IBM

Real Storage Manager's – Functional Evolution for Scalability and Performance

NY METRO NaSPA Meeting

New York, NY

April 24, 2007

Elpida Tzortzatos elpida@us.ibm.com

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

- DB2
- **eServer**
- **IBM**
- **IMS**
- **MQSeries**
- **MVS**
- **MVS/ESA**
- **OS/390**
- **RMF**
- **SRM**
- **System z9**
- **S/390**
- **VTAM**
- **WLM**
- **z/OS**
- **zSeries**
- **Z9**
- **Z990**

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries.

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

Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.

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

SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

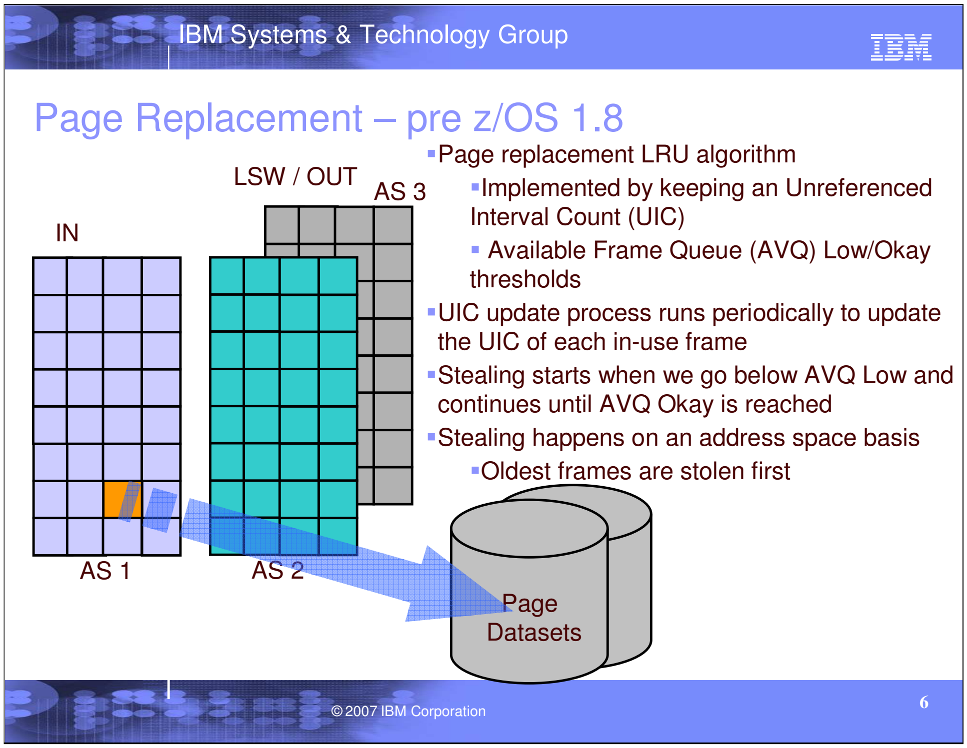
Table of Contents

Session Objectives

- **Things you will learn:**
  - This presentation will review the major RSM/SRM functions and how they have evolved over the last few years to provide for greater scalability. In particular it will cover changes made to:
    - Page Replacement and UIC
    - Physical Swap Processing
    - Pageable Storage Processing
    - Changes to RMF reports and RMF SMF records
  - Understand RSM changes made in z/OS 1.8 and what the migration issues are
  - What is being done in the performance area with Large Real Memory in z/OS 1.8
  - How to find related publications

Overview

- z/OS 1.7 supports up to 128G of central storage that can be configured to a single z/OS image.
- As processing power increases, demand for processor memory also increases. Currently the maximum storage supported by the hardware is as follows:
  - z990 - up to 256G of central storage
  - z9 - up to 512G of central storage
  - Next Generation – something larger than 512G
- The Greater than 128G Central Storage Support in z/OS 1.8 allows the operating system to support up to 4 terabytes of central storage.
  - The implementation is transparent to problem state applications and authorized programs.
  - Greater workload scalability within a single z/OS system



Currently page replacement in z/OS is done by a Least Recently Used algorithm, which keeps an Unreferenced Interval Count to signify the age of unreferenced frames. The process of selecting which pages to displace from central storage and paging them out to auxiliary storage is called page stealing/page replacement. Page replacement is accomplished by paging out to AUX the oldest frames (those with the highest UIC) owned by an address space.

The UIC update process runs periodically to update the UIC of each in use pageable frame owned by an address space.

The Real Storage Manager maintains a *free list* of page frames that are available to satisfy page faults. In RSM terms, this free list is often referred to as the Available Frame Queue (AFQ). Page frames are returned to the AFQ when storage is freed explicitly, or implicitly when a program ends. In most environments since the amount of virtual memory that is in use at any given instant may be larger than real memory, the system itself must replenish the AFQ. The AFQ replenishment is driven by the AVQ LOW and AVQ OKAY thresholds managed by the System Resources Manager. When the number of frames on the AFQ falls below AVQ LOW, SRM calls RSM to replenish the AFQ by removing some of the current page data from real memory. The system stops replenishing the AFQ when the AVQ OKAY threshold is reached. Stealing happens on an address space basis and the oldest frames from each address space are stolen first.

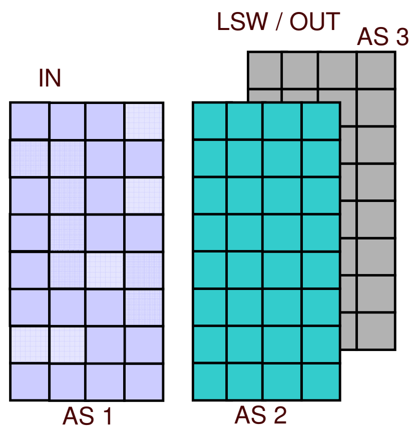
Page Replacement and MCCAFCTH values

- Frame Management/Frame Replenishment
  - MCCAFCTH=(avqlow value, avqokay value)
    - Specifies the initial low and the OK threshold values for central storage
    - Default MCCAFCTH=(400,600) without APAR OA14409
    - Defaults for MCCACFTH with APAR OA14409
      - avqlow – is the maximum of 400 and .2% of pageable storage
      - avqokay - is the maximum of 600 and .4% of pageable storage
  - SRM raises/lowers the AVQ LOW/OKAY thresholds by sampling low frame events in a measurement interval
    - Theshold values will never fall below the values specified in MCCAFCTH.
  - RSM signals SRM about availability of frames via SYSEVENTs
  - Suspendable callers are suspended when the number of available frames reaches 1/4 of the AVQ LOW threshold

IBM Systems & Technology Group

IBM

UIC Calculation in pre z/OS 1.8



- The UIC Update process is scheduled to run every 10 seconds
- This process analyzes every pageable frame from each address space on the Swapped-IN queue
- The analyzed frames are counted and are distributed into buckets depending on their age.
- The oldest frame in an address space represents the address space High UIC
- The highest address space UIC represents the Highest System UIC (MCVSTCRI)

In a pre z/OS 1.8 system the Highest System UIC gets calculated every 10 seconds. The System High UIC represents the oldest frame of all Swapped-IN address spaces. Every frame in use by each address space on the Swapped-IN queue needs to be processed in order to keep its UIC current.

While stealing happens on a demand basis, UIC update still needs to run. This process consumes a considerable amount of CPU time while holding address space locks, even in environments where the system is not paging at all. While UIC update is running for a particular address space, RSM can not perform any other functions for that address space (such as service page faults) due to the fact the UIC is holding the necessary locks.