IBM Education Assistance for z/OS V2R2

Item: RMF Monitor III PCIE Support
Element/Component: RMF

Material current as of May 2015

Agenda

- Presentation Objectives
- Overview
- Usage & Invocation: RMF Monitor III Data Gatherer Option
- Usage & Invocation: RMF Monitor III PCIE Activity Report
- Usage & Invocation: RMF DDS PCIE Activity Support
- Presentation Summary
- Appendix

Presentation Objectives

- Explain purpose/usage of RMF z/OS V2R2 Monitor III PCIE Support
- Data gathering option PCIE / NOPCIE was added to RMF Monitor III that can be used to control data collection of hardware features
    - RDMA (Remote Direct Memory Access) support for z/OS over Converged Enhanced Ethernet (RoCE)
    - IBM zEnterprise Data Compression (zEDC) capability using zEDC Express
  - New Monitor III PCIE Activity report allows short-term performance analysis of RoCE devices and zEDC hardware accelerators
  - RMF Distributed Data Server (DDS) supports new resource types *PCIE* and *PCIE Function* to allow performance analysis of RoCE devices and zEDC hardware accelerators by DDS API exploiters

Overview

- Problem Statement / Need Addressed
  - Performance problems on RoCE devices and zEDC hardware accelerators can only be analyzed after the facts using SMF 74.9 / RMF Postprocessor
- Solution
  - The new RMF Monitor III PCIE Activity Report and the Distributed Data Server API allow short-term performance analysis of PCIe devices and hardware accelerators
- Benefit / Value
  - RMF online monitoring can be used to identify performance problems related to PCIe devices and hardware accelerators

Usage & Invocation: RMF Monitor III Data Gatherer Option

- Users can control whether or not they want Monitor III to collect PCIE activity data by specifying data gathering option:

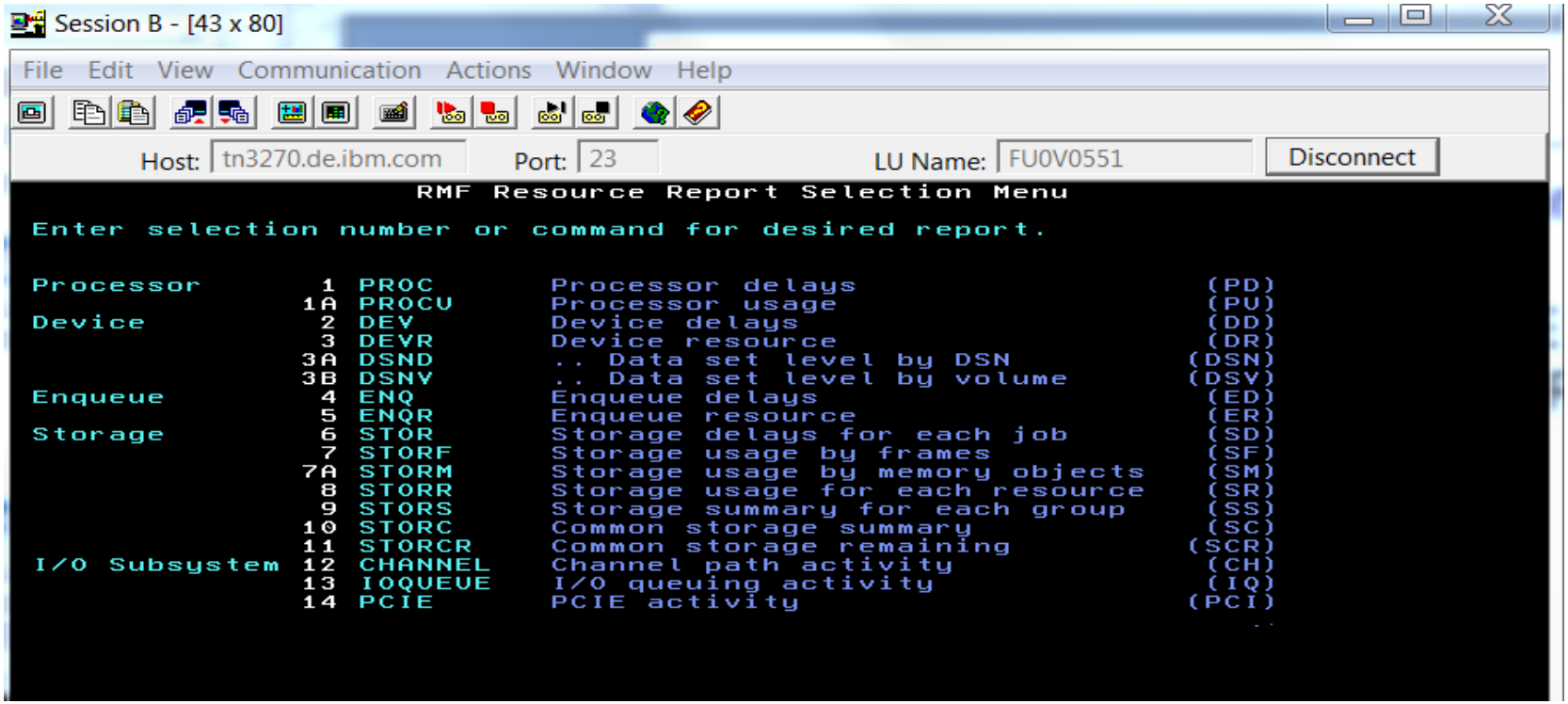
  PCIE | NOPCIE

- Default value PCIE is set in shipped PARMLIB member ERBRMF04

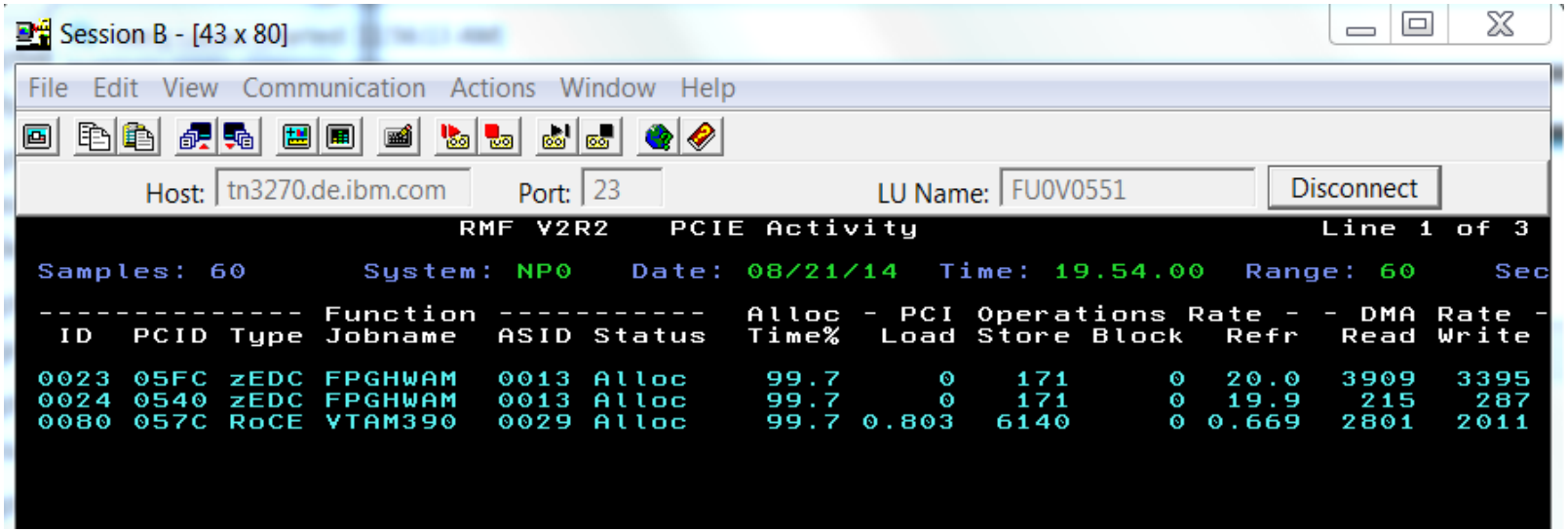
```
/*********************************************************************/
/* NAME:        ERBRMF04                                             */
/* DESCRIPTION: PARMLIB MEMBER WITH RMF MONITOR III GATHERER OPTIONS */
/*              (ALL OPTIONS ARE SET TO DEFAULTS)                    */
/*********************************************************************/
  CYCLE(1000)               /* SAMPLE EVERY SECOND (1000 MSEC)      */
  DATASET(STOP)             /* NO DATASET SUPPORT                   */
  DATASET(NOSWITCH)         /* APPEND TO LAST NON-FULL DATASET      */
  DATASET(WHOLD(7))         /* CONTROLS BUFFER PAGES IN STORAGE     */
  MINTIME(100)              /* LENGTH OF MINTIME                    */
NOOPTIONS                   /* DO NOT DISPLAY OPTIONS               */
  RESOURCE(*JES2,JES2)      /* SPECIFIES JES STARTED TASK NAME      */
NOSTOP                      /* RUN UNTIL OPERATOR ISSUES STOP       */
  SYNC(00)                  /* MINTIME SYNCHRONIZATION              */
  ....................................
  VSAMRLS                   /* ACTIVATE VSAM RLS GATHERING          */
  OPD                       /* ACTIVATE OMVS PROCESS DATA GATHERING */
  PCIE                      /* ACTIVATE PCIE DATA GATHERING         */
NOZFS                       /* NO ZFS DATA GATHERING                */
NOSGSPACE                   /* NO STORAGE GROUP SPACE GATHERING     */
NOLOCK                      /* NO LOCK DATA GATHERING               */
```

Usage & Invocation: RMF Monitor III PCIE Activity Report

- To request the Monitor III PCIE Activity Report, select a 3 from the Primary Menu, then select 14 from the Resource Report Selection Menu
  or
  enter command: *PCIE* or *PCI*



Usage & Invocation: RMF Monitor III PCIE Activity Report

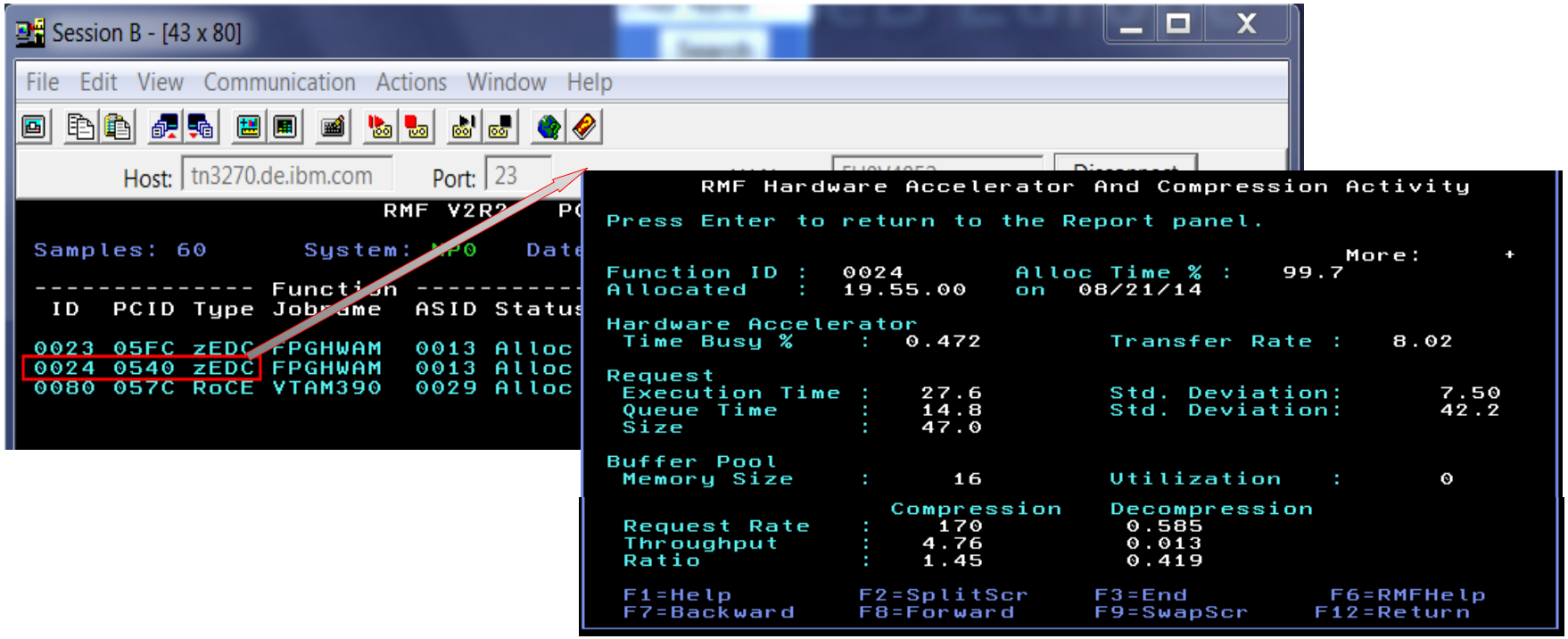
- The Monitor III PCIE Activity Report can be used to investigate performance problems that are related to PCI Express based functions
- On the main panel, metrics are displayed that are independent of the type of the exploited hardware feature and reflect the activity of the z/OS system on which RMF data collection took place
- 

Usage & Invocation: PCIE Activity Report Main Panel Fields

Field Heading	Meaning
Function ID	Hexadecimal identifier of the PCIE Function for which performance data is reported.
Function PCID	Physical channel identifier for the PCIE function.
Function Type	Device type for the PCIE function which can be one of the following: **HWA** Hardware Accelerator **Oth** Unknown device type **RoCE** Remote Direct Memory Management **zEDC** zEnterprise Data Compression
Function Jobname	Name of the job who allocated the PCIE function.
Function ASID	Address space ID of the job who allocated the PCIE function.
Function Status	The PCIE function status at the end of this reporting interval: **Alloc** The function is allocated and in use **Dpend** The function is in the process of de-allocation **Error** The function is in permanent error **DeAlloc** The function is de-allocated **Unknown** The function status is unknown
Alloc Time %	Percentage of reporting interval for which the PCIE function was allocated or in the process of de-allocation.
PCI Operations Rate Load	Rate of PCI Load operations that were executed within this reporting interval.
PCI Operations Rate Store	Rate of PCI Store operations that were executed within this reporting interval.
PCI Operations Rate Block	Rate of PCI Store Block operations that were executed within this reporting interval.
PCI Operations Rate Refresh	Rate of Refresh PCI Translations operations that were executed within this reporting interval.
DMA Rate Read	Number of megabytes per second that were transferred by DMA reads from all defined DMA address spaces to the PCIE function.
DMA Rate Write	Number of megabytes per second that were transferred by DMA writes from the PCIE function to all defined DMA address spaces.

Usage & Invocation: HW Accelerator And Compression Activity

- If cursor is placed on one of the cursor sensitive fields *Function ID, Function PID, or Function Type*, additional metrics are displayed for **zEDC Accelerators** on pop-up panel HW Accelerator And Compression Activity

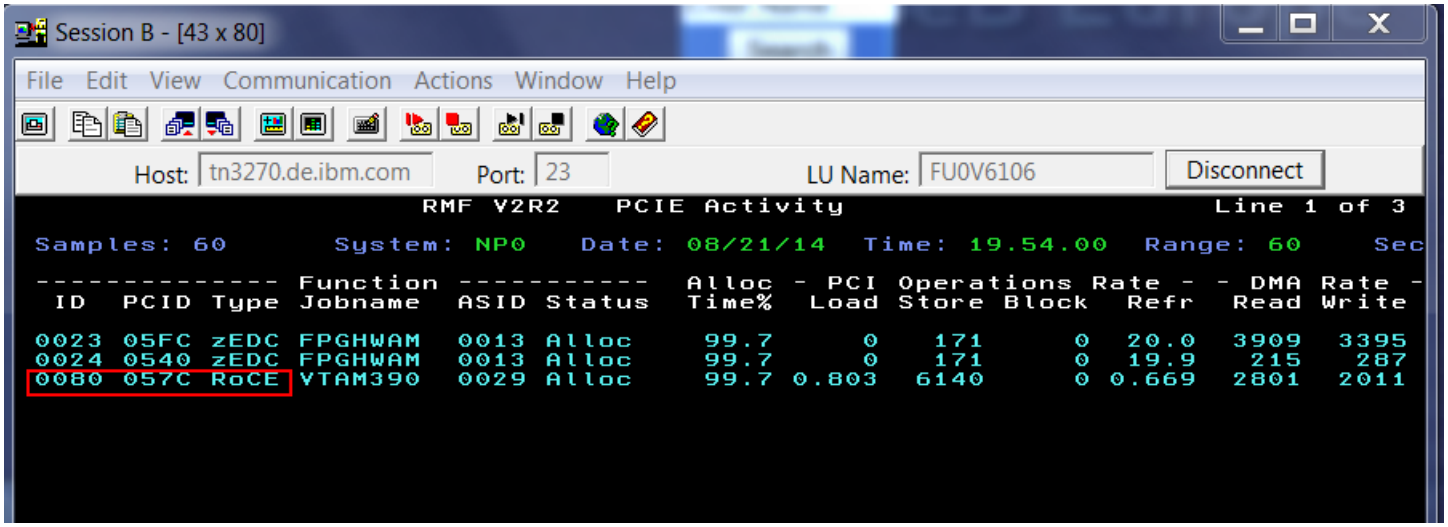


Usage & Invocation: HW Accelerator And Compression Activity Fields

Field Heading	Meaning
Allocated	Date and time when the Hardware Accelerator was allocated.
Hardware Accelerator Time Busy %	The percentage of time that this partition kept the Hardware Accelerator busy.
Hardware Accelerator Transfer Rate	The number of megabytes per second that were transferred by DMA operations.
Request Execution Time	The average time in microseconds the Hardware Accelerator took to process a request.
Request Execution Time Std. Deviation	The standard deviation of the request execution time.
Request Queue Time	The average queue time in microseconds that was spent for a request.
Request Queue Time Std. Deviation	The standard deviation of the request queue time.
Request Size	The average number of kilobytes that were transferred per request.
Buffer Pool Memory Size	The total size of memory in megabytes that is allocated to the buffer pool.
Buffer Pool Utilization %	The average utilization percentage of the buffer pool that z/OS kept for in-use buffers.
Request Rate	The number of compression or decompression requests per second.
Throughput	The number of megabytes that were compressed or decompressed per second.
Ratio	The ratio between input and output bytes that were compressed or decompressed within this reporting interval.

Usage & Invocation: RoCE Device Activity

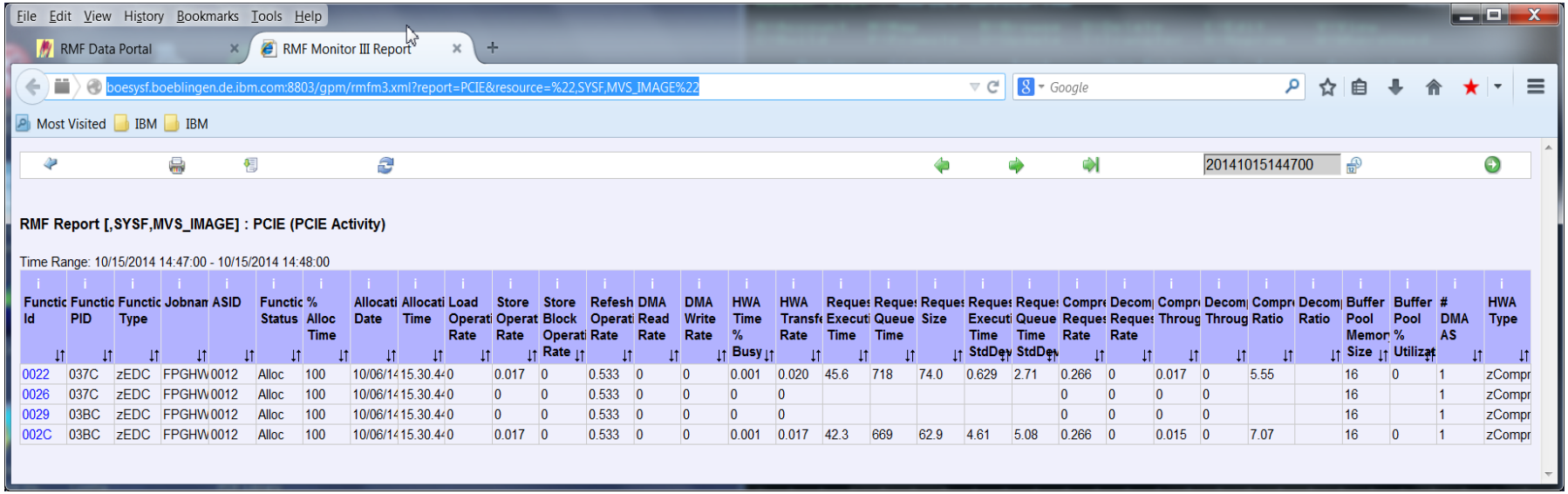
- If cursor is placed on one of the cursor sensitive fields *Function ID, Function PID, or Function Type*, message “No additional information available” is displayed when selected PCIE function is a **RoCE device**



Usage & Invocation: DDS Full PCIE Activity Report Support

- Alternatively, the browser based version of the report can be requested from the RMF Distributed Data Server (DDS) by using the following URL:

  http://hostname:8803/gpm/rmfm3.xml?report=PCIE&resource=,sysname,MVS_IMAGE



Usage & Invocation: DDS Resource Model Enhancements

- RMF DDS Resource Model represents a composition of resources that can exist in a Parallel Sysplex environment
  - MVS_IMAGE is a child resource of resource SYSPLEX
  - I/O_SUBSYSTEM is a child resource of resource MVS_IMAGE
  - Child resource PCIE is added to resource I/O_SUBSYSTEM

Children of: SYSF,*,I/O_SUBSYSTEM

Icon	Resource	Metrics	Attributes	Res-Type
	SYSF,*,ALL_SSIDS	Metrics	N/A	ALL_SSIDS
	SYSF,*,ALL_LCUS	Metrics	N/A	ALL_LCUS
	SYSF,*,ALL_CHANNELS	Metrics	N/A	ALL_CHANNELS
	SYSF,*,ALL_VOLUMES	Metrics	N/A	ALL_VOLUMES
	SYSF,*,ZFS	Metrics	N/A	ZFS
	SYSF,*,PCIE	Metrics	N/A	PCIE

  - Child resource PCIE_FUNCTION is added to resource PCIE

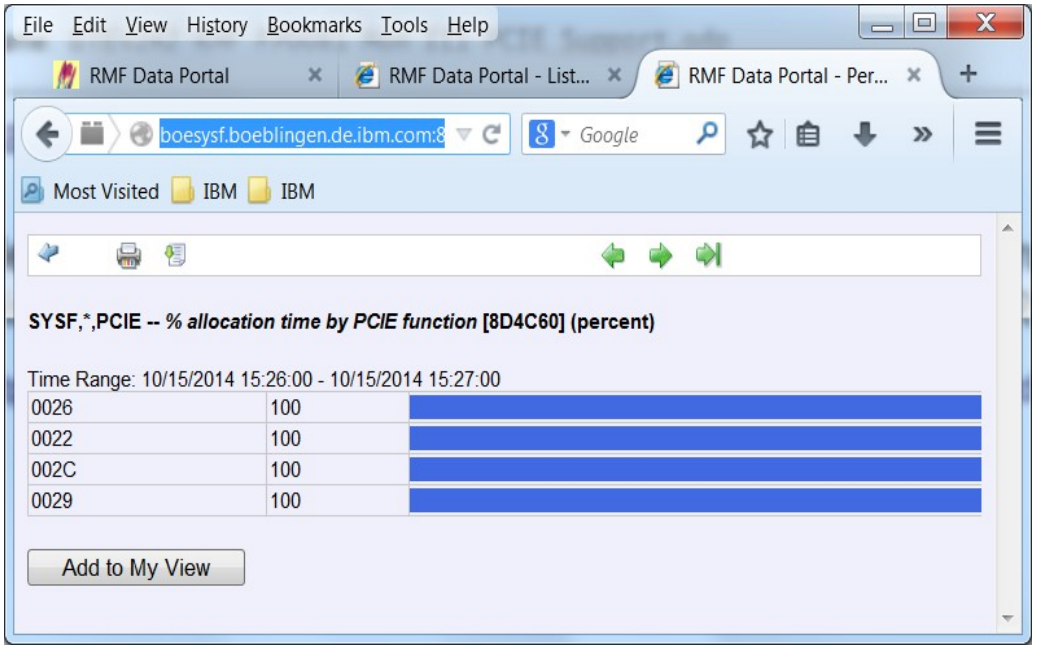
Children of: SYSF,*,PCIE

Icon	Resource	Metrics	Attributes	Res-Type
	SYSF,002C,PCIE_FUNCTION	Metrics	N/A	PCIE_FUNCTION
	SYSF,0022,PCIE_FUNCTION	Metrics	N/A	PCIE_FUNCTION
	SYSF,0026,PCIE_FUNCTION	Metrics	N/A	PCIE_FUNCTION
	SYSF,0029,PCIE_FUNCTION	Metrics	N/A	PCIE_FUNCTION

Usage & Invocation: DDS PCIE metrics

- A variety of metric values that are related to resource types PCIE and PCIE_FUNCTION can be requested from the RMF Distributed Data Server (DDS)
- Example: % Allocation Time by PCIE Functions allocated in z/OS system *sysname*

http://hostname:8803/gpm/perform.xml?/resource=sysname,*,PCIE,&id=8D4C60



% allocation time by PCIE function
% buffer pool utilization by PCIE function
% time busy by PCIE function
buffer pool memory size by PCIE function
compression ratio by PCIE function
compression request rate by PCIE function
compression throughput by PCIE function
decompression ratio by PCIE function
decompression request rate by PCIE function
decompression throughput by PCIE function
request execution time by PCIE function
request execution time standard deviation by PCIE function
request queue time by PCIE function
request queue time standard deviation by PCIE function
request size by PCIE function
transfer rate by PCIE function
DMA read rate by PCIE function
DMA write rate by PCIE function
PCI load operations rate by PCIE function
PCI refresh operations rate by PCIE function
PCI store block operations rate by PCIE function
PCI store operations rate by PCIE function

Presentation Summary

- Online monitoring of hardware features RoCE and IBM zEnterprise Data Compression (zEDC) was added to RMF in the V2R2 release
  - New Monitor III PCIE Activity report can be invoked in ISPF and through DDS API
  - DDS API exploiters can obtain PCIE related metric values
  - PCIE activity data is collected by RMF Monitor III if data gathering option PCIE is set
    -
  -

Appendix

- RMF homepage: www.ibm.com/systems/z/os/zos/features/rmf/
  - Product information, newsletters, presentations, etc.
  - Downloads
    - RMF Spreadsheet Reporter
    - RMF PM Java Edition
    - RMF Postprocessor XML Toolkit
- RMF email address: rmf@de.ibm.com
- Documentation and news
  - *RMF Report Analysis*, SC34-2665
  - *RMF User's Guide,* SC34-2664
  - Latest version of PDF files can be downloaded from:
    - http://www.ibm.com/systems/z/os/zos/bkserv/