zFlash Setup, Management, and Configuration

February 5, 2013

SHARE in San Francisco

Tom Mathias
IBM SE/HMC Development
mathiast@us.ibm.com

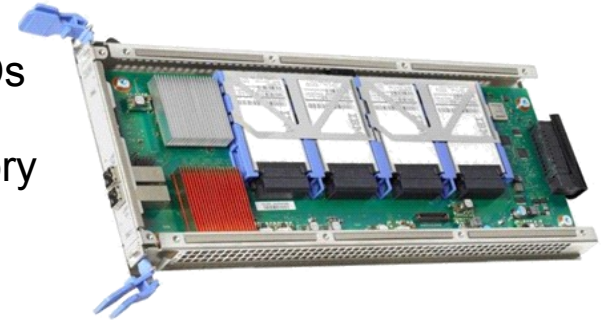
Elpida Tzortzatos
IBM z/OS Development
elpida@us.ibm.com

Abstract

The IBM zFlash introduces non-volatile memory in a disk-like package on an I/O card that can be used by Operating systems like z/OS for paging storage and to ultimately improve overall z/OS performance. Because zFlash is both non-volatile storage and an I/O device, there are some unique aspects of zFlash as compared to traditional DASD and I/O adapters. This “how-to” presentation will cover these unique aspects, including setup, configuration, management, monitoring and eventual discontinuance of zFlash by covering the z/OS, Hardware Management Console (HMC) and Support Element (SE) controls for zFlash.

What is zFlash (aka Flash Express)?

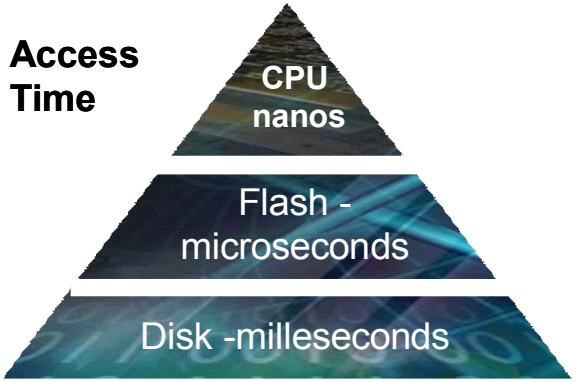
- FLASH Express
  - Flash Express is a PCIe IO adapter with NAND Flash SSDs
  - Physically comprised of internal storage on Flash SSDs
  - Used to deliver a new tier of memory- storage class memory
  - Uses **PCIe I/O drawe**r



One Flash Express Card

  - Sized to accommodate all LPAR paging
    - Each **card pair** provides **1.4 TB** usable storage (2.8 TB total)
    - Maximum 4 card pairs (4 X1.4=5.6 TB)
  - Supported on z/OS V1.13 plus web deliverable

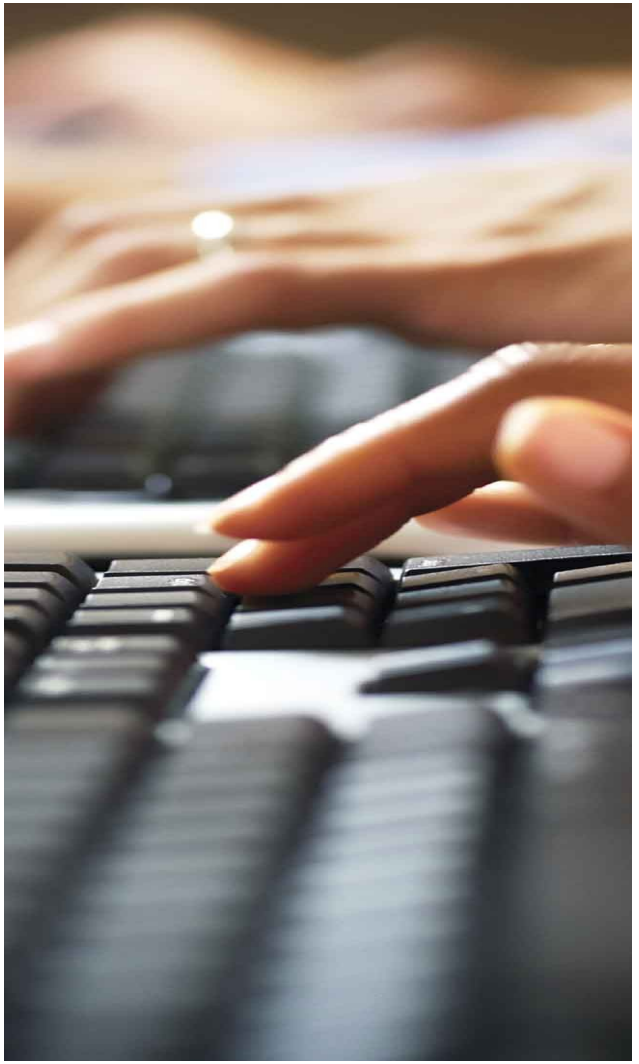
  - **Designed for continuous availability**
    - Concurrent Firmware update for service
    - RAID 10 design



  - **Immediately usable**
    - No capacity planning needed
    - No intelligent data placement needed

  - **Secured**
    - Flash Express adapter is protected with 128-bit AES encryption.
    - Key Management provided based on a Smart Card
    - Secure Cryptographic Erase meets audit requirements

Representative Use Cases - Flash Express



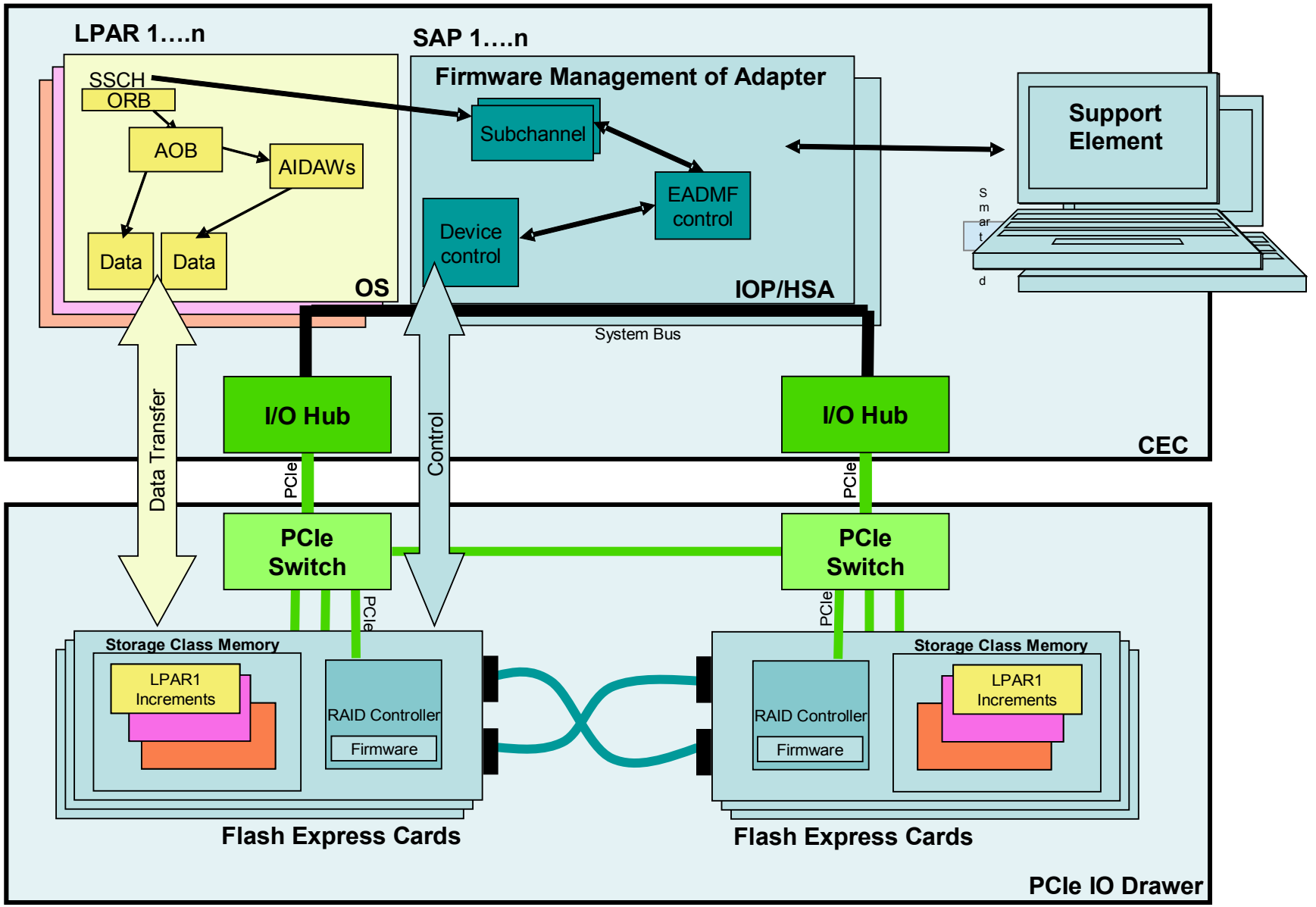
Flash Express can reduce latency delays from paging to bring system availability to new heights and improve overall service levels

Application related errors will require collection of diagnostics. These diagnostics can be collected faster with Flash Express, reducing paging related delays that can impact your overall system availability.

Having your working data resident in Flash can help accelerate start of day processing, and improve service for many industries at the busiest time of their work day- a time when they cannot afford disruptions.

DB2 and Java in memory buffer pools work to store and process application data. DB2 and Java can benefit from 1MB pageable large pages with Flash Express, improving overall performance.

z FLASH Implementation

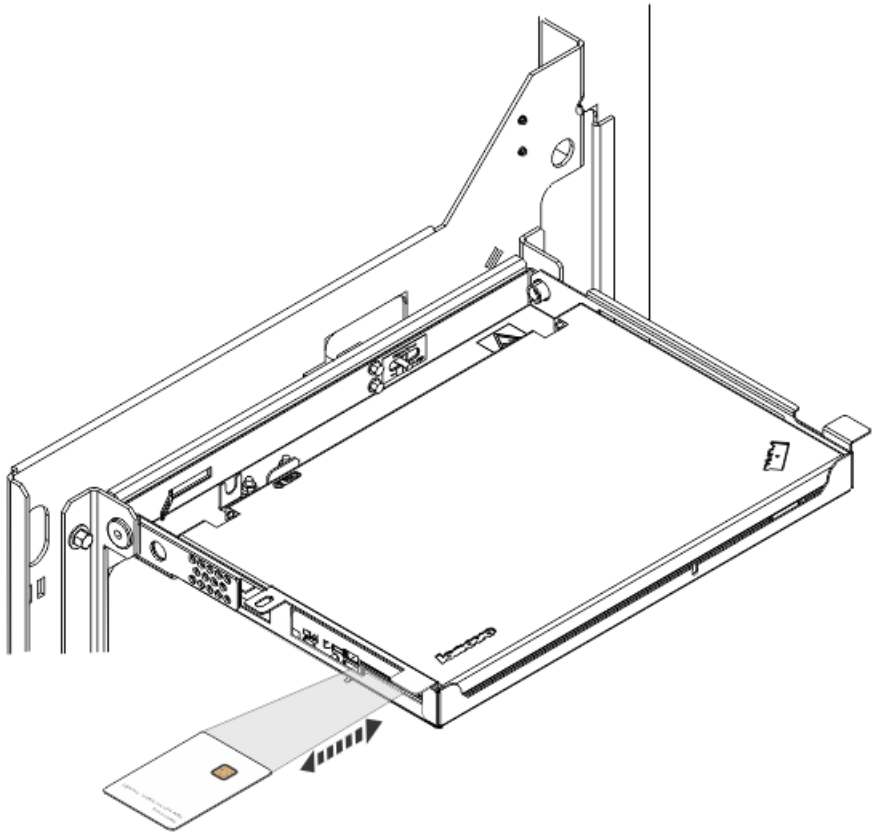


Outline

- Introduction to zFlash
- Initial Setup
  - Customer Service Representative (CSR) Portion
    - Install Smart Cards in Support Elements (SE)
    - Install zFlash cards if necessary
    - Create Pair
  - Customer Portion
    - No IOCDS changes
    - Allocate zFlash memory to partition(s)
    - Configure z/OS to use zFlash
- Management
  - Management of zFlash Allocations
  - zFlash PCHID details
  - View Partition to PCHID map
  - View Flash Allocations for a specific Partition
  - View Flash (details)
  - System Activity Display (SAD) / Monitors Dashboard
  - Console Events
  - Security Logs
  - Status (Service Personnel Only)
  - Configure On/Off (Service Personnel Only)
  - Service On/Off (Service Personnel Only)
- Terminating Flash
  - Change all instances of z/OS to no longer use zFlash
  - Disband all zFlash pairs
  - Remove SE Smart Cards and destroy (optional)

Install Smart Cards in Support Elements (CSR responsibility)

- Cards will be installed by the CSR
  - During machine installation if zFlash shipped with the machine
  - Before installing the first zFlash adapters (if the machine was not shipped with zFlash)



More on Smart Cards

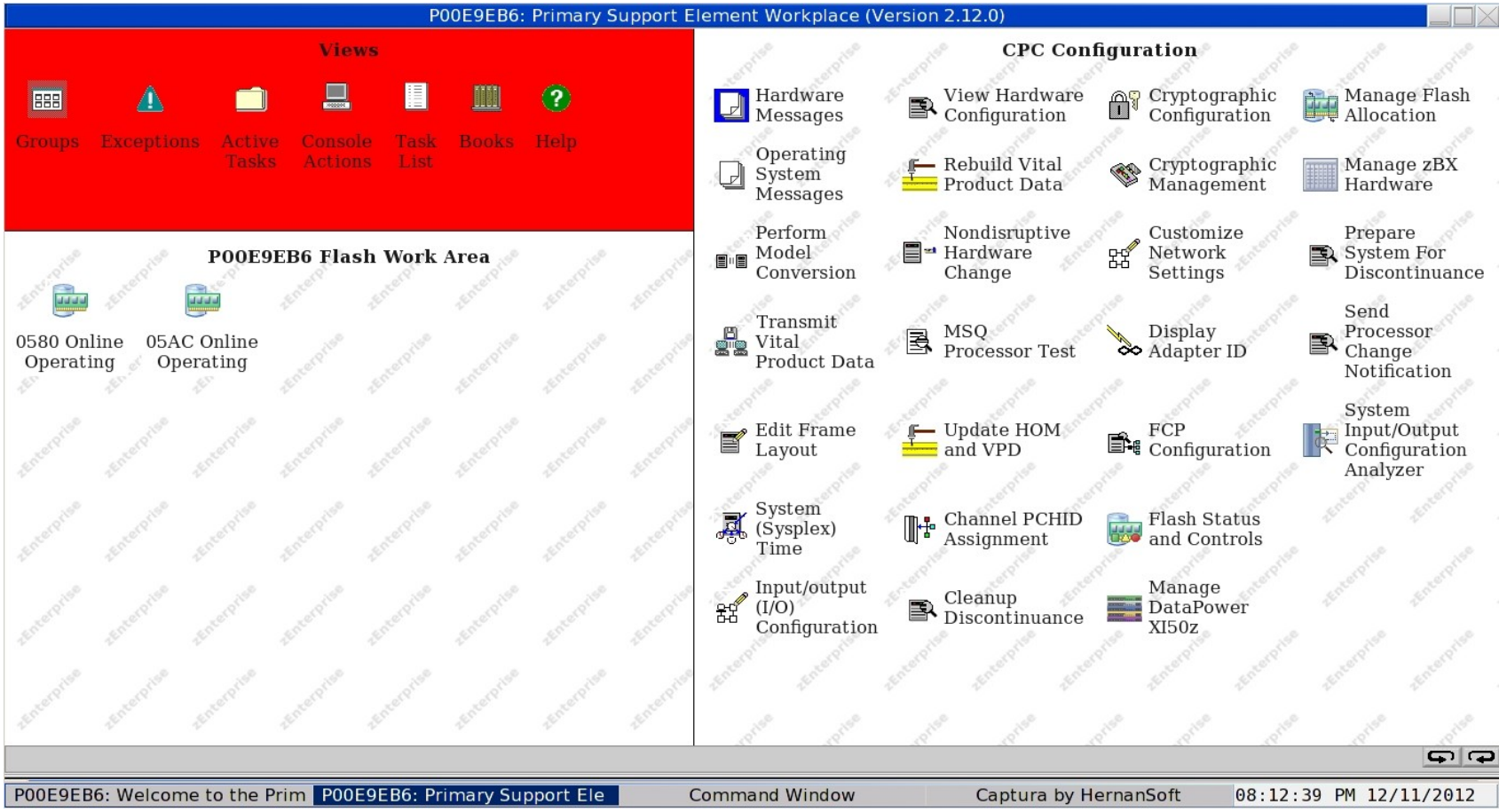
- The data on the zFlash cards is encrypted. This is done to prevent access to the data if a zFlash card is removed from the system, such as for a repair action or thru some malicious action (i.e. theft).
- The Smart Cards are an essential part of managing the encryption keys.
- The blank Smart Card is the same one used by the TKE device.
- The Primary SE will create an authentication key using the smart card and store it on the SE. The Alternate SE will uses the smart card in it to store the key sent from the primary.
- The smart card, the SE hardware, the CEC, and the generated Key are tightly coupled in order to prevent access to the data on the zFlash card in any place other than the CEC it was formatted for.
- If for some reason the smart card fails on the primary an automatic switch to the alternate will happen and a service call will occur to have the smart card or the SE serviced. There are procedures to ensure the repaired SE or Smart Card is properly updated with the encryption keys.
- The keys will not be preserved during migrations/upgrades. So, persistence of data on the zFlash adapters is thus not guaranteed. The zFlash adapters are therefore good for things like paging storage but should not be used for situations where persistence is required.
- Bottom line: The Smart Cards must be installed so that the SE is prepared to store and handle the encryption keys used to protect the data on the zFlash adapters.

Install zFlash cards if necessary (CSR responsibility)

- Installed in pairs in Seneca cages, one per I/O domain
- Pairs are cabled together with 2 SAS cables

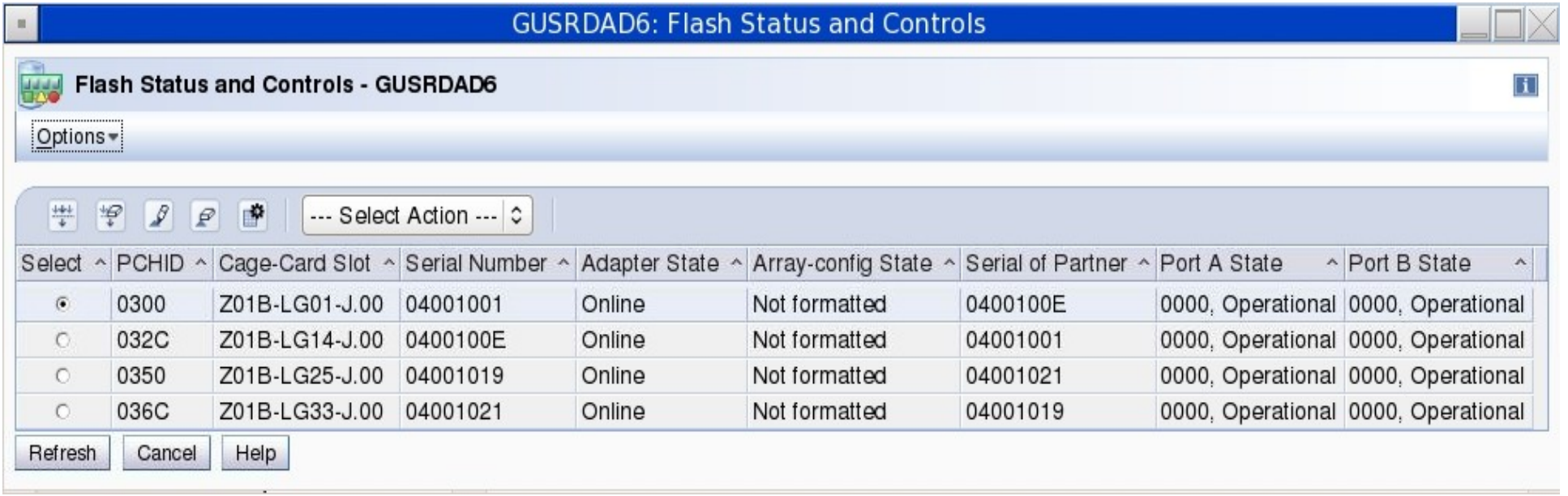
Install zFlash cards if necessary (CSR responsibility)

- Once installed, the cards are visible on the Support Element's User Interface as a PCHID



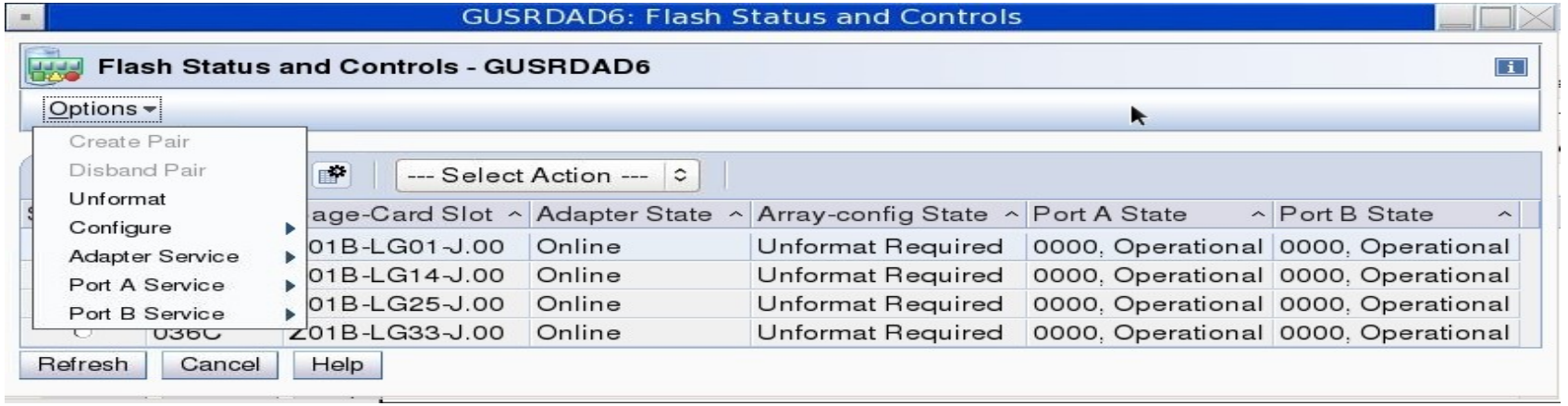
Create pair(s) of zFlash adapters (CSR Responsibility)

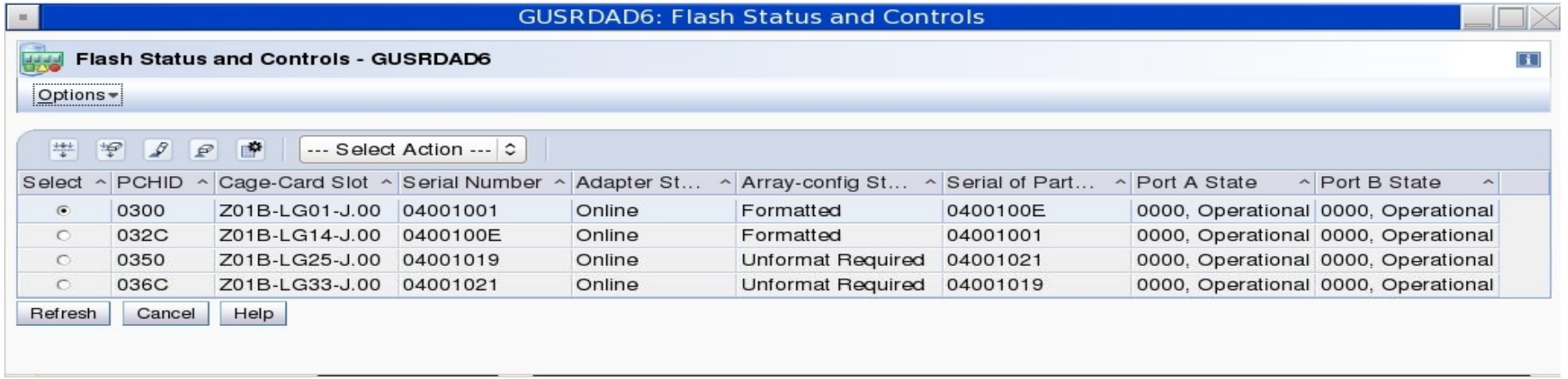
- A “create pair” operation must be performed that allows the paired adapters to initialize themselves into a pair and format the storage.
- Done via a new SE task, **Flash Status and Controls**
- Service Personnel only



Create Pair

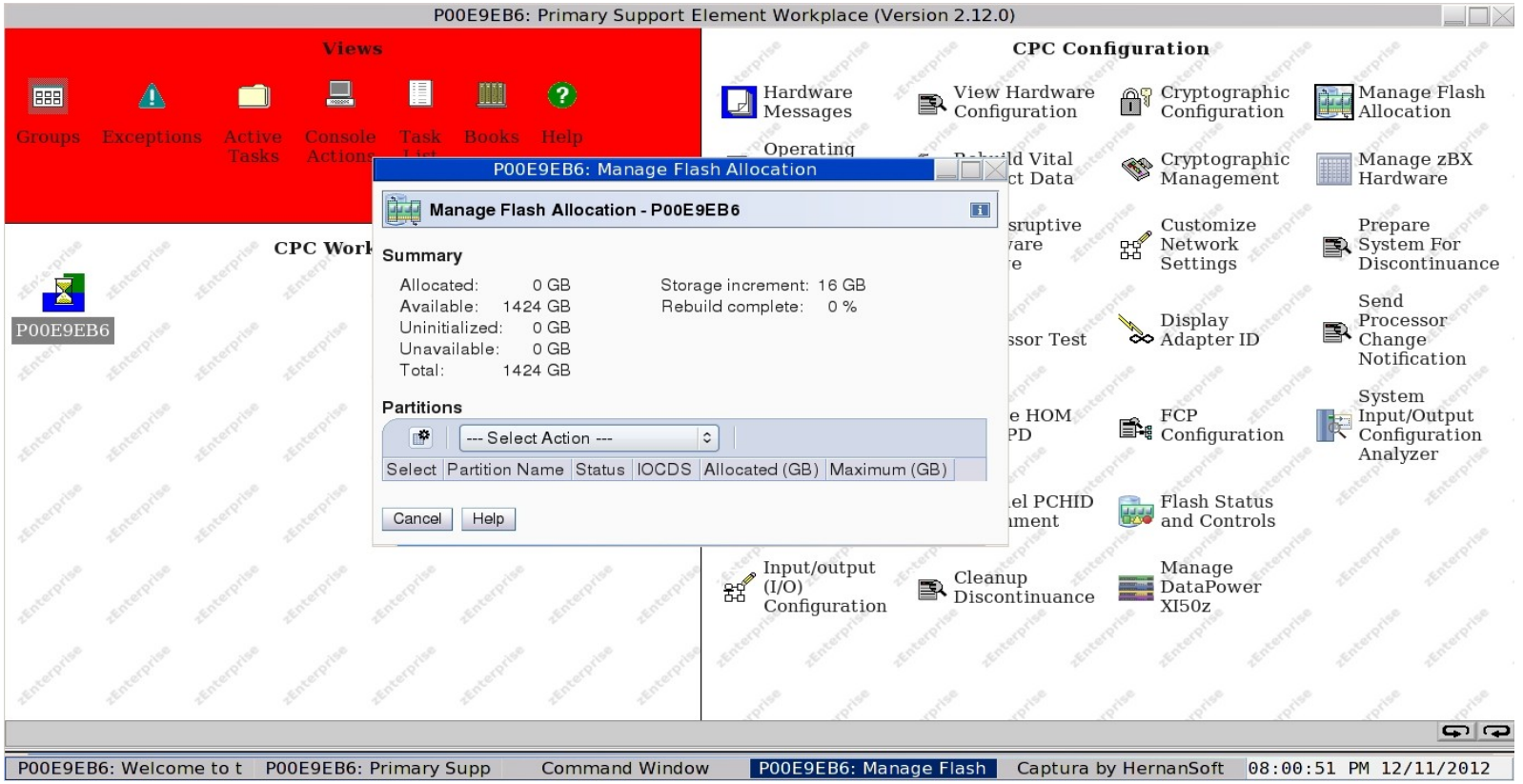
- Select **Create Pair** to create and format a pair
- Use the Refresh button to monitor the progress of the formatting.
- It takes a while (15 to 20 minutes) to complete the pairing/formatting operation.



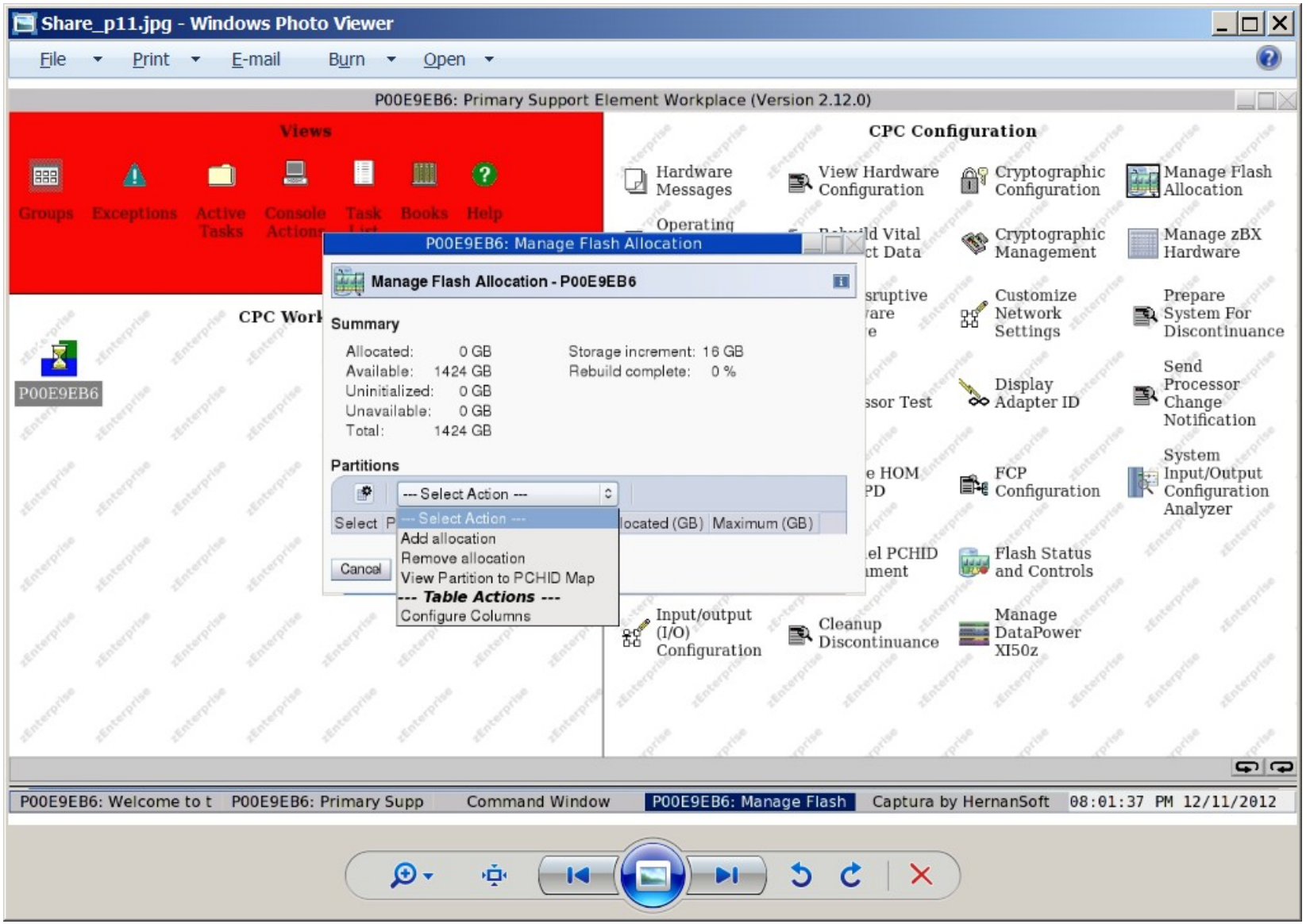


Allocate zFlash Memory to partition(s)
Manage Flash Allocation SE and HMC task

- Available on both the HMC and SE.
- Displays current summary Flash information for the system.
- Displays current Flash information by partition.
- Use to Add, Change or Remove allocations to a partition.

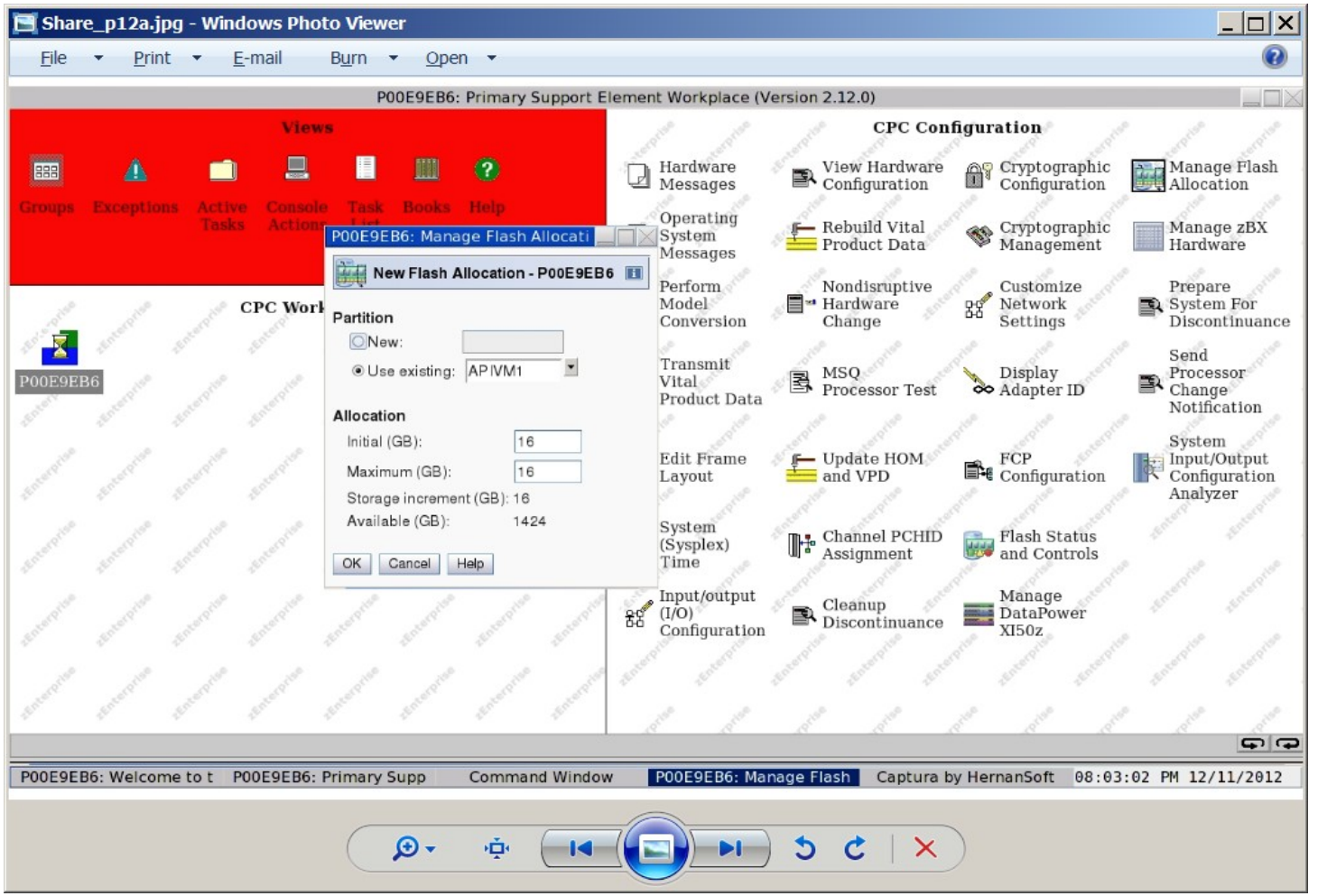


Manage Flash Allocation Task's Actions



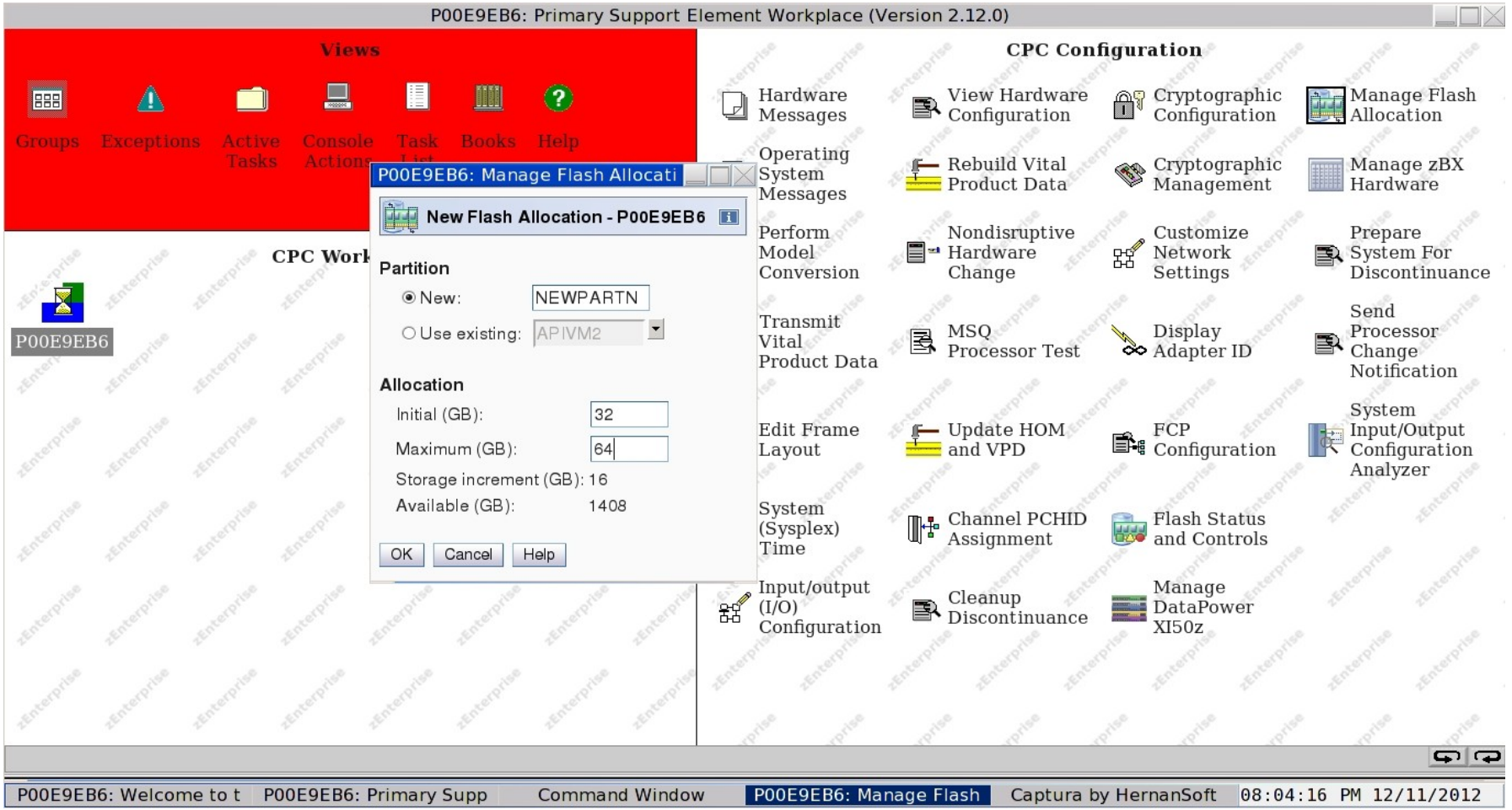
Manage Flash Allocation - Add zFlash Allocation

- Allocation can be done for a partition defined in any IOCDS or a partition not currently defined.
- May be done after initial zFlash setup as necessary (such as when a new partition is defined).
- Example of picking an existing partition:



Manage Flash Allocation - Add zFlash Allocation

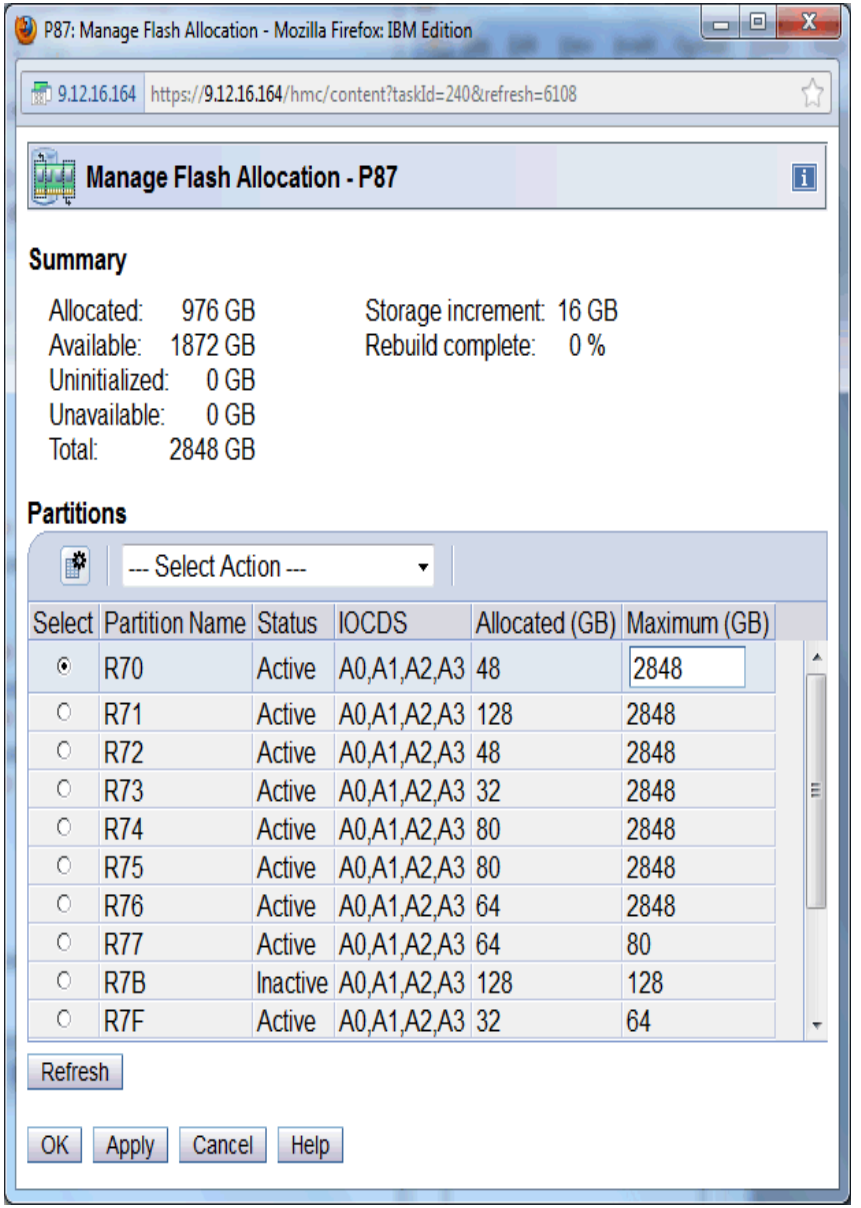
- Example of typing in a new partition name:



z/OS Configuration

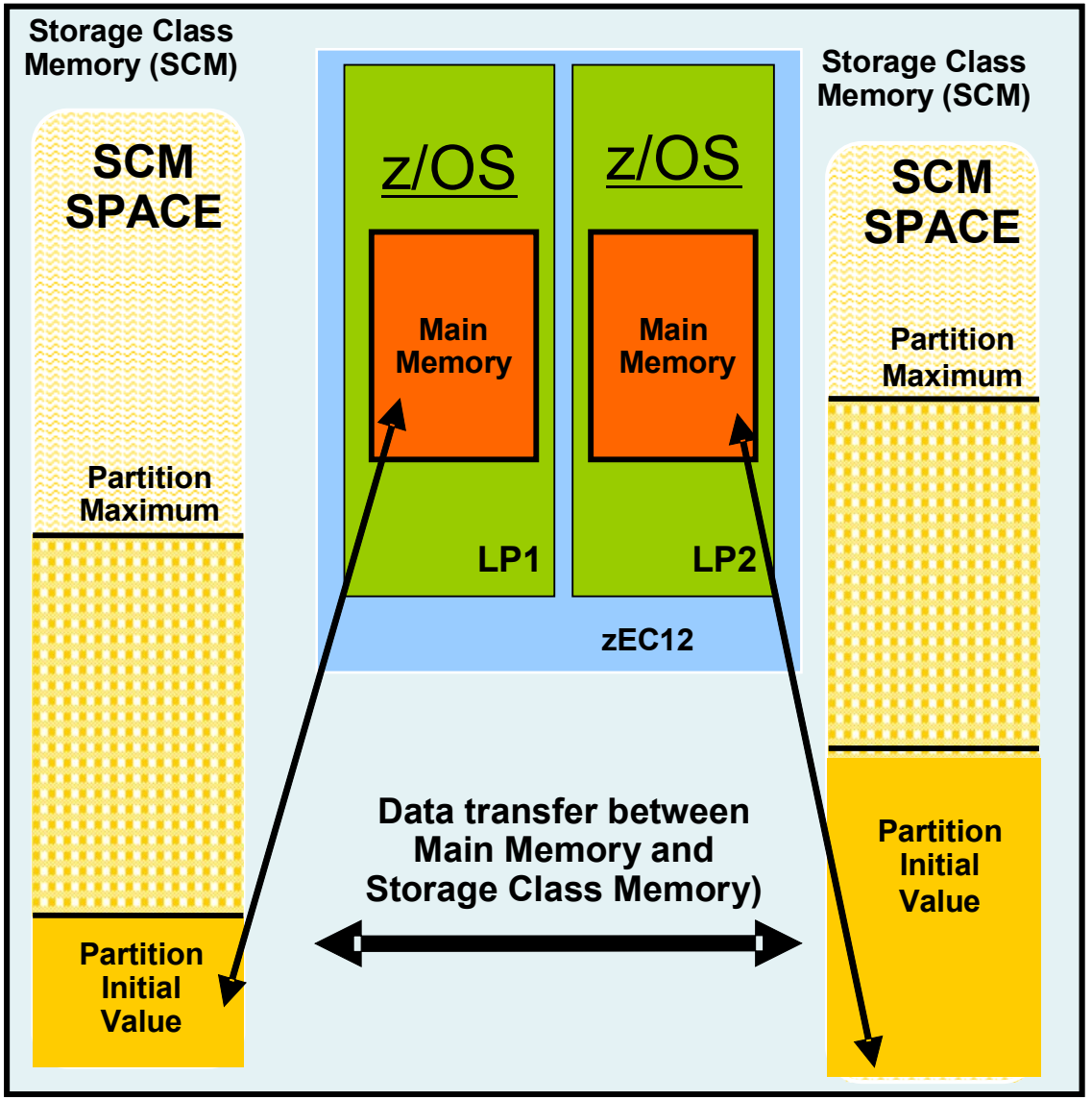
Allocating Flash to a partition

- The initial and maximum amount of Flash Memory available to a particular logical partition is specified at the SE or HMC via a Flash Memory Allocation panel
- Dynamically change maximum amount of Flash Memory available to a logical partition
- Additional Flash Memory (up to the maximum allowed) can be **configured online** to a logical partition dynamically at the SE or HMC
  - This can also be done via an operator command
- Can dynamically configure Flash Memory **offline** to a logical partition at the SE or HMC
  - For z/OS this can also be done via an operator command
- Predefined subchannels, no IOCDS needed



System z Flash Virtualization

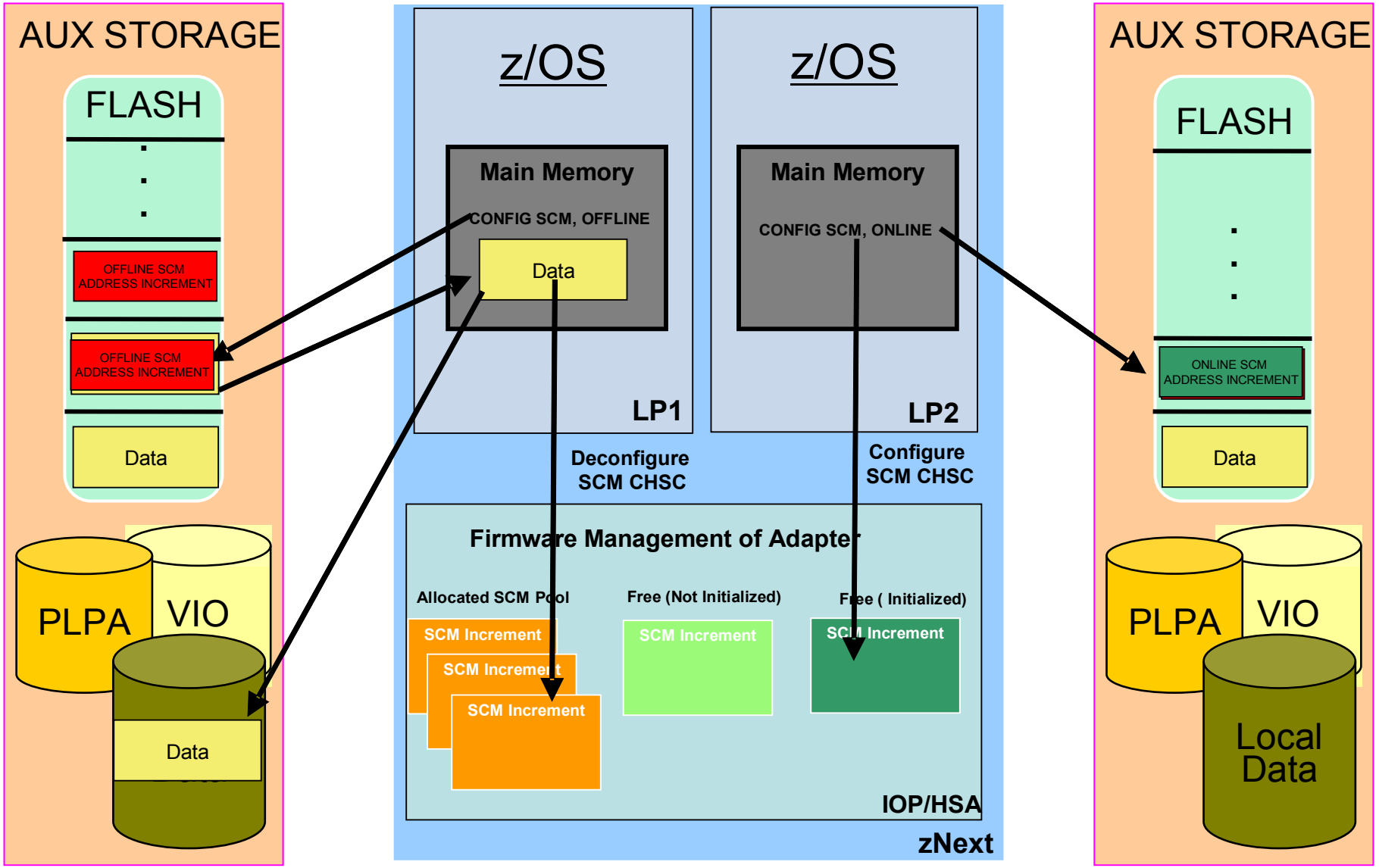
- Full virtualization of physical Flash cards across partitions
- To Software, Flash is an abstracted Storage Class Memory Space
  - Each LPAR can be configured with its own SCM address space
  - Allocate Flash to partitions by amount, not by card size
- Qualities of Service are Built in
  - Error Isolation, Transparent mirroring, Centralized diagnostics, etc.
  - Hardware Logging, FRU Call, Recovery: Independent of software
  - Underlying technology is transparent



zFlash for z/OS Paging Value

- **Flash Memory is a faster paging device as compared to HDD**
  - The value is NOT in replacing memory with Flash but replacing disk with Flash
  - Flash is suitable for workloads that can tolerate paging and will not benefit workloads that cannot afford to page
  - The z/OS design for Flash Memory does not completely remove the virtual storage constraints created by a paging spike in the system. (Some scalability relief is expected due to faster paging I/O with Flash Memory.)

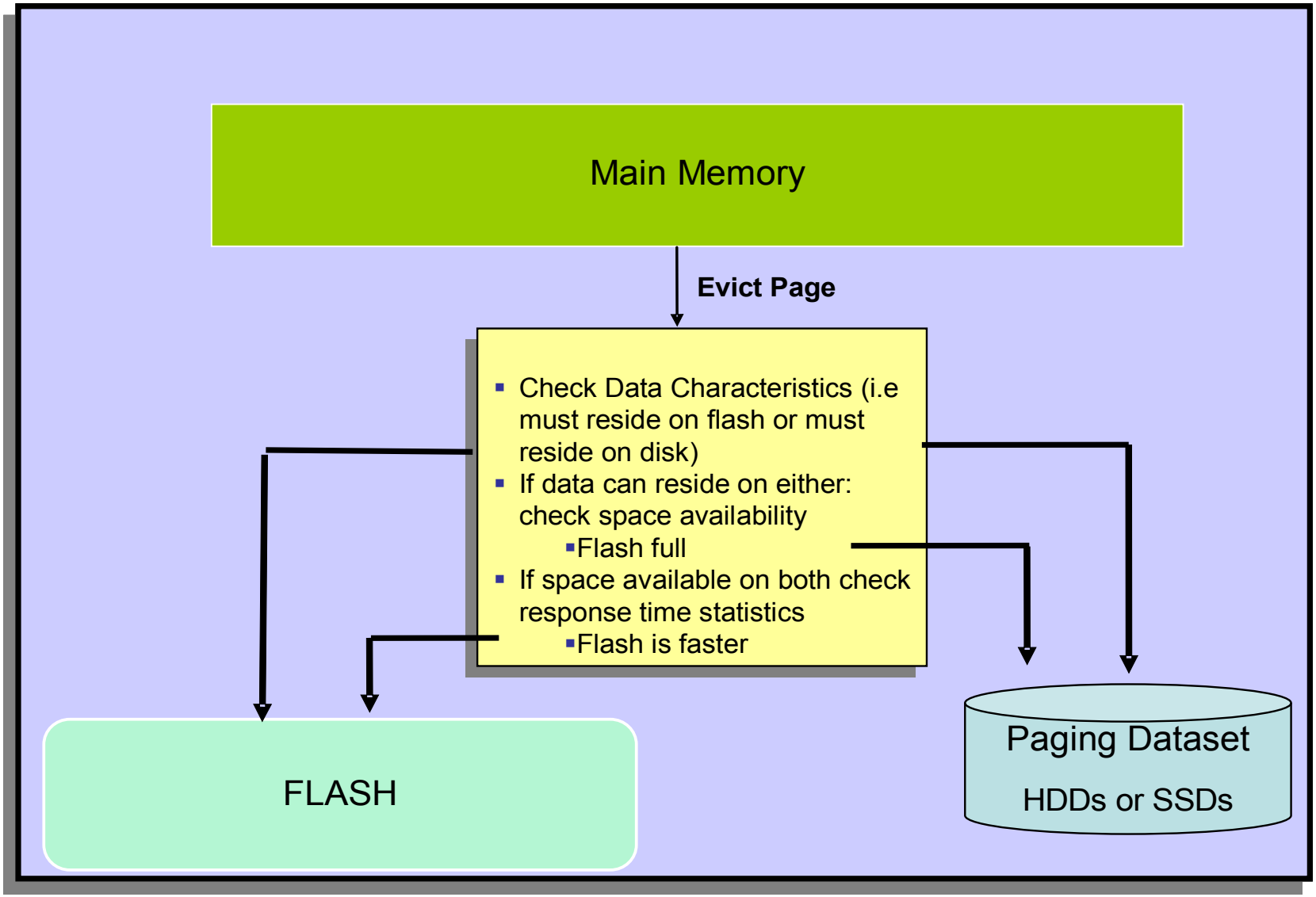
A z/OS zFlash Configuration



Typical Customer Configurations for zFlash

- Flash card pair memory size is 1.4TB
  - Min: 1 Card Pair
  - Max: 4 Card Pairs
- **Typical customer configuration is 6 to 8 LPARs per CEC and 40GB - 80GB for paging configuration dataset size**
- **Even with 10 LPARs per CEC, each LPAR has 160 GB of Flash Memory available for its paging datasets, more than double the current typical customer configuration.**
  - **All paging data can easily reside on Flash**
  - Data will preferably go to Flash and only go to disk (if any) when Flash is full
  - No intelligent placement of data on internal Flash needed

zFlash vs Disk Placement Criteria



Flash vs Disk Placement

Data Type	Data Page Placement
PLPA	**At IPL/NIP time PLPA pages will be placed both on Flash and disk.**
VIO	**VIO data will always be placed on disk (First to VIO accepting datasets with any spillover flowing to nonvio datasets)**
Pageable Large Pages	**If contiguous Flash space is available, pageable large page will be written to Flash.** **If Flash is not available in the system configuration pageable large pages will be backed with 4k page frames.**
All other data	**If available space exists on both Flash and disk then make a selection based on response time.**

z/OS zFlash Use Cases

- **Paging**
  - **z/OS paging subsystem will work with mix of internal Flash and External Disk**
    - Self Tuning based on measured performance
    - Improved Paging Performance, Simplified Configuration
  - **Begin Paging 1 MB Large Pages only on Flash**
    - Exploit Flash's random IO read rate to get CPU performance by enabling additional use of Large Pages. Currently large pages are not pagable.
  - **Begin Speculative Page-In of 4K Pages, 1MB Pages only on Flash**
    - Exploit Flash's random IO read rate to get Improved Resilience over Disruptions.
    - Market Open, Workload Failover

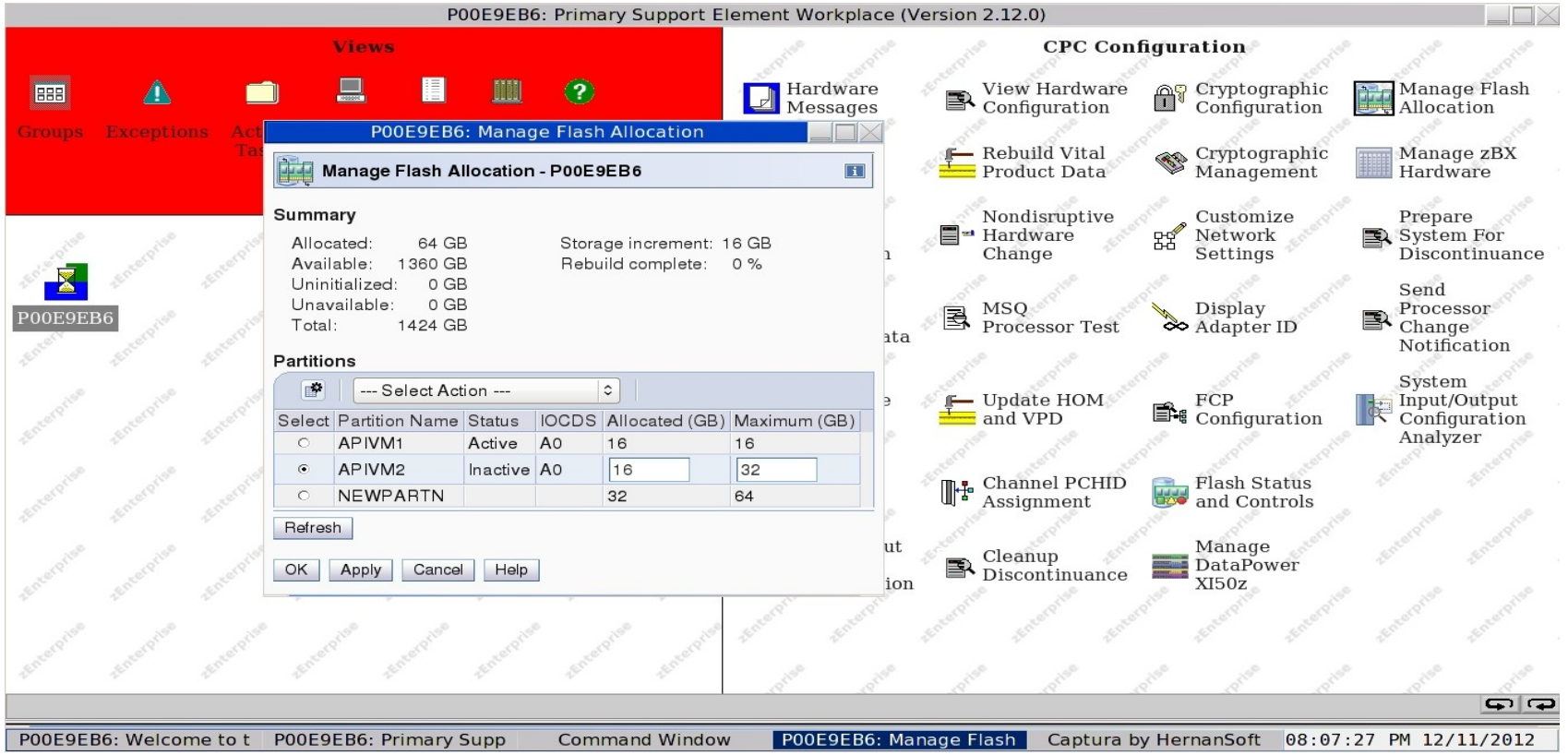
New z/OS Message

IAR031I USE OF STORAGE-CLASS MEMORY FOR PAGING IS ENABLED - PAGESCM=ALL, ONLINE=xxxxxxxxM

- Issued during IPL to indicate whether SCM is in use for paging, what value was used for the PAGESCM parameter, and the amount of SCM that is in use for paging

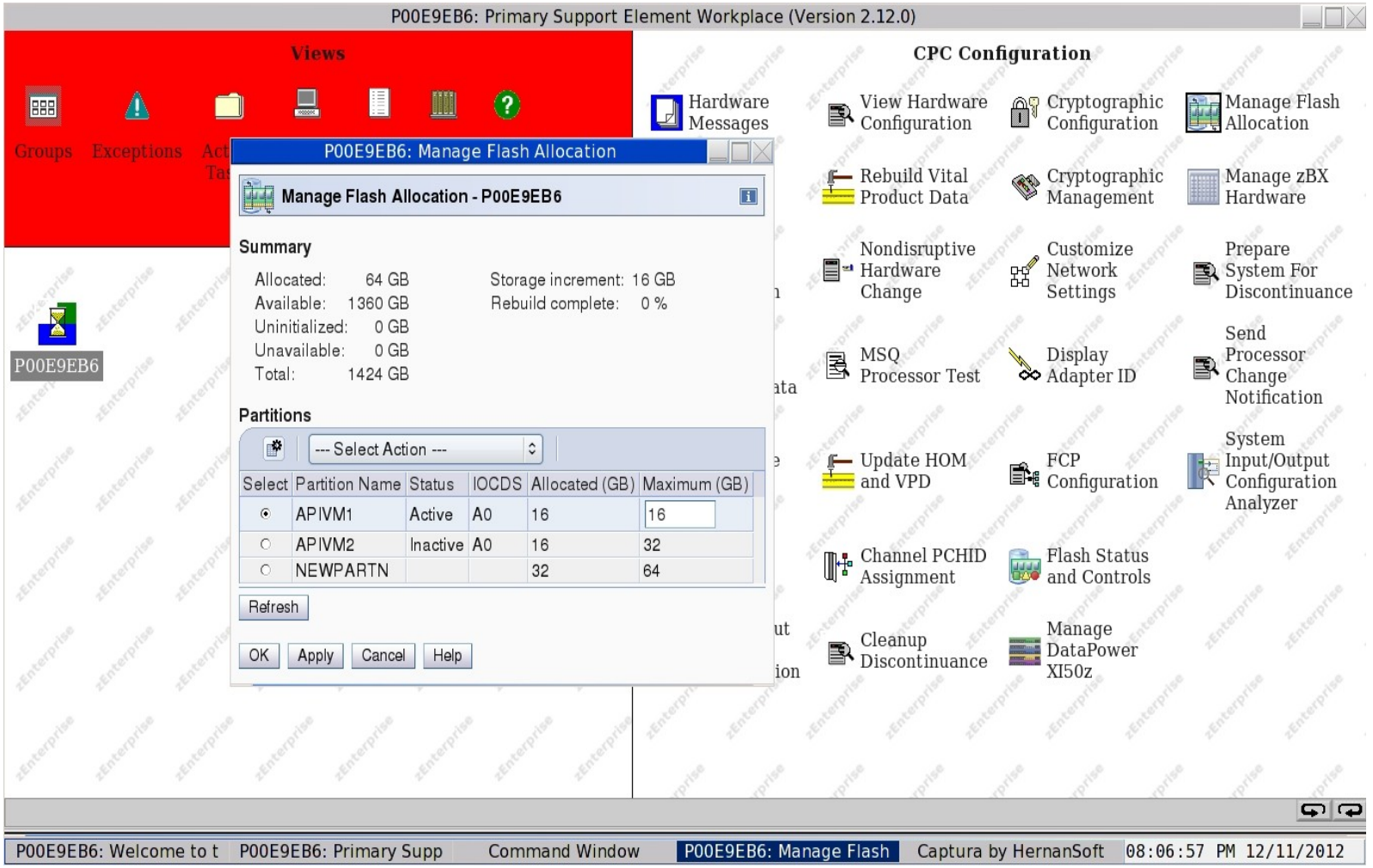
Manage Flash Allocation - Change zFlash Allocation

- Allocated can only be changed for inactive partitions (APIVM2) or undefined partitions (NEWPARTN)
- Changing Allocated results in loss of data
- Changing allocations for an inactive partition:



Manage Flash Allocation - Change zFlash Allocation

- Changing allocations for an active partition (notice only the maximum can be altered):

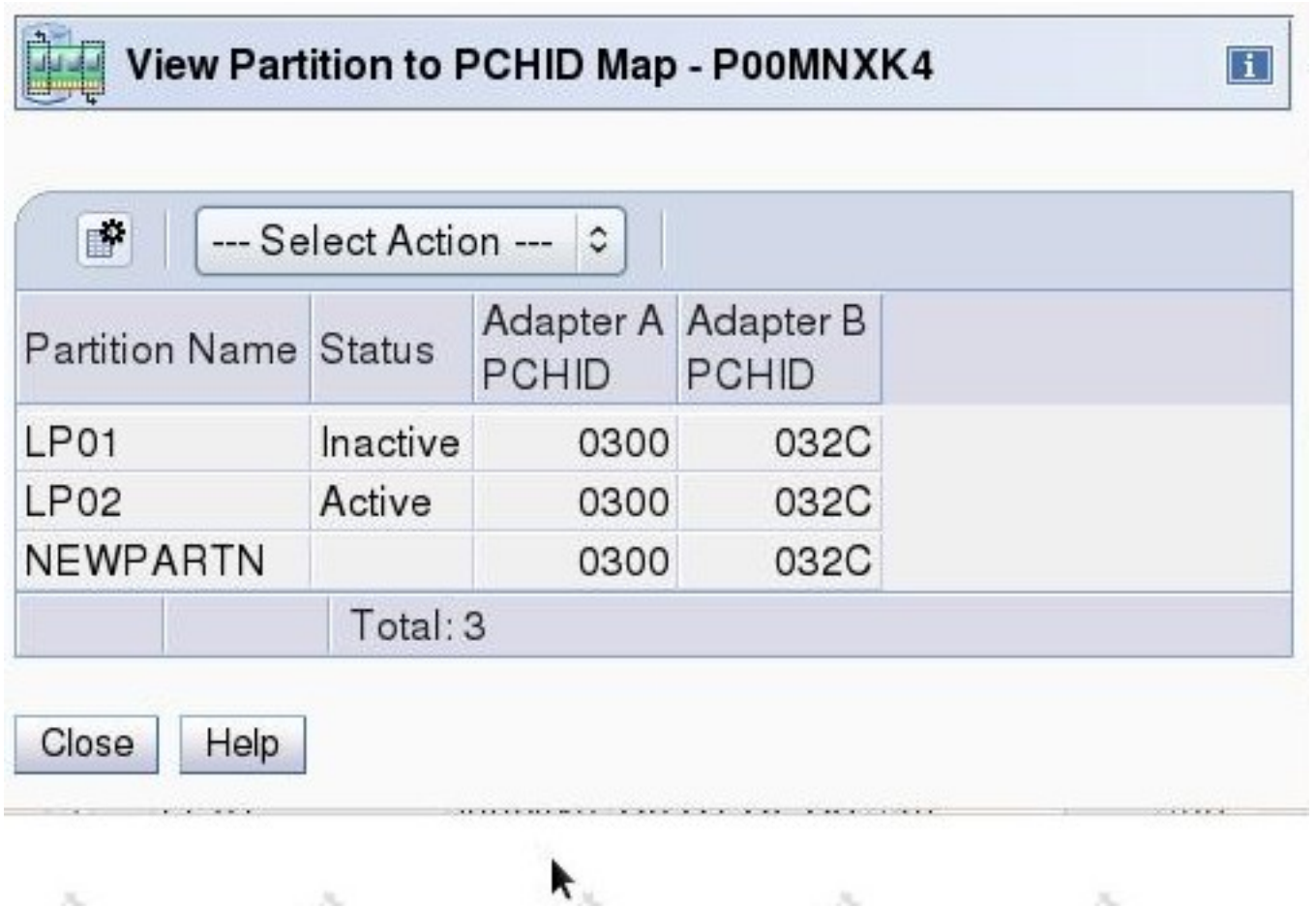


Manage Flash Allocation - Remove zFlash Allocation

- Remove Allocation can only be performed for an inactive partition
- All data will be lost
- A warning message will be issued and confirmation required before the Remove Allocation is done

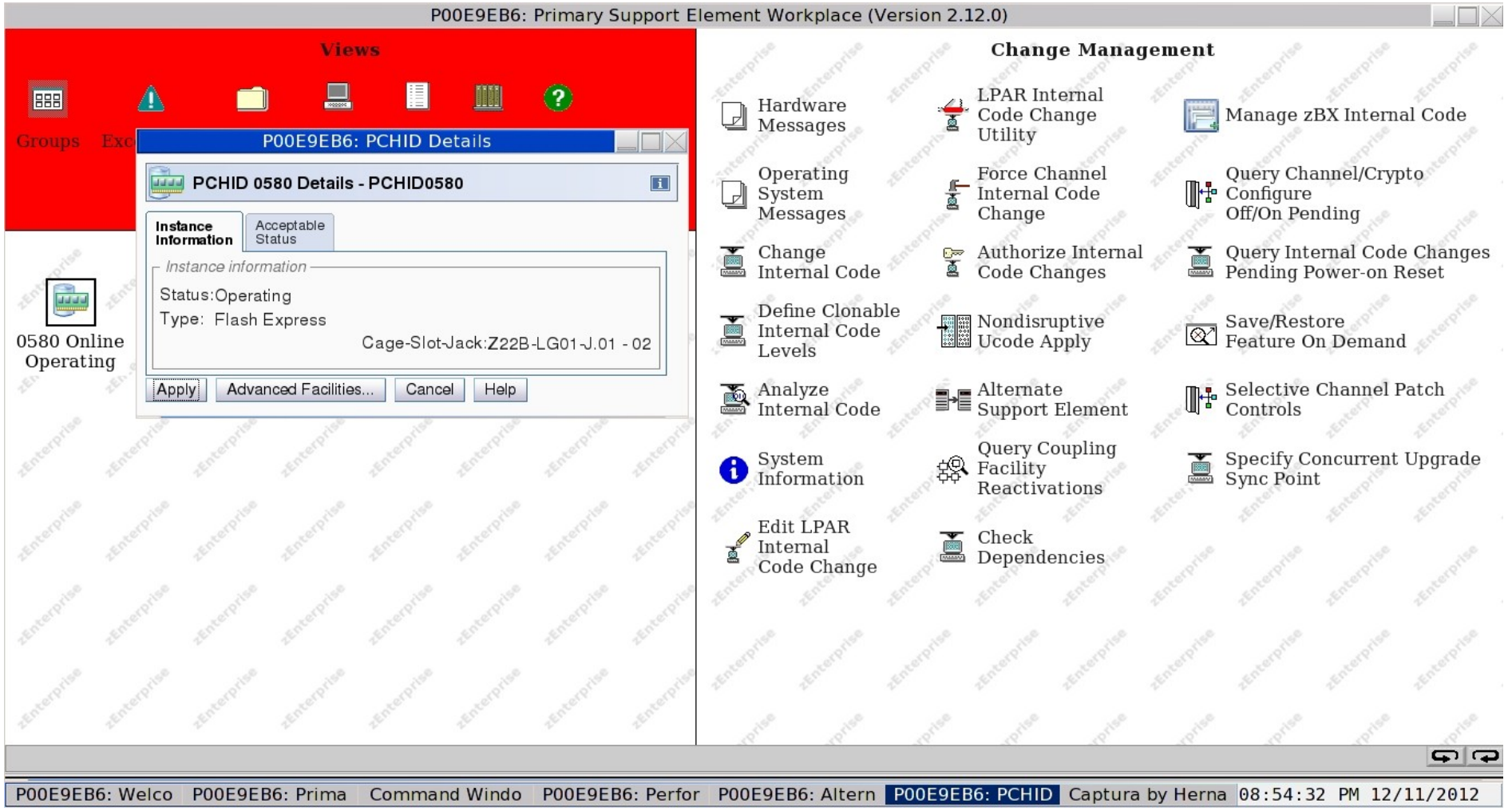
Manage Flash Allocation - View Partition to PCHID Map

- Shows information for all PCHIDs
- SE and HMC



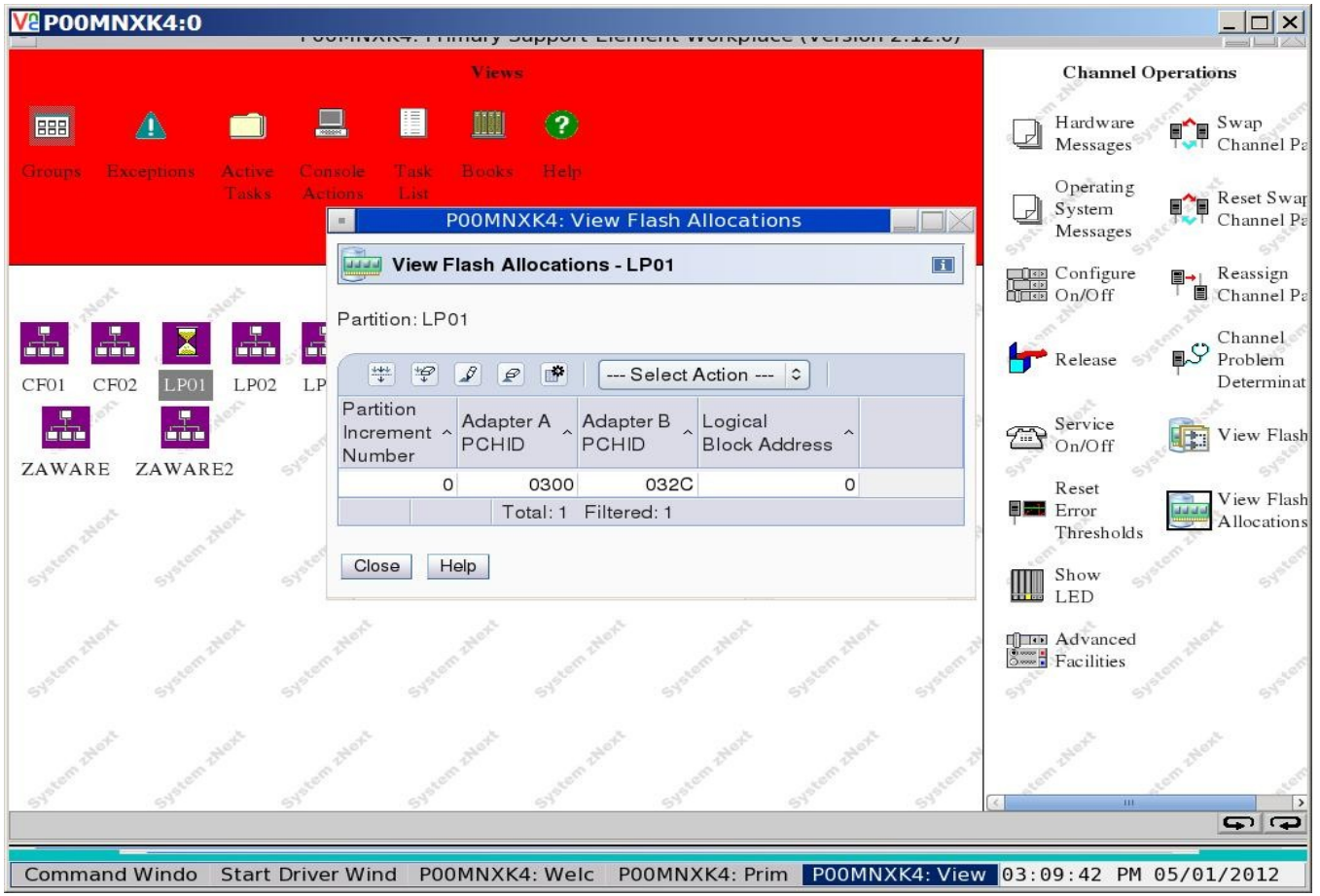
zFlash PCHID Details

- Display information for one PCHID
- SE only



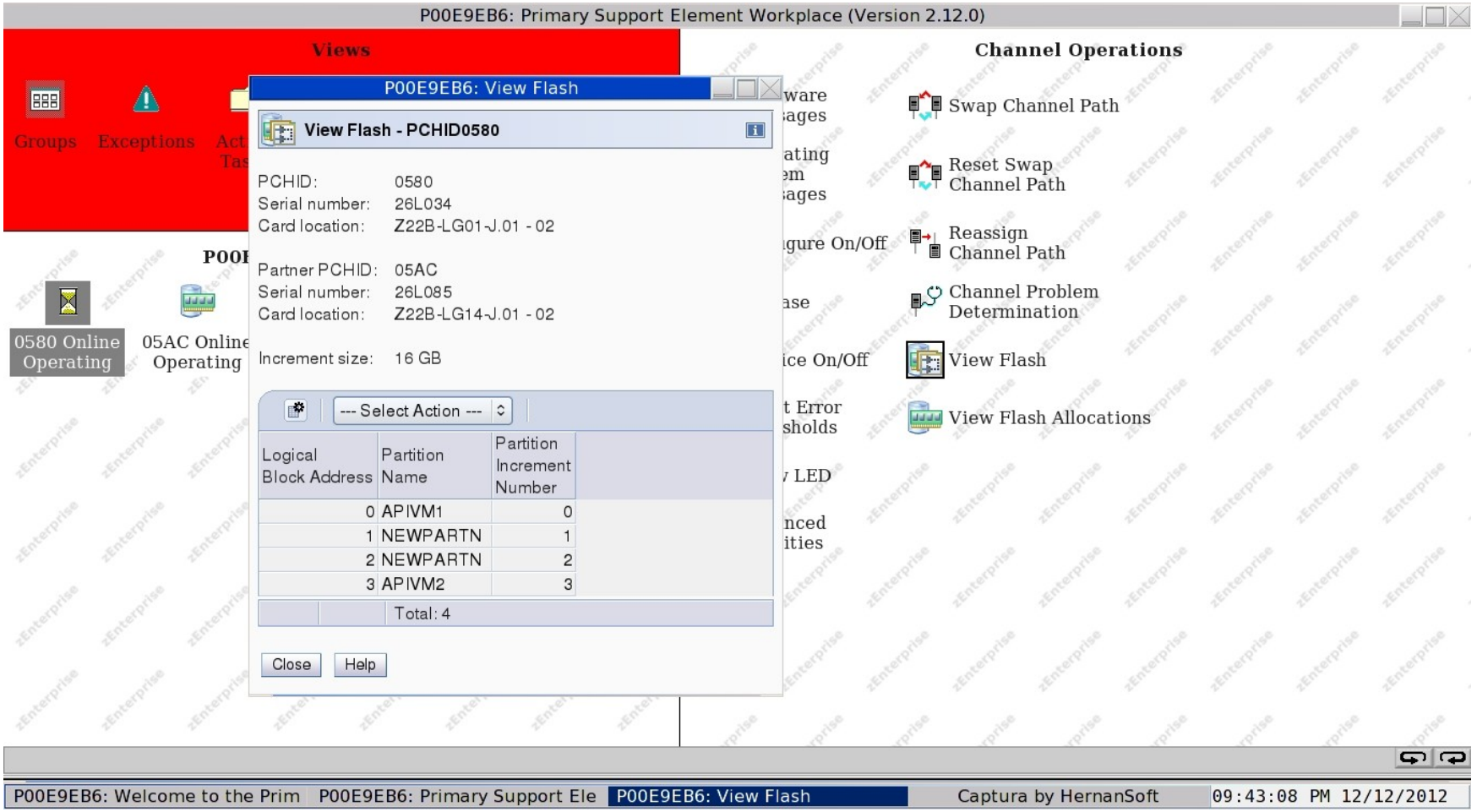
View Flash Allocations Task

- Display information for one partition
- SE only



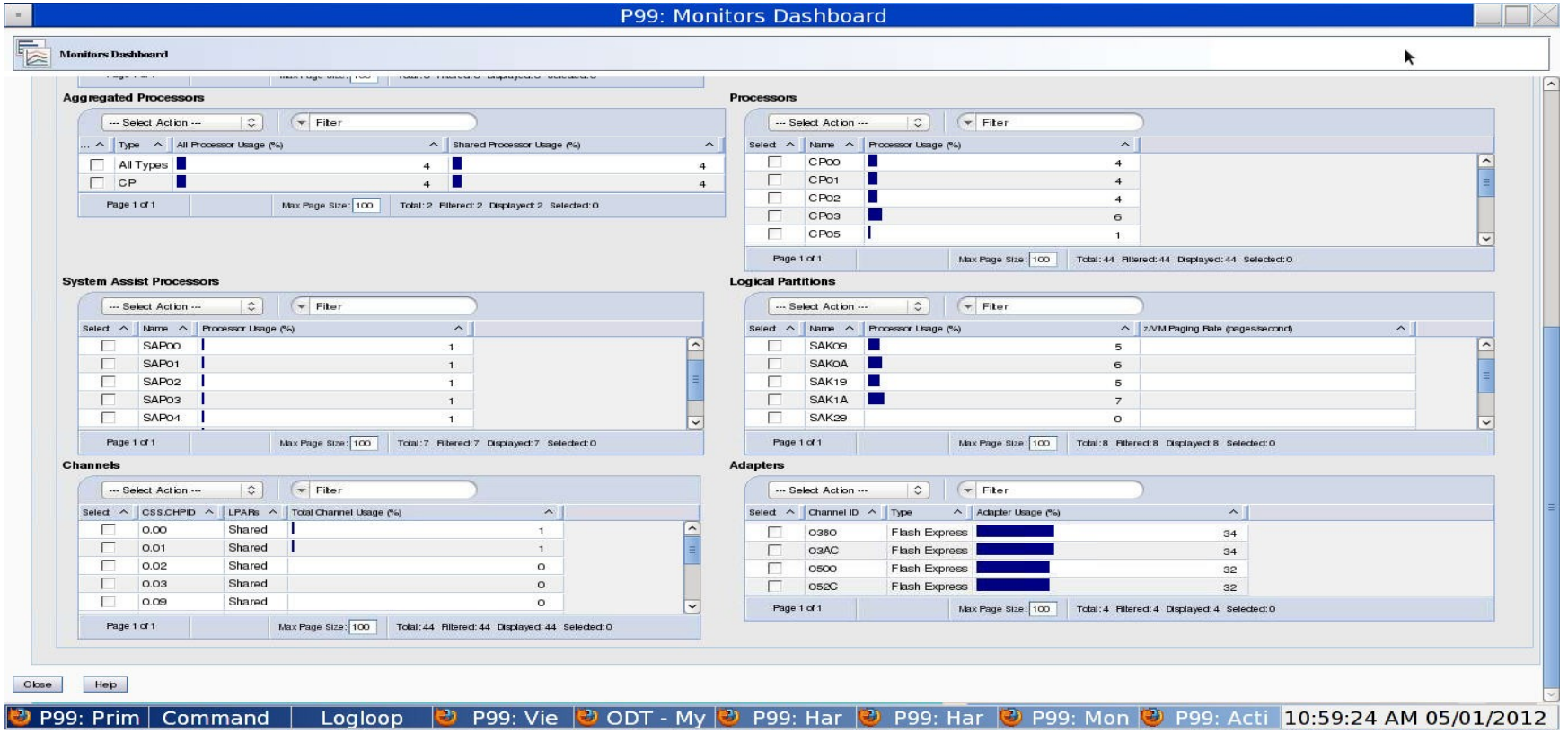
View Flash

- For the selected PCHID, shows you some physical and allocation details
- SE Only



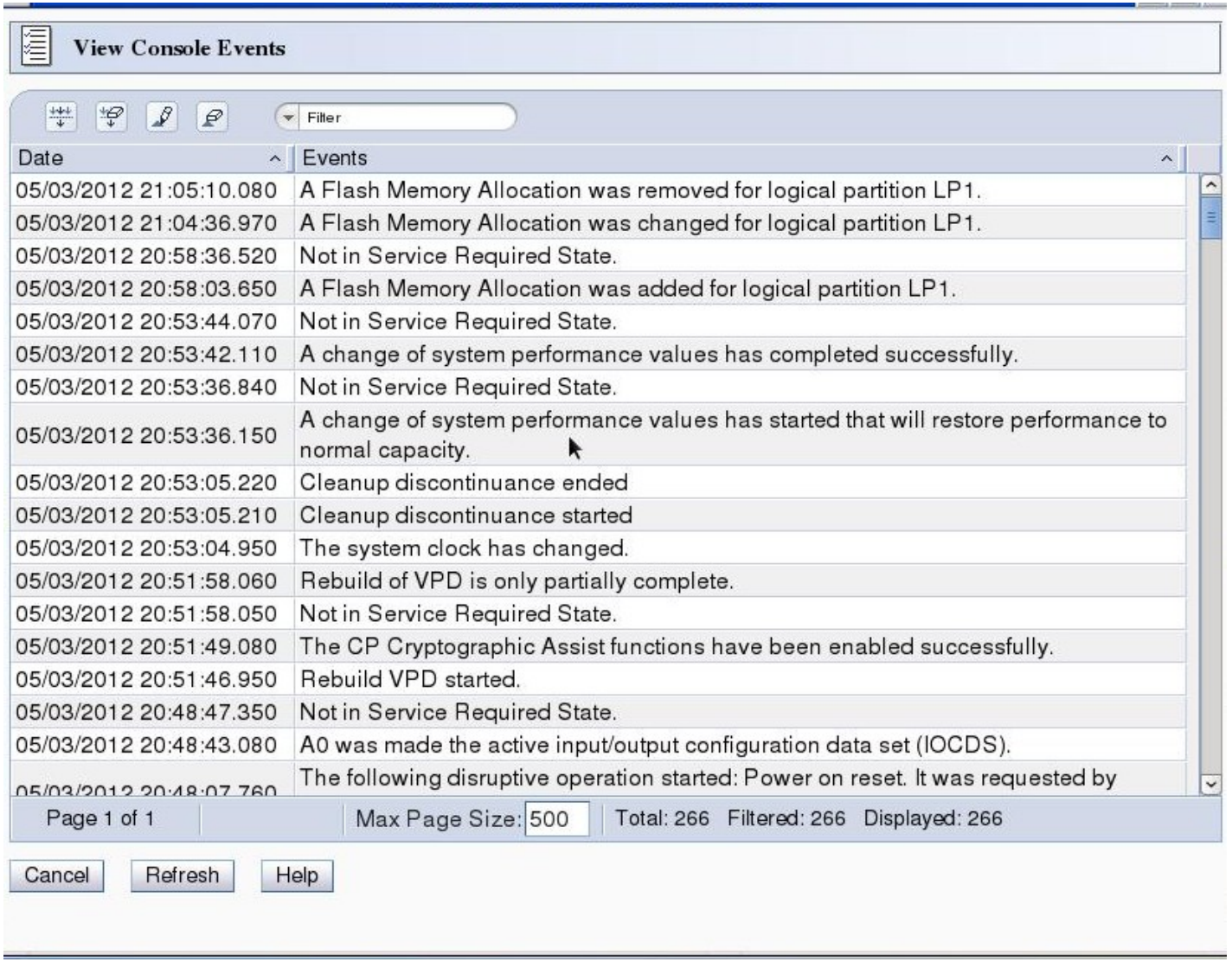
System Activity Display (SAD) / Monitors Dashboard

- Zflash is not supported by System Activity Display (SAD)
- It is supported by Monitors Dashboard. Refer to the new “Adapters” table in the lower right.



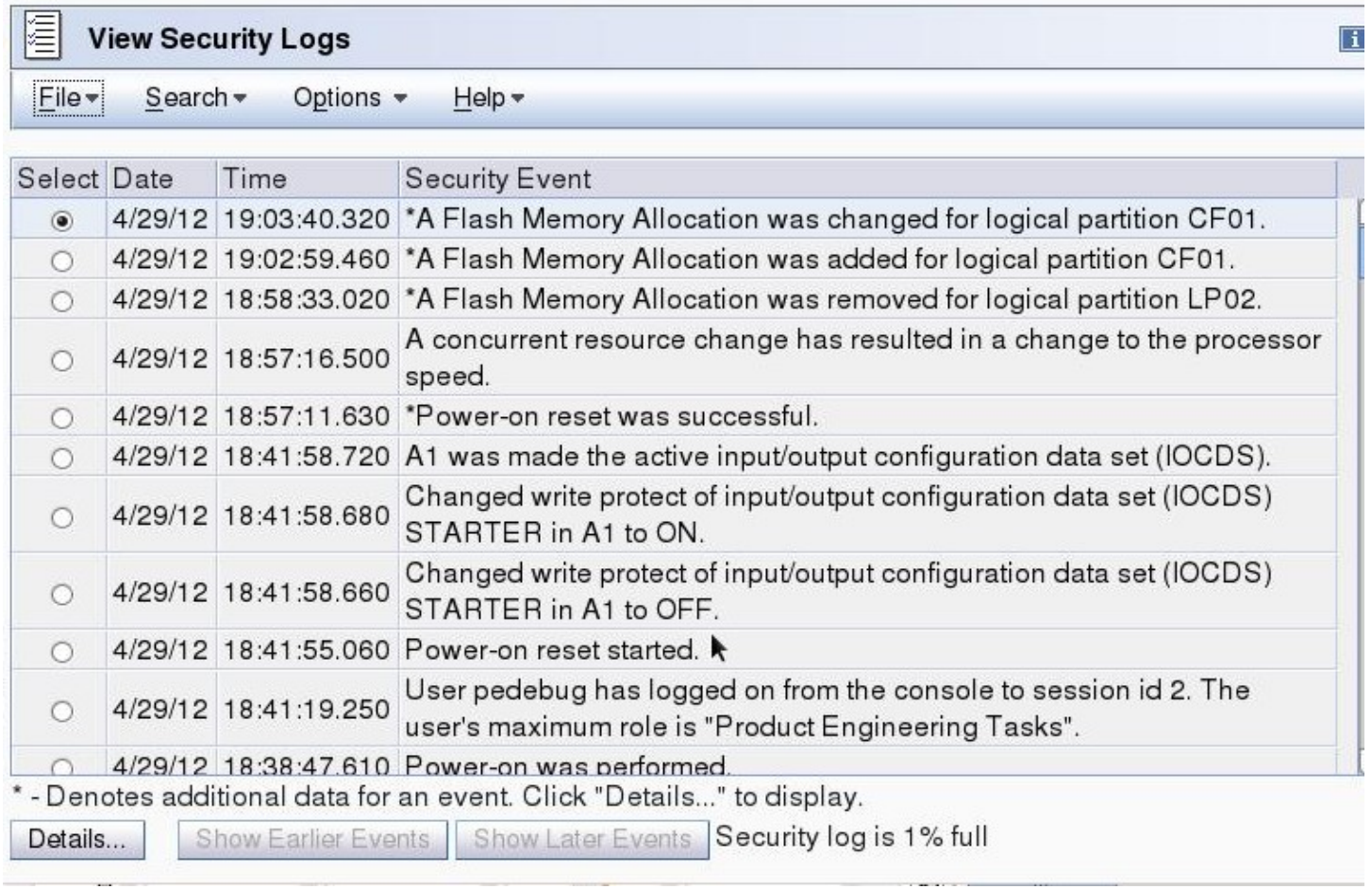
Console Events

- Event logs will be generated when a flash allocation is added, changed or removed:



Security Logs

- Appropriate security logs will be generated for zFlash-related actions.
- Examples:

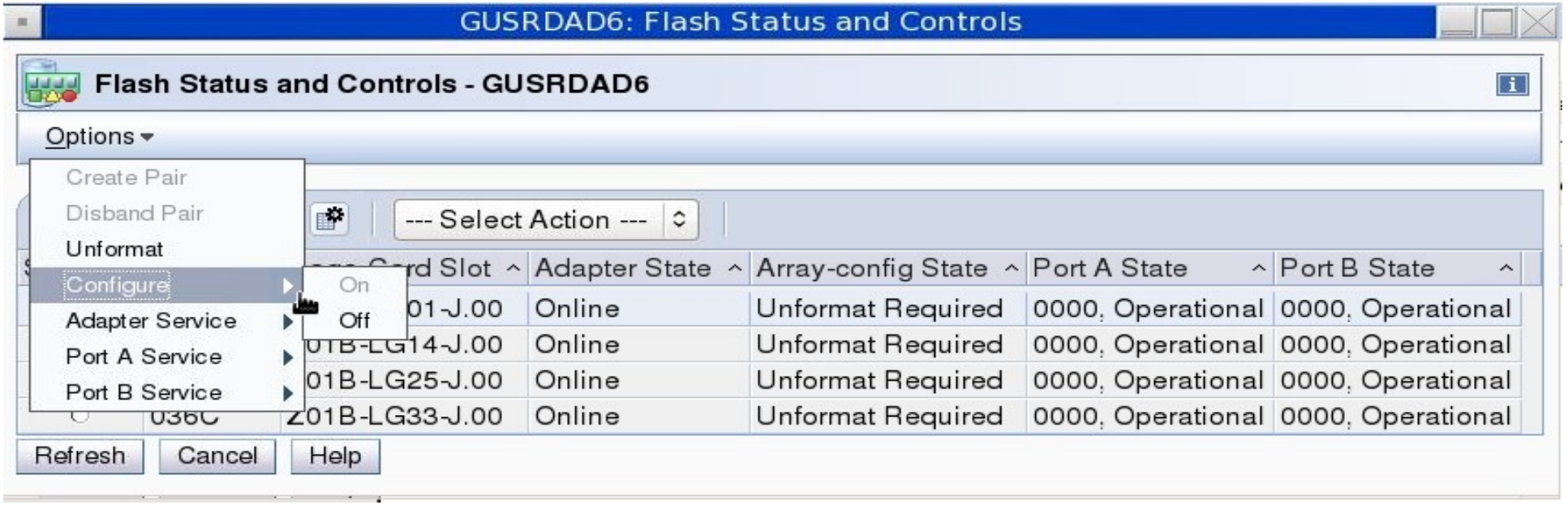


Flash Status and Controls States (Service Personnel Only)

- Adapter States:
  - Not Installed
  - Online
  - Online in progress
  - Offline
  - Offline check stopped
  - Offline in progress
  - Online check stopped
  - Service
  - Configuration error
- Array States:
  - Not formatted
  - Format in progress #% complete
  - Unformat in progress
  - Formatted
  - Configuration error
  - Rebuild in progress #% complete
  - Exposed
  - Unformat required
- Port States:
  - Unknown
  - Operational
  - Service
  - Dangling
  - Check stopped
  - Configuration error
  - Entering service mode
  - Exiting service mode
  - Repair in progress
  - Not installed

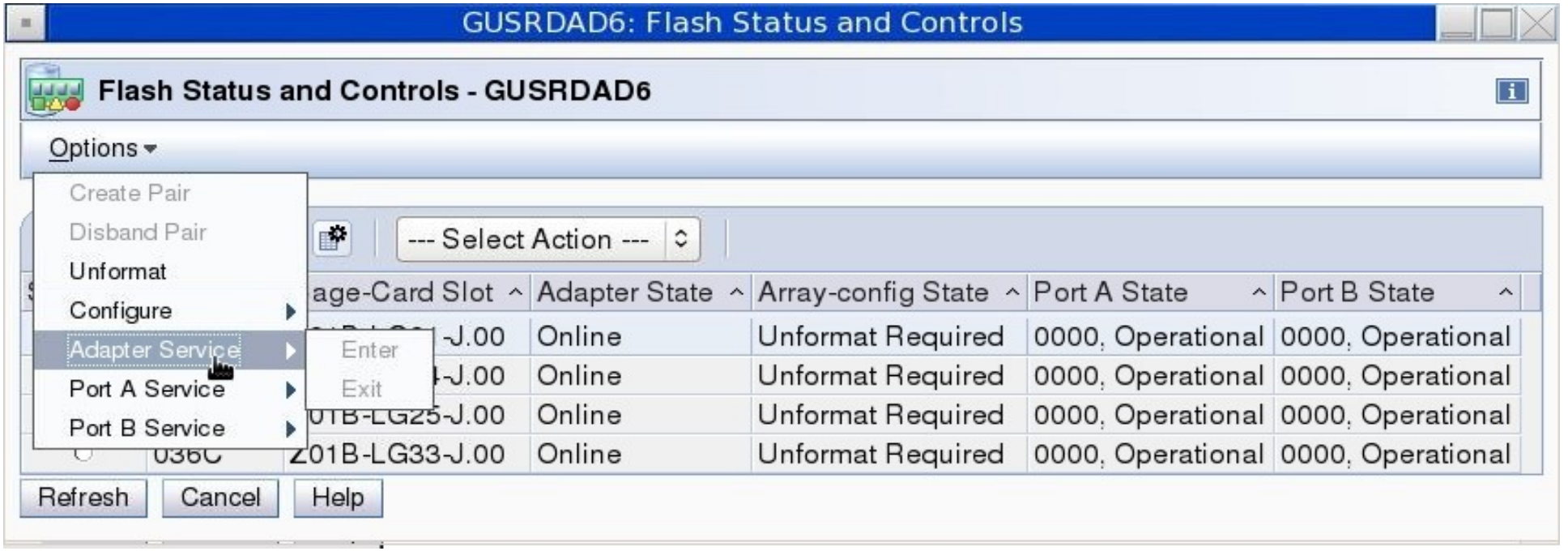
Configure the zFlash Adapter On/Off (Service Personnel Only)

Configure the adapter online or offline.



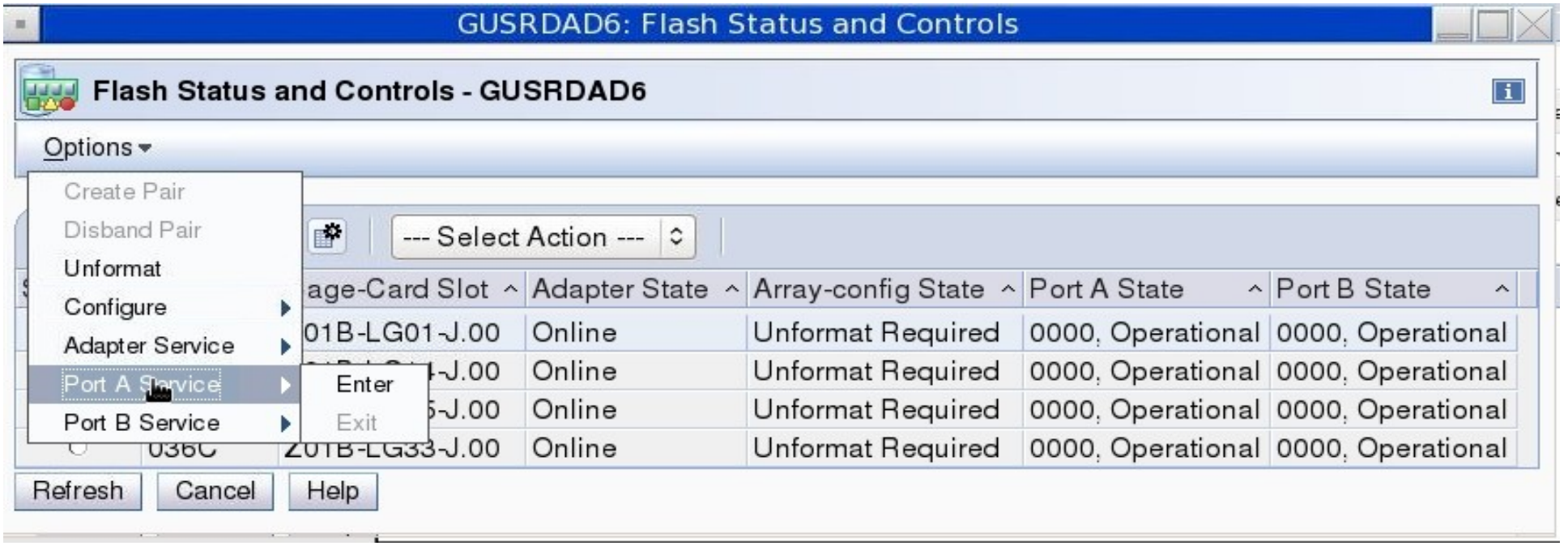
Adapter Service Mode (Service Personnel Only)

Enter adapter service mode or exit adapter service mode, they are both disabled right now because the adapter status in Online.



Port Service Mode (Service Personnel Only)

- Enter Port A service mode or exit Port A service mode, the exit is disabled currently because the port is currently in operational state.
- Port B Service behaves the same way.



You should now understand the steps to set up IBM zFlash.

Questions?

Backup Material

Registering for IBM Resource Link Access

- Registering for IBM Resource Link Access
- To view the documents on the Resource Link Web site. you need to register your IBM Registration ID (IBM ID) and password with Resource Link.
- To register:
  - Open the Resource Link sign-in page: http://www.ibm.com/servers/resourcelink/
  - You need an IBM ID to get access to Resource Link.
    - If you do not have an IBM ID and password, select the "Register for an IBM ID" link in the "Your IBM Registration" menu. Return to the Resource Link sign-in page after you get your IBM ID and password.
    - Note: If you're an IBM employee, your IBM intranet ID is not an IBM ID.
  - Sign in with your IBM ID and password.
  - Follow the instructions on the subsequent page.

Reference Documentation

- Available from “Books” group of Classic Style UI and the Welcome page of the Tree Style UI (& IBM Resource Link: Library->zEC12->Publications)
  - IBM SC28-6919: Hardware Management Console Operations Guide (Version 2.12.0)
  - IBM SC28-6920: Support Element Operations Guide (Version 2.12.0)
  - IBM SB10-7030: Application Programming Interfaces
  - IBM SC28-2605: Capacity on Demand User’s Guide
  - IBM SB10-7154: Common Information Model (CIM) Management Interfaces
  - IBM SB10-7156: PR/SM Planning Guide
  - IBM SA22-1088: System Overview
  - IBM SC27-2623 Advanced Workload Analysis Reporter (IBM zAware) Guide
- Available from IBM Resource Link: Library->zEC12->Technical Notes
  - System z Hardware Management Console Security
  - System z Hardware Management Console Broadband Remote Support Facility
  - System z Activation Profile Update and Processor Rules

Trademarks

IBM, the IBM logo, and ibm.com® are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at http://www.ibm.com/legal/copytrade.shtml.

Adobe is a registered trademark of Adobe Systems Incorporated in the United States, and/or other countries.

Linux is a registered trademark of Linux Torvalds in the United States, other countries, or both.

Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Tom Mathias

IBM
mathiast@us.ibm.com

Elpida Tzortzatos

IBM
elpida@us.ibm.com

Feb 5, 2013 - Session 13057

Please fill out the online session evaluation at either:

SHARE.org/SanFranciscoEval, or

Aim your smartphone at this QR code:

