

Хранение терабайт, или Как сохранить IT бюджет

Данная публикация — первая из серии, посвященная решениям на основе ленточных технологий, разрабатываемых и поставляемых одним из лидеров в этой области — компанией StorageTek (www.storageitek.com).

Цель серии — расширить представление читателей о возможностях и преимуществах, которые дают эти решения в области хранения и управления данными.

“Хранение данных часто представляется как унылая, но необходимая задача, — пишет Amanda Baker — аналитик StorageTek. — Но при этом все же недооцениваются возможности, которые это приносит.”

Введение

Увеличивающийся поток данных в организациях уже не вызывает никакого удивления, если знакомиться с оценками аналитиков роста хранимых данных. Ежегодный их прирост на 100% — самое обычное явление.

Поскольку всемирная “жажда” информации и уверенный рост данных не показывает никаких признаков замедления, происходящее “наводнение” данными продолжится и дальше. Вопрос, как организации сохранить все данные, которые поступают, комментирует Михаэль Фэт (Michael Vätth), эксперт по решениям и услугам хранения компании StorageTek: “Туп данных, которые мы сохраняем, также имеет большое значение. Так, приложения, генерирующие интенсивный поток данных, типа видео, электронной почты, медицинских изображений и программ CRM, заполняют объем ресурсов хранения действительно очень быстро. Это создает много проблем для предпринимателей в связи с необходимостью дополнительного финансирования или сокращения IT бюджета, соответственно, эффективное управление информацией становится серьезной проблемой”.

Для большого количества компаний удаление данных — не опция. Мало того, есть правительственные правила, регламентирующие постоянное увеличение временного периода хранения данных. Есть также необходимость длительного хранения бизнес-информации, которая может оказаться полезной в последующем. В любом случае, важно принять решение, что сохранить, а что уничтожить? Для большинства IT под-



Michael Vätth, эксперт по решениям и услугам хранения компании StorageTek

разделений административное бремя этого действия находится за пределами их возможностей. Ситуацию усугубляет еще и то обстоятельство, что, согласно исследованию, проведенному Университетом Беркли (University of Berkeley), 80% данных, сохраненных в цифровой форме — фактически скопированные. Это означает, что помимо “новых” данных, многократно размножается и сохраняется первичная информация.

Перспектива рисуется мрачной, но есть ли надежда для тех, кто пробует уравновесить требования по стоимости, производительности, защите и емкости? “Да, — отвечает Фэт. — Когда данные созданы, начинается их жизненный цикл, который показывает, что наибольшее число обращений за ними приходится только на первые 7 дней. После этого частота обращений к ним значительно падает, но они должны быть в состоянии готовности в случае необходимости. После 90 дней к данным вообще редко кто обращается. Странность ситуации в том, что вопреки этой статистике, данные продолжают и в дальнейшем хранить тем же способом и на тех же носителях, на которых первоначально они были записаны. Это означает, что есть значительные резервы для улучшения эффективности их хранения”.

Ситуацию обостряет еще и то обстоятельство, что, согласно исследованиям, проведенным Gartner/Credit Suisse, эффективность использования дискового пространства достигает всего лишь 35-40%. А благодаря последним усовершенствованиям работы лент, информация может быть найдена в течение 20-30 секунд. Добавьте к этому тот факт, что даже с падением цен на enterprise диски, хранение информации на этих носителях приблизительно в 85 раз (в пересчете на 1 Гбайт) дороже, чем при использовании автоматизированных ленточных устройств. Эти обстоятельства обеспечивают хорошую перспективу для лент при архивировании больших объемов данных.

“Стоимость носителей данных должна быть сбалансирована по частоте обращения к ним”

Jon Collins — эксперт из аналитической группы Quocirca соглашается, что есть еще возможность для улучшения: “Организация хранения не должна строиться в соответствии с концепцией “one-size-fits-all”. Стоимость носителей данных должна быть сбалансирована в соответствии с частотой обращения к ним, и размещать редко используемые данные архивного характера вместе с часто используемыми офисными транзакционными данными на высокопроизводительные диски крайне глупо. Различные типы носителей имеют свою нишу использования в соответствии с тем, как требования к данным изменяются в течение времени”.

Но как решить, какие данные где должны находиться? “Каждая организация имеет свои собственные индивидуальные потребности, — говорит Фэт. — Например, к данным, используемым в здравоохранении, обращаются значительно позже их создания — всякий раз, когда пациент посещает доктора. С другой стороны, в технической среде к данным наиболее часто обращаются во время разработки проекта и эти обращения резко падают с началом производства. Поэтому, прежде всего, важно установить функции распределения частоты обращения



Рис. 1. Автоматизированная библиотека StorageTek Powderhorn и часть конструкции робота.

к данным во времени для различных приложений и типов данных.”

В большинстве случаев, создание многоуровневых архивов состоит в построении алгоритма миграции данных на основе модуля управления жизненным циклом информации (ILM — Information Lifecycle Management). Физически — это комбинация дисков высокой производительности для данных, к которым доступ должен быть немедленным (ATA-диски*) высокой емкости), и ленточных технологий хранения для данных с редким обращением. Примером организации управления данными между двумя уровнями может служить пакет ASM** StorageTek — интеллектуальное ПО с низкими затратами на управление и вмешательством человека.

Поскольку ИТ инфраструктуры развиваются постепенно, часто по методу устранения текущих трудностей, то в результате системы хранения с точки зрения производительности и управления представляют одну сплошную головную боль для ИТ администраторов, в большинстве случаев при низкой эффективности использования общего объема хранения. Одна из причин этого — в плохой сбалансированности устройств в системе, поскольку некоторые из них перезагружены, а другие недогружены. Консолидация в данном случае, суть которой состоит в объединении разрозненных заархивированных данных в одной большой ленточной библиотеке, приводит к значительно более низкой стоимости хранения в пересчете на 1 Тбайт. При этом достигается улучшенная защита, снижаются затраты на персонал за счет уменьшения количества управляемых устройств и появляется возможность более эффективного использования ресурсов хранения, а также их масштабирования. По сообщениям одной европейской телекоммуникационной компании, сокращение стоимости составило более чем 50% (в пересчете на 1 Гбайт) после проведения объединения ее данных на ленточной библиотеке StorageTek Powderhorn.

“Автоматизация сама по себе — способ сокращения риска”

Резервное копирование и восстановление — одна из областей, где автоматизированное ленточное хранение играет жизненно важную роль. В копировании на ленту иногда больше смысла, чем в зеркалировании или других вариантах для защиты целостности данных. Низкая цена — не единственное преимущество, которое дает лента. Лента идеальна для того, чтобы обеспечить безопасную копию вне сайта информации компании. Например, если зеркалированный сервер атакован вирусом, данные будут испорчены и на зеркале. Ночное резервное копи-

*) ATA — первый интерфейс из ATA (AT Attachment) / IDE-семейства. Предназначен только для подключения двух жестких дисков на одном канале (прим.ред.).

** ASM — Application Storage Manager — интеллектуальное ПО перемещения данных, разработанное StorageTek для автоматизации управления данными (сохранение и восстановление) и их поиска в иерархической системе хранения на основе политик, определяемых пользователем (прим.ред.).

Каждый выпуск телевизионных программ, однажды показанный, сохраняется для будущего использования, и немногие зрители задумываются над тем, как телекомпания весь этот объем информации хранит.

Для основанной в Люксембурге широкоэмитательной компании для центральной Европы (Broadcasting Center Europe — BCE), архивирование видео — необходимая часть ее бизнеса. BCE — часть RTL Group и они передают 16 каналов приблизительно на 7 европейских стран. Леон Маркс (Léon Marx), вице-президент по оценкам и консультациям технологий BCE комментирует: “Мы имеем 750 000 событий в год только на наших 6 коммерческих каналах. Добавьте к этому 11 тематических каналов, которые мы передаем, и это — колоссальное количество видео, которое мы должны сохранить.”

“До недавнего времени, архивируя все это видео, мы израсходовали 2000 м² в штабе BCE, — продолжает Маркс, — мы хотели развивать новый бизнес, но не имели никакой площади для дополнительных каналов.

Если бы мы не решали эту проблему, это препятствовало бы росту BCE. Площадь была не только единственной проблемой.” Джин Лампак (Jean Lamrach), глава подразделения разработки и консультаций BCE добавил к этому проблему разрушающих последствий ошибки персонала: “При нашем количестве событий, которое мы обрабатываем, даже низкая норма ошибки — 1% — это приблизительно 7 тыс. проблем в год. Поэтому мы должны были быть всегда готовы к плану непредвиденных обстоятельств.”

Леон Маркс продолжает: “Особенностью TV технологий является невозможность вещания программ на несколько каналов одновременно. Так как мы имеем 16 каналов, то для нас это была большая проблема. Мы видели выход в решениях ИТ индустрии и обратили свой взгляд на другие компании, использующие хранение данных. В особенности банки имеют очень большие объемы хранимых данных, и они одновременно имеют подобные потребности высокой доступности и надежности. Все они использо-

вание позволяет сделать откат назад, хотя и с некоторой потерей данных, но не всех, как это может случиться в первом варианте. Автоматизация сама по себе, сокращая вмешательство человека, также уже представляет некоторый путь к сокращению риска.

Конечно, недостаточно просто резервировать данные — восстановление процесса также важная составляющая. Изучение Meta Group показало, что среднемировое снижение оборота в час от потери данных — приблизительно 1 млн дол. Для некоторых отраслей промышленности, типа энергетики, телекоммуникаций, эти потери измеряются более чем 2 млн дол. в час. Какие компании могут противостоять такому роду потерь? Очень немногие, как это оказывается. Действительно, другой деловой обзор — Eagle Rock Alliance — показал, что больше чем половина всех компаний считает свое банкротство крайне вероятным после того, как вре-

вали продукты StorageTek. Другими ключевыми критериями выбора были скорость плюс совместимость и не только с существующей инфраструктурой BCE, но также и с другими продуктами на рынке. Lamrach прокомментировал: “ Мы выбрали StorageTek из-за скорости, надежности и технологии приводов. ”

Решение по организации видеоархива, реализованное в BCE, состоит из накопителей на магнитной ленте StorageTek 9840 (выбранных исключительно из-за их скорости) и двух библиотек Powderhorn. Программное обеспечение от Front Porch Digital гарантирует, что управление архивом происходит “гладко” и требует небольшого количества ресурсов. В настоящее время BCE имеет заархивированные 25 000 часов видео, что составляет примерно 72 000 часов видео в реальном времени. Система полностью автоматизирована и позволила высвободить 16 человек, ранее работающих с этим архивом и необходимых были только для поиска и замены кассет. Эти люди теперь освобождены для других задач в BCE.

Выгоды для BCE были во многом. “Мы сэкономили деньги на обслуживании, на людях и площадях. На этапе оцифровки информации, мы имели порядка семи роботов и на каждый пять видеосистем записи изображений на ленту (VTR), которые из-за интенсивности работы приходилось менять каждые 2500 часов. Это в свою очередь также снизило высокие затраты на персонал, связанный с архивированием видео. Сейчас мы нуждаемся только в 40 м² площади по сравнению с 2000 м², и это — очень большое различие. Экономия площади имела большое значение для BCE, непосредственно давая возможность развернуть новый бизнес. До этого проекта мы не имели никакого запасного пространства — все было занято заархивированными кассетами. С тех пор, как освободилась часть площади, мы разместили на ней новое техническое оборудование и диспетчерскую, выиграв уже новые контракты и увеличив число каналов, ” — прокомментировал Маркс.

простая достигает 24 часов. Для 4% опрошенных это время еще меньше и составляет только один час. И в немногих ситуациях даже небольшое количество потерянных данных является критическим, например, для крупного банка, скажем, 1 час сделок. И все же, как с такими объемами сохраняемых данных, выполнить восстановление функционирования максимально быстро?

“Информация должна храниться в соответствии с приоритетами”

Эксперт по решениям и услугам хранения М. Фэт компании StorageTek рекомендует проведение анализа воздействия на бизнес каждого типа сохраненной информации для определения ее приоритетности: “Информация может быть грубо разделена на четыре категории: критическая, жизненная, чувствительная и некритическая. И эти категории должны располагаться и храниться по при-

StorageTek в России



Леонид Винокуров, руководитель Департамента информационных технологий ТехноСерв А/С.

SN. Как Вы оцениваете позиции StorageTek в мире и в России?

Л.В. Компания StorageTek занимает ведущие позиции на мировом рынке систем архивирования информации. Имея более 150 тыс. установок в мире, StorageTek контролирует более 70% рынка автоматизированных ленточных библиотек для Mainframe и более 35% рынка библиотек для открытых систем. При этом StorageTek, наверное, одна из немногих компаний – мировых лидеров в области информационных технологий, которая не очень хорошо известна в России. Во многом это обусловлено тем, что основная часть поставок оборудования StorageTek на российский рынок осуществляется на основе OEM-соглашений. Прямые партнерские соглашения со StorageTek имеют только две компании, представленные на российском рынке.

SN. Почему Ваша компания, системный интегратор, стала партнером StorageTek?

Л.В. Заключение партнерского соглашения между StorageTek и ТехноСерв А/С состоялось в 1996 году. Давайте вспомним, каково было в то время положение в области информационных технологий в мире и в России.

оритетам.” Критические данные – те, которые являются основными для ключевых бизнес-процессов, или те, которые определяются в таком качестве законодательством. Они необходимы, чтобы обеспечить минимально приемлемый уровень работы в случае катастрофы. Жизненные данные могут не требоваться немедленно, но могут обеспечить существенные инвестиции компании. Чувствительные данные – те, которые могут иметь альтернативные источники или могут быть восстановлены без значительных усилий, тогда как некритические данные – это очень легко восстанавливаем

В продолжение темы предыдущей статьи (см. стр. 16), редакция журнала обратилась с просьбой к компании ТехноСерв А/С (www.technoserv.ru) – единственной российской компании, являющейся авторизованным партнером StorageTek на рынках России и стран Ближнего зарубежья, дать свою оценку особенностям решений StorageTek, реализованных в России. На вопросы SN отвечает Леонид Винокуров, руководитель Департамента информационных технологий ТехноСерв А/С.

В начале 90-х годов в западной ИТ-индустрии происходит переход от сервероцентричной (ориентированной на серверы) модели – к инфоцентричной (ориентированной на данные) модели информационной системы. Именно корпоративные данные стали рассматриваться как консолидирующий, централизующий ресурс, причем не только из-за выросших объемов данных, но и из-за того, что многоплатформенность стала стандартом де-факто информационной системы. Ключевым элементом вычислительной инфраструктуры становятся системы хранения, как на дисках, так и на магнитных лентах, и как следствие – стремительное развитие этого ИТ-направления по всем параметрам.

С проблемой хранения данных и организации доступа к ним – проблемой сложной и многогранной, все чаще начинают сталкиваться и отечественные предприятия. По крайней мере, те из них, которые имели длительную историю и накопили данные за несколько десятков лет, или те, которые развивались наиболее динамично. При объемах корпоративных данных, достигающих 1 Тбайт, трудно обойтись традиционными решениями – требуются качественно новые решения.

Следует упомянуть еще один фактор. До середины 90-х годов на российском рынке были востребованы в основном малобюджетные проекты, связанные с секонд-хенд решениями. Именно такие решения позволяли бюджетам большинства предприятий. Затем происходит перелом, значительная часть предприятий начинает становиться в экономическом смысле “на крыло”, фактор стоимости ИТ-проекта перестает быть для них доминирующим. Симпатии заказчиков перемещаются от самых дешевых решений к решениям наиболее оптимальным.

Именно в это время ТехноСерв А/С включает целую серию стратегических соглашений с мировыми лидерами компьютерного рынка, в том числе, с компанией StorageTek – по автоматизированным ленточным библиотекам. И начинает поставлять на российский рынок те решения, которые были отработаны на Западе – решения по-настоящему мирового уровня.

Сегодня мы можем безо всяких преувеличений сказать, что те устройства и технологии, которые доступны западным заказчикам, доступны и любому российскому заказчику. Более того, наши заказчики, не обремененные грузом прошло-

мые или дублированные данные. И снова, ни автоматизированное резервное копирование на магнитную ленту и ни дисковые решения в одиночку не в состоянии предотвратить всех возможных неприятностей. Только их комбинация может отразить “падение” центра данных для большинства компаний.

И в заключение

“Россия быстро развивается, и много компаний уже имеют достаточно сложные информационные системы, — говорит Фэт . — И они уже должны уметь решать про-

блемы хранения, которые могут возникнуть. При этом необходимо учитывать уроки других организаций для того, чтобы с самого начала строить правильную инфраструктуру у себя. Не бойтесь говорить с вашими реселлерами, говорите с экспертами хранения об управлении вашей информацией в течение всего ее жизненного цикла, требуйте проведения социологических исследований, чтобы видеть, как другие преодолели подобные проблемы. И обращайтесь за консультациями.”

Аманда Бэйкер,
Компания “StorageTek”

го, часто предпочитают самые современные достижения ИТ. И сегодня можно констатировать, что российский ИТ-рынок является одним из наиболее инновационных в мире.

SN. В чем Вы видите преимущества прямого партнерства?

Л.В. Поставки по OEM-соглашениям являются совершенно нормальной практикой среди крупных западных компаний. В частности, с автоматизированными библиотеками StorageTek по такой схеме работают компании Sun Microsystems и Hewlett Packard, продвигающие их под своей маркой. Прямое партнерство с производителем позволяет нашей компании, во-первых, реализовывать решения, не предлагающиеся OEM-партнерами, а, во-вторых, обеспечивать более высокий уровень сопровождения проектов. Я имею в виду консультации, техническую экспертизу, сервис и т.д. Это, на мой взгляд, очень важный момент, особенно для российских заказчиков.

SN. Как развивалось сотрудничество между StorageTek и ТехноСерв А/С?

Л.В. ТехноСерв А/С последовательно реализует один из основных своих принципов – нашим заказчикам должны быть доступны самые современные технологии. С учетом этого принципа мы развили и наши взаимоотношения со StorageTek.

В течение 1997 года мы занимались в основном поставками больших роботизированных библиотек WolfCreek и PowderHorn для информационных систем на базе Mainframe. В 1998 году начались широкомасштабные поставки самых быстрых на тот момент 36-трековых лентопротяжек TimberLine и лентопротяжек RedWood, обеспечивающих хранение на одном картридже до 150 Гбайт информации – рекорд для того времени. Тогда же были выполнены первые успешные реализации многохостовых конфигураций. Ленточные библиотеки StorageTek становятся стандартным решением для всех наших крупных проектов. С появлением в 1998 году у StorageTek семейства компактных библиотек TimberWolf, удачного сочетания уникальных роботов StorageTek и носителей DLT-стандарта, мы начинаем продвигать это решение в проектах с использованием серверов открытых систем. В 1999 году нашей компанией был реализован один из первых в России проект по подключению библиотеки StorageTek в архитектуру SAN. Сейчас у нас на маркетинге весь спектр библиотек новой L-серии.

SN. Какие преимущества решений StorageTek Вы хотели бы отметить?

Л.В. Я отмечу то, на что в первую очередь обращают внимание наши заказчики: надежность, производительность и масштабируемость.

Решения StorageTek надежны потому, что хорошо разработаны и многократно

протестированы. Библиотеки рассчитаны на безостановочную работу в режиме 24x7x365, среднее время между отказами составляет 360 000 часов (более 40 лет непрерывной работы!).

Производительность библиотек StorageTek – одна из лучших в своем классе. Она обеспечивается механическим роботом уникальной конструкции, возможностью использования высокоскоростных драйвов, а также наличием FC-интерфейсов как у драйвов, так и у самой библиотеки.

Библиотеки StorageTek дают оптимальное решение для систем любой сложности и масштаба. Начав работу с минимально необходимой модели библиотеки, пользователь может наращивать ресурсы системы самыми различными способами: увеличением числа смесных носителей в одной библиотеке, установкой дополнительных драйвов или заменой их на более производительные модели, и, наконец, наращиванием конфигурации библиотеки дополнительными модулями хранения.

Особо хотелось бы подчеркнуть комплексность и открытость решений StorageTek. Автоматизированные библиотеки большой емкости и производительности предназначены для использования в составе сложных информационных систем – в гетерогенных операционных средах, с использованием различного программного обеспечения и с применением оборудования SAN различных производителей. И здесь особенно важна возможность совместимости библиотеки с другими элементами информационной системы.

Компания StorageTek имеет соглашения о совместимости с такими производителями компонентов вычислительной инфраструктуры, как Quantum Corporation, Brocade, CNT, Emulex Corporation, Network Appliance, Connected и др. И, что не менее важно, более 50-ти независимых производителей программного обеспечения сертифицировали для своих решений весь спектр библиотек StorageTek. В их числе – Veritas, Computer Associates, Legato, Oracle, Tivoli, SAP.

В этом смысле наши заказчики абсолютно защищены.

SN. Есть ли какие-либо особенности внедрения решений StorageTek у российских заказчиков?

Л.В. В первую очередь, это, конечно, масштабы проектов. Сегодня, по моим оценкам, стандартное решение для информационной системы масштаба предприятия – это 5-10 Тбайт ленточных ресурсов, что в несколько раз меньше, чем на Западе. Но, если вспомнить, что в 1997 году поставка решения объемом в один терабайт уже была событием, то мы видим, что масштабы российских ИТ-проектов растут очень быстро. Кстати сказать, это отразилось и на статусе нашей компании в рамках партнерской политики StorageTek. По итогам 2000 года ТехноСерв А/С был назван “Dis-

tributor of the year” в регионе Europe/Africa/Middle East, и с 2001 года работает со статусом “StorageTek Premier Partner”.

Второе, что следует отметить, это – структура проектов. До недавнего времени в подавляющем большинстве случаев решения StorageTek использовались в проектах по страховочному копированию. Страховочное копирование – традиционное и всем понятное использование автоматизированных библиотек. С ранних лет компьютерной эры резервные копии корпоративных данных хранились на магнитной ленте. Этот способ хранения по-прежнему остается одним из самых дешевых и самых надежных.

Но сегодня все больше появляется проектов по электронным архивам, позволяющим максимально автоматизировать хранение и поиск электронных объектов различной природы, проектов, реализация которых невозможна без использования автоматизированных библиотек. К таким решениям проявляют интерес предприятия государственного, телекоммуникационного, банковского и страхового секторов, а также нефтегазовой отрасли. Причем в последней наиболее интересен сегмент геологической разведки, где хранению подлежат поистине огромные массивы информации. Кроме того, есть горизонтальный слой заказчиков, без привязки к конкретной отрасли, – те, кто начал внедрение ERP-систем. У них также накапливается колоссальный объем информации по операциям электронного бизнеса.

Особый класс решений – это архивы мультимедийных данных. Такие архивы, например, в скором времени должны стать неотъемлемой частью информационных систем крупных медицинских учреждений, там, где объемы хранимой информации просто громадны. Ведь необходимо хранить сканированные изображения рентгеновских снимков, результаты медицинских исследований, истории болезни и т.д. К решениям такого рода проявляют интерес и некоторые отечественные телевизионные компании.

SN. Расскажите, пожалуйста, о самых интересных проектах с использованием продуктов StorageTek, выполненных Вашей компанией.

Л.В. Нашей компанией выполнено более 100 проектов с использованием продуктов StorageTek. Упомяну лишь несколько цифр по поставкам библиотек, характеризующих параметры наших проектов: 4 WolfCreek, 17 PowderHorn, 38 TimberWolf, более 50 библиотек L-серии (точнее можно будет сказать после подведения итогов 2002 года).

Многие из реализованных нами проектов интересны – и с точки зрения техники и технологии, и с точки зрения экономической выгоды. Многие из них заслуживают отдельного разговора.

SN. Спасибо за интересный разговор и позвольте в преддверии Нового года пожелать успехов Вам и процветания Вашей компании.