

Выбор решений ключевых проблем управления хранением информации на разных платформах зависит от различных факторов, в том числе и от экономических - и в стоимости создания распределенных приложений клиент/сервер, и в выборе аппаратного обеспечения, которые делают необходимыми программные средства в таких областях, как структурированное хранение информации (hierarchical storage management, HSM) и резервирование (Backup). Эти факторы глобальны по своей природе, и в соединении с административными и техническими проблемами формируют определенную совокупность требований к управлению распределенными данными.

Принципы управления хранением

Прежде всего я хотел бы сформулировать фундаментальное правило управления хранением.

Данные находятся в сохранности только тогда, когда они хранятся более чем на одном носителе и более чем в одном месте.

Это утверждение может показаться одной из самоочевидных истин, однако новые технологии его так легко вуалируют, что мы теряем ощущение реальности. Любой носитель может выйти из строя, будь это диск оперативного доступа, магнитная лента или оптический диск. Аналогично может быть разрушено любое место хранения - например стореть.

Первый принцип управления хранением - хранить данные в безопасности. Проблема, из-за которой мы уделяем так много внимания данной теме, состоит в том, чтобы сделать это наиболее доступным по цене способом. Вместо единого решения, которое подходило бы для всех вариантов хранения, существуют компромиссы. Обычно они возможны между стоимостью хранения, усилиями по выполнению резервирования и стоимостью простоя и потери данных. Стоит еще раз напомнить самим себе, - зачем мы делаем резервирование - чтобы защититься от случайностей (о, дорогой, я только что удалил не тот файл) и разных бедствий (разрушен диск или сторел центр обработки данных).

Существуют гибкие методики, такие как RAID, включая зеркалирование (mirroring), однако хотя они и повышают надежность системы, но не защищают ни от катастрофы (если у вас нет удаленной зеркальной копии), ни тем более от случайного удаления.

Два или три года назад все были твердо уверены, что перенос приложений в открытую среду клиент/сервер, сэкономит значительную сумму денег. В конце концов, мэйнфреймы и мини-компьютерные системы предлагались ограниченным числом производителей и дорого стоили. Ясно, что недорогие RISC-процессоры могли бы обеспечить тот же уровень производительности за гораздо меньшую цену.

Такое представление изменилось. Последнее исследование Information Week показало, что менее половины менеджеров информационных систем по-прежнему считают, - модель клиент-сервер экономит деньги. Однако почти треть считает, что цены могут даже вырасти!

Посмотрим, почему же произошел такой поворот на 180 градусов. Другое недавнее исследование показывает, что стоимость распадается на несколько ключевых компонентов, и для платформ клиент/сервер, включая UNIX и локальные сети ПК компонент, связанный с расходами на персонал, значительно выше, чем для более традиционных мэйнфреймов.

Это обусловлено полным отсутствием мощных и изощренных автоматизированных средств управления, подобных тем, которые последние два десятилетия развивались в среде мэйнфреймов, а также сложностью выполнения, присущей распределенным системам, включающим в себя большое число независимых и разделенных компонентов.

Таким образом, мы четко определили первый ключевой экономический фактор, влияющий на нашу стратегию выбора программного обеспечения. Он затрагивает стоимость трудовых затрат, и целью любого программного продукта, очевидно, должно быть снижение цены администрирования и управления распределенными системами.

Если модель клиент-сервер такая дорогая, почему бы просто не вернуться к мэйнфреймам? Простой ответ не всегда самый подходящий, и мы видим это - не случайно переход к клиентсерверным приложениям продолжается с высокой скоростью. Основная причина этого перехода состоит в том, что бизнесмены переосмысливают использование технологий, для обеспечения уровня сервиса, необходимого для того, чтобы обогнать конкурентов. Рассмотрим технологию обработки изображений, которая во многих отраслях производства и банках позволила сократить количество обрабатываемых и хранимых бумаг. Для этого типа приложений требуется графика высокого разрешения, быстрая скорость обработки и большой объем памяти; эти требования лучше всего обеспечивает использование технологии клиент-сервер. Хотя при этом цена может вырасти, реорганизация бизнеса приведет к повышению производительности и, следовательно, общей конкурентоспособности. Мы можем сделать вывод, что переход к открытым системам будет продолжаться с большой скоростью. Однако нельзя рассчитывать на значительное увеличение персонала.

Согласно другому недавнему исследованию в прошлом году было продано 30 миллионов дисков. Объем используемых данных увеличивается с нарастающей скоростью - растет интенсивность работы уже существующих пользователей, добавляются новые пользователи и приложения.

Размер отдельных файлов часто остается довольно небольшим, но совокупное число файлов, управляемых большинством систем,

огромно. Один гигабайт диска, стоящий заметно менее 2000 долларов, может содержать порядка 100000 файлов - чересчур много объектов, чтобы суметь обработать их с использованием доступных сегодня ручных методов.

Если мы посмотрим на сравнительные скорости роста, то увидим, что при увеличении требований к объему хранимых данных на 45% количество инсталлированных дисков должно возрасти на 20%. Однако вследствие высокой стоимости персонала появилась возможность увеличить штат всего на 5%.

Эти тенденции не оставляют другого выбора, кроме перехода к новым методам управления, благодаря которым можно автоматизировать функции резервирования, архивирования и перемещения данных. Если вы попытаетесь справиться с этими проблемами, используя методы, работавшие два или три года назад, скорости роста просто сокрушат вас и вы столкнетесь с неприемлемо высокими ценами на трудовые затраты.

Каковы технологии?

Рассмотрим технологии, доступные для развертывания в среде открытых систем. За последние несколько лет цены и на CD-ROM, и на магнитооптические диски формата 5.25 дюйма упали, в то время как их плотность выросла до такого уровня, на котором эта технология становится вполне привлекательной. Аналогичная картина (высокая плотность при низких ценах) наблюдается в технологии 8 и 4 мм магнитных лент.

Более современные магнитооптические дисководы формата 3.5 дюйма снизились в цене, но по-прежнему имеют очень ограниченную емкость, в то время как ленты VHS обладают большой емкостью и высокой скоростью, однако стоят дороже, чем ленты 8 и 4 мм. Мы ожидаем, что магнитооптические устройства достигнут необходимого уровня плотности и станут привлекательнее в ближайшие несколько лет. Сегодня технология VHS используется преимущественно в среде с очень большими объемами данных. По мере приближения этой технологии к нормальному уровню цен можно ожидать, что и она получит более широкое распространение.

Перечислив основные технологии устройств, обратимся к роботизированным версиям, автоматизирующим функцию передвижения носителей. Семейство оптических роботов компании HP очень сильно снизилось в цене за последние три года. Высокопроизводительный робот с множеством дисководов большой емкости теперь доступен менее чем за 100000 долларов. Самые известные роботы магнитных лент сегодня - это семейство Exabyte для лент 8 мм, также обладающее вполне приемлемыми ценами и включающее устройство самой большой емкости, значительно более дешевое, чем год или два назад. Очевидно, что роботизированные средства хранения теперь имеют достаточно низкую цену, чтобы стать подходящими кандидатами на роль популярного автоматизированного устройства хранения. Подводя итог нашего обзора технологических альтернатив, отметим, что каждую из них можно приобрести у широкого ряда поставщиков, часто под различными именами, но в действительности все они в данный момент стали товарами массового спроса. Кроме того, аппаратные решения бесполезны

без интеграции в него программного обеспечения соответствующего уровня.

Действительно, существует много поставщиков и много-много имен, однако обычно это одна и та же технология, либо просто с различными именами, либо в различной упаковке. Пока же вы не выбрали программное решение, аппаратура совершенно беспомощна. Программы управления хранением делают технологию жизнеспособной.

Какое программное обеспечение требуется?

Итак, что вам следует делать для использования технологий и какие характеристики должны присутствовать в любом программном решении? Аппаратура, вероятно, быстро станет массовым товаром. А вот как насчет программного обеспечения? Какие прикладные технологии необходимы, чтобы решить рассмотренные выше экономические проблемы? Несмотря на относительно низкую цену за гигабайт магнитных дисков оперативного доступа (порядка 1500 долларов), особенно быстрое снижение цены наблюдается при переходе к магнитооптическим дискам, роботизированным лентам и, наконец, к автономным устройствам хранения. В определенных условиях наиболее выгодной может оказаться стратегия перемещения файлов, которая создает иерархию носителей данных и переводит файлы на различные уровни, чтобы обеспечить самую низкую стоимость хранения.

Аналогично, рассматривая компонент цены за труд, мы увидим, согласно исследованию, проведенному в 1992 году и по-прежнему справедливому сегодня, что автоматизация резервирования является одним из наиболее важных факторов, необходимых для снижения общей стоимости функций администрирования системы. Свыше 2500 долларов в месяц (заработная плата администратора на полном жалованье) тратится на выполнение вручную этой жизненно необходимой функции.

В 1995 году можно ожидать начала развертывания продуктов управления распределенными системами в области хранения. Эти продукты обеспечат распределенные сетевые решения для перемещения файлов и управления перемещенными файлами на оптических дисках и магнитных лентах, наиболее популярных на рынке открытых систем. Продолжится разработка средств и на рынке традиционных мини-компьютеров или мэйнфреймов, которые решают те же задачи, что и сетевые открытые системы, но применяют традиционную аппаратуру типа мощных серверов, извлекая выгоду из масштаба централизованной системы и используя при этом мощность распределенной модели клиент-сервер.

На что будет похожа распределенная система HSM?

Теперь мне хотелось бы представить некоторые из ключевых проблем, вокруг решения HSM для среды открытых систем. Эти решения стоит сопоставить с результатами деятельности комитета IEEE Mass Storage. Существуют другие решения, но модель, предложенная IEEE, предлагает прекрасный базис для

сравнительного изучения. Полезно дать краткое описание модели IEEE, тем более что вы все чаще будете встречаться с ее терминологией в ближайший год или два. Затем я расскажу о том, как, по-моему, надо определять требования для продуктов HSM.

Модель IEEE

Терминология, предложенная IEEE, включает понятие битфайл (bitfile), которое определяет простой пакет с группой «бит», содержащий все атрибуты, персональную принадлежность, данные и любые другие компоненты файла, характерные для любой конкретной операционной системы, файловой системы или другого прикладного контекста, в котором обычно может использоваться файл. В конце концов, битфайл - это пакет, которым легко управлять, находясь за пределами области его оперативного размещения, и он становится ключевым объектом, управляемым системой HSM. Битфайл-клиент - это абстрактная функция, которой известна файловая система, где обычно размещаются битфайлы, и которая может создавать битфайл или восстанавливать его из HSM на диск оперативного хранения. Он связывается с битфайл-сервером, который понимает, как взять битфайл, дать ему имя и управлять им. Наконец, IEEE определяет абстракцию «сервер имен» (name server), поддерживающую отображение битфайла в приложение, так что мы можем определить, как декодировать битфайл и восстановить прикладной формат.

Продолжая рассмотрение терминологии IEEE, определим «сервер хранения» (storage server), который лучше всего представить как отображение, или функцию базы данных, хранящую информацию о том, в каком месте иерархии доступных носителей находятся перемещенные с диска оперативного доступа битфайлы; однако носители при этом абстрагируются до уровня «абсолютных» томов, поэтому представление их не зависит от физического размещения данных на носителях более низкого уровня. Это позволяет обеспечить независимость от устройств и изоляцию от конкретных технологий, поэтому система способна легче принимать новые варианты технологий, а также переходить к более «прогрессивным» носителям. «Хранилище физических томов» (physical volume repository) обеспечивает другой функциональный уровень, который отвечает за физическое управление самого носителя, отвечая за передвижение носителей с помощью роботизированных устройств, за постановку в очередь заданий для определенного устройства и т.п. Наконец, «менеджер узла» (site manager) задает компонент, который отвечает за конфигурирование, поддержку и функционирование системы.

Приведенный рисунок иллюстрирует взаимоотношения между перечисленными компонентами. Одна из ключевых рекомендаций IEEE - строгое разделение потоков управления и данных. Это позволяет программировать взаимодействия между битфайлклиентом, битфайл-сервером, сервером имен, сервером хранения и хранилищем томов с помощью любого универсального метода распределенного программирования, например стандартного протокола RPC (Remote Procedure Call - Вызов

Удаленной Процедуры). Использование RPC облегчает реализацию и делает доступными, присущие данному механизму, полезные моменты, включая такие функции, как XDR (внешнее представление данных - eXternal Data Representation). Как только открыт требуемый каталог, по путям управления выполняются размещение третичного хранения, выбор тома и другие аналогичные действия, затем может быть создано средство переноса битфайлов (Bitfile Mover), которое решает простую задачу переноса данных непосредственно с on-line диска на близлежащие или автономные носители без каких-либо дополнительных накладных расходов или передачи через промежуточные узлы распределенной сети. Просто открывается высокоскоростной путь, такой как канал подключения потока (stream socket) между сетевым узлом, управляющим назначенным носителем, и данные передаются.

Предложения по реализации

Итак, битфайл-клиент, битфайл-сервер и средство пересылки битфайлов организованы как HSM-клиент, обладающий магнитным диском оперативного доступа и желающий иметь возможность перемещать свои файлы. Сервер имен и сервер хранения - это функции базы данных, находящиеся еще где-либо в сети и реализующие отображение, которое сохраняет информацию о местонахождении любого битфайла в любой момент времени. Хранилище томов образует другой компонент; он расположен в сети на любом узле, обеспечивающем «почти оперативный» доступ либо посредством роботов оптических дисков или магнитных лент, либо посредством требующих дополнительного управления автономных дисководов.

Формат носителей

Одна из ключевых проблем в любой системе управления хранением - это формат записи данных. Он может быть каким угодно - от стандартного дискового файла операционной системы до формата ANSI магнитной ленты, закрытого формата резервирования или полуоткрытого стандарта, подобного tar и spio в UNIX-системах. Закрытые форматы нередко обеспечивают существенные преимущества в производительности, так как они создаются в расчете на конкретное приложение (резервирования, переноса или архивирования) и на формат файлов в конкретных ОС. Рассматривая открытый формат носителя, вы должны также учесть формат самих данных. Вряд ли стоит настаивать на действительно открытом формате носителя, если все, что потребуется писать, это данные закрытой системы. Существует также большая разница между стандартом ОС и открытым стандартом. Это можно наблюдать на рынке VMS, где формат, используемый утилитой VMS BACKUP, часто принимается за стандарт. Как только вы начинаете выходить за пределы чистой среды VMS, так называемый стандарт фактически становится помехой. Возможно, лучше было бы использовать продукт со своим собственным форматом, который переносится во множество ОС, чем полагаться на стандарт только одной из них.

Еще более очевидно, что при построении HSM для открытых систем данные следует записывать с использованием открытых стандартов. В Raхso мы следуем этим принципам. Для VMS, которую не принято считать самой открытой системой, в наших продуктах TapeControl и AutoStor мы приняли решение использовать закрытые форматы (либо Digital VMS, либо свои собственные форматы), что позволяет получить существенное преимущество в производительности. Для продуктов на основе открытых систем, например Backup-Unit, мы использовали открытые форматы spio и tar.

Структура на диске (файловая система)

Рассмотрев формат носителей, предназначенных для резервирования, перемещения или архивирования, обратимся теперь к структуре на диске. Мы считаем, что решение HSM никоим образом не должно изменять физический формат файловой системы. Это важно, чтобы избежать замкнутости на специфических версиях операционной системы и риска нанести вред при использовании других утилит, которые не понимают сделанных хитрых изменений.

Для среды UNIX это означает, что не должно быть никаких модификаций ядра UNIX или изменений любых команд пользователя, включая продукты резервирования или команды spio и tar. В настоящий момент это требование удовлетворяется далеко не во всех HSM-системах для ОС UNIX. Большинство нуждается в обширных изменениях файловой системы, и по сути каждому второму поставщику требуется лицензия на исходный код базовой ОС, чтобы поддержать значительные изменения в ядре.

Интеграция резервирования

Система HSM должна обеспечивать определенный уровень прозрачности доступа к файлам, так чтобы не вывести из строя любые существующие продукты резервирования других поставщиков или свои собственные. Необходима возможность использования любого средства резервирования, а в рамках HSM-решения должно быть известно о стратегиях резервирования. Ключевое соображение состоит в том, что продукт резервирования должен иметь возможность резервировать и восстанавливать перемещенные представления файлов, а не вызывать дополнительное перемещение, что не обеспечивает защиты диска оперативного доступа и может потребовать чрезмерного количества времени. Любой продукт HSM, который не обеспечивает этого уровня прозрачности, либо разрушит существующую систему резервирования, либо потребует приобретения всех технологий у одного поставщика. В то же время важно выполнение фундаментального правила управления хранением: данные хранятся на множестве физических объектов (носителях) во многих местах; система HSM защищает свои собственные данные и имеет возможность синхронизироваться с выбранной системой резервирования.

Управление «мусором»

Наконец, не забудьте, что данные имеют ограниченное время жизни. Должны существовать способы их удаления, лучше без изменения стандартных утилит - `del`, `rm`, `delete` или какую-либо еще - в зависимости от того, как они называются в выбранной вами ОС. Система HSM должна обнаруживать и улаживать такие удаления. Слишком часто одной из главных задач администратора является уборка устаревших файлов пользователей.

Управление миграцией

Если говорить о перемещении файлов, необходимо, чтобы система HSM обеспечивала возможность запускать перемещение либо непосредственно пользователем (по требованию в случае недостатка места), либо запускать перемещение периодически, так чтобы необходимые операции управления могли выполняться в течение выделенного времени, лучше в периоды низкой загрузки, так чтобы к началу рабочего дня была доступна адекватная емкость дискового пространства.

Перемещение файлов можно рассматривать как двухэтапный процесс. Во-первых, данные должны быть скопированы в безопасное место (хранение битфайлов), после чего копия, находящаяся на устройстве оперативного доступа, усекается и помечается таким образом, чтобы система HSM могла обнаруживать попытки доступа к ней и перемещать данные обратно. Нет причин для того, чтобы выполнять эти два этапа совместно. Можно до поры до времени ограничиваться первым этапом: это означает, что копирование данных выполняется, но усекаение откладывается до тех пор, пока не возникнет реальная потребность в свободном месте. Привлекательность этого метода состоит в том, что он обеспечивает очень быстрый ответ в случае нехватки места и при этом поддерживает оперативный доступ к файлам так долго, как это возможно.

Для управления процессом перемещения можно использовать две возможные модели: либо система оценок, определяющих необходимость перемещения, либо основанная на правилах система предсказуемого управления. Обе модели достойны внимания, и, возможно, в идеале обе следовало бы реализовать. Ясно, что, если диски пусты, вам не потребуется переносить много файлов, но, если пространство очень заполнено, придется перенести больше файлов. Система оценки обычно определяет «ватерлинии», например, 90% используемого дискового пространства как пороговое значение для начала полного перемещения, 70% используемого дискового пространства как сигнал остановки полного перемещения файлов и 50% как сигнал прекращения предварительного этапа. Для системы на базе правил аналогичный эффект достигается выполнением правил в зависимости от качества свободного пространства. Могут быть определены правила, которые применяются, только когда пространство на диске критично (скажем, менее 10%), другие для небольшого объема свободного пространства, и наконец, регулярные правила, которые применимы всегда. Конечно, фактические проценты, употребляемые для «ватерлиний», могут

быть различны для определенных систем, но обычно предусмотрены значения, используемые по умолчанию. В общем случае, системы на базе правил обеспечивают более сложные критерии и большую гибкость, в то время как системы оценок проще настраиваются. Обычно системы оценок предоставляют лишь небольшое число критериев для выбора перемещаемых файлов, такие как возраст или размер, а системы на базе правил позволяют учитывать практически все атрибуты файлов. Возможно перемещение вручную, даже перенос целого каталога. Наконец, для некоторых специальных файлов должна быть возможность защиты от перемещения.

Управление многоуровневыми носителями

Для устройств резервного хранения необходимо иметь возможность группировать перемещаемые файлы в соответствии со специальными требованиями. Возможно, мы захотим использовать высокопроизводительный оптический дисковод с автоматической сменой дисков только для определенных приложений или для файлов, используемых вполне определенным образом. Для приложений с более низким приоритетом нам, может быть, понадобится выполнять перенос на магнитные ленты. Очевидно, что при этом потребуется больше времени для перевызовов, но в таких случаях допустима более низкая производительность в обмен на более низкую общую стоимость носителей.

Для резервного хранения также существует необходимость определять путь переноса файла, перемещаемого на нижние уровни иерархии хранения. Например, файл может мигрировать с устройства оперативного доступа на оптический диск, затем, через некоторое время, на магнитную ленту в работе и наконец на автономную ленту, если к нему все еще нет доступа. Система HSM должна допускать добавление монтируемой файловой системы любого типа на произвольный уровень иерархии. Это позволяет добавлять в иерархию резервного хранения устройства, основанные на новых технологиях, по мере того как они становятся доступны на рынке. Очевидный пример – использование дисков, ориентированных на высокую производительность, на первом уровне, затем дисков, ориентированных на меньшую пропускную способность, затем оптических дисков, магнитных лент и автономных лент в качестве самых низких уровней.

Архитектура клиент-сервер

Вместе взятые, эти требования определяют систему HSM с действительной архитектурой клиент-сервер. В настоящий момент на рынке существуют продукты HSM, основанные на других технологиях, таких как NFS, для рассредоточения функций перемещения файлов, но они не обеспечивают действительно распределенного решения, так как контролируют управление хранением только на сервере и не делают этого на клиентских рабочих станциях. Они также не допускают совместного использования множеством серверов ресурсов запоминающих устройств большой емкости (роботов оптических дисков и

магнитных лент). Действительно распределенная архитектура обеспечивает значительно большую гибкость по сравнению с этими, более ранними архитектурами. Кроме того, разделение путей управления и путей данных обеспечивает более высокий уровень производительности по сравнению с тем, что можно получить от менее продуманной архитектуры.

Результатирующее решение, благодаря его общности, весьма популярно для сетевых сред. Множество роботов можно разместить в множестве мест; перемещение можно настроить различными способами, чтобы добиться наилучшего сервиса.

Стандартизация

Существует организация, носящая название Data Management Interfaces Group (DMIG), которая предпринимает попытки создать стандарт для интерфейса между системой HSM и нижележащей операционной системой. Любая система HSM должна находить способ управления определенными событиями в ней. Например, всякий раз, когда пользователь открывает перемещенный файл, управление должно передаваться системе HSM, чтобы инициировать прозрачное перемещение файла на диск оперативного доступа. К сожалению, сегодняшняя среда не обеспечивает стандартного способа сделать это, что мешает переносимости HSM-решений. DMIG – это консорциум поставщиков HSM и производителей ОС, которые пытаются создать эскиз нового стандарта, определяющего этот интерфейс. DMIG предложила представить этот стандарт для официального утверждения в X/Open, как только он станет доступен. Когда стандарт станет доступен в созданных различными производителями операционных системах типа UNIX, мы будем поддерживать его.

Резюме

Мы показали, что в мире открытых систем цены ведут себя неожиданным (не самым лучшим) образом из-за недостатка средств администрирования, влияющих на величину трудовых затрат. Несмотря на то, что цены на аппаратуру значительно снизились – фактически до уровня массового товара – уровень автоматизации, необходимый для обеспечения низкой стоимости трудозатрат, потребует программного решения.

Рассматривая эту проблему и усилия, предпринимаемые для ее решения, мы увидели, что резервирование и архивирование системы HSM являются ключевыми приложениями, которые потребуют дальнейшей разработки. Разработка должна будет опираться на клиент-серверную архитектуру и результаты работы по стандартизации.

РАЗОБЛАЧЕНИЕ МИФА

Более дешевый компьютер не обязательно означает более низкую стоимость вычислений. При добавлении прикладного программного

обеспечения или увеличении персонала цена в расчете на одного пользователя для распределенных систем становится на 50-130 процентов выше, чем для мэйнфреймов!

Так почему мы переходим к модели клиент/сервер?

1. Продуктивность для отдельных лиц.
2. Изменения парадигмы бизнеса.
3. Успех в общей конкурентной борьбе.
4. Гарантия корпоративной целостности.