CSIRO ASC's local DMF tools

Peter Edwards | Systems Support Manager

4 December 2012

CSIRO IM&T ADVANCED SCIENTIFIC COMPUTING

www.csiro.au

CSIRO ASC local DMF tools

Over the years, various tools have been created or enhanced at CSIRO Advanced Scientific Computing to ease the day to day administration of its DMF Data Store. Some of these are just simpler and easier ways to use standard DMF utilities, but others provide capabilities which are difficult to achieve otherwise.

This presentation will show off our Top Ten.

Most examples below have been edited to shorten line lengths, and colouring has been removed.

The Top Ten

- dmd a wrapper around dmdadm for listing database entries
- dmfstatus shows VG activity, both current and the day so far
- dmgrep grep multiple DMF logs and interleave them
- dmlookup convert BFIDs, fhandles and DCM paths/keys to file pathnames
- dmorder show current and queued recalls on a per-tape basis
- dmsilo add/remove tapes to a tape library in various ways
- dmv a wrapper around dmvoladm for listing tapes and altering hflags
- find_bfids show the BFIDs of file chunks on a tape
- logw a "tail -f" log watcher
- tpstat a wrapper around oper, tmstat, ov_stat, msgd & ps to show DMF status

dmd

a wrapper around dmdadm for listing database entries

```
cherax# dmd 4fcef489000003ffd2f 4fcef48900000401080
                    BFID   ORIG        ORIG   ORIG MSP      MSP
                            UID        SIZE    AGE NAME     KEY
-------------------------------------------------------------------
4fcef489000003ffd2f  38675   358914892     1d se2      4fcef489000003ffd2f
4fcef489000003ffd2f  38675   358914892     1d te3      4fcef489000003ffd2f
-------------------------------------------------------------------
4fcef48900000401080  38675   358914892     1d se2      4fcef48900000401080
4fcef48900000401080  38675   358914892     1d te3      4fcef48900000401080
```

Equivalent to

```
  dmdadm -c ‘list 4fcef489000003ffd2f; list 4fcef48900000401080’
```

dmfstatus

shows VG activity, both current and the day so far

```
cherax$ dmfstatus
                   Reads
           Current              Today            Hit Rate
VolGrp Queued MiB-Queued  Total   MiB-Total  %Recalls %Data
cache       3    43035.2   1163    501992.2      11     7
maid        1      253.1    179     43057.5       2     1
se3         1    52189.9     14    500786.1       0     7
te2         5   303288.2   9000   5638734.9      86    80
Total      66   428125.2  10497   7072795.9

                   Writes
           Current              Today
VolGrp Queued MiB-Queued  Total   MiB-Total
cache       0        0.0    124    217275.3
maid        0        0.0      0         0.0
se3         0        0.0     45   1915378.2
te2         0        0.0      0         0.0
Total      37    19568.4   9126   7711428.7
```

dmgrep

grep multiple DMF logs and interleave them

```
cherax$ dmgrep Req=839540,
Grepping DMF logs from 20121102

daemon  15:35:45 I 209955-dmfdaemon Req=839540,request=recall,fhandle=0100000000000018857
daemon  15:35:45 I 209955-dmfdaemon Req=839540,reply=recall,rp=Request deferred
daemon  15:35:45 V 209955-dmfdaemon Req=839540,bc67b0/1, msprq dest_off=0 off=0 len=48730
daemon  15:35:45 V 209955-dmfdaemon Req=839540, Get_File sent to MSP se2.
ls      15:35:45 I 210413-dmatls Req=839540,bc67b0,Get_File,key=4036b2770000000009f1a8c5,
......(34 secs later)
movers  15:36:19 V 106654-dmatrc queue_chunk: Req=839540,ef4cc0/bc67b0, zn=175, cn=4200
movers  15:36:19 V 106654-dmatrc pick_recall_target: Req=839540,ef4cc0/bc67b0, zn=175,
.........(48 secs later)
movers  15:37:07 V 106654-dmatrc send_chunk_done: done: Req=839540,ef4cc0/bc67b0, zn=175,
ls      15:37:07 I 210413-dmatls Req=839540,bc67b0,Get_File,key=4036b2770000000009f1a8c5,
daemon  15:37:07 V 209955-dmfdaemon do_GetReq_Done: Req=839540,bc67b0, received MspGetReq
daemon  15:37:07 V 209955-dmfdaemon do_GetReq_Done: Req=839540,bc67b0,
daemon  15:37:07 V 209955-dmfdaemon recall_mspreply: Req=839540,bc67b0, off 0, len 487303
daemon  15:37:07 V 209955-dmfdaemon recall2: Req=839540, region off = 0, len = 487303375
daemon  15:37:07 I 209955-dmfdaemon Req=839540,reply=recall,rp=Request completed
```

dmlookup

convert BFIDs or fhandles to pathnames

Convert BFIDs, fhandles and DCM paths/keys to file pathnames (or to BFIDs or fhandles) using sqlite3 database built each night.

```
cherax$ dmlookup 4fcef4890000000000074825d \
0100000000000018857751791553287200000000010910860fd0bd4f00000000
Scan date: Fri Nov  2 01:00:10 EST 2012
Database date: Fri Nov  2 06:36:46 EST 2012

/datastore/d/IPCC/CMIP5/output/data/climdex/FGOALS-s2_r3i1p1_1850-2005.nc
/datastore/asc/edw192/is220.tar.gz
```

dmlookup

the database

The database used by dmlookup and other scripts is generated each night, and contains three tables:

- All the output from a full dmscanfs, indexed by BFID, fhandle and UID
- BFID and VSN mappings from dmdump of CAT database, indexed by BFID
- VG and VSN mappings from dmdump of VOL database, indexed by VSN

At CSIRO ASC, for 21M files on hybrid SSD/HDD filesystem, the dmscanfs and build of the 15GiB database take 25 and 40 minutes resp. (Hybrid filesystem gave the dmscanfs a 2x speedup.)

dmorder

show current and queued recalls on a per-tape basis

```
cherax# dmorder
Volume Group: se2 (4 mounts)                    Longest tape wait is 0:01:43
    *G60379 edw192[16:52:37]
     G62544 edw192[16:52:37]

Volume Group: te2 (7 mounts)                    Longest tape wait is 0:01:43
    *G60100 edw192[16:52:37]
    *G60125 edw192[16:52:37]
    *G61837 root[16:52:37, 16:52:37]
    *G63884 cssat[16:44:46, 16:44:46, 16:44:46, 16:53:42, 16:53:42]
    *G64039 edw192[16:52:37]
     G63073 edw192[16:52:37]
     G63888 cssat[16:53:42, 16:53:42, 16:53:42, 16:53:42, 16:53:42]

        10        cssat     Edward Ming,0262465899
        8         edw192    Peter Edwards,0386013899
        2         root      Root on Cherax
```

See slides 7 – 10 in http://hpsc.csiro.au/users/dmfug/Meeting_Oct2009/Presentations/load_sharing/load_sharing.pdf

dmsilo

add/remove tapes to a tape library in various ways

dmsilo provides a variety of tape movement features including:

- TMF equivalent to dmov_loadtapes, to add tapes to a library and include them in a DMF VG
- TMF equivalent to ov_eject
- Move tapes in and out of a full library
- Move tapes to and from an off-site DR facility (optionally including filesystem backups)

At CSIRO ASC, the first two are currently the most important to us.

Note: dmsilo belongs to SGI and is unsupported; use at your own risk!

dmsilo

very abbreviated “usage”

```
cherax# dmsilo
Usage:  dmsilo [-e] [-i] [-p] [-s] [-t] [-v] [-a AG] [-c configfile] \
               [-n max_ejects] \
               {config|export|import|inject|list|swap|offsite|onsite}
      dmsilo [-e] [-i] [-p] [-s] [-t] [-v] [-a AG] [-c configfile] \
               [-n max_ejects] \
               {eject|export|swap|offsite} VSNs...

      where
          eject    eject tapes, without making any database changes
          export   eject tapes and mark them in the VOL database as HOA
          import   request tapes to be placed in the library,
                   identify them and clear their HOA flags
          VSNs     is a case-insensitive space or comma delimited list
                   of VSN ranges.
                   Eg: aaaaaa BBBBBB,cccccc cccc10-cccc12,dd0000 - dd0003
```

dmv

a wrapper around dmvoladm for listing tapes and altering hflags

```
cherax# dmv G63500 hv G62641
        VG/AG or    dataleft      datawritten       ec    eotzone     wfage
VSN    LS:VG/AG       dl              dw      th  eotchunk  ez   hflags   wa

G62556  ls:se2  1299648.699975 1494284.678509 86%   35034 1972 ---v--- 61d
G62641* ls:se2  1368397.927211 1676666.637941 81%   19841  128 -----u- 267d
G63500  ls:T1         0.000000       0.000000 ---       1    1 --r---- 815d

cherax# dmv bad
        VG/AG or    dataleft      datawritten       ec    eotzone     wfage
VSN    LS:VG/AG       dl              dw      th  eotchunk  ez   hflags   wa

T00487  ls:te3        0.000000       0.000000 ---       1    1 e------ 163d
G62556  ls:se2  1299648.699975 1494284.678509 86%   35034 1972 ---v--- 61d

cherax# dmv hl:off T00064
VSN T00064 updated.
Updated 1 record.
```

find_bfids

show the BFIDs of file chunks on a DMF tape

```
cherax# find_bfids -z G62556
vsn=G62556 zn=1971
303c2a0000000000237cbb
303c2a0000000000237cbc
303c2a0000000000237cc5
303c2a0000000000237cc6
```

logw

a "tail -f" log watcher

```
cherax$  logw -s daemon
Monitoring /data/flush/dmf_spool/daemon/dmdlog.20121108
16:00:39:819-I   cherax   209955-dmfdaemon Req=1531370,reply=status,rp=Request completed
.(5 secs later)
16:00:44:400-I   cherax   209955-dmfdaemon Req=1531371,request=settag,tag=0,fhandle=01000
16:00:44:400-I   cherax   209955-dmfdaemon Req=1531371,reply=settag,rp=Request completed
...............(76 secs later)
16:02:00:312-I   cherax   209955-dmfdaemon Req=1531372,request=usage
16:02:00:312-V   cherax   209955-dmfdaemon Req=1531372, Report_Usage sent to MSP cache.
16:02:00:312-V   cherax   209955-dmfdaemon Req=1531372,7fe68c0d1320, received MspReq_Done
16:02:00:312-V   cherax   209955-dmfdaemon Req=1531372,MspReport_Usage complete
16:02:00:312-I   cherax   209955-dmfdaemon Req=1531372,7fe68c0d1320, 46136644255744 byte
16:02:00:313-V   cherax   209955-dmfdaemon check_libsrv_usage: new DMF-managed byte total
....(20 secs later)
16:02:20:533-V   cherax   209955-dmfdaemon do_GetReq_Done: Req=1531237,7fe69000da20, reca
16:02:20:533-V   cherax   209955-dmfdaemon recall_mspreply: Req=1531237,7fe69000da20, off
16:02:20:533-V   cherax   209955-dmfdaemon recall2: Req=1531237, region off = 65536, len
16:02:20:534-V   cherax   209955-dmfdaemon Req=1531237, reg 0: off=0, len=6570246827, st
16:02:20:534-I   cherax   209955-dmfdaemon Req=1531237,reply=krclrea,bfid=4fcef489000000
```

tpstat

a wrapper around tmstat, ov_stat, msgd & ps to show DMF activity

In a self-refreshing screen, using oper, tpstat can show edited output from:

- tmstat
- ov_stat
- msgd
- ps
  - TMF-related
  - OV-related
  - backup-related
  - DMF-related

tpstat

sample output from tmstat, ov_stat & msgd

```
cherax$ tpstat -mp
user  session  group   a!stat+device stm rl* ivsn evsn   blocks req-details
               T1B     - idle dd8
root   194260  T1B     - assn dd9   15 in G63884 G63884   55816 G te2 #1175
               T1B     - down dd13
root   206860  T1B     - assn dd14   8 ob*                      G se2 #1177
root   194261  T1B     - assn dd15  13 in G63888 G63888   80428 G te2 #1176

Drive   Group   Disabled  S-State H-State   Occupied   PCL
C00d00  dg_c00      -      inuse  unloaded    true   C00A0A
C00d01  dg_c00      -      inuse  loaded      true   C00A09
C03d05  dg_c03      -      inuse  unloaded      -

1176    17:21   TM046 - Mount volume G61106(blp) ring-out, on device
               dd14 for root (206860) [G se2 #1177]  or reply cancel /
               device name
```

tpstat

sample output from ps

```
209955     1 Oct28                01:48:54 dmfdaemon
210023     1 Oct28                08:09 dmlockmgr -a dmlockmgr -z /dmf/home/RDM_LM
13422 13419 15:15 cron           00:00 dmmove -r 10 -d none cache
45962 45959 15:17 cron           00:00 dmmove -r 10 -d none cache
52343 52342 16:04 pts/66  hua03r 00:00 dmusrcmd -L              (dmget)
53219 53216 16:04 pts/66  hua03r 00:00 dmusrcmd -L              (dmget)
93578 93576 15:19 cron           00:00 dmmove -r 10 -d none cache
152048 152047 15:02 cron           00:00 dmmove -r 10 -d none cache
154199 154198 16:10 56336   ngu038 00:00 dmusrcmd -L            (dmget)
167120 167114 16:11 56331   ngu038 00:00 dmusrcmd -L            (dmget)
171044 171040 14:20 cron           00:00 dmmove -r 10 -d none cache
195741 195738 16:12 56329   ngu038 00:00 dmusrcmd -L            (dmget)
210411 209955 Oct28                01:43 dmdskmsp cache
210412 209955 Oct28                11:43 dmatls maid_ls
210413 209955 Oct28                36:43 dmatls ls
210425 209955 Oct28                02:58 dmfsmon
```

Note process type (cron, batch, i/a), the username and for dmusrcmd, the command

Thank you

Advanced Scientific Computing
Peter Edwards
Systems Support Manager

e peter.edwards@csiro.au
w http://www.hpsc.csiro.au/