more storage)
- All standard UNIX/Linux commands plus rsync
- 5. no vendor dependency
- Allows multiple backups, without large growth
  - for slowly changing file system, most files are hard-linked
- 9. space saving of about a factor of five compared with conventional backups

Rsync with --link-dest (feature 10)

- Lots of problems, nearly all solved
  - Number of inodes grows – directories not hard-linked
  - Number of file paths grows more – tolerable
    - Purging and thinning
    - Recycling, but rsync deficiencies – over-come.
  - Consolidation – reduces number of files – tar based
  - Schemes for superseded backup series – culling
  - Schemes for discontinued backup series – archiving
- 10. strategies for superseded and retired backups

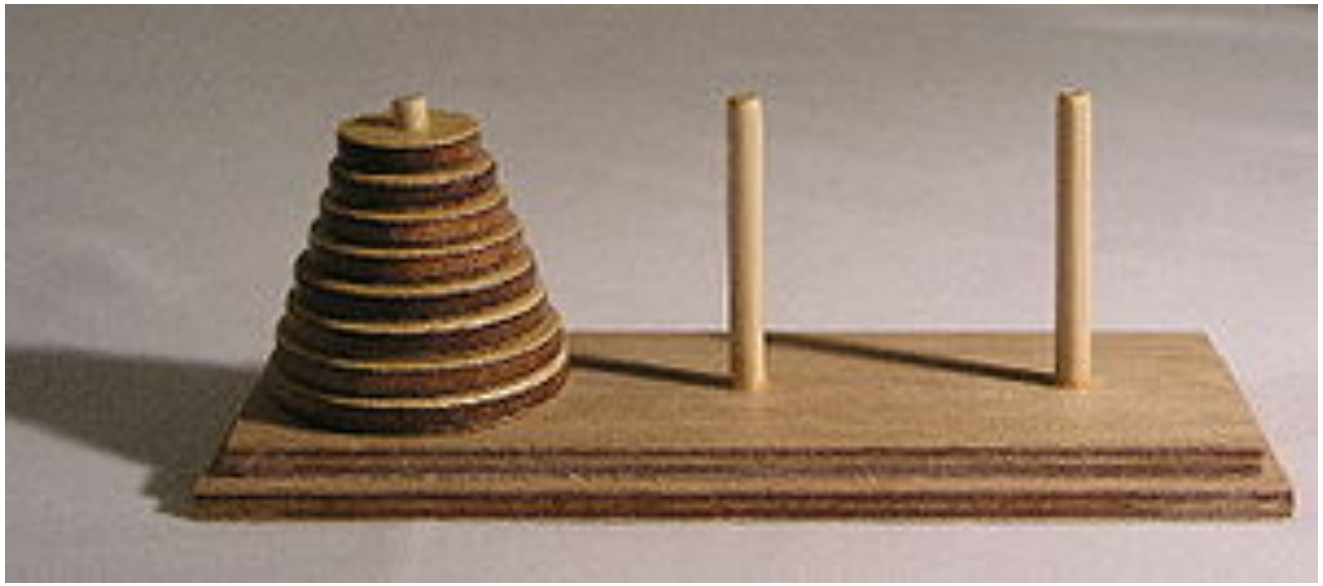
Rsync with --link-dest on DMF (feature 8)

- Use DMF & xfsdump to protect:
  - all eggs in one basket otherwise
  - use tuning of xfsdump and DMF to keep dump times tolerable
- Issue – only one physical copy?
  - Two copies in DMF, or restart backup series occasionally.
- Issue – scan times?
  - No longer need DMF to scan filesystem for migration candidates
    - Scheduled targetted migrations as part of the backups
- 8. no scanning of the HSM needed

Tower of Hanoi

- Can't keep every backup!
  - Need to make decisions about what to keep, in some way representative of the likelihood of the need to recover from the backup.
  - Intuitively, less likely to need backups the further back in time they were taken.
  - You hope most mistakes are discovered quickly, though not always the case.
  - Need management strategy for the backups.
  - Usually time based
    - Hourly, daily, weekly, monthly, quarterly, yearly backups kept (e.g. Apple's Time Machine)

Tower of Hanoi



Aim: to move all the rings from one tower to another, one at a time, with no ring placed on a smaller ring.
So, if we number the rings from number 1, the smallest, the moves are in the order:

ring 1, ring 2, ring 1, 3, 1, 2, 1, 4, 1, 2, 1, 3, 1, 2, 1, 5, 1, 2, etc.

Tower of Hanoi

Tower of Hanoi move number	Tower of Hanoi ring to move	
1	1	00001
2	2	00010
3	1	00011
4	3	00100
5	1	00101
6	2	00110
7	1	00111
8	4	01000
Backup sequence number	**Backup set number**	**Binary sequence number**

Tower of Hanoi

- Usage:
  - assign a sequence number, then a set number (from ToH)
  - keep one (directory, tape) of each set number

Assigning slash.20090823 to seq 32 to set 6
Assigning slash.20090909 to seq 48 to set 5
Assigning slash.20090917 to seq 56 to set 4
Assigning slash.20090921 to seq 60 to set 3
Assigning slash.20090923 to seq 62 to set 2
Assigning slash.20090924 to seq 63 to set 1

Tower of Hanoi (feature 1)

- Advantages
  - Easily programmed to determine which backup set to use.
  - No complication of days of the week, days/weeks per month, etc.
  - Can use a sequence number with a date, rather than just a date. This allows arbitrary spacing of backups.
  - The backups space out more, the further back you look. For example, with daily backups, there will be better than:
    - 2 backups in the last two days
    - 3 backups in the last 4 days
    - 4 backups in the last 8 days
    - 5 backups in the last 16, etc.
- 1. coverage back in time adjusting to the likelihood of recovery being needed

Tower of Hanoi (feature 9)

- Advantages
  - The number of backups kept grows as $\log_2 n$, where n is the number of backups taken.
  - Storage growth will be no more than that, and will often be far less because of rsync hard-linking.
- 9. space saving of about a factor of five compared with conventional backups
  - Coverage adjusts to the use of a system (e.g. Home PC)
    - since the keeping is based on the sequence number rather than time of day, gets it right when deciding what backups to keep when the system is intermittently working.
  - The system self-heals
    - Missing a day is not a disaster, unlike time based backups where missing a critical day can mess up the scheme.
    - Can easily resume interrupted backup, unlike unitary backup.

Tower of Hanoi - extensions

1. Baseline backup
2. Keep two set ones, to provide better cover of recent times
3. We make backups to a new area before discarding an old.
4. We can recycle old backup directories on disc
   - Better performance with rsync
   - (Never seen anyone else do that)

More ….

Tower of Hanoi - extensions

- 6. More coverage:
  - Can have hybrid scheme.
  - Can do higher density
- 7. Less coverage:
  - Keep only a set number of backups
  - Keep backups for only a fixed period, e.g. 3 years
- 8. Old backups:
  - Cull with respect to a replacement backup series
  - ‘Archive’: remove redundant information:
  - Consolidate: put files into containers, to reduce numbers of files.

Tower of Hanoi

- Usage:
  - Sequence and set number (from ToH)
  - keep one (directory, tape) of each set number

Assigning slash.20090722 to seq 0 to set 0
Assigning slash.20090823 to seq 32 to set 6
Assigning slash.20090909 to seq 48 to set 5
Assigning slash.20090917 to seq 56 to set 4
Assigning slash.20090920.recycle to seq 59 to set 1
Assigning slash.20090921 to seq 60 to set 3
Assigning slash.20090922 to seq 61 to set 1
Assigning slash.20090923 to seq 62 to set 2
Assigning slash.20090924 to seq 63 to set 1
Found 9 backups as expected up to sequence number 63

(Extensions to ToH in red or italics)

Tower of Hanoi

- Can still use for tapes – copies of dumps – hybrid scheme

Older tapes subject to Tower of Hanoi management
G61009 2009-03-15 0 0
G60581 2010-05-30 64 7
G60903 2010-09-26 80 5
G60592 2010-12-19 92 3
G60578 2011-01-09 94 2
G62450 2011-01-16 95 1
G60584 2011-01-23 96 6
Newer tapes kept as the most recent 13
G60576 2011-01-30 97 1
G61006 2011-02-06 98 2
.....
G61000 2011-04-10 108 3
G60571 2011-04-17 109 1

Hosts and clients

- Clients: just rsync – have run with Linux and Windows clients
  - Use rsync daemon to avoid root over the network
- Have used ToH over Apple Time Machine backups for a more rational management of holdings.
- Host: any system supporting rsync
  - If target is a plain file system, at risk of total loss: hence HSM
- Currently hosted on UV1000 with SGI's DMF HSM.
  - Could use any HSM on a file system supporting rsync.

11. hosted on Linux with SGI's DMF HSM:
    clients for Linux, Windows and partially for OS X

Features

The techniques in use provide:

1. coverage in time matching the likelihood of restores
2. full backups every time, for the cost of incrementals
3. simple visibility of the backup holdings
4. simple recovery for individual files and complete file systems
   (with fast option at the cost of more storage)
5. no vendor dependency (open source and in-house)
6. use of HSM as a target for backups to protect against ‘all the backup eggs in one basket’
7. centralised tape management
8. no scanning of the HSM needed
9. space saving of about a factor of five compared with conventional backups
10. strategies for superseded and retired backups
11. hosted on Linux with SGI’s DMF HSM:
    clients for Linux, Windows and partially for OS X

Implementation – another talk!

CSIRO

Search Go

Home | Contact Us | Facilities | Software | Support | News | Partnerships | User Information |

CSIRO ADVANCED SCIENTIFIC COMPUTING

User Filesystem Backup Status

Last updated Tue 2009-09-29 11:00:01 +1000

Error codes 23 and 24 are soft errors (files disappeared or open) - these can be ignored.

Sorted by date:

Host	Filesystem	Last backup	Status	Files Transferred	Total Files	Bytes Transferred	Total Bytes
cherax	/cs/datastore_0	Sun 2009-09-27 11:47	0	15555342	16188904	82704722344	4283195858944
burnet-fs0	/system/cs/home	Mon 2009-09-28 18:30	0	8526	846308	4442311132	313213832298
CBF	/current	Mon 2009-09-28 19:03	0	397	7244997	103613919	905086125446
sx-6	/cs/home	Mon 2009-09-28 20:48	23	1618	862158	29749655382	712029664160
cherax	/cs/datastore_9	Tue 2009-09-29 01:03	0	501940	16393816	3647467608	4554001555456

Thank you

CSIRO IM&T Advanced Scientific Computing
Robert C. Bell
Technical Services Manager

t +61 3 9669 8102
e Robert.Bell@csiro.au
w www.hpsc.csiro.au/