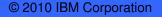# Best Operational Practices for IMS Availability

Kevin Stewart

IMS Technical Support

kevstew@us.ibm.com

Overview

- Operational Practices to prevent avoidable unplanned IMS service outages

    Parameters

    Exits

    Procedures

    Monitoring

    * simple and cost effective *

Outside of Scope

- ## Service/Maintenance Strategies
  - IBM recommended strategy and procedures covered elsewhere.

- ## Hardware Configuration
  - Very important, but here we will focus on operating what we already have.

- ## Subsystem Configurations
  - Datasharing, Shared Queues, etc.

- ## Planned Outages
  - Definitional changes, application, etc

Why?

- ## I was a customer once
  - Experience is something acquired **AFTER** its needed

- ## IBM L2 Experience
  - Outages that customers could have avoided
  - Outages that smart customers did avoid
  - Share the experience

Avoidable Outage Definition

- Outage that could be avoided, postponed, or shortened by operational action, parameter specification, or minimal customization

- Two major categories
  - Resource 'creep'
  - Runaway or errant applications

- For our purposes, software defects that crash the subsystem are for the most part, unavoidable
  - Use recommended maintenance strategy to reduce the chances of this happening

Agenda

- Storage growth

- Locking

- Message Queue problems

- Logging, DBRC, Restart

- z/OS Environmental

- Miscellaneous

Storage Growth

- ## Memory leaks
    - • Defects such that storage is allocated but never freed or never reused.
    - • IMS systems now routinely stay up for weeks or months.
        - • *Even a slow "leak" can cause big problems if not detected*
        - • *Not easy to detect in testing (IBM or Customer)*
    - • If detected early, unplanned outage can be avoided
        - • *Cycle subsystem at appropriate time*

Storage Growth

- ## Resource Utilization
    - ### Workload related storage
        - *Not a defect*
    - ### Most common is OTMA
        - *Legacy networks had limited number of terminals and users, implicitly limiting maximum number of IMS network related control blocks*
        - *The 'O' in OTMA is 'Open' – the client specifies resource names, usually with no direct relation to physical hardware*
    - ### Others
        - *Locks, etc*

Storage Growth

- Monitoring Storage

  - Vital to monitor virtual storage utilization of IMS Subsystem Address Spaces:
    CTL,DLS,DBRC,IRLM,CQS,SCI

  - Track PVT/LSQA, EPVT/ELSQA
    - ECSA/CSA if not done by z/OS team

  - OEM Monitors such as OMEGAMON/IMS, MainView, TMON
    - Real time alerts
      - *Set alert threshold slightly above observed normal peaks*
      - *NOT slightly below maximum available*
      - *Alerts useless unless acted upon*

Storage Growth

- Monitoring Storage

    - Configure RMF to collect data for IMS address spaces
        - ERBRMF00 in SYS1.PARMLIB
            - *VSTOR(IMSPCTL)*
            - *INTERVAL(15M)*
    - Run ERBRMFPP
            - *REPORTS(VSTOR)*
    - Use to configure OEM Monitor storage alert thresholds
    - Review "regularly"

Storage Growth

- Specifying IMS Storage Parameters
  - z/OS Region Size
    - No reason for arbitrary limit
      - *0M, or value close to maximum available*
      - *Allow for workload spikes and allow more time to deal with memory leaks*
      - *S40D + IPL risk not directly related to region specification*
  - IMS Pool Specifications
    - Pools allocated in CSA/ECSA
      - *Limit as appropriate to protect z/OS system*
    - Pools allocated in EPVT
      - *No reason for arbitrary limit, code no limit or largest possible value*
      - *Pools space obtain failures are generally not tolerated by IMS code and result in abends (U0732, etc)*
    - LRU Type Pools
      - *Some pools use LRU algorithms and may in some cases work better at smaller sizes, due to reduced overhead in pool management*
        - *CSAPSB,DLISPSB*

Storage Growth

- ## How to determine IMS Pool residency
  - IMS Macro DFSCBT00 in SDFSMAC contains definitions for DFSCBTS managed pools, including z/OS storage subpool
    - SP 231, 241 are Common Storage
  - DFSPOOL managed
    - As of V11 Common Pools are EMHB, EPCB,LUMC, CESS
      - *Specified in DFSSPMxx PROCLIB member*

OTMA and Storage Growth

- Synchronous (CM1)
    - Each instance of a message requires about 8K of IMS EPVT storage.
    - If arrival rate exceeds processing rate, queuing will cause IMS private area storage growth
    - Unlike legacy SNA processing, there is virtually no limit on the number of messages that may arrive and be queued
    - You will run out of IMS private storage before you run out of message queue space
    - IMS maps storage it allocates with small blocks of storage in LSQA and it is possible you will run out of low private before you run out of extended private. LSQA allocates are not subject to region size limits and are allocated top down. The result is usually an S40D and possibly an IPL to recover ECSA
    - For conversational transactions these control blocks remain for the life of the conversation
        - Some customers have MANY conversations are never terminated properly

OTMA and Storage Growth

- **Solution: OTMA Message Flood Protection**
    - Limit the number of YTIBs, which correspond to input messages
    - Per TMEMBER limit
        - In IMS V10, defaults to 5K per TMEMBER.
        - How many TMEMBERs (IMS Connects, or MQs) do you have?
        - Can be changed by command or set on OTMA Descriptor
            - */STA TMEMBER INPUT*
            - *DFSYDTx Client Desciptor INPT=*
    - IMS NAKs input with unique sense code when limit reached
        - MQ, but not IMS Connect, will suspend input
            - *ICON exit or CEX can suspend*
        - DFS1988W issued at 80% of limit and every 5% thereafter
        - DFS1989E issued at limit
        - Further input rejected until YTIB count drops to 50% of specified limit
    - Set the limit as *low* as practical for your environment
        - Messages may time out anyway if queued for a long time, depending on TCP/IP client logic – wasted processing and U0119 / U0711 abends
    - In periods of host (IMS) problems, consider turning off CM1 input rather than attempting to process it

OTMA and Storage Growth

- Asynchronous (CM0)
  - YTIB released when message queued, so no input queue issue
  - TPIPE name set by application client, often to arbitrary unique value
    - IMS Connector for JAVA (IC4J)
  - 1 or 2 YQABs per TPIPE, similar in size to YTIB
    - 2 if Client has HOLDQUEUE support (IMS Connect)
  - Storage problem if many unique TPIPE names
  - IMS TPIPE Cleanup at IMS System Checkpoint
    - No significant status (RESUME TPIPE AUTO)
    - No output queued to TPIPE
    - Shared Queues – TPIPE idle for 3 consecutive checkpoints
    - Not Synchronized MQ TPIPE (sequence numbers)
  - Synchronized TPIPEs are checkpointed and persist across IMS restarts
    - Only used by MQSeries CM0 messages
    - Cold-Start excepted
  - Non-synchronized TPIPEs are not restored during restart

OTMA and Storage Growth

- Solution: TPIPE Limit
    - Limit the number of TPIPEs
    - APAR PK79048/PTF UK57875 V10
        - Set on OTMA Descriptor
            - *DFSYDTx Client Desciptor MAXTP=*
        - No command support to alter online
        - Global limit is largest value for any member
        - TMEMBER Limit:
            - *DFS4382W at 80% DFS4384E at 100%, DFS4383I when back to 50%*
        - GLOBAL Limit
            - *DFS4385W at 80%, DFS4386I when back to 80%*
    - Limit depends on normal operation needs
        - Understand application use of TPIPEs and naming conventions
    - Recovering TPIPEs
        - /DEQ TMEMBER TPIPE + system checkpoint

Locking

- Protect IMS and innocent bystanders from aberrant application behavior

- Assist applications to resolve locking issues

Locking

- Use IRLM, not IMS PI
  - Capacity
    - PI: Locks in 31-bit ECSA
    - IRLM: Locks in 64-bit IRLM Private, CF structure
    - PI: Limit of 63 waiters for any lock
      - *Application U2478 abend*
      - *For FastPath systems, can cause U0113*
  - Capability
    - Deadlock Detection
      - *IRLM, Cyclical*
      - *PI, instant*
    - IRLM: Long Lock SMF Reporting
    - IRLM: LOCK Timeout (optional)
  - RAS
    - IRLM is separate address space
    - Much better diagnostic tools

Locking

- IRLM Long Lock SMF Reporting
  - IRLM reports via SMF 79.15 records when lock request has waited longer than specified LOCKTIME (IRLM TIMEOUT) interval
    - Record shows Blocker and Waiter(s)
    - Full Sysplex Datasharing Support
    - DXR162I issued when records produced
    - ERBSMF RMF macro maps SMF 79.15

- Useful in after-the-fact problem determination
  - Why did transaction XYZ run for 3 minutes at 4:00 am?
    - SMF 79.15 may have the answer

- RMF Monitor III can view data in real time
  - ILOCK ALL command

Locking

- LOCKMAX
    - Runaway application can acquire locks until:
        - Out of PI Lock Space
        - Out of IRLM Lock Space or CF Lock Structure Space
    - Innocent bystanders get hurt
        - U0778,U3300,U3307 abends
        - Contention and performance issues
        - Possible system failure
    - Limit number of locks a unit of work can acquire via LOCKMAX
        - U3301 abend if exceeded
    - Specify in PSBGEN, or (override) in dependent region JCL
        - No System Default -> some users code usermod to DFSRRA40
        - IMS Commit (x'37',x'5937'x'41') log records contain highwater mark lock obtained count.
    - Recommendation
        - Specify in JCL, distinguish between MPP,IFP and BMP

Message Queue Issues

- ## Defend the Message Queues
  - Runaway / Looping ISRTs
  - High input message arrival rates

Message Queue Issues

- ## Best solution may be OEM Tool
  - Very capable and flexible products on the market
  - Not inexpensive

- ## Simple and "free" options
  - SEGNO on TRANSACTION
  - DFSQSPC0 | DFSQSSP0 user exits

Message Queue Issues

- ## SEGNO
    - Option on TRANSACT macro
        - UPD TRAN ttt SET(SEGNO(x))
        - /ASS TRAN ttt SEGNO x
        - Default is 0 (No limit)
    - ISRT above SEGNO limit receives 'A7' status code
    - Only applicable for Message Driven processing
    - Specify for all transactions
        - 100 should more than cover most situations
        - Exceptions – application dependent

Message Queue Issues

- ## Queue Space Notification Exits
  - DFSQSPC0 (standard IMS)
  - DFSQSSP0 (Shared Queues)

- ## Capability
  - Warning
    - Issue message at threshold utilization
      - *QTL, QTH*
  - Action
    - Prevent ISRTs when appropriate

- ## IBM Samples Provided

Message Queue Issues

- ## Suggested implementation
  - Prevent ISRT ('A7') when number of queue buffers in use by application exceeds threshold value
    - Entry for Application Assign DRRN call
    - Included in IBM sample, but commented out
      - *QSPC0250 label*
      - *Sample threshold value is 5 buffers – too low?*
    - Works for BMP
      - *May need coding to exempt cases where application design is to insert many messages*

Logging, DBRC, Restart

- ## IMS System Restart

- DFS3804I LATEST RESTART CHKPT: yyddd/hhmmss

- Restart checkpoint cannot move beyond oldest UR in system

- Can result in extended restart time if IMS abends and has to mount and read many SLDS datasets
  - Automate on DFS3804I message
  - Verify that checkpoint time in message is changing (moving forward) within several consecutive system checkpoints
  - Verify checkpoint time in message vs current time

- No easy way at present to identify which dependent region is holding up the restart checkpoint
  - Can use x'47' checkpoint record field CAPUORTM
    - Mapped by DFSLOG47

Logging, DBRC, Restart

- # BMP XRST
  - SLDSREAD simplifies BMP restart
  - CHKPT=LAST works even if OLDS containing checkpoint has been overwritten
  - No requirement to manually determine SLDS dataset and code IMSLOGR DD statement
  - Restriction: BMP must restart on the same IMS where it originally abended.
    - This IMS knows the location of the x'18' log records

Logging, DBRC, Restart

- ## Reserve OLDS
  - Archive problems / Unexpected log volume
  - Define additional DFSMDA Members for extra OLDS Dataset pairs
  - Pre-Allocate at least one (large) extra OLDS dataset pair
  - Automate on DFS3258A LAST ONLINE LOG DATA SET IS BEING USED
    - /STA the pre-allocated extra OLDS
    - Alert support

Logging, DBRC, Restart

- ## Dual WADS vs Single WADs
  - DUAL protects against read errors during restart not write errors during normal operation
  - Spare WADS allocated and available
    - Otherwise degrade to no WADS
  - Have several spare WADS datasets available
    - Allocated and DFSMDA created and specified in WADSDEF

Logging, DBRC, Restart

- ## RECON Dataset Placement and Sizing
  - Follow System Administration Guide
    - Each RECON defined in own ICF catalog
    - RECON and catalog on same volume
    - Only dataset defined in the ICF catalog
    - Exercise caution with other datasets on volume
  - Allocate different sizes for each RECON, with initial SPARE being the largest

Logging, DBRC, Restart

- ● **DBRC Automatic RECON Loss Notification**
    - Most useful when a lot of DL/I Batch or batch utility processing
        - All DBRC instances immediately aware of RECON switch
    - Use Dynamic Allocation for RECONs in all cases
    - Requires SCI, CSL setup
        - CHANGE.RECON IMSPLEX
        - Use DSPSCIX0 to specify IMSPLEX name
            - *No JCL changes*

Logging, DBRC, Restart

- ## RECON Cleanup
    - IMS automatically compresses PRILOG
        - *16MB maximum size*
        - *DSP1150I issued when compression fails (Archive)*
        - *DSP0287W when SIZEALERT or LOGALERT threshold in RECON reached*
    - DELETE.LOG INACTIVE
    - V11 – CLEANUP.RECON

z/OS Environmental Considerations

- ## S322 | S522 (CPU Time, Wait Time) abends
  - Asynchronous
  - Will cause U0113 abend if they occur in dependent region DL/I call path

- ## Z/OS IEFUTL Exit
  - Determine if target address space is an IMS dependent region
    - Extend CPU limit, or Extend Wait Time
    - Issue informational WTO
  - IMS Sample DFSUTL in DFSSAMP

z/OS Environmental Considerations

- ## JES Initiator Storage and IMS EOM
  - IMS "Phantom Region" problems
  - Storage problems may occur in dependent region for various reasons, including application errors
  - IMS EOT processing may not be able to run to notify Control Region of failure, resulting in some or all of the dependent region control blocks remaining.
  - EOM processing runs in Master ASID and would recover this situation
    - Never runs in JES batch job case, as initiator address space does not terminate
  - JES2
    - Use Subsystem Interface Job Selection Exit (Exit 32) to specify that initiator is to end (drain) or end and restart, when job terminates.
      - *Bits 4 and 5 in Response Byte from Exit*
  - JES3
    - specify either DYNAMIC or DEMAND for the UNALOPT option of the EXRESC parameter in initiator's GROUP definition

z/OS Environmental Considerations

- ## PS EDI
    - z/OS function to prevent multiple open-for-output of PS datasets allocated DISP=SHR
    - IFGPSEDI member in SYS1.PARMLIB
    - IMS disables for OSAM and manages for OLDS, WADS in XRF, FDBR environments
    - Consider implementation
        - *MODE(WARN) | MODE(ENFORCE)*
        - *Can exclude by DSN*
    - Overwritten OLDS can happen – wrong DFSMDA member, etc.

Miscellaneous

- ## DB2 Resource Unavailable / Timeout
  - IMS handles as deadlock – U0777 and reschedule message if IFP, MPP
  - Often with identical results
  - Often for prolonged periods
  - DSNT376I, DSNT375I, DSNT501I issued by DB2
    - Contain CONNECTION ID – can automate if ID is IMS
    - DFSNDMX0 user exit gets control for U0777.
      - *Can send DFS555I to input device and discard message*

Miscellaneous

- ## APPC Timeout
  - ### APPC/IMS does not provide Message Flood protection for synchronous APPC
    - Same issues as OTMA Synchronous, 1 TIB per message instance
  - APPCIOT=(x,y)
    - X = timeout value for APPC verb (waiting in APPC)
    - Y= timeout value in minutes for waiting message
    - If Y is specified, IMS can cleanup TIBs and release storage for messages in a long wait to be processed. Client receives a DEALLOCATE_ABEND. Message not deleted, will process asychronously.

Miscellaneous

- ## Dynamic Allocation for ACBLIB
    - New V11 Function
    - Can now scratch and reallocate inactive ACBLIB without IMS outage
        - Resize
        - Relocate

Miscellaneous

- ## OTMA ACK Timeout
  - New V10 Function
  - Send-then-Commit (CM1)
    - Application response is sent before syncpoint
      - *For SYNCLVL SYNCPOINT or CONFIRM, dependent region will wait for end client to respond*
      - *Locks held*
  - Timeout specified on OTMA Client Descriptor
    - Default is 120 seconds
      - *Very large ?*
  - IMS V11 has ACK Timeout for Commit-then-Send (CM0) messages

Miscellaneous

- **Transaction Expiration**
  - New V11 Function
    - Specify EXPRTIME on TRANSACT macro
    - UPD TRAN / CREATE TRAN
    - OTMA client can set in OTMA prefix
      - *Per message*
  - Checked at IOPCB GU
    - Abend U2043 if expired
  - OTMA also checks at
    - XCF receipt
    - Message Enqueue
    - NAK if Expired

Miscellaneous

- ## Shared Queues Local First Optimization
  - Reduce SQ Overhead and False Scheduling
  - Only applies to Input Messages, not Message Switches
    - Local First if
      - *Not REMOTE, not SERIAL, PSB not SERIAL (SSPM)*
      - *Single Buffer Message*
      - *Not Priority Zero*
      - *Not already scheduled on arrival IMS*
        - *Unless WFI region available (includes PWFI)*
  - For SQ systems, consider PWFI=Y
    - Increase Local First Optimization
    - Beware of Serial Reusability
      - *No different than PROCLIM > 1 though*

Kevin Stewart

IBM

kevstew@us.ibm.com