



Building a 3-D DRAM Architecture for Optimum Cost/Performance

By Gene Bowles and Duke Lambert

As systems increase in performance and power, magnetic disk storage speeds have lagged behind. But using solid-state disk systems to create a third dimension in the DRAM architecture can increase overall performance and reduce system costs.



Gene Bowles (left) is president & CEO of Database Excelleration Systems, and founder of Storage Dimensions (owned by Artecon). Duke Lambert is vice president of sales, with seven years' experience in database application environments.

Performance of relational database management systems has improved steadily over the past ten years. They are more sophisticated and capable of performing greater quantities of work, while simultaneously delivering more complicated result sets. This has generated tremendous pressure on the hardware platform to perform. Running under high loads, Sybase's Adaptive Server is capable of consuming 12 or more CPUs while producing thousands of I/O requests per second.

But the raw speed of the CPU inside host platforms is keeping up. CPUs have followed Moore's law and doubled in performance every 18 months, representing a ten-year performance improvement of over ten times. During the same time, however, magnetic disk storage has increased in speed by a factor of only three times. This has resulted in an imbalance in system design that often results in both excessive costs and disappointing performance.

The number one challenge in disk storage performance lies in the inherent mechanical delays caused by the access times and latencies of disk drives. Even with complex striping and partitioning techniques employed to improve throughput, if all I/O requests are serviced by physical magnetic disks alone, service times will escalate. Accordingly, DRAM, in the form of main memory and disk cache, is frequently utilized to have the data of interest available quickly to the database engine. The key to optimized data performance is creating a systems archi-

tecture that can assure the highest chances that the data of interest is being read from or written to DRAM, not magnetic disk.

The objective of this article is to outline a 3-D strategy for system DRAM deployment that provides an optimized system architecture for Sybase—one that achieves the highest possible percentage of I/O requests fulfilled directly from DRAM without magnetic disk interaction. This strategy provides new alternatives for configuring a range of systems with superior performance relative to 2-D architectures.

Characteristics of Two-Dimensional Architectures

The most common application of DRAM, the first dimension, in most hardware architectures is inside the host system as main memory. Main memory offers very fast access and very high bandwidth for all data that is located there. However, this memory is completely volatile. If the host system experiences power loss or must be rebooted for any reason, all data in main memory will be lost and must be recreated. Because of volatility, main memory is used only for storing data being read and for write buffers waiting to be flushed to disk. Primary uses are buffers, global cache, named cache, and operating system management. A higher performance system configuration may have 1–4GB of DRAM configured in main memory.

The second dimension of DRAM deployment usually consists of controller-based disk cache, as an I/O buffer to the disks and/or

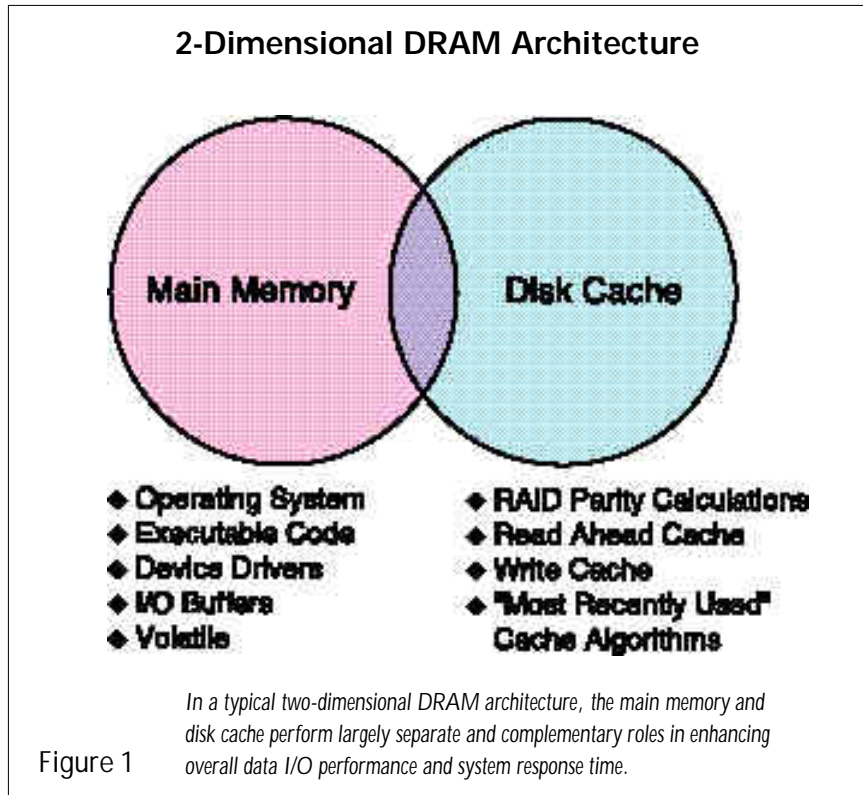
RAID arrays. DRAM mounted on the disk controller is moderately fast (slower than main memory due to necessary directories for cache content lookup) and in most cases non-volatile, having battery and magnetic disk backup to protect data in the event of power loss. Controller-based DRAM is used for write buffering, read-ahead cache, LRU/MRU algorithms, and RAID parity calculations. Performance is based upon the ability of the algorithms to keep the data of interest in the cache area so it does not have to be retrieved from magnetic disk drives. Typical disk controller cache sizes run from a single MB on a standalone disk drive, up to 256MB on open-systems RAID arrays, and as much as 4GB on mainframe storage systems.

This system design is referred to as having two dimensions. 2-D systems are simple to design, easy to understand, and straightforward to implement. However, in actual usage they often fall far short of achieving the required performance levels to support the Sybase server. The reason for this is that they depend upon the data caching algorithms for performance. These systems are "statistically" fast; that is, they depend on the probability that the data is in the cache when needed by the CPU. Otherwise, the system must wait for the disk drives to retrieve the data, subjecting the system to inherent mechanical delays due to access times, latencies, and data throughput rates.

Figure 1 illustrates this design and shows that while there is a small area of functional overlap, each area of DRAM utilized has specific tasks for which it is best suited.

2-D Performance Enhancement with Data Striping

Within 2-D systems there are hidden costs which can be quite high. These manifest themselves in several ways. The algorithms that are used to manage the two dynamic cache areas in 2-D systems are often not effective enough to provide sufficiently high cache hit ratios. Based upon testing performed on hundreds of different host system configurations running Sybase Adaptive Servers, we know that data can be delivered from either main memory or disk cache DRAM fast enough to saturate the server engine. When the server is waiting on I/O, it is because physical disk access is being required.



The latency and seek times of the disk drives themselves are the root cause of the I/O troubles. Systems where the desired data is always in cache do not experience I/O bottlenecks. Therefore, I/O troubles indicate that physical drives are being accessed. That, in turn, indicates there is an insufficient cache hit/miss ratio.

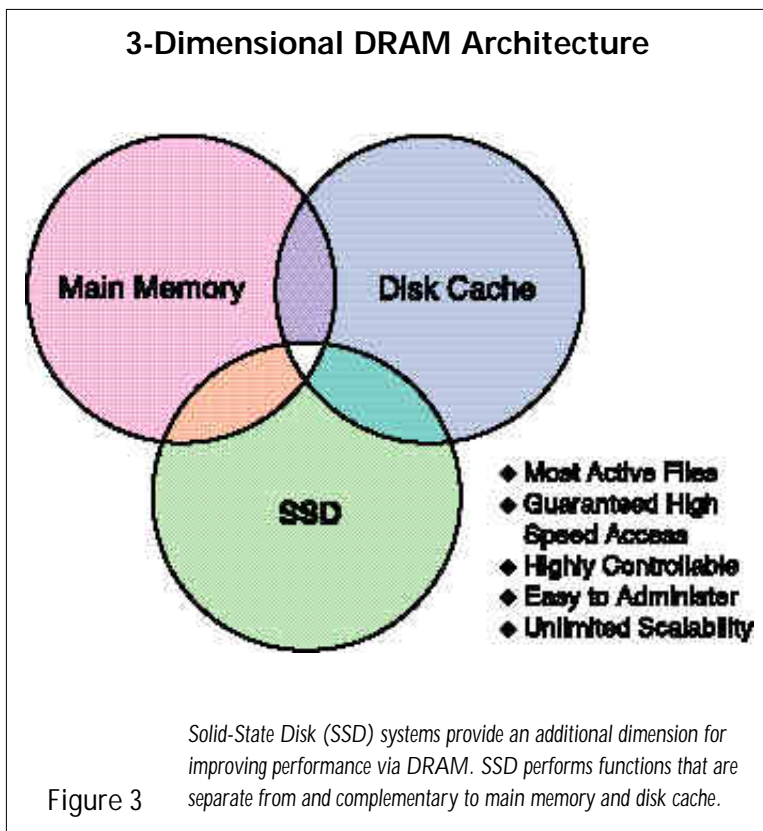
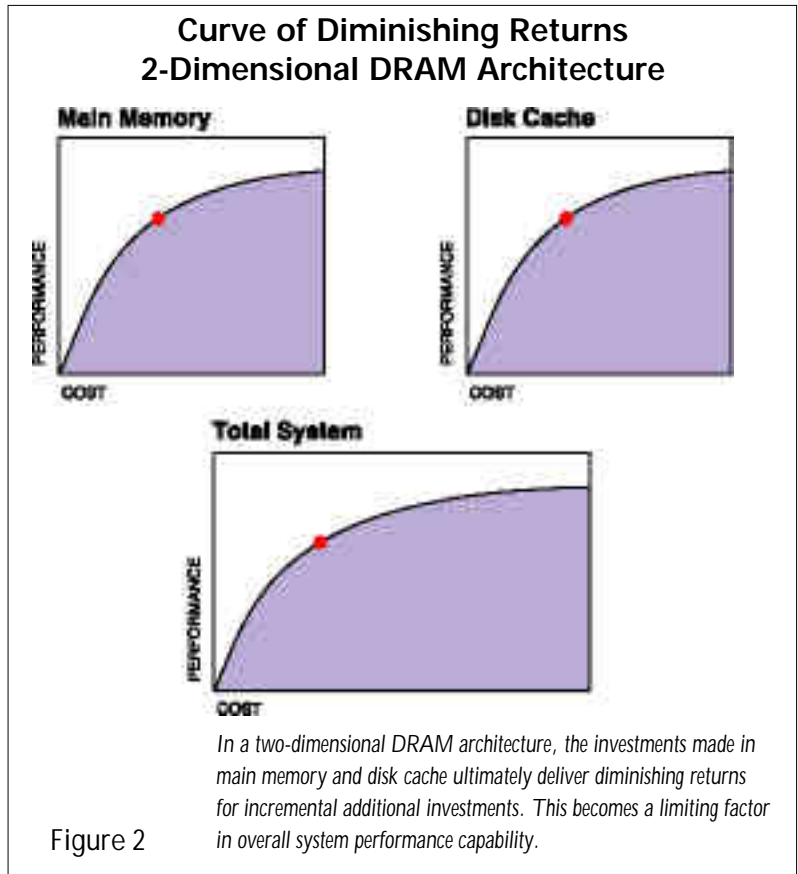
When cache hit ratios are low, the next step most system administrators take is to begin spreading the data over many disk spindles in order to minimize contention and increase aggregate throughput. Sophisticated techniques involving various levels of striping and mirroring are employed to minimize the impact of the cache miss. However, these techniques come with significant penalties. Complicated stripe sets can require days or even weeks of careful data analysis and layout.

Once the stripe is configured, it will not be easy to determine where the logical data segment is physically located. This further complicates follow-on tuning, since locating the particular data segment that is creating a performance penalty may be difficult or impossible. As system data storage requirements increase, the data analyst will be forced to completely re-stripe data sets, a challenging and time-consuming process.

Diminishing Returns from Additional DRAM Investments

Savvy systems administrators know that since the penalty for a cache miss is high, systems should be configured with very large amounts of both system memory and disk cache. However, Pareto's law applies: The initial 20% of the memory will handle 80% of the I/O activity. Beyond that point, as memory is added, the resulting performance advantage declines. A system configured with 1GB of system memory may receive a 20% improvement by adding another 1GB of memory. Adding a third GB will add only 15% more, and adding a fourth GB may result in only 10% improvement.

In each case, the cost of the 1GB of memory was the same, so the net value received for dollars invested declines dramatically as the total amount of memory increases. The effect of this is that a 2-D architecture for high performance is very expensive, since the net performance gain for each dollar invested is declining; yet the system administrator must highly configure the system to achieve acceptable performance.

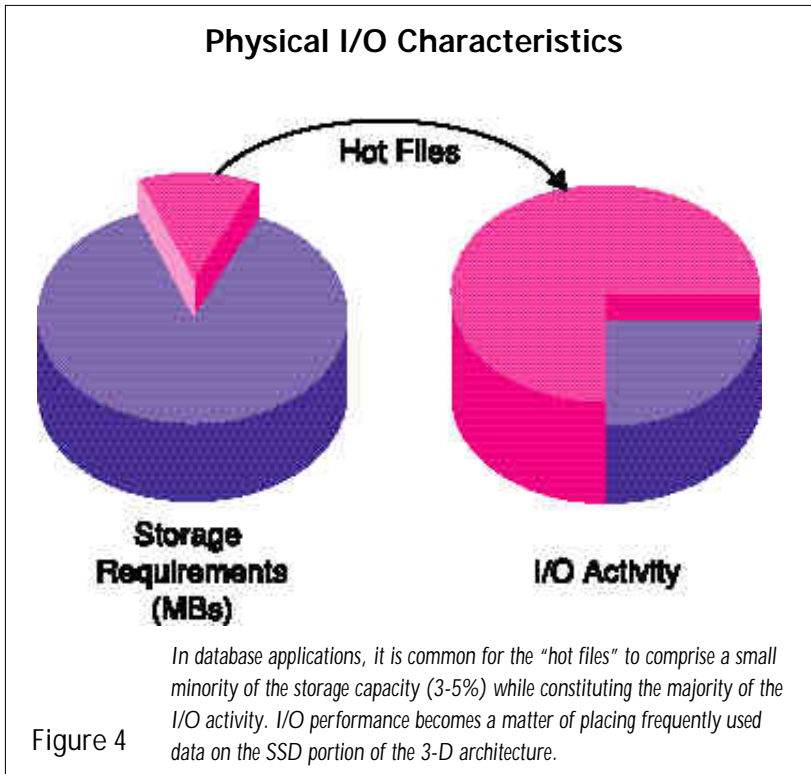


In a system constrained by having only two dimensions, it is often necessary to be far out on the performance curve in order to achieve the level of system performance required by the users. The result is that the costs of configuring the system are driven further up, while relatively small performance gains are experienced by system users.

Three-Dimensional Architecture

Using intelligent Solid-State Disk (SSD) systems as a third dimension creates a third area of DRAM that has unique strengths. The SSD is characterized by very fast performance with complete non-volatility and the ability to permanently locate data on the DRAM. All access of the DRAM is performed via direct addressing logic, allowing data to be accessed consistently in 18 useconds.

3-D architecture takes advantage of the I/O characteristics of Sybase Adaptive Server. With Sybase, as with all other relational databases, a relatively small amount of data receives the vast majority of I/O requests. Many studies have been



DRAM Scalability

2-D architecture is scalable until the point where adding additional DRAM stops improving the cache hit ratio, or until there are no additional card slots for DRAM expansion. Following the curve of diminishing returns we outlined above, in most cases the cache hit ratio stops improving long before the system has been fully configured with memory.

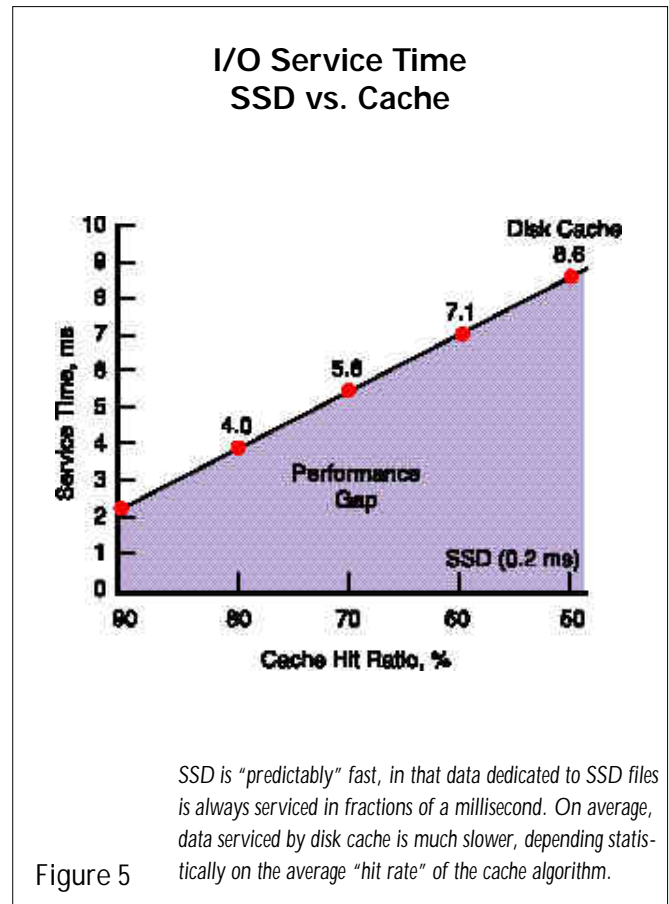
As the load increases, the 2-D system becomes increasingly difficult to maintain for two reasons:

1. Main memory and disk cache DRAM configurations grow beyond the optimum point of cost effectiveness. More dollars are spent initially buying hardware, while less and less return is realized in performance as the configuration moves out of the value curve.
2. As the DRAM cache effectiveness falls off, the system administrator is forced to create complex data stripes inside the disk array in

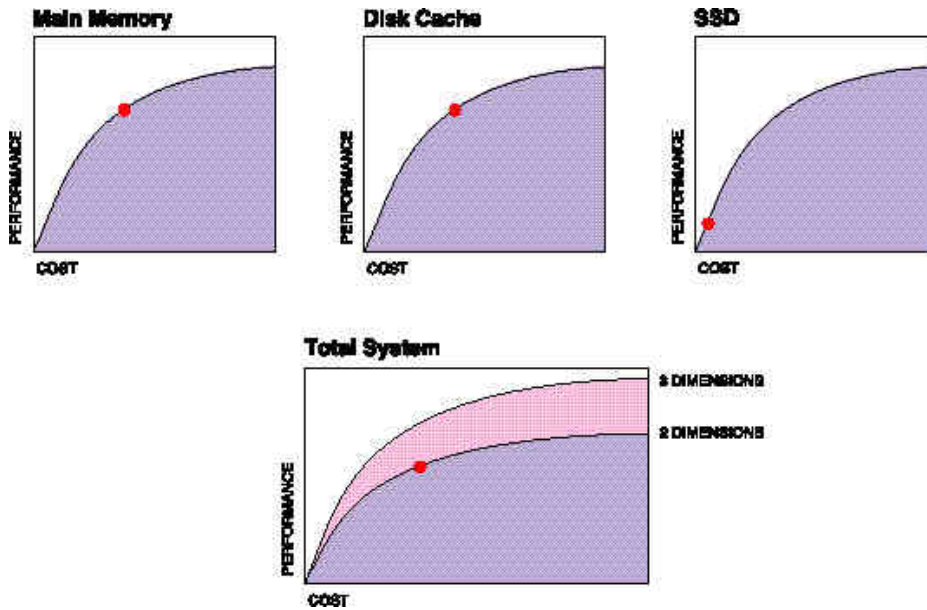
conducted on this subject and we have monitored the I/O performance characteristics of dozens of systems ourselves. The results show that 3-5% of the data typically represents over half of the I/O requests. This is represented in the two graphs in Figure 4. A small percentage of the data is receiving and processing a majority of the data requests.

The third dimension is used to systematically place the "hot" data spaces in DRAM. By permanently caching this data, it is possible to move over half of the total physical I/O load to the SSD DRAM area. The remaining 40-50% of physical I/O activity leaves a light load for the disk cache to handle. As a result, the majority of the physical I/O activity is serviced in under .5 ms, and most of the remaining 30% is serviced at disk cache speeds (approximately 4ms). Figure 5 shows the effect that changing disk cache hit ratios have on disk service time and compares them to SSD service time.

A properly designed 3-D system will offer much higher performance. Reducing average service times by three-fold yields very positive results in the performance of Sybase Adaptive Server (or any other RDBM). The "hot" data spaces tend to be those that online users or off-hours batch jobs are depending upon for their next I/O request. Eliminating wait on this data is one of the desired effects of a 3-D architecture.



Curve of Diminishing Returns, 3-Dimensional DRAM Architecture



When SSD is first added to the architecture, the performance return on additional DRAM investment is large. Moreover, the addition of SSD provides a new dimension of DRAM for increasing system cost and/or lowering cost per user.

Figure 6

an attempt to ward off the performance impact of physical disk access. The cost of this striping is significant both in time delays prior to deployment and in ongoing support. Optimal disk striping is typically specific to one particular type of operation. For instance, in a data warehouse, it is not possible to optimize the stripe for both data load and for random access. Compromises must occur which increase cost by reducing effectiveness.

Thus, while there may be theoretical scalability remaining in the 2-D architecture, in practice the limit is reached prior to the limit of the physical capacity. For the host, this typically means 2-4GB of DRAM, and for the disk cache, from 256MB to 2GB. For the majority of Adaptive Servers, expanding the DRAM in 2-D architectures beyond these capacities will increase cost but will not significantly increase performance.

3-D systems offer infinite scalability in I/O performance. In 3-D, it is possible to configure as much SSD DRAM as desired, as well as to physically locate specific data on that DRAM. I/O performance becomes a matter of placing the frequently used data on the SSD portion of the 3-D architecture and leaving the balance of the data on the magnetic disks, where it will be cached by the disk cache DRAM.

Using 3-D in this manner enables the system designer to achieve maximum efficiency, allocating enough DRAM in each of the three dimensions to reach, but not exceed, the point of diminishing returns. This provides a system that has an equivalent cost but offers significantly greater performance than a 2-D system.

It is also possible to design for cost savings using a 3-D architecture. As the effectiveness of the 2-D system diminishes, the cost per unit of performance increases. Since the third dimension adds a segment of DRAM that was not configured in the 2-D system, there are very large performance gains for a small incremental cost.

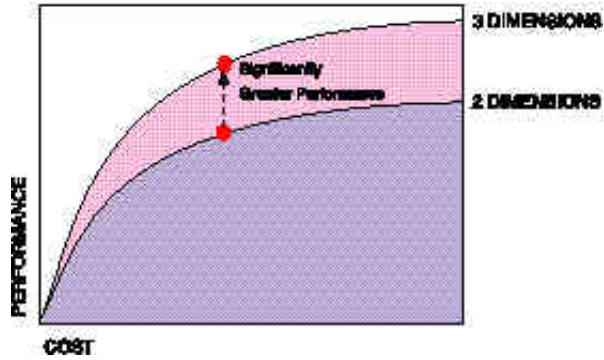
A balanced 3-D system provides the ability to configure a

system for additional performance at the same cost or for equivalent performance at a lower cost.

In summary, 2-D architectures offer simple design and ease of use for systems that do not face demanding loads. When 2-D systems are stressed by Adaptive Servers that are being used for heavily loaded applications supporting significant work loads, they lose their effectiveness and simplicity. Three-dimensional architectures provide a much simpler design that is infinitely scalable. Support costs are reduced, since complex striping is greatly lessened or eliminated completely. Administration is easier and more cost-effective, since extremely complex data layouts are not required. Assuring that the data of primary concern is always located in cache also provides optimal performance and prevents over-configuring the system in other areas.

In addition, data warehouses and other VLDB systems no longer need to worry about exceeding the effectiveness of their disk cache. Placing the "hot" areas (indexes, data marts, tempdb) on SSD in the third dimension assures that Adaptive Server will operate efficiently, because the critical data will always be accessed at DRAM speeds (<.5ms).

3-Dimensional DRAM Architecture Performance Improvement Advantages

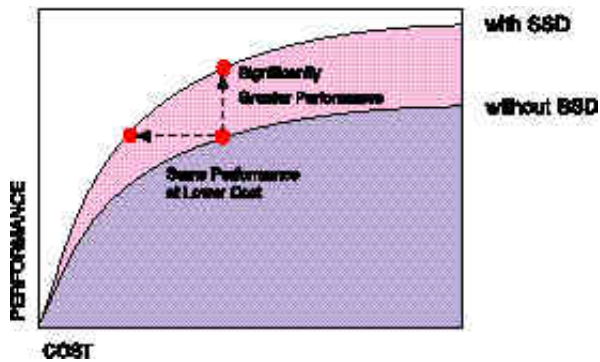


With 3-D architecture, the data files with the most I/O activity are dedicated to the SSD system, which does not have the mechanical access and latency delays of disk drives. The result is a level of system performance unachievable with a 2-D implementation of the same hardware platform.

Figure 7

As shown in Figure 9, 3-D architecture allows a new range of performance options that increases performance, lowers cost and expands scalability. These features are particularly beneficial in corporate environments that have standardized on a single server platform for the majority of their Sybase applications and requirements.

Benefits of 3-D More Performance for Less Cost



The 3-D architecture, using SSD for hot files, can deliver greater performance for the same cost or the same performance at lower cost, compared to systems without SSD.

Figure 8

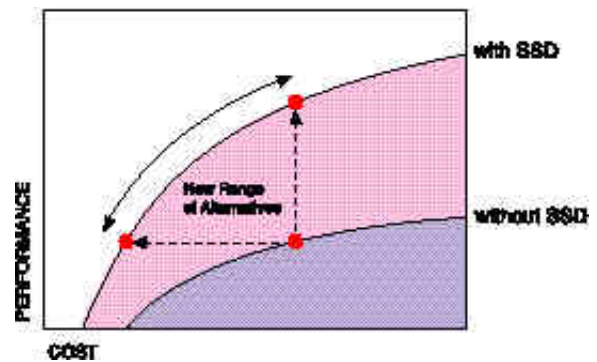
Conclusion

DRAM has traditionally been deployed in open-systems hardware platforms in two dimensions — main memory and disk cache. Solid-State Disk systems offer a third dimension for utilizing DRAM to increase the performance and lower the cost of Sybase's Adaptive Server applications.

SSD provides functionality that is different from, and complementary to, main memory and disk cache. Specifically, by dedicating strategic files to SSD, the system designer can insure that the highest activity data in the application will always be read from or written to DRAM, not magnetic disk.

As a result, systems incorporating 3-D DRAM architecture are frequently faster and more cost-effective than 2-D systems. By incorporating SSD architecturally, system integrators will be able to configure a new range of alternatives that deliver superior cost/performance characteristics relative to traditional approaches. ■

Summary Benefits of 3-Dimensional DRAM Architecture



The third dimension of DRAM, comprised of SSD storage, provides new alternatives for configuring a range of systems with superior cost/performance characteristics.

Figure 9