Memoirs of a HALDBA
The challenges of IMS
HALDB administration at TNT

Neil Price

Virtual IMS User Group
7 February 2012
Disclaimers

- Some knowledge of HALDB basics is assumed
- Conversion to HALDB is not specifically covered - see our 2005 presentation (details later)
- Any mention of specific IBM or ISV software products used at TNT does not constitute a recommendation
- The challenges and solutions presented may not be applicable to your environment and applications
- Some of the techniques adopted may have since been superseded by features in IMS or related software
- The logic behind some of the decisions made may be obscure, open to debate, or downright wrong!
Agenda

- Introduction
- Challenges we’ve avoided
- Our Challenges
  - in Design
  - running Utilities
  - with Monitoring & Tuning
  - related to Indexes
- Summary
- Questions?
Who are we?

- **TNT Express** [www.tnt.com](http://www.tnt.com)
  - Based in Amsterdam
  - 83,000 employees
  - 50 aircraft and 30,000 road vehicles
  - 2,635 depots in the network
  - Delivering to 200+ countries
  - €7 billion Revenue

- A company in its own right since May 2011
  - Formerly a division of TNT N.V.
Who are we?

Express ICS

- Dedicated provider of Information & Communication Systems and Services to TNT Express
- Based in Atherstone, Warwickshire, UK
- Around 600 employees in the UK
- 24x7 Service, year-round
- 20,000+ users in 900+ locations worldwide
Our Mainframe Hardware

- **z196 Model 607**
  - 4,624 MIPS and 2 zIIPs
  - Production service

- **z9 Model 405**
  - 895 MIPS and 2 zAAPs
  - Network, Dev & Test service
  - 1 small production LPAR

- **DS8000 DASD subsystem**
  - 18TB total capacity
  - Global Mirror to DR site in London
Global Link Application Suite

- International shipments applications
  - First implemented 1993-95 (TNT EW)
  - 11,000+ concurrent CICS users, 320 trans / sec
  - 4,500+ MIPS at peak (08:00 – 11:00 UK time)
  - Online down 90 minutes a week, rest quarterly.

- Software:
  - IMS 10 DBCTL
  - z/OS 1.11, CICS TS 3.2 (20 regions)
  - DB2 9, WebSphere MQ V7
  - Enterprise COBOL 3.3
  - PROIV for Screen Handler
  - WebSphere Application Server (WAS) 6.1
Skymaster

- Airport systems applications
  - Implemented 2001
  - 24x7, modest usage, peak 22:00 – 02:00 UK time
  - Uses LPARs on both machines

- Software
  - DB2 9 data sharing (2 members)
  - MQ V7 queue sharing
  - WebSphere Message Broker
  - z/OS, CICS and COBOL
    - as for Global Link
  - 1 CICS region per LPAR
  - No IMS, WAS or PROIV
TNT’s IMS Databases

- 163 Databases totalling 600GB addressable
  - 64 HDAM / OSAM
  - 90 HIDAM / OSAM
  - 19 PHDAM / OSAM

- 6 HIDAM Message Databases, 7.5 GB each
  - Part of home-grown middleware
  - Root-only
  - Key deliberately “randomised”

- 51 HIDAM Queues of various sizes
  - Root-only, Variable length, Timestamp key, PTR=NT

- Other HIDAM small and/or non-volatile
Application Partitioning - CCDB

Track & Trace
Central Consignment Data Base implemented as 15 IMS databases

14 Consignment (Con) Databases

Consignment (Con) Master Index

The Master Index database enables searching across multiple Consignment DBs

12 Secondary Indexes
HALDB at TNT

- First HDAM database converted in 2005
  - All converted for size reasons

- 2 key-range partitioned, 17 using exit
  - Mostly 6 partitions, previously 16

- All of CCDB converted

- Total of 28 PSINDEXes (on 5 HALDBs)
  - 13 indexes have multiple partitions
Current CCDB Configuration

All databases converted to HALDB

Con Master Index = 6 OSAM + 6 ILDS

Con Master Index

total datasets: 113
(6 OSAM + 107 KSDS)

14 Con Databases = 84 OSAM datasets + 84 ILDS

6 Parts each

12 PSINDEXes = 101 KSDS

1-20 Parts each
Challenges We’ve Avoided
Challenges we’ve avoided

- Always had DBRC in Development & Test and BMP-only access to a single copy of DBs
- Retained application-implemented partitioning
  - 14 “cloned” databases converted to 14 HALDBs
- Simple PSE (Partition Selection Exit) logic common to all exit-partitioned databases
- No Partition-wise operations
  - DB name kept as Master DB name
  - /DBR, /STA, Reorg etc. always on whole DB
  - Avoids application impact due to ‘BA’ status etc.
Challenges we’ve avoided (2)

- Logical relationships
  - We have 1 internal to a DB and 1 between 2 DBs
  - None of these have (yet) become HALDB

- Secondary dataset groups

- Application impact from /SX fields
  - One index affected, not (yet) converted to HALDB

- PHIDAM
  - Only candidates have timestamp-based keys
    - Potential locking problem
  - Conversion to DB2 probably a better option
Challenges we’ve avoided (3)

- Pointer healing
  - PSINDEXes are always rebuilt and ILDS are empty

- Duplicate ILKs and similar pitfalls
  - Most scenarios don’t apply

- IMS Reorgs of PSINDEX datasets
  - As for non-HALDB indexes
  - Many rebuilt regularly during full DB reorg
  - VSAM reorg used for the rest

- HALDB Online Reorg (OLR)
  - Not a significant advantage to us
Challenges in HALDB Design
A Design Challenge - PSE

- PSE = Partition Selection Exit routine
- Thought about splitting the existing RAA without changing the sequence of records
  - but discovered it wasn’t really feasible

HDAM RAA – DFSHDC40 Randomizer

PHDAM Partition RAA with DFSHDC40
Decided to use one-line mod to sample PSE (DFSPSE00) to get even spread

Documented in Chapter 7 of Redbook SG24-6945 ‘The Complete IMS HALDB Guide’

Easy to determine partition strings
- Divide max value ‘999’ by no. of partitions

More about the performance implications later
Choosing Key-Range Partition Limits

- PHDAM databases
  - Limits taken from DFSMAID0
  - Adjusted following performance analysis
  - More about this later as well

- Indexes
  - Records up to more than 3x previous size, e.g. 26 → 88 for Index key 20 & Root key 34
  - Early DBDGEN bug could add another 2!
  - Many PSINDEXes exceed 4 GB total size
  - All have to be key-range partitioned
  - DFSMAID0 works with indexes too!
PSINDEX Partition Sizing

- Not particularly critical in general
- SMS controls placement, not DBA
- Volatile KSDS grow rapidly towards 4 GB
- Must take account of reorg frequency
- Generally take action when size is > 2GB
- No. of partitions kept as low as possible
- Compromise between size and number
- Using small primary if volume fluctuates
Challenges with Utilities
Utilities and HALDB

- Before HALDB with DBRC=N we could:
  - Use or create a copy database
  - Unload from an online database
  - Run DFSMAID0 against the online database

- With HALDB it’s not so easy:
  - DBRC is required
  - Datasets must match RECONs
  - Unload and DFSMAID0 require exclusive use

- New Tricks are needed!
  - Features of vendor products?
  - Use a copy of the RECONs?
  - Register live DB datasets in another system?
Utility Challenges continued

- A feature without a HALDB equivalent
  - DBD can be changed in non-destructive reorg
  - HALDB RAA etc. cannot be changed – requires Unload / DBRC / Reload / Index Build
  - Can change PSINDEXes in previous step
  - Vendor(s) may have this in hand!
Re-partitioning Challenges

Limitations due to Partitioning Scheme

- With our PSE, changing no. of partitions requires complete DB Unload/Reload in general
- Parallel unloads save time
- Reload in groups sometimes possible – creates non-empty ILDS (temporarily)
- Currently no “discard” processing on Reload – but maybe someone will fix this soon?
  - Cf. DB2
Unload and Load Compatibility

- Require Input File for Test Data tool
  - Unload utility is the most efficient option
  - Using Load utility for target database
- HD Unload file format
  - Different for HALDB
  - Contains information about partitions unloaded
  - Not suitable for sub-selection and reload
  - Created user-format files instead
  - Reconverted the target system HALDBs
Monitoring & Tuning Challenges
The Monitoring Challenge

- Exceptions previously reported daily by email based on report from vendor-supplied program
- Former infrastructure did not support HALDB
  - Separate repository
  - No batch reporting
- SAS code post-processes Pointer Check reports from separate HALDB Image Copy jobs
- Thresholds implemented in the SAS code
  - Currently only checking free space
  - Difficult to add metrics to monitoring
- Since installed add-on vendor product
Monitoring Key-Range Partitions

- Partitions need individual monitoring
  - Should be treated like separate databases
- Data volumes likely to grow at different rates
  - Other measures may also vary
  - Might want different warning thresholds
- One database had wide variations in both growth rate and average database record length
Performance – DB Physical Sequence

- Keys from sequential scan used in calls to other databases
  - Keys obtained directly or from file
- DFSHDC40 keeps HDAM physical order
  - Target database(s) accessed in block sequence
  - Minimizes I/O, maximizes caching
  - May activate OSAM SB for multiple databases
- In PHDAM, same keys no longer guarantee same sequence
  - Order only matches with identical partitioning
Physical Sequence mismatch

PHDAM database 1 (3 Partitions)

HDAM

PHDAM database 2 (4 Partitions)
Performance Challenges – OSAM SB

- Extensive use of OSAM Sequential Buffering
  - Complete sequential scans
  - Sequential file derived from a DB scan
  - File pre-sorted in DB physical sequence
- Non-HALDB sequential detection by dataset
- Operates at DB (dataset group) level in HALDB
- Not like DSGs or application partitioning
- Access in HDAM sequence does not trigger SB
  - E.g. between conversions of Con & Master Index
Physical Sequence Sorting Challenge

- Many jobs sort files into Physical Sequence of target database
- Using DFS0ASRT for HDAM – E61 Exit from old IMS Primer system
- HALDB equivalent not available
- Wrote E61 for PHDAM sort
  - Started with DFS0ASRT source
  - Quite a complex task
  - Program available on request
Buffer Pool Assignments

- DBs explicitly assigned to Buffer Pools
- VSPEC DBD statement changes for HALDB
  - Dataset number becomes letter
  - but no change to DBD statement for index!
- Changed statement required after IMS restart
  - Subpool assignment is preserved across change to or from HALDB if IMS is not restarted
- Kept both versions in member across several restarts
- Might assign partitions across multiple pools
  - Not practicable prior to IMS 12
An Aside on DB Tuning

“Bucket” sizes and the (P)HDAM Home Block
• Different bucket sizes imply different max packing
• Otherwise I/O per DB record may vary

![Graph showing the relationship between %Packing and %Displacement for different bucket sizes.]

- 3 per block
- 5 per block
- 10 per block
- 20 per block
Key-range PHDAM Optimisation

- Initial partitioning was not optimal
  - Used DFSMAID0 with specific no. of partitions
  - More roots displaced than expected
  - Due to variation in average DB record length

- Rebalancing is complex
  - Run DFSMAID0 with revised limits
  - Calculate average record from output
  - Apply appropriate packing density limits
  - Adjust limits if necessary & try again
  - Use more limits at a time to reduce no. of runs
The Key-Range I/O Challenge

- I/O rates can vary widely between partitions
  - Can make the previous exercise academic!
  - HDAM gave an even spread across the RAA
  - Variations may not cause any problem nowadays

- Index I/O “Hot spots”
  - May show up when the index is partitioned
  - Requires monitoring of catalog statistics
    - but only if indexes are taken offline sometimes!
  - KSDS statistics provided by scheduled LISTCATs
PSINDEX Challenges
Another Aside – IMS Index Inserts

- KSDS Insert Mode
  - Normal inserts cause 50-50 splits
  - Not good for ascending keys
  - Could use INSERT=SEQ on VSPEC OPTIONS
  - FREESPACE=YES on DBD statement is better
  - Use FREESPACE(0 0) in cluster definition
  - Would also work with descending keys
  - Might be applicable to “pseudo-ascending” keys but difficult to assess.
Index Buffering Challenges

- PSINDEX records much larger than non-HALDB
  - E.g. 32 → 90 bytes for large & busy indexes
  - Fewer records per CI, more CIs and index levels
  - Increased pool sizes to compensate
  - Pro-rata increase prevented by 32K buffers limit and storage constraints

- Using FREESPACE=YES
  for both Primary and Secondary Indexes
  - Initially didn’t work with PSINDEX
    - Impact sufficient to delay first conversion
  - Some indexes reverted following analysis
A Special Index Partitioning Issue

- Shifting Index Keys, e.g. date-based
  - Needs regular partition limit changes
- Initially set up for annual review
  - Equally-spaced partition limits
  - Extra partitions to last between changes
- E.g. limits from May 2009 for 10 partitions:
  - 1\textsuperscript{st} partition limit 200902
  - No data initially beyond 3\textsuperscript{rd} partition
  - Limits 2 months apart
  - Dates from 201007 onwards went to last partition
- Reviewed in October 2010
  - last partition had exceeded 2.5GB
Shifting Keys (continued)

- Automated re-partitioning
  - Introduced REXX to adjust limits weekly
  - Uses fixed numbers of days from current
  - Halved the number of index partitions

- E.g. limits set on 29/10/2011:
  - 20110805 = 85 days previous
  - 20110930 = 29 days previous
  - 20111028 = 1 day previous
  - X'FF'

- “Day of year” indexes still have fixed limits
  - Small primary allocation specified
Index Partition VSAM Parameters

- Choose Freespace to minimise splits
- Insert rates may vary considerably between PSINDEX partitions
- Requires differing Freespace for optimal performance
- Delete rates might also be relevant
- May be able to isolate Hot Spots and improve performance
Summary

- You might be able to convert databases to HALDB without impacting the Programmer - but the DBA will definitely notice!
- Conversion is just the first challenge.
- Ongoing maintenance can be challenging
- Partitioning can bring tuning advantages - not just problems.
- "Your challenges may vary"!
Further information available from …

- IBM Redbook ‘The Complete IMS HALDB Guide’ SG24-6945
- BMC book ‘Best Practices for HALDB’ by Peter Armstrong
- ‘HALDB or not HALDB - that is the question...?’
  - Presentation describing our early experiences (2005)
  - On request from me (contact details later)
- “HALDB Administration” presentation by Steve Nathan, IBM
  - From www.gse.org.uk (IMS meeting June 2011) or me
- The IMS manuals!
- Various other presentations, articles & documents online
- IMS-L discussion list and archives
  - imslistserv.bmc.com/archives/ims-l.html
- Other users
- Your friendly vendor support staff
Thanks for listening!

Neil Price
neil.price@tnt.com
+44 1827 710702