

## **SEPATON S2100 AND THE DATABASE BACKUP CHALLENGE**

**NOVEMBER 2011**



Enterprise data centers are experiencing huge challenges around massive data growth. There is no stopping the data flow: digital information will only grow and expand over time. Data retention management is important to controlling the size of older and inactive data volumes, but the enterprise is left with the challenge of storing and protecting very large data volumes on production systems. This environment requires enterprise IT to provide high levels of storage performance, protection, scalable capacity and control to mission-critical data environments – but they must do it in the face of constrained space, budgets and management overhead.

Much of this storage challenge occurs around business-critical databases. Large enterprise depends on databases for their business applications, and the databases depend on fast and secure backup and recovery service level agreements (SLAs). Critical information must be carefully protected and frequently backed up for fast restores and must remain highly available for disaster recovery and compliance. Adding to the challenge, multiple large business databases are the enterprise norm. From very large ERP (Enterprise Resource Planning) databases like SAP and Oracle to massive data warehouses, when the data environment depends on tape for database backup (and many of them do) then backup and restore performance may be badly affected.

This is why SEPATON purpose-built the S2100 platform as the only backup appliance dedicated to managing and optimizing large database backup and recovery in enterprise environments. SEPATON engineered the S2100-ES2 to reduce complexity and cost in the enterprise data center and extends that value to the edge with the S2100-DS3. The SEPATON S2100 appliance serves large enterprise storage environments with high backup and recovery performance, replication, deduplication and scalable capacity. This enables enterprise IT to cost-effectively meet critical database SLAs with a single unified storage system.

This Solution Profile highlights the SEPATON S2100 platform, which cost-effectively meets stringent SLAs around massive database protection requirements in the largest enterprise environments. Client scenarios will demonstrate how the S2100 can transform challenging database environments into a cost-effective, simplified and high performance storage system.

### ***Big Data in the Enterprise***

Enterprise storage administrators commonly report data growth rates of 30% and higher year over year. Some of this growth comes from unstructured data like corporate file shares or SharePoint content, but much of it occurs in the large mission-critical databases that are the information lifeblood of the enterprise. Both types of data need to be backed up and restored but it is the databases that often present the most difficult challenge for timely backup and recovery.

Changing backup from tape-based to disk-based is a good start but databases like Oracle, DB2 and SQL offer specific challenges that even disk-based backup alone cannot provide. Database SLAs

require backup with high performance ingest and deduplication along with very fast recovery speeds. At the same time IT operates within budgetary and space constraints, and must control purchasing and maintenance costs around management overhead, storage provisioning, technology refreshes and data center space/energy costs.

This is not an easy problem to fix but serious pain points are driving the enterprise towards change. Let's drill down into some of the most common challenges that affect backup and recovery environments charged with protecting large databases.

- **Massive data volumes.** Managing backup windows and storage capacity for growing databases is complicated and time-consuming. Frequent provisioning characterizes data centers whose massive backup volumes threaten capacity thresholds. Another challenge in these environments is the speed of backup, restores and vaulting to tape, which may not be fast enough to meet SLAs. For example, important databases may be backed up twice a day... but if backup takes 15 hours and is supposed to occur twice in a 24-hour period; then you do the math. Administrators must either sacrifice optimal backup schedules or deploy multiple storage targets with all those attendant headaches.
- **High management overhead.** In their effort to back up big data, IT often relegates reporting and monitoring tasks to the back burner. This makes it difficult to know if a backup operation has even completed properly, let alone optimizing the process. Other management tasks also go begging including disaster recovery (DR) planning and testing, meeting regulatory compliance requirements, and proactively preparing for eDiscovery requests. And the management tasks that IT is able to do are labor-intensive and highly repetitive.
- **Sprawling backup storage.** Sprawling backup environments are expensive to manage and provision. Just keeping the backup running becomes the only real driver with little thought put into a maximum return on technology investments. Over-purchasing and wasting existing technology resources are endemic problems in these data center environments, as are higher energy costs from poorly controlled storage growth. And some data centers have the added restriction of overcrowded and aging regional energy grids, which is no joke in some dense population centers.

### ***SEPATON S2100: Meeting the Database Backup Challenge***

A high-performance, cost-effective and scale-out storage system is key to providing effective backup and recovery for large enterprise databases. SEPATON S2100-ES2 is a grid-based data protection platform offering fast backup, restore performance and deduplication resulting in high capacity storage gains. The S platform is the only backup appliance that is purpose-built to manage and optimize large database backup and recovery in enterprise environments. The S2100 eliminates the slow performance of tape-based backup, and the uncertain performance of commodity disk-based backup for enterprise databases.

The S2100-ES2 deduplication system is highly scalable, offers flexible multiprotocol support and provides fast deduplication and virtualized storage pooling. DeltaStor deduplication operates across multiple SEPATON Replication Engine (SRE) nodes offering concurrent backup, deduplication, replication, and restores. Backup speeds reach up to 5.4 TB/hour per node with concurrent deduplication rates up to 25 TB per day per node. DeltaScale grid architecture lets administrators modularly add SRE nodes for more processing power or capacity with additional disk shelves. The S2100-ES2 can provide 1.6 PB of usable capacity before deduplication in a single system.

The S2100-ES2 system is able to use economical SATA disk while preserving deduplication performance. While SATA performs well enough out of the box on single stream backup, it is not as effective with the intensive data I/O that characterizes database backup. This leads to a dilemma: cost-effective SATA disk keeps the cost of disk-based storage down but disk speeds may retard to unacceptable levels when backing up large sequential streams. SEPATON developed its Dynamic Disk File System (DDFS) to efficiently backup and restore both large and small data volumes using SATA disk. DDFS handles large, sequential I/O particularly well. It does not break up database backup between virtual cartridges but instead dedicates entire virtual volumes optimized to the type of incoming backup stream.

The disk file system also efficiently balances I/O streams across multiple disks. Built-in features monitor disk array and controller performance to determine the optimal path for the waiting I/O in the deduplication queue. This architecture enables SEPATON to deliver 4-Gb/sec FC or 10Gb/sec Ethernet performance at backup time. SEPATON's ContentAware engine further fine-tunes deduplication and reporting by capturing information such as the backup application, file types, and more from the backup stream.

### SEPATON FOR THE ENTERPRISE DATABASE ENVIRONMENT

Let's take a closer look at the SEPATON 2100 in the large database environment. Growing databases have far outstripped traditional tape-based backup. Backing up to disk is a huge improvement over tape but large databases have their own set of requirements that can quickly overwhelm a basic disk-based backup product. In these circumstances administrators are forced to purchase more and more systems just to meet backup window requirements.

In contrast, SEPATON S2100 is based on a grid-based scale-out architecture that is optimized for large database backup environments. Databases such as Oracle may store data in sub-8 KB blocks. There is a good deal of duplicate data stored in these tiny blocks but if the deduplication engine only works on segment or chunk sizes, then they will not recognize less common data in a smaller than 8 KB-sized block. This will miss many deduplication opportunities, which will result in worsened deduplication ratios. The S2100 is purpose-built to recognize common data in sub-8 KB blocks, and deduplicates at the byte level for greater database backup space savings and maximal WAN optimization for replication.

SEPATON DeltaStor running on the S2100 compresses and deduplicates databases to the tune of 24:1 or higher and it does it without impacting backup performance. Also there is no need to compensate for database multiplexing or for tools like Oracle RMAN. RMAN's changes to backed up Oracle data can throw off deduplication ratios in other systems. This does not occur in the SEPATON environment.

The S2100 is characterized by performance and capacity scalability, and administrators may easily and non-disruptively add extra nodes and SATA disk shelves. The S2100's high throughput of up to 1500 MB per second per node enables compliant backup windows for even the largest databases. Additional nodes scale ingest speeds. The platform's DeltaCache Recovery feature keeps the most recent backup in cache making it available for immediate restores and tape vaulting as needed.

Replication is another crucial aspect to protecting critical data. SEPATON's DeltaRemote software only replicates new data at the byte level, which provides high bandwidth reduction and fast transport speeds. The S2100-DS3 can replicate from remote and branch offices to the S2100-ES2 data center system. The S2100-ES2 also provides high data availability for critical business databases where high application priority demands nearly instantaneous disaster recovery in case of corruption or loss.

S2100-ES2 is exceptionally available with 99.999 percent data availability, RAID-6 with hot spare disks, redundancy and internally mirrored drives and more. (The S2100 also seamlessly integrates with tape for a tertiary level of long-term data retention.) S2100-ES2 supports storage pooling, which allows administrators to manage database backup storage with discrete virtualized pools that can be segregated by database type or multi-tenancy. DeltaView's monitoring and reporting tools will work on individual pools, such as reporting on data movement or compliant proof of deletion for highly regulated database information.

## ***SEPATON in the Real World***

### **CUSTOMER STUDY #1: MAJOR SCIENTIFIC ASSOCIATION**

This customer is a premiere members association that serves an important scientific discipline in academia, business and government. They publish scientific journals and magazines requiring extensive online content storage, offer publication subscription and publishing management, and provide business services to dozens of member and affiliated societies.

A centralized Oracle database houses content, business operations and publication management applications. The association used EMC NetWorker backup application and a tape library for backup and restore operations on their critical databases. They were satisfied with NetWorker but backup to tape was another matter when it came to restoring data. Tape was regularly cycled to offsite tape storage and getting this data back often took 6 hours or more to transport and restore. They decided that a VTL would allow them to retain NetWorker backup and tape settings but provide much faster restore performance.

For the association, the SEPATON S2100-ES2 VTL fulfilled its two primary requirements: extensibility and performance. Deployment was immediate and the VTL slotted directly into their existing backup infrastructure. They found that they could locate and restore any data they needed within minutes instead of hours. They plan to grow speed and capacity with the S2100's highly scalable system. They retain an off-site tape vault for disaster recovery purposes but plan on replacing it with DeltaRemote replication. They will also add DeltaStor deduplication for even better system capacity.

### **CUSTOMER STUDY #2: FORTUNE 500 ENERGY COMPANY**

This customer is one of the largest energy businesses in the U.S. The company builds energy infrastructure and operates utilities, and offers many additional energy-related products and services.

The company depends on SAP whose business-critical databases must remain highly available and protected. It uses Symantec NetBackup to create incremental nightly and full weekly backups. Large SAP databases account for 80% of backup made up of 5-20 TB incremental backup and 25-30 TB full backup. This large amount of data was backed up to a physical tape library. The incremental nightly backups occurred within backup windows but the full weekly was passing the 48-hour mark. This was unacceptable and the company decided to explore disk-based VTL technology.

They reviewed several leading VTLs and were most impressed by SEPATON DeltaStor deduplication. The customer bought two S2100-ES2 systems with DeltaStor. They installed the three-node 240 TB system in one data center and the one-node 60 TB systems in a second data center for disaster protection, using DeltaRemote for replication. The SEPATON systems have enabled the customer to slash backup times and reduce storage capacity requirements, while greatly improving their disaster protection.

### CUSTOMER STUDY #3: LARGE WIRELESS COMPANY

The U.S.' largest wireless company serves 80 million customers with voice and digital wireless services.

They manage many of their critical business functions using large Oracle databases. They use Tivoli Storage Manager (TSM) to backup the massive databases to physical tape drives in three large data centers. A major data center reorganization process included establishing multiple detailed backup policies. This along with massive data growth compromised the backup to tape process, which could no longer meet SLA requirements. The company countered tape's sub-par performance by using high performance Fibre Channel disk in the TSM disk pool but other challenges remained. These included large data center footprints, backup and recovery infrastructure sprawl and no replication to remote sites.

After testing leading VTLs, the SEPATON system proved the best fit for the wireless company with high performance, best single-system capacity and robust remote replication. The company bought two S2100-ES2 systems, each with 168TB and six nodes for a total of 336 TB of physical usable capacity before compression and deduplication. They also utilize DeltaStor deduplication, which increases physical capacity many times over. SEPATON is now the standard data protection platform for the company's divisions and subsidiaries. Benefits include much smaller backup and restore times, the ability to meet Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO), and exceptionally high backup reliability.

They have continued to grow their SEPATON solution by installing another system at the third data center.

### Benefits

SEPATON S2100 offers many benefits to the backup and restore environment in general and also offers specific benefits to the challenging enterprise database environment. These benefits include cost-effective data protection, recovery and management, lowered risk for mission-critical applications, and efficient and compliant backup management as well as the high performance, high availability and massive scalability that high priority data protection requires.

- **Cost-effective data protection and lowered risk.** Efficient capacity optimization makes continual storage purchases unnecessary. When the enterprise does need to purchase new capacity then they simply purchase additional nodes for processing or cost-effective SATA disk for capacity. SEPATON greatly lowers the risk of failing service levels from inadequate backup windows or changing configurations. Multi-protocol features help SEPATON to flexibly fit into evolving backup environments. A grid-scalable, modular architecture also enables non-disruptive technology migrations and helps to leverage existing technology investments. And by reducing the amount of storage systems to be managed, SEPATON reduces management overhead, lowers power and cooling costs, and relieves the pressure on limited data center space.
- **Less management overhead.** SEPATON greatly simplifies backup management tasks around critical databases. Provisioning and storage virtualization are efficient operations, and automatic balancing between backup disks eliminates manual backup optimization procedures. A central policy engine makes it simple to update policies around backup targets, performance, replication and schedules. 24x7 monitoring and reporting enables IT to produce compliance-based reports for regulated environments and internal governance.

- **High performance and availability.** It is not unusual for critical databases to be backed up twice daily. High throughput and fast recovery are essential for enabling this demanding schedule. SEPATON provides fast deduplication by launching deduplication passes as soon as a single data stream has completed backup. Immediate restores are accomplished with DeltaCache Recovery, which retains the most recent full backup in cache. The SEPATON system is highly reliable and available by monitoring and remediating common causes of backup failures.
- **Highly scalable and flexible.** SEPATON's scalability allows administrators to easily expand backup and recovery performance, grow storage capacity and accelerate replication. In addition, flexible multi-protocol features let users adopt the type of protocols the enterprise data center needs instead of letting their VTL lock-step them into less optimal choices. The result is greatly improved ROI and leveraged technology investment.

### *Taneja Group Opinion*

The storage environment in the enterprise data center is changing rapidly. The highest percentage of data growth is unstructured data, which the S2100 handles perfectly well. But the truly critical storage challenge belongs to the large databases that enable crucial business applications. Storage administrators for massive databases need to lower storage risk, cost and complexity while improving backup speeds and capacity utilization.

This is not an easy task and may be an impossible one for traditional tape-based backup. Tight RPO and RTO frequently require multiple daily backups, fast replication and mirroring; not features that tape is likely to support. But SEPATON can.

Its purpose-built S2100 lowers the complexity, cost and risk of protecting massive databases as well as unstructured data. The grid-like architecture yields fast backup and restore, economical scaling for smooth capacity growth, and centrally managed storage pools. Deduplication creates highly efficient storage and global replication extends data protection across the enterprise. The resulting capital and ongoing savings result in excellent ROI.

The world of big data has many storage entrants but only a few are capable of providing acceptable levels of performance, capacity, data protection and economy. We find that the purpose-built SEPATON S2100 meets these requirements and much more – making the S2100 a winner in the enterprise database protection race.

---

NOTICE: The information and product recommendations made by Taneja Group are based upon public information and sources and may also include personal opinions both of Taneja Group and others, all of which we believe to be accurate and reliable. However, as market conditions change and not within our control, the information and recommendations are made without warranty of any kind. All product names used and mentioned herein are the trademarks of their respective owners. Taneja Group, Inc. assumes no responsibility or liability for any damages whatsoever (including incidental, consequential or otherwise), caused by your use of, or reliance upon, the information and recommendations presented herein, nor for any inadvertent errors that may appear in this document.