There is a serious re-hosting effort going on in data center storage as flash-filled systems replace large arrays of older spinning disks for tier 1 apps. Naturally as costs drop and the performance advantages of flash-accelerated IO services become irresistible, they also begin pulling in a widening circle of applications with varying QoS needs. Yet this extension leads to a wasteful tug-of-war between high-end flash only systems that can’t effectively serve a wide variety of application workloads and so-called hybrid solutions originally architected for HDDs that are often challenged to provide the highest performance required by those tier 1 applications.

Someday in its purest form all-flash storage theoretically could drop in price enough to outright replace all other storage tiers even at the largest capacities, although that is certainly not true today. Here at Taneja Group we think storage tiering will always offer a better way to deliver varying levels of QoS by balancing the latest in performance advances appropriately with the most efficient capacities. In any case, the best enterprise storage solutions today need to offer a range of storage tiers, often even when catering to a single application’s varying storage needs.

There are many entrants in the flash storage market, with the big vendors now rolling out enterprise solutions upgraded for flash. Unfortunately many of these systems are shallow retreads of older architectures, perhaps souped-up a bit to better handle some hybrid flash acceleration but not able to take full advantage of it. Or they are new dedicated flash-only point products with big price tags, immature or minimal data services, and limited ability to scale out or serve a wider set of data center QoS needs.

Oracle saw an opportunity for a new type of cost-effective flash-speed storage system that could meet the varied QoS needs of multiple enterprise data center applications – in other words, to take flash storage into the mainstream of the data center. Oracle decided they had enough storage chops (from Exadata, ZFS, Pillar, Sun, etc.) to design and build a “flash-first” enterprise system intended to take full advantage of flash as a performance tier, but also incorporate other storage tiers naturally including slower “capacity” flash, performance HDD, and capacity HDD. Tiering by itself isn’t a new thing – all the hybrid solutions do it and there are other vendor solutions that were designed for tiering – but Oracle built the FS1 Flash Storage System from the fast “flash” tier down, not by adding flash to a slower or existing HDD-based architecture working “upwards.” This required designing intelligent automated management to take advantage of flash for performance while leveraging HDD to balance out cost. This new architecture has internal communication links dedicated to flash media with separate IO paths for HDDs, unlike traditional hybrids that might rely solely on their older, standard HDD-era architectures that can internally constrain high-performance flash access.

Oracle FS1 is a highly engineered SAN storage system with key capabilities that set it apart from other all-flash storage systems, including built in QoS management that incorporates business priorities, best-practices provisioning, and a storage alignment capability that is application aware – for Oracle Database naturally, but that can also address a growing body of other key enterprise applications (such as Oracle JD Edwards, PeopleSoft, Siebel, MS Exchange/SQL Server, and SAP) – and a “service
provider” capability to carve out multi-tenant virtual storage “domains” while online that are actually enforced at the hardware partitioning level for top data security isolation.

In this report, we’ll dive in and examine some of the great new capabilities of the Oracle FS1. We’ll look at what really sets it apart from the competition in terms of its QoS, auto-tiering, co-engineering with Oracle Database and applications, delivered performance, capacity scaling and optimization, enterprise availability, and OPEX reducing features, all at a competitive price point that will challenge the rest of the increasingly flash-centric market.

RIPE FOR DISRUPTION

Oracle has a long history of designing enterprise-class storage and is actually a top 10 storage supplier to data centers worldwide, although for many IT storage folks Oracle may not be top of mind. With the heritage of Sun storage (as evidenced by the equally impressive Oracle ZFS NAS storage appliance), Pillar Axiom, and purpose-built appliances for Oracle Database and business applications, it’s almost inevitable that Oracle would combine best of breed thinking from all that inspirational ancestry to design a high performance, low cost, flash SAN storage system for the modern data center. While it inherits a great deal of storage experience, the FS1 is a really a new storage beast with a flash-first design point and over 5M lines of code built on the latest 64-bit Intel multi-core chips.

From our perspective, it seems Oracle intends to solve at least three storage industry challenges with the FS1 design:

1. Make effective use of flash to maximize application performance
2. Hold the line on cost and support effective capacity scaling with smart, QoS auto-tiering across multiple tiers of flash and disk media
3. Improve management at scale with virtual storage “sub-array” domains, storage-side application awareness, and the default inclusion of HA/enterprise needed features and capabilities

If faster, cheaper, and better weren’t enough, Oracle uniquely incorporates one more key capability - deep co-engineering with Oracle Database and applications. The FS1 is one of Oracle’s new “engineered systems” – all built with the objective of combining Oracle hardware and software IP into practical datacenter solutions. You don’t need Oracle databases to love the FS1, but if you have any in your datacenter, then the FS1 provides even greater value. (See sidebar on HCC.)

The FS1 has an all-flash configuration option with two flash tiers (performance flash and capacity flash), so it is an all-flash solution in that respect. This version starts with dual redundant controller nodes and up to 900TB flash. But likely folks will want to deploy it with more tiers of storage that include HDD shelves to be able to use it for a wider variety of QoS needs such archiving colder data, performing backups, and providing storage for less performance sensitive applications. The resulting mix of storage can be carved up as desired into different storage domains or assigned as resources application by application.

Fine-Grained Auto-tiering

Generally we’ve come to expect that top-end flash performance comes from systems purpose-built for flash from the ground up. And this is what we’ve seen from the majority of solutions aiming at the enterprise top tier storage market; all-flash designs easily outperforming older generation systems that just substitute in SSDs for HDDs. Likewise, so-called hybrids are in general designed to use flash to accelerate all workloads in a general purpose manner while requiring HDD storage for capacity, helping the low to mid-market companies get more mileage out of smaller flash budgets.
The FS1 is available as an all-flash version (with a flash-first design), but is also highly engineered to use up to four tiers of storage to optimize performance and cost:

1. **Performance Flash** – Today these are 400GB high performance flash drives.

2. **Capacity Flash** – Higher capacity 1.6TB flash drives (with more limited write cycles used more as “read” flash)

3. **Performance Disk** – These could be high performance 300GB or 900GB 10K SAS drives.

4. **Capacity Disk** – Slower but larger 4TB disks

There are a couple of very interesting aspects to the FS1 tiering implementation. First, the chunk size for data movement in tiering is finer grained than in most other tiering approaches. Oracle has determined with their unique engineering perspective that the sweet spot between Oracle databases, storage, and today’s CPU-rich platforms is 640K. Other solutions might be paging with a 256MB chunk (e.g. HP 3PAR StoreServ, EMC VNX2) or larger, which according to Oracle is like using a dump truck to optimize shovelfuls of data. And, taking aim at EMC XtremIO, Oracle claims that a 4k chunk size is too small and will cause wasteful overhead and likely write amplification on the flash layer. Of course, proof is in performance. As we’ll see later the FS1 in this respect is nailing the competition.

The FS1 QoS Plus/auto-tiering facility tracks data skew using built-in heat mapping based on access frequency combined with a “usage profile” of the chunk of data being accessed (random/sequential, read or write) and the data's defined business priority. Typically users set a tiering evaluation scan interval to every 24 hours, but with the FS1 this can be configured down to minutes, or if desired much longer - weeks, for example. The FS1 uses a sticky algorithm to avoid thrashing, and perhaps most differentiating, understands and tiers preferentially based on designated application quality of service (QoS). Thus it's possible for different workloads to share storage, but avoid having lower priority apps hog the flash tier just because they roll through a lot of data in the course of a day. Fundamentally, the FS1 tiers on both the usage profile and the business value of the data to ensure performance and optimize cost.

Normally this tiering is applied across all four supported storage tiers, but if, for example, there are only two tiers present or configured (i.e. in the all-flash FS1), the FS1 still pretends to tier to four internally. Why? So it can then report on “what if” you had more tiers to help advise you with optimizing future capacity planning investment decisions – where the best opportunities exist to extend the storage mix at the lower cost capacity tiers (or not), and if there is any contention, what is competing for the performance tier of storage.

**Large Scale Flash**

Physically the FS1 starts with two active-active HA nodes (controllers) and will be able to grow to 16 controller nodes. Each controller is configured with a 4-processor (24 cores) latest gen Intel E5 chip. These controllers have a large memory cache (either 64GB or 384GB) which is fully protected with super-caps and flash-backed DIMMs. As enterprise storage, the FS1 supports 16GbFC and 10GbE.

The FS1’s all-flash version is spec’d to grow from 912TB of flash in a 2 controller configuration to a massive 7 petabytes of flash and 6.6TB of SDRAM in a scaled-out 16 node version. The initial FS1 model (FS1-2) is designed to mix flash and disk and host up to 30 drive enclosures, growing to 2.9PB total capacity. Note the point is to tier storage. To us the FS1 architecture is an all-flash array design with the bonus of capacity tiers of storage built-in, rather than a hybrid array that might have design compromises penalizing top-end performance. For example, we've been told that if the FS1 were fully populated with HDDs, the controllers would use only 10% of the available CPU resources.

**QoS Plus, Plus, Plus**
Central to the FS1 design is a focus on ensuring QoS to application workloads. As introduced above, the whole “adaptive” auto-tiering facility is driven to deliver on QoS goals set for storage, balanced against utilizing the most cost-effective media to achieve those goals, driven by business value.

But smart tiering is just half of the FS1’s “QoS Plus” implementation. The other half is that it also prioritizes the IO queue. By prioritizing IO according to QoS, the highest priority IO (e.g. OLTP) will get served in the most performant timeframes, while lower priority IO (e.g. archive) will wait its turn. Of course, “wait” is relative and we’ll see that the FS1’s performance (along with QoS Plus) is such that mixed workloads that do not co-exist well on other enterprise storage can be easily hosted together here.

The FS1 is also unique in that it is the only flash storage system to provide IO prioritization at the application layer in real-time with adaptive technologies. The FS1 applies QoS at the compute and storage layers, separating and isolating apps from one another. This is far superior to QoS implementations that are based on a simple application “throttle back” approach that just constrains certain workloads once they’ve reached designated limits – this neither assures good performance under contention to high priority workloads nor makes efficient use of available resources.

Specific workload QoS can be set with one-click provisioning based on application profiles and can be dynamically changed. There are pre-defined application profiles to make this useful right out of the box starting with Oracle Database.

![QoS Plus: Priority Aligned Autonomous Tiering](image)

**Figure 1. QoS Plus in Oracle FS1**

**Scale, Capacity, and Management**

A key part of the FS1 storage system, and one that shows both the storage heritage Oracle has and their enterprise qualifications for service providers at the largest scale, is FS1 Storage Domains. With Storage Domains, an admin can dynamically subset out what are effectively multiple virtual arrays from within a single FS1. These Storage Domains are actually carved out of physical resources by allocating “drive groups” from one or more tiers of storage into a virtual storage system.

Interestingly, an FS1 Storage Domain consists of fully partitioned (i.e. not shared) media, within the shared management of the whole FS1. Service providers can then basically offer a complete “isolated”
array service to a client, guaranteeing that the storage media is not shared, while benefiting from the multi-tenancy through holistic management (and other OPEX factors) and the ability to dynamically alter each storage domain as needed (i.e. to grow, shrink, etc.).

Within each storage domain, QoS Plus settings can be customized for that domain and its specific mix of workloads. These domains can be carved out of storage from any tier in any combination (including single-tier, all-flash, or mixed), with QoS Plus settings tailored for that mix.

Storage Domains are useful not only for multi-tenant situations or to create unique QoS services, but also to isolate or protect specific sets of data for privacy, regulatory compliance, or other needs.

**Performance and Price**

Clearly the Oracle FS1 has a great performance story. Oracle claims the two node FS1-2 all flash (on FC) not only can hold 173TB of performance flash (greatly out-scaling the likes of EMC XtremIO), but delivers 265k IOPs (at 50/50 r/w mix), with tremendous throughputs of 5.4 GB/s write, 10.8 GB/s read.

The FS1 architecture can grow to 16 HA controller nodes, interconnected via high-speed Infiniband. Oracle tests claim that the FS1 can provide over 2M IOPs (50/50 r/w mix) with up to 80GB/s throughput.

The price point of the FS1 appears attractive too, with Oracle including all in-the-box software in the base price, including QoS Plus and auto-tiering, Storage Domains, Application Profiles, thin provisioning, efficient snaps and clones (and the SnapDelta file system feature for efficient replication), multi-system management, and their “fast forwarding file system”.

The FS1 also works well for the virtualization world too, including support for VMware SRM, VAAI, VASA, and vSphere plugins.

**Application Engineered Storage**

In addition to overall enterprise capabilities, Oracle has taken the opportunity to co-engineer the FS1 with specific features that accelerate Oracle Database performance and reduce storage requirements. Their “application engineered storage” approach clearly doesn’t limit the FS1 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 on this front. Other storage vendors will have a difficult time competing head-to-head with the FS1 for the

**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 format, 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 and application development/testing, 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 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 to deliver both capacity savings and increased performance.
Oracle Database workload. Two of the unique values for Oracle databases include:

- **Hybrid Columnar Compression** – HCC helps compress and accelerate database data IO and analysis (see sidebar) across the data lifecycle including in and out of storage. HCC usage can reduce capacity on disk by factors ranging from 10x to 50x (obviously depending on data composition and usage patterns). As a multiplier on the flash investment, this can be pretty impressive.

- **Predefined storage Application Profiles** – Oracle FS1-2 Flash Storage System comes with predefined Application Profiles that provide tuned and tested out-of-the-box storage optimization for Oracle Database and key enterprise applications. With one-click provisioning customers can set up, optimize and deploy Oracle Database and applications with a minimum of administration.

For example, the FS1 works with Oracle Database to inherently recognize storage “components” of the database on the array side and treat them differently. For example, database tables, index files, redo logs, and temp files might all get different QoS parameters. Index files would be given storage and priorities appropriate for random/read, while redo logs will be aimed for sequential/write service. In an actively multi-tiered system, this is a huge optimization step that is automatically built-in. In addition to Oracle Database, Application Profiles are available for Oracle applications (E-Business Suite, JD Edwards, and PeopleSoft), as well as third-party applications including Microsoft Exchange/SQL Server and SAP.

**TANEJA GROUP OPINION**

The Oracle FS1 is posting some seriously competitive performance numbers (which EMC and other competitors seem to avoid discussing). While the FS1 array is new, it is built on five generations of Oracle flash hardware and software innovation (e.g. Exadata and ZFS Storage Appliance), plus it inherits Pillar’s battle hardened comprehensive set of data services. The result is a demonstrably faster, arguably much less expensive flash storage system that offers better capabilities in a number of key areas of enterprise concern: all-flash performance, inherent support of mixed workload environments, secure multi-tenancy, and Oracle and other application awareness.

Perhaps the key metric for IT buyers today is price. Oracle tells us it is currently half the price of comparable EMC XtremIO offerings and that isn’t even accounting for the complete set of software Oracle offers today with the FS1. We expect that with FS1 Storage Domains for secure multi-tenancy, impressive total IOPs and huge scalability with self-optimizing auto-tiering, cloud and managed service providers will see the FS1 as an opportunity to deploy enterprise class flash-tuned storage for their clients.

Oracle today is number 3 in NAS worldwide, but ranked number 7 for SAN. We expect that the FS1 will help Oracle climb the SAN charts. In their favor, massive and critical Oracle databases lurk at the heart of many datacenters, giving Oracle storage a direct path to engagement (unlike small strip mall storage startups).

Bottom-line—the Oracle FS1 represents a great value for customers looking for both performance and capacity to support highly demanding business-critical applications. It’s an obvious choice for Oracle Database environments where performance and/or consolidation are top concerns, inevitably leading to market share gains. But also with a competitive price, unified and multiprotocol support, and the ability to replace multiple aging storage systems with a forward-looking platform, we expect to see the FS1 competing head-to-head with the top storage platforms to support other datacenter workloads. Given Oracle’s enterprise-level relationships and impressive storage system portfolio, we expect to see more organizations standardizing on Oracle for their critical storage infrastructures.

**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 that may appear in this document.