# Enterprise NAS systems, scale-out NAS systems, and software defined NAS storage

Dedicated storage systems serving file protocols (aka NAS or Network Attached Storage) have been popular in the enterprise IT field for more than 20 years. Even though the names did not change much, the architecture inside and ecosystem surrounding them went through several rounds of evolution. New architectures are proposed to satisfy the needs of more versatile demands, such as file sharing to edges, single name space for a large capacity, higher performance, and stricter disaster recovery needs. In this section, we will investigate the pros and cons of a few architectures and hopefully that will help our readers in the decision-making process.

## NAS overview

When we look at the NAS systems available on the market, we can approximately separate them into three categories: enterprise NAS systems, scale-out NAS systems, and software defined NAS. The enterprise NAS systems inherited the years of engineering optimizations and thus are still the best choice for many use cases. They are commonly delivered as a set of file servers (aka NAS gateways or filers), and since they are designed to serve this focused task, the latency is often optimized. Even though the internals have been updated several times with modern hardware (Faster CPUs, Flash drives, etc.) and modern software architecture (log-structured, multi-core optimization, etc.), an enterprise NAS system still provides services to the traditional file interfaces (NFS, CIFS, etc.) as previous generations. The workload running on enterprise NAS systems are often referred to as transactional NAS, including home directories, databases, etc. Thanks to technology progress, today an all-flash system can reach a very high performance so that a single system can often serve the whole department, or even the whole enterprise. Its management style is often traditional and easy to start, while new enhancements are added to fit new management styles. For example, plug-ins are provided so that information can be collected to a cloud-based management and they can be monitored and managed via remote links. Through all the years, the enterprise NAS system market bracket is always the core piece of the NAS market, as customers found it versatile to satisfy different scenarios. OceanStor V6 NAS is one example of modern-day enterprise NAS platform.

In contrast, the scale-out NAS systems have a shorter history and they were invented to cope with the need for scalability. The scalability needs came from two sources: the need for large capacity (storage capacity, IO throughput, and file system size) and the need for flexibility. First, let us discuss the need for capacity. One example is the media industry. With the higher demand of movie special effects and high-definition TV, the size of file is increasing in a rapid pace. Therefore, some of the traditional optimizations towards smaller files are no longer valid. Many of the scale-out NAS systems are optimized towards throughput, total size of a file system, and multiple protocol support. As its name suggests, a scale-out NAS system can be scaled out quickly by adding new nodes. After adding new nodes, the system can support more IOs and can process a larger capacity, usually under the same namespace. Some examples of the workload running on scale-out NAS systems include 4K HD video, cloud scale computing, etc. Both Dorado V6 and 100D can handle these workloads. Depending on the capacity needs, OceanStor V6 can scale to 32 node and 100D can be scaled to 4096 nodes.

Finally, the modern-day IT environment is asking for agility and flexibility. The nodes used for enterprise NAS systems are custom built, but some customers would like to use their existing server farms. In the last 10 years, software defined storage (SDS) concept has been proposed to address this need. Users would like to build nodes with commercial off the shelf servers. Many vendors started to provide software defined, and sometimes converged and hyper-converged systems. The SDS solutions can be built on physical servers but not required. Many allowed customers to build storage services on virtual machines and even inside of public cloud vendors. Compared to traditional NAS solutions, the SDS ones are not aiming to excel in performance and capacity. Rather, they focus on flexibility and give users the ability to quickly bootstrap, teardown, and change the storage services in many environments. OceanStor 100D can be used in an SDS way but can also use custom-built nodes.

A good NAS system should cover at least two requirements well, being excellent in one and good in the other. At the same time, trying to cover the third requirements. For example, OceanStor V6 NAS is excellent serving as an enterprise NAS platform, at the same time good at scale-out to 32 nodes. For agility, it may not be as flexible as SDS but service programs are designed to increase agility (e.g., reduce the turnaround time for system expansion). On the other hand, OceanStor 100D is excellent in scalability as it can scale from 3 nodes all the way to 4096 nodes. OceanStor 100D also has a good support for agility by supporting SDS. But for the use cases where OceanStor V6 NAS is excellent (e.g., low latency and 99.9999% reliability scenarios), 100D is the second choice.

## Scale of data

Scale-out NAS systems addresses the explosive data and performance demands with cluster of nodes (usually general-purpose servers) on connected networks. A cluster can be scaled to thousands of nodes and massive capacity.

In a scale-out architecture, new nodes can be easily added and configured to support changing business requirements. Instead of overprovisioning from the start, users can simply add new capacity as the storage needs increase, which makes it a cost-effective storage option.

On the other hand, traditional enterprise NAS is known as scale-up NAS, because it is upgraded by adding performance and capacity (RAM, CPU, disks) to an existing architecture (several controllers).

OceanStor 100D NAS systems are typical scale-out systems. They support 4096 nodes per cluster, 256 nodes per filesystem, and 200 PB capacity per filesystem.

OceanStor Dorado enterprise NAS is both scale-up and scale-out. It can scale out up to 32 nodes in a system, allowing a flexible combination of smart disk enclosures and controller engines, providing up to 6400 disks for capacity and 20 million IOPS, without forcing customer to over-investment in computing power like many other competitors.

## Performance

OceanStor Dorado enterprise NAS system is built on top of Huawei’s flagship storage hardware to take full advantage of the bleeding edge technology and offer best in class performance. Dorado NAS support 100GE RDMA (RoCE) network for both backend and front end, enables industry leading throughput and latency. I/O latency can be as low as 0.1ms for block access on average under typical mixed read/write workload and can nearly saturate the 100GE connections.

OceanStor 100D NAS takes a quite different focus on scalability instead of extreme performance. While the software components share a lot of common designs, overall performance is limited by the hardware you choose. OceanStor 100D is a software defined storage, it supports a few different flavors of hardware, including but not limited to generic x86 server, generic ARM server, Huawei OceanStor Pacific node and Atlantic node. Pacific nodes are HDD based and optimized for highest capacity density and lower cost, while Atlantic nodes are SSD based and offer best performance density. For example, using Pacific hardware, OceanStor 100D NAS can achieve the following performance with a 5U 8 controller system: *aggregate read throughput: 144GBps, aggregate write throughput: 96GBps, aggregate IOPS on typical mix of small I/O: 1.6 million IOPS with average latency around 1.0ms*. Generic servers offer best flexibility; however, performance and cost could be worse than the other two Huawei specialty hardware.

One advantage of scale out system like OceanStor 100D is, when you add more nodes, total aggregate throughput can grow near linearly, this can be helpful when there are many hosts connected to the system, even if a single host does not benefit from the scale out, the total solution does.

## Availability

Due to the difference of their design goals, OceanStor Dorado and 100D have different availability goals. Both have high availability numbers: 99.99999% for Dorado and 99.9999% for 100D respectively. Dorado is designed for the highest availability needs such as core banking systems, whereas 100D is designed for other mission critical workloads.

## Features

The feature set provided by OceanStor Dorado NAS and OceanStor 100D NAS are very similar. They both provide a strong set of features, including snapshot, clone, async replication, tiering, quota, WORM, anti-virus, NDMP support, dedupe, QoS, etc. These features provide a strong base to support customers’ data service needs and lower the TCO. In addition, OceanStor Dorado and 100D has solutions to migration customers’ existing data from existing servers, including third party equipment.

These two platforms belong to the same product family and it is natural to have both share many common features. An effect is that data can flow natively from one platform to the other, so together they can form a bigger scale solution to optimize customer’s TCO and provide more choices to satisfy customer’s needs.

## **INDITEX Requirements and our recommendations**

INDITEX needs a high-performance, scale out, high-availability NAS system to support its core business needs. After evaluating capacity, availability, performance, scalability, and features, we recommend OceanStor V6, a mix of high-performance enterprise and scale-out NAS product, as the best platform to satisfy INDITEX’s needs. OceanStor V6 is designed for high performance, high availability, and enterprise-scale core business needs. In the following subsections, we examine each of the areas in details.

### Capacity

INDITEX has the current IO profile as follows:

* 12B files and 60% (7000M) are small files (1KB – 200KB) with a total capacity of over 1PB. With increasing business needs, the total capacity could be increased to 2PB in 5 years.

OceanStor Dorado V6 provides up to 32PB (6400 disks). OceanStor 100D provides up to 200PB per filesystem. Both systems satisfy INDITEX’s capacity requirements. Even though OceanStor V6 is not cloud scale, with up to 32 nodes and 32PB limit it can support INDITEX’s business needs many years down the road.

### Performance

* INDITEX requires 1ms access time.

OceanStor Dorado V6 provides the best of the class performance with well under 1ms latency guarantee and can reach 0.1ms latency with the high-performance settings. V6 also has a large battery backup DRAM write cache for extreme performance, which is much faster than NVDIMM used in OceanStor 100D. Compared to 100D, the architectural benefits of Dorado V6 are clear. Many of the optimizations such as RDMA high speed network are built-in in the V6 architecture, whereas for 100D users must make efforts to install and maintain such an environment themselves. In order to get closer to V6 performance numbers, 100D systems must be equipped with high-end nodes and be carefully tuned.

* INDITEX requires “performance and low latency for hot data and metadata.”

For both Dorado V6 NAS and 100D NAS, metadata are evenly distributed to all nodes, avoiding the bottleneck of metadata server. Local read cache with various prefetching algorithms is used for hot data. Compared to 100D, Dorado V6 uses DRAM to serve hot data so that latency is minimized.

* INDITEX has 60% small files (1KB – 200KB)

Both Dorado V6 NAS and 100D NAS have performance improvement for small files, including:

1. Small IO aggregation;
2. I/O workload model of intelligent service-learning implements cross-file pre-reading.
3. NVDIMM for write latency and file aggregation.
4. 8 KB block size for small files, improving read and write performance.

### Scalability

INDITEX requires “Scale-out. Equitable load distribution among nodes.”

For both Dorado V6 NAS and 100D NAS, metadata and data are evenly distributed to all nodes. Adding disks or nodes is easy. Auto rebalancing is fast and does not affect host IOs.

OceanStor Dorado V6 NAS supports scale-out up to 32 nodes. OceanStor 100D NAS supports scale-out up to 4096 nodes per cluster and 256 nodes per filesystem. Based on our estimation, a 4 node high end Dorado V6 system should be our initial recommendation so 32 nodes will meet the future needs.

### Availability

Dorado V6 systems have a 99.9999% availability guarantee because of the coordinated efforts of hardware and software because the fault domains are designed for high availability. 100D has a lower HA claim because of the flexible options and the complexity comes with the options.

### Access Protocols

INDITEX requires SMB (v1, v2, v3) and NFS (v3, v4).

Dorado V6 NAS and 100D NAS do not support SMBv1 and NFSv4 yet. Customers will get support in later releases.

### Features

Both OceanStor Dorado V6 and OceanStor 100D inherit a rich set of features from OceanStor 9000. Because they have a different architecture than 9000, the features will be enhanced and optimized on the new platform. According to the plan, a large set of features will be available on Oct 2020. More features are under development to be available in 2021. The Dorado V6 and 100D platforms will get more features added during their life cycle and users will get non-disruptive upgrades to get benefits from these new features and enhancements.

### Summary

Here is a high-level list of INDITEX requirements with Dorado and 100D numbers. Some of the features are in active development and will be available soon (please contact us for updated release schedules).

Based on our investigation, Dorado V6 fits INDITEX’s needs in that it offers excellent performance, enough scale out scalability, and rich features. **Roadmap part needs to be confirmed by HQ.**

|  |  |  |
| --- | --- | --- |
| Requirements | OceanStor Dorado V6 | OceanStor 100D |
| Performance – throughput | >20GBps per controller | 10(w)-16(r) GBps per controller |
| Performance - IOPS | ~600k IOPS per controller | 200k IOPS per controller |
| Performance - latency | 0.1ms-1ms | ~1ms |
| Capacity > 1PB | 6400 disks (32PB) | 200PB per filesystem |
| Scale out | Up to 32 nodes (8 engines) | Up to 4096 per cluster, 256 per filesystem |
|  |  |  |
| Synchronous Replication | 21Q2 | Roadmap |
| Asynchronous Replication | Oct 2020 | 21Q1 |
| Snapshot replicas to third cabins | Oct 2020 | 21Q1 |
| Inline deduplication | Oct 2020 | 21Q3 |
| Tiering/archiving of old data | Oct 2020 | Oct 2020 |
| Snapshots | Oct 2020 | Oct 2020 |
| Clone | Oct 2020 | Roadmap |
| QoS support | Oct 2020 | Oct 2020 |
| Encryption | Oct 2020 | Oct 2020 |
| NDMP support | 21Q3 | 21Q3 |
| WORM | 21Q3 | 21Q3 |
| Cloud tiering | 21Q3 | 21Q3 |