

# **An In-depth Look at the RamSan-630 Flash Solid State Disk**

---

Texas Memory Systems



April 2010

# Contents

---

<b>Introduction</b> .....	<b>1</b>
<b>Performance</b> .....	<b>3</b>
<b>Connectivity</b> .....	<b>6</b>
<b>Reliability</b> .....	<b>9</b>
<b>Availability</b> .....	<b>12</b>
<b>Scalability</b> .....	<b>13</b>
<b>Manageability</b> .....	<b>14</b>
<b>For More Information</b> .....	<b>16</b>

## Introduction

The Texas Memory Systems RamSan line of solid state disks has always dominated the performance end of the storage spectrum. That performance dominance continues with the RamSan-630, which offers up to 10TB of usable capacity in a 3U chassis that draws under 450 Watts of power.



Figure 1: The RamSan-630

The RamSan-630 takes advantage of generations of backplane, interface, and reliability experience designed into RamSan solid state disks (SSD). The RamSan-630 has the following key features:

- **The World's Fastest Storage®:** Sustained I/O rates up to 500,000 IOPS (at 4K block sizes) for reads or writes and sustained, random bandwidth up to 8 GB/second.
- **4Gbit Fibre Channel:** The RamSan-630 supports up to five dual ported 4Gbit Fibre Channel controllers.
- **QDR InfiniBand:** The RamSan-630 can be equipped with one to five QDR InfiniBand ports<sup>1</sup>.
- **Fault Tolerant Flash (FTF):** The Flash used in the RamSan-630 is enterprise grade SLC Flash. Each Flash card in the RamSan-630 includes on-board RAID protection to protect against chip failure. In addition, an advanced chip-level ECC provides multiple regions of bit correction per block.
- **Wear Leveling:** The RamSan-630 includes wear-leveling technology that protects the underlying Flash media by spreading out writes and maximizing the write life of the system.
- **Completely Non-Volatile:** Flash memory is inherently non-volatile. Ultra capacitors are incorporated into the design anywhere RAM buffers are used with enough power to synchronize the buffers to Flash.

---

<sup>1</sup> 2010 Q3 planned availability.

- **Unparalleled Scalability:** A single 3U chassis upgrades from 4 to 10TB of Flash storage and 10 Fibre Channel ports or 5 InfiniBand ports (when available). Multiple units can be added to meet any capacity or performance requirement.
- **Low Power:** The RamSan-630 draws less than 450 Watts of power.

## Section 2

---

### Performance

Performance is key to a quality SSD. The RamSan-630 is the latest entry in the TMS World's Fastest Storage® product line. The RamSan-630 has the following key performance metrics:

- 80 microsecond access time writes (4K)
- 250 microsecond access time reads (4K)
- 500,000 sustained random IOPS (4K Read)
- 400,000 sustained random IOPS (4k Write)
- 8GB/second peak bandwidth

The key measure of the RamSan-630's capabilities is its ability to sustain extreme inputs/outputs per second (IOPS) even with mixed read/write workloads. This is important because most applications have a mixture of reads and writes with a great deal of random access. Most other Flash SSDs and hard disk drive (HDD) solutions suffer significant performance degradation from either random access patterns or write accesses, if not both. For example, most Flash SSDs suffer disproportionately if they experience even a small percentage of writes, to the point that their mixed read/write performance may actually be less than their peak write performance. In practice, this means that a RamSan-630 will deliver better sustained "real world" random read/write application performance than any other Flash SSD.

Even though cheaper disks can be scaled to increase throughput, the RamSan-630 provides an advantage in performance costs. Consider high performance 15K RPM Fibre Channel (FC) HDDs can achieve at best 300 IOPS. At that IOPS rate, it would take well over 1,600 HDDs to equal the transactional performance of one RamSan-630. This means 120 disk enclosures in nine full racks, all of which would draw approximately 30,000 Watts, while one RamSan-630 uses 450 watts!

The densities and performance achieved by the RamSan-630 can often make the move to RamSan cheaper than disks. An application needing only a fraction of the performance available from the RamSan-630 will often use many shelves of 15K RPM disks just to reach 10TB of capacity and acceptable speeds. At many kilowatts of power usage and leased rack space, coupled with the expensive maintenance costs for the equipment, customers are finding traditional disks too expensive when so much more performance can be achieved at similar pricing.

## **Bandwidth**

Maximum data throughput is another key performance metric. The RamSan-630 delivers its bandwidth to hosts through up to ten 4-Gb Fibre Channel interface ports or five QDR InfiniBand ports (coming soon). Its high bandwidth offers two main benefits:

- High bandwidth enables administrators to share the RamSan-630 across multiple hosts without affecting performance.
- Some applications require high bandwidth, including high performance computing, data warehousing, batch processing, and non-linear editing of HDTV.

Competing SSDs offer far less bandwidth than the RamSan-630. Competing SSDs also cannot support all of the applications the RamSan supports and cannot be shared with as many hosts.

The RamSan-630 provides a massive 8 GB/s bandwidth in a 3U form factor that also offers up to 10TB of capacity drawing less than 450 Watts and weighing less than 60 lbs. An HDD-based system would need at least 200 drives to match this bandwidth, take up one full rack, and draw many kilowatts in power. An important note, as you compare the RamSan to HDD RAID: our solid state disk can sustain its bandwidth numbers with random data streams. Conversely, HDD RAID systems can only sustain high bandwidth numbers with sequential data streams.

## **I/Os per Second (IOPS)**

The RamSan-630 can deliver 500,000 random IOPS to your applications. We stress "random IOPS" because HDD RAID manufacturers occasionally cite their IOPS performance and discerning buyers should be aware that these quoted numbers are almost always sequential IOPS performance. Similarly, if you see Flash SSD performance quoted, beware that many times these numbers only reflect read performance, are often based on cache hits, and are not sustainable without extremely high thread counts to the system. Further, some Flash manufacturers quote numbers that are only achievable with odd capacity-constrained configurations.

The problem with sequential IOPS is that almost no "real world" applications actually result in sequential small block disk access. So why do you need high random IOPS? They are needed to process database transactions. Database transactions have two main characteristics: they are small (averaging around 8K), and they are random. Small random file accesses thrash HDD. In fact, really good HDD can only provide around 300 random IOPS. Fast HDD RAID systems would need hundreds of drives to match the RamSan-630's performance, if the controller didn't become a bottleneck first. The fastest cache built into expensive HDD storage arrays can handle at most 150,000 IOPS, still well below the RamSan-630's performance direct to non-volatile Flash.

This brings us to an answer to the original question – it is important to support a high number of IOPS because your servers can produce a high number of IOPS. When your processor is faster than your storage, then your processor literally waits on storage for data needed to perform calculations – this is called I/O wait time. If your processor is waiting, then your users are waiting too. In addition, you are wasting a significant investment in the latest processors and software licenses.

By supplying very high random IOPS, the RamSan-630 eliminates I/O wait time for all of its stored files. Because its random IOPS capabilities are so extraordinary, a single RamSan-630 can provide I/O acceleration for multiple host servers, allowing you to scale-out your processing demand as needed

### **Latency and I/O**

Peak application performance is affected by the combination of response time (latency) and peak I/O at certain thread counts, block sizes, and read/write patterns. HDD RAID systems typically offer 4-8 millisecond access times. To lower response times and increase performance, HDD RAID systems add RAM or Flash caches to their array controllers. These caches can lower latency to around 0.5 millisecond access times when the requested data resides in the cache, but when it doesn't and the data request goes through to the back-end disks, the latency returns to 4-8 milliseconds. The RamSan-630, on the other hand, offers 0.08 millisecond access times for writes. This is much faster than HDD RAID systems. The RamSan-630 provides 0.25 millisecond access times for reads. This is at least 20 times faster than most HDD RAID systems as well. In fact, in most cases the Flash-based RamSan-630 will return data faster than the cache on RAID subsystems (regardless of the type of storage used for the cache).

For some customers, this combination of low latency and sustainable I/O is the key to improving application performance.

When you consider our combined bandwidth, random IOPS performance, and extremely low latency, it is clear why we call the RamSan "The World's Fastest Storage®."

## Connectivity

### **SAN Connectivity**

The field of solid state disk options has expanded dramatically as the cost has come down. This expansion has included SSD connectivity. Now you can buy SSDs that both look like and connect using the same protocols as HDD. You can also find SSD solutions such as the RamSan-630 that connect using Fibre Channel or Infiniband interfaces directly to servers or through Storage Area Networks (SAN) and even Network Attached Storage (NAS) architectures. Most recently, PCIe attached SSD options such as the RamSan-20 have appeared on the market, offering the ability to put a high performance, high capacity SSD solution entirely within a server.

PCIe connectivity can be a powerful feature for high-end workstations, smaller reporting database systems, and other environments where an internal solution is an advantage. However, it isn't the best solution for every storage requirement. When deploying PCIe-based solutions, the SSD capacity becomes locked up in the server and limited to the number of PCIe expansion slots supplied by the manufacturer. Plus, the SSD cannot be shared across applications.

Proponents of PCIe SSDs claim that locating the SSD nearer the server CPU increases performance by avoiding network and storage array performance penalties. But notice, the latencies advertised by PCIe SSD vendors, including the figures for the TMS RamSan-20, are actually 4x higher than the latency of our own RAM-based FC-attached products. The truth is, connecting SSDs using a high speed low latency interface such as 4 Gbit/s Fibre Channel or SDR Infiniband adds <15 microseconds to an access. The Flash storage medium itself, the SSD controllers, and the PCIe interface add significantly more latency than this. So you can, in fact, configure externally attached, shared SSD with lower latency and higher IOPS and bandwidth than can be achieved with a PCIe SSD implementation.

The RamSan-630 is an SSD attached externally to servers. Because of this it does not suffer from the limitations of PCIe solutions. By leveraging SAN and NAS architectures the RamSan-630 offers some important advantages:

- Shareability
- Redundancy
- Ability to provision the exact SSD capacity that an application requires
- Centralized management.

These SAN/NAS-derived features enormously broaden the range of SSD solutions over PCIe products. For example, consider an organization that has 40 servers, each needing 250GB of SSD capacity. The systems administrator could consider two options: first deploy forty 450GB RamSan-20s, one for each server; or second, deploy a RamSan-630 and provision forty 250GB LUNs – one for each server. In the first scenario each server must have maintenance scheduled, take downtime to install the RamSan-20, and then have 200GB of excess capacity locked in the server. In aggregate, this leads to 8 TB of excess SSD capacity! And consider if the application storage requirements change down the line. With a RamSan-630, storage requirement changes can easily be accommodated, while an internal solution is effectively locked up forever.

The RamSan-630 offers the flexibility of shareable externally attached storage that can be integrated with host side or SAN/NAS side software features such as data deduplication, data replication, and snapshots. The RamSan-630 installs in minutes and appears to the server operating system as simply another storage drive, albeit a blazingly fast one. The RamSan-630 connects to its host or SAN/NAS fabric using Fibre Channel or Infiniband to support its very high performance and low latency capabilities.

### Fibre Channel

The RamSan-630 uses Fibre Channel interfaces to connect to up to ten host devices directly and to storage networks for connection to a virtually unlimited number of devices.

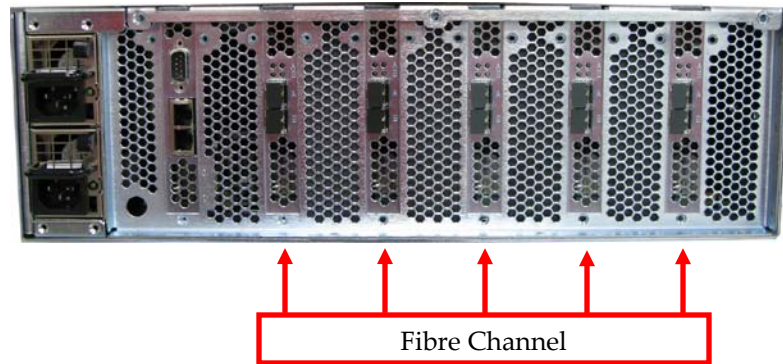


Figure 2: Back Connectivity of the RamSan-630

Our newest Fibre Channel interface, the FC-341, includes two 4-Gbit Optical (LC) Fibre Channel interface ports. The FC-341 supports both 4- and 2-Gbit devices and auto-negotiates to operate in the proper mode. The RamSan-630 can support up to five FC-341s in a single chassis. All ten ports can be active, enabling LUN sharing across servers and various multipathing and clustering schemes. The FC-341 supports the following Fibre Channel topologies:

- *Point-to-point*: The RamSan-630 properly implements the point-to-point (n-port to n-port) topology for Fibre Channel. Through the point-to-point topology, the RamSan-630 can be attached directly to up to ten host servers.
- *Switched Fabric*: The most popular way to attach a RamSan-630 is with the switched fabric topology. The switched fabric topology implies that there is a storage network switch between the host and the RamSan-630. Switched fabric topologies can be used to provide unparalleled scalability and redundancy for storage networking. Because Fibre Channel switches tend to add very little latency to storage transactions, the switched fabric topology is still an excellent choice for high performance storage.
- *Arbitrated Loop*: In many ways, arbitrated loop is the legacy topology for Fibre Channel. Arbitrated loop operates similarly to the old token ring local area networks and tends to be somewhat slower than the other topologies. In addition, arbitrated loop can support far fewer devices than the switched fabric topology. This functionality allows the RamSan-630 to be connected directly to host bus adapters and storage devices that work best with arbitrated loop.

### **InfiniBand**

The RamSan-630 will soon support one to five single ported QDR InfiniBand controllers. The system communicates with IB HCAs and switches with the SRP upper layer protocol.

---

## Reliability

It's commonly believed that Flash-based SSD is not ready for the enterprise, partly due to issues of data reliability. In the case of RamSan products using Flash, this simply is not true. But it is true that Flash, as a storage medium, presents certain challenges for storage engineers.

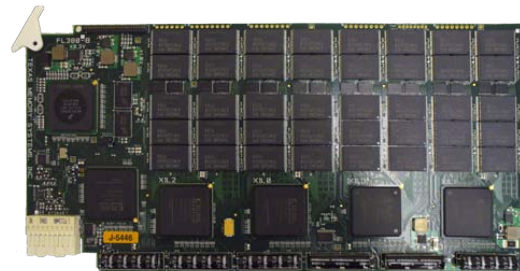
The storage medium in the RamSan-630 is Single Level Cell (SLC) NAND Flash. SLC NAND Flash is a requirement for enterprise applications where there is any kind of mixed read-write workload and heavy sustained use. Some SSD vendors have experimented recently with Multi-Level Cell (MLC) NAND Flash chips. But MLC is slower and offers only 10% of the reliability of SLC. From the TMS perspective, use MLC in mixed workload enterprise environments at your own risk.

In order to meet enterprise reliability requirements, the RamSan-630 includes many key reliability features in the FTF cards:

### **Three Layers of Flash Memory Protection**

#### **Layer 1: ECC for Flash Memory**

A proprietary Error Correction Code (ECC) implementation detects and corrects single bit errors and detects multi-bit errors for each 128 byte chunk of data. This error correction algorithm is far more aggressive than industry standard solutions that provide this level of protection across 512 byte chunks.



#### **Layer 2: RAID for Flash Memory (board level)**

The RamSan-630 uses a modified RAID-5 algorithm to ensure that the failure of a Flash memory chip does not result in data corruption. The modification to the RAID-5 makes an important leap forward over HDD-based RAID-5 write performance. On HDD RAID-5 implementations, any time even a small block of data is written, the RAID-5 controller must read back the entire data stripe and the parity bits, then rewrite the data, and finally rewrite the parity. But the RamSan-630 always writes to a new location on the Flash medium as part of its wear leveling algorithm, so a read of the old data and parity before a write is not required. This makes the RamSan-630 RAID-5 implementation twice as efficient as traditional RAID-5.

### **Layer 3: Active Spare (system-level) *optional***

At the system level, the RamSan-630 allows you to designate one of the cards inside the system as an active-spare that works hand-in-hand with the chip level RAID on each of our boards. If one of the cards experiences a failure that degrades its RAID protection, the system will immediately migrate the data on that card to the Active Spare to return to a fully redundant state. At that point, the failed card will be inactive and can be swapped during the next scheduled maintenance downtime. Having the Active Spare means that a faulty card can be swapped in a matter of minutes without the need to migrate all 10TB of data off and back on to the RamSan.

### **Flash Wear Leveling**

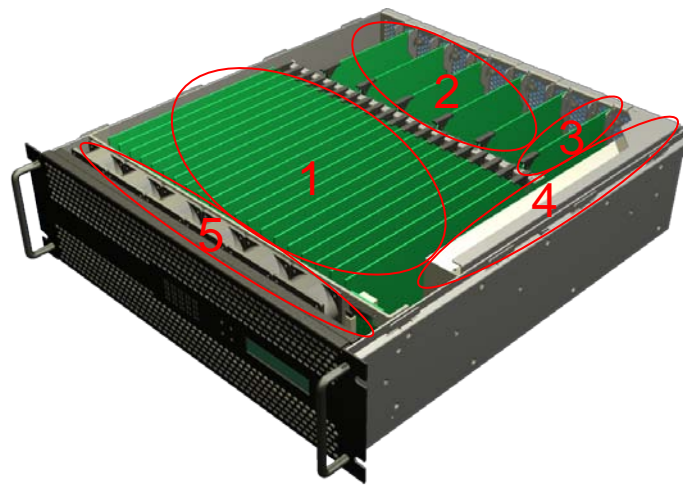
Wear leveling is a common feature in Flash SSD solutions. Wear leveling is required because the underlying Flash memory chips are specified to handle a limited number of writes. The RamSan-630 uses high quality SLC NAND Flash memory specified to handle 10x more programs/erases (writes) than the lower quality MLC NAND Flash used in inexpensive consumer devices. Even with the use of SLC Flash memory, it is still important to implement wear leveling strategies in order to prolong the write endurance of the system. The RamSan-630 includes algorithms to maximize its write life and is designed to operate for over 12 years even with a heavy stream of writes.

### **Ultra-Capacitors**

The RamSan-630 uses a small amount of RAM on each of the Flash controllers in the system. This RAM is used to store the Flash management tables for the card and provide a small write buffer. In order to ensure the buffered data is written to Flash, each card in the RamSan-630 includes an ultra-capacitor. If external power is lost, the ultra-capacitors will automatically engage and provide ample amounts of power to back up the RAM contents to the Flash storage area. This ensures that the RamSan-630 is an entirely enterprise-grade non-volatile storage device.

### **Fault Isolation**

The RamSan-630 is designed to isolate any failures within the system. The major components are illustrated below. A passive backplane provides multiple connections between the FTF cards, the multiple controllers in the system, the redundant power supplies, and an out-of-band management card. The management card coordinates system startup /shutdown, monitors system environmentals, sends out fault alerts, and provides a management interface. It is not in the data path and a failure here does not impact data availability. This design allows any failure on a card to be isolated to that card only and allows the rest of the system to operate normally.



- 1 – Twenty 500GB FTF cards
- 2 – Five I/O interface cards
- 3 – Out of Band Management card
- 4 – Redundant power supplies
- 5 – Redundant fan banks.

Figure 3 - Inside the RamSan-630

## Availability

TMS has been designing and building high availability enterprise solid state storage for over three decades. The Ramsan-630 is no exception. TMS anticipates a few common deployment models for the RamSan-630:

- **Just a Bunch of Flash (JBOF):** A separate LUN can be presented for each FTF card in the RamSan-630. This flexibility allows you to mirror LUNs for maximum data protection. These LUNs can then be accessed externally by servers or storage appliances.
- The RamSan-630 units can also be deployed as single modules/arrays. In this mode LUNs are created that stripe across all of the FTF cards. This mode provides the maximum capacity and performance but limits the availability of the system as each LUN is created as a RAID 0 across multiple highly reliable FTF cards.

The RamSan-630 includes redundant, hot swappable power supplies and redundant cooling fans. SNMP alarms, the browser-based management console, and the system's front panel simultaneously alert administrators to any failures. To assist in this process, Texas Memory Systems offers a variety of support options, including next day or 4-hour onsite support, advance parts replacement, and on-site spare kits for all serviceable components.

For more information on Texas Memory Systems support options, visit [www.RamSan.com/warranty.htm](http://www.RamSan.com/warranty.htm).

## Scalability

The RamSan-630 provides 10TB of usable capacity (13.7TB raw), and users can scale capacity from 4TB to 10TB. The RamSan-630 can present a single 10TB LUN for application use. Enterprises starting with 2TB configurations can expand storage capacity as their databases grow.

An added benefit of having higher capacity per chassis is decreased total cost of ownership for large SSD configurations. Total cost of ownership comes from the lower price per capacity available at high capacities in a single chassis and from the lower management costs associated with having fewer discrete systems to manage.

For customers with much larger data sets, multiple RamSan-630 systems can be arrayed. For example, a single rack containing 14 RamSan-630 units can provide 140TB of usable Flash storage capacity and sustain 112GB/second of random data transfers.

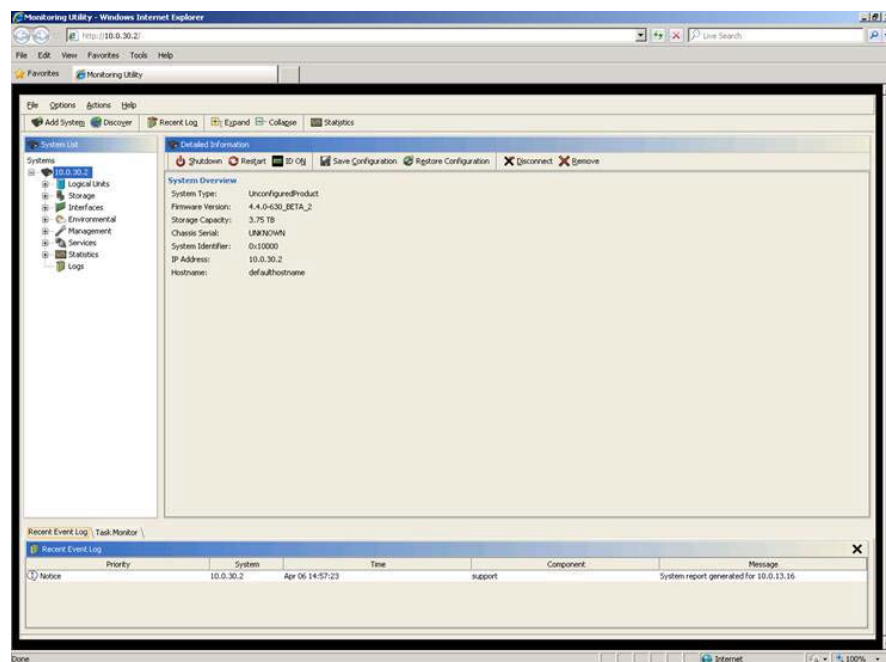
The RamSan-630 includes slots for five dual ported 4-Gbit Fibre Channel interfaces or five QDR InfiniBand interfaces. Some of our customers use this capability to provide active: active multipathing configurations, while others use additional slots to directly connect to multiple servers. Each controller added to the RamSan-630 increases the performance available to performance-starved applications.

# Section 7

## Manageability

The RamSan-630 includes the most management features of any SSD to ensure the most effective and efficient use of your investment.

Administrators can access the RamSan's management capabilities over the network or directly from the unit. From either a Telnet or a browser interface (all password protected), administrators can monitor the RamSan's various sensors and status alerts and fully configure the system.



**Figure 4: Screenshot from the Browser Configuration Interface (system information)**

From these interfaces, administrators can:

- Configure from 1 to 1024 LUNs of variable capacity that can be used by a single server or shared across multiple servers
- Configure Fibre Channel or InfiniBand controller settings
- View all SNMP and email alerts, system temperatures, and the status for every major component
- Specify the IP address for the RamSan
- Monitor system health.

Unlike many other storage appliances, the RamSan-630 also incorporates a front panel display and basic control scheme. The front panel display shows Fibre

Channel or InfiniBand port activity during normal operation and displays any system-generated warnings. Using the buttons next to the front panel display, the user can configure the RamSan's IP address and initiate a manual shutdown of the unit.

The RamSan-630 is fully compatible with existing SNMP monitors.

## For More Information

The RamSan-630 is the densest high-performance SLC Flash-based enterprise storage on the market. It is completely non-volatile and incorporates a suite of reliability and error-correction strategies. The RamSan-630 offers a faster, smaller, and more reliable alternative to traditional HDD RAID solutions.

Now that SSD at this price, performance, and capacity has arrived, the days of the exclusively HDD-powered enterprise data center are over.

Texas Memory Systems specialists are available to discuss what the RamSan-630 can do in your particular application or environment. Call the main office in Houston, Texas, at 713-266-3200 or do one of the following:

- For more in-depth information, visit [www.RamSan.com](http://www.RamSan.com)
- Existing customers contact [support@RamSan.com](mailto:support@RamSan.com).
- Potential customers contact [sales@RamSan.com](mailto:sales@RamSan.com).