

SATA/SAS

СИСТЕМЫ ХРАНЕНИЯ

— состояние и особенности организации

С середины 2004 г. начнутся массовые поставки продуктов и решений на основе последовательных интерфейсов. Этим процессом будут охвачены все компоненты, продукты и решения, ориентированные на массовый и средний рынок. В частности, речь идет не только об интерфейсах SATA и SAS – для систем хранения, но и о реализации системных блоков на основе PCI Express и всего семейства продуктов с использованием последовательных интерфейсов в поддержку этой технологии (SN № 2/16, 2003). С кратким представлением особенностей построения “серийных” систем хранения – наш обозреватель.

Введение

Эволюция развития HDD выделила 3 класса дисков с ориентацией на свое нишевое применение: Mobile, Desktop, Enterprise (см. таблицу и рис. 1). ATA-интерфейсы, соответственно, занимали преобладающее положение в первых двух группах, а SCSI-интерфейсы – в последней, где каждая группа имела свои специфические требования. Для дисков, используемых в настольных/переносных компьютерах, определяющими были характеристики: цена/производительность, стоимость за один мегабайт, простота использования; для SCSI-накопителей (серверы и рабочие станции) – готовность, производительность, расширяемость.

Повышение требований к IT-системам в части масштабируемости, пропускной способности, гибкости и одновременно экономичности систем хранения (прежде всего, в секторе SMB) заставило разработчи-

Характеристики	Mobile	Desktop	Enterprise
групп	3600, 4200, 5400	5400, 7200	10К, 15К
Ср. время поиска	12-14 мс	8,9-9,5 мс	3,2-7,4 мс
Производительность	N/A	79-136	146-366
Кэш (Мбайт)	2	2-8	2-8
Емкость (Гбайт)	10-80	40-250	18,36,72,144,180
Надежность	300К час MTBF	500К час MTBF	1,2М час MTBF
Потр. мощн. (W)	2,5	10	15
Цена (\$)	73-160	75-240	160-1400
Интерфейс	ATA/66, ATA/100	ATA/100, ATA/133	Ultra 160 SCSI, Ultra 320 SCSI, FC

Данные на середину 2002



Рис. 1. Изменение диаметра дисков в зависимости от скорости вращения.

ков во многом изменить отношение по жесткому разделению систем хранения на базе HDD двух классов, а также обеспечить большую архитектурную гибкость и расширение функциональности как ATA-, так и SCSI-дисков, при возможности совместного использования HDD двух классов в одной системе.

Накопители с параллельными интерфейсами PATA и PSCSI производятся с 80-х годов и за это время выдержали целую серию усовершенствований, но дальнейшее развитие данных интерфейсов с конца 90-х уже сдерживалось особенностями их физической организации. Можно выделить следующие причины, послужившие основанием для разработки нового ATA-стандарта:

- малое количество устройств (до двух), подключаемых на один контроллер;
- короткий интерфейсный кабель (< 0,5 м);
- низкая пропускная способность (< 133 Мбит/с);
- ненадежный, неэкранированный 80-проводный плоский кабель;
- отсутствие возможности “горячего подключения”;
- невозможность гибкого масштабирования системы.

Результатом совместных усилий явилось опубликование в 2001 г. стандарта Serial ATA 1.0, обеспечивающего поддержку совместимости с существующими драйверами для ATA, долговременную (> 10 лет) совместимость будущих поколений SATA, а также простоту и доступность, аналогичных существующим PATA. Уже первая версия SATA – Serial ATA 1.0 обеспечи-

- возможность горячего подключения (разноуровневые контакты);
- повышенную надежность (32-bit CRC для всех передач данных);
- First Party DMA (предоставление прямого доступа к памяти хоста);
- управление питанием (частичное – уменьшение энергопотребления; спящий режим – отключение PLL).

Эволюция параллельных SCSI-интерфейсов

Поколение	Стандарт	Год	Скорость	Ключевые особ.
SASI		1979		
SCSI-1	SCSI-1	1986	~ 2 MB/sec	Asynchronous, narrow
SCSI-2	SCSI-2	1989	10 MB/sec	Synchronous, wide
SCSI-3	Split command sets, transport protocols, and physical interfaces into separate standards			
Fast- Wide	SPI/ SIP	1992	20 MB/sec	
Ultra	Fast- 20 annex	1995	40 MB/sec	
Ultra 2	SPI-2	1997	80 MB/sec	LVD
Ultra 3	SPI-3	1999	160 MB/sec	DT, CRC
Ultra 320	SPI-4	2001	320 MB/sec	Paced, Packetized, QAS

Эволюция параллельных ATA-интерфейсов

Поколение	Стандарт	Год	Скорость	Ключевые особ.
IDE		1986		Pre- standard
	ATA	1994		PIO modes 0- 2, multivord DMA 0
EIDE	ATA-2	1996	16 MB/sec	PIO modes 3- 4, multivord DMA modes 1- 2, LBAs
	ATA-3	1997	16 MB/sec	SMART
	ATA/ ATAPI-4	1998	33 MB/sec	Ultra DMA modes 0- 2, CRC, overlap, queuing, 80- wire
Ultra DMA 66	ATA/ ATAPI-5	2000	66 MB/sec	Ultra DMA mode 3-4
Ultra DMA 100	ATA/ ATAPI- 6	2002	100 MB/sec	Ultra DMA mode 5, 48- bit LBA
Ultra DMA 133	ATA/ ATAPI-7	2003	133 MB/sec	Ultra DMA mode 6

Эволюция последовательных SATA-интерфейсов

Поколение	Стандарт	Год	Скорость	Ключевые особ.
Serial ATA	ATA/ ATAPI-7	2002	150 MB/sec	
Serial ATA II	ATA/ ATAPI-8?	2003	300 MB/sec	Enhanced queuing
Serial ATA III	ATA/ ATAPI-9?	?	600 MB/sec	

Стратегия развития продуктов с SAS-интерфейсом

Поколение	Стандарт	Год	Скорость	Ключевые особ.
SAS 300	SAS 300	2004	300 MB/sec	
SAS 600	SAS 600	2007	600 MB/sec	
SAS 1200	SAS 1200	2009	1200 MB/sec	

Интерфейс SCSI, изначально ориентированный для использования в составе серверов и рабочих групп, также испытывает значительное влияние как со стороны ATA-интерфейсов (и ATA-систем), так и со стороны FC-интерфейса. Занимая промежуточное положение между этими нишевыми группами, его развитие требует сохранения преемственности и ценового диапазона с параллельным семейством интерфейсов и одновременным расширением функциональности и скоростных характеристик в сторону FC-систем. Основные изначальные требования при его разработке следующие:

- реализация поддержки устройств SATA с целью обеспечения большей гибкости и экономичности при построении систем хранения;
- поддержка драйверов протокола SCSI и его ценовой группы;
- возможность многократного увеличения существующей (320 Мбайт/с) скорости интерфейса;
- конструктивная реализация меньших и более простых соединений;
- поддержка двухпортовости с целью обеспечения большей доступности;
- значительно более (> 16) высокая связность;
- лучшая масштабируемость.

Архитектурные особенности SATA/SAS систем хранения

Итак, какие преимущества предлагают новые “серийные” технологии. Во-первых, это возможность строить более экономичные системы хранения с расширенным диапазоном характеристик, чем существующие (рис. 2), за счет возможности объединения

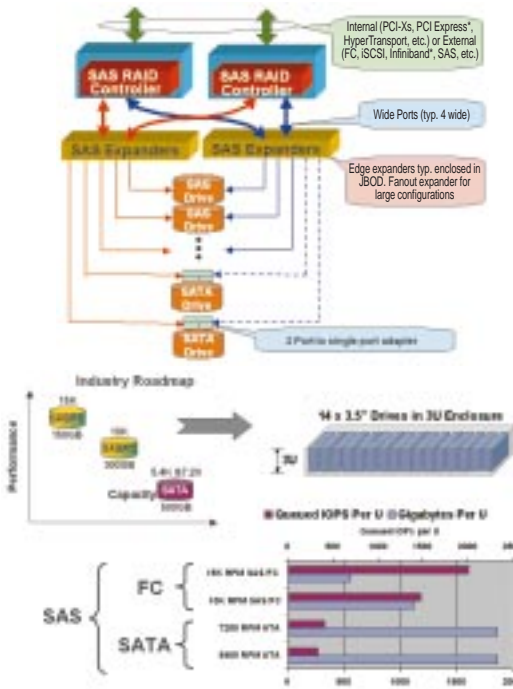


Рис. 2. Возможность объединения в одном конструктиве SATA- и SAS-дисков дает возможность гораздо более широкого варьирования показателей емкости и производительности системы хранения.

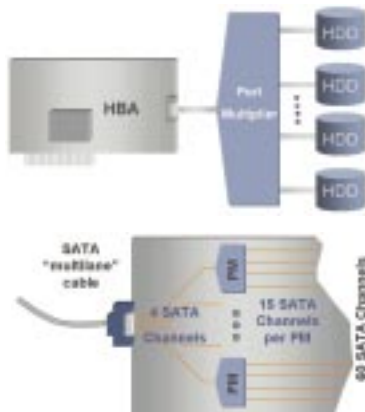


Рис. 3. За счет расширителя портов к одному активному хосту можно подключить до 15 дисков SATA, а за счет кабеля и разъема SATA II “Bulkhead” увеличить число HDD до 60.

в одном конструктивном модуле (на одном разъеме или на одной панели) как SATA, так и SAS-дисков.

Во-вторых, за счет расширителя портов на один “линк” (один активный хост) можно “повесить” до 15 дисков SATA. В дополнение к этому, за счет использования кабеля и разъема SATA II “Bulkhead” (каждый кабель поддерживает 4 канала SATA) совместно с расширителем портов очень просто увеличить число накопителей в подсистеме (через “один кабель”) до 60 (рис. 3).

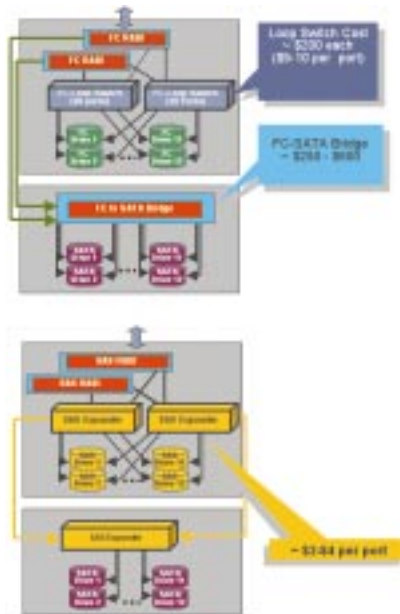


Рис. 4. Пример вариантов организации в одном конструктиве различных по доступности подсистем хранения.

В-третьих, в SAS-дисках предусматривается встроенная двухпортовость, а для SATA-дисков предусматривается условная двухпортовость за счет интеграции простого коммутатора в устройства для подключения дисковых накопителей (а НЕ в сами накопители), который управляется протоколом или внешним сигналом. Это дает возможность повысить надежность системы в случае отказа одного из хостов.

В-четвертых, SAS-SATA-системы хранения дают возможность поддерживать в одном конструктивном модуле приложения различной степени доступности (рис. 4): от on-line до mid-line & near-line (при этом поддержка SATA-протокола в SAS устраняет

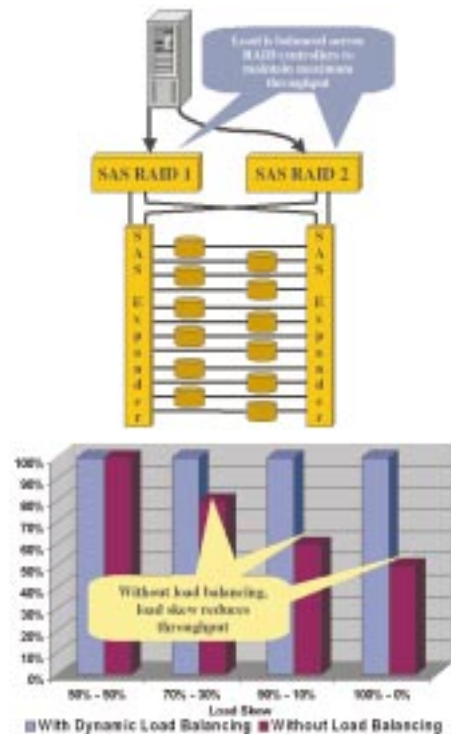


Рис. 5. Автоматическая балансировка нагрузки в 2-канальных SAS-системах обеспечивает повышение общей производительности до 50%.

ет потребность в соответствующих чипах моста).

В-пятых, SAS-системы поддерживают автоматическую балансировку нагрузки с двухвходными системами (рис. 5). Сейчас это поддерживается такими средствами, как: EMC PowerPath, HP SANworks Secure Path, Hitachi Dynamic Link Manager, IBM Data Path Optimizer. Из-за перекоса по отдельным каналам снижение общей пропускной способности может достигать 50%.

В-шестых, за счет архитектурной гибкости SATA-SAS и SATA/SAS системы обеспечивают реализацию систем всех уровней – DAS/NAS/SAN, от SMB до enterprise (рис. 6).

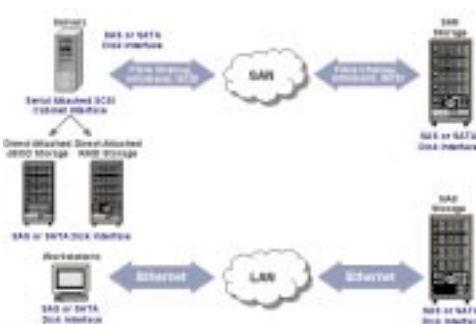


Рис. 6. SATA-SAS и SATA/SAS-системы за счет архитектурной гибкости обеспечивают реализацию систем всех уровней – DAS/NAS/SAN, от SMB до enterprise.

Состояние, перспективы SATA/SAS систем хранения

Продажи SATA уже начались, а SAS-системы начнут поставляться с середины с.г. (рис. 7), но в полной мере воспользоваться преимуществами организации систем на новых интерфейсах можно только с конца с.г., когда для SATA-дисков будет: 1) реализова-



Рис. 7. Ближайшая перспектива SAS-систем.

на скорость 3 Гбит/с (обеспечение совместимости с SAS); 2) реализована возможность очередизации команд (повышение пропускной способности и доступности), а в SAS будет: 1) поддерживаться двухпортовость; 2) обеспечиваться совместимость по соединениям с SATA-дисками.

Наиболее активно системы с новыми интерфейсами продвигает компания Maxtor. И по ее прогнозам, к 2008 г. в секторе корпоративных систем хранения FC-диски будут занимать не более 20% рынка (рис. 8).

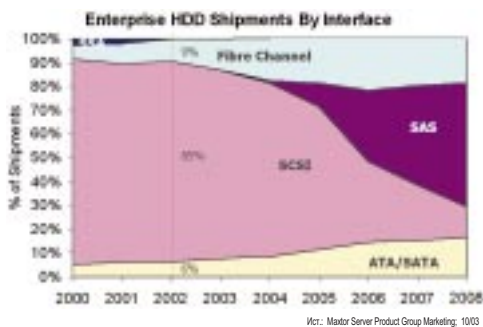


Рис. 8. Долгосрочный прогноз продаж HDD по интерфейсам систем хранения масштаба предприятия.

Производительность систем хранения с SATA-дисками

Вследствие того, что SAS-системы будут активно появляться на рынке только с III кв. с.г., то сравнение по производительности возможно лишь с SATA-системами.

Приложения по производительности классифицируются как комбинация случайных



Рис. 9. Сравнение производительностей SATA- и FC-дисков на потоковых и случайных операциях ввода/вывода.

или/и последовательных обращений к данным. Случайный доступ к данным измеряется в операциях ввода/вывода в секунду (IOPS) и в основном определяет работу транзакционных приложений типа OLTP и баз данных. Последовательный доступ к данным измеряется в мегабайтах в секунду (Мбайт/с) и критичен для приложений с высокой интенсивностью потоковых операций (например, медиа-приложения, обработка сейсмоданных и др.).

SATA- и Fibre Channel-диски имеют очень различные профили по производительности. Fibre Channel-диски разработаны для высоконагруженных приложений при одновременной обработке множества запросов к дискам, подключенным к общему интерфейсу с минимальными задержками.

SATA-диски – более медленные и SATA-интерфейс не предусматривает параллельную работу нескольких накопителей одновременно на общем интерфейсе. Поэтому для большинства критичных транзакционных приложений они имеют ограниченное применение, в то время как для потоковых применений (при достаточно большом числе накопителей) они приближаются к FC-системам (рис. 9, по данным IBM).

Потоковая производительность критична для приложений с большими блоками последовательных операций ввода/вывода. Определяющими требованиями систем хранения для таких приложений являются: максимальная потоковая производительность интерфейса и накопителей, максимальный размер блока передаваемых данных и очередизация команд. И хотя Fibre Channel-диски имеют преимущество над характеристиками SATA-приводов, оно не столь ощутимо и позволяет достичь максимальной производительности контроллера. При небольших и средних конфигурациях обычно требуется удвоенное количество SATA-дисков, чтобы достичь производительности FC-дисков. Например, в системе FASiT600 выравнива-

	FASiT900	FASiT600T
Burst I/O rate "cache reads (512 byte)	148,000 IOPS	77,500 IOPS
Sustained I/O rate "disk reads (4k)	53,200 IOPS	25,000 IOPS
Sustained I/O rate "disk writes (4k)	17,000 IOPS**	6,000 IOPS**
Sustained throughput "cache read (512k)	10,900 IOPS	5,200 IOPS
Sustained throughput "disk read (512k)	3,800 IOPS**	1,250 IOPS**
Sustained throughput "disk write (512k) (cache mirroring off)	800 MB/sec	800 MB/sec
Sustained throughput "disk read (512k)	800 MB/sec	800 MB/sec
Sustained throughput "disk write (512k)	795 MB/sec	400 MB/sec
Sustained throughput "disk write (512k) (cache mirroring on)	780 MB/sec	320 MB/sec**
Sustained throughput "disk write (512k)	625 MB/sec	300 MB/sec
Sustained throughput "disk write (512k)	600 MB/sec	250 MB/sec**
Sustained throughput "disk write (512k)	305 MB/sec	159 MB/sec
Sustained throughput "disk write (512k)	300 MB/sec	130 MB/sec**

ние производительностей происходит при 28 SATA-дисках, для FASiT900 – при 56 SATA-дисках. Сравнение двух систем – FASiT900 и FASiT600 в комплектации с FC и SATA-дисками (синим цветом) приведено в таблице (по данным IBM).

Стоимость систем хранения с SATA-дисками

В общей стоимости системы хранения диски занимают долю от 30% до 80%. Использование SATA-дисков, которые в пересчете на единицу хранения стоят от 1/3 до 1/2 стоимости FC-накопителей, позволяет снизить общую стоимость систем хранения от

Емкость системы хранения	Экономия при комплектации SATA-дисками вместо FC-дисков
3 Тбайт	29 - 42%
5 Тбайт	34 - 45%
10 Тбайт	44 - 52%
20 Тбайт	50%+

30% до 50%. В таблице (по данным IBM) приводится возможная экономия на системах хранения при комплектации их SATA-накопителями в зависимости от емкости системы.

Сроки поставок SAS/SAS-SATA систем хранения

Поставки SATA-систем уже идут, SAS/SAS-SATA-системы должны появиться в начале

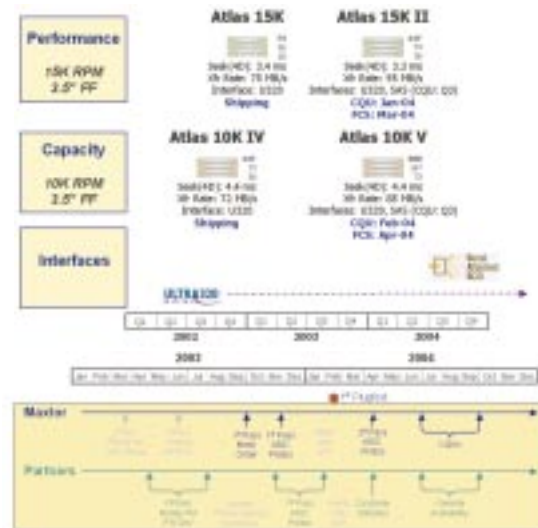


Рис. 10. Roadmap продуктовых линеек Maxtor, в которой будет развиваться SAS-стандарт.

года с.г., а активные их продажи ориентировочно планируются на август. Roadmap продуктовых линеек одного из активных "лоббистов" последовательных интерфейсов компании Maxtor – на рис. 10.

Дополнительная информация: Международный комитет по стандартам информационных технологий (International Committee for Information Technology) – www.incits.org; рабочие SCSI-стандарты (T10) – www.t10.org; SAS-соединения (SFF (соединения) – www.sffcommittee.org; рабочие ATA-стандарты (T13) – www.t13.org; Serial ATA II спецификации доступны на www.t10.org и www.serialata.org; Fibre Channel стандарты (T11) – www.t11.org; SCSI Trade Association – www.scsita.org; Serial ATA Working Group – www.serialata.org; SNIA (Storage Networking Industry Association) – www.snia.org; отраслевые новости – www.infostor.com, www.byteandswitch.com, www.wvpi.com, <http://searchstorage.com>.