Veridictus Associates Inc
14 Via Sierra Grande
Manitou Springs, CO 80829

303.588.6674
www.veridictusassociates.com
blog: www.storagetopics.com

MAID — "Massive Array of Idle Disk".

A discussion of the technology, its derivatives and its business value.

Bill Mottram
Veridictus Associates, Inc.
January 2009

Abstract: MAID delivers the ultimate in energy efficient, storage density, on-line storage architectures. However, not all that is called MAID is equal, nor is much of what is called MAID actually MAID, at least as it was originally defined. A number of companies and industry observers are using the MAID terminology to describe software features imposed on traditional architectures. These features can deliver some energy efficiencies by drive spin down but the free use of the term MAID is misleading. This paper is an attempt to set the record straight and provide a straight forward description of MAID, what a MAID solution is and is not, and where such a solution is likely to deliver the most benefit.

Introduction.

The concept of MAID (massive array of idle disks) is the brainchild of a research team from the University of Colorado who hypothesized that MAID as a storage structure would deliver the density of tape with performance similar to that of disk and with a very small power envelope.[1]

The motivation for such an architecture was to deliver a solution that exploited relatively inexpensive SATA disk technology to create a commercially viable, enterprise class, mass storage solution that exhibited much of the performance, data access and data integrity characteristics of a disk array but with the economics of a tape library. The sweet spot for this technology, and where it will deliver the most benefit, is in the storage and management of persistent data. Persistent data is data that will rarely if ever be changed, but where data serving applications need faster access to individual files than magnetic tape can deliver. Complementing these performance characteristics are significant cost savings when compared to traditional disk array storage.

To effectively exploit MAID and enjoy its full benefit a user has to first understand the nature of the data that is populating their datastore. It is only when the storage and access characteristics of the data are known that storage managers will understand how this innovative technology or its derivatives will fit into their data management schema.

Because not all "MAID" labeled solutions are created equal, the purpose of this paper is to highlight some of the key differences and provide pertinent information to those considering the value of MAID or a MAID derivative for their storage infrastructure needs, especially during acquisition consideration.

Understanding your data.

In most organizations data is not growing, it is exploding! The problem with data is that once it is created it is rarely eliminated whether by choice (corporate governance), indolence (poor data management practices) or compliance (government legislation). With the accretive nature that characterizes data growth, the problem will become more acute, particularly if current storage practices continue.

Data makes up the digital history of an enterprise. It is a corporate asset which can hold considerable value or liability. The value of data is variable being content, time and activity dependent. Value can be influenced by age, by access requests, and while not a characteristic per se, its find-ability, accessibility, and integrity are all critical influencers. With that being said, there is also much worthless data being stored and managed in an enterprise, with its retention and management being a measurable expense. Hence the need for robust data classification and management practices.

[1] The Case for Massive Array of Idle Disks (MAID): Dennis Colarelli, Dirk Grunwald and Michael Neufeld, Dept of Computer Science, University of Colorado, Boulder. January 7th, 2002.

Data types can be simply classified as either highly active, transactional data, or inactive historical/reference data becoming labeled as persistent data. One of today's greatest storage management inefficiencies is the practice by default, of placing the right data on the wrong storage type.

1. ***Transactional Data*** **–** the tradition view of data and the view that has molded today's disk storage architectures. This data is being captured or created and tends to be highly dynamic, drive high IOPs, is random in nature and is likely to have a short shelf life. This explains the motivation that has influenced the evolution of today's traditional, transactional storage.

2. ***Persistent Data*** **–** data that once created is rarely accessed or rarely, if ever modified. This data does not demand the same response time or IOPs and tends to have a low temporal access locality, meaning caching is a wasted expense. Persistent data is apt to have a long term retention requirement, is bandwidth centric with data integrity concerns, and is likely to be event driven and immutable. Persistent data is the fastest growing segment of today's digital information. The surprising fact is the 60% to 80% of the data in a data center fits this "persistent data" description.

Compliance driven retention periods.	
tax returns -	Indefinitely
expense reports -	6 yrs
bank statements -	3 yrs
production correspondence -	8 yrs
terminated personnel files-	6 yrs
employees exposed to toxic substances -	30 years
medical records -	Life + 2 yrs
monthly trial balances -	6 yrs
insurance claims post settlement	10 yrs
manifests -	4 yrs

Differentiating data types does not enhance or diminish the relative importance or value of data. It simply improves the chances for its effective management, availability and use. There is agreement that persistent data is the fastest growing data type, in terms of data volume in the data center. The reason is that much of this data is subjected to minimum retention periods that are dictated by one compliance regulation or another. Not only must this data be retained but when requested it must be available in a timely manner. Significant financial penalties have historically been levied on companies who failed to deliver data in the time required by law.

By appreciating that different data types have different value and be willing to execute even on the simplest of classifications enables astute data storage managers to effectively match the value and requirements of their data to the cost and performance of the "hosting" storage technology. This realization is the epiphany opening the door of opportunity to significant cost savings while surviving today's phenomena of explosive data growth.

Purpose built architectures.

Purpose built to meet the demands of managing inactive persistent data, MAID meets the requirements of long term data storage. The sidebar contrasts the design principles of traditional, transactional storage with those architectures focused on the long term storage needs of persistent data.[2]

The key characteristics of a MAID solution that delivers user benefit.

The MAID storage architecture is characterized by power management and very dense drive packaging, hence its name, Massive Array of Idle Disk drives that are powered down individually or in groups when not required.[3] A MAID system suitable for an enterprise data center application is identified by unique design characteristics that are not noticeable in traditional array design.

1. ***Very high drive packaging density.***
   In a MAID architecture the number of drives that can spin at any one time is limited. This allows extremely dense packaging not possible in conventional architectures. Example would be a single COPAN frame that can support up to 896 drives, Data Direct Networks at 600 drive per rack, Isilon at 252 drives and the HP ExDS 9100 at 328 and 492 plus an extension cabinet with a capacity of 672 drives. The EMC Infiniflex 10000 aka Hulk tops single frame tops out at 360 drives. What about Atrato who can pack 1440 drives into a single rack? Admittedly they are 2½" drives with a maximum capacity of 500GB but this is a trend to be watched.

2. ***The number of drives that can spin concurrently is limited and will not exceed 50% of the total number of drives installed.***
   This is the original definition as presented by the University of Colorado researchers. COPAN currently limits this number to 25% of the physical maximum, the key being that not all drives spin at any one time. Thus reducing maximum power requirements and reducing heat generation in turn reduces the necessary cooling infrastructure and aides in the elimination of rotational vibration issues.
   There are many variations on the MAID theme that either slows spindle speed or stops them completely. While these implementations deliver energy efficiencies, the savings are not as dramatic as a native MAID architecture.

[2] Persistent Data Storage Architecture: COPAN Systems, September, 2006

[3] The Dictionary of Storage Networking Technology, Storage Networking Industry Association (SNIA), 2005/2006.

Transactional Array	Persistent Data Array
• Very low latency response • High IOP • Read-Write-Modify Model • Optimized to provide access to data at most if not all times • Exhibits temporal or spatial locality –cache friendly • Optimized to small-grained data access • Data retention tends to be short lived • Design principles that drive cost	• Once created, data is rarely accessed or modified • Often immutable • Low temporal access locality, caching tends to be a wasted expense • Only a small portion of the data repository is accessed at any one time • Often access is to medium to large grained data • Long term data retention • Design principles that drive lower design and production cost

3. ***The cabinet's power is limited and will not support all the drives spinning at the same time when fully populated.***
   A defining characteristic of a native MAID architecture is the fact that there is insufficient power in the frame to power up all the drives at the same time. A limited power budget not only drives power efficiencies but helps manage the overall provisioning of data center power. It also prevents any misguided attempt to simultaneously power up an illegal number of drives.

4. ***As the total available power requirement is significantly less than in traditional arrays, less power supplies, power converters, fans, etc. will be required.***
   Reduced component count (should) equate to lower cost and improved system reliability.

5. ***Access to data on drives that are powered down will be 15 seconds or greater; longer than expected from a traditional disk subsystem but significantly less than an off board tape system.***
   Applications must be MAID aware. A request for data on a powered down LUN will experience a delayed response back to the application and if the application is not MAID aware the delay may trigger a time out or recovery action.
   This is one of the major reasons hybrid MAID solutions have evolved. By not removing all the power from drives when they are spun down completely the response time to a data request is improved.

6. ***To meet data center, enterprise class expectations the solution should have embedded data and device integrity checking and self healing capabilities.***
   The value and usefulness of data lives much longer after its creation and its initial period of activity. Corporate governance and government compliance regulations are causing data to be stored for increasingly longer periods of time. This accretive process is fueling the explosive data growth issue. Any solution targeting the storage of long term data must be architected to deliver self diagnostic and self healing capabilities, ensuring data integrity and data availability. The more automated the processes the better.
   This is not a unique characteristic of MAID architectures but applies to all storage solutions intended for long term data storage.

Drive spin down.

Drive spin down is not MAID. It is better classified as a power efficiency technique for traditional architectures rather than a legitimate MAID implementation.

Spin down is defined as reducing the spindle speed from 7200 rpm to 4000 rpm or to zero. This reduces the power needed when the drives are in this idle or sleepy state. Note: data is only accessible when drives are spinning at their rated speed. Even implementations that actually stop spindle rotation, still have power feeding the drive electronics, impacting total energy efficiency. Actual energy saving expectations will be discussed later.

The following table illustrates the key differences between MAID and spin down technologies

MAID	Spin Down	Comment
"Hard MAID"		A term used to describe the implementation of the original concept of MAID, referring to a hard limit on the total power available meaning not all drives can spin at the same time. It also implies that power is completely removed from the drive contrary to other approaches where the electronics on the drive remain hot. However, by physically limiting the power available within the cabinet the greatest power savings are realized. Other benefits of limiting the number of drives that can spin include increased packaging density, improved scalability leading to improved management capabilities.
	"Soft MAID"	A software implementation and the simple addition of a spin down feature into an existing,, transactional array architecture. This means the data center power and cooling budgets must be able to cope with the peak delivery requirements of the storage systems.
Power drives down completely		The defining feature of MAID, by powering drives down significant power savings can be realized. Vendors claim up to 80% savings. This is also one of the areas of greatest debate.
	Partially spinning down drives	Architectures that simply spin drives down do not deliver significant power savings nor do they meaningfully reduce cooling requirements. Vendor speak suggests a 20% savings at best could be realized.
Limiting the number of drives that can spin concurrently		This translates to smaller power supplies, less cooling fans, less power converters meaning less heat. Less heat also means greater packaging densities are possible and with only a subset of drives rotating rotational vibration is eliminated. All characteristics that improve reliability and lower TCO.

Companies who offer a spin down include Fujitsu, HDS, NEC, Data Direct Networks, Nexsan, Xyretex with EMC and Pillar Data promising a future deliverable. There is an interesting new player, Greenbytes, who recently introduced an innovative hybrid solution joins Nexsan as one of the "near MAID" solutions.

Where does MAID fit?

MAID is a storage solution intended to resolve the challenges of storing very large amounts of long term, persistent data. It fits with backup and recovery, hosting aged transactional data, image capture and store, record and document management and archive. Archive is probably the greatest opportunity for MAID and the least served.

To understand where MAID fits in a data center, it is necessary to appreciate not all data is equal nor does data hold equal value to an enterprise, and some data has no value what so ever and should be discarded. There is a danger of misusing any large, relatively inexpensive data repository as a catch all data bucket to simply store all persistent data whether it has value or not. It is a novel thought that the differentiating characteristics of data should be used to influence the choice of the hosting storage technology. Data classification, retention and migration practices have to be defined and in place if any investment in MAID or such technology is to be optimized.

Where MAID does not fit.

MAID cannot support Tier I or Tier 2 applications such as OLTP, databases, exchange, ERP solutions and other mission critical type applications.

As with any new technology the compatibility question should be asked. MAID has a similar latency as a tape library, meaning if the application in use is not "MAID aware", or is expecting a disk-like response when the data may be on a powered down LUN, problems will occur as the application will likely time out and potentially flag a drive failure. Worth noting: while the vendor community is working feverishly to expand into the file and block world, VTL (virtual tape library) is the application that dominates MAID implementations.

Any consideration of MAID or other power efficiency features must be tempered with this question of application compatibility. In short, applications must be MAID aware.

MAID – a proven energy efficient solution.

MAID by its nature only powers up disk drives when a request to access data is made. As unneeded disks are powered off, significant energy savings can be realized, not only as a consequence of powering down the drives but by limiting the cooling requirements.

"Certainly, there are environmental reasons for going green, but a green focus also can result in significant savings."

— eWEEK.com

Energy usage in the data center is becoming an increasingly relevant topic in executive offices and with storage requirements currently representing about 30% of the total IT consumption, it is not unreasonable to expect an increasing focus on energy efficiency at the storage device level. However, while the volume of chatter regarding the greening of the data center may be increasing, it is not driven by any altruistic inclinations but a desire to control the ever growing cost and availability of data center power and cooling, driving the broad focus on energy efficiency. If you asked the data center manager what keeps him awake at night, his response is likely to reference meeting his corporation's quality of service (QoS) expectations rather than any focus on reducing utility bills. However, a technology that can address both operational as well as executive concerns should at least receive consideration as a possible solution.

Energy savings whether it is a MAID or a spin down solution can be substantial as the published numbers would suggest:

1. ***20% for spin down:*** These would include solutions from Fujitsu, HDS, NEC, Data Direct, EMC and Pillar Data.

2. ***15% to 56% for what is called MAID 2.0:*** Nexsan offers a range of possible efficiency gains depending on the mode selected. The range is broad, from 15% to 56% or even 70%. The conflicting data on the Nexsan WEB site (at time of writing this paper) is confusing but the data in the Schutz paper guides me to suggest that they can deliver up to a maximum of 56% savings. For more information on the Nexsan solution see the Greg Schultz, January 2008 paper on MAID 2.0.[4] Xyretex is another interesting option with claims of up to 40% power savings when their basic engineering innovations are complemented with their APM feature.

3. ***85% maximum:*** COPAN Systems were the first to do a deep dive on practical power savings measurement as Aloke Guha, former CTO of COPAN in his paper will testify5. While there are many variables, the published maximum power saving of the COPAN box when compared to a conventional array is 85%. COPAN technology also has the distinction of being eligible for energy rebates from PG&E and Austin Energy.

[4] MAID 2.0: energy Savings without Performance Compromises; Greg Schultz, Senior Analyst, the StorageIO Group, January 2nd.

[5] Solving the energy Crises in the Data Center Using COPAN Systems' Enhanced MAID Storage Platform; Aloke Guha, CTO COPAN Systems, December 1st 2006.

What about the reliability of a MAID solution?

No matter how interesting the technology- if it is not robust then it should not take up space in a data center. MAID and its derivatives have navigated a number of challenges questioning the suitability of SATA drives in the data center and the potential negative consequences of powering down drives. While no words in a white paper, or for that matter a product data sheet, will replace actual vendor manufacturing and field data the following are some thoughts on MAID and the reliability question.

- ***Limited power budget and power managed RAID groups***

  The characteristics of MAID enabling highly dense packaging while avoiding the challenging cooling issues, a normal consequence of a large population of spinning drives. An architecture that limits power consumption also means fewer components such as power supplies, fans converters etc and less heat translates to less thermal stress, all of which should lead to increased reliability.

  This is all very well, but when reduced to practice does theory become fact? Only the actual reliability statistics can prove how the theory plays out in practice. Before any purchase decision is made vendors should be asked to provide real world data, both from their manufacturing records as well as field data. If this data is not made available, the buyer beware.

- ***Effective Drive MTBF***

  Does field data support the published MTBF? There is a body of work that suggests not. [6][7] However in the Google study there was one very interesting observation: the failure rate statistics across the 100,000 test population of drives showed no difference whether it was a SATA, SCSI or FC Drive.[8] Guess misery loves company but it does offer support to the notion that SATA drives as used, are as reliable as the more traditional enterprise offerings. One empirical piece of data from Fujitsu which surfaced when their spin down was announced. was that no impact on reliability was observed in their testing of drive spin down configurations. COPAN has reported data which is much more dramatic and suggests the drive (effective) MTBF is increased by four to six times due to much lower power on hours.

- ***Power cycling drives (frequency of on/off cycles)***

  The initial reaction to this practice is to be convinced that frequent power cycling will reduce drive reliability. However both Seagate and HDS specify their drives to a 50,000 power on/off cycles which is far in excess of the stress drives will be subjected to in a

[6] Failure Trends in a Large Disk Drive Population; Pinheiro, Weber, Barroso; Google Inc; Presented at 5th USENIX conference, February, 2007.

[7] Disk Failures in the Real world; what does an MTBF of 1,000,000 hours mean to you? Schroeder, Gibson; computer science department, Carnegie Mellon University; Presented at 5th USENIX conference, February, 2008.

[8] Google et al

MAID environment. The argument remains however, and there is academic work that raises concern. So again- get the empirical data from the vendor before any commitments.

- ***Data integrity***
  Any system that is positioned as a solution for long term data storage must have the ability to self monitor and self heal. If this functionality is not available, look elsewhere for your storage solution.

Some last thoughts:

1. MAID will not replace Tier 1 or Tier 2 storage in mission critical applications or where the data is highly volatile (High IOP demand). Leads to point 2.

2. Data classification is a critical first step. MAID or any mass storage solution is not (should not be) a cover for sloppy data management practices. Understand your data and your migration strategy before including MAID in your storage architecture.

3. With the economics of MAID disk storage challenging that of tape, consider what data can be migrated from tape. Why? Improvements in data access, find-ability and data robustness delivering a higher QoS for vast quantities of persistent data traditionally labeled Tier 3 or Tier 4.

4. MAID delivers the highest storage density resulting in meaningful floor space savings. By consolidating your persistent data onto a dense MAID platform you open the possibility to reallocate more expensive storage for its intended purpose (highly active, transactional data) and perhaps delay an expensive purchase. Another possibility is eliminating aging space and energy hog, freeing up floor space, capturing energy savings and reducing operational expense.

5. Understand that MAID and spin down derivative technologies deliver different quality of service and environmental advantages.

6. Any application being used with a MAID device must be MAID aware.

Biography

Bill Mottram compliments his role as the managing partner of Veridictus Associates Inc, a high technology marketing consultancy, as a Principal Contributor for the Wikibon Project and as an analyst practitioner in the data storage industry with Data Mobility Group. He has over 25 years product development, marketing and sales experience in the data storage, information technology and medical device industries. Experienced with Fortune 500 companies such as StorageTek, Compaq and HP and smaller, entrepreneurial enterprises including Pillar Data Systems and COPAN Systems, he was the marketing leader responsible for the development, introduction and "go-to-market" activities for a number of innovative and highly successful data storage solutions. Although primarily a marketing professional Mottram has considerable expertise in technology management, corporate acquisition and public company funding.

A graduate of Paisley College of Technology (electronic engineering) and holds a Masters of Business Administration from Pepperdine University.

Mottram can be contacted at bill@veridictusassociates.com