# FULL DATABASE PROTECTION WITHOUT THE FULL BACKUP PAIN
## Oracle's Cloud-Scale Zero Data Loss Recovery Appliance

### DECEMBER 2015

**THE NET NET**

Today's tidal wave of big data isn't just made up of loose unstructured documents – huge data growth is happening everywhere including in high-value structured datasets kept in databases like Oracle Database 12*c*. This data is any company's most valuable core data that powers most key business applications – and it's growing fast! According to Oracle, in 5 years (by 2020) most enterprises expect 50x data growth. As their scope and coverage grow, these key databases inherently become even more critical to our businesses. At the same time, the sheer number of database-driven applications and users is also multiplying – and they increasingly need to be online, globally, 24 x 7. Which all leads to the big burning question: How can we possibly protect all this critical data, data we depend on more and more even as it grows, all the time?

We just can't keep taking more time out of the 24-hour day for longer and larger database backups. The traditional batch window backup approach is already often beyond practical limits and its problems are only getting worse with data growth – missed backup windows, increased performance degradation, unavailability, fragility, risk and cost. It's now time for a new data protection approach that can do away with the idea of batch window backups, yet still provide immediate backup copies to recover from failures, corruption, and other disasters.

Oracle has stepped up in a big way, and marshaling expertise and technologies from across their engineered systems portfolio, has developed a new Zero Data Loss *Recovery Appliance*. Note the very intentional name that is focused on total recoverability – the Recovery Appliance is definitely not just another backup target. This new appliance eliminates the pains and risks of the full database backup window approach completely through a highly engineered continuous data protection solution for Oracle databases. It is now possible to immediately recover any database to any point in time desired, as the Recovery Appliance provides "virtual" full backups on demand and can scale to protect thousands of databases and petabytes of capacity. In fact, it offloads backup processes from production database servers which can increase performance in Oracle environments typically by 25%. Adopting this new backup and recovery solution will actually give CPU cycles back to the business.

In this report, we'll briefly review why conventional data protection approaches based on the backup window are fast becoming obsolete. Then we'll look into how Oracle has designed the new Recovery Appliance to provide a unique approach to ensuring data protection in real-time, at scale, for thousands of databases and PBs of data. We'll see how zero data loss, incremental forever backups, continuous validation, and other innovations have completely changed the game of database data protection. For the first time there is now a real and practical way to fully protect a global corporation's databases— on-premise and in the cloud—even in the face of today's tremendous big data growth.

## FULL BACKUPS ARE BECOMING OBSOLETE

Databases need to be protected – that's not in doubt. What has to change is the reliance on a backup window approach, whether it leads to a daily full backup scheme or a weekly full with daily incrementals or any other practical variation. Ensuring that solid and reliable backups are always taken

(and can actually later be used for a good recovery) has always been a challenge, but there are a number of factors making the traditional backup window approach increasingly less viable today including:

- Database growth prolongs backup time, leading to missed windows and straining backup resources
- Increasing database access and availability requirements (e.g. global, always on)
- Incomplete database coverage as the number of databases and critical applications grows
- Fragile recovery procedures that aren't guaranteed to always work
- Backup data corruption, damage, or loss from media issues or other faults at scale
- Risk of losing as much as 24 hours of data with daily full backups
- Lengthening recoverability delays (when recovery is possible)

At even just the basic level of analysis, the idea of taking time out for a full backup window is becoming obsolete. With enough database growth any given backup window will eventually be exceeded. Applications are becoming more global (24x7), demanding higher performance, and constantly adding data – all of which mean full backups become difficult to schedule and complete. Even if a particular database implementation can support a large backup in the background while remaining technically online, simple physics says any large backup process will impact the performance and availability of other concurrent applications.

Already the reality is that many critical databases aren't actually fully protected properly in production today. A disturbingly high percentage of those that are covered by some backup scheme aren't likely to actually be recoverable if and when necessary. And of those that could possibly be recovered, the time and effort to assemble a full backup (first applying incrementals and available redo logs as necessary) would likely negatively impact operations.

Many companies are now faced with an unpalatable choice - stop growing their critical databases (e.g. through more severe aggregation/archive/pruning) and thereby limit value, opportunity and even innovation, or go farther at risk of not being able to recover in the face of very real and possible disaster scenarios.

## *Continuous Database Protection*

The basic idea for continuous data protection isn't new per se (e.g. journaled filesystems, synchronous replication, etc). but it has taken Oracle's unique industry capabilities and resources to co-engineer the world's top production database with arguably best-in-class IT infrastructure (i.e. servers, storage and networking leveraged in engineered systems such as the Exadata platform) to fully embed and automate continuous database protection into an easily deployable and scalable appliance. It's a complex engineering challenge internally, but the new Recovery Appliance makes continuous database protection easy to apply in practice, ensuring databases are actually recoverable to a sub-second point in time.

At a high level, Oracle has turned the idea of backups on its head. Instead of relying on "nightly full backups", only incremental backups ever need be taken – an "incremental forever" approach. And those incrementals are optimized to only read changed blocks before transmission out to the Recovery Appliance. In addition one can then stream a database's "redo blocks " in real time directly from memory to the appliance to eliminate any data loss window. The Recovery Appliance intelligently captures, arranges, and stores this stream of incremental changes. Then, whenever requested on demand, it can dynamically and immediately serve back out a virtual "full backup" image. This eliminates several major pain points for companies of all sizes. The big question of being able to recover a good copy of the database from last week or yesterday is no longer an issue. DBAs can now simply select a specific backup time they want to recover any database to and the Recovery Appliance can immediately serve the requested recovery image. In this sense, the Recovery Appliance works like an Apple Mac OSX Time Machine for enterprise databases (or like a DVR for Enterprise databases).

The Recovery Appliance can send any necessary archived log backups to rebuild to any desired point-in-time. This is elegant architecture, one that transcends the limitations of single controller data deduplication appliances. Each of those traditional appliances maintains its own repository. Unfortunately these older approaches, lacking scale-out capabilities and high availability design inevitably lead to dedupe sprawl and storage fragmentation, not to mention an increased potential for data loss.

## ORACLE'S ZERO DATA LOSS RECOVERY APPLIANCE

First, the Zero Data Loss *Recovery Appliance* is named specifically to emphasize its role in practical, effective 100% recovery – the whole point of implementing database protection. The Recovery Appliance was designed with the end goal of providing the best possible recovery rather than just incrementally trying to improve on an outdated backup process. By focusing on delivering recovery, Oracle realized that they had all the pieces available in their arsenal to engineer and build a continuous protection solution that could deliver rapid complete recovery images on demand.
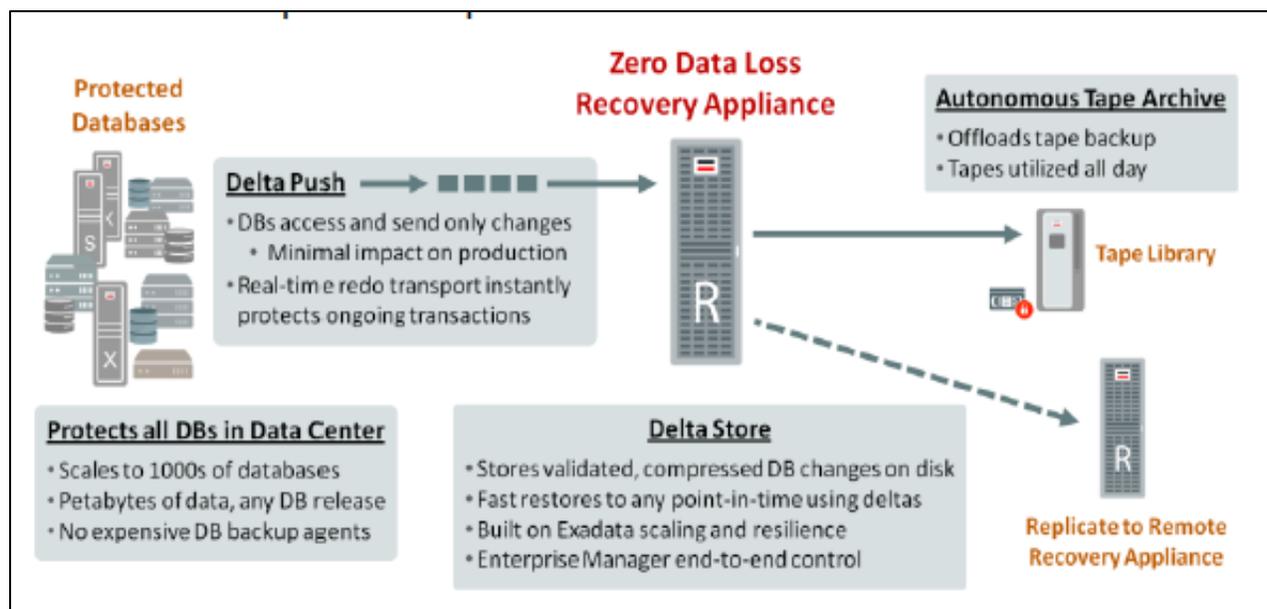


**Figure 1 – Oracle's Zero Data Loss Recovery Appliance**

First of all, because of all the Oracle co-engineering and internal development, the Recovery Appliance is an easy database protection solution to deploy and manage. For any company using Oracle Recovery Manager (RMAN), they're already familiar with using the Recovery Appliance. Before now, many companies have tried to assemble database protection schemes by stitching together multiple 3rd party solutions. This tends to create complex data flows and hard-to-manage processes. And, as with most complex multi-vendor solutions, the resulting protection is often too expensive, incomplete and ultimately becomes unreliable. In contrast, this new Recovery Appliance was engineered by the same group that developed the Oracle Database so it organically understands Oracle databases, easily deploys as a self-contained appliance, scales to protect large numbers of databases (each potentially large in size), and provides sub-second RPO. This new backup and recovery solution itself runs Oracle Database 12c as its central RMAN Recovery Catalog and backup meta data store – leveraging a world-class database to help customers protect and recover other databases. While others may have taken a more mechanical approach to recovery, it's clear that by owning the database IP Oracle has been able to take a software-centric approach that makes a lot sense – databases intimately know about database objects.

## *A Recovery Point Objective (RPO) Less Than One Second*

Oracle perhaps has an unfair advantage because they know Oracle databases intimately. This knowledge certainly works in favor of the Recovery Appliance. To virtually eliminate data loss exposure, they needed to stream the database "redo blocks" in real-time out to the appliance. Oracle was able to leverage their Real-time Redo Transport functionality originally developed for the high-availability Data Guard solution. This redo transport capability ensures that the client's ultimate exposure to data loss is now reduced to a zero/sub-second level. This is light-years ahead of alternatives such as EMC Data Domain appliances and NetApp filers when considering the hours or days of exposure to data loss that such traditional backup window-based approaches offer.

## *No More Full Backups*

Not only does this approach absolutely minimize the risk of data loss, it also eliminates the dreaded full backup window. The Recovery Appliance does take an initial full backup upon deployment, but then only requires incrementals forever after. And as introduced above, these incrementals are fully minimized and deduplicated on the server before transmission. This capability is called "Delta Push", leveraging intimate knowledge of how Oracle Database works internally.
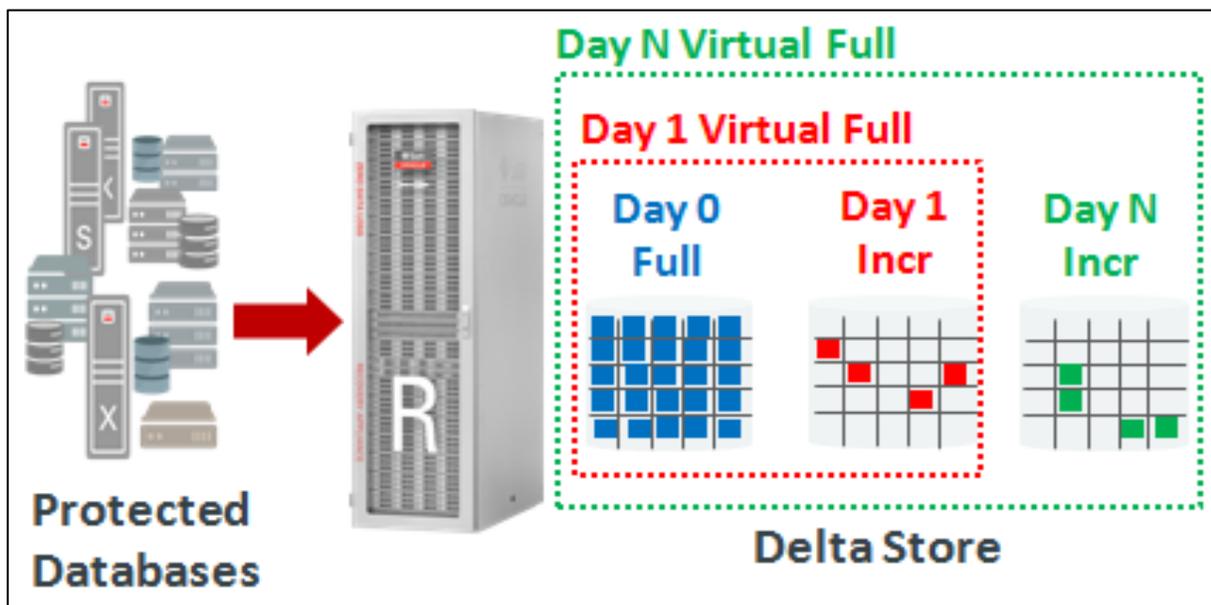


**Figure 2 - Oracle Recovery Appliance - Incremental Forever**

Since there is never any need to take another full backup (scheduled or otherwise), the full database backup window goes away completely providing some really outstanding benefits including:

- **Savings on Server and Network Resources** - No more server overhead incurred for full backups
- **Better Performance** - No more full backup contention for the database resources
- **Expanded Availability** - No more database offline time required for a full backup window
- **Real-Time Data Changes are Protected** - Never again take a full backup and impact production (or fail that backup)
- **Save People Resources** – No need for daily backup management and oversight

In addition, any further downstream backup processing, like the copying of database backup to tape for long-term offsite storage, is also offloaded from production database hosts to the appliance. And if replicating from one Recovery Appliance to another, only the same minimized/deduped changed blocks need to be sent, saving time and network resources.

## Full Recovery On Demand

Perhaps the best new trick of the Recovery Appliance is that internally it indexes, compresses, and efficiently stores the incoming changed data blocks in such a way that when asked, it can quickly produce a full backup on demand. Internally it creates a highly space efficient "virtual full backup" for each incremental point-in-time, collectively known as the "Delta Store". Whenever requested, it can effectively and rapidly serve out a full backup (the same as might have otherwise been taken as a full backup) for any point-in-time increment it has kept in store. Because these virtual full backup images are highly space efficient, the appliance can store 10x more full backup recovery points online than traditional approaches, providing a wider window/greater choice for fast recovery.

Even better, because the Recovery Appliance can also host all the recent transaction information from the real-time redo log transport capability, it can provide for a current "zero loss" full recovery. And due to the way the Recovery Appliance is engineered to scale, serving out backups for restore doesn't degrade any ongoing protection processing for other databases.

There are numerous benefits from being able to generate a full backup on demand, but the first should be obvious – the backup image is reliably going to exist!  The probability of recovery with the Recovery Appliance is multiple times more likely to be successful compared to any method in which the correct backup component files need to be restored from deep archives, then applied correctly as a series of full and incremental recoveries.  The benefits we see here include:

- **Reliability of Recovery** – As indicated, the Recovery Appliance produces an immediate full backup without fumbling/messy scripts/human intervention.  It's fully automated and eliminates any room for human error.
- **Timeliness to Recover** – The Recovery Appliance can produce any requested full backup on demand, without the need to externally assemble, correlate, or serially apply multiple recovery files
- **Recovering the Right Thing** – Because the Recovery Appliance internally manages all the incoming protected data, there is no room for human (or other) error in retrieving or applying the wrong files
- **Reducing Recovery Guesswork** – For all databases under management, the Recovery Appliance displays in real-time the current recovery status by database (both current recovery window and data loss exposure), along with out-of-the-box reports, eliminating the guesswork as to whether or not backups can be recovered.
- **Choice of Many Points in Time** – Easily recover to the most optimal incremental point-in-time (e.g. to reset to a point before corruption or hacking occurred)
- **Eliminating the Need to Babysit Backups**– Reducing management headaches, the Recovery Appliance has comprehensive monitoring and alerting when data loss exposure for any database exceeds the user-defined threshold

It's worth pointing out that recovery with the Recovery Appliance is virtually assured as backupsets are validated on the appliance itself (equivalent to *RMAN RESTORE VALIDATE* without the overhead on database servers). In our field research, we still find that many large database recovery plans have never been successfully tested/executed on key production databases, leaving those IT shops and businesses with a serious amount of protection risk. Which brings up another pervasive source of recovery failure – lack of data integrity.

## Recovery Appliance's Continuous Validation

Within any complex backup-based protection scheme, there are numerous opportunities for data corruption to creep in, especially at scale. From low-level media issues to mis-named files to a random neutron hit, flipping even a single bit anywhere can doom a recovery attempt. The larger the files and the longer they are stored, the more likely integrity will become an issue. While technology and media are constantly improving, data growth is perhaps exceeding per-GB improvements when it comes to

ensuring full database backup integrity. Too many IT folks have first-hand tales of failed recovery attempts when it mattered.

When Oracle designed the Recovery Appliance, the company made validation a continuous process such that data integrity would be paramount to ensure successful recoveries. By continuously internally checking and correcting for corruption, recovery status for all databases is known and reported in real-time. Because the Recovery Appliance understands internal Oracle Database block formats, all data and redo blocks in motion are validated as they are received, copied to tape replicated, and recovered. All data at rest is also periodically scanned and repaired.

## *The Recovery Appliance*

What's inside the Recovery Appliance itself? Fundamentally Oracle used their Exadata design approach so that the Recovery Appliance is fully scale out. The base configuration has two high availability compute servers and three storage servers connected with InfiniBand providing 94 TB of usable capacity. Of course, the Recovery Appliance is designed to be able to scale to fully protect thousands of business-critical databases.

The base configuration can be expanded to 18 storage servers in one rack for a total of 580TB of usable capacity. Assuming general parameters for database change rates, Oracle reports this can deliver the equivalent of 5.8 PB using the virtual full backup capability. Ingest rates for one full rack are nominally 12TB/hr, but as that would be handling minimized changed block data using Delta Push, it would effectively equal 120TB/hr of actual full backups, assuming a 10% daily change rate. The 12TB/hr per full rack performance is for backup and/or restore so the Recovery Appliance delivers similar backup and restore performance. This is very different than traditional backup appliances which often claim fast backup but never discuss restore performance because it is only a fraction of their backup speeds.

The Recovery Appliance can then be expanded further to 18 IB-connected racks, each rack starting with two compute and three storage servers. A full deployment would contain 10PB of usable storage capacity storing more than 100PB of virtual full backups. The full ingest rate of an 18 rack solution could equal 2PB/hr of virtual full backup protection and 216TB/hour restore.

For offsite protection, each Recovery Appliance can be set up to replicate to other remote Recovery Appliances, optimally still only transmitting changed blocks to minimize WAN requirements. Full recovery images can also be produced directly by remote appliances if necessary (e.g. the primary site became unavailable).

The benefit of transmitting only changed blocks should not be underestimated. For example, as a new customer (Brazilian Justice Tribunal of Santa Catarina) recently pointed out at Oracle OpenWorld, it was taking them up to 90 hours to complete full backups with its EMC Data Domain appliances and only 20 minutes to send just changed blocks to the Oracle Recovery Appliance each day. The customer purchased two Recovery Appliances and highlighted that acquiring this new solution was still less expensive than expanding on its existing solution to meet their total backup requirement.

## *Unified Management*

Oracle has also addressed the challenges of ensuring protection as experienced in the past when multiple tools with multiple consoles and user interfaces used by multiple users (e.g. DBAs, storage admins) were required to be involved in the company's data protection environment.

The Recovery Appliance is managed by Oracle Enterprise Manager Cloud Control, which provides a complete end-to-end perspective and control point from the time backups are initiated at the database (by RMAN) all the way through the rest of the Recovery Appliance recovery lifecycle. There is an easy wizard for adding a new database to the Recovery Appliance and specifying its required protection

policy. These policies then control which databases get which guaranteed recovery windows (how many days back should be stored online).

The system will dynamically prune or archive backups as necessary to best optimize the Recovery Appliance's available capacity to meet user-defined recovery window goals for all databases. Overall, the Recovery Appliance intelligently and automatically manages storage space at the level of an SLA type policy (for the stated recovery window for a database) rather than forcing lower level management based on managing opaque storage volume sizes. Enterprise Manager also enables easy capacity planning, with storage space predictions broken down by database (based on historical usage and stated goals). This SLA policy enables companies to deliver what might be called Data Protection-as-a-Service (DPaaS) across their enterprise or as external service providers. Customers are looking to the Recovery Appliance to bring about a significant reduction in infrastructure costs across storage, network, database and server resources, increase database performance, enable more focus on revenue generation and provide greater application availability. And actually useful recovery services to the sub-second RPO. We believe this will set a new standard for data protection services that CIO's will soon not only greatly desire, but absolutely require.

Additionally, the Recovery Appliance can be configured to work (over 16GB FC) with and smartly tier to tape libraries or the Oracle Public Cloud on the backend. The Recovery Appliance comes with Oracle Secure Backup software pre-installed, and will autonomously manage tape and cloud archiving and recall operations for Oracle tape solutions (other tape managers can be used). This offloads archival tape operations from production and can leverage the tape or cloud on a 24-hour basis (rather than only in batch windows) to get more ROI out of the desired architectural mix. With a public cloud that it owns and operates, Oracle is in a completely different position than its principal competitors in the backup space. As we've noted previously, HP has shut down its Helion public cloud. EMC and Dell don't currently have an analog to AWS—and the two companies will necessarily need to focus on internal reorganization for some time to come as they merge. Neither Hitachi nor NetApp has a cloud platform (nor Symantec/Veritas or CommVault). By offering end customer services on its own public cloud, Oracle is able to provide end-to-end visibility from on-premises infrastructure deployments to public cloud targets. If customers move data from their data center to the Oracle Public Cloud, Oracle can offer diagnostics on both ends. The customer doesn't have to deal separately with a chain of vendors – a database provider, a backup device provider and a public cloud provider – and then try to make sense of "where the data went" issues. The burden to manage all of these different offerings shifts from the customer to Oracle Support.

The fact that the Oracle Public Cloud, the Oracle Database and the Oracle Recovery Appliance were designed by the same organization within Oracle means tight co-engineering and commonalities across all three. Who better than the group that actually designed the Oracle Database to develop a way to backup and recover it?

## TANEJA GROUP OPINION

Protecting multiple business-critical databases, at scale and as data grows is a key and fundamental IT mandate. Just considering the data loss risks that many companies are otherwise facing, Oracle's new Recovery Appliance should be an easy solution for any enterprise serious about protecting its Oracle databases to adopt.

Databases aren't just a bunch of files although traditional data backup solutions treat them as such. This new fully automated recovery appliance purpose-built for Oracle Database demonstrates that there are miles of differences between how files get backed up versus how databases can best be protected.

The Recovery Appliance is almost too good to be true in that it makes fully protecting any Oracle Database or thousands of Oracle databases dead simple. You also gain an ability to fully recover any

database at will and along the way lower the impact on production database resources. And the Recovery Appliance is easy to deploy, the policy-driven unified management is an operator's dream, and the engineering is proven and world-class.

The policy-driven Recovery Appliance is so scalable and automated that it could readily be used to build and offer DB-Recovery-as-a-Service or Data-Protection-as-a-Service capabilities, even if just to internal lines of business. It truly offers cloud-scale and cloud service level functionality that will help transform IT's ability to offer distinct and reliable cloud-like services to its clients. And the fact that Oracle will offer their own public cloud services with the Recovery Appliance means direct customers are investing in a future-proof, cloud-ready architecture. Oracle has a well thought out lifecycle strategy. This new Recovery Appliance is not just an isolated backup device in search of a cloud story.

Bottomline – If finally eliminating the pain of outdated full backup approaches or the increasing risk of data loss at scale doesn't move you, the many benefits in operational ease and manageability shown in this report are significant enough to easily justify the Recovery Appliance's acquisition and deployment by Oracle Database owners. There are simply no more excuses for not fully protecting your critical Oracle Database assets as they should be protected.