GLOBAL PLACEMENT MODEL

Tasha Crannell and Chris Richardson
Science Applications International Corporation
10260 Campus Point Drive
San Diego, California 92121

ABSTRACT

The optimal performance of a M application is dependent upon the proper placement of globals into distinct balanced volume sets. M performance measurement tools, provide usage estimates of the real application environment on a per global basis. This paper discusses how the Global Placement Model (GPM) uses the output of these tools to estimate the demand for global resources and to balance the globals such that I/O contention for each volume set is minimized, optimizing disk I/O performance.

INTRODUCTION

The Composite Health Care System (CHCS) is a Department of Defense hospital information system. It is being deployed at over 600 military treatment facilities that vary from 1- or 2-bed clinics to larger than 1,000-bed full facility hospitals. CHCS includes both Digital Standard M (DSM) based VAX/AXP and Micronetics Standard M (MSM) based PC architectures, all running the same applications.

CHCS Applications

CHCS has a variety of applications that offer the health care providers in the hospital departments the required functionality. CHCS keeps track of the number of applications performed and I/O rates, as well as other vendor specific performance data. The number and types of physical accesses vary depending on the hospital's functionality needs. To achieve optimal performance, each hospital requires a slightly different global placement scheme to effectively balance the volume set utilization.

For a new hospital, the impact of the applications can only be estimated. However, with enough information from the existing facilities, a statistically significant estimate of the application's impact can be projected for a given workload situation.

Volume Sets

The globals are placed in volume sets to optimize the application's access to the disk I/O subsystem. The rate of access and the fragmentation of each global are accounted for in order to provide optimal I/O speed of the M application. Globals that are required to be grouped by application are placed in the same volume set. The remaining globals are systematically distributed to minimize I/O contention at the volume sets.

Paper Organization

The organization of the paper is as follows: The data sources and the storage and access of the data are described in the Global Placement Model (GPM) Inputs section. The formulation and the algorithm design to solve the GPM are discussed in the GPM section. The GPM output reports are described in the GPM Output Reports section.

GPM INPUTS

To parameterize the GPM, M application data are gathered from the CHCS hospitals. These data are stored in File Manager (FileMan) Files for use as inputs to the GPM.

Data Sources

Two data collection sources, Global Efficiency (%GE) and RTHIST, are invoked at CHCS sites in order to obtain a complete picture of the application's I/O use during the peak usage periods. An additional FileMan data file is required to ensure that those globals that must be placed together are assigned in the same volume set and those globals that must be locally mounted are placed in locally mounted volume sets.

Block Size

The unit of a M database assignment is the block. The size of the M block in both DSM and MSM is 1024 characters, which figures prominently in sizing estimates.

The Global Efficiency Utility (%GE)

%GE provides three important pieces of global information: number of pointer blocks, number of pointer levels, and number of data blocks. This utility is available on both DSM and MSM platforms. Effectiveness is a measure of the fan-out rate of the pointers to get to the data blocks. Effectiveness is defined as the number of pointer blocks divided by the number of data blocks. Since most implementations have an overhead of 2 pointer blocks (1 header block and 1 pointer block per global), 2 is the minimum number of pointer blocks defined.

The RTHIST Utility

RTHIST is a tool provided by both DSM and MSM to aid the systems manager in the tuning of the M environment to the application.

The VAX DSM RTHIST report provides data per second on numbers of logical reads, sets, and kills for a particular global during the given time frame and its 'cost' in columns. The rows of the report are global names. From this set of parameters, the reads can be extracted directly and the writes can be computed. Writes are the sum of sets and kills.

$$writes = sets + kills$$

The cost of a global to the system is a function of it's effectiveness times the sum of the reads plus 4 times the writes. M writes are complex including several processes such as garbage collection and other demon processes. The logical to physical ratio is used to provide a rough measure of the number of physical disk accesses. The writes are estimated to be 4 times the cost of a read because a write typically takes twice as long as a read, and physical writes must update both the primary and the redundant (shadow) disks. This is a rough rule of thumb that has shown excellent results when applied to the CHCS application.

The next release of DSM RTHIST (Version 6.3 c) collects the physical I/O rates for each global, but does not include this information in the standard reports.

Basic Global Configuration File

A flat file, Basic Global Configuration File, is used to identify all of the globals and their grouping. In this file, a flag is used to indicate if the global must be assigned to the same volume set as the others in it's group.

Data Storage and Access

A series of FileMan files are created to store %GE, RTHIST, and the Basic Global Configuration data. The data from the hospitals are made available as the output from these files. The access rate is expressed by the number of reads and writes to each global as determined from the above data. An estimate of the effectiveness of a global is calculated from %GE.

SYLDGE - Load Global Efficiencies

The SYLDGE utility reads the %GE output reports and extracts the globals and their volume names. It also provides the sum of blocks assigned for pointers and data, as well as the depth of the pointer structure. Separate formats are employed for DSM and MSM outputs.

SYLDGR - Load Global Groupings

The loader for the Global Groupings provides the ability to cluster globals together and apply mandatory assignments of certain globals to specific volume sets. The input file is hand built and loaded into a FileMan file. Another entity in this file is the grouping of globals into User Class Identifiers (UCIs). The two major UCI groups of globals under the DSM and MSM configurations are PRD (production) and MGR (manager). These two UCIs have linkage to all the other groupings of globals that define the different applications that make up CHCS.

SYLDRT - Load RTHIST

RTHIST provides the greatest variability of parameters, frame of reference, and format between the two implementations. Currently, VAX DSM reports global references, sets, and kills on a per global basis, but these values need to be carefully extracted from the RTHIST global summary report. MSM provides a value for the overall impact of a volume set and the percentage of impact each global has had, but no finer granularity for references, sets, or kills. A method of evaluating the MSM RTHIST global to provide similar statistics to the DSM RTHIST global statistics is being researched.

THE GPM

The GPM evaluates the data collected from the CHCS hospitals, weighs the impact of the globals, and assigns them to volume sets such that the impacts of the application I/O accesses to the globals are balanced.

The inputs are selected at the time of the model run. The set of %GE and RTHIST reports are selected from the FileMan files, and the number of volume sets to be loaded are input by the user.

GPM Formulation

The GPM is represented in a mathematical formulation. Three main criteria are considered for placing globals into volume sets:

- The reads and writes to the global,
- The levels of pointers and effectiveness in which the global can be accessed, and
- The size of the global.

The I/O rate and the effectiveness of the global are used in conjunction to determine the cost of the global to the system. Since the size of the volume set is limited, the size of the global is used to ensure the group of globals in a volume set does not violate the size limitation.

Reads and Writes

The reads value is the average number of references per minute for each global during the RTHIST collection period, extracted from the RTHIST report. The sets value is the average number of sets per minute to each global for the RTHIST collection period. The kills value is the average number of kills per minute to each global for the RTHIST collection period. The writes value is the sum of the sets and kills per minute to each global for the RTHIST collection period.

Pointer Blocks - PNTER

The pointer blocks provide the pointer tree structure for the global. The global effectiveness provides the number of pointer blocks at each pointer level and the total number of data blocks. The pointer blocks represent the structure that the first global reference must traverse to resolve a global reference. The pointer blocks usually exist in a ratio of 1:50 with the data blocks. In certain cases (very small globals), the pointer blocks may equal the number of data blocks (usually 1:1). Occasionally, the pointer levels of a global expand to three or more levels deep with many data blocks and then the data blocks are killed off except for a single data block. These cases are filtered and the effectiveness calculation is adjusted to a ratio of 1:50 pointer blocks to data blocks.

Data Blocks - DATA

The data blocks contain the data stored in each populated node of the global. Once a reference is accessed at the data level, the global structure can be accessed sequentially. The model takes a worst case approach to the data block access with the assumption that every data block access traverses all pointer block levels. M uses an "M-way, B-tree" structure.

Effectiveness - EFF

The effectiveness is a measure of the ease with which the applications access the data blocks. It is calculated by dividing the pointer blocks by the data blocks and multiplying by 100.

$$EFF = \frac{PNTR}{DATA} * 100$$

Speed of Access - SPEED

The speed is a relative estimate of the time it takes to reach a data block. It is a measure of the effectiveness times the pointer levels for the global. The greater the number of

levels, the more pointer blocks that have to be fetched to get to the data block.

$$SPEED = EFF * PL$$

Cost of the Access to the System - COST

The cost is a measure of the total impact of the global to the volume set. It is calculated from the reads, the writes, and the speed. As described earlier, the write impact is estimated to be 4 times that of a read. The cost is the reads plus 4 times the writes times the speed.

$$COST = (reads + (4 * writes)) * SPEED$$

GPM Algorithm

The GPM algorithm distributes globals across the volume sets such that the contention and response time at each volume set are minimized. The cost of each global and each group of globals is calculated as defined. The globals are then sorted in order of cost.

Initially all available volume sets are considered void of globals and have costs of zero. The global with the heaviest cost is placed in an empty volume set. The cost of the volume set is incremented by the cost of the global.

The next global with the heaviest cost is placed in the volume set with the minimum cost. The cost of this volume set is incremented accordingly. This process is repeated until all globals are placed into volume sets.

The GPM algorithm assigns the globals into volume sets such that the costs of all volume sets are evenly distributed. This results in reduced contention at the volume sets and improved access times to the disk I/O subsystem.

GPM OUTPUT REPORTS

The reports generated from the model provide the details of the global placements into volume sets. They report the input values used in determining the cost of the globals and the volume set where the globals are assigned.

The Global Assignment Report

The global assignment report provides a listing of the global entities; size in blocks, effectiveness, cost, group, and the volume set assignment. There is a flag (asterisk after the global name) to indicate that the global assignment to a group is mandatory. Figure 1 shows a sample output from the global assignment report.

Global		Volume Set	Group	Size	Effic.	Speed	Reads	Sets	Kills	Writes	Cost
%EDIHELP	*	MSTVOL	MST*	2	2	2	0	0.01	0.01	0.01	0.02
%EDT	*	MSTVOL	MST*	2	2	2	0	0.01	0.01	0.01	0.02
%ET	*	MSTVOL	MST*	2	2	2	0.06	0.1	0	0.1	0.2
%HELP	*	MSTVOL	MST*	2	2	2	0	0.01	0.01	0.01	0.02
%MENU	*	MSTVOL	MST*	2	2	2	0	0.01	0.01	0.01	0.02
%REPLACE	*	MSTVOL	MST*	2	2	2	0	0.01	0.01	0.01	0.02
%RESTORE	*	MSTVOL	MST*	2	2	2	0	0.01	0.01	0.01	0.02
%ZIS	*	MSTVOL	MST*	374	2.47	4.93	31.34	0.02	0	0.02	1.64
%ZOSF	*	LGAVOL	LG*	3	2	2	10.14	0	0	0	0.2
%ZTSCH	*	ZTMVOL	ZTM*	1017	1.4	2.79	19.79	7.12	4.29	11.41	32.4
%ZTSK	*	ZTMVOL	ZTM*	58652	2	7.99	0	0	0	0	0.01
APDS		KKKVOL	APDS	45	2.27	2.27	0	0	0	0	0.01
BDCVCK		LLLVOL	BDCVCK	12	9.09	9.09	0	0	0	0	0.01
CH		JJJVOL	CH*	695602	2.11	10.53	2.29	1.6	0	1.6	17.09
CHECK		HHHVOL	CH*	13	8.33	8.33	0	0	0	0	0.01
CHRA		LLLVOL	CH*	198	2.06	4.12	0.06	0	0	0	0.01
CHTU		GGGVOL	CH*	86240	2.09	8.36	0.18	0.02	0.02	0.04	0.35
CPA		JJJVOL	CP*	1291	2.22	4.43	0.07	0	0	0	0.01
CPC		GGGVOL	CP*	72	1.41	1.41	0.05	0	0	0	0.01

Figure 1: Sample Global Assignment Report.

The Group Assignment Report

The group assignment report provides a summary of the accumulated global size and cost of each group. A volume set report identifies the location of each global in the group. Figure 2 presents sample output from the group assignment report.

The Volume Set Assignment Report

The volume set assignment report provides the list of volume sets, the capacity of each volume set (currently defaulted to 16 gigabytes), the number of blocks consumed by all of the globals assigned, their cost to the volume set, and the list of globals assigned. Figure 3 shows a sample from the volume set assignment report.

Group	Volume Set	Size	Cost	Volume Set	Size	Cost
C*	<Total>	163	0.04	MSTVOL	11	0.01
	GGGVOL	148	0.01	IIIVOL	2	0.01
	KKKVOL	2	0.01			
CH*	<Total>	782053	17.45	GGGVOL	86240	0.35
	HHHVOL	13	0.01	JJJVOL	695602	17.09
	LLLVOL	198	0.01			
CP*	<Total>	5678	0.11	MSTVOL	3061	0.01
	GGGVOL	74	0.02	HHHVOL	106	0.01
	IIIVOL	106	0.02	JJJVOL	2327	0.03
	KKKVOL	2	0.01	LLLVOL	2	0.01
D*	<Total>	13	0.13	HHHVOL	13	0.13
DA*	<Total>	9547	0.09	MSTVOL	1381	0.01
	GGGVOL	3	0.01	HHHVOL	2	0.01
	IIIVOL	4	0.01	JJJVOL	4	0.02
	KKKVOL	2	0.01	LLLVOL	8151	0.02
DD	<Total>	10761	6.82	MSTVOL	10761	6.82
DG*	<Total>	59313	0.58	MSTVOL	59254	0.54
	GGGVOL	50	0.01	HHHVOL	2	0.01
	IIIVOL	2	0.01	KKKVOL	5	0.01
DI*	<Total>	28362	1.91	GGGVOL	2	0.01
	HHHVOL	362	0.04	IIIVOL	8067	0.02
	JJJVOL	5073	0.43	KKKVOL	4602	0.74
	LLLVOL	10256	0.66			
DIAU	<Total>	1062387	0.75	JJJVOL	1062387	0.75
DIC	<Total>	38988	2.73	JJJVOL	38988	2.73
DIET	<Total>	547700	1.75	MSTVOL	547700	1.75
DOD*	<Total>	12071	1.3	MSTVOL	10880	0.02
	GGGVOL	120	0.01	HHHVOL	4	0.02
	IIIVOL	66	0.03	JJJVOL	2	0.01
	KKKVOL	5	0.02	LLLVOL	994	1.19

Figure 2: Sample Group Assignment Report.

Generating the Translation Table

The translation table for a given site enables an application to locate a M global. CHCS references the translation table to determine the actual UCI and volume set of the global.

The outputs of the GPM are used to generate the translation table.

Volume Set	Capacity	Used	Cost	Global	Global	Global	Global	Global	Global
AAAVOL	10400000	1708215	98.13	PS					
BBBVOL	10400000	2159302	86.98	PSRX					
CCCVOL	10400000	29470	50.28	LA					
DDDVOL	10400000	10274	44.71	SDAV					
GGGVOL	10400000	510560	35.18	GTTL	SC	CHTU	XUSEC		
				DIJ	SDE	XMBS	CPXREF		
				DACNVT	DGDRG	DIROU	DODP		
				GSSY	GTARY2	H3QADM	[H3QUTIL]		
				JASBUG	LRZCYER	RZDUPBG	MIMI		
				PSDDIG	[PSDFDB]	[SAVE]	XUSPLCVT		
				XQH	[ZTJMVD]				
HHHVOL	10400000	4132370	35.18	INRHB	RT	RTV	RABTCH		
				DWA	NST	DIW	GTPSADR		
				GSFUN	[GSRFW14]	GSUTIL	GTGARY		
				INTHERL	INVD	[LAZBG3]	LRZDUP3		
				ORO1	ORTSK	PSDARE	PSI		
IIIVOL	10400000	2743415	35.18	ORB	RARPT	SDPT	ORD		
				PSDRUG	INRHT	LAZHIT	DODN		
				CPAM	CPT	DACNV	DGCONERR		
				FHRD	GSOR	[GSSC1]	GTARY		
				INLHV	INTHL7S	[IOCXLG]	LRZ		
				OPR	ORO4	PSD1	PSDDIE		
				SYPF	VVTRPT	XUB	XAPPLOT		
OREVOL	10400000	6586797	100.22	OR					
ORTVOL	10400000	5868549	119.61	ORT					
XMBVOL	10400000	1141616	45.58	XMB					
ZTMVOL	10400000	119340	64.83	%ZTSCH	%ZTSK				

Figure 3: Sample Volume Set Assignment Report.

CONCLUSION

SAIC has applied the GPM to several CHCS hospitals. Initial results indicate significant improvement in the response times of the M disk I/O subsystem. By using the data available in the %GE and RTHIST utilities, as well as applying the techniques presented in this paper, the user can greatly improve the effectiveness of M applications.

ACKNOWLEDGMENTS

Particular thanks are due to Jessica Reed of SAIC for providing the initial concept of the GPM and assisting in the problem formulation.

We are grateful for Patrick Lanclos of SAIC for contributing to the understanding of global placement and refining the GPM. Mr. Lanclos was responsible for insuring that GPM was successfully implemented at CHCS hospitals.

ADDITIONAL INFORMATION

P. Gill, W. Murray, and M. Wright, Practical Optimization (San Diego, CA: Academic Press, 1981).

SAIC, "VAX Systems Architecture Guide: Supplement to the Computer Operation Manual" (SAIC, Health Care Technology Group, 1994).

H.P. Williams, Model Building in Mathematical Programming (West Sussex, England: John Wiley & Sons Ltd., 1990).