

SSD: итоги 2012 года, прогнозы на 2013 год

Обзор основных тенденций развития технологий flash-памяти для enterprise-рынка, включающих: расширение доли flash с архитектурой 3d NAND, повышение надежности flash-памяти, обеспечение гарантированной производительности flash-памяти.



Сергей Платонов – менеджер по продуктам, компания AVRORAID.

Тенденции на рынке flash-памяти

2012 год ознаменовался стремительным ростом количества решений, использующих flash-память для enterprise-рынка. Согласно исследованиям IDC, рынок flash вырастет с \$1,7 млрд в 2011 г. до \$5,6 млрд в 2016 г. Технология flash сегодня является одной из наиболее спорных среди всех существующих решений хранения данных предприятия. Мы видим множество подходов в использовании данного вида памяти:

- гибридные и полностью основанные на flash CХД;
- кэширование и использование NAND-памяти в качестве уровня хранения;
- SLC-, MLC-аппараты и попытки заменить NAND-память;
- NVMe и SOP-PQI;
- накопители, подключаемые по SAS и PCI-E, появление SCSI-Express и SATA-Express;
- использование flash на стороне CХД и также в среде SAN.

Flash Memory Summit вырос в этом году более чем на 30% по количеству участников и посетителей. Основными темами форума стали следующие.

1) Снижение литографии NAND-памяти.

Производители flash-памяти стремятся улучшить характеристики GB/\$ и характеристики энергоэффективности. При этом, несомненно, меняется подход к архитектуре памяти. Считается, что в 2013 г. будет достигнут предел снижения размеров плавающего затвора и начнет развиваться архитектура 3d NAND (рис. 1).

2) Рост внимания к сохранению целостности данных и долговечности памяти.

Снижение размера ячеек, увеличение их плотности, все большее применение flash-памяти в коммерческих ЦОД делают необходимым работу над защитой данных при сбое питания, обнаружением и коррекцией ошибок.

Явно выделяется класс накопителей, предназначенный для интенсивных операций записи.

3) Обеспечение гарантированной производительности.

Производительность твердотельных накопителей сейчас очень высока, но когда речь идет об обеспечении гарантированной производительности, требуемой в современных ЦОД мы сталкиваемся с некоторыми сложностями. Процессы коррекции ошибок, wear leveling, выделения свободного пространства, сбора му-

сора могут значительно влиять на производительность, особенно при заполнении пространства. Также возникают проблемы при смешанном цикле работы приложений, при котором не достигается производительность чистых записи и чтения.

Какие изменения мы увидели на рынке в 2012 г. и чего нам стоит ожидать в 2013 г.?

All-Flash CХД

Заметен рост числа CХД, полностью построенных на flash-памяти. Данные системы перестают быть чисто нишевыми решениями, применяемыми только для ограниченного круга задач. Конечно, для приобретения уже некоторое время доступны модели CХД среднего и выше среднего сегментов от ведущих вендоров, имеющие в качестве носителей только SSD-диски. Но их показатели \$/IOps и тем более \$/GB оставляют желать лучшего. Цена одного гигабайта в таких решениях достигает 30\$, что в 15 раз превышает цену при использовании вращающихся дисков enterprise-класса. В первую очередь стоит обратить внимание на стартапы, которые изначально разрабатывали решения под flash-память и внутренние разработки all-flash CХД крупнейших производителей:

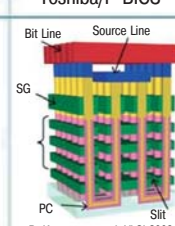
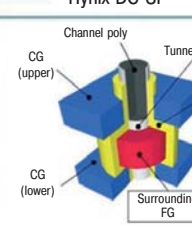
	Gate First		Gate Last
	Toshiba/P-BICS		Hynix DC-SF
			
Type of 3D NAND			Samsung/TCAT
Transistor	Gate all around; Salicided Poly Si gate	Gate all around; Salicided Poly Si gate	Gate all around; Damascene metal gate
Storage	Charge trap	Floating gate	Charge trap

Рис. 1. Архитектуры 3D NAND flash-памяти (ист. IMW 2011).

Nimbus Data, Pure Storage, SolidFire, Skyera, Tegile System, Kaminario, Violin Memory, Whiptail Technologies, XtremIO. Их производительность может достигать до миллиона операций ввода-вывода в секунду при низкой латентности, не превышающей нескольких миллисекунд.

Указанные производители работают над достижением максимальной производительности решения и снижением стоимости хранения гигабайта информации. Одним из таких способов снижения стоимости является переход с дорогой памяти SLC-аппаратов на eMLC.

При создании решения, полностью построенного на SSD, необходимо пересматривать подход к архитектуре СХД и набору обязательных функций. Первым же встает вопрос: использовать ли SSD как замену HDD с теми же интерфейсами и форм-факторами или отказаться от наработок в этой области? Большинство производителей пошли по второму пути, что, во-первых, позволило снизить затраты на R&D, а, во-вторых, использовать уже существующие на рынке SSD-накопители.

Компания Nimbus Data System и Violin разработали собственные устройства, содержащие память (eMLC), контроллерные модули и NVRAM.

Существующие алгоритмы кэширования становятся неэффективными в СХД, которые полностью построены на flash. Некоторые производители отказываются от использования WriteBack и ReadAhead-кэширования, а уже упомянутые компании Nimbus Data System и Violin перенесли кэширование на сторону запоминающих устройств.

Производительность современных SSD-дисков способна полностью утилизировать канал SAS 6G, поэтому конфигурация бэкэнд-интерфейсов должна быть изменена.

В качестве хост-интерфейсов могут применяться исключительно самые современные и производительные виды транспорта. Сейчас используются 56Gb Infiniband, 40G Ethernet и 16Gb Fibre Channel.

Многие производители реализуют технологии, аналогичные wearleveling, в устройствах SSD на уровне целых массивов для обеспечения балансировки IO по дискам и продления их долговечности.

Методы DataReduction, такие как Primary Deduplication и Online сжатие, необходимы в all-flash массивах, т.к. позволяют значительно снизить интенсивность записи на запоминающие устройства и при этом не создают тех же проблем с производительностью как на массивах с вращающимися дисками. Обязательным требованием является поддержка QoS на all-flash СХД.

Поскольку подобные устройства перестают быть чисто нишевыми решениями, в них начинают появляться функции, присущие всем СХД среднего класса — поддержка HA, DR, мгновенные снимки/копии, Thin Provisioning.

Что же можно сказать об основных играх рынка?

В мае 2012 г. компания EMC приобрела израильского разработчика горизонтально-масштабируемых all-flash СХД XtremIO. На VMworld EMC показала раннюю версию устройства, построенного с участием технологий XtremIO, известного как Project X. Появление законченного решения ожидается в 2013 г.

Поскольку был представлен прототип, конечное решение может отличаться от описанного ниже¹⁾.

Горизонтально-масштабируемая СХД состоит из так называемых X-Bricks, которые могут быть кластеризованы между собой. Каждый "кирпич" предлагает 4 8-ми гигабитных порта Fibre Channel и 4 порта iSCSI, использующих транспорт 10Gb Ethernet. В каждый модуль устанавливается 25 SSD на MLC-аппаратах. Суммарно каждый X-Brick предоставляет 7 Тбайт пространства. В качестве интерконнекта используется Infiniband.

Снижение нагрузки на накопители достигается за счет дедупликации, которая работает постоянно. Высокая производительность дедупликации достигается за счет вычислительных мощностей — каждый X-Brick содержит 24 процессорных ядра. Все операции внутри массива выполняются 4к блоками. Кроме дедупликации, массив обладает следующими функциями:

- горизонтальное масштабирование с близким к линейному росту производительности;
- Thin Provisioning;
- поддержка VAAI;
- мгновенные снимки и копии.

При этом гарантируется высокая производительность, независимо от степени заполненности массива или того откуда идет чтения — основного тома или снапшота.

Несколько месяцев назад СТО компании NetApp Jay Kidd сообщил, что в планах компании нет создания all-flash СХД, но в интервью, данном в середине декабря, он признался, что NetApp планирует выйти на этот рынок в 2013 г. При этом рассматриваются варианты поглощения одного из стартапов (вероятно, упомянутого в этой статье) и использования исключительно собственных технологий.

Компания IBM — уже на рынке all-flash СХД с продуктом, полученным после приобретения Texas Memory Systems. TMS предлагает несколько моделей, работающих на flash- и RAM-памяти. СХД Ramsan обеспечивает максимальную производительность при минимальных

затратах на пространство и электроэнергию. TMS Ramsan используют накопители на SLC- и eMLC-аппаратах²⁾.

HP представили адаптированные под SSD модели — StoreServ (3PAR) и StoreVirtual (LeftHand).

Компания Hitachi Data Systems планирует создать решение исключительно на собственных технологиях. Во время проведения Flash Summit компанией Hitachi Data Systems была объявлена стратегия использования технологии flash в рамках большинства Enterprise IT-продуктов. HDS был разработан пул технологий, выдвигающий SSD на совершенно новый уровень надежности, производительности и вместительности.

Одним из первых стал доступен Hitachi Accelerated Flash, использующий технологию многоядерного контроллера флэш-памяти от Hitachi, Ltd. Применение данной технологии позволило использовать flash на MLC-аппаратах для приложенного корпоративного уровня, превосходя по производительности имеющуюся память на SLC-аппаратах, и обеспечить при этом низкую стоимость в расчете на бит. Функция сжатия записываемых данных позволяет ускорить ее и повысить срок службы накопителей.

Модуль устанавливается в шасси высотой 8U и масштабируется от 6,4 до 76,8 Тбайт. Hitachi Accelerated Flash поддерживает все функции платформы Hitachi VSP, в том числе Hitachi Dynamic Tiering (HDT). Платформа Hitachi VSP может вмещать до четырех корзин с флэш-накопителями, что позволяет получить более 300 Тбайт флэш-памяти на одну систему³⁾.

Переход от SLC к MLC

В 2012 г. мы видим переход к использованию устройств на памяти MLC-аппаратах. Появление MLC и eMLC на корпоративном рынке было замечено еще в 2011 г., но прошедший год стал знаковым. Необходимо было бороться за основную характеристику накопителей, которая не позволяла пользователям отказываться от использования вращающихся дисков — GB/\$. Так, переход на MLC позволил стартапу Skyera построить решение стоимостью 3\$/GB, что превышает стоимость хранения на HDD корпоративного класса не более чем в 1,5-2 раза. Производители приближают выносимость и производительность данного типа памяти к показателям SLC. Добиваются они этого за счет усовершенствований контроллеров памяти, улучшая алгоритмы записи, оптимизируя механизмы отслеживания и коррекции ошибок, делая их незаметными для основных задач. Сейчас много говорят о появлении TLC-памяти, но данная технология еще не будет готова к появлению на корпоративном рынке в ближайшие 2 года.

В отчетах лабораторий нередко упоминаются успехи в адаптации PCM, STT-MRAM и RRAM. Но реальных устройств, поддерживающих данные технологии, стоит ожидать не ранее 2016 г.

1) http://www.storagereview.com/emc_project_x_hanson/;

http://chucksblog.emc.com/chucks_blog/2012/08/when-flash-changed-storage-xtremio-preview.html;

<https://community.emc.com/docs/DOC-20054>;

<http://www.emc.com/collateral/campaign/global/forum2012/chicago/flash-1st-the-storage-strategy-for-the-next-decade.pdf>.

2) https://www.ibm.com/developerworks/mydeveloperworks/blogs/DataCenter7/entry/ibm_flash_storage2?lang=en;

<http://www-03.ibm.com/systems/storage/news/announcement/flash-20120731.html>;

<http://www.ramsan.com/products/rackmount-flash-storage/ramsan-820>.

3) <http://www.hds.com/corporate/press-analyst-center/press-releases/2012/g120823.html>

Кэширующие устройства на стороне серверов

Кэширующие устройства на стороне серверов стали одним из самых инновационных подходов применения flash в средах предприятий. Перенос кэша сети хранения данных на сторону хостов, очевидно, стал новым подходом. Сейчас подобные решения планируются к анонсу у всех крупнейших производителей систем хранения данных: EMC, NetApp, Dell, HP и др. Часть независимых производителей базовых компонент таких решений – FusionIO, LSI, OSZ, Intel, SanDisk и Velobit уже анонсировали их (см., например, SN № 2/50 и 3/51, 2012 с описанием продукта LSI Nytro WarpDrive, прим. ред.).

В первую очередь, подобные решения применяются для ускорения СУБД и финансово-аналитических систем. Они имеют ряд ограничений использования, но дают максимальный прирост производительности подсистемы хранения, т.к. акселераторы устанавливаются непосредственно в шину PCI-E и взаимодействуют напрямую с вычислительными модулями, исключают также задержки интерфейсов и имеют поразительную пропускную способность. Кроме того, производительность акселераторов постоянна и прогнозируема (в том числе, благодаря использованию специализированного ПО).

В отличие от all-flash, СХД-акселераторы требуют некоторого времени до начала повышения производительности – от 30 минут до 1 часа.

В конце августа 2012 г. компания EMC сообщила, что расширяет возможности продукта VFCache, добавив в него поддержку дедупликации. Дедупликация позволяет повысить показатель \$/GB и обеспечит большую выносливость flash-памяти (<http://www.emc.com/about/news/press/2012/20120828-02.htm>).

В течение 2013 г. можно ожидать расширения областей применения акселераторов, в первую очередь это будет касаться работы в кластерных средах и возможности кэширования записи. Ожидаемый в 2013 г. Dell Fluid Cache позволит выполнять кэширование записи данных. Разработка Fluid Cache стала возможной после поглощения Dell стартапа RNA Networks, разработчика распределенного между хостами пула памяти.

Несомненно, мы увидим увеличение объемов и производительности акселераторов, а также постепенное исчезновение акселераторов на SLC-аппаратах и расширение линейки моделей, использующих eMLC.

Также стоит ожидать все большую интеграцию акселераторов с СХД с точки зрения управления и обслуживания.

В 2013 г. появятся первые устройства, использующие новые интерфейсы: SCSI Express, NVMe и SATA-Express.

NVMe будет обеспечивать уровень задержек в 3 раза меньший, чем SAS/SATA и уже сейчас прототипы устройств достигают 1 млн IOPS.

Новый универсальный коннектор SFF 8639 позволит использовать устройства, подключаемые к бэкплейн-системы с возможностью горячей замены.

При этом бэкплейн сможет поддерживать 12Gb SAS, 6Gb SATA, SATA Express, SAS Express, NVMe/SOP-PQI, многолинковый SAS (4-портовые устройства).

Кроме производителей полностью законченных решений, появляются исключительно программные решения, позволяющие использовать накопители третьих производителей.

Компания QLogic создала решение Mt. Rainier, которое представляет собой симбиоз HBA и хранилище промышленного стандарта на флеш-NAND, привнося в SAN-производительность серверное SSD-кэширование. SSD-накопители, подключенные к Mt. Rainier, объединяются в пул высокопроизводительного хранения данных внутри кластера серверов, который может быть выделен для различных мультисерверных приложений, сохраняя при этом преимущества управляемого хранилища SAN.

Ожидается появление продукта компании EMC, носящего в настоящее время название "Thunder". "Thunder" представляет собой распределенный серверный flash, работающий в так называемой Server Area Network, который может быть использован в кластерной среде и масштабируется как вертикально, так и горизонтально. Кроме того, он может использоваться с blade- и cloud-серверами.

Thunder – это решение стоечного фактора высотой 2U и 4U, обеспечивающее производительность в несколько миллионов IOPS при сохранении низкой латентности (<http://www.emc.com/collateral/campaign/global/forum2012/chicago/flash-1st-the-storage-strategy-for-the-next-decade.pdf>).

Акселераторы Fusion-io также могут быть использованы в качестве сетевого устройства, благодаря разработанному компанией ПО ION Data Accelerator software.

Сергей Платонов,
компания AVORRAID

RAIDIX объявила о поддержке интерфейса Fibre Channel 16Gb

Декабрь 2012 г. – Компания RAIDIX, производитель программного обеспечения для высокопроизводительных систем хранения данных, объявила о совместимости сетевых карт 16Gb Fibre Channel с программным продуктом RAIDIX 3.0 SAN software.

RAIDIX – один из первых производителей, представивших на рынке 16Gb FC как часть технологического решения. Это дает потребителям возможность поддерживать новейшие сетевые интерфейсы промышленного стандарта. В медиаиндустрии, для нагрузок которой оптимизирован RAIDIX 3.0, использование СХД с поддержкой новых карт 16Gb FC позволит создавать видеоконтент форматом 4K и 8K и обеспечивать непрерывный вещательный процесс.

"Партнеры RAIDIX ждали сообщения о поддержке в нашем продукте интерфейса 16Gb Fibre Channel еще с момента появления новых карт на российском рынке в начале зимы, – комментирует Сергей Разумовский, генеральный директор RAIDIX. – Это позволит им расширить линейку предлагаемого оборудования, "возглавив" распространение новой технологии".

Первыми устройствами 16Gb FC, сертифицированными с программным обеспечением RAIDIX, является семейство ATTO Celerity 16Gb Fibre Channel HBAs, использующее новейшую технологию PCIe 3.0.

RAIDIX



СХД для монтажных студий

- ✓ До 6ГБ/с в RAID6 на чтение и запись
- ✓ Неограниченный размер LUN
- ✓ Стабильная производительность
- ✓ Поддержка A/V приложений