

New Database Technologies Usher in New Approaches to Data Protection and Disaster Recovery

September 2007



Although Disaster Recovery (DR) and data protection require cross-IT coordination and planning, the accountability and responsibility for implementing DR and data protection frequently falls to the storage team. Technologies such as replication, snapshots, and backup have become the “bread and butter” tools in the storage administrator’s toolkit.

Traditionally, storage managers have utilized storage array based capabilities to safeguard their mission-critical applications. However, with the development and maturing of application-integrated and data-aware protection technologies over the last few years, storage teams face a much more complex decision-making process. Nowhere is this choice more important than database-centric storage environments.

Given storage teams’ pivotal role in DR and data protection, storage administrators and managers need to be cognizant of database-centric DR and data protection approaches and carefully analyze whether to use them instead of traditional storage-centric technologies to achieve the desired Recovery Time Objective (RTO) and Recovery Point Objective (RPO). In this article, we examine how new product capabilities at the database level are altering the calculus of data protection and continual access to information stored in the database.

Two Concepts Revolutionizing Data Protection and Disaster Recovery

In light of recent events, end users have had to revisit how to safeguard and plan for local and catastrophic outages. Over the last few years, two salient technology trends have emerged that we think alter how users should think about protecting their most mission critical systems:

- 1. Data Protection & DR Approaches Become Data-Aware.** Traditionally, most DR and data protection technologies

were “dumb” – copying bits to disk or across the wire without knowledge of their meaning, or structure. However a new class of intelligent data protection and disaster recovery approaches has emerged that understand the underlying data structures of the application. As a result, application vendors are subsuming more storage management, data protection, and disaster recovery capabilities into their applications obviating the need for expensive storage-centric technologies at the array or network level. In short, this new class of protection and DR tools can significantly

speed recovery of data and ensure better efficiency than approaches that simply protect data at the bits and bytes level.

- 2. Continuous Data Protection (CDP) Changes the RTO/RPO Calculus.** CDP has emerged as a major industry buzzword. However, the core concept of continuously capturing data changes coupled with the ability to create Any Point In Time images of a volume has the power to compress RTO and RPO to seconds from hours and days. Despite the hype, CDP, if implemented properly, has tremendous merit for IT users.

Taken together, these trends redefine what is possible in terms of data protection and DR and require storage administrators to reassess their current approaches.

Traditional Storage-Centric Approaches to Data Protection and Disaster Recovery

Today, storage teams rely on a combination of copy creation and mirroring/replication techniques to meet the desired RTO and RPO. They use snapshots to non-disruptively create point-in-time copies of the database that then can be backed up to disk or tape. A snapshot provides a consistent recovery point from which the database can be restarted. In addition, storage teams employ asynchronous or synchronous mirroring in order to create copies of the database in other locations and protect their systems from catastrophic events.

The Drawbacks with Storage-Centric Disaster Recovery and Data Protection Technologies

Storage-centric technologies provide a simple horizontal solution for safeguarding data across all applications and data types. However, with this lowest-common-denominator approach, storage teams make several crucial sacrifices that impact the overall manageability and recoverability of the underlying storage infrastructure. Here are the most common tradeoffs and issues that we hear from users:

- **Limited functionality and automation** – Storage-centric technologies have no knowledge of the underlying data structures or consistency points of the database. As a result, after a failover, manual steps have to be invoked so that the database can perform crash-consistent recovery before it is online and able to process requests. This lengthens the recovery time. Similarly, if the underlying bits get corrupted because of physical or logical failures, storage-centric technologies will not recognize this condition and will mirror the corrupted bits to the target storage volumes as well; thereby, severely compromising recovery.
- **Bandwidth Inefficient** – Since storage-centric replication approaches do not understand the database's data structures, they often end up replicating substantially more write operations than are necessary, even when they replicate only the incremental changes. This lack

TECHNOLOGY BRIEF

of intelligence results in more bandwidth consumed and a higher telecom bill for the datacenter.

- **Cost Prohibitive** – Storage vendors have typically charged additional money for high-end software capabilities such as replication, snapshot (in some cases), and CDP. This can greatly increase the overall cost of a high-end storage system.
- **Hardware Lock-in** – End-users often want to use high-end storage to store their database data in their primary data center, but prefer to use more cost-effective storage at their secondary site. Unfortunately, many storage-centric approaches require users to replicate from one high-end storage system to another of the same make and model, forcing users to be locked in to a single vendor or product family.

The Case for Database-Integrated Disaster Recovery and Data Protection Approaches

Intelligent data protection and disaster recovery technologies integrated within the database kernel mitigate the aforementioned drawbacks of the storage-centric approaches. Let's examine the key advantages to leveraging replication, snapshots, and CDP at the database level, instead of at the storage device level:

- **Data Aware** – A database, unlike a storage system, can create continuously consistent snapshots of the data that do not require crash recovery procedures. As a result, database-centric approaches offer greater granularity in terms of recovery and enable hot standby operations that storage-centric approaches cannot provide. For example, a DBA using native database protection capabilities can rollback modifications to a table, a transaction, or the entire database to any point in time. Furthermore, in the case of disaster recovery, a production database can replicate data to a hot standby database, which is online and ready to process requests. Thus database replication obviates the need to perform a lengthy crash recovery operation before startup. Finally, DR solutions integrated within the database offer more resilient data protection because they do not mirror corrupted bits to the target storage volumes. Since the data is already validated, end users do not have to wait until the disaster failover time to determine whether the data in the protected storage volumes is valid or not. In short, data aware protection technologies compress RTO and RPO SLAs.
- **Hardware Agnostic** – Database data protection and DR technologies interoperate with any type of storage system. Unlike array-based replication technologies, database-centric storage technologies allow users to choose a high-end storage system at the primary and a more cost-effective midrange storage system at the secondary site to reduce overall cost of the deployment.
- **Bandwidth Efficient** – Unlike some storage-centric replication technologies,

database replication technologies propagate only selected incremental changes to the secondary site, allowing end users to reduce their telecom bill and utilize scarce bandwidth efficiently. This is more efficient than storage-level change tracking because sending the logical description of a change is usually more compact than sending the bits.

- **No Distance Limitation** – Database resident replication technologies can ship data to any location around the world using WAN IP links and are not limited by Fibre Channel distance limitations as with some storage-centric mirroring solutions.

Database-centric protection approaches do have some drawbacks. They only focus on what is stored within the database, without considering data or files stored outside the RDBMS. Moreover, a database approach only works with a particular database vendor's software and is not extensible across other RDBMS and applications.

Assessing the Database Vendors' Data Protection and DR Capabilities

We will now turn our attention to the three major database vendors, Oracle, Microsoft, and IBM, and examine their product strategies and capabilities as they relate to DR and data protection. All these database vendors for several years have supported log-shipping as a baseline mechanism for data protection and disaster recovery. Log-shipping is the process of copying the database transaction logs (that capture the

database changes) on the primary database to a secondary (or standby) database and reapplying the transactions to the secondary system. Although a common and simple approach, basic log-shipping requires some manual oversight for managing the DR configuration.

Over the past few years, database vendors have begun to subsume more advanced recovery capabilities previously thought to be the domain of file systems, volume managers, and storage systems into their offering. As a result, we have witnessed database-centric solutions increasingly incorporate technologies, such as CDP, snapshots, replication, and storage virtualization into their products – often for free. In short, the database, like file systems, has become an extension of the storage infrastructure and as a result the storage administrator needs a working knowledge of what each RDBMS kernel can provide.

ORACLE DATABASE 10G & 11G

Oracle provides the most comprehensive and sophisticated suite of data protection, disaster recovery, and storage management capabilities of any of the major database vendors. Storage administrators operating in an Oracle environment need to be aware of at least three core Oracle storage technologies – Oracle Data Guard, Oracle Flashback Technology, and Oracle Automatic Storage Management (ASM).

Oracle Data Guard. Data Guard, a standard feature of Oracle Database 10g or 11g, synchronously or asynchronously replicates granular database changes on the

T E C H N O L O G Y B R I E F

production database to one or more remote or local hot standby databases. Data Guard validates all changes before they are applied to the standby database, preventing physical corruptions that occur in the storage layer from causing data loss and downtime. If the production database becomes unavailable because of a planned or an unplanned outage, Data Guard automatically switches a standby database to the production role, allowing client applications to automatically connect to the new production database. This approach offers high levels of both RPO and RTO.

Rolling upgrades using a Data Guard standby database significantly reduce downtime required to upgrade to new database releases or to perform other maintenance tasks. Also, because of its underlying knowledge of Oracle data structures, Data Guard transmits the database changes much more efficiently over the network compared to storage-centric mirroring, enabling even synchronous replication over hundreds of miles.

Active Data Guard, a new capability introduced in Oracle Database 11g, allows standby databases to offload real-time queries, reporting, fast backups, and testing from the production database, increasing the ROI in standby systems without compromising RPO. In summary, for availability and protection of Oracle databases, Data Guard represents a much higher performance and a much more bandwidth-efficient substitute to traditional storage-centric mirroring technologies.

Oracle Flashback Technology. Oracle's Flashback capabilities provide historical analyses and CDP for the database. To identify problematic changes, users can query a set of rows at a past point-in-time, view the row changes (and associated transactions) between two points-of-time, and view all changes effected by a particular transaction. To repair the problems, Flashback Technology provides a near instantaneous, non-disruptive mechanism for recovering the entire database, a table or row within the database, or rolling back a transaction (including its dependent transactions) to any point in time.

Oracle Flashback Technology provides a level of granularity and performance optimization that storage-centric CDP simply cannot match. While storage-centric CDP solutions can recreate a volume at any point in time, Oracle Flashback allows administrators to easily rollback individual logical elements of the database (e.g. tables, transactions) and rewind operator errors that might occur. Finally, while storage-centric CDP solutions rely on I/O-intensive copy-on-write technology, Oracle Flashback uses an I/O-optimized algorithm because of its awareness of the underlying data structures, and this has minimal impact on the production system.

Oracle Automatic Storage Management. Automatic Storage Management (ASM), another integrated feature of the Oracle database, obviates the need for third-party file systems and volume management for managing Oracle database files. Using ASM, Oracle optimizes the underlying storage performance by striping

and mirroring the Oracle datafiles over all available spindles.

ASM automatically rebalances the data, in an online manner, across all spindles in the event a disk is added or removed. This makes it feasible to add or remove disks in a just-in-time manner, depending on business requirements. ASM also improves storage efficiency by enabling multiple databases to share a common storage pool, sharing the pool's input/output processing (IOP) capacity across multiple applications in much the same way pooled servers enable sharing peak CPU capacity across applications. This enables ASM to be the foundation of Oracle's grid-storage vision. Also, because ASM is Oracle data-aware, it can differentiate between different types of database files and tune its striping and mirroring algorithms to optimize performance and availability. Policies are set by file type, so certain files can have higher levels of redundancy or perhaps a finer stripe size.

In summary, ASM provides a simplified, automatic, intelligent and integrated approach to data layout and resiliency for storage and database administrators of the Oracle database.

MICROSOFT SQL SERVER

With the release of SQL Server 2005, Microsoft also began subsuming traditional remote mirroring and copy creation technologies into the RDBMS, although the capabilities are not as rich in functionality as that of Oracle. Microsoft offers two capabilities – database mirroring and database snapshots – of which storage teams

need to be aware. Database Mirroring transfers transaction log records from one server to another allowing quick failover to the standby server. In the event of a failover, client applications can automatically redirect their connection to the standby server. Unlike Data Guard, SQL Server database mirroring is limited to a single pair of database servers – production and standby – and does not support a “hub and spoke” deployment model.

SQL Server database snapshots allow administrators to create a read-only, static view of the database at a single point in time. Database snapshots can be used to recover from operator errors or for reporting purposes. However, SQL Server does not provide a CDP mechanism that allows administrators to rollback the database or individual database objects to any point in time.

IBM DB2

Of the three major database vendors, IBM DB2 has subsumed the least amount of advanced data protection and disaster recovery capabilities into the database engine. DB2 offers an integrated replication capability called High Availability Disaster Recovery (HADR), but does not provide snapshots or CDP-level functionality in the database kernel. HADR is similar in capabilities to Microsoft SQL Server database mirroring. HADR replicates data changes, synchronously or asynchronously, from a primary database to a single standby database. In the event the primary database fails, clients can be rerouted to the standby database at another location.

Taneja Group Opinion

We believe that storage managers' focus has shifted from how to backup and protect data to how to recover data quickly and at increasing granularity. This emphasis on recovery is widespread throughout the data protection and disaster recovery markets. In that context, we believe that users should consider four points as they plan their database infrastructure:

First, users need to assess their RTO and RPO objectives on an application-by-application basis. Storage teams must break out of the "one size fits all" dogma and expand their understanding of different protection and recovery approaches available today.

Second, database-centric data protection and disaster recovery approaches offer compelling ROI benefits over storage-centric approaches and should be leveraged when feasible. Conventional wisdom has held that there is an implied tradeoff between cost and SLAs (e.g. low RTO and RPO). With the advent of application-integrated and data-aware data protection and DR technologies, users can get better RTO and RPOs at a lower cost point than what has been possible with traditional storage-centric approaches.

Third, in assessing how to protect their database infrastructure, storage teams must weigh the RTO and RPO for the database, the budget to achieve that RTO and RPO, the native capabilities of the chosen database, and how they compare to storage-centric technologies. End users who have

standardized on a particular database platform with rich data protection and DR functionality should strongly consider using database-centric DR and data protection capabilities. For example, Oracle's rich CDP, replication, and storage management capabilities make it a natural fit for any Oracle shop; end-users benefit by having a fully integrated stack, one technical support contact, etc. Furthermore, since most of the database vendors do not charge additional money for these features, end-users who are looking to contain costs can utilize them in lieu of more costly storage centric approaches.

Lastly, even though the mantel of data protection and disaster recovery may fall upon the storage administrator, storage teams must increasingly coordinate activities with DBA teams in order to jointly deliver the desired RTO and RPO objectives. This is particularly true as planning and cross-domain IT coordination become central to meeting ever increasing RTO and RPOs.

Today, database-centric disaster recovery and data protection technologies allow users to meet higher SLAs at a lower cost point than previously possible. As a result, storage teams face a new challenge. They must reassess their current approach to disaster recovery and data protection for their database infrastructure and determine whether to utilize CDP, replication, and snapshots capabilities at the database layer. With innovation comes added risk, but in the case of database-centric DR and data protection technologies, we think that end-users will find these new intelligent, data-aware technologies to be up to the challenge.



T E C H N O L O G Y B R I E F

NOTICE: The information and product recommendations made by the TANEJA GROUP are based upon public information and sources and may also include personal opinions both of the 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. The 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 which may appear in this document.