

# Удаленное резервирование и производительность

*В дисковых подсистемах IBM реализован ряд функций, которые могут использоваться при создании центров обработки данных с повышенной устойчивостью к катастрофам. Об особенностях конфигурации и характеристиках производительности одной из них — PPRC — данная публикация.*

Функции удаленного копирования служат для ведения альтернативной копии данных на удаленной резервной дисковой подсистеме. Наличие такой копии может существенно сократить время на восстановление вычислительного процесса в случае потери работоспособности основной дисковой подсистемы.

Например, функция PPRC (Peer-to-Peer Remote Copy) позволяет синхронно и прозрачно для приложения обновлять данные на основной и резервной дисковых подсистемах. Расстояние между ними может достигать 103 км.

Аналогичная функция XRC (Extended Remote Copy) обеспечивает асинхронное копирование данных на резервную дисковую подсистему. Небольшая временная задержка между обновлением данных на первичной и резервной дисковой подсистемах компенсируется тем, что эта функция не влияет на производительность приложений, работающих с этими данными. Кроме этого, XRC позволяет использовать оборудование, снимающее ограничение в 103 км между дисковыми подсистемами.

Другие функции — FlashCopy, Concurrent Copy — позволяют уменьшить время останова базы данных до нескольких секунд для создания резервной копии всей базы в согласованном состоянии. Функция параллельного копирования — Concurrent Copy — позволяет создавать резервную копию базы данных параллельно с основной обработкой. Функция моментального копирования — FlashCopy — позволяет делать “снимок” (моментальную копию данных) для последующей обработки, не влияющей на основной процесс. Например, архивирование, резервное копирование на ленточные носители.

Цель данной публикации — дать представление о характере производительности и информации о конфигурации предлагаемой IBM одной из перечисленных функций резервного копирования — PPRC, реализованной на ESS “Shark” на стандартной Telecom сети с использованием INRANGE технологии на базе устройства 9801 SNS (Storage Networking System).

Приводимая информация о производительности была получена на бенчмарках IBM и может значительно отличаться на реальной нагрузке конечного пользователя.

IBM поддерживает максимальное расстояние для ESS PPRC — в настоящее время 103 км. INRANGE технология позволяет увеличить это расстояние, но такое расширение возможно только после оценки конкретного состава конфигурации и приложений.

## Введение

IBM поставляет PPRC на базе ее новой линейке ESS — “Shark” с конца 2000 года.

Новая версия PPRC, которая выполняется на “Shark”, значительно улучшена в сравнении с версией, которая реализовывалась на RAMAC: в дополнение к намного более быстрому кэшу и дискам IBM также улучшило сам PPRC протокол. Это не только значительно увеличило производительность PPRC, но и улучшило его возможности как в локальном, так и расширенном окружении.

PPRC — позволяет обеспечить, в первую очередь, решение проблемы точного и быстрого восстановления данных после катастрофы.

PPRC-копия данных в резервной подсистеме памяти синхронизирована с операциями I/O основного центра. Это означает, что, ко-

гда прикладная система записывает данные в первичную подсистему памяти, эти данные одновременно передаются в кэш (с защитой по питанию) и управляющая память посылает команду “конец канала” приложению основной системы. Затем управляющая память основной системы инициирует канальную программу I/O для передачи обновленных данных в резервный центр. И когда данные записаны в нестираемый кэш на резервном центре, управляющая память передает команду “конец устройства” прикладной системе. После этого прикладная система уведомляет прикладную программу, что операция закончена.

Другими словами, операция записи на основном центре завершается только после того, как записываемые данные переданы в резервный центр, что вызывает увеличение времени реакции прикладной системы основного центра.

Связь “первичной” и “вторичной” памяти осуществляется через ESCON канал.

Общая проблема, связанная с увеличением длины канала передачи, состоит в том, что для синхронных приложений производительность тем меньше, чем больше расстояние. Это происходит из-за увеличивающихся временных задержек, возникающих при передаче: команд “установления связи” и фреймов. В новой ESS версии PPRC протокола уменьшено число команд “установления связи”. В результате — лучшая производительность поддерживаемая на большем расстоянии, чем было возможно с RAMAC.

## Конфигурация при тестировании

Первичный хост или сервер “не понимают”, что имеется зеркалированный диск. PPRC канал между этими двумя контроллерами дисков — ESCON, независимо от соединения с CPU. Например, RS/6000 может быть

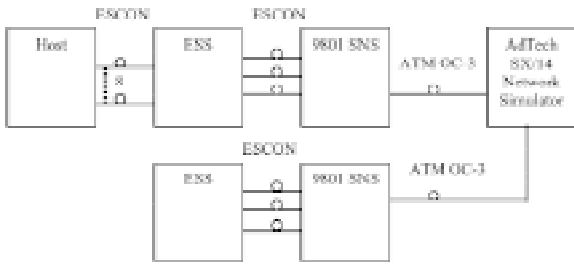


Рис. 1. Конфигурация оборудования при тестировании

связан с ESS через интерфейс SCSI или fibre channel, но PPRC, выполняющийся между ESS, — ESCON.

Все функции зеркалирования формируются в контроллерах. Хост в удаленном сайте может существовать, но необходимость в соединении с CPU возникает лишь при катастрофе.

9801 SNS с ESCON FramePassing™ платой действует, как ESCON модем. Очень упрощенное представление состоит в том, что ESCON фреймы конвертируются в сетевые фреймы и затем обратно в ESCON фреймы. По этой причине 9801 SNS может расширять любой ESCON протокол, потому что он полностью прозрачен. Устройство 9801 слегка усложнено тем, что 9801 поддерживает сжатие данных и некоторые особенности управления сетью, но базовая концепция ESCON модема полностью реализована.

Способ расширения канала — Frame-Passing — идеален для синхронных протоколов типа PPRC, но из-за увеличения задержек, связанных с длиной канала, 9801 SNS не очень хорошо подходит для больших расстояний. С учетом этого фактора и того, что другие протоколы не столь эффективны, как PPRC, INRANGE технология использует подход эмуляции, который в настоящее время доступен только с 9095 ENS, для расширения на большие расстояния потоковых устройств.

Сетевые интерфейсные платы доступны в ATM OC-3, T3, SEPT-3 и HSSI (HDLC). INRANGE предлагает 4:1 инверсный мультиплексор, который устанавливается на заводе в 9801 SNS стойку, который соединяется с OC-3 NIC для интеграции T1 и E1 поддержки.

В дополнение к вышеупомянутому ESCON FramePassing контроллеру, HSSI канальная интерфейсная карта с компрессией данных также доступна и может использоваться для разделения PPRC сети с другим сетевым оборудованием, которое поддерживает HSSI HDLC выход, таким как роутер и другие INRANGE Technologies продукты.

### Некоторые замечания по тестированию с Telecom FramePassing

Производительность при увеличении длины канала с помощью 9801 SNS непосредственно зависит от доступной ширины пропускной способности канала и расстояния.

Расстояние влияет на производительность из-за того, что некоторые части PPRC протокола требуют подтверждения получения

перед посылкой следующего фрейма. Это часто упоминается как “установка контакта”. Когда расширение происходит по сети, фрейм, требующий подтверждения, и само подтверждение должны пересечь сеть, и поэтому добавляется сетевая задержка распространения.

PPRC оптимизирован так, что в течение стадии передачи данных DIB размер и буферы канала не являются ограничивающими факторами. Управляющие команды, начало ввода/вывода и завершающие последовательности все еще требуют “установления контакта”. Передачи с большими блоками данных поэтому дают большую производительность, потому что накладных затрат на передачу при этом меньше.

Данные сжимаются на интерфейсных канальных картах и сетевая полоса пропускания используется только пока данные передаются по каналу. В течение вышеупомянутого “установления контакта” сеть может использоваться для передачи данных от других каналов. Предполагается, что при тестировании имеется достаточное число каналов, чтобы заполнить это время и полностью загрузить сеть. И, в целом, минимальное количество каналов, необходимое для оптимальной производительности, зависит от доступной ширины полосы пропускания, коэффициента компрессии, размера блока и расстояния.

### Результаты тестирования

#### Общая информация по конфигурации

Все тесты выполнялись со следующей конфигурацией оборудования (рис. 1): один тестируемый канал, два из PPRC канальных интерфейсов на каждом ESS были заблокированы. Все измерения с нулевой дистанцией проводились с дистанционным имитатором расстояний в режиме “bypass mode”. Заметим, что это отличается от производительности при локальном подключении, потому что оборудование расширения канала добавляет циклическую (прохождение сигнала туда и обратно) задержку, что эквивалентно задержке при передаче на 125 миль.

Измерения были сделаны с установкой задержки на AdTech SX/14 — сетевым имитаторе времени эквивалентном расстояниям 0, 10, 20, 30, 50, 75, 100, 150, 250 и 300 км. Скорость передачи в линии всегда была полной OC-3 (155 Мбит/с), выполняющей ATM протокол.

#### Последовательная передача данных

**Описание тестирования.** Последовательные записи; 3 трека, переданные одновременно по одному каналу, при размере блока 4 Кбайт. Тест повторен при размере блока 27 Кбайт (рис. 2).

**Комментарии.** Компрессирование для 4 Кбайт блока в испытаниях было 9,2:1. Для блока 27 Кбайт и тех же данных степень сжатия увеличилась до 10:1. Хотя одни и те же данные использовались для обоих испытаний, аппаратные средства компрессирования были более эффективными с большими потоками данных. В дополнение к лучшей степени сжатия, при блоке 27 Кбайт испытание показало на 2 Мбайт/с лучшую производительность.

**Специальные замечания относительно расстояния.** Имеются задержки, свойственные оборудованию расширения канала, и поэтому нулевая дистанция — не то же самое, что прямое fiber соединение. Прямое подключение PPRC для тех же самых испытаний дало бы лучшую производительность.

#### “Установка”

**Описание тестирования.** С PPRC полные треки (55,872 Кбайт) всегда используются, чтобы “установить” дисковую пару в длинное CCW. В этом тесте один PPRC канал использовался, чтобы установить зеркальную

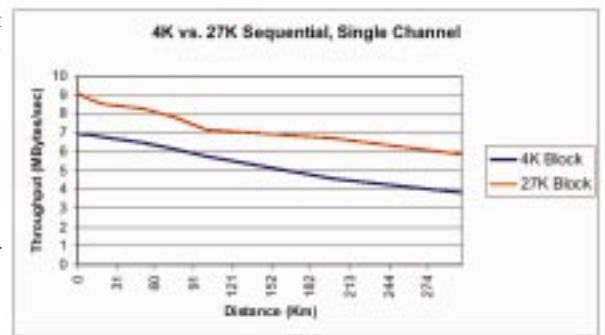


Рис. 2. Пропускная способность при последовательной передаче данных при одном канале

копию на удаленном диске. Испытание было повторено при тех же самых данных для установления зеркальной копии на удаленном диске с тремя PPRC каналами (рис. 3).

**Комментарии.** При нулевом расстоянии 3 канала теста насыщали сеть (OC-3 имеет пропускную способность — 155 Мбит/с, ATM насыщается при скорости передачи данных 130 Мбит/с или 16 Мбайт/с). Компрессирование в течение этого испытания составляло 2,5:1. Поскольку расстояние увеличилось, индивидуальные каналы замедлились из-за задержек “установления контакта”, таким образом освобождая сетевую ширину пропускания. При нулевом расстоянии составная пропускная способность из 3 тестируемых каналов составила только 2,3 раза

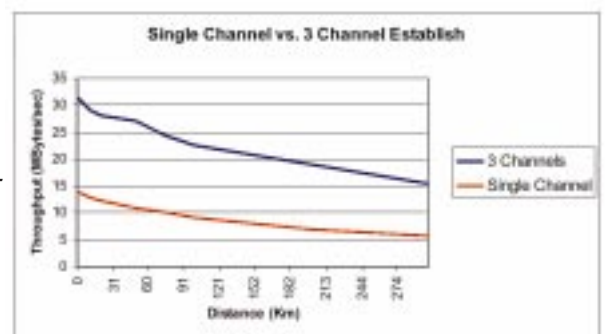


Рис. 3. “Установочная” пропускная способность при одном и трех каналах



Рис. 4. Пропускная способность при операциях случайной записи и одном канале

в сравнении с одним каналом (т.к. сеть насыщалась). При 300 км составная пропускная способность из 3 каналов составила 2,7 раза в сравнении с одним каналом, потому, что сеть больше не насыщается и компрессирование имеет преимущество использования увеличивающейся полосы пропускания.

Чтобы проверить управляемость на длинных расстояниях, один канал был установлен на 1800 км (1100 миль). Пропускная способность в этом случае составила 1,39 Мбайт/с.

#### Случайный ввод/вывод

**Описание тестирования.** Шестнадцать томов 3390-3 осуществляли случайные операции записи I/O. PPRC канальные программы проводили запись блоками по 4 Кбайт.

**Комментарии.** Случайный ввод/вывод обычно измеряется количеством операций I/O в секунду. Хотя тестировался только один канал, составная пропускная способность может быть экстраполирована простым умножением количества каналов на соответствующее значение на диаграмме (рис. 4), до тех пор, пока сеть не подойдет к пределу пропускной способности.

## Дисковая память с соединениями FC для мэйнфреймов

1 августа 2001 года (Сомерс, США) — IBM анонсировала новые продукты и технологии для корпоративных систем хранения данных.

IBM — первый поставщик устройств хранения данных, который предлагает пользователям мэйнфреймов средства, позволяющие подключаться к их высокопроизводительным дисковым системам по интерфейсам Fibre Channel (FICON). Благодаря этому уникальному предложению заказчики, работающие с серверами хранения данных IBM TotalStorage Enterprise Storage Server (“Shark”), смогут наращивать конфигурацию своих систем по мере расширения бизнеса, причем скорость доступа и обработки данных увеличится до шести раз, а емкость системы хранения данных — в четыре раза по сравнению с максимальной, доступной на сегодняшний день.

Впервые в отрасли, обеспечивая собственное соединение FICON с дисковой системой, Shark позволяет увеличить скорость передачи данных до 100 Мбайт/с, в то время как современный промышленный стандарт Enterprise Systems Connection (ESCON) работает на скорость 17 Мбайт/с.

В декабре 2000 года IBM впервые обеспечила собственную поддержку FICON для накопителей на магнитной ленте, выпустив контроллер Magstar 3590 A60. В сочетании с Shark, IBM предлагает полное решение FICON для хранения данных, которое позволяет увеличить производительность, максимальное расстояние, расширить возможности общего доступа к данным и сократить расходы на инфраструктуру за счет замены нескольких каналов ESCON одним каналом FICON.

Благодаря появлению систем хранения FICON данных для мэйнфреймов заказчики впервые смогут подключать мэйнфреймы непосредственно к тем же открытым сетям хранения данных (SAN), с которыми работают серверы UNIX и Windows/NT. Эта общая инфраструктура SAN поможет упростить администрирование сетей хранения данных и позволит заказчикам быстро внести изменения в конфигурацию при изменении потребностей бизнеса.

“FICON позволит значительно увеличить производительность многих приложений, обрабатываемых на мэйнфреймах, — заявил Джон Макартур (John McArthur), вице-президент аналитической компании IDC по исследованию технологий хранения данных. — Новаторские разработки IBM в области поддержки FICON: не только Shark, но и Magstar 3590, еще больше увеличат обороты быстрорастущего бизнеса устройств хранения данных IBM”.

Компания IBM объявила также о дальнейшем усовершенствовании систем Shark — теперь они оснащаются кэшем объемом 24 Гбайт, что предоставляет заказчикам дополнительный, промежуточный вариант выбора системы с кэшем 8, 16 и 32 Гбайт, максимально соответствующей уникальным потребностям каждого набора приложений и требованиям к производительности. Это предложение дополняет недавнее объявление о поддержке в решениях Shark функций Peer-to-Peer Remote Copy (PPRC) и Flash-Copy на серверах iSeries.

Кроме того были объявлены две новые модели сервера виртуальных лентопротяжных устройств IBM TotalStorage Virtual Tape Server (VTS) — первые ленточные системы в индустрии, использующие сверхсовременную технологию медных проводников IBM. Новая технология позволяет вдвое увеличить производительность и снизить энергопотребление.

Две новые модели сервера IBM TotalStorage Virtual Tape Server (VTS) — модели B10 и B20 — первые представители нового поколения ленточных накопителей для электронного бизнеса и являются первыми системами ленточной памяти, использующими микропроцессорную технологию медных проводников IBM. Эти серверы хранения данных используют программную технологию мэйнфреймов IBM Geographically Dispersed Parallel Sysplex (GDPS), а также технологию самовосстановления IBM Shark.

Впервые увидевшие свет в исследовательских лабораториях IBM, микропроцессоры IBM PowerPC с медными проводниками, используемые в VTS, отличаются от алюминиевых чипов целым рядом преимуществ: они меньше, допускают более высокую степень плотности элементов, работают быстрее и выделяют меньше тепла. Кроме того, они надежнее и дешевле. Это новшество позволило вдвое увеличить общую скорость чтения и архивации данных, снизить энергопотребление и стоимость эксплуатации.

GDPS — разработанная IBM технология, применяемая для соединения центров данных между собой и защиты данных путем сведения к минимуму (в идеальном варианте — к нулю) негативного эффекта, вызванного авариями или плановыми отключениями. Технология самовосстановления, заимствованная из проекта Shark, предоставляет диагностический инструмент и функцию вызова администратора по пейджеру в случае неполадок в системе, давая в дальнейшем возможность выявить с помощью ресурсов Web возникшую или потенциальную проблему.

“Наш центр данных значительно “похудел” благодаря применению решений IBM, — говорит Дон Вем (Don Veum), второй вице-президент по информационным технологиям финансовой компании Fortis, Inc., использующей системы ленточной памяти IBM. — Установив VTS, мы смогли сократить количество кассет с лентой с 450 000 до 100 000, причем, по нашим прогнозам, к концу года их останется всего 50 тысяч. За счет этого мы смогли высвободить дорогостоящее пространство и сотрудников для других проектов”.

Описанные технологические инновации являются самыми значительными за всю историю развития систем VTS, начиная с их появления в 1998 году. Кроме мощных решений для крупных корпораций, новые модели позволяют строить системы среднего уровня. Они также способствуют сокращению эксплуатационных расходов благодаря прозрачной и эффективной работе с ленточными ресурсами и сокращению необходимого количества кассет с лентой, лентопротяжных устройств и систем автоматизации. В комплексе эти преимущества позволят компаниям добиться намного более высокого уровня автоматизации хранилищ данных, а также повысить производительность, сократить занимаемый объем пространства и в конечном счете снизить полную стоимость владения.

Кроме новых систем дисковой и ленточной памяти, IBM также анонсировала:

- контроллер сетей хранения данных IBM TotalStorage SAN Controller 160 — многоцелевой последовательный контроллер дисков, предназначенный для подключения по линиям Fibre Channel к UNIX-серверам производства IBM, Sun и HP, а также к серверам Windows NT/2000;