

Xyratex ClusterStor6000 & OneStor

Proseminar „Ein-/Ausgabe –
Stand der Wissenschaft“

von Tim Reimer 26.07.2013

Betreuer: Julian Kunkel

Structure

1. About Xyratex	3
------------------------	---

2. OneStor

2.1 OneStor SP	3-4
2.2 OneStor AP	4
2.3 OneStor Enclosure	5
2.4 "Green" OneStor	6

3. ClusterStor6000

3.1 Hardware Architecture	6-8
3.2 Software Architecture	9-10

4. Conclusion	11
---------------------	----

1. About Xyratex

Xyratex has over 25 years of experience in research and development relating to disk drives, storage systems and manufacturing process technology. It was founded in an MBO from IBM in 1994 and has headquarters in the UK. Moreover it has operational facilities in North America, Asia and Europe.

2. OneStor

2.1 OneStor SP

OneStor Storage Platform is an enclosure by Xyratex which delivers ultra dense storage capacity.

It provides maximum availability through comprehensive fault diagnosis, fault logging and fault monitoring. It has also n+1 power cooling modules and dual 6Gb/s I/O modules and dual data paths to all drives in an enclosure.

OneStor further includes the Unified System Management by Xyratex, a tightly to the hardware coupled embedded software.

Another advantage of OneStor is that it supports OEMs by accelerating their market introduction and simplifying development and testing.

Additionally it gives OEMs the opportunity to tailor their own unique brand requirements like labeling, logo printing and product packaging. Therefore it is ideal for enterprise-class applications such as big data analysis and digital media.

2.2 OneStor AP

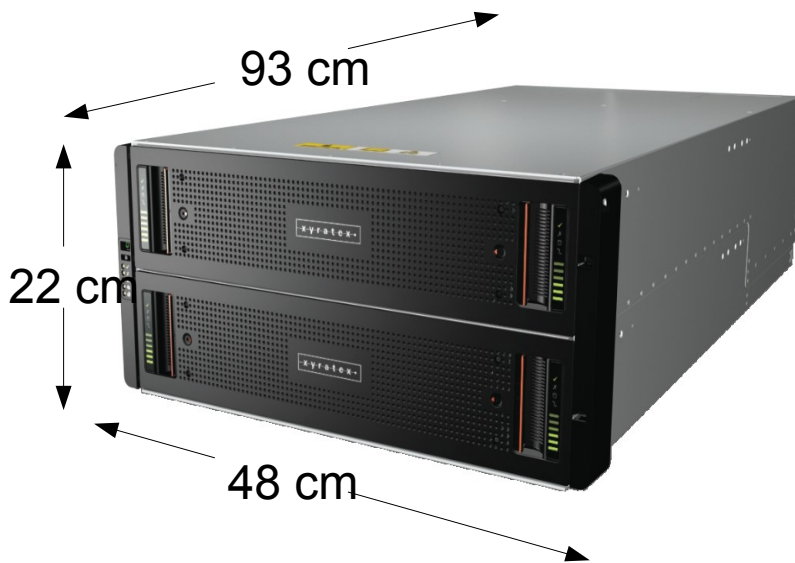
The OneStor Application Platform delivers an ultra dense storage server building block for OEMs so that they can integrate their unique software solution stack into cloud computing markets, big data analytics and digital media. It mainly consists of a scale-out storage server architecture, two embedded server modules (ESM) and the Unified system management API (USM).

With the scale-out server architecture the application performance scales along with capacity increases. The ESMs offer server-level processing capabilities directly on board and deliver more scale-out processing where it is needed.

And the USM gives OEMs the ability to design their own management and diagnostic systems which are universal across their products lines and also helps them to bring new products to market.

2.3 OneStor Enclosure

In one of those enclosure is space for a maximum amount of 84 hard disk drives. It can be chosen from 2TB drive up to 3 and 4TB drives which enables a maximum storage capacity of 168, 252 or 336TB per enclosure depending on the chosen capacity for a single disk drive.



OneStor SP enclosure

Weight: 128 Kg



OneStor SP enclosure – inside view

2.4 "Green" OneStor

OneStor has an individual drive power control and advanced adaptive cooling technology in form several additional fans for each enclosure. Its green design meets worldwide strict recycling requirements for environmental friendliness.

Furthermore it has got a High Efficiency 80 PLUS Gold Certification (which is a registered trademark of Ecova Plug Load Solutions) for over 80 percent of efficient power transformation.

3. ClusterStor 6000

3.1 Hardware Architecture

The hardware of ClusterStor 6000 consists of four components including the Cluster Management Unit (CMU), which is the central point of management, the Scalable Storage Unit (SSU), which is the crucial part for Xyratex' scale-out storage architecture and the Network Fabric Switches and the Management Switch as general hardware management units for the SSU. The hardware features of the CMU are, among other things, an embedded RAID, the ClusterStor Manager Software and a pair of servers. The management server provides and receives information and the meta data server makes meta data from its targets, which store the data on disk, available.

Inside of one SSU there is space for up to 84 hard disk drives with a capacity of two to four TB per drive. The SSU hosts two object storage server nodes, which provide file I/O service for their storage targets.



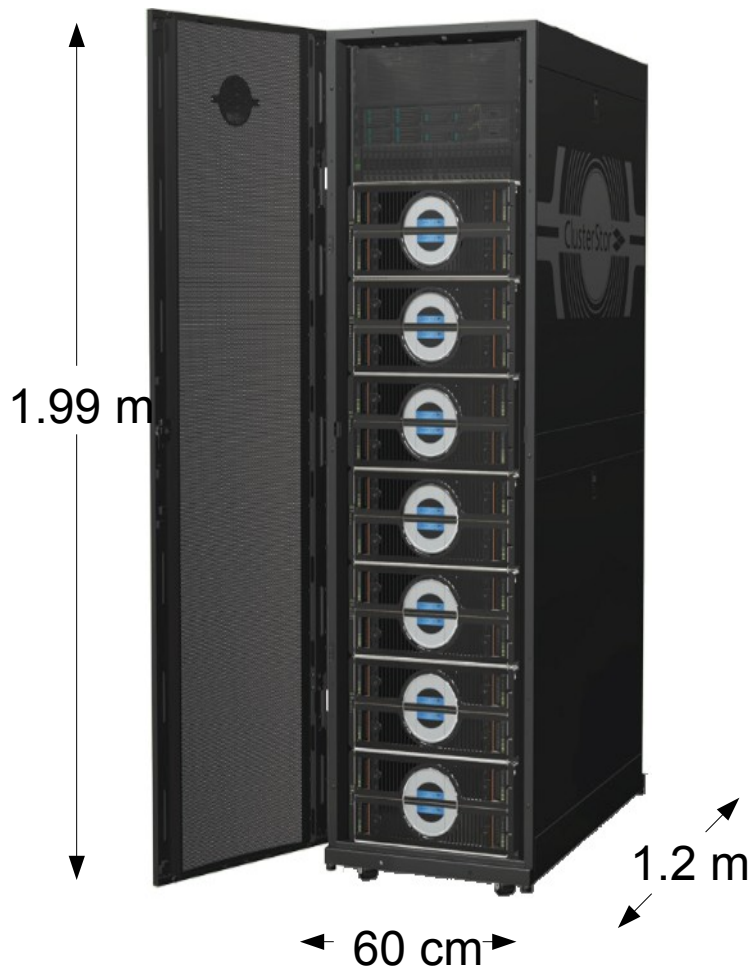
Scalable Storage Unit

It also contains two Embedded Server Modules (ESM), which can directly access all drives and grant active-active failover. This means both ESMs are active and if one of them fails the other one manages its object storage targets. Furthermore the ESMs are connected to all drives in the SSU with a common midplane delivering a high speed interconnect.

They are also connected to several Network Fabric Switches, which are managing the I/O traffic. These are usually realized with InfiniBand or a 10GbE or a high speed 40GbE to maximize the network availability and reliability.

The Management Switch enables the ClusterStor Manager to power-cycle the ESMs and consists of a local network with 1GbE.

Into a ClusterStor 6000 rack fits a maximum of seven SSUs with a possible capacity of 588 drives and 2,352 TB of storage. It has a maximum throughput of 1 TB/s.



Weight: 1,141 Kg

ClusterStor 6000 rack

3.2 Software Architecture

The Software Architecture of ClusterStor 6000 consists of four major parts, which are the Lustre File System, a Data Protection Layer, a System Management Software and the ClusterStor Manager. The Lustre File System is a server based architecture for large-scale computing and powers the world's top HPC clusters. It provides petabytes of storage and hundreds of gigabytes per second of I/O throughout. In addition it processes meta data, manages free space and also presents its file systems to clients and is therefore perfectly shaped for the ClusterStor 6000 storage solution.

The Data Protection Layer protects against double disk failures with a RAID 6 array. It supports two hot spares where data rebuilds on if one disk fails and it also consists of write intent bitmaps, which aid the data recovery and so reduce the recovery time from hours to just a few seconds.

Furthermore there is the Unified System Management Software to mainly retain the management system health. The software runs on each ESM in a SSU and controls and monitors the SSU's hardware infrastructure by controlling power of hardware subsystems, efficient adaptive cooling, event capturing for post failure analysis and status monitoring.

The software which finally runs on Linux OS is the ClusterStor Manager where you have a simply structured view of ClusterStor's infrastructure and also has got a browser based GUI component. This component enables you to start and stop file systems, collect and browse performance data and it monitors the node status.

Software overview:

The screenshot displays the ClusterStor Manager interface for host 'dvtrack202'. The top navigation bar includes 'Node Control', 'Performance', 'Log Browser', 'Support', 'Terminal', 'Dashboard', 'Health', and 'Config'. A status bar at the top indicates 12 UP, 0 DOWN, 0 UNREACHABLE, 0 PENDING, and 0/12 TOTAL nodes. The main content area shows a 'Service overview for "dvtrack202"' with a 'Host: dvtrack202 Service: Host Perfdata' and a '25 Hours' time range. Two performance graphs are visible: 'Round Trip Times' (Ping times) and 'Packets Lost'. The 'Round Trip Times' graph shows a peak of 0.42 ms and an average of 0.20 ms. The 'Packets Lost' graph shows 0% loss. A sidebar on the left contains navigation menus for 'General', 'Status', 'Problems', 'System', and 'Monitoring'. A 'Services' list on the right includes Host Perfdata, Current Load, Network statistics, RAM usage, Root Partition, Swap Usage, and Total Processes. The footer shows '© 2012 Xyratex Technology Limited All Rights Reserved', '2012-02-20 02:28 PDT', and 'ClusterStor Manager 1.1 by build #10145 01/05/2012'.

The screenshot shows the 'Terminal' view in ClusterStor Manager. The terminal session starts with 'Login: admin' and 'Last login: Mon Feb 20 01:10:35 2012 from 10.0.101.74'. The user runs the command 'sudo /opt/xyratex/bin/cwcli show_nodes -c fal'. The output is a table with the following data:

Hostname	Node type	Power state	Lustre state	Targets	Partner	RA Resources
dvtrack200	nds	on	mfs	0 / 0	dvtrack201	None
dvtrack201	nds	on	mfs	2 / 2	dvtrack200	Local
dvtrack202	oss	on	mfs	4 / 4	dvtrack203	Local
dvtrack203	oss	on	mfs	4 / 4	dvtrack202	Local
dvtrack204	oss	on	mfs	4 / 4	dvtrack205	Local
dvtrack205	oss	on	mfs	4 / 4	dvtrack204	Local
dvtrack206	oss	on	mfs	4 / 4	dvtrack207	Local
dvtrack207	oss	on	mfs	4 / 4	dvtrack206	Local

The footer of the terminal window shows '© 2012 Xyratex Technology Limited All Rights Reserved', '2012-02-20 02:24 PDT', and 'ClusterStor Manager 1.1 by'.

ClusterStor Manager Software

4. Conclusion

All in all OneStor and ClusterStor 6000 by Xyratex are ideally shaped for Big Data and High Performance Computing. While OneStor is focused on enterprise class Cloud Computing including big data analytics and digital media, ClusterStor 6000 is used for scientific research, climate modeling and simulation and energy exploration.

With this two data storage solution Xyratex has developed efficient and innovative possibilities to store petabytes of data without losing performance and with a simple software to manage it.

Sources:

<http://www.xyratex.com/>

http://www.xyratex.com/sites/default/files/files/field_inline_files/OneStor_SP2584_DS_1-0_0.pdf

<http://www.prnewswire.com/news-releases/xyratex-announces-onestor-2584-the-highest-density-storage-platform-solution-shipping-in-the-market-today-143647516.html>

http://www.xyratex.com/sites/default/files/files/field_inline_files/ClusterStor%206000%20Datasheet.pdf

http://www.ecmwf.int/newsevents/meetings/workshops/2012/high_performance_computing_15th/Presentations/pdf/Kling_Petersen.pdf

http://www.xyratex.com/sites/default/files/files/field_inline_files/Xyratex_white_paper_ClusterStor_The_Future_of_HPC_Storage_1-0_0.pdf