White Paper:

z/OS 1.8 Large Memory Exploitation

An IBM and Software AG Joint Project

Author:

Marianne Hammer

Elpida Tzortzatos

Rainer Herrmann

Contents

Preface

This document captures the information gathered from an IBM and Software AG joint project, funded by IBM, in which a real memory stress test was performed. The success of this project can be attributed to the team effort between IBM and Software AG. Each team member made significant contributions in order to achieve the project's objective. Peter Harris and Dave Tipler (Software AG) and Elpida Tzortzatos and Leticia Cruz (IBM) would like to acknowledge the following IBM and Software AG team members for their outstanding efforts on this joint project:

IBM Team	Software AG Team
Tom Carielli	Peter Harris
Victor Chao	Rainer Herrmann
Leticia Cruz	Alexander Hulm
Marianne Hammer	Dan McClard
Mei Lin	Dave Tipler
John Mullin	Arno Zude
Mariama Ndoye	
Darren Swank	
Elpida Tzortzatos	
Mark Wisniewski	

Performance is based on measurements and projections using a test case for Adabas 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 real 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.

1 Introduction

1.1 Overview

In z/OS V1.8, IBM supports memory configurations up to 4 terabytes of real storage

One of Software AG's flagship products, Adabas, is a database management system capable of very high transaction rates and support for very large databases housing mission-critical applications. As a System z Development Partner, Software AG has historically been on the leading edge of exploiting new IBM hardware and software functions. Software AG's Adabas Parallel Services provides a shared global cache capability among different Adabas address spaces. Adabas Parallel Services' exploitation of 64-bit shared memory serves as a precondition of its ability to exploit large real memory configurations and will enhance its information sharing capacity among different address spaces.

The objective of this joint project is to obtain information on memory performance based on the Adabas application test cases in an Adabas test environment. Prior to the start of this project, the Adabas development team completed functional and system testing.

1.2 Adabas Parallel Services

Adabas is a high-transaction rate database management system (DBMS) from Software AG. With the add-on product Adabas Parallel Services, it runs multiple fully symmetric instances of the DBMS engine (called nucleus in Adabas terminology) in a single z/OS image. The Adabas nuclei share a global cache by using the z/OS data space functionality. Adabas Parallel Services has provided significant performance improvements by utilizing multiple processors (each Adabas nucleus runs under a single TCB only) and by reducing the amount of I/O activity because of the global cache.

However, a data space is limited to 2GB virtual addressability. Software AG has determined that its large customers often want to take advantage of the latest IBM technologies such as the 64-bit hardware and larger memory configurations to keep more data in memory. The 64-bit shared memory function meets Software AG's customers' need for larger than 2GB shared memory for the global cache. Therefore, Adabas Parallel Services was enhanced to exploit 64-bit shared memory as an option for global cache. Software AG Adabas Parallel Services V7.5 provided virtual storage constraint relief by exploiting the z/OS V1.5 64-bit shared memory function so that multiple Adabas nuclei can share greater than 2GB virtual global cache in a single z/OS system image. Since the cache is frequently referenced, it must be backed up by large amounts of real memory and z/OS V1.8 will allow for more than 128 GB of real memory.

1.3 Project Objective

z/OS 1.8 introduced a new memory management operating system, which is designed to manage single z/OS images with up to 4 Terabytes of real storage. The purpose of this joint project is to evaluate the new memory management infrastructure of z/OS 1.8 and identify any system constraints that might impede scalability or performance. Both IBM and Software AG have an interest in verifying that the new memory management infrastructure of z/OS 1.8 allows for scaling up with increased workloads and exhibits advantageous performance characteristics in single system images with large memory configurations. Software AG's high memory workload software was used in this joint test to drive performance and scalability tests on IBM's new System z9 hardware Model EC S38 running z/OS 1.8.

2 The Real Memory Stress Test

2.1 Adabas Parallel Services Global Cache

Software AG Adabas Parallel Services V7.5 will provide virtual storage constraint relief by exploiting z/OS V1.8 larger memory configurations so that multiple Adabas nuclei can share greater than 128 GB virtual and real global cache in a single z/OS system image.

Because Software AG does not have the sufficient hardware configuration for a high-volume, large 64-bit shared virtual global cache test; the stress tests were conducted at IBM's System z Benchmark Center in Poughkeepsie, New York.

2.2 Stress Test Environment and Configuration

2.2.1 Test configuration

2.2.1.1 Hardware configuration

The Benchmark Center system that was used for this stress test consisted of:

- System z9 EC S38, up to 32 CPs per a single LPAR
- 1 LPAR
- Real storage, 256 GB, of which a maximum of 234 GB were assigned to the LPAR where tests were run
- One 3490 tape drive
- DASD, 9 TB 2107-922(384 volumes) with 64GB cache and PAV
- Hipersockets between LPARs, 16K frame size and 8K TCP/IP MTU size.

2.2.1.2 Software configuration

The software products include IBM and Software AG products and one third-party product.

2.2.1.2.1 IBM products

- z/OS V1.8
- UNIX System Services, LE and C/C++ compiler available
- RACF
- RMF Monitor I, II, III
- WLM
- TCP/IP

2.2.1.2.2 Software AG products

- Adabas V8.1
- Adabas V7.4.4
- Adabas Parallel Services V7.5.1
- Adabas Transaction Generator (TXG)
- SMON, Software AG monitoring tool.

2.2.2 Measurement

The stress test measurement objective is to collect data with measurement points at different Real Storage sizes: 1) 64GB; 2) 128GB; 3) 192 GB; and 4) 234 GB.

The data is collected when the workload steady state is at about 80-90% CPU utilization. The high-level goal is to obtain measurement data for the following category:

- Based on the transaction workload drive a System z9 single system image with 16 CPs to a high CPU utilization for 5-10 minutes.

2.2.3 Test design and assumptions

2.2.3.1 Test workload

Software AG’s internal transaction generator, TXG, generates the stress test workload. The database and workload design align with the well-known TPC-C OLTP benchmark. The TXG program can be easily configured to simulate and generate workloads based on number of users, and the ratio of updates to read-only transaction for individual test cases. For each of the four Real Storage sizes (64, 128, 192, and 234 GB) three tests were run in which the TXG program simulated 400 users, 2000 users, and 4000 users, giving 12 different measurement points. The sum of the tables in the test database has more than 400 million records. Between 128GB and 192 GB, the size of the database was significantly increased by changing the amount of free space in the blocks.

2.2.3.2 Test measurement metrics

The test measurement data collected are gathered from SMF records and RMF data. The TXG and the Adabas nuclei provide some additional statistics. In general, the primary metrics are CPU consumption, I/O activities, Adabas transaction throughput, and Adabas command response time.

2.3 Stress Test Result and Recommendation

2.3.1 Test results

The actual stress tests were performed at the System z Benchmark Center in Poughkeepsie from May 22, 2006 to May 25, 2006. The following table summarizes the test case measurements.

CPU z9 738-S38

Benchmark run was using 16 CPs

One logical partition on a z9 190 with z/OS 1.8 Operating System

MIPs (16 CPs): 6413.00

Note: MIPs numbers are derived from "System z Competitive Sales Support" team of IBM Denmark.

Real Storage (GB)							
A. 64							
64x1	Users	Date	Time Interval hhs:mms – hhe:mme	Runid Ddddhhmm	CPU busy%	Total MIPs usage	ETR cmds/sec
	400	25-May	11:22 – 11:32	D1451121	48.60%	3116.72	93,178
	2000	25-May	09:58 – 10:08	D1450938	59.22%	3797.78	94,384
	4000	25-May	13:32 – 13:42	D1451312	66.24%	4247.97	93,004
Average MIPS					58.02%	3720.82	93,522

B.128							
64x2	Users	Date	Time Interval hhs:mms – hhe:mme	Runid Ddddhhmm	CPU busy%	Total MIPs usage	ETR cmds/sec
	400	24-May	16:51 – 17:01	D1441651	56.60%	3629.76	118,824
	2000	24-May	16:00 – 16:10	D1441541	78.38%	5026.51	156,731
	4000	24-May	17:53 – 18:03	D1441733	82.51%	5291.37	157,123
Average MIPs					72.50%	4649.21	144,226

C. 192							
64x3	Users	Date	Time Interval hhs:mms – hhe:mme	Runid Ddddhhmm	CPU busy%	Total MIPs usage	ETR cmds/sec
	400	23-May	15:10 – 15:20	D1431508	59.68%	3827.28	98,398
	2000	23-May	14:04 – 14:14	D1431353	79.01%	5066.91	142,726
	4000	23-May	18:47 – 18:57	D1431553	83.93%	5382.43	146,083
Average MIPs					74.21%	4758.87	129,069

D 234							
	Users	**Date**	**Time Interval hhs:mms – hhe:mme**	**Runid Ddddhhmm**	**CPU busy%**	**Total MIPs usage**	**ETR cmds/sec**
	400	22-May	15:39 – 15:45	D1421537	54.79%	3513.68	100,536
	2000	22-May	14:22 – 14:32	D1421403	83.45%	5351.65	162,000
	4000	22-May	16:51 – 17:01	D1421631	89.46%	5737.07	159,041
Average MIPs					**75.90%**	**4867.47**	**140,525**

Table 1: Test Measurement Results

Legend:

- Runid=Ddddhhmm where ddd=day of year, hh=hour, and mm=minute when test case started.
- Time interval is the measurement time period where hhs=hour when the measurement starts, mms=minute when the measurement starts, hhe=hour when the measurement ends, and mme=minute when the measurement ends.
- “CPU%” is the percentage of CPU time used in the test case.
- “ETR cmds/sec” is the number of transactions (Adabas commands) per (elapsed time) second during the steady state period of the peak workload.
- MIPs numbers are derived from “System z Competitive Sales Support" team of IBM Denmark.

2.3.2 Test results analysis

The test results can be examined in two categories: number of users and Real Storage configuration:

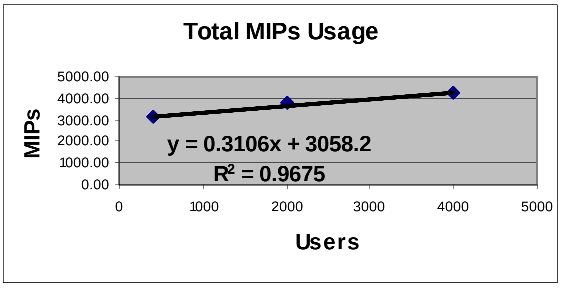
To interpret the results correctly it must be stressed that the tests with Real Storage configurations of 64 GB and 128 GB were run with a smaller database, with the intent to eliminate the potential of an I/O - 11 -bottleneck. For the Real Storage configurations of 192 and 234 GB, the major file (table) Order-Line was increased from 58 GB to 144GB by a corresponding increase of free space per block (padding factor). This file (table) is randomly accessed by the application, which explains why the ETR between 128 GB and 192 GB did not improve. Apart from that, the External Throughput Rate (ETR) increased with an increase in Real Storage.

2.3.2.1 MIPS Usage versus number of user

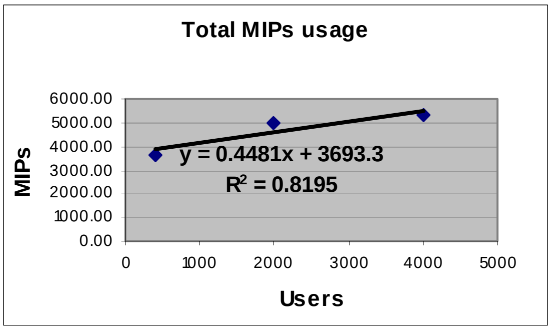
The test with 400 users served as a baseline, and ETR and MIPs increased only slightly as Real Storage increased. 400 users were insufficient to stress the system. “User think time” was set to zero throughout to raise ETR as much as possible.

With 2000 and 4000 users, demand on memory as well as ETR and MIPS usage increased as Real Storage increased. 2000 users were capable of stressing the system. With 4000 users versus 2000 users internal contentions increased only slightly and throughout the tests the results for 2000 and 4000 users were similar.

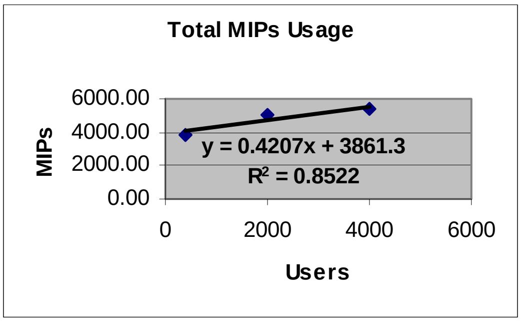
1. A Analysis (64 GB)



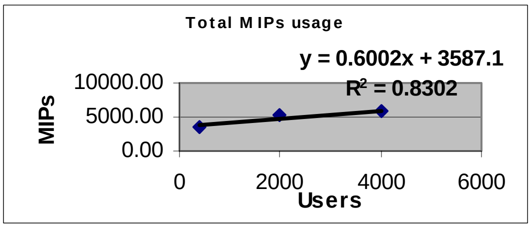
2. B Analysis (128 GB)



3. C Analysis (192 GB)



4. D Analysis (234 GB)



2.3.2.2 Impact of larger Real Storage Configurations on I/O performance

With increases in Real Storage, Read I/Os drop significantly for the workload under measurement and ETR increases, while MIPS/TR remains constant or decreases with less I/O necessary per transaction.

The results between 128 Gigabyte Real Storage and 192 Gigabyte Real Storage are not directly comparable, since the major file size was increased in the database.

Read hit rate in the cache control unit (64 gigabytes) was generally poor and, consequently, average DASD response time was quite high.

CPU% for users/real storage

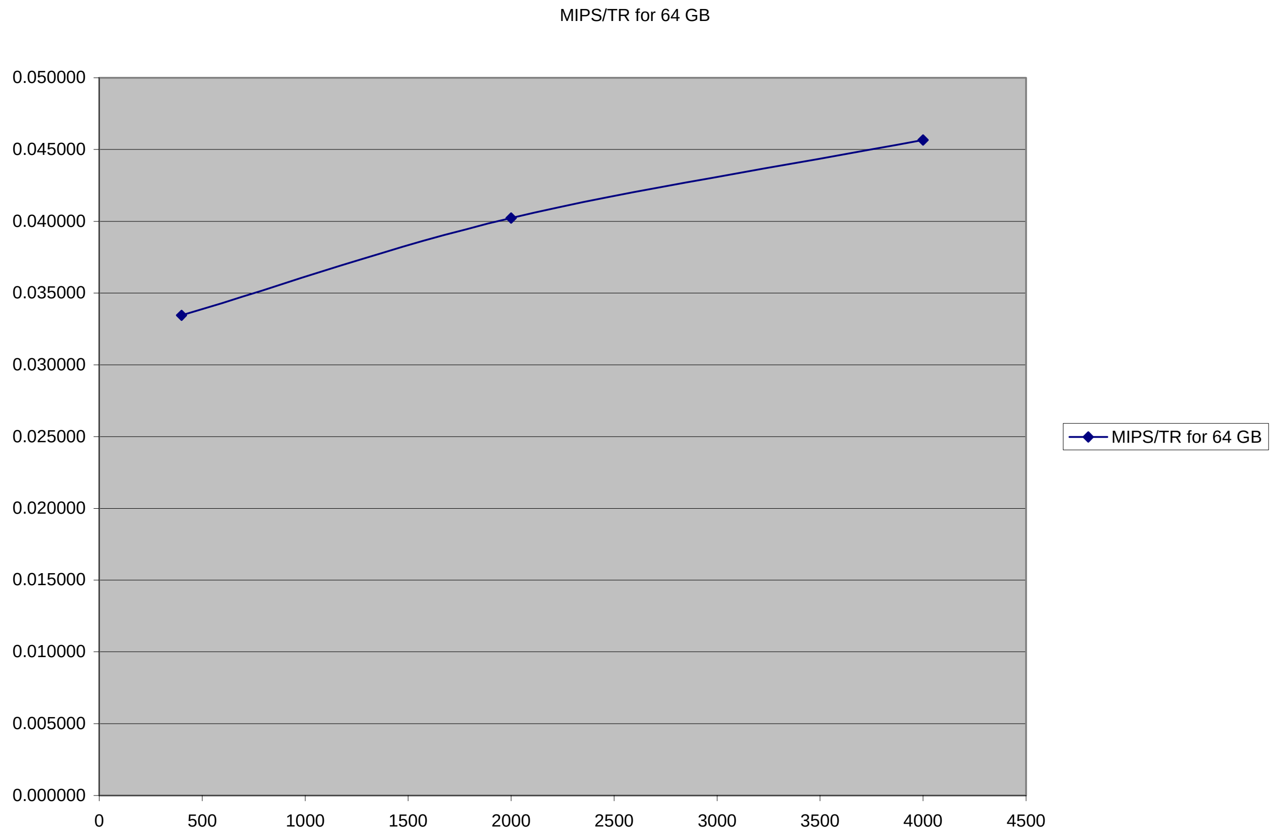
Users	CPU busy% /64GB	Total MIPs usage/64GB	ETR	DASD I/O	DASD Resp in ms	MIPS/TR
400	48.60%	3116.72	93178	17863	60.4	0.033449
2000	59.22%	3797.78	94384	16459	26.4	0.040238
4000	66.24%	4247.97	93004	17595	48.9	0.045675
average	**58.02%**	**3720.82**				**0.039787**

Users	CPU busy%/128GB	Total MIPs usage/128GB	ETR	DASD I/O	DASD Resp in ms	MIPS/TR
400	56.60%	3629.76	118824	13882	46.1	0.030547
2000	78.38%	5026.51	156731	11400	35.1	0.032071
4000	82.51%	5291.37	157123	13779	43.3	0.033677
average	**72.50%**	**4649.21**				**0.032098**

Users	CPU busy%/192GB	Total MIPs usage/192 GB	ETR	DASD I/O	DASD Resp in ms	MIPS/TR
400	59.68%	3827.28	98398	12985	40.3	0.038896
2000	79.01%	5066.91	146651	11976	28.5	0.034551
4000	83.93%	5382.43	146083	13150	16.9	0.036845
average	**74.21%**	**4758.87**				**0.036764**

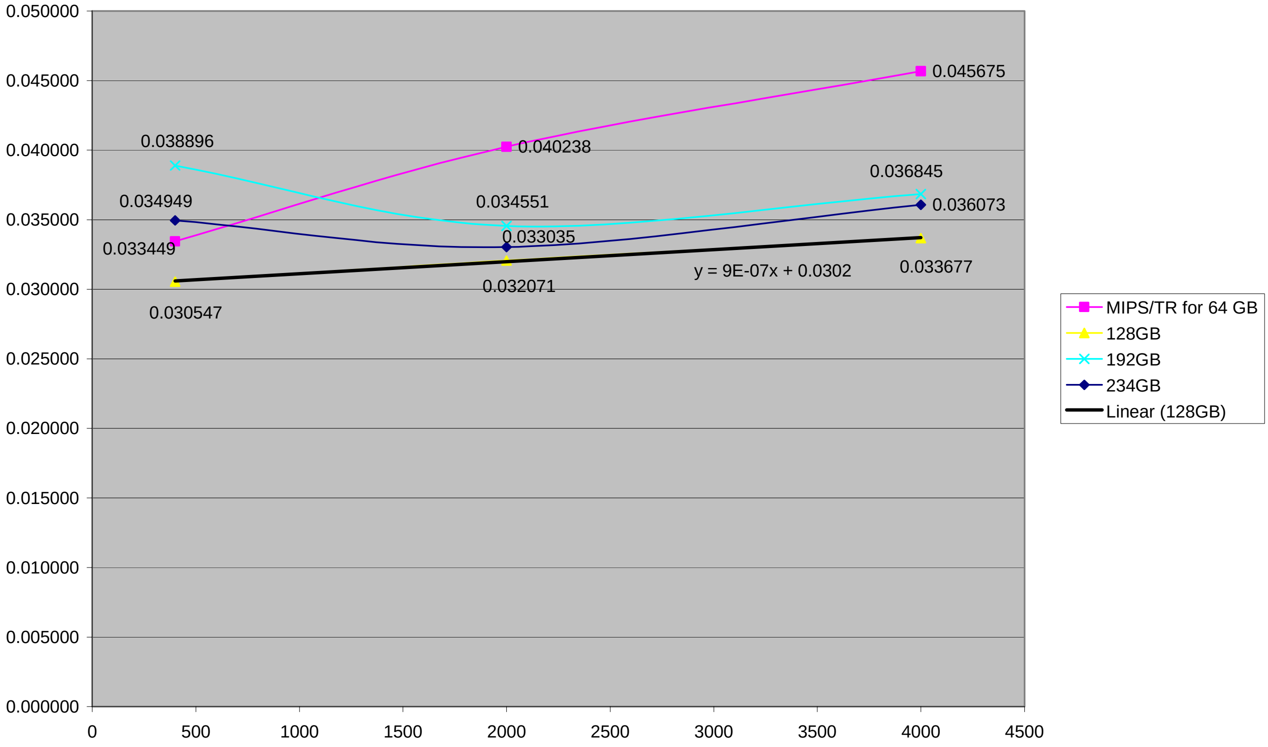
Users	CPU busy%/234GB	Total MIPs usage/234GB	ETR	DASD I/O	DASD Resp in ms	MIPS/TR
400	54.79%	3513.68	100536	13272	17.4	0.034949
2000	83.45%	5351.65	162000	11302	35.7	0.033035
4000	89.46%	5737.07	159041	13272	14.7	0.036073
average	**75.90%**	**4867.47**				**0.034686**

2.3.2.3 MIPS to External Throughput Ratio (MIPS/TR)



For the 64 Gigabyte Real Storage, the MIPS/TR increased with the number of users, while the ETR did not improve significantly. The contention for storage and the Read I/O rate was high**.**

This was different or far less pronounced for larger Real Storage configurations:



For Real Storage Configurations of 192 and 234 Gigabytes, MIPS/TR initially goes down and ITR goes up - 15 -with the number of users before increasing much more moderately thereafter, even though the database was much larger – indicating that larger Real Storage configurations may save CPU for applications with many concurrent users.

2.3.2.4 Internal Throughput Rate (ITR)

Results show that the Internal Throughput rate (ITR, or transactions per CPU busy second) increase when increasing the size of the data buffer in memory. The increase in ITR means the machine can process more transactions without increasing the overall CPU utilization, and we get a reduction in CPU-time per transaction.

The ITR results follow the same trend as the ETR results. The ETR increased by 66% and the ITR increased by 26% when going from a 36GB buffer size (with 64GB real storage) to a 100GB buffer size (with 128GB real storage). The reason for this improvement in ETR and ITR seems to be related to a 75% reduction in read I/O to DASD when increasing the buffer size. It appears this workload had a relatively large I/O content since we got such a large ITR benefit from the increased buffer size.

There is a small drop in both ETR and ITR when increasing the buffer size from 100GB to 160GB and this seems related to the fact that the read I/O rate to DASD actually increased by 4%. The data base layout was changed to add more free space per block for the tests with 192GB and 234GB real storage which may explain the small increase in read I/O when testing with 160GB buffer (and 192Gb real storage). The tests done with 36GB and 100GB buffer size cannot be directly compared to the tests done with 160GB and 200GB buffer sizes because of the difference in the data base layout.

When we increased the buffer size from 160GB (with 192GB real storage) to 200GB (with 234GB real storage) we got a 33% reduction in read I/O, the ETR increased by 11% and the ITR increased by 5%.

There seems to be a direct relationship between the amount of reduction in read I/O and the improvement in ETR and ITR for this workload. The larger increase in buffer size from 36GB to 100GB (a 2.7 times increase) resulted in the largest reduction in read IO to DASD and showed the most increase in both ETR and ITR. When we increased the buffer size from 160GB to 200GB (25% increase) we saw a smaller reduction in read I/O and we got a relatively smaller net increase in ETR and ITR.

Our analysis showed that about half of the ITR improvement was gained from a reduction in CPU time in the Adabas component and the other half from the z/OS Supervisor component, mainly related to I/O processing.

Here is an example of the relationship between ETR, ITR and the buffer size for the tests done with 2000 users. The tests with 400 and 4000 users follow a similar trend.

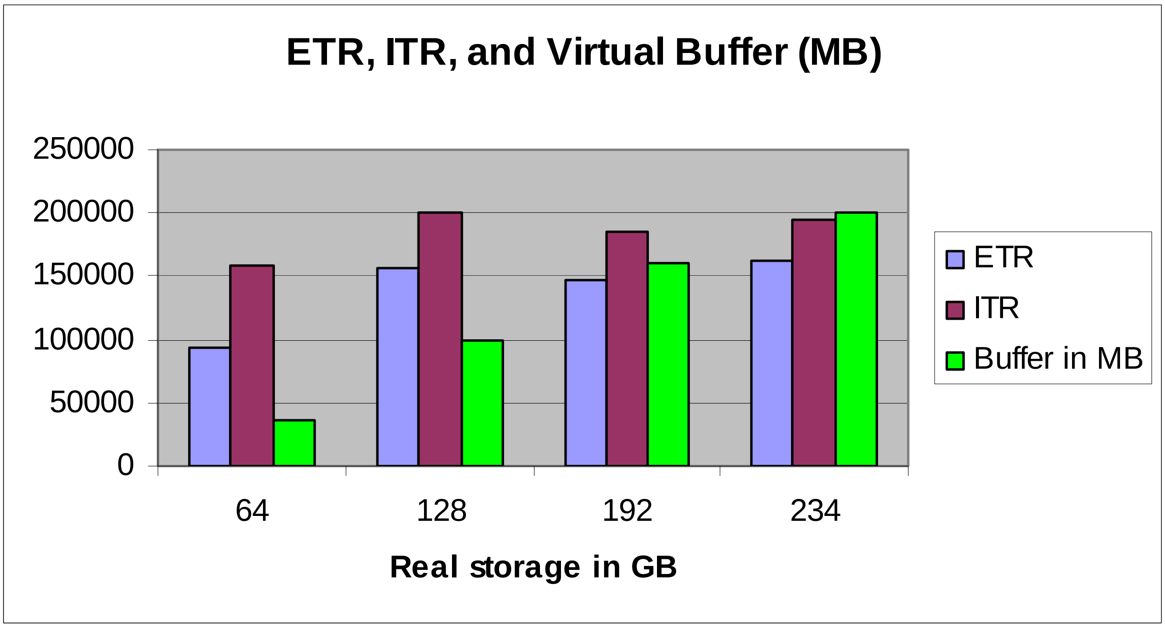
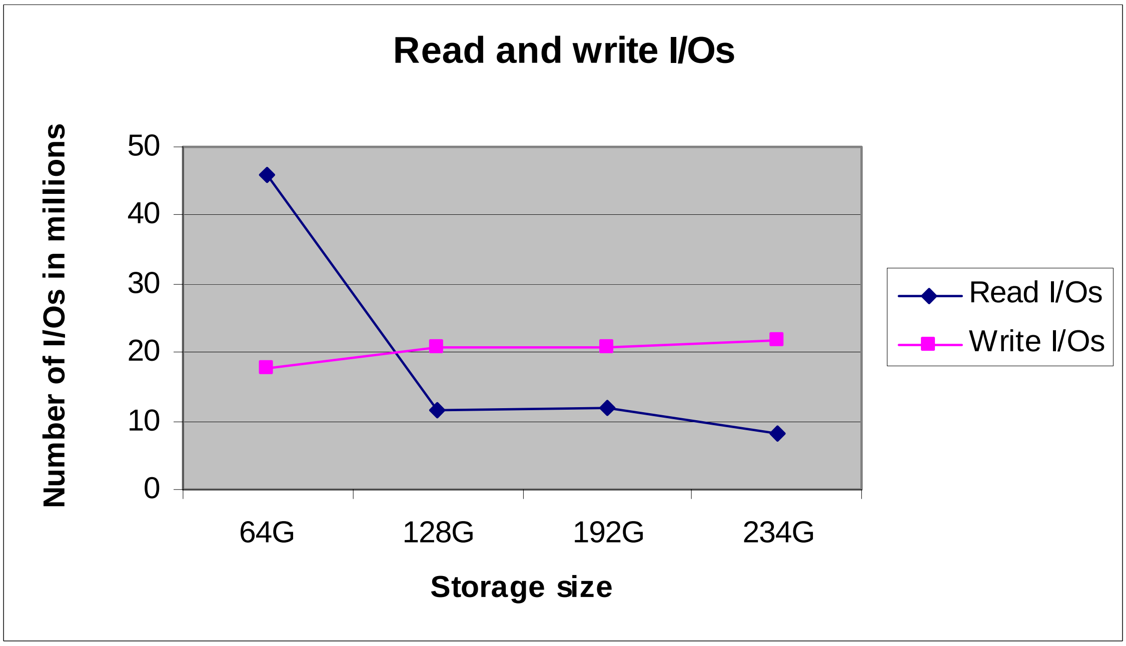


Chart showing read and write I/O rates for the 2000 user test



2.4 Conclusion

The purpose of more Real Storage is to provide the capability to define a larger buffer pool in memory in order to reduce Read I/O to DASD and therefore increase the ETR. These tests clearly show that this is what happens. It also has a moderate but significant impact on increasing the ITR for workloads that have a fairly large I/O content. In Adabas Parallel Services a shared cache space acts as the global database buffer pool.

With 64 GB Real Storage a shared cache space of 36 GB was set up. The hit rate in the buffer pool was 97.0% with 4000 users. The total number of READ I/Os were 91.4 million,

With 128 GB Real Storage a shared cache space of 100 GB was set up. The hit rate in the buffer pool was 99.3% with 4000 users. The total number of READ I/Os dropped to 21.5 million

With 192 GB Real a shared cache space of 160 GB was set up. The hit rate in the buffer pool was again 99.3% with 4000 users but this time the database was much larger. The total number of READ I/Os was 21.1 million

With 234 GB Real Storage a shared cache space of 200 GB was set up. The hit rate in the buffer pool was 99.6% with 4000 users. The total number of READ I/Os dropped to 12.6 million.

The impact on I/O performance is less clear. Of course the total number of I/Os will decrease strongly with larger memory configurations.

The remaining Read I/O from much larger cache spaces tends to have a much lower “locality of reference.” As a result, a low cache hit rate for Read I/Os was found for the cache (64 gigabytes) in the I/O control unit, and high I/O response times were observed throughout the tests. The cache in the I/O /control unit acts as a secondary cache to the primary cache of the buffer pool in real storage. After all, a Read I/O will only be done if blocks are not found in the buffer pool. Thus, for a secondary cache to be effective, it is recommended that the secondary cache be much larger than the primary cache.

Because Write I/Os due to updates are less affected by larger buffer pools and cache spaces and Write I/Os increase with an increase in ETR, the ratio of Read to Write I/Os starts to drop significantly. Scenarios with large buffer pools will see a higher percentage of Write I/Os. This will increase demand on nonvolatile storage in the cache control unit, and an increase in non-volatile storage size may further improve Write I/O throughput in large Real Storage configurations.

The result of this joint project also confirmed our original assertion that it is a win-win project for IBM and Software AG. Using Software AG’s Adabas Parallel Services, IBM was able to get positive verification on the large memory design points. Using IBM’s System z benchmark system for the stress test, Software AG was able to gain better experience to gain better experience and expertise on the workings of Adabas under heavy transaction loads. It was a great project with excellent teamwork between IBM and Software AG.

3 Supporting Material

3.1 Test Environment Diagram

3.1.1 Adabas Parallel Services components

Figure 1 depicts the Adabas Parallel Services component structure and its various caches and database relationships.

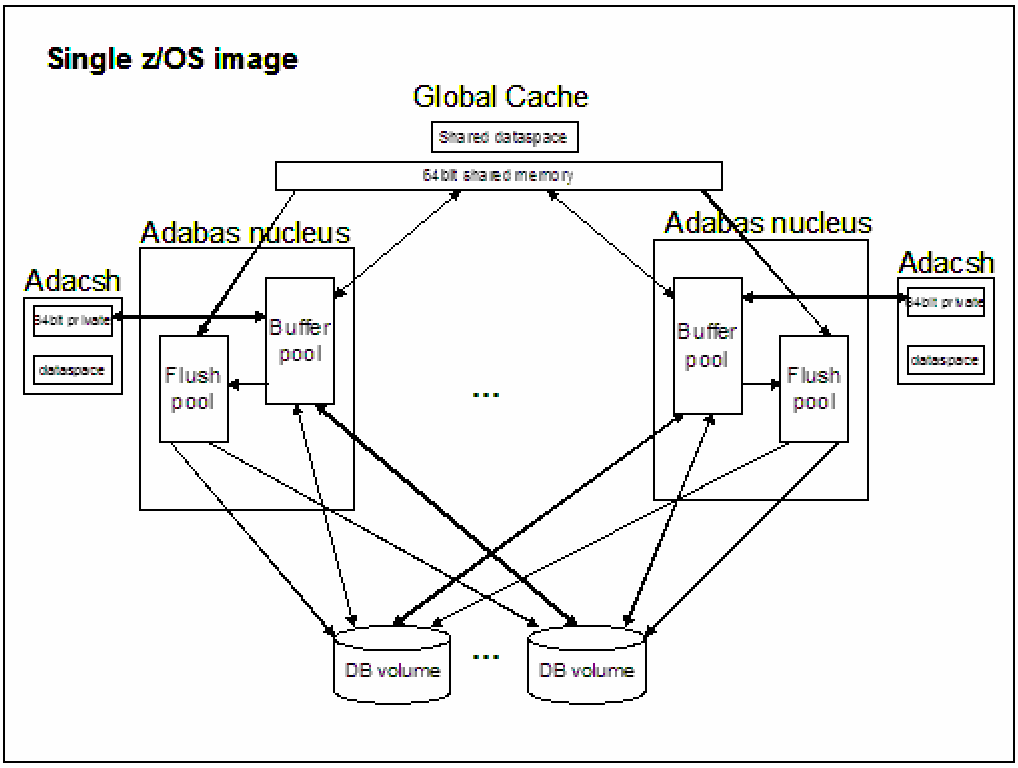


Figure 1: Adabas Parallel Services component structure

"END OF DOCUMENT"

Trademarks

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

z/OS*

System z

zSeries*

IBM*

IBM logo*

*Registered trademarks of IBM Corporation