Session 17423

Common z/OS Problems You Can Avoid

SHARE in Orlando

August 12th, 2015

Patty Little John Shebey
IBM Poughkeepsie
plittle@us.ibm.com jshebey@us.ibm.com

Trademarks

- **The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.**
  - **MVS**
  - **OS/390®**
  - **z/Architecture®**
  - **z/OS®**

* Corporation Registered trademarks of IBM

Contents

- SADMP Best Practices
- Runtime Diagnostics: RTD
- Dealing with a hung job
- zFS Large Directories
- Dynamic I/O Activation
- ERASE-on-Scratch
- SMSVSAM Buffers and Caches
- HMC Change Logical Partition Controls
- SDUMP RAS Enhancements

z/OS Best Practices: SADMP

Problem: Unprepared for Stand-Alone Dumps

- No one wants to take a SAD, but when you do, be prepared
- z/OS Best Practices: Large Stand-Alone Dump Handling - Version 4

http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TD103286

z/OS Best Practices: SADMP

- **<u>What is discussed</u>**:
  - Stand-Alone dump data set definition and placement
  - IPCS performance considerations
  - Preparing documentation to be analyzed
  - Sending documentation to IBM support
  - Testing your Stand-Alone dump setup

z/OS Best Practices: SADMP

What is discussed:

- PDUU, COPYDUMP, AMDSADDD, compression, encryption, transmission, etc
- Most recently, in z/OS V1.13, the Problem Documentation Upload Utility allows you to transmit large files much more quickly and encrypts the data all in the same process.

See *MVS Diagnosis: Tools and Service Aids (GA22-7589)*, Chapter 4

z/OS Best Practices: SADMP

New in V4 of the Best Practices:

- Use of INITAPPEND option of COPYDUMP in z/OS V2.1 to append initialization dump directory to dump dataset
- This will shorten the process of dump initialization by the IBM Support Center
- Recommended for full SAD only (not for subset dump)

z/OS Best Practices: SADMP

Want to practice? (highly recommended)

- Take a Standalone-Alone dump
- Open a Software Usage PMR using the component of AMDSADMP
- FTP the dump to IBM using PDUU
- Ask the support center to check the dump to make sure it is good

z/OS Best Practices: SADMP

Other considerations

- Problem with using PDUU to FTP SAD?
  - Check for any networking issues in your system
- FTP initially anything that is smaller (and therefore faster)
  - Subset dump
  - Any SVC dumps taken prior to the failure? (check for IEA794I or IEA611E)
  - They may show the problem!

Runtime Diagnostics: RTD

- **Problem:**
  - When z/OS runs into problems such as hangs, loops, delays, missed SLAs, numerous error messages, ENQ contention, it is often difficult for sysprog to determine the source(s) of problem

Runtime Diagnostics: RTD

- **Solution**: Runtime Diagnostics is a diagnostic tool to run in your environment when your system experiences symptoms that require its use. Run it when you experience system degradation or if you want to check for potential problems; do **not** run when the system is operating normally.

- After you start Runtime Diagnostics (S HZR,SUB=MSTR), you can analyze the system by entering the following MODIFY command.
- F HZR,ANALYZE

Runtime Diagnostics: RTD

Runtime Diagnostics does many of the same tasks you might typically do when looking for a failure, such as:

- Reviewing critical messages in the log
- Examining address spaces with high processor usage
- Looking for an address space that might be in a loop
- Evaluating local lock conditions
- Analyzing various types of contention that include ENQ, GRS latch contention, and z/OS UNIX file system latch contention.

Runtime Diagnostics: RTD

Runtime Diagnostics was enhanced at z/OS V2R2!

- JES2 Health Exception
  - Utilizes JES2 SSI call to obtain health monitor information
    - Surfaces critical JES2MON messages and suggested corrective actions
  - **Shipped in V2R1 via SPE APAR OA46531 (PTF UA76362)**

- Server Health Event
  - Exploits IWM4QHLT and returns address spaces with health value < 100.

Runtime Diagnostics: RTD

```
HZR0200I RUNTIME DIAGNOSTICS RESULT 568
SUMMARY: SUCCESS
REQ: 003 TARGET SYSTEM: SY1 HOME: SY1 2010/12/21 - 13:45:49
INTERVAL: 60 MINUTES
EVENTS:
FOUND: 02 - PRIORITIES: HIGH:02 MED:00 LOW:00
TYPES: HIGHCPU:01
TYPES: LOCK:01
------------------------------------------------------------------
EVENT 01: HIGH - HIGHCPU - SYSTEM: SY1 2010/12/21 - 13:45:50
ASID CPU RATE:99% ASID:002E JOBNAME:IBMUSERX
STEPNAME:STEP1 PROCSTEP: JOBID:JOB00045 USERID:IBMUSER
JOBSTART:2010/12/21 - 11:22:51
ERROR: ADDRESS SPACE USING EXCESSIVE CPU TIME. IT MIGHT BE LOOPING.
ACTION: USE YOUR SOFTWARE MONITORS TO INVESTIGATE THE ASID.
------------------------------------------------------------------
EVENT 02: HIGH - LOCK - SYSTEM: SY1 2010/12/21 - 13:45:50
HIGH LOCAL LOCK SUSPENSION RATE - ASID:000A JOBNAME:WLM
STEPNAME:WLM PROCSTEP:IEFPROC JOBID:++++++++ USERID:++++++++
JOBSTART:2010/12/21 - 11:15:08
ERROR: ADDRESS SPACE HAS HIGH LOCAL LOCK SUSPENSION RATE.
ACTION: USE YOUR SOFTWARE MONITORS TO INVESTIGATE THE ASID.
```

Dealing with a hung job

Problem: what to do if a job is hung?

- Take a dump of the job first (else we would not know why)
- If needed, use Cancel/Force to terminate the job
  - be prepared for an IPL if FORCE is used
- Only use this with the guidance of IBM support:
  - There is a new way in z/OS V2.1 to terminate a TCB in an address space (if this TCB is related to the hang)
  - *See **FORCE** command with new **TCB*** parameter in z/OS V2.1 System Commands
  - Some address spaces do not support the FORCE TCB command

zFS Large Directories

Problem: Poor performance when accessing zFS version 1.4 aggregate (file system) with large directory

- Slow response time or elongated job run time when reading or writing to zFS aggregate
- High CPU in user or system address space (running PRIMARY in zFS) when reading or writing to zFS aggregate

zFS Large Directories

Environment: Following conditions together can lead to problem

- zFS version 1.4 aggregate (v4 directory)
  - Fast Lookup Cache (FLC) buffers not implemented
- Directory being accessed in zFS file system has large number of objects (hundreds of thousands or millions) or is large in size
  - 10,000+ objects in directory

    - or -

  - Size of directory >= 1024K (1 MB)

zFS Large Directories

Symptoms: Check for evidence of v4 directory size problem

- MODIFY ZFS,QUERY,KN operator command or z/OS UNIX **zfsadm query -knpfs** command
  - Look at **Avg Time** field on lines for operations that require zFS to search through directory objects (i.e., zfs_lookup or zfs_create)
    - If times are larger than a few milliseconds (ten to a hundred times larger), could be dealing with large directory problem

zFS Large Directories

Symptoms (continued):

- **largedir.pl** command (perl script) can be downloaded from z/OS UNIX Tools & Toys (http://www.ibm.com/systems/z/os/zos/features/unix/bpxa1ty2.html)
  - Uses **find** command to determine large directories
    - Directories over 1 MB in size

- **NOTE: Space not reclaimed when objects removed from v4 directory, so must look at size, rather than current number of objects**

zFS Large Directories

Solution: At z/OS 2.1, can convert zFS file system to version 1.5 aggregate (v5 directory)

- Systems at earlier z/OS release levels cannot access version 1.5 aggregate
  - **Only use version 1.5 if all systems that access aggregate are running z/OS V2R1 (or higher release)**
- If unable to convert to version 1.5 aggregate, remove large directory, and try to spread objects across multiple directories

zFS Large Directories

Solution (continued):

- Otherwise, performance may be improved for version 1.4 aggregate by creating Fast Lookup Cache (FLC)
  - IOEPRMxx configuration option ‘flc’
  - Implemented via zFS APAR OA40530
    - Valid for z/OS V1R13 and z/OS V2R1

- Refer to **z/OS V2R1.0 Distributed File Service zFS Administration** manual for more details about this issue and its mitigation

zFS Large Directories

Problem: zFS dynamic grow delays with similar symptoms to zFS large directory

- Delayed response to users/applications accessing zFS aggregates.
  - Correlated with various zFS aggregate dynamic grow messages (i.e., IOEZ00312I)
- NFS server delays (GFSA1030E and GFSA1033E)
- Problem most predominant with very large zFS
  - 10's or 100's of GB in size

Solution:

- **HIPER APAR OA46665**

Dynamic I/O Activation

Problem: The system does not accept an IODF during dynamic I/O configuration in S/W-only or in S/W and H/W mode

- From console:
  - **IOS500I ACTIVATE RESULTS**

    ACTIVATE FAILED - ERROR MESSAGE(S) ISSUED

    REASON=A816,NO CONFIGURATION CHANGES ARE ALLOWED AT ALL. THE CURRENTLY ACTIVE I/O DEFINITION DOES NOT MATCH IODF IODF.IODFxx.
- From HCD:
  - **CBDA816I No configuration changes are allowed. Active I/O definition does not match IODF *dsname*. IODF token: *iodf_token date_1 time_1* active token: *active_token date_2 time_2***

Dynamic I/O Activation

Background:

- Dynamic activate can be invoked from HCD panels or via ACTIVATE command on console
- Tokens for currently active IODF must match between storage and DASD
- Token in storage reflects most recent IPL or ACTIVATE

Dynamic I/O Activation

Solution:

- If DASD backup copy of IODF is available that has same token as in storage copy
  - Restore this copy to DASD and retry activate
- If backup copy of IODF is not available, then IPL will be required.
- See INFO APAR **II13721** for more details

ERASE-on-Scratch

- **Problem:**
  - Certain jobs started exhibiting the following symptoms
    - Experiencing much greater elapsed time
    - Taking much more CPU
    - Experiencing many, many times more EXCPs (I/O operations)

ERASE-on-Scratch

- **Solution:**
  - Problem jobs may be using data sets defined with the Erase-on-Scratch option
  - You can deactivate Erase-on-Scratch processing via the NOERASE operand on the SETROPTS command.

ERASE-on-Scratch

- **Background:**
  - Erase-on-scratch processing erases the content of:
    - Any deleted data sets
    - Any scratched or released DASD extents
  - In order for it to be performed against a data set:
    - Erase-on-Scratch processing must be activated via the ERASE operand on the SETROPTS command
      AND
    - The ERASE option must be specified in the data set profile (through use of the ADDSD or ALTDSD command)

ERASE-on-Scratch

- **RACF ERASE sub-operands** for installation override of user specifications
  - **SECLEVEL** allows you to specify a security level at which all data sets at this security level or higher are always erased, regardless of the erase indicator in the profile.
  - **NOSECLEVEL** specifies that RACF is not to use the security level in the data set profile when it decides whether data management is to erase a scratched data set.
  - **ALL** specifies that *all* data sets (including temporary data sets) are always erased, regardless of the erase indicator in the data set profile. *Installation exits cannot override!*

For more info, see z/OS Security Server RACF Security Administrator's Guide
- Erasing Scratched or Released Data (ERASE Option)
- Erasing of Scratched (Deleted) DASD Data Sets

SMSVSAM Buffers and Caches

Problem:

Jobs accessing SMSVSAM datasets may experience

- Elongated response times
- CPU overhead

SMSVSAM Buffers and Caches

Background:

- SMSVSAM uses buffers (in private storage) and cache structures (in coupling facility).
  - Data set access uses buffers locally and keeps cache copy up to date, allowing other systems to access that data without doing I/O to DASD.

- If buffer sizes too small, they fill up
  - More requests need to go to DASD to retrieve data, elongating response times
  - Least Recently Used (LRU) routines need to run frequently to clean up buffers, causing increased CPU in SMSVSAM

SMSVSAM Buffers and Caches

Background:

- If cache sizes too small, buffers will become invalidated before the system is done using them, causing the next request to have to be re-read from DASD.
  - Can drive up response times
  - Adds CPU overhead

SMSVSAM Buffers and Caches

How to Check:

Buffers:

- To determine whether buffer sizes are too low, check SMF 42, subtype 19 reports
  - Watch Buffer Mgmt Facility (BMF) Hit Percentage, Usage vs Goal, and number of times where normal algorithms bypassed to reclaim buffers.

Caches:

- To determine whether cache sizes are too low, check SMF 42, subtype 15 reports
  - Check for rising False Invalid rate
- RMF Monitor III also reports this on RLSSC panel

SMSVSAM Buffers and Caches

Solution:

Buffers**:**

- IGDSMSxx in parmlib
  - **RLS_MAX_POOL_SIZE**
    - Set by first system that comes up (plex-wide setting)
    - SETSMS cmd to change dynamically
  - **RLSABOVETHEBARMAXPOOLSIZE**
    - Can be different for every system
    - SETSMS cmd to change dynamically

Caches:

- **Redefine** SMSVSAM cache structure in CFRM polity
  - Does not use AUTO ALTER
  - Stays at INIT size
- **Rebuild**

SMSVSAM Buffers and Caches

For more details, please attend session:

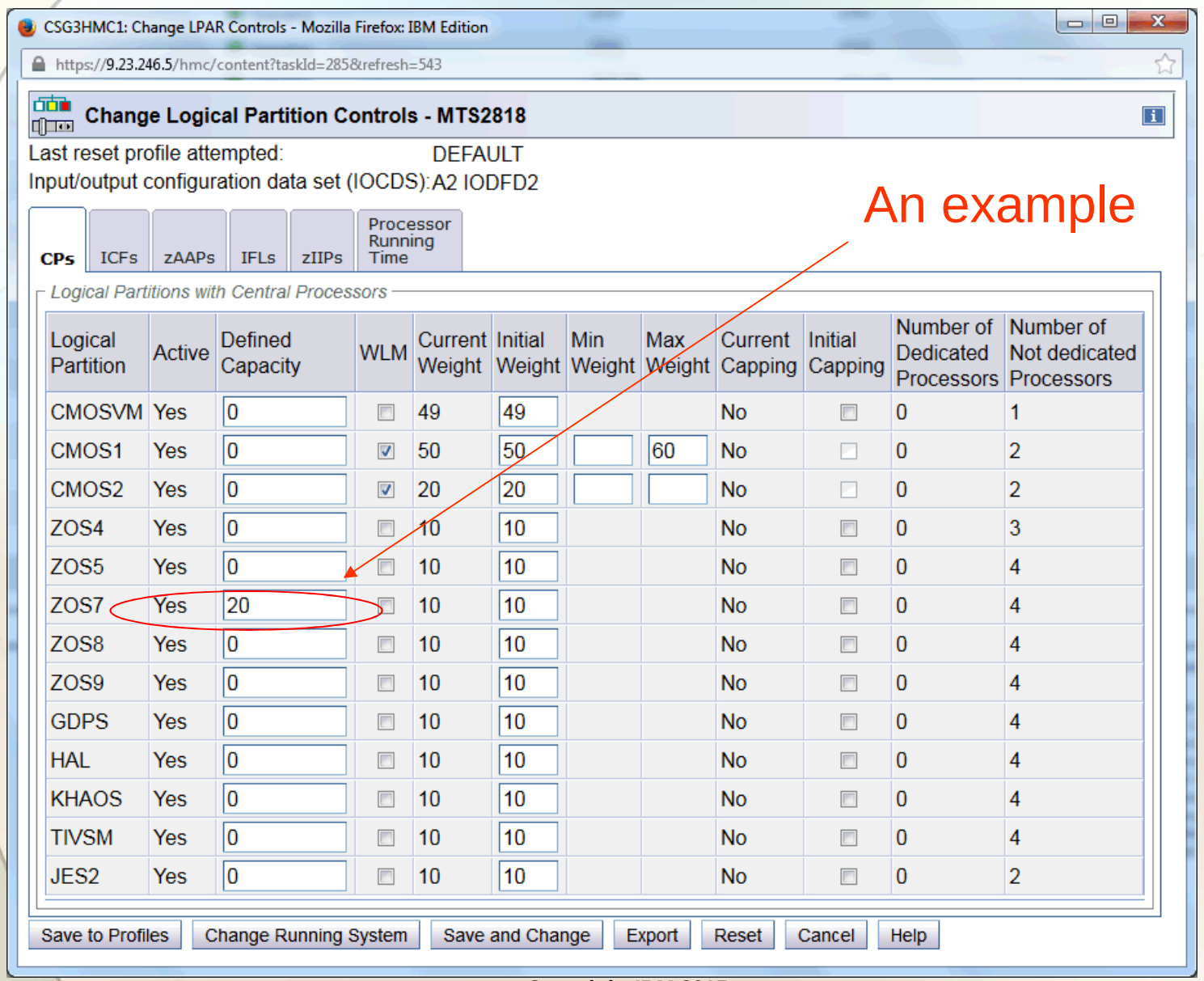
17832: Preparing SMSVSAM for HSM and Catalog - VSAM RLS Performance and Tuning

Friday, Aug 14 – 10:00 am

HMC – Change Logical Partition Controls

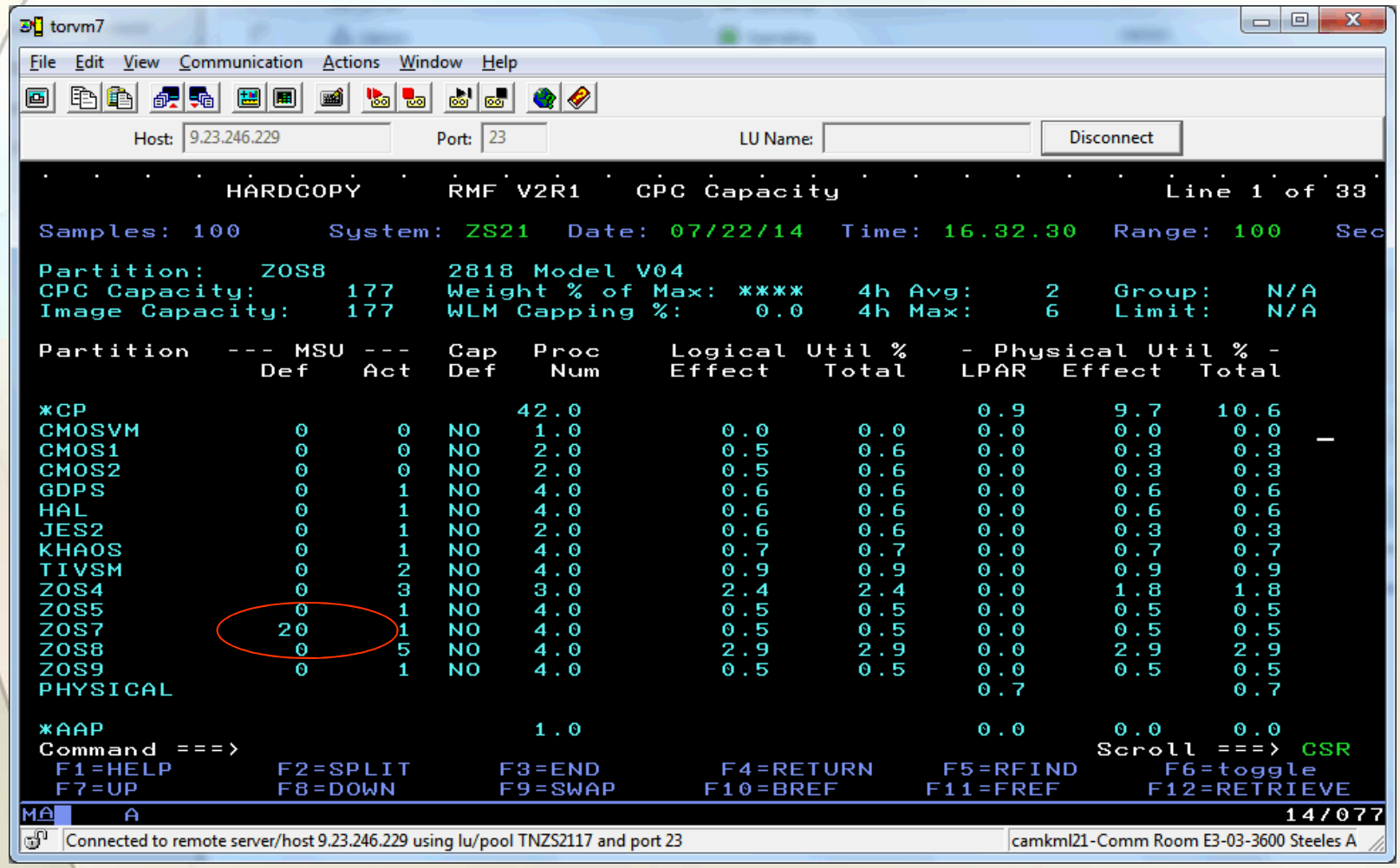
- **Problem:**
  - Incorrect usage of ‘Defined Capacity’ in HMC ‘Change Logical Partition Controls’ panel
  - ‘Defined Capacity’ on this panel does not mean number of CPUs or LPAR weight
  - Instead it is the MSU soft-cap 4-hour rolling average limit of the LPAR
  - A small unintended value in Defined Capacity will most likely result in a performance issue

HMC – Change Logical Partition Controls



HMC – Change Logical Partition Controls

RMF III CPC Report



SDUMP RAS Enhancements

- **Problem:**
  - Quiz: Name one component for which you cannot take an SVC dump when it has a problem.
  - Answer: SDUMP!

SDUMP RAS Enhancements

- **Problem:**
  - An SVC dump plays a **critical role in diagnosing** a wide variety of problems on the operating system.
  - When SDUMP processing hangs or abends, the result is often an **unusable dump**.
  - When SDUMP performs slowly or hangs, this can result in **system performance impact**.
  - ➢ Because we cannot take a dump of SDUMP processing, diagnosis of a problem in SDUMP is time-consuming and often unsuccessful, leaving the system exposed to both a recurrence of the SDUMP issue as well as a recurrence of the issue for which the SVC dump was being taken.

SDUMP RAS Enhancements

- **Solution**: We are delighted to announce the arrival of our new SDUMP CTRACE!
  - Available in base z/OS V2R2 OR

    at z/OS V2R1 via APAR OA45912 (PTF UA90790)

- Features:
  - Traces all SDUMP events as the default **(**recommended**)**
  - Captured above the bar so it does not consume valuable below the bar resource
  - Included in console dump of DUMPSRV (SDUMP addr space)
  - Useful for debugging hangs, abends, and performance issues within SDUMP processing
  - Contains at least 2 SVC dump's worth of history (current dump and previous dump)

SDUMP RAS Enhancements

- **Details**:
  - Parmlib member CTIDMP00 shipped with recommended defaults:

```
TRACEOPTS
   ON
   OPTIONS('ALL')  /* trace all SDUMP events */
   BUFSIZE(4M)    /* can be as large as 32M */
```

  - If this meets your installation's needs, you don't need to do anything but IPL!
  - Activated during IPL at DUMPSRV start up
  - Writes data to storage buffers in Highvirt Common

SDUMP RAS Enhancements

- **Details**:
  - If CTIDMP00 settings do not meet your installation's needs, you can set up your own CTIDMPxx parmlib member
    - See MVS Initialization and Tuning Reference, chapter on component trace parameters (CTncccxx), for component trace options including writing to an external writer.
    - Activate CTIDMPxx member via a command in COMMNDxx:

      COM='TRACE CT,ON,COMP=SYSDUMP,PARM=CTIDMPxx'

SDUMP RAS Enhancements

- **<u>But wait, there's more!</u>**
  - In addition to this amazing SDUMP CTRACE, we have thrown in a SUMDUMP buffer that is now nearly 2Gig, (yes, 2Gig!) in size, perfect for all your DB2 dumping needs!
    - Should significantly reduce the number of SVC dumps that are partial due to the summary dump buffer being filled
    - Exploited implicitly by DB2, as well as other components; you don't need to do a thing