White Paper

Oracle's Zero Data Loss Recovery Appliance: A Transaction DVR for the Enterprise

Sponsored by: Oracle

Phil Goodwin
November 2016

IDC OPINION

Business leaders expect two things from IT: keep mission-critical applications available and high performing 24x7 and, if something does happen, recover to be back in business quickly and without losing any critical data so there is no impact on revenue stream. Of course, there is a gap between this de facto expectation from nontechnical business leaders and what current technology is actually capable of delivering. For mission-critical workloads, which are most often hosted on databases, organizations may choose to implement high availability (HA) technologies within the database to avoid downtime and data loss. In addition, where highly demanding applications are deployed, IT organizations should consider specialized best-of-breed systems to provide the necessary service levels and improve data protection at the best possible price. Imperatives for data protection solutions include:

- Uniquely engineered for mission-critical business applications
- Policy-based, tiered recovery goals to enable enforcement of business SLAs per database
- Agility and flexibility, such that the solution is equally efficient in on-premise deployments, cloud deployments, or hybrid deployments
- Automation, such that data protection activities become part of normal operations

IN THIS WHITE PAPER

In this white paper, IDC discusses the inherent difficulties associated with traditional backup schemes and the changing dynamics of data protection strategies. We examine Oracle's Zero Data Loss Recovery Appliance (ZDLRA) and the role it can play in providing significantly improved service levels for all types of Oracle databases.

November 2016, IDC #US41634716

DEFINITIONS

- **Recovery point objective (RPO)** is the time interval between data protection events. For example, if snapshots are taken once per hour, then the RPO is one hour. RPO indicates the maximum amount of time during which data can be lost in the event of a failure.
- **Recovery time objective (RTO)** is the amount of elapsed time from the point of a failure to the time when application services are restored.
- **Service-level agreement (SLA)** is specification of the level of service that the IT group agrees to provide to business units. RPO and RTO are the building blocks of data protection SLAs. For example, an SLA may specify service levels of a one-hour RPO and four-hour RTO. SLAs typically vary by causal event and application criticality.
- **Purpose-built backup appliance (PBBA)** is a disk-based system designed specifically as a backup target; designed for data capacity rather than IOPS performance with specialized software, particularly deduplication. Due to the lower IOPS (due to high capacity/lower performance HDDs) and deduplication, PBBAs are usually not designed to support application workloads.

SITUATION OVERVIEW

After years of consolidation, data protection is in a period of disaggregation. During most of the client/server era, during which SAN/NAS has been the dominant storage architecture, the mantra has been to consolidate storage and eliminate silos. At the same time, data protection has evolved from a tape-based once-per-day backup event to a group of technologies, including snapshots, mirrors, and replication to supplement traditional backup/recovery (B/R). Each of these technologies addresses different data loss scenarios and combines to achieve a continuum of data protection. More importantly, each technology (product) provides added granularity to both RPO and RTO; working together, these technologies provide lower RPOs and RTOs to organizations. During the SAN/NAS era of consolidation, IT organizations sought to minimize the number of similar data protection products that they deployed. For example, whereas some organizations had as many as seven backup and recovery software products deployed, best practices recommended organizations to consolidate to no more than two.

The advent of x86 virtualization, cloud computing, federated storage, object storage, and architectural platforms, such as Hadoop, has caused organizations to change the way they approach data protection. While traditional backup/recovery is still the mainstay for traditional workloads, it is inadequate to meet the SLAs required by business users for most applications. Moreover, traditional backup and recovery is labor intensive; prone to frequent failures that must be analyzed and recovered manually; is costly in terms of hardware, software, and maintenance; all leading to missed backup/recovery windows, very long RPOs, of course, and unmet SLAs. Indeed, traditional backup/recovery now mostly commonly serves as a worst-case backstop in the event that other data protection methods fail. The result is high employee costs and even higher business costs of prolonged downtime

Because traditional backup/recovery is inadequate and increasingly obsolete, IT organizations are deploying multiple methods to different environments in an effort to ensure that business needs are

properly served. Organizations may retain the incumbent backup/recovery vendors, but many have added others to address virtual infrastructure and cloud, as well as for specific workloads such as Oracle databases and Hadoop. In most cases, this means deploying a best-of-breed product strategy toward workloads rather than the common denominator compromise of the consolidation era.

This disaggregation of data protection has the potential to complicate the data protection topology within organizations, impacting vendor management, user training, and integration. Nevertheless, business needs take priority. To combat this potential challenge, organizations have found that deploying integrated purpose-built backup appliances (PBBAs) solves some of the most common backup challenges. PBBAs, when hardware and software are bundled together, simplify deployment for IT organizations by providing precertified configurations that are easier to install, implement and manage.

PBBAs have a number of inherent advantages over non-integrated backup technologies. For example, instead of managing multiple vendors with different requirements and timings for hardware and software patching, upgrades, optimizations, etc., users have a single, unified solution comprised of the hardware and software together.

When evaluating data protection and availability solutions, IDC recommends that IT organizations start by quantifying their downtime costs per application and/or line of business. Our research has found that the overall industry average cost of downtime is $100,000 per hour, though that amount can be substantially higher for business critical OLTP applications. Data protection schemes are driven almost entirely by the cost of downtime. This cost has the following two major components:

- Data loss costs can be both tangible and intangible, but the tangible costs are most easily quantified. Tangible costs quantify the dollar amount of labor costs required to recreate the data along with any direct economic loss from being unable to recreate the data. Intangible costs include reputational loss when customers learn that their order was lost, bad press, stock price hit, etc.
- Direct downtime costs' most tangible examples include the lost employee productivity who cannot perform their work and the transactions that cannot be processed (i.e., customer orders and ATM transactions). Often, organizations know the average economic value per transaction upon which this cost of downtime is calculated. Direct downtime can also contribute intangibly to reputational loss and can lead to negative publicity if the outage is lengthy or affects many users.

When this total cost of downtime is calculated and known, it is very simple for corporate managers to compare the cost of downtime with the cost of a given data protection solution and calculate the ROI against the average amount of downtime incurred per year. If the cost of the solution is less than the cost of downtime, then obviously implementing the solution makes sense. Intangibles also factor into the management decision and may tip the scales in favor of action.

A fully capable data protection scheme requires a continuum of methods and technologies. Almost all organizations use array-based snapshots, mirrors (clones), and local and remote replication. In addition, cloud is playing an increasingly prominent role in data protection. According to IDC's research, more than half (53%) of all IT organizations today use the cloud for some aspect of their data protection scheme; we expect that percentage to increase to over 80% by 2018.

Each of the following technologies protects against different data loss scenarios and also provides different RPO/RTO:

- **Snapshots** provide recovery from routine data loss events, such as user error. Snapshots do not protect against array failures because the snapshots are normally on the same array and RAID group. Snapshots often provide the lowest RPO since organizations may take snapshots as often as every 15 minutes due to their relatively low overhead. RTO may also be relatively short. It is important to note, however, that snapshots may not be of as much assistance when recovering databases because database recovery from snapshots is often a manual process and may require staging, thereby extending restore times and delivering an unknown recovery status.
- **Mirrors, or clones,** are complete replicas of a data set or a database. They protect against storage system failures as they are made to a separate one. RTO can be very low because recovery may be as simple as mounting the mirrored copy to the server. However, due to high overhead (time) and space required (a full copy), organizations do not make mirrors or clones very often. Therefore, RPO is often relatively high, 12 hours to 24 hours or longer.
- **Remote replication** provides protection against almost any failure, including loss of an entire datacenter. RPO can be adjusted based on distance and bandwidth considerations, but in most cases, it ranges from 5 minutes to 1 hour. RTO, however, can be lengthy because the time to recover an entire workload remotely may be substantial. Remote replication is usually accomplished with storage system based software between like systems.
- **Cloud backup repositories** have many of the same benefits as remote replication in terms of protection, RPO, and RTO, but without the constraint of using like systems. Cloud may also offer significant cost savings compared with managing one's own datacenter. Cloud replication is often accomplished through some sort of gateway that will make the translation from the local system to a cloud protocol, such as S3 or OpenStack. RPO and RTO are likely to be on the high side, depending to a great extent on network bandwidth.

Even in cases where organizations use snapshots and remote replication to reduce RPO, the interval between snaps or replication points may represent thousands of transactions. Organizations learn to live with the loss because they don't currently have a better solution. However, it is fair to say that most organizations would like to minimize that loss and associated cost to the smallest amount possible.

The next step in the data protection evaluation recognizes that users will want to map your various technology options to the SLAs needed for different applications, as not all applications or workloads are of equal business value yet the current landscape of generic PBBAs treat them as though they are. Not only do databases house mission-critical application data, they are highly evolved, relational entities that function very different from flat files. As a result, databases demand a higher level of sophistication than other workloads, such as email. Database recovery also presents its own unique challenges which generally require more sophisticated techniques than file systems. These sophisticated technologies specific to Oracle Database high availability include:

Oracle Real Application Clusters (RAC): RAC provides high availability by sharing the database across a pool of clustered servers. Therefore, if a server in the cluster goes down, the other instances are still available and database access continues uninterrupted. RAC is usually applied only to mission-critical applications, and therefore not to all databases in an environment. In addition, RAC most likely will not

be implemented to protect against datacenter failures. (Implementing RAC across geographically separated locations is rare, though technically feasible.)

Logs: Oracle Database tracks transactions with the use of logs which create a historical record of the changes and actions taken within the database. Users take advantage of both redo and undo logs to allow transactions to be rolled forward or backward to regain a point of database consistency in the event of system failure. When redo logs are stored or replicated to other storage systems or alternate locations, they can be very useful in recovering to a specific point in time, even for remote recoveries. However, like snapshots, redo logs usually are not configured to protect against system failures.

Data Guard: is a high availability feature in which a standby database is created for failover. This technology uses logs, whereby the logs are sent to a secondary location enabling failover to that other location with the ability to be up-to-date with recent transactions. In addition, with Active Data Guard, the secondary site can be used for activities other than failover such as reporting.

Recovery Manager (RMAN): For Oracle Database backup specifically, RMAN has been the backbone of the data protection scheme for many years. RMAN provides backup, restore and recover capabilities specific for the database and is well understood by DBAs and backup administrators alike. RMAN is also supported by nearly every backup software vendor in the marketplace.

In summary, IT organizations are being driven by the cost of downtime and data loss toward service levels that demand shorter and shorter RPOs and RTOs. Oracle users have evolved a system of tools and techniques that are well known and can meet the majority of data protection requirements specifically for Oracle Database. In order to address even more demanding future requirements, companies will look for the ways to simultaneously reduce downtime and data loss in a manner that meets their business SLAs while optimizing for Oracle Database workloads, simplifying, and consolidating data protection capabilities.

FUTURE OUTLOOK

Oracle offers a full continuum of data protection products and has set the standard for Oracle Database backup and recovery over the years, including NAS, tape and cloud backup technologies. Oracle has worked with third-party data protection software suppliers to provide compatibility of products and solutions.

Recently, Oracle introduced the Zero Data Loss Recovery Appliance, a special-purpose database recovery appliance uniquely engineered for Oracle databases as opposed to a generic general purpose PBBA. In a nutshell, the Recovery Appliance functions like a DVR for Oracle databases, continuously capturing transaction logs and allowing them to be played back at a very granular level.

Recovery Appliance embodies two important capabilities: transaction-level data protection and extended data protection using remote appliances. This combination protects organizations from a broad range of data loss threats, including data corruption, transaction errors, and system and datacenter failures. Moreover, as a single integrated appliance, it reduces the number of products that must be managed and simplifies the customer's deployment effort.

Oracle touts five key benefits to the users of Recovery Appliance:

- Virtual elimination of data loss exposure, thereby making zero RPO a real possibility
- Minimal impact backups by shifting backup processes from production systems to the appliance itself and by sending only changed blocks automatically (no deduplication required since the database is already tracking this). Oracle notes that unlike conventional PBBAs such as Dell-EMC Data Domain, the Recovery Appliance can increase production database server performance by 25% as critical backup processes are offloaded to the appliance. In contrast, Data Domain PBBAs charge a "backup tax" and impact production database servers up to 25% with all of the backup processes and agents that they burden servers with. Database-level validation and recovery, meaning that the Recovery Appliance continually validates the backup blocks, knows the exact recovery window (and displays it, for peace of mind), and can recover anything from a single transaction to an entire database.
- Scalability capable of protecting all Oracle databases in the datacenter and delivering these advanced recovery capabilities to all databases, not just mission critical, with the use of policy-driven SLAs.
- Cloud insurance to future-proof today's investments with equivalent Oracle Public Cloud offerings for when cloud dominates the deployment options.

Recovery Appliance is based on Oracle Database 12*c* and is thus a fully integrated hardware and software appliance that can be deployed in private and public clouds. Oracle has also retained RMAN as the backbone of the Recovery Appliance process. Indeed, RMAN is the backbone of what Oracle calls Maximum Availability Architecture by providing the data movement management from the database to tape, disk, cloud, and Recovery Appliance. This means that operations for Recovery Appliance, while different, will retain key familiarity for DBAs and storage administrators. Some of the new capabilities are "delta push" technology, which sends changed blocks from the production database to the Recovery Appliance. Oracle describes this as a "database monitoring a database."

Among the most interesting manifestations of this technology is the ability to support a real-time redo transport that immediately protects in-process transactions. This is a significant advance as previous technology only protected committed transactions. In environments where hundreds or thousands of transactions may be in process at any given point in time, the lost partial transactions could be significant: Recovery Appliance can help recover even these partial transactions. The delta store function of Recovery Appliance is a compressed, validated store of changed blocks. This function provides any-point-in-time recovery. Think of this as being able to rewind a DVR to any word in a movie, not just the beginning of a section.

Recovery Appliance's policy engine allows organizations to provide "data protection as a service" to their organization in both public and private cloud environments. This policy engine allows a great deal of granularity when defining SLAs for specific business units, databases, or applications. Whereas many organizations apply a default data protection profile to databases, Recovery Appliance enables cost/benefit to be matched with business requirements.

As noted earlier, a key aspect of Recovery Appliance is that, according to Oracle, it can increase the performance of the production database up to 25%. This is because it offloads the data protection functions from the database server. Other backup solutions may add a processing load to the

database server because of backup agents or backup operations such as compression, deletion, and validation. Best of all, this applies to all databases under management across the enterprise.

To complete the continuum of data protection scenarios, Recovery Appliance can enable disaster recovery services as well. The replication capabilities of the appliance are flexible enough to allow almost any configuration.

- Unidirectional, as one might do for a single-purpose disaster recovery site
- Bidirectional, in which two Recovery Appliances replicate to each other as one might do when two datacenters are used for redundancy
- Hub-and-spoke, whereby multiple remote sites can replicate to a single central site
-

Any of these configurations can include stops between a private datacenter and a device in a public cloud. Of course, Oracle also supports replication to tape when low-cost archiving of backups is needed.

The replication between Recovery Appliances can be done in real time, thereby virtually eliminating data loss even in the most dire scenarios. The self-healing capabilities of the Recovery Appliance include automatically reconciling the backup catalogs between appliances. Data is validated at ingest, and any bad blocks are automatically repaired. In addition, all databases are mirrored and striped and stored with redundancy; backup catalogs are triple mirrored.

Recovery Appliance is based on Oracle's Exadata Database Machine, with the same scaling and no single point of failure resiliency. The base configuration includes 2 compute and 3 storage nodes, though compute and storage nodes can scale independently. A maximally configured system would fill 18 racks, housing up to 100PB of virtual full backups and up to 216TB per hour of backup throughput.

Comprehensive monitoring and management cannot be overlooked. Recovery Appliance includes Oracle's Enterprise Manager for oversight and end-to-end visibility of the backup operation from the active database to the back-end storage, whether it is tape, disk, cloud, or Recovery Appliance itself. Enterprise Manager provides monitoring to help the operations team stay on top of the backup environment. It also gathers and reports on key metrics and statistics while issuing warnings and alerts based on preestablished triggers and thresholds.

CHALLENGES/OPPORTUNITIES

The Oracle Zero Data Loss Recovery Appliance is very much a special-purpose system, even within the category of purpose-built backup appliances.

The opportunity to consolidate and simplify Oracle Database backup and recovery, the ability to apply policies and protect thousands of databases enterprisewide, plus deliver a near-zero data loss solution to the make Recovery Appliance a compelling value proposition. In addition, when factoring the added performance of the database servers and the reduced man-hours of management, and the cloud

insurance that Recovery Appliance provides these organizations may find a pleasingly-quick payback on their investment and the purchase justification becomes easier.

Clearly, companies don't want to be in the business of losing data, dealing with outages and coping with the stressors of locating a viable backup copy. With the Recovery Appliance, the stress of having a good backup copy available is eliminated with real-time recovery status updates that take the guesswork out of recovering your data and enable IT professionals to "sleep at night."

CONCLUSION

Oracle's Zero Data Loss Recovery Appliance (ZDLRA) represents a best-of-breed-type market entrant for Oracle Database data protection. Although structured data may not be growing as rapidly as unstructured data, Oracle databases continue to host some of the most demanding and business-critical applications and Oracle commands 46% of the relational database market and 40% of the overall database market—making the addressable market opportunity for the Recovery Appliance significant. These applications frequently serve OLTP environments and may process thousands of transactions per minute. The tangible and non-tangible costs of both data loss and downtime, even when systems go down for a few minutes, can be surprisingly high. Recovery Appliance addresses both data loss (RPO) and downtime (RTO). The cost of data loss, employee productivity, manual labor to repair failed backups and to recover databases, as well as corporate reputation can easily eclipse the cost of the appliance solution.

Recovery Appliance addresses the broad range of data loss scenarios, from simple error and data corruption to catastrophic datacenter loss. IT organizations will no longer need to deploy, integrate, and manage a diverse set of snapshot, mirror, and replication tools to meet their data protection needs in an Oracle Database environment. Those organizations that struggle to meet their Oracle service levels, want to simplify their database data protection scheme, or want to proactively plan for future requirements of cloud deployment will want to investigate Recovery Appliance.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com