

# Converting iSeries Disk Arm Requirements Based on Protection Type

Clark Anderson  
I/O Performance  
IBM Rochester Lab  
January 17, 2005

More server system users are realizing that different data storage disk device (disk unit, or Direct Access Storage Device, or DASD, or arm, or drive) protection levels, available with IBM iSeries™ and AS/400™ systems, provide different benefits from one another. The various protection levels are provided by different RAID (Redundant Array of Independent Disk) configurations. Certain applications and system management requirements benefit more from some types of RAID than other types of RAID. Because of this, there are more upgrades/changes being proposed in the field that involve switching disk RAID protection levels for some, or all, disks attached to a given server. This has led to an increased number of people asking how they can determine how many disks to configure with a different RAID level than they are currently using, while maintaining acceptable disk I/O subsystem performance. The intent of this paper is to help guide people responsible for determining how many drives are needed, make choices that will meet capacity and performance objectives.

## Table of Contents

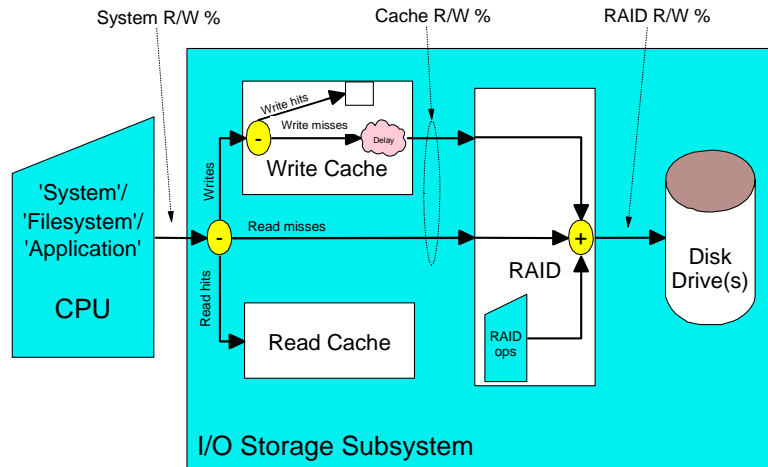
<i>1 Background information</i> .....	2 .....
<i>2 Relationship to Other Sizing Tools</i> .....	2 .....
<i>3 Various RAID Types Available For Use</i> .....	3 .....
<i>Unprotected (RAID 0) Environments</i> .....	3 .....
<i>Mirrored (RAID 1) Environments</i> .....	3 .....
<i>Parity (RAID 5) Environments</i> .....	3 .....
<i>Parity (RAID 6) Environments</i> .....	4 .....
<i>4 Method to Convert Number of Disk Drive Requirements</i> .....	5 .....
<i>5 Author Contacts and Additional Information</i> .....	9 .....

# 1 Background information

Typical iSeries I/O subsystem designs include features that alter workloads sent to the physical disk drives as compared to the I/O request workload the system sends to the I/O subsystem. The two largest contributors to workload alteration are data caches and RAID data protection schemes. The following diagram shows the request, or command flow, through a typical iSeries I/O disk drive storage subsystem and the components that can change the workload. In this case, the ‘workload’ is defined by the ratio, or mix, of read versus write requests, thus ultimately impacting the number of requests the disk drives are sent assuming a given system requested throughput (io/s) required by the system/filesystem/application(s).

The concept that not only capacity, but also speed, needs to determine the number of arms is premised on the notion that “a certain number of disk arms is needed for optimum performance on each processor level. This number is independent of the quantity of drives needed to meet the desired storage capacity.”<sup>1</sup>

## DASD I/O Command Flow



You can read the white paper, if available, that quote was gathered from to obtain more detailed information about this topic. But the third section of this paper, containing excerpts from that paper, suffices for serving as a background for understanding how to convert the number of drives from one protection type to the required number of drives of another protection type.

WARNING: The process described here should never be used alone to make purchase decisions.

# 2 Relationship to Other Sizing Tools

One key tool to use in helping determine how many drives are required is named “Workload Estimator”. Workload Estimator assists IBM approved IBM Business Partners, and customers in projecting a model that meets capacity requirements within CPU % utilization objectives. The *IBM eServer Workload Estimator* is available online at <http://www.ibm.com/eserver/series/support/estimator>.

Workload Estimator now has the ability to convert the number of drives required to meet speed, or performance, objectives for all supported RAID types. Workload Estimator has algorithms based upon those described in section 4. Thus if Workload Estimator is used, there is no extra work required to take advantage of the process defined in section 4 of this paper. If not using Workload Estimator, or another sizing tool that properly accounts for the affects of caches and RAID, you can apply the algorithms defined in section 4 to a current system with acceptable performance characteristics to project the number of drives for a new system with identical technologies and workload, but with a different storage protection level.

<sup>1</sup> “iSeries Disk Arm Requirements Based on Processor Model Performance” located on the web at <http://www-03.ibm.com/servers/eserver/series/perfmgmt/diskarm.html>

---

## 3 Various RAID Types Available For Use

---

### 3.1 Unprotected (RAID 0) Environments

For disk I/O subsystems that do not employ any protection, system I/O requests are often translated into identical, or similar, requests sent to the individual disk drives attached to the I/O subsystem. This is couched with the word “often” because there could be, and are, in typical iSeries I/O subsystems, features such as caches and operation reordering algorithms that can alter the Read/Write mix, transfer lengths and temporal & spatial locations (i.e. all aspects that define a workload) of the requests actually sent to the disk drives as compared to the request workload sent to the I/O subsystem from the system.

The algorithms in section 4 refer to this level of ‘protection’ as RAID 0. Even though there is no redundancy employed, the iSeries Single Level Storage architecture allocates storage across multiple disks, effectively ‘striping’ the data, which is recognized as “RAID 0”.

.

---

### 3.2 Mirrored (RAID 1) Environments

In a mirrored or RAID 1 environment, system requested writes must occur on both disk drives of a mirrored pair. This increases the number of disk writes that are generated. Therefore, in a mirrored environment, the number of arms required must be altered by the change in disk requests that are disk writes.

For example, assume a system with a system I/O read-to-write ratio of 3-to-1. This means the system does 3 (75%) reads for every 1 (25%) write, which totals 4 (100%) disk requests. In a mirrored environment, an extra disk access is required for every system write request. This results in 5 disk requests occurring (3 reads plus 2 writes). This is 25% more activity than for an “unprotected” environment. Therefore 25% more disk drives may be required to support that workload when mirrored, than when unprotected.

Since i5OS systems generate the redundant write requests ‘before’ the caching functions in the I/O subsystem, the writes generated to both disk drives of a mirrored pair are included in the System R/W %. In order to simplify the method used to compute the converted number of disk drives, if the ‘current’ protection level is RAID 1, the System R/W %’s are normalized to R/W ratios equivalent to all other protection levels, by reducing the number of writes by a factor 2.

---

### 3.3 Parity (RAID 5) Environments

In a RAID 5 environment, (sometimes referred to in iSeries literature simply as ‘RAID’ because other RAID protection levels were not initially available when RAID 5 was introduced) writes must occur not only on the drive containing the customer data, but also on another drive in the array containing parity information. Not only must an extra drive write be performed but 2 extra drive reads are also performed, to each of the drives containing customer and parity data. This increases the number of potential disk drive requests for every system I/O write request by 3. Therefore, in

RAID 5, the number of drives required must be altered proportionally by the changed number of drive accesses.

For example, assume a system has an I/O system workload read-to-write ratio of 3-to-1. This means the system does 3 reads (75%) for every 1 write (25%), which totals 4 requests (100%). In a RAID 5 environment, 3 extra drive requests are required for every system write request, thus 7 disk requests occur (5 reads plus 2 writes). This is 75% more activity than for an unprotected configuration. To support that extra disk activity, 75% more disks may be required for a RAID 5 configuration than for a RAID 0 configuration. Again, we couch that with the word 'may', because the controller caches may alter that increase.

---

### **3.4 Parity (RAID 6) Environments**

In a RAID 6 environment writes must occur not only on the drive containing the customer data, but also on 2 other drives in the array containing parity information. Not only must 2 extra drive writes be performed but 3 extra drive reads are also performed. This increases the number of potential disk drive requests for every system I/O write request by 5. Therefore, similarly to RAID 5, in RAID 6 the number of drives required must be altered proportionally by the changed number of drive accesses.

For example, assume a system has an I/O system workload read-to-write ratio of 3-to-1. This means the system does 3 reads (75%) for every 1 write (25%), which totals 4 requests (100%). In a RAID 6 environment, 5 extra drive requests are required for every system write request, thus 9 disk requests occur (6 reads plus 3 writes). This is 125% more activity than for an unprotected configuration. To support that extra disk activity, 125% more disks may be required for a RAID 6 configuration than for a RAID 0 configuration. Again, we couch that with the word 'may', because the controller caches and actual workload R/W ratio may alter that increase.

---

## 4 Method to Convert Number of Disk Drive Requirements

The process to convert the number of required arms from one protection level to another involves 4 steps. Those 4 steps are...

1. Gather input data.
2. (a) Derive normalized, if RAID 1, and intermediate R/W ratios associated with functions located in a typical iSeries I/O subsystem. Then (b) 'reverse engineer' the minimum number of disk drives needed to meet speed requirements for the 'current' system configuration assuming no protection.
3. Apply the intermediate R/W ratios derived in step 2 for the 'new' system configuration to the 'unprotected' required value also computed in step #2 to determine the minimum number of drives needed to meet speed requirements.
4. Select the larger of either the minimum number of disk drives needed to meet capacity (from step #1) or speed requirements (from step #3).

The inputs required are:

1. The RAID protection level of the 'current' system configuration.
2. The RAID protection level of the 'new' system configuration.
3. Minimum number of disk drives needed to meet capacity requirements of the new system configuration.
4. Minimum number of disk drives needed to meet speed requirements of the current system configuration.
  - The number of disk arms required to support a given workload on any CPU model is affected by the combination of controller (IOP/IOA) and DASD features selected. A method, outside the scope of this paper, must first be used to determine this value. It is highly recommended that either information from an existing system, assumed to be performing satisfactorily with respect to speed, or Workload Estimator (with RAID 5), be used.
5. Controller read cache efficiency.
  - For existing iSeries systems you can get this value directly from the 'Cache hit Statistics, Controller Read' column in the 'Disk Activity' section of a Performance Tools 'Component Report'.
  - If this data is not available, either use a best guess or a default value of 0%.
6. Controller write cache efficiency.
  - For existing iSeries systems you can get this value directly from the 'Cache hit Statistics, Write Effic' column in the 'Disk Activity' section of a Performance Tools 'Component Report'.
  - If this data is not available, either use a best guess or a default value of 0%.
7. Percentage of read requests that the system sends to the I/O subsystem.
  - For existing iSeries systems this value can be computed by dividing the "Average Reads / Sec" column value from the 'Disk Utilization Summary' section of a Performance Tools 'Resource Interval Report' by the corresponding 'Average I/O / Sec' column value from the same report..
  - If this data is not available, either use a best guess or a default value of 50%.

If Performance Tools reports are used to gather inputs 5-7, the data is collected on a disk basis. You will need to choose an average value from all of the disks being converted. However if default values are used, depending on the situation, additional disks may need to be added once the actual read-to-write and cache efficiency ratios have been determined.

In the following table, the yellow boxes (slightly grey if printed in black & white) identify the inputs to the process. Note that  $SW\% = 1 - SR\%$ .

Inputs	Sample Value	Value name	Possible values
<b>New System</b>			
Protection Level	1	NEW PROT LEVEL	0,1,5,6
# of drives for capacity requirements	50	NUMDRIVES4CPCTY	Drive model and customer dependent
<b>Current System</b>			
Protection Level	5	CUR PROT LEVEL	0,1,5,6
# of drives for speed requirements	100	CURR NUM DRIVES	System dependent
controller Read Cache efficiency	1%	RCEFF	0%-100%
controller Write Cache efficiency	50%	WCEFF	0%-100%
System request Read Percentage	70%	SR%	0%-100%
System request Write Percentage	30%	SW%	Derived from SR%

Use the following lookup table to set values in the WMx constants used in the derived equations of step 2. For example, if the 'new' configuration is RAID 5, use the column #2 value ('4') for the WMN value, located in the same row as the '5' in column #1.

### Constants

Write op disk request RAID Multiplier lookup table:

Protection level (NEW or OLD PROT LEVEL)	WMx value
0	1
1	2
5	4
6	6

*Table name*  
W\_MULT\_TABLE

Where x = 'C' for current configuration and 'N' for new configuration.

The intermediate derived equations are as follows :

Derived Values (Normalized for no Mirror)	Equation	Value name
No mirror System request Read Percentage	@IF(CUR PROT LEVEL=1, SR%/(SR%+SW%/2),SR%)	NSR%
No mirror System request Write Percentage	100%-NSR%	NSW%
<b>Derived Values (Cache Output)</b>		
Cache output Read request Percentage	(NSR%*(100%-RCEFF))/((NSR%*(100%-RCEFF))+(NSW%*(100%-WCEFF)))	CR%
Cache output Write request Percentage	100%-CR%	CW%
<b>Derived Values (RAID Output)</b>		
New write op disk request multiplier	@VLOOKUP(NEW PROT LEVEL,W_MULT_TABLE,1)	WMN
Current write op disk request multiplier	@VLOOKUP(CUR PROT LEVEL,W_MULT_TABLE,1)	WMC
current RAID output Read request Percentage	+CR%/(CR%+(CW%*WMC))	RR%
current RAID output Write request Percentage	100%-RR%	RW%
Extra drives to support current RAID level	+CURR NUM DRIVES*(100%-(RR%+RW%/WMC))	EXTRA CR DRIVES
Number of drives required for no protection	+CURR NUM DRIVES-EXTRA CR DRIVES	NO PROT DRIVES

The final equation combines the 3rd and 4th steps of the process as follows:

New # of drives needed	@MAX(NUMDRIVES4CPCTY,@ROUNDUP((NO PROT DRIVES) * (CR%+CW%*WMN),0))
------------------------	--

This process allows the translation of the required number of drives of any of the current iSeries RAID protection levels to any other current RAID protection level. It assumes the current configuration results in acceptable response times at the desired throughput(s).

Note that this paper does not address the computation of the minimum number of disk drives needed to meet capacity requirements. This was intentionally left out because the available drive capacity models varies too frequently and the method to determine that is hopefully obvious to the intended audience of this paper. Also note that this process assumes that all devices within the group of disk drives being analyzed have identical storage capacities.

The next page shows one example of converting from RAID 5 to RAID 1 using this process.

## Example:

Inputs	Sample Value	Value name	Possible values	Default Values
<b>New System</b>				
Protection Level	1	NEW PROT LEVEL	0,1,5,6	
# of drives for capacity requirements	50	NUMDRIVES4CPCTY	Drive model and customer dependent	
<b>Current System</b>				
Protection Level	5	CUR PROT LEVEL	0,1,5,6	
# of drives for speed requirements	100	CURR NUM DRIVES	system dependent	
controller Read Cache efficiency	1%	RCEFF	0%-100%	0%
controller Write Cache efficiency	50%	WCEFF	0%-100%	0%
System request Read Percentage	70%	SR%	0%-100%	50%
System request Write Percentage	30%	SW%	Derived from SR%	

### Constants

Write to disk request RAID Multiplier lookup table:

Protection level (NEW or OLD PROT LEVEL)	WMx value
0	1
1	2
5	4
6	6

*Table name*  
W\_MULT\_TABLE

Where x = 'C' for current configuration and 'N' for new configuration.

Derived Values (Normalized for no Mirror)		Value name
No mirror System request Read Percentage	70.00%	NSR%
No mirror System request Write Percentage	30.00%	NSW%

Derived Values (Cache Output)		Value name
Cache output Read request Percentage	82.21%	CR%
Cache output Write request Percentage	17.79%	CW%

Derived Values (RAID Output)		Value name
New write op disk request multiplier	2	WMN
Current write op disk request multiplier	4	WMC
current RAID output Read request Percentage	53.60%	RR%
current RAID output Write request Percentage	46.40%	RW%
Extra drives to support current RAID level	34.80	EXTRA CR DRIVES
Number of drives required for no protection	65.20	NO PROT DRIVES

### Output

Current # of drives needed =	100	for RAID level	5
New # of drives needed =	77	for RAID level	1

A Lotus 123 spreadsheet that automates the above process and algorithms can be found in IBM TechDocs. Search for RAID\_Swizzler.123 at the link shown below.

Business Partners: <http://partners.boulder.ibm.com/src/atmastr.nsf/Web/Techdocs>

IBMers: <http://w3-03.ibm.com/support/techdocs/atmastr.nsf/Web/Techdocs>

---

## 5 Author Contacts and Additional Information

Additional performance information and tuning suggestions are available in the latest edition of the iSeries “Performance Capabilities Reference” document on the internet at:

<http://www.ibm.com/servers/eserver/series/perfmgmt/resource.html> .

**Questions** relating to the information in this document should be directed to

Clark Anderson ([clarkand@us.ibm.com](mailto:clarkand@us.ibm.com)),


The author would like to thank the following people, who have contributed to this document: Sue Baker, James Cioffi, Allan Johnson, Brian Podrow, Alexei Pytel, Farnaz Toussi and Keith Zblewski and others who have reviewed this.

This document last updated 1/17/2006.

## Trademarks and Disclaimers

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both:

AS/400	DB2	IBM	Redbooks
AS/400e	DB2 Universal Database	IBM (logo)	xSeries
OS/400	eServer	iSeries	

The following terms are trademarks or registered trademarks of Lotus Development Corporation and/or IBM Corporation in the United States, other countries, or both:

Lotus	Domino	Lotus Notes
-------	--------	-------------

Other company, product, and service names may be trademarks or service marks of others.

Information is provided "AS IS" without warranty of any kind.

All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information in this presentation concerning non-IBM products was obtained from the supplier of these products, published announcement material or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the supplier of those products.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your local IBM office or IBM authorized reseller for the full text of the specific Statement of Direction.

Some information in this presentation addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in IBM product announcements. The information is presented here to communicate IBM's current investment and development activities as a good faith effort to help with our customers' future planning.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance 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 or performance improvements equivalent to the ratios stated here.

Photographs shown are of engineering prototypes. Changes may be incorporated in production models.