Overview: Current SMF Data Flow

- Program requests to write a SMF record
- Locates appropriate buffer in SMF A.S. to write the record
- When ready to write, writes full buffers to the SMF dataset

Note: Each buffer is numbered to correspond to a particular record in the SMF dataset
Replace SYS1.MAN Datasets with Logstreams

SMF Data Flow Using Log Streams

- Program requests to write a SMF record
- SMF locates correct dataset
- Locates appropriate buffer to write the record
- If full, buffer passed to task to be written to logstream
### Usage and Invocation

- Define new logstreams in system logger
- Defining new keywords in SMFPRMxx
  - LSNAME(IFASMF.q1.q2,TYPE(xx:yy))
  - DEFAULTLSNAME(IFASMF.q1.q2)
  - RECORDING(DATASET|LOGSTREAM)
    - SETSMF operator command can be used to toggle recording settings
- Creating new JCL to use IFASMFDL with new logstreams
- Update processes to use data from logstreams, if necessary
- Activate PARMLIB changes via IPL or SET SMF=xx command

### SMF Processing

- Relative data processing in IFASMFDL intended to mirror typical GDG processing
- RELATIVEDATE keyword
  - Specify DAILY, WEEKLY, or MONTHLY range and number of units
- IFASMFDL LSNAME OPTIONS to dump and/or delete data from logstream (vs. waiting for retention period to expire)
  - DUMP
  - DELETE
  - ARCHIVE (DUMP and DELETE)
- SMFPRMxx MAXDORM applies to SMF log streams (in addition to dataset recording)
Usage and Invocation

- The support for ARCHIVE, DELETE and RELATIVEDATE is invoked by the IFASMFDL program. The support for MAXDORM is invoked by updating your SMFPRMxx.

- RELATIVEDATE Parameter
  - Used to specify a date range based on the current day, week or month
  - RELATIVEDATE(u, x, y)
  - u = BYDAY, BYWEEK or BYMONTH
  - x = Number of units to move back
  - y = Number of units to gather

- DELETE/ARCHIVE Option
  - LSNAME(IFASMF.LS1,OPTIONS(ARCHIVE))
  - LSNAME(IFASMF.LS1,OPTIONS(DELETE))

Relationship of SMF Records to Log Blocks

![Diagram of SMF records and log blocks]
Log Blocks in a Multi-System CF Logstream

- Avoid reading to end of logstream

  - IFASMFDL starts reading a logstream at a point (approximately) representing a specified time
    - SMARTENDPOINT keyword to specify that IFASMFDL should stop reading a logstream before the end
      - SMARTEPOVER specifies amount of time added to end date/time (default is two hours)
        - Avoids reading to end of logstream

- Allow entire logstream to be archived or deleted

  - Treat logstreams as though they were SMF datasets
  - Will reset logstream starting point to next new block
z/OS Ver 2.1 - SMF Logger Updates

- Specify log stream buffer sizes with new DSPSIZMAX parameter in SMFPRMxx
  - Support for DSPSIZMAX to be used when SMF is initialized also available for z/OS V1.12 and V1.13 with the PTF for APAR OA35175
  - z/OS V2.1 supports dynamic changes via SET SMF and SETSMF

- SMF also supports the use of data compression on zEC12 and zBC12 systems with the zEDC Express feature and the zEnterprise Data Compression (zEDC) feature for z/OS V2.1.

IBM z Enterprise Data Compression

**What is it?**
- zEDC Express is an IO adapter that does high performance industry standard compression
- Used by z/OS Operating System components, IBM Middleware and ISV products
- Applications can use zEDC via industry standard APIs (zlib and Java)
- Each zEDC Express sharable across 15 LPARs, up to 8 devices per CEC.
- Raw throughput up to 1 GB/s per zEDC Express Hardware Adapter

**What Changes?**
- Disk Savings: Many people are already getting value from CMPSC compression and software compression today
- Performance: High throughput alternative to existing System z compression for large or active files.
- Industry Standard: Low cost compressed data exchange across all platforms
- Pervasive: Standard APIs allow quick adoption by middleware products running on System z

**What is the Value?**
- QSAM/BSAM compression can save disk cost
- Business Partner Data Exchange can have higher throughput with lower CPU cost
- Sterling Connect:Direct saves additional link bandwidth, elapsed time.
- ISV Products delivery expanded customer value
- Java transparently accelerates java.util.zip
- IBM Encryption Facility for standard compliant data exchange
- Improved availability with SMF compression
### Compression Coprocessor (CMPSC) vs. zEDC

#### Use Cases

- **Small object compression**
  - Rows in a database
  - VSAM for better disk utilization
  - DB2 for lower memory usage
  - The majority of customers are currently compressing their DB2 rows

- **Large Sequential Data**
  - QSAM/BSAM Online Sequential Data
  - QSAM/BSAM for better disk utilization and batch elapsed time improvements
  - SMF for increased availability and online storage reduction
  - DFSMSdss for better disk and tape utilization for backup data
  - DFSMSshm for improved CPU, disk reduction

- **Industry Standard Data**
  - Cross Platform Data Exchange
  - Java for high throughput standard compression via java.util.zip
  - Encryption Facility for z/OS for better industry data exchange
  - IBM Sterling Connect: Direct for z/OS for better throughput and link utilization
  - ISV support for increased client value

### zEDC Configuration Overview

- **Operating system requirements**
  - Requires z/OS 2.1 or later and new zEDC Express for z/OS feature
    - PTF Support for BSAM/QSAM
  - z/OS V1.13 and V1.12 offer software decompression support only

- **Server requirements**
  - Exclusive to z13, zEC12 and zBC12
  - New zEDC Express feature for PCIe I/O drawer (FC#0420)
    - Each feature can be shared across up to 15 LPARs
    - Up to 8 features available on zEC12 or zBC12 and later
  - Recommended high availability configuration per server is four features
    - No additional software cost is incurred by increasing the number of hardware features
    - Provides high availability during concurrent update (half devices unavailable during update)
    - Recommended minimum configuration per server is two features
  - Hot pluggable
  - For best performance, feature is needed on all systems accessing the compressed data

- **Capacity Planning**
  - The z Batch Network Analyzer now reports on potential zEDC usage for QSAM/BSAM data sets
SMF Exploitation of zEDC Express

- SMF compression support is for both CF or DASD logstream
- Specify the new **COMPRESS** option on one or more logstream definitions LSNAME or DEFAULTLSNAME
  - Option to specify amount of memory to permanently fix for performance
  - Same SMF records can be directed to multiple log streams and compression can be enabled on one of them
- IFASMFDL requirements
  - No changes required if zEDC is available; it will be used automatically
  - Specify the SOFTINFLATE option to process compressed data when there are no zEDC devices available
    - Requires z/OS PTF to provide software inflate (decompression) capability for z/OS 1.12 and 1.13 systems
  - If the SOFTINFLATE option is not specified on a system with no zEDC device, an error will occur and no records will be deleted from the SMF logstream
- Following SMF records to collect performance information:
  - SMF 23 – SMF buffer usage, number of records written etc.
  - SMF 88 - System logger log stream size, frequency of offload
    - Collection enablement is via SMFPRMxx

zEDC and SMF Logstream Data

- New SMFPRMxx **COMPRESS** keyword on LSNAME and DEFAULTLSNAME
  - A buffer of SMF records is compressed by zEDC Express before it is written to the system logger
  - SMF data is only compressed while it is resident in the system logger
  - **PERMFIX** to specify amount of storage used for SMF buffers that can remain permanently fixed
  - When compressed data is processed by IFASMFDL, it decompresses the SMF records for selection and writing
  - **SOFTINFLATE** parameter to process compressed SMF records using software algorithm, for a pre-z/OS V2.1 system or no zEDC Express
SMF Data Flow Overview

Application Generating SMF Records

SMF

SMF Data Space 64k Buffers

LOGR

Compression Output Buffers

End User Applications

Long-term Storage

SMF Buffers

IXGWRITE

zEDC

Offload Data sets

End User Applications

With compatibility PTFs Software Inflate can be done on downlevel z/OS or pre-GA2 /zBC12.

SMF with zEDC

This example shows a DASD-Only logstream used for SMF recording

Reduced Logger overhead

Improved Dump Performance

• Compression SMF logstreams reduce the amount of data in System Logger up to 4x and reduce the elapsed time to extract IFASMFDL data up to 15%

• zEDC compression must be available on all systems that will access zEDC compressed SMF logstreams

• Setup from SMFPRMxx either globally or per Logstream

SMFPRMxx in SYS1.PARMLIB
DEFAULTLSNAME(DEFAULT,…,COMPRESS)
LSNAME(SMF30,TYPE(30),…,COMPRESS(PERMFIX(10M))
LSNAME(RMF,TYPE(70:79),…,COMPRESS)
Logstream BufferParms

- Program requests to write a SMF record
- SMF locates correct dataspace
- Locates appropriate buffer to write the record
- If full, buffer passed to task to be written to logstream

Obtain PCIe Information by Command

- Use the console command D PCIE for general status information
- Use the console command D PCIE,PFID=xxx for adapter details
Obtain PCIe Information via API

- IQPINFO – Obtain PCIe Information
  - The IQPINFO service provides PCIe related information, including any performance statistics
  - The service is described in *MVS Programming: Authorized Assembler Services Reference*
  - The response data area of the IQPINFO service is mapped by the macros
    - IQPYPERF PCIE Performance Data Return Area
    - IQPYPFMBPCIE Function Measurement Block

- RMF Monitor III Data Gatherer collects PCIe performance statistics frequently and writes new SMF Record Type 74 Subtype 9
- The new RMF Postprocessor PCIE Activity Report provides detailed information about PCIE Express based functions. Currently supported functions are:
  - z Enterprise Data Compression (zEDC)
  - Shared Memory Communication via RDMA (SMC-R)

RMF Postprocessor PCIE Activity Report

- Basic PCIe Metrics, e.g. PCI Load/Store and DMA Operations
- Common Request Statistics across all Personalities (Compression and Future Personalities)
- Compression related Statistics
New RMF report shows the utilization of each device.

**RGF Postprocessor Interval Report: PCIe Activity Report**

<table>
<thead>
<tr>
<th>Function ID</th>
<th>Time Busy</th>
<th>% Request Execution Time</th>
<th>Std Dev for Request Execution Time</th>
<th>Request Queue Time</th>
<th>Std Dev for Request Queue Time</th>
<th>Request Size</th>
<th>Transfer Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>9001</td>
<td>50.9</td>
<td>0.17</td>
<td>0.12</td>
<td>0.085</td>
<td>2.95</td>
<td>23.9</td>
<td>21.8</td>
</tr>
<tr>
<td>9011</td>
<td>49.7</td>
<td>0.16</td>
<td>0.10</td>
<td>0.085</td>
<td>2.95</td>
<td>23.9</td>
<td>21.8</td>
</tr>
</tbody>
</table>

**Hardware Accelerator Activity**

- Function ID: 9001
- Time Busy: 50.9
- % Request Execution Time: 0.17
- Std Dev for Request Execution Time: 0.12
- Request Queue Time: 0.085
- Transfer Rate: 23.9

**Hardware Accelerator Compression Activity**

- Function ID: 9001
- Compression Request Rate: 810
- Compression Throughput: 11.5
- Compression Ratio: 20
- Decompression Request Rate: 0
- Decompression Throughput: 0
- Decompression Ratio: 0

The percent of this interval where this specific zEDC Express device was executing requests.

Compression ratio of all requests serviced by zEDC. This will span all users of this device.

Average request queue time in Microseconds for this device.
z13 – RMF PCIE Enhancements

- z13 introduces new PCIE performance measurements for RDMA-over-converged-ethernet (RoCE Express) and zEnterprise data compression (zEDC) devices
  - Existing DMA read/write measurements are replaced by new PCIE function type specific measurements
  - For RoCE Express devices, there are four new measurements
    - Received-Bytes: No. of bytes received on the external ethernet interface
    - Transmitted-Bytes: No. of bytes transmitted on the external ethernet interface
    - Received-Packets: No. of packets received on the external ethernet interface
    - Transmitted-Packets: No. of packets transmitted on the external ethernet interface
  - For zEDC devices, there are two new measurements
    - Consumed-Work-Units: No. of work units processed by the PCI function
    - Maximum Work Units: Maximum no. of work units that the PCI function is capable of processing per second
- With zEC12 / zBC12, the existing DMA Read/Write metrics are still maintained
  - DMA Reads: No. bytes transferred from DMA address spaces to PCIE function
  - DMA Writes: No. bytes transferred from PCIE function to DMA address spaces

RMF Postprocessor PCIE Activity Report

New PCIe Statistics for z13
RMF Postprocessor PCIE Activity Report

New PCIe Statistics for z13

<table>
<thead>
<tr>
<th>PCI Store Memory Operations Rate</th>
<th>PCI Store Block Operations Rate</th>
<th>Refresh PCI Translations Operations Rate</th>
<th>LMA Address Space Count</th>
<th>Read Transfer Rate</th>
<th>Write Transfer Rate</th>
<th>Packets Received Rate</th>
<th>Packets Transmitted Rate</th>
<th>Work Units Processed Rate</th>
<th>Adapter Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.91</td>
<td>0</td>
<td>2.91</td>
<td>1</td>
<td>62.9</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.91</td>
<td>0</td>
<td>2.92</td>
<td>1</td>
<td>143</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>62.9</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

z/OS V2.2 – RMF Monitor III PCIE Activity Report

- RMF Monitor III Data Gatherer collects PCIe performance statistics frequently
- z/OS V2.1 introduced RMF Postprocessor PCIE Activity Report for after-the-facts analysis with SMF 74.9 data
- The new RMF Monitor III PCIE Activity Report provides detailed short-term information about PCIE Express based functions. Currently supported functions are:
  - z Enterprise Data Compression (zEDC)
  - Shared Memory Communication via RDMA (SMC-R)

New Monitor III data gatherer option PCIE | NOPCIE

<table>
<thead>
<tr>
<th>NAME</th>
<th>EBRMII04</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESCRIPTION</td>
<td>PARMLIB MEMBER WITH RMF MONITOR III GATHERER OPTIONS</td>
</tr>
<tr>
<td></td>
<td>ALL OPTIONS ARE SET TO DEFAULTS</td>
</tr>
<tr>
<td>CYCLE(1000)</td>
<td>SAMPLE EVERY SECOND (1000 MSEC)</td>
</tr>
<tr>
<td>DATASET(STOP)</td>
<td>NO DATASET SUPPORT</td>
</tr>
<tr>
<td>DATASET(NOSWITCH)</td>
<td>APPEND TO LAST NON-FULL DATASET</td>
</tr>
<tr>
<td>DATASET(MOLDI7)</td>
<td>CONTROLS BUFFER PAGES IN STORAGE</td>
</tr>
<tr>
<td>MINTIME(100)</td>
<td>LENGTH OF MINTIME</td>
</tr>
<tr>
<td>NOOPTIONS</td>
<td>DO NOT DISPLAY OPTIONS</td>
</tr>
<tr>
<td>RESOURCES</td>
<td>Specifies ES started task name</td>
</tr>
<tr>
<td>NOSTOP</td>
<td>RUN UNTIL OPERATOR ISSUES STOP</td>
</tr>
<tr>
<td>SYNC(90)</td>
<td>MINTIME SYNCHRONIZATION</td>
</tr>
<tr>
<td>DPD</td>
<td>ACTIVATE OWS PROCESS DATA GATHERING</td>
</tr>
<tr>
<td>PCIE</td>
<td>ACTIVATE PCIE DATA GATHERING</td>
</tr>
<tr>
<td>NOZFS</td>
<td>NO ZFS DATA GATHERING</td>
</tr>
</tbody>
</table>

PCIE Data collected by default
z/OS V2.2 – RMF Monitor III PCIE Activity Report

RMF V2R2   PCIE Activity
Samples: 60      System: TRX1 Date: 04/18/15 Time: 10.56.00 Range: 60 Sec

<table>
<thead>
<tr>
<th>ID</th>
<th>PCID</th>
<th>Type</th>
<th>Jobname</th>
<th>ASID</th>
<th>Status</th>
<th>Time%</th>
<th>Load</th>
<th>Store</th>
<th>Block</th>
<th>Refr</th>
<th>Read</th>
<th>Write</th>
</tr>
</thead>
<tbody>
<tr>
<td>00A2</td>
<td>013C</td>
<td>RoCE</td>
<td>VTAM390</td>
<td>002C</td>
<td>Alloc</td>
<td>100</td>
<td>0.113</td>
<td>5999</td>
<td>0</td>
<td>1.69</td>
<td>0.267</td>
<td>288</td>
</tr>
<tr>
<td>006C</td>
<td>0204</td>
<td>zEDC</td>
<td>FPGHWAM</td>
<td>0013</td>
<td>Alloc</td>
<td>0.3</td>
<td>0</td>
<td>102</td>
<td>0</td>
<td>14.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>007C</td>
<td>025C</td>
<td>zEDC</td>
<td>FPGHWAM</td>
<td>0013</td>
<td>Alloc</td>
<td>0.3</td>
<td>0</td>
<td>102</td>
<td>0</td>
<td>14.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RMF Hardware Accelerator And Compression Activity
Press Enter to return to the Report panel.

Function ID : 006C       Alloc Time % : 99.7
Allocated   : 22.03.22 on 02/18/15
Hardware Accelerator
Time Busy % : 0.286        Transfer Rate : 4.87
Request Execution Time : 28.0 Std. Deviation: 0.87
Queue Time   : 65.7     Std. Deviation: 140
Size         : 47.6     
Buffer Pool
Memory Size   : 16        Utilization   : 0
Compression   Decompression
Request Rate  : 102      Throughput    : 2.91
Throughput    : 0.809    Ratio         : 2.99

SMF 30 support for zEDC

- Using zEDC, CPU time for job could go down
- How much zEDC capacity a job used?
- APAR OA45767 adds zEDC usage info to Type 30 records
  - Total number of zEDC requests
  - Total queue time in microseconds
  - Total execute time in microseconds
  - Deflate and inflate statistics for number of bytes input and output
Configuring and Using SMF Logstreams with zEDC Compression

Thank you for attending!

Summer SHARE
August 2015
Session 17644