

HPSS 5.1 Conversion at ECMWF

Experiences converting from HPSS 4.5 to HPSS 5.1

Mike Connally

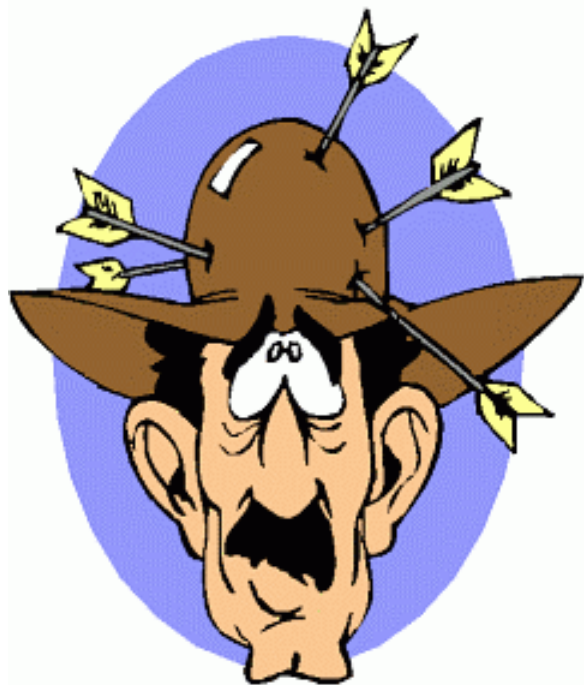
Mike.Connally@ecmwf.int

European Centre for Medium-Range Weather Forecasts

Reading, England

<http://www.ecmwf.int>

Pioneers take arrows



What's it all about?

- **Preparing for conversion**
 - From HPSS 4.5 SFS-managed metadata**
 - To HPSS 5.1 DB2-managed metadata**
- **Running the conversion**
 - 1.3M bitfiles, 3M nsubjects, 2M tapesegs, few disksegs**
- **Planning, practicing, pitfalls**
- **Not about code defects that are now resolved**

What we did

- **'Vanilla' mkhpss installation on test system (07/03)**
- **DB2 admin course**
- **Custom DB2 setup on production system (09/03)**
- **Trial conversions on production system (09/03-11/03)**
- **Build-up of test bed system (11/03)**
- **Repetitive conversions on test bed**
- **Live conversion on production system (11/03)**

'Vanilla' installation

- On Nighthawk high node with
 - 8 CPUs 4GB RAM**
 - 4 100GB SSA RAID-5 arrays for hpssdb**
- Nuisance aliases created by mkhpss
 - Databases HCFG, HSUBSYS1, etc.**
 - Database aliases CFG, SUBSYS1, etc.**
 - Instance HPSSDB**
 - Instance alias ATHOS5#E (hostname athos5-eth)**

DB2 Admin Course

- **Not a black box (and shouldn't be)**
- **DB2 UDB Admin Workshop (CF21U)**
- **DBA skills extremely useful**
- **SFS occult, arcane, opaque, hard to administer**
- **DB2 open, well-known, transparent, easy to administer**
- **DB2 discovery features: db2cc (Control Center GUI)**
- **DB2 performance monitoring, snapshots, db2expln**

Custom DB2 setup on production system

- On p660 6M1 node with
 - 6 CPUs 8GB RAM
 - FAStT700 disk
- Install & configure DB2 outside mkhpss
- Instance HPSSDB
- DBs CFG, SUBSYS1, SUBSYS2, SUBSYS3
- No aliases
- Auto-start & persistence via db2fmcd

DB2 disk setup

- **VG hpssdbh - 17GB on RAID-1 15K 34GB(1+1)**
 - /hpssdb_home (2GB)**
 - /hpssdb_logs (10GB)**
- **VG hpssdbd - 136GB on RAID-1 15K 34GB(1+1) x 4**
 - DMS tablespace containers (85GB)**
 - Each tablespace on 4 LVs across all drives**
- **VG hpssdbb - 120GB on RAID-5 10K 137GB(8+P)**
 - /hpssdb_backups (120GB)**

DB2 tablespace allocation

	CFG	SUBSYS1	SUBSYS2	SUBSYS3
SYSCATSPACE	512MB (4KB)	1GB (4KB)	1GB (4KB)	1GB (4KB)
TEMPSPACE1	1GB (4KB)	2GB (4KB)	4GB (4KB)	4GB (4KB)
USERSPACE1	1GB (8KB)	2GB (4KB)	6GB (4KB)	60GB (4KB)
TEMPSPACE2	1GB (8KB)			

Trial conversions on production system

- Weekends – 8 to 12 hour outages of HPSS
- Shutdown HPSS leaving SFS & DCE up
- Run conversions serially (60-90 mins)
- Don't do db_convert_dce_cds
- Run verifications in parallel

Full checks could take > 16 hours

Do partial checks to fit in time slot, e.g.:

```
db_convert_ns_check -f 50000 -l 50000
```

```
db_convert_address_check -e 1000 -i 35000
```

Test bed system

- **Four conversions on production system**
- **Still had problems in conversion/verification utilities**
- **Deadlines approaching**
- **Decided to build a dedicated test bed on another platform**
- **Ran conversions/verifications 24x7**
- **Got there!**
- **Useful too for full checks before and after final conversion**

Building test bed system

- SFS space for dd copies (100GB)
- DB2 space for everything (250GB)
- DCE, SFS, DB2 binaries and setup (Jae & Co)
- HPSS conversion utilities
- Bare-metal restore of DB2 from backups

Bare-metal restore of DB2

- **/etc/passwd & /etc/group**
- **Create & populate /hpssdb_home, /hpssdb_logs & /hpssdb_backups**
- **Create LVs for tablespace containers**
- **Sort out glitches with db2start, license, etc.**
 - /var/hpss/hpssdb/sqllib/db2nodes.cfg (hostname)**
 - db2licm (/usr/opt/db2_08_01/adm/db2ese.lic)**
 - /etc/services**
 - Java 1.3.1 for db2cc**
 - DBM CFG DIAGPATH /logfiles/hpssdb/diag**
 - DBM CFG SPM_NAME <hostname>**

Bare-metal restore of DB2

- **db2start**
- **db2 list history backup since <date> for <db>**
- **db2 restore db <db> from <dir> taken at <timestamp>**
- **db2 rollforward db <db> to end of logs and complete**
- **Entire bare-metal exercise took 2 hours start to finish**
Including all glitch resolution (thanks to Jae & Co)

Repetitive conversions on test bed

- Preserve dd clone of SFS database
- Convert until clean
- Verify until clean
- Observe timings
 - > 16 hours for full verification
- Repeat
- Re-clone SFS database before final conversion

Final conversion

- Re-clone SFS database to test bed
- Serial conversions on both systems
- Partial verifications on production system
- Full verifications on test bed
- Setup for running 5.1
 - db_convert_dce_cds & set ACLs (IG 7.1.7)
 - SSM ID conversion & config (IG 7.2.6)
 - HPSS.conf & inetd.conf (IG 7.2.6)

Precautions

- **Migrate all disk-tape hierarchies to tape (& purge)**
- **Active tapes in VV condition RWC – set to RO**
 - EOM is irreversible**
 - Put them back to RWC when confident**
- **Prevent repacks, reclaims, shelves**
 - No exec on repack, reclaim, shelf_tape, shelf_tape_util**
- **Prevent MPS running**
 - No exec on hpss_mps**

Gotchas

- Apart from code defects found & resolved ...
- DB2 Fix Pack broke db2start – needs db2iupdt
- Extended Shared Memory
- MINCOMMIT



Extended shared memory

- SQL1224N Database agent could not be started
- Maximum 10 shared memory segments per process
- db2set DB2ENVLIST=EXTSHM
- /var/hpss/hpssdb/sqllib/userprofile
EXTSHM=ON
export EXTSHM
- /opt/hpss/config/hpss_env
export EXTSHM=ON
- Gotcha! /etc/environment (for db2fmcd started by init)
EXTSHM=ON

MINCOMMIT

- Delay logging commits until MINCOMMIT or 1 sec
- Was set to 2 or 3 (by DB2 Config Wizard?)
- Caused significant delays in various areas
- Subsys3 core server startup time

Mount 72 disk volumes – very slowly

Up to 90 minutes with MINCOMMIT = 3

9 seconds with MINCOMMIT = 1

The Cavalry

- Jae Kerr, IBM
- Dennis Tran, IBM
- Jason Hick, LANL
- Jim Daveler, LLNL
- Dave Fisher, LLNL
- Mike Gleicher, Gleicher Enterprises



Other peacemakers

- **Benny Wilbanks, IBM**
- **Danny Cook, LANL**
- **Per Lysne, LANL**
- **Jim Minton, LLNL**
- **Deryl Steinert, ORNL**
- **Vicky White, ORNL**



HPSS 5.1 Conversion at ECMWF

Experiences converting from HPSS 4.5 to HPSS 5.1

Mike Connally

Mike.Connally@ecmwf.int

European Centre for Medium-Range Weather Forecasts

Reading, England

<http://www.ecmwf.int>