

# STORAGE SWITZERLAND

## USING ALL-FLASH ARRAYS TO SOLVE TIER-1 DATABASE PROBLEMS



George Crump, Lead Analyst

To solve tier-1 database performance problems, it is important to understand the nature of tier-1 applications. Standard definitions of tier-1 include: (i) extremely high cost, extremely high performance applications – sometimes referred to as “tier-0” (e.g., Wall Street trading platforms) and (ii) business critical applications with very high (not extreme) performance requirements, where high availability and manageability are key. Not surprisingly, the majority of tier-1 applications are in the latter category.

True tier-1 solutions should deliver the highest levels of performance, but also high availability and ease of management. They need to meet these requirements assuming a defined budget and a shared IT staff. They are not tier-0 solutions, which optimize for performance but require high operational expenses for manageability and high availability.

All-Flash Arrays have the potential to eliminate database performance problems for tier-1, but selection of the right All-Flash Arrays can deliver more far reaching benefits in efficiency, manageability and availability.

### Tier-1 Database Performance Drives Complexity

Potentially the single biggest challenge facing the tier-1 database and storage administrator is architecting a design that will deliver on the performance and availability required. Achieving those objectives with disk based

arrays leads to a very complex design. Typically, administrators need to create different volumes and RAID groups based on the I/O characteristics of the data. For example redo logs are often on a RAID-10 volume (for performance and reliability), while data files are on a RAID-5 volume.

The core database itself is typically placed on a separate volume. This volume often needs to be concatenated from a maximum number of available hard disk drives so that it can respond well to read/write requests. This often means using RAID 5 or 6, thereby forcing another management challenge - mixed data protection formats (Mirroring vs. RAID).

The consequences of this is that each time a database is created or moved a whole series of volumes and striping schemes have to be carefully constructed in order to gain maximum performance from hard drive based technology. This process can take hours if not days.

Even when the configuration creation process is completed and all the data is loaded, the work is still not done. Because these disk based systems were stretched to their maximum potential, all the volumes have to be carefully monitored and constantly tuned to insure maximum performance. In fact, additional disk drives are often bought not for their capacity but for the extra performance they may deliver.

Oracle ASM and other database applications or file systems support something called short stroking. This is the process of only formatting the faster, outer portion of a group of disks sacrificing the rest of the capacity. This intentional low utilization further drives the price per GB of the solution higher.

The problems described above become exacerbated when a cluster for availability or compute performance is factored into the requirement. You now have twice the waste and twice the complexity in the way volumes are created and shared.

While some large enterprises may have a member of the IT team solely focused on tier-0 database storage performance, it is more typical that the tier-1 database environment does not. In fact even tier-0 applications are finding it difficult to dedicate personnel to just this task.

Increasingly all database applications, regardless of importance and performance demand, are managed by IT generalists that have a wide range of responsibilities to keep track of everything and every application. In reality, they can't be everywhere and database performance often suffers, many times without warning. It is the unpredictability of performance that is of greatest concern.

### The Hybrid Solution

Sitting between All-Flash Arrays and Hard Disk based systems are Hybrid Storage Arrays and dedicated caching devices. While these solutions may help with performance, they don't help reduce complexity. Most of these systems have the ability to "pin" or lock certain data sets into cache to achieve performance isolation. But it is often done at the volume level. This means that separate volumes, as described above, still need to be created so that certain data types like log files can be isolated to a particular location. In other words, complexity in design could still be a problem.

Even after initial implementation, the volumes that are pinned to cache need to be monitored to make sure they are most deserving of the cache resources. Most organizations have multiple database applications and each of these

applications tend to demand performance at different times. With a hybrid system, the database or storage administrator may have to constantly be pinning or unpinning data from cache.

Of course an alternative is to let the caching or tiering software manage the database files, treating all data types equally. The problem with this approach is lack of predictability. Will the right database components be cached at the right time and if multiple databases have a near simultaneous demand for performance, how will the cache respond?

Most caching vendors report a 40% to 60% cache accuracy rate; that means that at least 40% of the time there will be a cache miss. Hybrid or caching solutions can be effective but they have to be managed. In tier-1 workloads, there may not be the time available to perform that management.

It is also important to remember that Oracle has been optimizing their algorithms for 25+ years and still reports very conservative cache hit ratios. Optimizing applications for mixed storage types is a difficult problem and it is unlikely that a new company with a hybrid solution and no access to Oracle application code will be able to do better.

### All-Flash Arrays

#### The Performance Sledge-Hammer

When it comes to database performance, All-Flash Arrays essentially flatten the performance concern. Everything operates at the same speed. And for a tier-1 database that may be exactly what is needed. A single volume can be created that stores the database, its various log and index files, as well as the application; all with performance to spare. This single volume also makes sharing in a clustered database environment significantly easier.

#### Performance for All Databases

One of the ongoing concerns raised when IT planners consider All-Flash Arrays is will my database application be able to take advantage of it? The simple answer is yes.

Almost any database that is of appreciable size and interacted with by users will see a performance improvement. Moreover, the performance improvement is consistent for all I/O, without pinning or stripe management. But do you need this performance? The simple way to verify this is if a hard drive has ever been bought for the environment to address a performance concern rather than a capacity concern. If you have added more drives to a RAID group to increase performance, you can benefit from an All-Flash Array.

It is not just the potential performance improvement of a single database application either. All-Flash Arrays are designed to be a shared resource so that multiple database and non-database applications can use them. A single system can meet the demands of database environments, virtualized desktop and server environments as well as business analytics applications. In each case, the accelerated performance leads to increased users, virtual instances or more rapid decisions -- all of which decrease operational costs.

#### Overcoming Database-specific Write Amplification

Performance over time, however, can be a challenge for All-Flash Arrays, due to the way databases use their various log files. These files are essentially circular in nature. Flash storage by itself has a challenge with this type of operation. Typically log updates are relatively small in nature, much smaller than a flash's cell block.

The way flash updates data is by reading an entire block into memory, changing a small section of the block and then writing the entire block to the flash NAND. This phenomenon is called write amplification, a common problem in flash systems that is magnified in database environments. It causes flash systems to wear out more quickly and significantly reduces flash performance while under heavy write loads.

Some companies are solving this problem by making sure that they are always writing to virgin flash blocks. This process means that the [Pure Storage](#) software scans the available flash capacity that it has and clears out old blocks of data, making them immediately ready for new data. This

saves time on flash updates and physical wear on the flash itself. Most importantly, it keeps performance consistent while eliminating flash write amplification.

#### All-Flash Availability

Another concern for tier-1 databases is availability. Many tier-1 workloads have implemented clustering and even those that have not, need high availability out of their storage system. Many All-Flash vendors propose RAID as their HA solution but that only protects against storage device failure, not network, power supply or controller failure. When pressed, they suggest buying two systems and allowing the database software to replicate between the two. This is not only expensive it can also actually reduce performance.

Some All-Flash Arrays, like those from Pure Storage, are now offering complete high availability. Not only is everything redundant, it also can run at full performance while in a failed state. Even storage system upgrades, like software updates or controller updates, can be done without interruption or performance loss.

This treatment of high availability maintains predictable performance in several ways. First, the database application does not have the added load of providing protection. Its full resources can be focused on performance. Second, because failure does not cause a performance drop, user response times are always consistent. This means that upgrades and drive replacements can be done during the day and the database or storage administrator can go home at night.

#### All-Flash vs. Disk Cost Considerations

Once the performance capabilities can be rationalized, the final barrier is cost. Part of the problem is that All-Flash Arrays are often compared to the cost per GB of raw off the shelf hard drives; not the fully burdened cost of a hard drive in a storage system that has been tuned for maximum performance. In reality, even if the operational time savings mentioned above are factored out, All-Flash Arrays are often less expensive and perform better than high performance hard-drive or hybrid storage systems on a usable \$/GB basis.

As mentioned above, databases have multiple components, each often requiring their own volume. These volumes can't be allocated perfectly to the right capacity, and often involve IT requests that span teams, and so they are often over-allocated.

In addition, some of the volumes leverage very wide striped RAID groups (a large number of drives per RAID set) to achieve performance, often adding drives just to increase performance with no need for the additional capacity. Compounding this issue, in some cases these drives are only formatted to 50% (or less) of capacity (short-stroking), so that data only resides on the fastest outer edge of the platter.

The net impact is that disk based arrays that host databases are horribly inefficient from a capacity utilization standpoint. Each individual drive costs money, space and power. Moreover, these inefficiently used hard drives are the most expensive hard drives on the market; not the cheapest ones.

All-Flash Arrays, on the other hand, can utilize all their available storage capacity. They don't need to create islands of capacity because of a need for separate volumes, nor do they need to be short stroked for performance. Finally, they can run at a much higher total utilization rate because there is no slow section on the device to avoid placing data. In short, all flash is fast, all the time.

Beyond this, All-Flash Arrays are now delivering deduplication and compression to further reduce the cost. Both techniques perform very well in All-Flash environments and in most cases, any performance impact is not seen by the user or database administrator.

It is important to note that having both deduplication and compression is important, especially in database environments. Redundant data (deduplication) is less common but easy to compress, textual data (compression) is very common. The impact of the combined storage efficient

technologies is conservatively, a 3X reduction in database capacity consumption, but can reach 8X.

Virtual environments can get 9X and All-Flash Arrays can easily support mixed workloads. No longer do specific storage systems have to be dedicated for the various environments that a data center has.

Additionally, flash systems are now delivering data services like thin provisioning, snapshots and clones. While common features on hard disk based systems, some flash vendors have been slow to bring them to market. When they do appear, as we see in Pure Storage's Flash-Arrays, they are able to deliver these services with no performance impact. This means they are used more heavily, which leads to a further reduction in storage consumption and better performance when using near-production copies for analysis or development.

#### Conclusion

All-Flash Arrays bring tremendous value other than their performance. They can simplify operations, eliminate tuning and do so at a price point that is less expensive than disk based and even hybrid systems. Frequently, All-Flash customers remark at the simplicity of operations and the time returned to the IT staff to work on other business requirements.

That said the performance of All-Flash Arrays should not be overlooked. Not only for how it can meet today's problems, because it will, but for the potential that they open up for application development in the future. No longer is the database designer constrained by the capabilities of the storage infrastructure. With that freedom they should be able to design environments that scale higher, handle a larger workload set, cost less, are simpler to maintain, and solve problems previously thought unsolvable by a single storage system.

### **About Storage Switzerland**

**Storage Switzerland is an analyst firm focused on the virtualization and storage marketplaces. For more information please visit our web site: <http://www.storage-switzerland.com>**

*Copyright © 2013 Storage Switzerland, Inc. - All rights reserved*