May 3, 2013
Jeremy Li

Credit: NexGen Storage Technical Overview for County of Los Angeles

NexGen Storage Inc. (http://www.nexgenstorage.com/) was founded in late 2009 by John Spiers and Kelly Long, the same founders of LeftHand Networks that was purchased by HP in 2007. Its mission was to solve shared storage performance problems with a new architecture built from the ground up for Quality of Service (QoS). Recently, NexGen was acquisition by Fusion-io. Below is the link to view its press releases:

http://www.fusionio.com/press-releases/fusion-io-acquires-nexgen-storage

Note: HP's Leafhand storage product was renamed by HP as P4000 series based on iSCSI technology in a similar fashion as Dell EqualLogic PS Series Storage.

As a startup company, NexGen has about 70 customers within 10 months after its first Hybrid Storage System, N5-100, was introduced in the market in the middle of 2012 and received a few distinguished awards: (1) CRN Partner Program Winner 2012; (2) 2011's Five Hot Storage Startups; and (3) Silver Cup in the enterprise storage systems category in Storage magazine's & SearchStorage.com's 2012 Products of the Year competition.

The hybrid storage solution delivers the performance of an all flash array at the price of hybrid system that can address both the performance and capacity issues at the same time with its unique design by putting the solid-state drives (SSDs) next to the PCIe in a hybrid architecture, as illustrated in Figure 1 below while most storage vendors designed SSDs behind SAS (designed for high latency disk drives) and share the SSDs with the SAS drivers in the same back plane, thus, reduce the storage capacity, and the RAID Controller might be the source of the bottleneck, as shown in Figure 2 below:

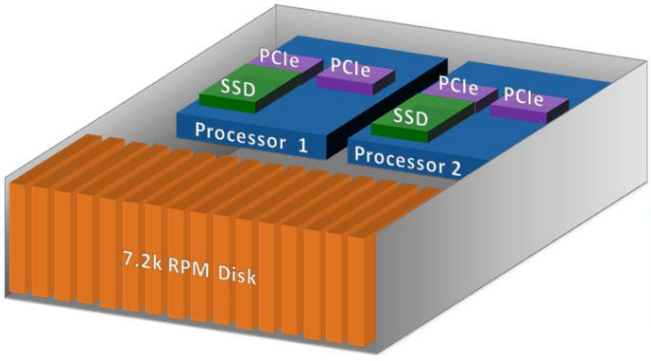


Figure 1 – NexGen Solution

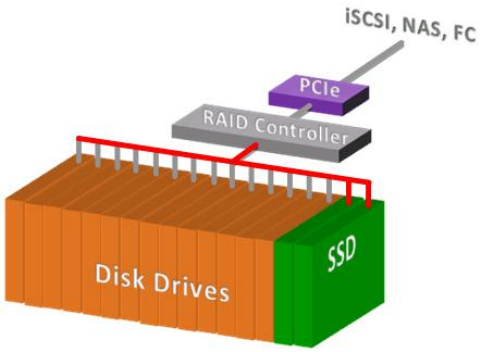


Figure 2 – Other Vendors' Solution

Therefore, NexGen solution can achieve: (1) Solid-state at computer CPU and RAM speeds, extreme low latency; (2) HDD slots are not wasted on SSD means

more storage capacity - Balanced Performance & Capacity; (3) Active-active means resources aren't idle; (4) Read/Write solid-state delivers high performance for all workloads; (5) QoS; (6) Dynamic Data Placement and (6) Data Reduction, as illustrated in the screenshot below:

ioControl Operating Environment™		
Quality of Service	Dynamic Data Placement	Data Reduction
Volume QoS	Real-time, not batch	Inline pattern matching
Service Levels	Heuristics-based	Volume level dedupe
Reporting	N-Tier architecture	Thin provisioning
Live Policy Migration	QoS driven	Variable block ingest

n5 Storage System
ioControl Operating Environment™
Active-active storage processors
Redundant disks, fans, and power supplies
48 GB RAM
1.28 TB PCI-e Fusion-io solid-state
32 TB 7.2K raw, 22 TB usable
4 10 GbE or 16 1 GbE data ports, iSCSI
Optional performance pack (640 GB solid-state)
Optional capacity pack (32 TB disk)

Dynamic Data Placement is different than the auto tiering. NexGen has studied the tiering algorithms of yesterday and re-designed them to address the shortcomings of the legacy auto tiering referenced below:

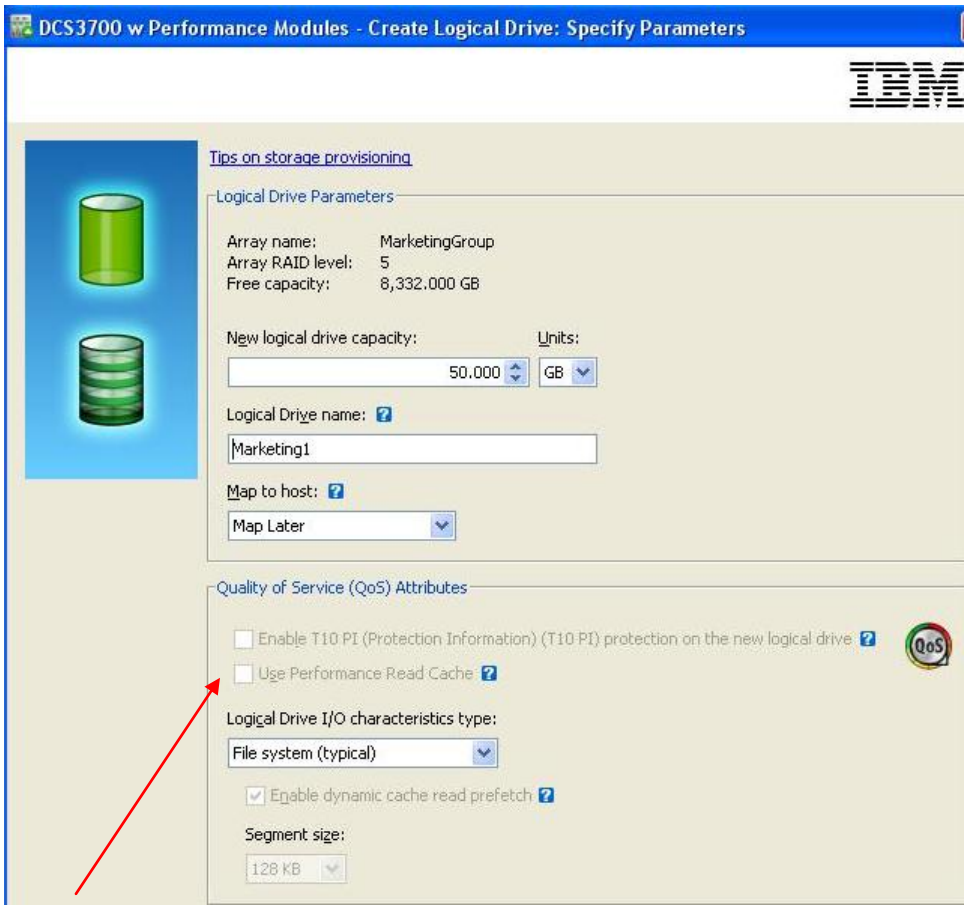
1) Does not use the realtime
2) Adds management complexity
3) Less predictable on performance

NexGen claims that the auto-tiering vendors (e.g., HP 3PAR, Dell Compellent and EMC) use reactive automation algorithms, while it uses the new dynamic algorithms to address the shortcomings that often causes customers to turn the auto-tiering off over time and go back to managing their system like they used to.

Below are the Real-Time Decision Factors used in NexGen systems:

- Current performance
- QoS setting
- Dedupe ratio
- Last accessed & frequency

Note: EMC, the world number one storage vendor, and other storage vendors such as IBM, NetApp and HP, can deliver the QoS for the storage capacity only, but cannot be able to deliver the QoS for the performance, as illustrated in the screenshot below:



Quality of Service (QoS)

According to EMC Press Release, placing Flash technology on a PCIe card can accelerate performance up to 4000X faster data access than 15K HDD - Source:

EMC² ▲ EMC Press Release

– "Putting Flash drives (SSD's) into the storage array achieved... faster data access than 15K HDD. Now, placing Flash technology on a PCIe card can accelerate performance up-to... 4000X faster data access than 15K HDD."

Note: Early adopters of the VDI technology must pay for a large amount of dollars due to the high IOPS requirement for the VDI solutions. Thanks for the Intel's leadership, the price of the flash (SSDs) has been reduced by a half today in comparison of the price one year ago. The trend will be continuing. Therefore, the VDI solution via the 2nd generation of the PCIe SSDs might be on the way with a reasonable price and performance ratio. Click on the link below about the SSD strategy with price and timing issue under item 5 on page 4 and 5 for details: http://www.lacaaea.com/vendors/SSD-Q1-2012.pdf

Copy-on-Write (COW) Snapshot vs. Redirect-on-Write Snapshot

NexGen uses Deferred Redirect-on-Write snapshot. The new writes are deferred by its data tiers (SSD and SAS) since all writes always hit flash before being committed to disk.

Note: IBM has eliminated the cascaded COW in the newer revision firmware on both DS2500 and DCS3700. Thus, the COW reduced the performance impact significantly, but the ROW snapshot is still better with 1000+ copies without impacting the performance of the arrays.

Boteh DS2500 and DCS3700 came from LSI. Therefore, the COW snapshot technology has been inherited from LSI. NetApp is not able to convert the COW snapshot into the ROW snapshot due to two different operating systems.

The demo is comprised of the followings:

1. **Control: Take ioControl**

Guarantee performance for applications with:

- **QoS: Set QoS based on Each application's Need**
- **Service Levels: Prioritized Performance during Degraded Mode Operation**

2. **Etc.**

Conclusion

NexGen N5 storage is a next generation storage system that uses the PCIe SSDs in a hybrid architecture, as illustrated in Figure 1 above to fully utilize flash performance to address both the performance and capacity issues, in addition to avoiding legacy array controller's potential bottleneck when the SSDs are placed along with SAS HHDs in the same enclosure, as illustrated in Figure 2 above.

NexGen plays in the Small Medium Enterprise (SME) market with its unique QoS, service levels and guaranteed performance of Mission Critical volumes / LUNS with an application. When the SSD, SAS or SATA drives are mixed in the same enclosure with a standard SAS/SATA backplane, the bottleneck from the array controllers might occur, thus, impact the performance. GexGen places its SSD's on the Fusion PCI-e which connects directly into four (4) 10GbE ports. As a result, Fusion PCI-e processors are in a true active/active mode and guarantee no degradation of performance if the QoS is applied.

Therefore, any SMB or SME with 1,000 to 10,000 employees with about 100 physical servers and 100's TB's of total storage can consider NexGen's solution.

However, it is not designed for large enterprise (e.g., Fortune 500) due to the lack of clustering capability and storage capacity with big overhead from four (4) RAID 6 dedicated parity drives for each 16-HDDs appliance.

NexGen N5 Series can be a good fit at some vertical markets for block-level applications (e.g., Exchange, SQL and VDI) due to having higher IOPS and QoS.

Challenge

1. The Lack of Clustering Capability

NexGen does not support the clustering capability. Therefore, a failover from one appliance to another is not possible. Therefore, there will be some challenges in the business continuity planning (BCP) and Disaster Recovery (DR) scenarios in the big enterprises.

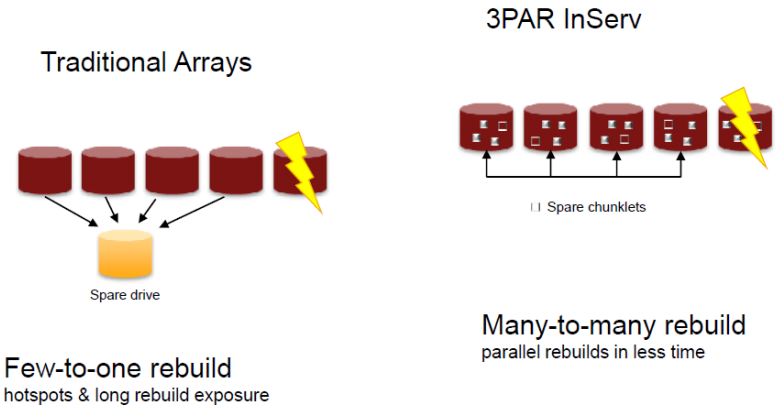
2. Dedicated Spare Disk Drives vs. Distributed Sparing Drives

The n5 Storage Systems do not use spare drives. The system is configured with two RAID 6 disk sets, better known as two RAID 6 groups, with 8 HDDs per set - 6 HDDs + 2 Parity HDDs. Therefore, each appliance will lose the capacity of four (4) HDDs dedicated for the purpose of the RAID 6 parity.

Rebuilding time for a 3TB drive in N5 Series will be much longer due to two dedicated parity drives per RAID-6 group. According to NexGen, the system can rebuild the RAID sets at ~40 MB/s. This allows the system to rebuild in less than 24 hours without significantly impacting system performance.

According to a field test, the average rebuilding time is at about 22 hours when the system might be at the capacity of 20% full. The rebuilding technique is a traditional technique – two parity HDDs will be acting like hot spots as a bottleneck when the rebuilding time occurs.

On the other hand, HP 3PAR uses a Distributed Parity Drives to rebuild the data for a lost drive, as illustrated in the screenshot below:



According to IBM and NetApp, a rebuilding time with traditional arrays might take 4 or 5 days with one 3TB drive of 80% capacity full, while IBM DS5000 or DCS3700 can finish the rebuilding drive under 96 minutes for most critical data and 12 hours for a full recovery as long as Dynamic Data Protection (DDP) in its newer firmware (FW) is enabled. This is because a RAID group on a pool where the 8+2 HDDs are randomized during the rebuilding process to all available HDDs, instead of a particular RADI group.

Note: IBM DS3500 and DCS3700 are OEM products from NetApp (E Series) which purchased from LSI that is based on block level, not file level (NAS) years ago. Therefore, the E Series is a true SAN and totally different from NetApp FAS series that is based on NetApp Data ONTAP 8 Operating System (OS), which has the longest OS history in storage arena. The E-Series has 160,000 systems installed worldwide and has 60% Federal Government Market Share as of today.

Caution: Depending on the application, the DDP will not be suited for Database and Exchange environments because logs must be written in a sequential pattern in order to achieve the optimal result. As a result, HP 3PAR or Dell Compellent method is a better choice if you would like to reduce the rebuilding time dramatically. It really deepens on an environment.

3. VMDK Files Whitespace Elimination

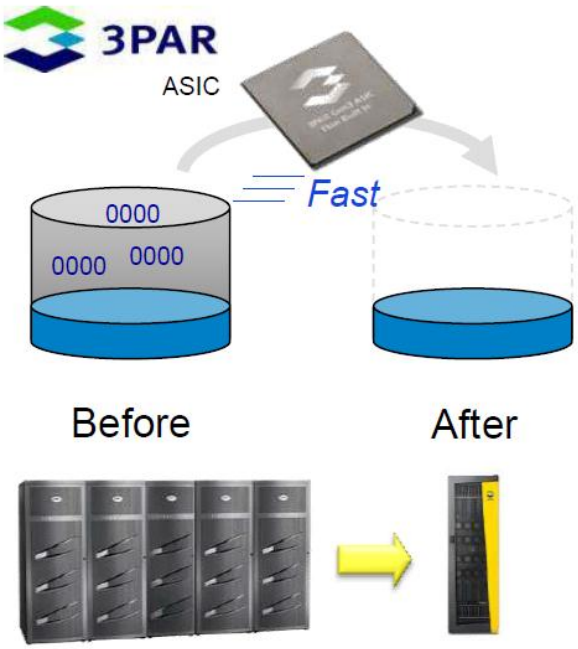
According to NexGen, a 4KB block size on SSD and 1.5 MB on SAS are used in N5 Series. This is more efficient, and reduces latency on spinning drives. The current ratio of the whitespace elimination is at 2 to 1 ratio. In the future, the efficiency will be increased in order to reclaim some unused disk space.

Since the N5 Series system is targeted to the block level applications (e.g., Exchange, SQL and Oracle), the whitespace might be high after a long period time in production due to massive log files deletion daily. In other words, any block size below 1.5MB will not be able to be converted into a usable space, reclamation.

NexGen has a tool to identify the efficiency of the storage regarding the ratio of the whitespace and usable storage space.

Note: HP Unique 3PAR ASIC with built-in zero detection, as illustrated in the screenshot, below, can deliver:

- Simplicity and speed – eliminate the time & complexity of getting thin
- Choice - open and heterogeneous migrations for any-to-3PAR migrations
- Preserved service levels – high performance during migrations



Special Note:

NetApp has a 9-Month Payback Guarantee Program to guarantee a customer by reducing the storage space by 50% by moving data from one vendor's storage into its own storage via its V-Series (e.g., V6240c).

However, in order to qualify this guarantee program, any customer must pass an initial assessment by NetApp to determine whether NetApp can reduce its customer's storage space by 50%.

NetApp has a special assessment tool to determine: (1) how many does the whitespace have; (2) how many documents can be compressed; (3) how much can the block-level de-duplication on primary storage be achieved.

4. **Startup Company**

NexGen has many competitors. Challenging incumbent market share leaders such as EMC, NetApp, IBM and HP will be difficult.

Recommended Reading:

1. New York Gov. Andrew Cuomo says the state will consolidate about 50 data centers at the State University of New York. (Source: GovTech)

2. State of Texas Moves More Than 100,000 State Employees to Microsoft Cloud

The State of Texas is moving more than 100,000 employees onto Office 365 at a cost of about $3.50 per user, per month, making it the largest statewide deployment of email and collaboration services in the U.S.

3. How New York City is going to Consolidate 50 Data Centers from 40 City Agencies into One Location

Acknowledgement

Thanks for Leo Kameya, Regional Account Manager, Fusion-io, Inc., for presenting the NexGen Storage Technical Overview for County of Los Angeles.