SAN-виртуализация – тенденции, проекты, мнения

Резкий всплеск интереса к решениям на уровне SAN обусловлен заявлениями ведущих вендоров, поставляющих SAN-коммутаторы, о завершении проектирования разработки т.н. интеллектуальных коммутаторов со значительно более широкими возможностями по управлению томами и данными в SAN-среде, а также объявлениями рядом мелких фирм уже о начале тестирования и поставок таких продуктов. К концу с.г. ожидаются анонсы о продажах от крупных вендоров. Данная публикация – продолжение темы, начатой в двух предыдущих номерах SN. С кратким обзором основных тенденций и состояния рынка с акцентом на разработки Brocade – наш специальный обозреватель.

Введение

Лидерами процесса SAN-виртуализации (или, проще, увеличение функциональности SAN-коммутаторов) являются компании Brocade и Cisco, одними из первых объявившие о своих планах и основных положениях проектов. Все ведущие игроки на рынке систем хранения (EMC, IBM, Sun Microsystems, HP), кроме Hitachi Data Systems, а также основные разработчики ПО управления в среде SAN, обозначили свое отношение к данному "процессу" и уже сделали ряд объявлений по проводимым разработкам. Особо хочется отметить Cisco, которая только в этом году вышла на рынок со своими FC коммутаторами, но уже к июлю с.г. они были сертифицированы и включены в листы поставок всеми ведущими компаниями high-end систем хранения.

EMC анонсировала свои партнерские взаимоотношения с Cisco и Brocade по совместной разработке интеллектуальных коммутаторов. При этом EMC активно внедряет мультипротокольные адаптеры Cisco в свои продуктовые линейки и ведет с ней проект по развитию iSCSI до уровня FC.

HP сотрудничает с Brocade и Cisco, основное внимание в большей степени уделяя первой, о чем было заявлено еще в начале года. Объединенное решение HP и Brocade (HP VersaStor технология с Brocade SilkWorm Fabric Application Platform) расширит существующую функциональность CASA (HP OpenView Continuous Access Storage Appliance), обеспечивая интеллектуальное перераспределение сетевого трафика в соответствии с потребностями пользователей по доступу к данным. Кроме того, это объединенное решение будет "бесшовно" интегрироваться с существующим HP и Brocade SAN оборудованием, а также с HP OpenView SAM (Storage Area Management) ПО и, как ожидается, начнет поставляться во второй половине с.г.

IBM в вопросе SAN-виртуализации делает основной упор на собственные разработки: TotalStorage SAN Volume Controller (SN № 1/15-2/16, 2003) и TotalStorage SAN File System. В контексте данной публикации к SAN виртуализации имеет отношение в основном только SAN Volume Controller, отгрузки которого начались уже с июля с.г.

Sun Microsystems также самостоятельно развивает SAN-виртуализацию на основе коммутаторов компании Pirus Networks, приобретенной ею в сентябре 2002 г.

Hitachi Data Systems (HDS) пока не обнародовала собственную стратегию развития в этой области.

Как сообщает ByteandSwitch, Cisco и VERITAS после длительных обсуждений выбрали в качестве прототипа совместно реализуемого проекта по SAN-виртуализации Veritas SAN Volume Manager (SAN VM), работающий на модуле виртуализации Cisco MDS 9500. Новый продукт после завершения бета-тестирования к концу лета с.г., как планируется, начнут продавать во второй половине с.г.

McData Corp.заявила о готовности поставлять продукты, аналогичные развиваемым Brocade и Cisco в 2004 г. Между тем ряд более мелких фирм, таких как Sanera Systems Inc., MaXXan Systems Inc. и ряд других, заявили об интересе к этому сектору рынка и уже вышли на него со своими оригинальными разработками.

Данная публикации – продолжение темы, начатой в двух предыдущих номерах SN, с более подробным описанием проекта, развиваемого компанией Brocade – признанного классика SAN-FC-коммутаторов.

Зачем нужна SAN-виртуализация

Появление сетевых хранилищ было своего рода водоразделом, позволившим снять физическую и логическую зависимость данных от серверов и приложений и выделить хранение данных в отдельную компоненту IT-инфраструктуры, прежде всего, как SAN. Отделение данных позволило свободно разделять их в корпоративной среде между приложениями, избегая многочисленного дублирования и обеспечивая централизованное управление, что, в свою очередь, повысило их доступность и надежность хранения.

Но одновременно с увеличением требований целостности, безопасности, доступности растет и сложность SAN. Кроме того, т.к. серверы и выполняющиеся на них приложения могут решать совершенно разные бизнес-задачи, требуется обеспечение различного доступа и типа хранения. При этом сложность повышается с увеличением количества и размера пулов данных, требующихся для поддержки прикладных серверов, а также с увеличением уровня гетерогенности SAN.

Чтобы уменьшить эти сложности, разработчики средств управления хранением используют методы виртуализации: трансляцию реального адреса хранения данных в логические номера устройств – LUN (Logical Unit Number), которые передаются прикладным серверам как специализиро-

ванные динамические пулы хранения. Реальная память, которая представляется этими LUN берется из одного большого физического storage-пула, который управляется SAN-менеджером.

Виртуализация на уровне коммутатора (SAN-виртуализация) создает непрерывные пулы гетерогенной памяти, которые могут централизованно управляться, значительно повышая эффективность управления и использование ресурсов (человеческих и аппаратных). Требуемая память может выделяться приложениям автоматически, без нарушения их работы и небольшими порциями.

Уровень SAN-виртуализации также устраняет барьеры для перемещения данных. Поскольку возможно объединение совершенно различных контроллеров массивов хранения, данные могут копироваться или перемещаться к/от совершенно различных запоминающих устройств. Это имеет большое значение для автоматизированного хранения и управления данными приложений, в частности, при резервном копировании/восстановлении, организации катастрофоустойчивых решений, иерархическом хранении и управлении жизненным циклом данных.

Проект Brocade

О начале проекта разработки интеллектуальных коммутаторов Brocade объявила более года назад, но конкретные шаги по его реализации стали понятны только осенью прошлого года после приобретения ею компании Rhapsody Networks (завершение сделки состоялось в начале с.г.).

Официально заявлено о восьми партнерах, с которыми Brocade подписала соглашение о поддержке ими платформы Rhapsody средствами развиваемого ими ПО управления хранением: Alacritus Software Inc., CommVault Systems Inc., FalconStor Software Inc., Incipient Inc., InterSAN Inc., StoreAge Networking Technologies Ltd. и Topio. К этому альянсу в начале мая с.г. присоединилась VERITAS, которая будет развивать volume management и storage resource management ПО технологии для новой платформы Brocade на базе Brocade XPath API интерфейса, что крайне важно для Brocade, т.к. VERITAS контролирует около 70% рынка этого сектора ПО.

В основе новой платформы – Brocade SilkWorm Application Platform (SilkWorm AP) – лежит XPath технология (развиваемая ранее Rhapsody Networks), при разработке которой преследовались следующие цели:

- использование распределенных и параллельных методов, чтобы снять ограничения в производительности и масштабируемости при обработке потоков данных, препятствующих развитию механизмов виртуализации на уровне коммутаторов;
- создание платформы, не зависящей от протоколов и поддерживающей как Fibre Channel, так и IP (FC-IP и iSCSI), одновременно обеспечивающей динамическое реконфигурирование возможностей

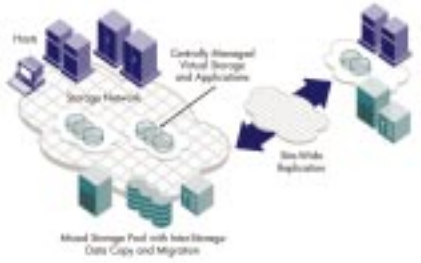

Рис. 1. Основные преимущества SAN, достигаемые при внедрении XPath технологии.

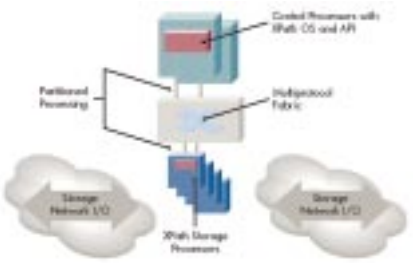

Рис. 2. Основные компоненты XPath технологии.

порта через графический интерфейс на основе браузера;

- обеспечение доступности OEM-партнерам и независимым разработчикам приложений использования возможностей XPath технологии через API интерфейс и соответствующий инструментарий к нему.

Перемещая функции хост-storage-приложений непосредственно внутрь SAN-коммутатора, XPath технология дает возможность одному приложению или нескольким его копиям разделять множество SAN-связанных хост- и storage-систем. Эта объединенная модель развертывания уменьшает затраты управления при расширении функциональных возможностей и гибкости приложений в противовес существующим подходам, при которых эти операции приводили к техническим и экономическим трудностям, снижавшим эффективность их использования.

Например, “in-band” устройства, использующие стандартные компьютерные платформы, не масштабируются эффективно, потому что они требуют универсального сервера, чтобы обрабатывать каждый поток данных “in-band”.

Точно так же, “out-of-band” устройства распределяют основные storage-функции программным агентам на клиентских HBA или драйверам операционных систем хостов, с тем чтобы избежать узкого места, связанного с отдельным путем данных. Однако такие функции, как совместное использование тома (storage volume sharing), реплицирование и перемещение данных, должны выполняться на off-хост-платформе с подобными ограничениями, как в случае “in-band” устройства.

Кроме того, инсталляция и обслуживание клиентских драйверов или HBA на каждом хосте вводят новый уровень управления хостом и снижают его производительность.

Целевая инфраструктура XPath технологии – следующее поколение SAN с характеристиками (рис. 1), делающими прикладные платформы не зависящими от хостов и систем хранения:

- расширенные возможности управления и эффективность использования ресурсов для большого количества серверов и систем хранения;
- значительно увеличенные возможности по репликации и перемещению данных для всей иерархии систем хранения, что существенно улучшает защиту данных и их катастрофоустойчивость;
- произвольная смесь технологий соединений, таких как Fibre Channel и Ethernet/IP, ориентированных на различную производительность, совместимость и функциональность.

Компоненты Brocade XPath технологии

XPath технология имеет 4 основных составляющих (рис. 2):

- *XPath Partitioned Processing (XPath разделенная обработка):* использование распределенного управления и data path процессоров с целью масштабирования сетевого ПО хранения;
- *XPath Storage Processors (XPath процессоры хранения):* обеспечение обработки хранимых данных на основе применения набора storage-оптимизированных аппаратных механизмов ускорения;
- *XPath Multiprotocol Fabric (XPath мультипротокольный коммутатор):* поддержка универсального протокола соединения с малым временем задержки и связывание всех компонент (“любой-к-любому”) с неблокирующей пропускной способностью;
- *XPath OS и API*: OS объединяет распределенную платформу, а API обеспечивает инструментальными средствами и услуга-

Табл.1. Модель разделенной обработки, используемая в XPath технологии.

Уровень обработки	Гибкость ПО	Производительность и масштабируемость	Метод выполнения	Пример функций
Control and Management Software Processors	очень высокая	ограниченная	Centralized general-purpose processors for control and boundary logic of applications; simplifies software adoption and systems management	• Volume configuration • I/O error recovery • Network device discovery
Data Path Software Processors	высокая	высокая	Replicated “in-line” processors scale performance logic of applications, leveraging hardware acceleration	• Virtual I/O translation • QoS functions • Emerging protocols (such as iSCSI)
Data Path Custom Hardware	None	очень высокая	Replicated custom hardware provides well-defined transport and application assist functions	• Fibre Channel link protocols • Virtual I/O translation assists

ми сторонние приложения с целью доступа к возможностям XPath технологии.

XPath-разделенная обработка

Чтобы эффективно масштабировать storage-приложения, XPath технология объединяет ключевые методы функционального разделения и распределенной обработки на множестве уровней обрабатывающих элементов. Различные типы функций распределяются в иерархии обработки (табл. 1). Например, функции storage транспортного уровня реализованы прежде всего в пользовательских аппаратных средствах (custom hardware); функции приложений – в "смеси" клиентского аппаратного обеспечения, микропрограммного обеспечения и универсального ПО; функции управления сетью хранения – в ПО, выполняемом на универсальном процессоре управления.

Разделение обработки управляющих и data path операций

Раздельная обработка разного типа операций позволяет развить основные прикладные функции нового коммутатора с точки зрения повышения его производительности и масштабируемости. Это разделение обработки изолирует наиболее часто встречающиеся и чувствительные к производительности функции и физически распределяет их группе data path процессоров, оставляя более сложные функции координации конфигурации меньшему числу централизованных управляющих процессоров.

SilkWorm AP отделяет управляющий поток от потока данных, которые обычно объединены во входном storage I/O path (рис. 3), непосредственно обрабатывая большинство data path операций внутри процессора. Data path операции, которые не могут быть успешно обработаны, направляются управляющему процессору для завершения. Data path поток содержит следующие типы команд: SCSI I/O read/write, block copy/mirror и фреймы трансляции виртуального I/O (virtualized I/O translation frames). Поток управления содержит фреймы, которые используются для: Fibre Channel entity login, SCSI операции запроса и команды конфигурирования виртуального тома (табл. 2).

Чувствительные к производительности data path операции, такие, как операции I/O по чтению/записи, формируют бóльшую часть трафика в сети (~ 95%). Относительно простая логика этих операций может быть уменьшена до функциональных примитивов (применимых ко многим типам приложений хранения), которые и оптимизируются на аппаратном уровне в data path процессорах (также называемых XPath Sto-rage процессоры).

Табл. 2. Сегментация сетевого трафика операций хранения.

	Частота и чувствительность к производительности	Функциональная сложность	Пример операций
Control Path операции	низкая	в основном высокая	• Fibre Channel entity login • SCSI inquiry operation • Virtual volume configuration
Data Path операции	высокая	в основном низкая	• SCSI I/O read/write • Block copy/mirror • Virtualized I/O translation

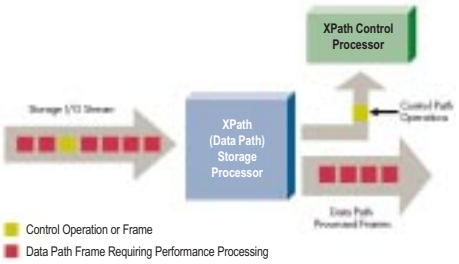

Рис. 3. Разделение в реальном времени управляющих и data path операций.

Остаток от сетевого трафика состоит из control path операций обеспечения функций конфигурирования, администрирования и ряда других. В общем потоке трафика эти операции являются относительно нечастыми (менее 5% общего трафика ввода-вывода) и могут быть обработаны меньшим числом универсальных control path процессоров. Тип управляющего потока зависит от приложений на хостах и в общем виде включает фреймы, которые могут содержать: конфигурацию тома и его размещение; команды восстановления при ошибках; метаданные файловой системы; команды обработки ошибок; дублирование и перенос данных; команды обеспечения защиты; топологию.

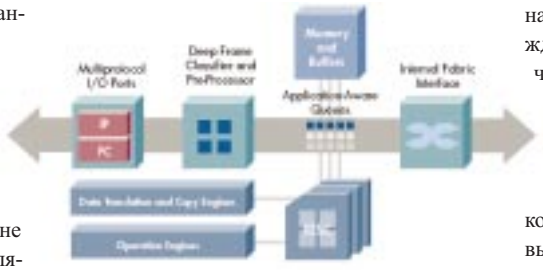

Рис. 4. Логическое представление одного XPath Storage процессора.

XPath storage процессоры

Поток данных характеризуется операциями, которые являются относительно частыми (например, SCSI операции ввода-вывода по чтению/записи), но относительно простыми. Поэтому Brocade использует однородные параллельные процессоры и заказные аппаратные средства, чтобы выполнить виртуальную трансляцию ввода-вывода, обработку протокола и функции QoS (Quality of Service – качество услуг). Архитектура SilkWorm AP распределяет data path обработку на множестве однородных функциональных блоках, называемых XPath Storage процессоры (или SilkWorm AP процессоры), которые имплементируют бóльшую часть XPath технологии, интегрируя следующие особенности (рис. 4):

- мультипротокольные с многогигабитной пропускной способностью сетевые порты I/O;
- встроенные RISC процессоры с локальной памятью и буферами фреймов на каждый порт для программной обработки потока в режиме "in-line";
- широкий набор операционных механизмов с аппаратной поддержкой для обработки фреймов с учетом особенностей приложений.

Мультипротокольное транспортное устройство

Сдвоенные блоки, поддерживающие Fibre Channel и Gigabit Ethernet транспортные протоколы в одном XPath Storage процессоре, делают возможной на одном I/O порту SilkWorm Fabric AP поддержку любого протокола и переход от одного протокола к другому, не прерывая трафик на других портах. Эта интеграция обеспечивает гибкость развертывания в будущем и снижение полной стоимости системы, в то время как эквивалентная гибкость и функциональные возможности в стандартной платформе требуют двух установленных PCI I/O интерфейсных карт с требуемой поддержкой протоколов.

Программная обработка на каждый порт

Полностью конвейерная многопроцессорная RISC-система и память соединены с каждым портом ввода-вывода с целью обеспечения встроенной возможности программной обработки, позволяющей значительно увеличить масштабируемость, производительность и гибкость в сравнении с другими традиционными архитектурами коммутаторов за счет следующего. Во-первых, параллельная обработка потоков данных в каждом сетевом порту ввода-вывода позволяет избежать возникновение конфликтов, приводя к большему масштабированию. Во-вторых, близкая интеграция процессора и совместное использование памяти с сетевым портом ввода-вывода устраняют программное управление и перегрузки при передаче данных. Напротив, устройства с отдельными основными процессорами и интерфейсами ввода-вывода могут снижать производительность при копировании данных и от перегрузок шины между HBA и основной памятью процессора.

Обработка фреймов с учетом особенностей приложений

Эти возможности значительно увеличивают производительность и эффективность data path операций за счет объединения модели обработки данных драйвером и аппаратными средствами с заранее запрограммированными сведениями об особенностях верхнего уровня хранения данных.

Конвейерная обработка фрейма начинает использоваться как только фрейм поступает в XPath Storage процессор. Эта методика увеличивает производительность ввода-вы-

Прим. – одна виртуальная операция I/O определяется как одна законченная операция, требующая передачи 512-байтного I/O запроса по чтению через сетевой обрабатывающий элемент к физической "цели".

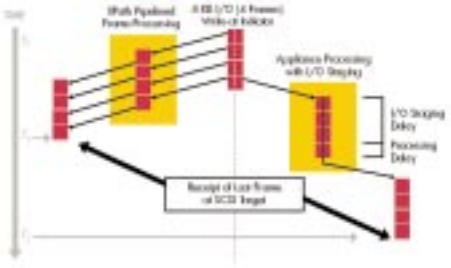

Рис. 5. Сравнение конвейерной (слева) и стандартной (справа) обработки фрейма.

вода и уменьшает время ожидания по сравнению со стандартной обработкой.

Например, на рис. 5 показано сравнение конвейерной обработки фрейма для 8 Кбайт I/O, который обработан конвейером (слева) и традиционным способом (справа), иллюстрирующее дополнительные задержки процессоров на стадии ввода-вывода, которые возрастают при увеличении размера фрейма.

"Механизмы" глубокой классификации анализируют каждый входящий фрейм данных и применяют к нему управляющие функции без выдачи прерываний операционной системе.

Результирующее влияние этих особенностей – значительное улучшение производительности и эффективности при обработке storage-приложений на XPath storage процессорах в сравнении с другими подходами. Например, XPath storage процессор добавляет менее 20 мкс общей задержки при обработке фрейма по сравнению с более чем 200 мкс на обычных коммутаторах. Точно так же виртуализация и операция реплицирования для одной SCSI операции ввода-вывода на запись SCSI требовали бы до дюжины прерываний операционной системы и других накладных затрат, что приводило бы ко многим десяткам тысяч циклов центрального процессора в сравнении с нулевым числом прерываний и несколькими сотнями программных инструкций в XPath storage процессоре.

Модификация данных и механизмы копирования

Модификация данных и механизмы копирования ускоряют обычные функции хранения, используя ряд методов трансляции данных и предотвращения копирования. Поскольку дополнительные копии данных приводят к заметным накладным затратам при многогигабайтных потоках, механизмы оптимизации копирования включены в архитектуру XPath технологии.

XPath мультипротокольный коммутатор

XPath мультипротокольный коммутатор – высоко масштабируемое протокольно-нейтральное соединение, оптимизированное для высокой полосы пропускания с низкой задержкой storage-трафика. Это соединение дает возможность данным и процессорам пути управления координировать и обменивать данные без влияния на первичный storage-трафик. Это соединение обеспечивает следующие преимущества перед существующими подходами:

- *линейная системная масштабируемость:* линейное масштабирование связности и пропускной способности делает возможным неограниченный рост сетевой обработки хранимых данных;
- *нейтральный (универсальный) протокол транспортировки:* XPath мультипротокольный коммутатор гибко транспортирует и соединяет множественные протоколы "по требованию", устраняя потребность в отдельных мостах протоколов и маршрутизаторах и упрощая мультипротокольные решения хранения типа катастрофоустойчивых решений и сетевого резервного копирования/восстановления;
- *высокая системная интеграция:* высокая интеграция коммутатора с другими подсистемами XPath технологии снижает затраты и увеличивает производительность по сравнению с кластерами устройств, построенными с разными шинами ввода-вывода и дискретной взаимосвязью компонентов.

XPath OS и API

XPath операционная система (Operating System – OS) и интерфейсы прикладного программирования (Application Programming Interfaces – API) упрощают развитие, развертывание и управление новыми и существующими приложениями хранения за счет масштабирования и преимуществ XPath технологии, главные особенности которых включают:

- *единый системный вид:* XPath OS и API объединяют множественные уровни распределенной обработки в XPath технологии в единую систему для администраторов и разработчиков программных средств, предоставляя им возможность использовать все преимущества новой технологии без необходимости понимания всех ее подробностей;
- *широкие услуги для разработки приложений хранения:* XPath API обеспечивают разработчикам программ легкий доступ к порту, развитие приложений хранения для платформ с XPath технологией, а также поддерживают прозрачное инкорпорирование будущих технологических усовершенствований. Кроме того, XPath OS обеспечивает storage-ориентированные приложения "строительными" блоками, которые экономят время на разработку и позволяют разработчикам сосредотачиваться на специализированных задачах. Например, предварительно подготовленная инфраструктура для SCSI блоковых приложений позволяет XPath OS разработчикам сосредотачиваться на инновациях, а не на проблемах протокола нижнего уровня;
- *полная коммутационная функциональность:* XPath OS позволяет SilkWorm Fabric AP участвовать непосредственно как элементу инфраструктуры сети хранения внутри Fibre Channel и IP сетевых доменов, расширяя и упрощая прикладные функции. Полнокоммутационная функциональность позволяет более гибко управлять сетевыми функциями, такими как маршрутизация и адресация для поддержки функций хранения, в частности, сетевой виртуализации.

Реализация

По сообщениям информационных агентств, поставки нового коммутатора – SilkWorm Fabric Application Platform – начнутся в 4 кв. с.г. как 16-портового блока (сначала – OEM-партнерам). Два главных элемента коммутатора: 1) центральный процессор (PowerPC), работающий под управлением NetBSD операционной системы; 2) связанные с портами XPath Storage процессоры или XSP – специализированные интегральные схемы с 3 млн элементов. В середине – перекрестный коммутатор с пропускной способностью 1 Тбит/с.

Операции в режиме split-mode позволяют центральному процессору выполнять приложения, в то время как XSP (каждый из которых может обработать 50 000 IOS) обеспечивает высокоскоростную передачу инструкций от приложений хостов на центральный процессор.

Каждый порт коммутатора может работать со скоростью 1 или 2 Гбит/с по интерфейсу FC или IP. В то же время он будет поддерживать как iSCSI, так и FCIP (Fibre Channel over IP) протоколы в зависимости от приложения, с которым работает.

Вместо заключения

Варианты виртуализации на уровне коммутаторов обсуждаются, в основном, на уровне конкретной схемы ее реализации: синхронной или асинхронной, т.е. внутри/вне потока data path ("in-band (in-the-data path)" или "out-of-band (outside-the-data path)"). Архитектура SilkWorm AP представляет собой гибридную схему, которая объединяет отделение (out-of-band) управляющего потока от общего с виртуализацией (in-band) потока данных, вбирая в себя преимущества от обоих подходов.

К преимуществам Brocade SilkWorm AP подхода можно отнести и то, что: 1) не требуется установка никакого дополнительного ПО на хостах; 2) отсутствует необходимость замены традиционных HBA на специализированные, разработанные для реализации функций виртуализации; 3) не требуется, чтобы весь I/O трафик был виртуализирован, при необходимости SilkWorm AP может быть сконфигурирован так, что он будет виден как стандартный Brocade FC коммутатор, позволяя виртуализировать только часть SAN (или отдельные массивы и/или серверы), оставляя другую часть SAN нетронутой.