TOP PERFORMANCE ON MIXED WORKLOADS, UNBEATABLE FOR ORACLE DATABASES
Oracle ZFS Storage Appliance – ZS3 Series

SEPTEMBER 2013

THE NET NET

There is a storm brewing in IT today that will upset the core ways of doing business with standard data processing platforms. This storm is being fueled by inexorable data growth, competitive pressures to extract maximum value and insight from data, and the inescapable drive to lower costs through unification, convergence, and optimization. The storage market in particular is ripe for disruption. Surprisingly, that storage disruption may just come from a current titan only seen by many as primarily an application/database vendor —Oracle.

When Oracle bought Sun in 2009, one of the areas of expertise brought over was in ZFS, a "next generation" file system. While Oracle clearly intended to compete in the enterprise storage market, some in the industry thought that the acquisition would essentially fold any key IP into narrow solutions that would only effectively support Oracle enterprise workloads. And in fact, Oracle ZFS Storage Appliances have been successfully and stealthily moving into more and more data centers as the DBA-selected best option for "database" and "database backup" specific storage.

But the truth is that Oracle has continued aggressive development on all fronts, and its ZFS Storage Appliance is now extremely competitive as scalable enterprise storage, posting impressive benchmarks topping other comparative solutions. What happens when support for mixed workloads is also highly competitive? The latest version of Oracle ZFS Storage Appliances, the new ZS3 models, become a major contender as a unified, enterprise featured, and affordable storage platform for today's data center, and are positioned to bring Oracle into enterprise storage architectures on a much broader basis going forward.

In this report we will take a look at the new ZS3 Series and examine how it delivers both on its "application engineered" premise and its broader capabilities for unified storage use cases and workloads of all types. We'll briefly examine the new systems and their enterprise storage features, especially how they achieve high performance across multiple use cases. We'll also explore some of the key features engineered into the appliance that provide unmatched support for Oracle Database capabilities like Automatic Data Optimization (ADO) with Hybrid Columnar Compression (HCC) which provides heat map driven storage tiering. We'll also review some of the key benchmark results and provide an indication of the TCO factors driving its market leading price/performance.

RIPE FOR DISRUPTION

The potential market disruption stems from the way Oracle has gone about the ZFS Storage Appliance evolution. On one hand, they've continued evolving the appliance into a full-blown enterprise solution, able to compete and win against incumbent storage vendors on price and performance. But on the other, they've also elevated it to power something they call "application engineered" storage. In plain language, this is storage that when coupled with the likewise "engineered" version of the Oracle Database, greatly accelerates database applications through intentional storage-side design features that are both aware of the database and directly leveraged by the database. This in theory is

similar in capability but very different in practice from convergence approaches that might customize hardware and tightly integrate a specific application or database which leads to expensive, dedicated, closed and quickly obsoleted infrastructure. In the ZS3 series we have the best of both worlds – cost-effective enterprise storage services, and application optimized storage infrastructure.

Broad Application

The ZS3 has several deeply engineered facilities for accelerating and optimizing Oracle Database performance. We will examine several of these later in the report, but together they provide tuning, troubleshooting, and performance optimizations that lead to application owners, DBAs, and the lines of business readily choosing the ZS3 as a storage platform target. Since it's from a large trusted vendor and has great support and management, IT has little cause in these business-driven designs to interfere. The interesting development is that once a ZFS appliance is up and running, IT finds it can readily support and accelerate many other business-critical workloads.

The ZFS appliance presents unified storage – both block and file – over perhaps one of the widest multiprotocol selections available including 1/10Gb Ethernet, Fibre Channel, and 40Gb InfiniBand. Because it has built-in dedupe, compression, thin provisioning, and other enterprise features, and can naturally deliver high transactional style IO performance, it's easy to imagine how other business critical workloads "slide" over onto the ZFS appliance. What IT folks discover is that the ZS3 is a rock-solid mid-range to enterprise storage, particularly for NAS use cases, that scales performance at a lower footprint and price than traditional dual-controller arrays and popular scale out storage solutions.

In particular, we highlight the available option of InfiniBand as a front-end interface. There are several places in the "next gen" datacenter where InfiniBand networking provides significant advantage – backing up Oracle Exadata to be sure, but also to support massively dense VM hypervisor clusters, large-scale VDI implementations, and HPC-like computing solutions. Reaching from 1GbE to 40Gb InfiniBand, the ZS3 can become a valuable enterprise-wide storage platform supporting all kinds of mixed workloads, from business-critical databases to home directories to highly virtualized environments with massive VM farms and content serving at scale.

ZS3 SERIES

The new ZS3 series has two basic models, an entry-level design, and the enterprise performance version, each of which is available in single or HA dual controller node configurations:

- **ZS3-4** – The ZS3-4 supports 80 cores, 25TB cache, 3.5PB of raw uncompressed capacity per cluster
- **ZS3-2** – The ZS3-2 supports 32 cores, 15TB cache, 768TB of raw uncompressed capacity per cluster

The server platforms here are not small cheap Linux or Windows boxes, but are some of the world's fastest computers—including up to 80 high-end Intel multi-core processors—running Solaris, a highly threaded symmetric multi-processing operating system that performs extremely well in VM environments, particularly since VMware is an SMP workload. This operating system enables the ZS3 Series to process hundreds of thousands of threads concurrently in contrast to traditional NAS filers which can become saturated at only a few hundred VMs each causing filer sprawl.

Performance

The first thing we note about the ZS3 is its blazing fast performance on critical, high-write, mixed workloads in addition to the engineered performance for Oracle databases. The ZS3 derives its performance numbers from three main sources:

- **Multi-level Caching** – The appliance can host a massive amount of DRAM, which unarguably makes the fastest cache. With its intelligent caching algorithms, the ZS3 can serve 70-90% of IO requests from DRAM depending on workload (and we note that DRAM is still much faster than flash cache). But then

there is also a large amount of MLC flash configurable for read cache in the controllers, and SLC flash for write cache in the drive bays. Intelligent caching pre-fetches pages, drains DRAM writes asynchronously, and recognizes large incoming synchronous writes to send directly to write flash or large streaming writes to send directly to the disk drive pool.

- **Intelligent Tiering** – Oracle's Hybrid Storage Pool (HSP) architecture, which holistically includes the above caching scheme, provides for the dynamic migration of data between DRAM, flash, and across multiple classes of SAS disk. In HSP, flash isn't just something tacked on the side, but is treated as an integral part of the fundamental design. HSP's Adaptive Replacement Caching has been improved with in-memory de-duplication and parallel access sequencing to further leverage relatively expensive cache resources, providing an improved performance of 2x over the previous generation.

- **Optimized Compression** – The ZS3 Series also provides customers with the ability to combine the benefits of Oracle Database 12c Advanced Data Optimization (ADO) with Hybrid Columnar Compression (HCC), to automatically determine what data sets in Oracle Database 12c should be compressed for query-only or deep archival and placed on capacity-optimized storage and which should be left uncompressed for frequent data access from performance-optimized storage pools. Hybrid Columnar Compression (HCC), available only with Oracle storage, compresses Oracle Database data from 10-50X and is specifically designed to work intimately with RMAN, unlike third party dedupe solutions for which the RMAN backup block is opaque, resulting in low levels of deduplication.

ZFS For the Enterprise

Obviously a big strength of the ZFS Storage Appliance is the ZFS file system. ZFS was designed as an incredibly scalable file system with massive potential capacity (128-bit addressing), and with matching built-in data protection features to ensure large-scale storage. Based on taking dynamic advantage of storage pooling rather than pre-allocations of volumes, ZFS has no need for external volume management, eliminating much storage management overhead. ZFS was also designed from the ground up with a distinct focus on data integrity, for example maintaining three copies of metadata (compared to just one copy in NetApp WAFL). In addition, ZFS validates the entire I/O path and can rapidly recover from bit rot, ghost writes, misdirected reads and writes, alpha particle hits, memory cell degradation, DMA parity errors, accidental overwrites and flash failures.

With all the built-in capabilities ZFS is designed to scale across hardware, utilizing what is given to it—scaling from small memory compute footprints to large—and

Hybrid Columnar Compression

Oracle's Hybrid Columnar Compression (HCC) provides a great example of "application engineering" across product lines. Fundamentally, HCC compresses sets of database records into a "compression unit" that is organized in a columnar fashion and then compressed. By grouping related records, compression is higher. By organizing the data in columnar organization, analytical queries are faster.

The unique value here is that HCC data is kept in compressed form throughout its subsequent lifecycle – into and out of storage, through downstream query processing, even into RMAN. Capacity savings are preserved and the data doesn't need rehydrating (unless being modified). Savings in space, processing time and performance accrue at every step.

HCC supports several levels of compression useful under Automatic Data Optimization processing. Based on the heat map analysis which dynamically categorizes data based on recent access patterns (that might correspond to data age), HCC can be applied at levels from low "query" to "high" archive. Data in query compressed format, while intermediate in size compared to the archive format, provides better performance for analytical queries, like in a data warehouse scenario, than even the original uncompressed data records.

This might seem paradoxical, but compressed data requires fewer IOs to get into and out of storage, less network bandwidth, and requires less memory. The columnar organization also speeds things up. You get both capacity savings and increased performance!

the ZS3 Storage Appliance provides plenty of horsepower. In this respect Oracle has been ahead of other vendors for years, some major players in the market are just now waking up to the idea of taking advantage of relatively cheap and constantly improving CPU power to accelerate their solutions (and many of those are caught on firmware, customized O/S, and other pace slowing architectures).

Multi-Protocol

One of the great things we note immediately useful to an enterprise IT shop is the wide integration that the ZS3 brings to a data center. It features broad multiprotocol networking support. At the physical networking level you can use 1GbE, 10GbE, Fibre Channel, and InfiniBand.

The complete range of networking protocols are supported with ZS3. There is nothing "bolt-on" in this technology—everything is integrated. Block storage access includes FC, iSCSI, iSER and SRP. NFSv3/4, CIFS, WebDAV, HTTP, S/FTP are all offered for file storage. It's hard to imagine not being able to leverage this unified ZS3 in a data center storage replacement or refresh project.

We especially note the availability of InfiniBand on the front end to support large clients like an Exadata cluster or other HPC-like computing platforms. And iSER, SRP, IP, and RDMA over InfiniBand are all in the mix. The ZS3 can provide that supercomputer-like storage if and when you need it! While the ZS3 might be seen as a great way to offload storage workloads or take backups from more expensive assets, we think there is an opportunity for it to be featured front and center on a broad range of emerging big data solutions, especially in the way it provides for a large number of workflow and data sharing requirements coming from different platforms and environments.

Oracle Acceleration

In addition to overall enterprise capabilities, Oracle has taken the opportunity to engineer the ZS3 with specific features to accelerate Oracle databases. Their "application engineered storage" approach clearly doesn't limit the Storage Appliance solely to the database workload, but instead empowers it to handle both databases and other highly demanding business critical applications at the same time.

There is plenty of proprietary IP here. Other storage vendors will have a difficult time competing head-to-head with the ZS3 on the Oracle Database workload. And enterprises are bringing in the ZFS Storage Appliances, based on their database capabilities, which show no sign of letting the competition catch up. These features include:

- **Hybrid Columnar Compression (HCC)** – On top of database-managed, row-level compression, the system-wide HCC facility organizes data into "compression units" that add column orientations over sets of rows for enhanced results of both compression ratios and query performance. HCC provides four advancing levels of "columnar" compression intended for different levels of data usage (low/high query, low/high archive). The key here is that data stays in HCC compressed format during all stages of transfer, storage, networking, query processing, and backup (with RMAN), optimizing every downstream phase of operation.
- **Automatic Data Optimization (ADO)** – ADO tracks database data as it ages and becomes less referenced transactionally and/or increasingly static (see Heat Maps illustration). As data "cools", ADO applies the various HCC compression levels to balance current access needs with optimal space savings. Not only does this result in 5-50x reduced storage capacity for any given chunk of data, but data subsequently referenced analytically can actually increase in performance under this scheme through both less disk IO and columnar organization.
- **Oracle Intelligent Storage Protocol (OISP)** – OISP enables the Oracle Database to directly configure allocated storage in the appliance. By automating the steps of setting parameters for low level database/storage alignment (e.g. record sizes, logbias, etc.), manual database tuning time is more than

halved and misconfiguration potential is completely eliminated. Think of this as a speed dial offered from the Oracle Database directly to the ZS3.

Overall, these capabilities on top of the ZS3 enterprise storage features lead to best-in-class Oracle Database performance with optimal capacity usage less silo-oriented resources are needed for operations and fewer mistakes can be made—all adding up to unmatched database hosting efficiency. The ZS3 is essentially designed to help customers accelerate time to insight, mitigate risk with highly granular data center analytics and run database queries faster. With ZS3, customers can expect better answers, get more data in less time, all faster than with comparative midrange systems.
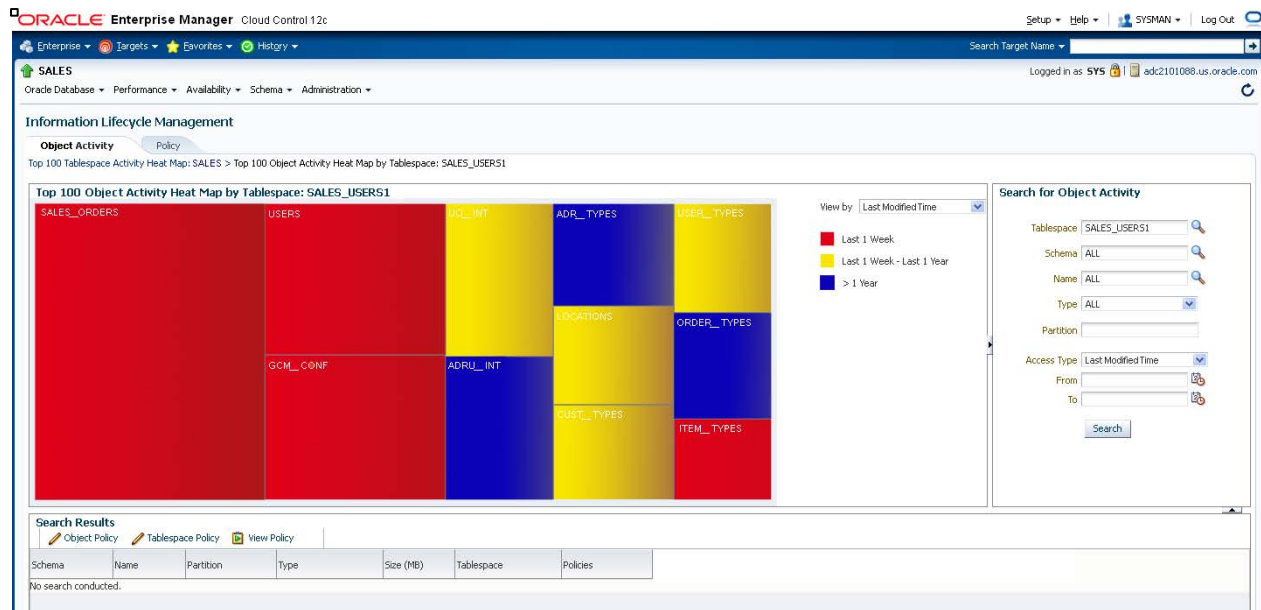
Figure 1 - Oracle Heat Map Tiering

Engineered for Management

Note that these application (i.e. Oracle Database) specific capabilities result from uniquely being able to engineer across both the application and the storage infrastructure. This engineering is intentionally designed from both (all) sides to work intimately in concert, but is loosely coupled in operation—other Oracle solutions like Exadata, SuperCluster and Axiom also support HCC. We think this is the best kind of convergence, and in important ways, represents what other forward-thinking solution vendors will strive to achieve in the future with their Software Defined Storage (SDS) plans—once they are in a position to "also" engineer the application side to take advantage of the storage "control plane" that will be exposed in their SDS visions (many still in PowerPoint form).

When designed together as a complete system, a solution can provide much faster troubleshooting and direct application-level storage visibility. The ZS3 comes with the DTrace Analytics for integrated visual-based problem solving across multiple layers of storage and application components. In more siloed systems, storage misalignment, contention, and errors can disrupt business-critical applications, sometimes degrading their performance over long periods of time without obvious connection.
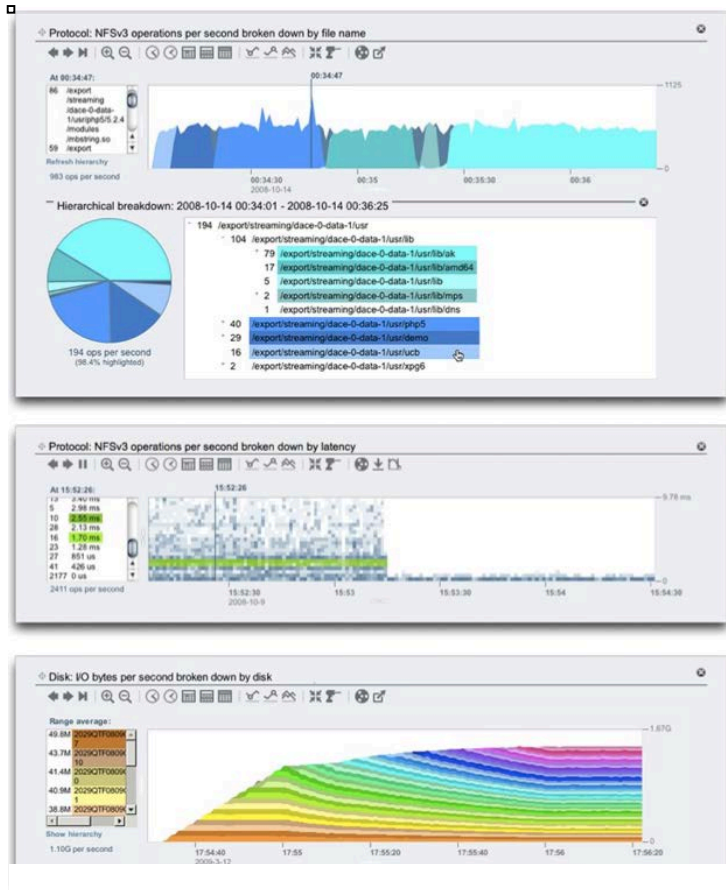
Figure 2 - DTrace Analytics

TOP PERFORMANCE TODAY, BRIGHTER FUTURE AHEAD

The ZS3 is designed for highly virtualized data centers and for use in both private and public clouds, for direct attach to high-speed, high-throughput compute clusters. Any environment that is write-intensive with demanding requirements for concurrent throughput, latency and IOPs is an ideal fit for its DRAM-fueled performance. The ZS3 is also designed to provide peak performance to Oracle Database OLTP, data warehousing, development and test environments, and as data protection for Oracle's Engineered Systems lineup including Exadata and Supercluster T5-8.

Unlike many other storage specific vendors, Oracle appreciates that infrastructure is only productive when used by client applications and is willing to publicly prove it. To that end, Oracle has carefully engineered for its key applications and has a deliberate program to test out its beta solutions at large scale in its own massive production operations first (Oracle Cloud & Oracle IT) before going to customers. And to prove application performance, Oracle has not shied away from publishing public benchmark results, and on multiple benchmarks at the same time. It doesn't hurt that they are able to top the charts on SPC-2and SPECsfs vs. competitors such as NetApp, IBM and HP.

The broad application suite performance results are impressive, but Oracle Database performance is even more so. We believe that with the architecture and potential of the ZS3, Oracle has, and will continue to maintain dominance on providing the fastest queries (GB/s) with the fastest response times (lowest latencies) for Oracle Database (with already world class software-engineered database performance).

The underlying performance and capacity optimizations also lead to great savings on backup/restore processes—space, time, utility and % success—and lead to smaller footprint, power and cooling requirements. When combining the operational numbers with a financial analysis, even a rough TCO estimate can show that at expected scale accounting for all the efficiencies, the ZS3 delivers outstanding $/IOPS and $/GB/s. For Oracle applications the TCO is undeniable, but IT shops should also compare the ZS3 with other enterprise storage under consideration for a broader range of file and even block use cases.

For example, we've been told a ZS3 can readily support over 2000 virtual machines with only 33% CPU utilization. The VM IO blender effect looks a lot like high-write small-block random IO similar to multiple high volume OLTP database application loads, so the ZS3 architecture is well aligned as virtual environment storage.

Looking Ahead

The ZS3 initially faces off in the hotly contested midrange to entry-level enterprise storage market against systems from NetApp, which are aging and in need of a refresh, IBM's XIV and V7000 arrays, and the recently refurbished EMC VNX-2. The ZS3 will offer stiff competition on performance, price and TCO, but significantly, the competition all lack specific Oracle Database integration capabilities giving the ZS3 an open door to almost every account.

Without compromising any internal plans, it's obvious that the Oracle ZS3 line has a bright future. On the hardware side, as compute and memory increases continue to improve (i.e. Moore's law) further bending the computing power curve, storage systems like the ZS3 that are intentionally designed to take advantage of increasingly cost-efficient compute, memory and flash will improve as well. And ZFS itself has a rather robust roadmap.

One clue to future capabilities comes from considering what OISP could evolve into. OISP's current incarnation supporting application direct storage tuning is clearly just a beginning to what might be possible when storage arrays are engineered to provide application-oriented services. This protocol will no doubt expand over time to provide more powerful application-storage integration features such as application hinting.

TANEJA GROUP OPINION

With Oracle storage having already passed the billion dollar mark and with over 10,000 deployed ZFS Storage Appliances, Oracle storage solutions are gathering impressive steam versus the competition. Adoption is already significant as an optimizing storage solution for Oracle Database environments but next we predict that IT on a broader scale will start to take notice that ZS3 storage is a great value for a wider class of enterprise needs. As its broader usefulness becomes known, we expect Oracle to show up on the short list of an increasing number of enterprise storage infrastructures in general.

It's worth pointing out again that Oracle shakes out its ZFS Storage Appliances in their own production cloud services and IT services at large scale—over 120,000 employees—before any customers are subjected to beta testing. While Oracle may have some work to do on their infrastructure support and service "impression", we trust that a company of this size and resources, dedicated to bullet-proofing critical operations, is aiming to further elevate its world-class reputation.

Bottom-line—the ZS3 represents a great value for customers looking for both performance and capacity to support highly demanding applications at the heart of the data center. It's an obvious choice for Oracle Database scenarios with demonstrated fastest database queries. This will inevitably lead to market share gains, particularly in Oracle Database environments coping with traditional filer sprawl. But also with a low $/IOP, unified and multiprotocol support, and the ability to replace multiple aging storage infrastructures with a forward looking platform, we expect to see the ZS3 competing head-to-head with the top storage platforms to support other mixed workloads. Given Oracle's enterprise-level relationships combined with the fact that the ZS3 series can be configured with a wide variety of adapters, disk types and capacities, including a backup appliance version, we expect to see more enterprises standardizing on Oracle for their critical storage infrastructures.