

# Файловые СХД HP в эпоху cloud computing

Обзор архитектурных и функциональных особенностей, произошедших в линейках NAS-систем компании HP за прошедший год (“3 этапа эволюции сетевых файловых СХД”, SN № 4/37, 2008).



Александр Грубин – ведущий специалист, Департамент систем хранения данных, HP Россия.

## Введение

Линейка NAS-систем HP StorageWorks, описанная в статье “3 этапа эволюции сетевых файловых СХД” (SN № 4/37, 2008), значительно эволюционировала после приобретения в начале августа 2009 г. компании IBRIX и на сегодняшний день имеет новые продукты как в начальном, так и в интерпрайз-сегментах.

NAS-системы начального уровня, построенные ранее на ОС Windows Storage Server 2003, сменили аппаратную платформу (теперь это серверы ProLiant G6) и саму ОС – теперь это Microsoft Windows Storage Server 2008 x64 Edition. Так же, как и раньше, есть варианты непосредственно файлового хранилища и отдельного шлюза, к которому можно подключать внешние системы хранения как от HP, так и от других вендоров. Модели серии X1000 это пол-

ноценные файлы, модели серий X3000 это шлюзы. Также появилась модель начального уровня X500 для рынка SOHO. Создание файловых ресурсов, настройка и управление доступом к ним производятся с помощью встроенного инструмента Automated Storage Manager (HP ASM), позволяющего просто и быстро выполнять все необходимые операции из единой консоли управления. Как и многие другие продукты, системы X1000–X3000 интегрируются с ПО HP Systems Insight Manager (HP SIM) и поддерживают пакеты HP ProLiant Support Pack.

Решение HP 4400 Scalable NAS File Services, построенное на технологии HP Scalable NAS, снято с производства. Вместо него были анонсированы 2 решения HP StorageWorks X5000 Network Storage System Gateway и HP PolyServe Software for Microsoft SQL Server.

Что касается решений Extreme Data Storage (ExDS 9100) на базе кластерной масштабируемой файловой системы Polyserve File System, то эти решения эволюционировали в продукты серии X9000, которая содержит как высокоемкие (от 82 Тбайт), высокоскоростные (1 или 10 Гбит/с Ethernet или Infiniband) хранилища на базе серверов HP BladeSystem и дисковых полок MDS600 (модели X9720), так и более доступные, но построенные по той же архитектуре модели X9320 и шлюзы X9300. Файловая система в этих масштабируемых NAS-системах также поменялась, теперь она позволяет создавать единое пространство имен (файловую систему)

общим объемом до 16 Пбайт и более, а количество узлов кластера, из которого формируется решение не ограничено 16-ю узлами, как раньше, а масштабируется горизонтально (scale-out) – теоретически до бесконечности (рис. 1).

Кроме того, под управлением файловой системы IBRIX Fusion File System в составе NAS-линейки появились такие опции, как автоматическая миграция данных с быстрых носителей на медленные (например с SAS на SATA), в зависимости от политик и параллельная удаленная репликация на файловом уровне. Естественно, есть возможность онлайн-расширения емкости, балансируя при этом данные равномерно – по добавленным дисковым ресурсам. Поддерживается одновременный доступ к ресурсам как по протоколам NFS, так и CIFS, кроме того, есть возможность высокопроизводительного доступа к использованию фирменной технологии Ibrx Client.

## Тенденции рынка

Основными современными драйверами рынка, способствующими развитию архитектуры файловых СХД, являются:

- стремительный рост объемов корпоративных хранилищ. Согласно оценкам IDC, несмотря на экономический спад, объем общемирового сетевого хранения данных в 2010 г. вырастет на 51%<sup>1)</sup>. Тенденции рынка последних лет свидетельствуют о резком увеличении доли неструктурированных данных в общем объеме данных, к которым можно отнести: цифровую графику/видео/аудио, компьютерные модели, результаты компьютерного моделирования, справочную информацию, исследования генома человека, медицинские рентгеновские снимки, трехмерное сканирование организма, моделирование недр земли и обработку данных геологоразведки и др. Исследования, проведенные ESG, показали, что общий объем постоянных данных в 2010 г. составит более 26000 Пбайт при соотношении постоянных данных к



### Малый бизнес

**X500**  
Small businesses with 10 or less PC's and Mac's who need to a simple solution to secure and share data on a limited IT budget

### Малый и средний бизнес

**X1000**  
Small, medium-sized, or workgroup IT environments that need to quickly add or consolidate storage to a single unified storage system

**X3000**  
Medium-sized, workgroup IT or enterprise environments that need to add services (file, iSCSI, mgmt) to an array/SAN, or need to build affordable HA unified storage

### Уровень корпораций

**X5000**  
Medium-sized or enterprise customers needing scalable unified storage system for enterprise array/SAN or iSCSI environments

**X9000**  
Large enterprise and scale-out environments that need to seamlessly serve, store and scale terabytes to multiple petabytes of unstructured data

Рис. 1. Линейка продуктов HP StorageWorks X Series Network Storage Systems.

<sup>1)</sup> IDC Worldwide Enterprise Storage Systems 2009–2013 Forecast: Economic Crisis Driving Quest for Storage Efficiency and Emerging Architectures: May 2009–Doc # 218252.

“динамическим” как 10:1. Однако уже в ближайшей перспективе это соотношение может возрасти до 100:1;

- ужесточение требований по минимизации затрат как на капитальные издержки СХД, так и на их администрирование, а также снижение стоимости единицы хранения данных;
- предъявление более высоких требований по масштабируемости СХД как по емкости, так и по производительности в условиях высокого градиента роста объемов данных, сокращения ИТ-бюджетов и необходимости поддержания высокого уровня непрерывности бизнеса;
- возрастание роли глобальных географически распределенных СХД, при использовании которых клиент не знает о конкретном месте хранения данных.

### Обзор архитектурных особенностей NAS-систем HP серии X9000

В наибольшей степени инновации затронули старшие линейки NAS-систем HP – X9000. Прежде всего, это связано с расширением диапазонов масштабируемости по емкости и производительности (без остановки ввода-вывода приложений), автоматизацией перемещения данных между уровнями хранения, автоматизацией балансировки нагрузки и использованием емкости по всем приводам СХД. С переходом на файловую систему IBRIX Fusion File System стало возможным поддержание системой HP X9000 файловой емкости до 16 Пбайт (теоретически неограниченно) в едином адресном пространстве.

Переход с файловой системы PolyServe в качестве базовой платформы в NAS-системах среднего и старшего класса на IBRIX Fusion File System позволил не только достичь большей масштабируемости по емкости и производительности, но и консолидировать файловые данные от Windows-, MAC-, Linux- и UNIX-серверов в едином пуле независимо от протокола (CIFS/NFS). При этом достигнута большая функциональность и надежность решения. Использование в качестве основной единицы хранения SAS/SATA-дисков в системах X9000 обеспечило минимальную стоимость хранения – менее 2 \$/Гбайт при значительно более высокой производительности в сравнении с альтернативными системами.

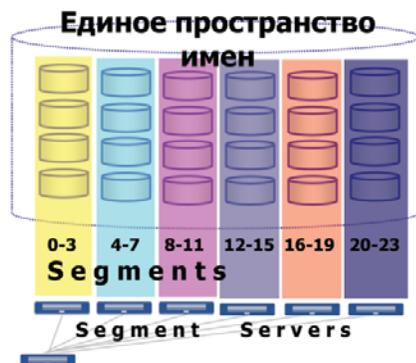


Рис. 2. Базовой единицей архитектуры IBRIX-кластера является сегмент.

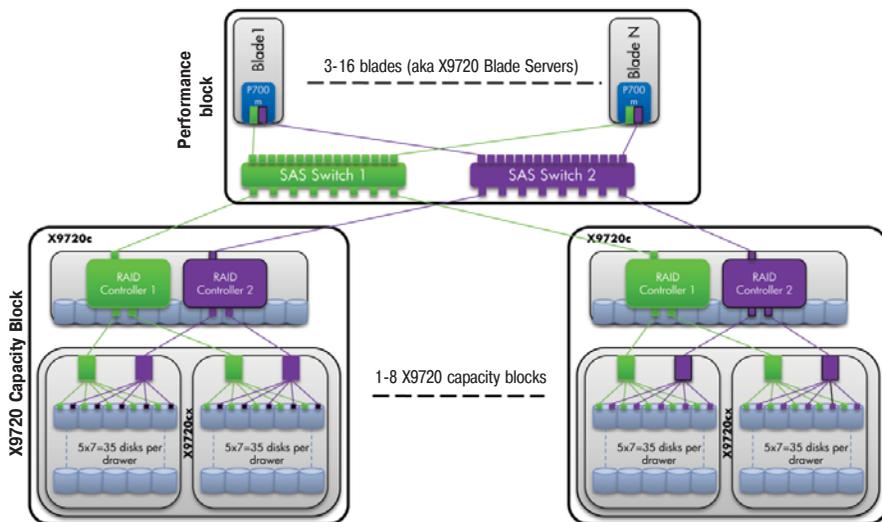


Рис. 3. Принципиальная схема SAS backend HP StorageWorks X9720.

В основе платформы IBRIX Fusion File System лежит распределенная параллельная сегментированная файловая система с распределенными метаданными и единым пространством имен, поддерживающая такие интегрированные функции управления данными, как: автоматизированное управление перемещением данных между уровнями хранения (Data Tiering), репликация данных между сайтами (Site-to-Site File Replication), мгновенные снимки (Snapshot), клонирование файлов (File Cloning) и др.

Базовыми единицами архитектуры IBRIX-кластера являются (рис. 2): сегмент (Segment), серверы сегментов (Segment Servers) и управляющий сервер (Fusion Manager Server). Сегмент это:

- минимальный элемент хранения данных (в 90% случаев это LUN на СХД). Например, в модели X9720 сегмент представляет собой группу дисков (10 HDD x 1 Тбайт), сконфигурированных по схеме RAID-6 с полезной емкостью 7,45 Тбайт, а в моделях X9320 серии Performance сегменты образованы из групп RAID-5 с полезной емкостью 2 Тбайт (6 HDD x 450 Гбайт);
- минимальный элемент файловой системы IBRIX;
- хранилище файлов и директорий.

Сервер сегментов выполняет следующие функции:

- управление элементами сегментов:
  - распределение данных;
  - хранение метаданных;
  - управление одновременным доступом к файлам (Lock Manager);
- обеспечение файловых сервисов:
  - протоколов NFS, CIFS, IbricClient;
  - File System и/или Segment Cache;
- предоставление клиентам кэша;
- взаимодействие с другими серверами сегментов (metadata information/operations; file information and access (NFS & CIFS); cluster status and other information).

Управляющий сервер может быть выделенным или распределенным (начиная с последней версии – май 2010 г.) по серверам сегментов. Он выполняет следующие функции:

- распределение отдельных сегментов серверам сегментов;
- обработка сбоев/отказов и управление серверами сегментов, обеспечивая функционал высокой доступности (HA);
- обеспечение подключения и взаимодействия с администратором;
- управление репликацией между площадками.

Конструктивно топовые модели линейки X9720 строятся на базе двух основных блоков – Performance Block и Capacity Blocks (рис. 3). Первый это блэйд-серверы от 3 до 16 – BL460c G6 (шасси c7000) в комплектации:

- Dual Quad Core Intel Xeon Processors 2.53 Ghz;
- два Internal 2.5-inch 300 GB 10K SAS drives;
- Dual 10Gb NICs (PCA);
- Dual 3 Gb/sec SAS Connections (Mezzanine Slot);
- 16GB Memory.

Блэйд-серверы выполняют функции серверов сегментов и управляющего сервера. Каждый блэйд-сервер обслуживает свой Performance Block, но логически соединяется с двумя Performance Block в

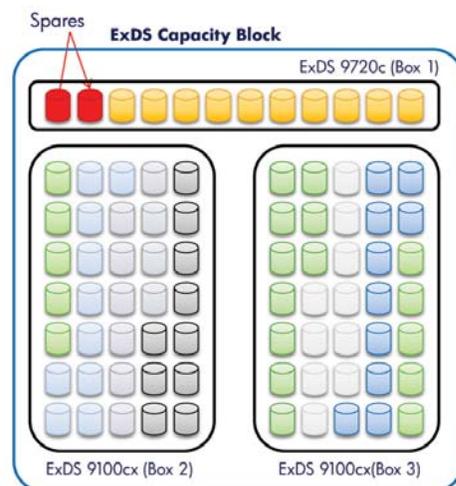


Рис. 4. Каждый Capacity Block имеет минимальную емкость – от 82 Тбайт – и состоит из 82 дисковых RAID-групп (82 x 1ТВ Midline SAS-диска 12+35+35, 59 TB usable).

режиме active-active и в случае отказа основного блэйд-сервера, связанного с Performance Block, берет на себя обслуживающие запросы к его сегментам.

Каждый Capacity Block имеет минимальную емкость – от 82 Тбайт, расширяется до 656 Тбайт, масштабируется по 82 Тбайт (рис. 4) и состоит из:

- 82 дисков в 7U (82 x 1TB Midline SAS диска 12+35+35 (59 TB usable);
- двух глобальных дисков горячей замены (первые 2 диска в контроллере);
- 8 дисковых групп (RAID групп):
  - каждая дисковая группа имеет 10 дисков;
  - уровень защиты группы – RAID6;
  - каждая дисковая группа имеет 1 логический диск LUN (сегмент);
  - полезное пространство сегмента – 7,45 Тбайт (78% storage efficiency);
- дублированные RAID-контроллеры (Dual-Redundant Active-Active Controllers);
- отказоустойчивые подключения дисковых полок (Redundant 3Gbit SAS Connect).

В составе моделей X9000 используются т.н. Midline SAS-диски, интегрирующие лучшие особенности SAS- и SATA-дисков: технологию дисковых пластин SATA-дисков, а электронику SAS-дисков (рис. 5). Это позволяет:

- обеспечить высокую емкость;



Рис. 5. Диски SAS Midline позволяют соединить лучшие свойства SAS- и SATA-дисков.

- поддерживать двухканальный SCSI-интерфейс;
- использовать алгоритмы контроля ошибок;
- иметь низкую стоимость за терабайт;
- использовать средний режим нагрузки;
- иметь совместимость с SAS/SATA-стандартами.

Максимальная производительность одного Capacity Block – 5,4 Гбайт/с при последовательном чтении и 2,6 Гбайт/с при последовательной записи.

Основным преимуществом NAS-систем нового поколения – X9000 в сравнении с традиционными решениями является автоматическое поддержание максимальной производительности нагрузки по всем компонентам СХД (как по емкости, так и по производительности). Одновременно автоматически

оптимизируется (по показателю \$/Гбайт) хранение данных между различными уровнями, которых может быть 3.

### Функциональные особенности NAS-систем HP серии X9000

Что делает NAS-системы под управлением IBRIX Fusion File System привлекательными? Прежде всего два обстоятельства: во-первых, максимизация использования СХД за счет балансировки нагрузки на всех уровнях, что дает возможность практически линейно наращивать производительность (в рамках фиксированной конфигурации) по мере роста нагрузки, вплоть до ее насыщения; во-вторых, минимизация затрат на СХД как капитальных (за счет низкой удельной стоимости хранения), так и операционных (за счет высокого уровня автоматизации всех файловых сервисов).

#### Равномерное распределение файлов одной директории по всем сегментам кластера

Благодаря патентованной технологии все файлы одной директории/поддиректории равномерно распределяются по всем сегментам (по сути, физическим дискам) с учетом равномерного их использования по емкости. При этом полностью устраняются недостатки иерархичной файловой системы, т.е. когда возникает “перекос” по интенсивности обращения к определенным дисковым группам из-за того, что одна директория перегружена запросами. В системах X9000 “узкого места” не возникает (рис. 6), используется неиерархичная файловая система (Non-hierarchical architecture). Существует несколько алгоритмов, позволяющих управлять размещением файлов на сегментах, все они основаны на т.н. политиках (Allocation Policy). Есть следующие политики:

- *Random* – произвольно выбирается любой сегмент;
- *Round Robin* – сегменты выбираются циклично;
- *Directory* – все файлы пишутся в тот же сегмент, что и название директории;
- *Automatic* – выбирается сегмент, с наибольшим свободным пространством;
- *Sticky* – используется один определенный сегмент;
- *Local* – выбираются в первую очередь сегменты, принадлежащие серверу сегментов, который получил запрос клиента.

Для различных приложений, в зависимости от характера нагрузки и типа файлов могут применяться разные политики. Например, политика Random подходит для работы с большим количеством мелких файлов, когда

есть достаточно большое количество сегментов. По умолчанию используется политика Random. Управляет политиками т.н. Ibrix Distribution Engine (IDE), который работает на каждом сервере сегментов, обслуживающим запросы пользователей.

Таким образом, в отличие от традиционных NAS-систем, существует возможность избежать узких мест в результате обращений только к одной дисковой группе и задействовать все доступные дисковые ресурсы для размещения файлов.

Еще одним преимуществом архитектуры Ibrix Fusion является возможность прозрачного расширения файловой системы без остановки ввода-вывода, путем добавления новых сегментов или Capacity Block.

Когда добавляются новый Capacity Block автоматически запускается политика и нагрузка также выравнивается с учетом нового блока. Уровень утилизации всех сегментов, формирующих файловую систему, поддерживается одинаковым. При добавлении новых сегментов происходит процесс балансировки (Rebalancer), позволяющий динамически перераспределить файлы на вновь добавленные сегменты и тем самым уменьшить нагрузку на уже существующие сегменты.

#### Оптимизация затрат на хранение файлов

С целью минимизации затрат на хранение файлов в рамках группы сегментов под управлением одного сервера сегментов/узла возможна организация сегментов на разных дисках (SSD/SAS/SAS Midline/FC/SATA) и типах RAID (рис. 7). При этом перемещать файлы с одного уровня/сегмента хранения на другой можно, руководствуясь характеристиками файла или его атрибутами. Политики для перемещения файлов выполняются по расписанию – незаметно для пользователей.

В качестве параметров при формировании политик для перемещения файлов могут выступать:

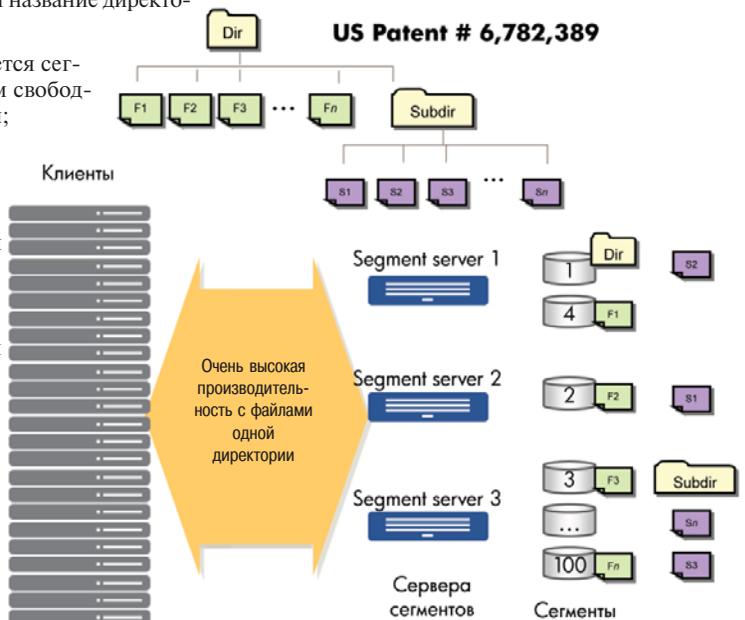


Рис. 6. Благодаря патентованной технологии все файлы директорий/поддиректорий равномерно “размазываются” по сегментам.

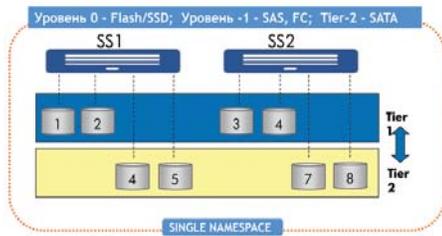


Рис. 7. X9000 дает возможность перемещать файлы с одного уровня/сегмента хранения на другой внутри сервера сегмента, руководствуясь характеристиками файла или его атрибутами.

- время создания;
  - время изменения файла;
  - размер файла;
  - имя файла или его тип (jpg, wav, exe и др.);
  - владелец файла.
- Можно совмещать несколько параметров при формировании политики.

#### Непрерывная параллельная удаленная репликация

- Удаленная репликация файлов между двумя кластерами позволяет:
- обеспечить быстрый доступ к файлам в случае катастрофы одного из кластеров;
  - использовать второй кластер параллельно с первым для повышения доступности файлов/данных;
  - осуществлять постоянное отслеживание и копирование изменений из локальной файловой системы на удаленную (требует начальной синхронизации

ции обоих файловых систем). При этом репликация происходит параллельно между серверами сегментов каждой файловой системы.

#### Оптимизация моделей серии X9000 по вариантам использования

В настоящее время в составе линейки X9000 представлены 3 варианта моделей, оптимизированные для различных вариантов использования (рис. 8).

**X9320:** модели, оптимизированные для высокой производительности или повышенной емкости.

**X9720:** модели, имеющие самую большую плотность хранения данных (12ТВ/1U высоты шкафа) и самую низкую удельную стоимость — менее \$2/GB.

**X9300:** шлюз для подключения внешних массивов. Позволяет создавать гибкое решение для построения NAS-кластера с уже существующими сетями хранения SAN (StorageWorks MSA, EVA, P4000 и др.).



Рис. 8. Начальные конфигурации HP StorageWorks X9000, оптимизированные по емкости и производительности.

#### Заключение

На сегодняшний день портфель файловых СХД (NAS) HP включает в себя решения как для малого и среднего бизнеса, так и решения масштаба предприятия, способные предоставлять пользователям доступ к ресурсам с единым пространством имен объемом до 16 Пбайт. Решения могут использоваться для консолидации удаленных офисов, создания высокопроизводительных, высокоемких, высокодоступных (НА) и прозрачно расширяемых для работающих приложений хранилищ неструктурированных данных. Благодаря уникальной параллельной сегментированной файловой системе, используемой в решениях серии X9000, возможно горизонтально масштабировать как производительность, так и емкость систем без прерывания доступа клиентов. Также немаловажно отметить высокие скорости доступа к ресурсам при работе клиентов с большим количеством мелких файлов благодаря запатентованным технологиям размещения файлов и наличию фирменной технологии доступа Irbix Client. Расширенные функции, такие, как многоуровневое хранение (Data Tiering), удаленная репликация, мгновенные снимки, позволяют в полной мере закрыть потребности предприятий по управлению неструктурированными данными, снизить стоимость хранения и обеспечить единую отказоустойчивую среду доступа и работы с неструктурированными данными.

Александр Грубин,  
HP Россия & СНГ

## HP: новые ленточные решения и СХД

Март 2010 г. — Компания HP представила ленточные решения нового поколения на базе технологии LTO-5 Ultrium (5,25" картридж, емкость — 3 Тбайт с компрессией 2:1, скорость передачи — 1 Тбайт/ч, интерфейс — 6 Гбит/с SAS или 8 Гбит/с FC, встроенное шифрование — AES 256 бит).

“Ленточные решения HP StorageWorks LTO-5 Ultrium выгодно отличаются от предыдущего поколения почти двойным увеличением емкости накопителей и расширенным функционалом защиты информации — теперь на HP LTO-5 можно с меньшими затратами хранить больший объем данных, параллельно обеспечивая их безопасность”, — прокомментировал Константин Паршин, руководитель отдела Систем хранения данных HP в России.

Ленточный накопитель HP StorageWorks LTO-5, использующий картриджи HP LTO-5 Ultrium, предоставляет следующие преимущества:

- повышение эффективности деятельности компании за счет ускорения резервного копирования и извлечения данных — до 3 Тбайт информации менее чем за три часа. По сравнению с накопителями предыдущих поколений производительность копирования удваивается, а количество требуемых носителей сокращается на 25% по

сравнению с ленточными накопителями половинной высоты HP StorageWorks LTO-3;

- экономия затрат на аппаратные средства для хранения информации. При использовании HP LTO-5 Ultrium хранение 1 Гбайт данных обойдется на 52% дешевле (в \$) по сравнению с ранее выпускавшимися накопителями HP StorageWorks LTO-3, а в случае применения ленточных библиотек HP MSL LTO-5 экономия составит 36% (по ср. с LTO-3);
- повышение отдачи от ранее сделанных инвестиций благодаря широкому возможностям интеграции в различную среду при необходимости повышения емкости или эффективности хранения данных. Накопитель LTO-5 совместим с накопителями LTO-4 (считывание и запись) и LTO-3 (считывание);
- новая утилита HP TapeAssure обеспечивает проактивный мониторинг состояния, производительности, степени использования и исправности накопителей, а также средств резервного копирования. ПО, доступное для бесплатного скачивания, заблаговременно выявляет потенциальные сбои и гарантирует непрерывную доступность и безопасность информации;
- упрощение резервного копирования и восстановления данных с помощью ленточных библиотек HP StorageWorks MSL, которые осуществляют автоматическое резервное копирование и консолидацию накопителей посредством организации легко управляемой сети;

- повышение надежности данных с помощью функции HP Data Rate Matching (DRM), синхронизирующей скорость работы накопителя со скоростью хост-системы, что обеспечивает максимальную производительность потоковой передачи данных;
- снижение износа ленточных носителей путем оптимизации перемещения данных по магнитной ленте, чтобы добиться должного уровня эффективности резервного копирования при интенсивных рабочих циклах.

Ранее HP представила новое поколение СХД HP StorageWorks P2000 G3 Modular Smart Array (MSA) и HP StorageWorks P4000 G2 SAN.

Среди новых опций MSA G3:

- улучшенные гибкость и управление данными благодаря комбинированному контроллеру FC/iSCSI Combo с портами FC 8 Гбит/с и iSCSI 1 Гбит/с. При этом данные могут передаваться по сети FC SAN через порты 8 Гбит/с, а в то же время порты iSCSI могут использоваться для репликации или подключения в iSCSI SAN для тех отделов, которым не нужна высокая производительность;
- репликация на основе Remote Snap. Этот функционал позволяет создавать “мгновенные снимки”, перемещая данные на второй массив P2000 G3 для восстановления после сбоев. Новое ПО Volume Copy автоматически создает локальные копии данных, а уже включенная в состав решения, ли-