# OrangeFS DDN SFA12K Architecture

**Testing Performed at:**

**Clemson Center of Excellence in Next Generation Computing**
**Evaluation & Usability Labs**

**October 2013**



October 2013

**[Table of Contents]**

# Executive Summary

The DDN SFA12K platform with the addition of OrangeFS can be leveraged in a wide variety of high-performance, data-intensive computing environments. The growing list of industries that can benefit from this level of computing accounts for the increasing demand for a storage solution as flexible and scalable as the OrangeFS DDN SFA12K combination.

This solution guide describes use and horizontal scaling of the DDN SFA12K with SSD drives surfaced to hosts via fiber channel and leveraging OrangeFS for client file system access. It also outlines the scaling of OrangeFS with SATA drives.

Using 4 SSD Logical Unit Numbers (LUNs) each in a 4+1 RAID 5 configuration we were able to sustain 3 GB/s or 24Gb/s aggregate over the switched 10G network. Similar tests scaled to nearly 12GB/s with 16 storage nodes with spinning disk (LUNs). The scaling from 2 to 4 nodes provided a doubling in aggregate utilization. The linear scaling of both the DDN SFA12K and OrangeFS provides a solid solution for HPC, Big Data, scientific modeling, render farms, financial modeling, oil and gas simulation, genomics applications and many other applications.

# 1. Introduction

In the evolving world of high-performance, data-intensive computing, managing massive amounts of data efficiently and effectively has become a top priority. Many different kinds of workloads are pushing storage demands, including scientific applications such as scientific modeling, render farms, financial modeling, oil and gas simulation, genomics applications and many others.

In this white paper we highlight the key concepts behind creating a successful deployment pattern of OrangeFS using the DDN SFA12K (see Figure 1). We describe how you can achieve high performance with very few SSD drives as part of the configuration, as well as higher scaling with the scale-out architecture of the SFA12K, standard servers and OrangeFS.

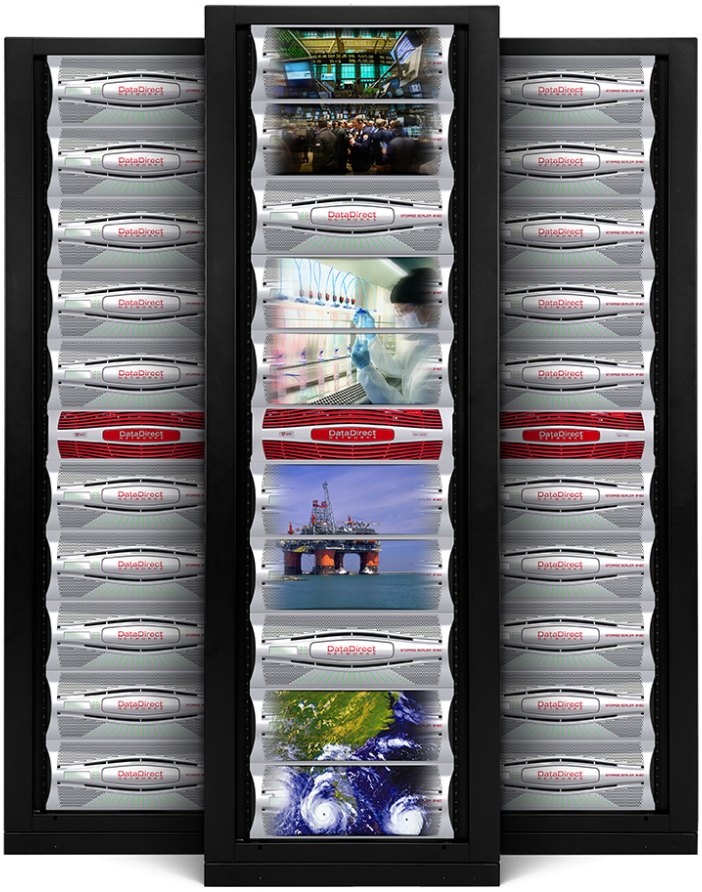

**Figure 1**

SFA12K™

# 2. Configuration

## 2.1. Hardware Configuration

The DDN SFA12K can use up to 1680 drives (SSD, SAS or SATA) and a maximum capacity of 6.72PB of storage. Drive enclosures come in three varieties:

- SS8460, 4U, 84 Drive Enclosure
- SS7000 4U, 60 Drive Enclosure
- SS2460 2U, 24 Drive Enclosure.

For network connectivity it can have up to 16 x 56Gb/s (FDR) Infiniband ports and 32 x 16Gb/s (FC16) Fibre Channel Ports.

For testing we used a DDN SFA12K-20 and several DDN SS2460 enclosures, with a mix of SSD and Sata HDD. For SSD testing we created 4 4+1 RAID5 LUNS. We disabled read-ahead cache and enabled write-back cache, which is the vendor-recommended configuration for SSD.

Dell R720s with E5-2600 Intel® Xeon® processors and PCI v3 bus, Mellanox 56Gb/s Infiniband Adapters and QLogic Fiber Channel Adapters are the servers used for testing. They were connected to the storage servers via Fiber Channel and to the clients via the Infiniband Network as shown in Figure 2.

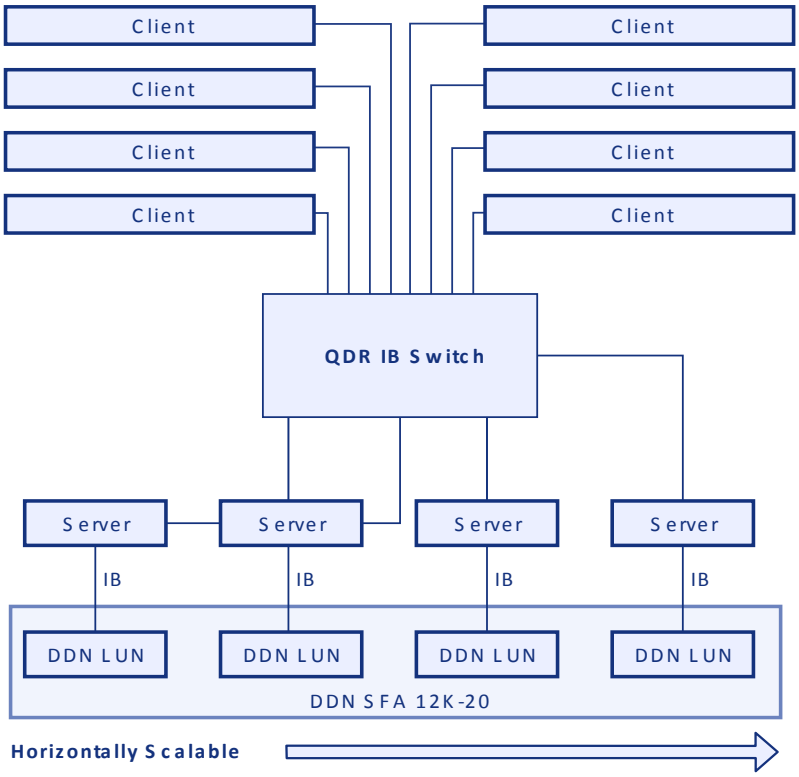


**Figure 2**

Another configuration tested for OrangeFS was comprised of an additional 12 servers, for a total of 16, with 2 5+1 SATA HDD RAID5 raidsets LVMed together with XFS.

### 2.2 Software Configuration

**Scientific Linux—**A Linux release put together by Fermilab, CERN, and various other labs and universities around the world. It exists primarily to reduce duplicated efforts and maintain a common install base among its scientific community of users.

**XFS**—A high-performance journaling file system originally created by Silicon Graphics for their IRIX operating system and later ported to the Linux kern
el. XFS is particularly proficient at handling large files and offering smooth data transfers.
**The following installation and configuration steps were performed:**

1) Installed SL 6.1 with Kernel 2.6.32-131.12.1 x86_64; disabled iptables and other services not needed for a private cluster environment.

2) Installed OpenFabrics Enterprise Distribution (OFED) + 6.1.0.0.72.

3) Configured XFS 3.1.1-4 on each SFA12k SSD LUN surfaced.

- Created the XFS filesystem:
  mkfs.xfs <lvname>

### 2.3 Additional Configuration and Methodology

**InfiniBand Network**—To facilitate the network, the InfiniBand QDR Switch was used to connect the PCI Express Adapters installed in the SFA12K storage array and 8 load driving clients.

To produce a simulated workload, IOzone was run concurrently on up to 8 client nodes over OrangeFS.

# 3. Software Overview

### 3.1 OrangeFS Overview

OrangeFS is a next-generation parallel file system based on PVFS for compute and storage clusters of the future. Its original charter—to complement high-performance computing for cutting edge research in academic and government initiatives—is fast expanding into a versatile array of real-world applications.

The general class of distributed parallel file systems ensures scalable, virtually unlimited growth. Simply take the highest performing single storage system at your site and add more of them. Clients now seamlessly write data across all of them.

Thanks to the unique architecture of OrangeFS, the performance of additional systems translates to proportional increases in capacity and throughput. OrangeFS clients also leverage the existence of single to multiple storage servers in the file system. In fact, the client can decide the order and number of storage servers to use when accessing a file, allowing the client to stripe a single file across multiple storage servers. OrangeFS metadata can also be stored on a single storage server, in conjunction with the file data, or spread across multiple storage servers (configurable on system, directory or file basis).

In comparison to similar parallel files systems, OrangeFS offers two significant advantages:

- Based on the powerful and modular PVFS architecture, the performance enhancement potential of OrangeFS has always been a fluid, ongoing process. This has allowed an ever-advancing design to incorporate distributed directory metadata, optimized requests, a wide variety of interfaces, and features. *It is well-designed.*
- OrangeFS is an extremely easy file system to get, build, install and keep running. PVFS has been used in numerous educational, experimental and research settings and has formed the basis of many graduate theses. *It is very usable.*

Primary targeted environments for OrangeFS include:

- High performance computing in all scales
- Hadoop/Big Data
- Rendering farms
- Financial analytics firms
- Oil and gas industry applications

For many years PVFS development focused primarily on a few large scientific workloads. At the same time members of the community used PVFS as a research tool to experiment with different aspects of parallel file system design and implementation. OrangeFS is broadening that scope to include production-quality service for a wider range of data intensive application areas. This has led to re-evaluating a number of assumptions that were valid for PVFS but may or may not be appropriate for these other workloads. Thus, a new generation of development is underway to address these new scenarios.

OrangeFS design features include:

- Object-based file data transfer, allowing clients to work on objects without the need to handle underlying storage details, such as data blocks
- Unified data/metadata servers
- Distribution of file metadata
- Distribution of directory entry metadata
- Posix, MPI, Linux VFS, FUSE, Windows, WebDAV, S3 and REST interfaces
- Ability to configure storage parameters by directory or file, including stripe size, number of servers, replication, security
- Virtualized storage over any Linux file system as underlying local storage on each connected server

OrangeFS has been hardened through several years of development, testing, and support by a professional development team. Now it is being deployed for a range of applications with commercial support, while still maintaining complete integrity as an open source project.

# 4. Evaluation

### 4.1 Evaluation

IOzone version 3.397 was used to test the read and write performance with an increasing number of clients and storage servers.

First, a local IOzone benchmark of single LUN was performed with an increasing number of local processes performing IO, (see Figure 3). The following commands were used, with *NUM_PROCESSES* being incremented (1, 2, 3, … , 8):

```
# Write
iozone -i 0 -c -e -w -r 256k -s 4g -t $NUM_PROCESSES -+n
# Read
iozone -i 1 -c -e -w -r 256k -s 4g -t $NUM_PROCESSES -+n
```

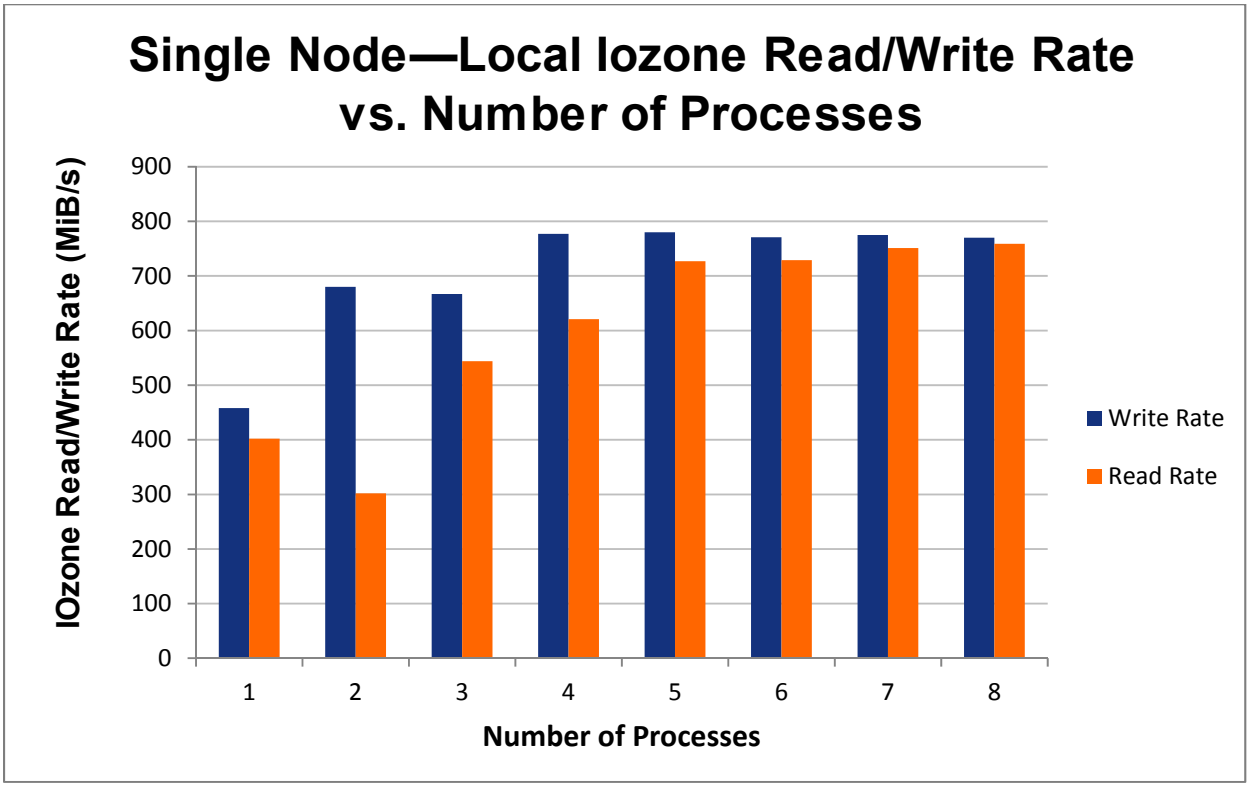


**Figure 3**

Next, IOzone benchmarks were performed against 1, 2, and 4 OrangeFS server configurations using remote clients (1 process per remote client). The IOzone record size was 256k for the single-server configuration, 512k for the 2-server configuration, and 1024k for the 4-server configuration. The record size was doubled as the number of OrangeFS servers doubled, to leverage all constituent drives of each LUN. Since the stripe size used in the 4+1 RAID5 configuration was 64k, then for a single LUN: 4 x 64k = 256k would be optimal. This optimal number was multiplied by the number of OrangeFS servers involved to determine the record size for a particular IOzone benchmark. Each OrangeFS configuration was benchmarked using 1, 2, 4, and 8 remote clients(see Figure 4). The following commands were run, with variable parameters shown in italics:

```
# Write
iozone -i 0 -c -e -w -r $RS -s 4g -t $NUM_PROCESSES -+n -+m $CLIENT_LIST
# Read
iozone -i 1 -c -e -w -r $RS -s 4g -t $NUM_PROCESSES -+n -+m $CLIENT_LIST
```

The results are shown in Figure 3. Read performance with 4 servers reached just under 3GB/s, and write performance reached just under 2.5GB/s.

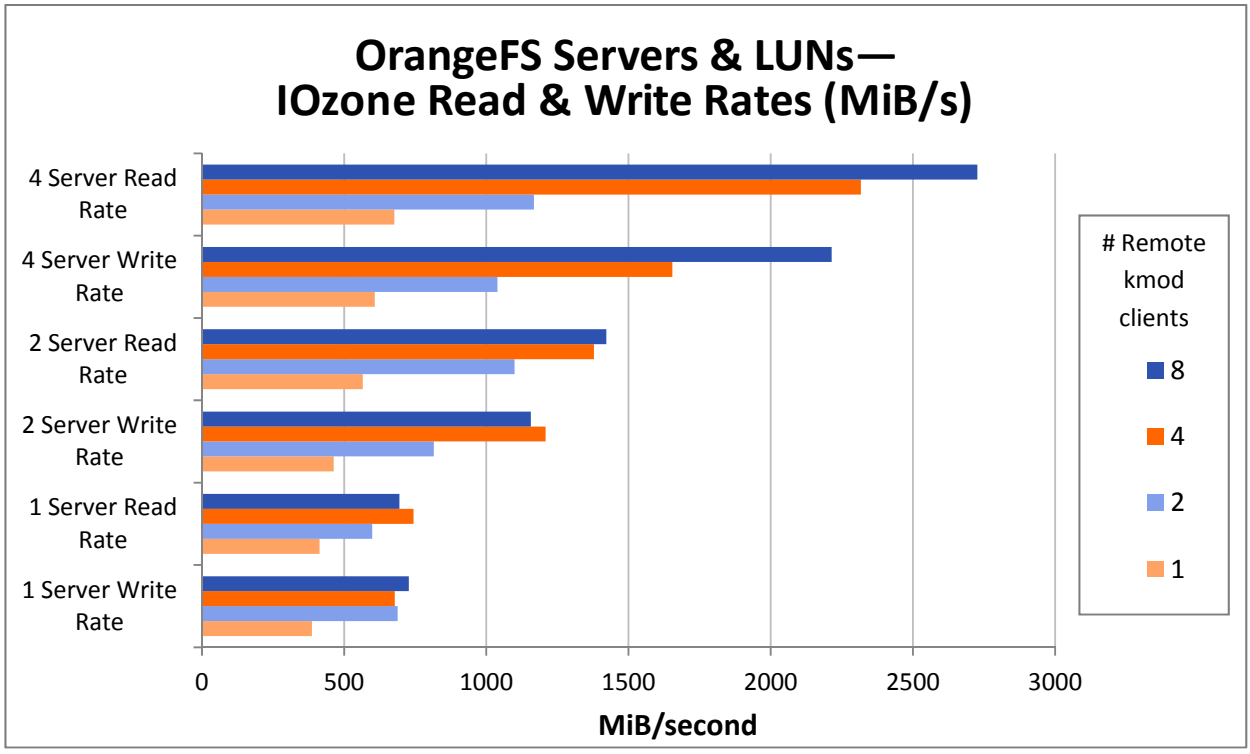


**Figure 4**

For the HDD Scaling evaluation 16 storage servers were tested with up to 32 clients, with read performance reaching nearly 12GB/s and write performance reaching nearly 8GB/s, See Figure 5.

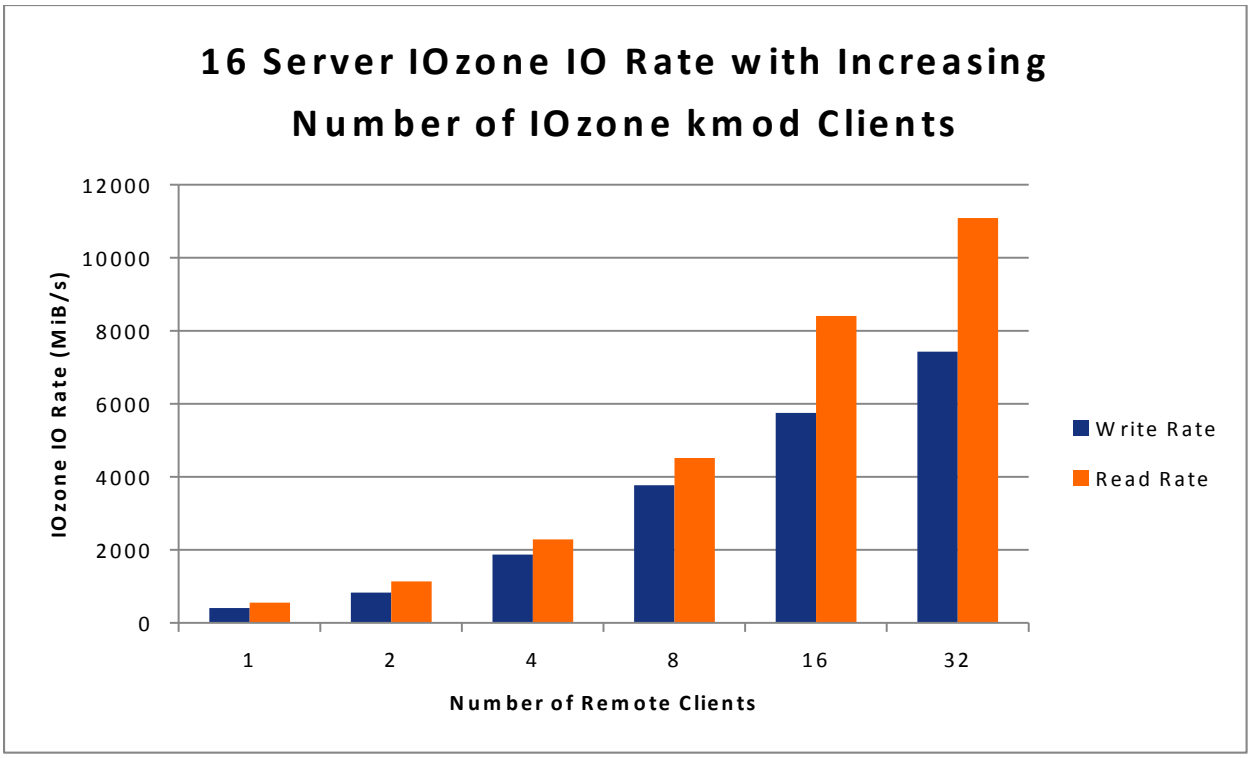


**Figure 5**

# 5. Conclusion

In the evolving world of high-performance, data-intensive computing, managing massive amounts of data efficiently and effectively has become a top priority.

The SFA12K Storage Platform with OrangeFS is an optimum solution for scale-out storage. The tests performed show excellent scaling, which can be leveraged for a wide assortment of applications and workloads including scientific applications such as scientific modeling, render farms, financial modeling, oil and gas simulation and genomics applications.