Визитка

АЛЕКСЕЙ БЕРЕЖНОЙ, системный администратор. Главные направления деятельности: виртуализация и гетерогенные сети. Еще одно увлечение помимо написания статей – популяризация бесплатного ПО

Резервное копирование

Теория и практика. Краткое изложение. Часть 3

В этой части статьи речь пойдет о практических нюансах резервного копирования, а также будут разобраны наиболее часто встречающиеся ошибки

В предыдущих частях [1, 2] были рассмотрены общие вопросы и технологические аспекты процедуры резервного копирования. Построение системы резервного копирования и обеспечение ее безотказного функционирования – чересчур сложная и ответственная задача, чтобы ее решить одним махом, лишь изучив техническую информацию, без какой-либо дополнительной подготовки. Вопросам грамотной организации сохранения информации и посвящена данная статья.

Методика организации резервного копирования

Прежде чем приступать непосредственно к организации резервного копирования, нужно выработать четкий план действий. Планирование должно охватывать весь процесс: от стадии проектирования системы до расписания замены съемных носителей. Это поможет избежать ненужных затрат еще на стадии приобретения оборудования, грамотно организовать процесс сохранения данных и обеспечить восстановление информации в допустимые сроки и без потерь.

Документировать систему резервного копирования нужно до ее создания, а не когда она уже есть и никто точно не знает, как она работает. Конечно, в каждой бизнес-среде есть свои нюансы, но существуют некоторые общие рекомендации, на которых мы и остановимся.

Таблица 1. Сводная таблица устройств и картриджей (кассет) стандарта LTO

Поколение	LTO1	LTO2	LTO3	LTO4	LTO5
Физическая емкость в Мб	100	200	400	800	1500
Примерная емкость при аппаратном сжатии в Мб	200	400	800	1600	3000
Скорость записи в Мб/сек	20	40	80	120	140
Поддерживает накопители следующих стандартов в режиме чтения-записи (RW)	LTO1	LTO2 LTO1	LTO3 LTO2	LTO4 LTO3	LTO5 LTO4
Поддерживает накопители следующих стандартов в режиме только чтения (RO)	LTO1	LTO2 LTO1	LTO3 LTO2 LTO1	LTO4 LTO3 LTO2	LTO5 LTO4 LTO3

1. Проведите аудит исходящих данных

Как много проблем удается избежать, если предварительно собрать сведения о том, что именно предстоит копировать и может потребоваться восстановить.

Необходимо четко знать, какие данные предстоит сохранять, на каких ресурсах они находятся, их объем, структуру, частоту обновления, обращения к файлам, а также такой несколько расплывчатый параметр, как важность для бизнеса. Например, корпоративная веб-страничка может содержать множество мелких файлов, архив бухгалтерской системы – всего один, но очень большой.

Конечно, бывает весьма затруднительно определить, насколько тот или иной файл важен для бизнеса. Как говорится, «какой палец не укуси, а все равно больно». Но в большинстве ситуаций вполне можно обозначить некоторые приоритеты. Например, среднестатистическая торговая компания без эскиза нового рекламного баннера как-нибудь выживет (или дизайнер нарисует еще один), а вот при потере всей бухгалтерской базы данных придется тяжко. Совершенно нелишним будет провести дополнительные консультации со специалистами или руководителями других подразделений. И, разумеется, при решении такого щепетильного вопроса последнее слово остается за руководством компании.

Всю собранную информацию необходимо занести в некий документ, который лучше оформить в виде таблицы. В итоге получаем свод исходных данных, из которого будет видно, откуда, в каком объеме, с какой частотой и в какое время нужно сохранять информацию.

2. Выполните анализ инфраструктуры

После процедуры аудита, описанной в предыдущем пункте, необходимо сосредоточить свое внимание на особенностях существующей ИТ-инфраструктуры.

Вот приблизительный перечень вопросов, на которые следует найти ответ:

- Сможет ли сеть или ее участок обеспечить требуемый поток данных за указанное время? Возможно, нужно приобрести новое оборудование или хотя бы использовать два интерфейса на сервере для удвоения скорости передачи?
- Выдержат ли дисковые подсистемы на серверах периодически возникающую высокую нагрузку при резервном копировании?
- Хватит ли быстродействия у серверов, чтобы обеспечить выполнение бизнес-задач в то время, когда запущены задания резервного копирования?
- Есть ли достаточный запас мощности электропитания? Не будут ли перегружены UPS? Сможет ли поставщик предоставить вам дополнительную мощность?
- Сколько имеется свободного места в серверных стойках для размещения дополнительного оборудования? На какой объем можно рассчитывать?
- Если вы собираетесь организовать передачу данных off-site (на хранилище за пределами предприятия) по сети, хватит ли ширины канала для копирования информации за указанное время?
- Если же данные будут перевозиться на съемных носителях, есть ли возможность доставить их обратно за требуемый промежуток времени? (На дорогах бывают пробки, а процедура получения требуемого носителя из места хранения вне офиса может занять какое-то время.)

По окончании такого анализа должно сложиться четкое представление о том, нужно ли приобретать что-то еще помимо непосредственно оборудования для резервного копирования, какие имеющиеся в наличии резервы можно задействовать для создания системы. Помимо этого, данный пункт позволит сделать приблизительный набросок схемы будущей системы сохранения данных.

Примечание: систему резервного копирования можно создать полностью своими силами или поручить компании – системному интегратору. И вопрос, с чего начнут свою работу приглашенные специалисты: будут ли проводить аудит или сразу станут навязывать уже готовое решение, является своего рода лакмусовой бумажкой, по которой можно судить о профессионализме и компетентности интегратора. В первом случае выбранной компании скорее всего можно довериться, во втором (когда без предварительного анализа предлагают что-то купить) – лучше не стоит рисковать.

3. Определите схему ротации

Имея на руках результаты аудита, уже можно определяться со схемой ротации данных. Почему это надо делать на первых шагах создания системы резервного копирования, когда даже не выбрано оборудование, не определен график резервного копирования и т.д.? Дело в том, что оборудование необходимо приобретать, руководствуясь схемой ротации, а не схему ротации подстраивать под существующее оборудование. В последнем случае можно получить множество проблем «на выходе».

Например, куплен одиночный ленточный накопитель с типовым набором картриджей, но при этом не все данные умещаются на один носитель, а полный бэкап возможно выполнить только в выходные, когда некому этот самый носитель заменить. Закупка ленточной библиотеки могла бы устранить проблему, но встает вопрос: «Куда девать одиночный накопитель, и зачем его вообще приобретали?»

Еще пример: в большинстве случаев выбирают схему «дед-отец-сын». Как я уже писал в [2], для полного резервного ежемесячного копирования нужно 12 носителей, для еженедельного – четыре, и для инкрементного по будням понадобится пять носителей. Но объем, записываемый на ленту, будет разный. Если для полной копии потребуется приобрести накопитель LTO4, то для записи инкрементной копии хватит и LTO3. Конечно, использование кассет меньшей емкости и, следовательно, менее дорогих – не самая большая экономия, но необходимо учесть, что при частом употреблении накопители быстрее выходят из строя. Картриджи, рассчитанные на большой объем данных, более критичны к условиям хранения и механическим повреждениям.

4. Составьте график резервного копирования: что, когда, за кем?

Имея на руках результаты аудита данных, подлежащих копированию, а также схему ротации носителей, можно определить график резервного копирования, в котором следует указать порядок и время сбора информации с серверов

Формат LTO Ultrium

Трудно представить ситуацию, когда каждый производитель аппаратного обеспечения для систем хранения данных начнет выпускать оборудование и сменные носители, несовместимые с продукцией других производителей. Помимо того, что это просто неудобно, такая обособленность в конечном счете может привести к неблагоприятным последствиям. Представьте, что у вас имеется архив резервных копий на ленточных носителях, но накопитель вышел из строя, аналогичные устройства уже больше не выпускаются. Именно поэтому технические средства, используемые для хранения данных, стандартизированы. LTO Ultrium – одна из наиболее известных спецификаций.

Сейчас уже выпущено пять поколений устройств и картриджей (кассет) стандарта LTO: начиная с LTO1 (уже не используется) и заканчивая LTO5. Данную линейку планируется развивать и дальше, то есть следует ожидать LTO6, LTO7. В каждом следующем поколении удваивается емкость носителя и в полтора-два раза увеличивается скорость записи. Исключение составляет только LTO5 (см. таблицу 1).

На практике нужно быть более осторожным при расчете количества носителей для записи требуемого объема данных. Указываемая в рекламных целях емкость с учетом сжатия является усредненной величиной. Например, файлы формата jpg, mp3, архивы типа rar, 7zip практически не сжимаются. Также следует учесть, что различные программы резервного копирования используют часть объема накопителя (до 20%) для записи служебной информации.

Обратите внимания на весьма примечательный факт: накопитель нового поколения может работать в режиме чтения-записи только с картриджами своей или предыдущей версии. С более старыми форматами такие устройства работают только в режиме чтения (см. таблицу). Например, накопители LTO4 могут считывать и записывать кассеты LTO4, LTO3 и считывать информацию с LTO2.

Помимо обычных перезаписываемых картриджей, существуют еще картриджи типа WORM (англ. Write Once, Read Many – картриджи со специальной электронной схемой, допускающей только однократную запись и многократное чтение) и картриджи для очищения считывающего устройства (UCC, англ.Universal Cleaning Cartridge).

на устройство резервного хранения. При этом учитываем, что различные участки ИТ-инфраструктуры в одно и то же время испытывают различную нагрузку. Поэтому вполне возможно, что у вас не будет единого «окна бэкапа», а только несколько небольших «окошек», в которые необходимо уложиться.

Особенно это актуально при создании полных резервных копий. Тщательно составленный график с учетом всех деталей позволит эффективно использовать время, отведенное для выполнения работ по сохранению данных и минимизировать затраты (например, избежать дополнительной модернизации сетевого оборудования, просто разнеся потоки данных во времени).

5. Создайте план восстановления

Даже самые свежие и полные резервные копии бесполезны без механизмов восстановления. Но мало иметь технические средства, необходимо знать, что и как сделать. Определиться с этим вопросом поможет соответствующий план.

Правильнее сказать, что таких планов у вас будет как минимум два: один на случай выхода из строя одного или нескольких серверов, другой для полного восстановления всей инфраструктуры, на случай самого тяжелого форс-мажора, например, крупного пожара.

Данные документы должны содержать детальную информацию о том:

- где хранятся резервные копии;
- какова процедура их получения (порядок обращения в службу безопасности, к руководству, правила получения ключей к шифрованным копиям и другую необходимую информацию);
- каким образом работает система восстановления, а также полное техническое описание этого процесса (включая такие аспекты, как восстановление данных из зашифрованных носителей).

Также нужно привести список ответственных лиц с необходимой контактной информацией, очередность восстановления серверов и других объектов инфраструктуры, требования к безопасности и т.д.

В итоге должно получиться некое руководство к действию, с помощью которого любой технический специалист, обладающий необходимой квалификацией, смог бы восстановить работоспособность сервисов и другую информацию, необходимую бизнесу.

6. Начертите будущую схему сети с учетом нового оборудования

В принципе этот пункт не является обязательным, но крайне желательно иметь такую схему. Она поможет охватить всю картину в целом и учесть дополнительные нюансы. Например, куда будет подключен новый сервер для резервного копирования с соответствующим дополнительным аппаратным обеспечением (например, ленточной библиотекой). Не помешает и чертеж размещения оборудования в стойке. Возможно, придется выполнить перекомпоновку, например, чтобы хватило длины SCSI-кабеля.

На общей схеме резервного копирования желательно указать следующую информацию:

- Расположение серверов с названиями, краткими характеристиками (название, IP, роль сервера, используемое программное обеспечение) ресурсов, подлежащих копированию.
- Методы соединения с системой резервного копирования (сеть, SCSI кабель и т.д.).
- Направление потоков резервного копирования с указанием дат, объемов, пропускной полосы и т.д.
- Более-менее детальное изображение системы резервного копирования с описанием ресурсов, указанием характеристик, таких как название сервера, IP-адрес, технические данные, например размер дисковых массивов, тип подключенного ленточного накопителя, а также используемое программное обеспечение.
- Метод передачи данных off-site (интернет-канал, носители) и дополнительные характеристики, например, ширина интернет-канала, расписание передачи.

7. Соберите как можно больше информации по оборудованию

Имея на руках результаты аудита, можно приступать к выбору оборудования. Но, чтобы он был правильный, необходимо проанализировать как можно больше сведений.

К сожалению, трудно выработать некий универсальный рецепт. Например, для системы, основанной на Windows Server 2008 с установленным Symantec Backup Exec, понадобится более мощный сервер, чем для аналогичного решения на базе FreeBSD с Bacula. Поэтому придется ограничиться некоторыми общими рекомендациями, представленными ниже.

Помимо таких тривиальных пунктов, как стоимость, условия по гарантии и т.д., вы должны знать ответы на следующие вопросы:

Технические аспекты общего плана: емкость, скорость чтения-записи, передачи данных по сети, интерфейс управления, взаимозаменяемость узлов и элементов.

О съемных носителях: сколько стоят картриджи, и какой у них срок службы. Придерживается ли производитель имеющихся стандартов. (Следует помнить, что от этого зависит не только показатель взаимозаменяемости картриджей и носителей различных производителей, но и уверенность в том, что такие носители будут выпускаться и после снятия оборудования с продажи.)

О дисковых подсистемах и сетевых хранилищах: возможно ли расширение дисковой подсистемы, и как это происходит. Какие жесткие диски и какого объема можно использовать.

Собранные данные по существующим устройствам лучше всего свести в один документ табличного вида, чтобы было проще анализировать преимущества и недостатки того или иного аппаратного решения.

8. Определите процедуры эксплуатации

Именно на этом этапе, когда еще ничего не куплено и не установлено, необходимо создать соответствующий документ (или целый ряд таких документов). Следует описать, какие работы требуется проводить для обслуживания системы резервного копирования. Необходимо предусмотреть решение таких вопросов:

- замена картриджей;
- контроль за использованием дискового пространства;

- контроль за своевременным запуском заданий и успешным завершением заданий;
- процесс перемещения данных за территорию периметра сети (off-site).

Сразу же надо определить ответственных за каждый пункт. При этом нужно быть готовым к выявлению интересных нюансов. Например, в удаленном филиале некому заменить ленточный картридж или невозможно организовать своевременную доставку носителей за пределы офиса, поскольку компания не обладает соответствующими транспортными средствами, или система корпоративной безопасности строго запрещает вынос любых носителей информации за территорию предприятия. Все это может наложить определенные ограничения не только при эксплуатации системы, но и повлиять на дальнейший выбор приобретаемого оборудования.

9. Выбор оборудования

Имея на руках оформленную таким образом документацию, задача подбора, закупки и монтажа оборудования не должна вызвать особых проблем. Как я уже упоминал выше, трудно составить некую универсальную инструкцию, которая подходила бы абсолютно для всех ситуаций и явилась бы панацеей от любых проблем. Но такое, увы, невозможно, поэтому дальнейшие шаги придется оставить на усмотрение читателя.

На что нужно обратить внимание:

Доступность. Вы должны быть уверены, что купленное аппаратное обеспечение, а также расходные материалы не являются редкими птицами в ваших краях. Иначе можно не только попасть на крючок к партнеру-монополисту, но и запросто остаться с неработающим решением в случае, если поставщику стало неинтересно работать с этим продуктом.

Масштабируемость. Всегда надо помнить о возможных изменениях, происходящих в компании. Поэтому желательно приобрести накопители с запасом, чтобы поддерживали запись на носители большей емкости, RAID-массив имел возможность гибкой перестройки на увеличенный объем при замене дисков, а приобретаемое программное обеспечение могло поддерживать растущую и усложняющуюся многоуровневую инфраструктуру в расчете на возможное открытие филиалов.

Работа над ошибками

Мы разобрали основные этапы проектирования и эксплуатации системы резервного копирования уровня предприятия. Теперь имеет смысл сосредоточиться на наиболее часто встречающихся упущениях.

Игнорирование проблемы роста объемов данных при закупке оборудования

Такую ошибку обычно допускают на стадии проектирования. Стремясь сэкономить некоторую сумму при закупке, параметры оборудования выбирают, что называется, впритык, чтобы удовлетворить сиюминутные нужды. Выше я уже писал о необходимости предусмотреть возможность масштабирования. Но иногда стремление сэкономить приводит к тому, что в расчет берутся только текущие нужды, без учета будущих изменений.

Приведу пример. При создании системы предполагалось, что для еженедельного полного копирования хватит съемного картриджа LTO3, а для промежуточной копии – LTO2. Поэтому был закуплен накопитель LTO3 с соответствующим набором кассет. Но за полгода прирост данных составил порядка 20% от заявленного объема. В итоге емкость существующих накопителей перестала удовлетворять. Возник выбор: приобрести еще один комплект носителей или новый накопитель, поддерживающий формат LTO4, и компенсировать недостаток объема дополнительной покупкой картриджей LTO4. В итоге было принято решение пойти по первому пути, как более экономичному на тот момент. Но через полгода объем данных возрос еще на 30%, и возникли проблемы с пропускной способностью, процесс полного резервного копирования перестал укладываться в установленные временные рамки. Так как накопитель был напрямую подключен к файл-серверу, на котором и концентрировалась большая часть пользовательских данных, то узким местом в схеме оказалась... скорость записи накопителя LTO3. И тогда все равно пришлось приобрести накопитель LTO4 с полным набором картриджей LTO4. Соответственно купленные ранее картриджи LTO2 можно было использовать только для чтения (см. [3]), а картриджи LTO3 – только для промежуточного бэкапа.

Разумеется, можно было избежать ненужных трат, если сразу проектировать систему и закупать оборудование с расчетом на перспективу. Оптимальный вариант: приобретать оборудование и расходные материалы с расчетом на удвоение объема данных в течение периода от года до двух лет.

Отсутствие регулярной проверки резервных копий

Эта ошибка наиболее часто встречается у начинающих системных администраторов. Суть проблемы в том, что даже при полной автоматизации процесса резервного копирования необходим постоянный контроль резервных копий. Не секрет, что любой процесс не гарантирован от случайных сбоев. Например, может оказаться нерабочим LTO-картридж. Или в связи с плохой работой накопителя, лента не будет качественно записана. В идеале необходимо провести полное восстановление содержимого носителя на идентичную тестовую инфраструктуру, а также сравнение по контрольным суммам, работоспособности сервисов и т.д. Но далеко не всегда бизнес готов поддерживать сложную и дорогостоящую процедуру. Чаще всего приходится ограничиваться автоматической проверкой контрольных сумм (такая возможность заложена во многих программах резервного копирования) при создании копии и восстановления нескольких файлов в другое местоположение в целях проверки целостности архива.

Но тем не менее такие тесты должны производиться регулярно. Например, при реализации месячной схемы «дед-отец-сын» проверке должны подвергаться все «деды» (ежемесячные резервные копии).

Также полезно время от времени тестировать резервные копии, находящиеся на хранении. Дело в том, что носители могут утрачивать свою работоспособность. Ведь магнитная лента подвержена размагничиванию, файловая система на дисковом хранилище может содержать ошибки. Регулярные проверки позволят вовремя определить про-

блемы и своевременно их устранить. Не забываем, что архив резервных копий – это своего рода история компании, а к истории нужно относиться бережно.

Отсутствие контроля за копируемым контентом

Еще одна частая ошибка. Данные имеют тенденцию постоянно меняться и, как правило, увеличиваться в объеме. Постепенное наращивание объемов обрабатываемой информации характерно для развития большинства компаний. Через какое-то время обнаруживается, что существующая система резервного копирования не справляется с возложенными на нее задачами. Гигабайты файлов и баз данных перестают помещаться на носителях. Особенно это характерно для дисковых хранилищ. В отличие от ленточных накопителей и библиотек, куда без труда можно вставить чистый картридж и продолжить запись, с дисковым RAID-массивом такой номер так легко провести не получится.

Помимо нехватки объема, существует и дефицит времени для выполнения заданий, которым не хватило отведенного периода («окно бэкапа»). Приходится задумываться о приобретении нового дорогостоящего оборудования для кардинальной перестройки всей системы резервного копирования.

Но есть и хорошая новость: данную проблему можно обойти или хотя бы отсрочить. Выход прост: нужно время от времени подвергать тщательному анализу данные, подлежащие резервированию. При этом рекомендуется запастись приличной долей скепсиса.

Например, безжалостно исключить из резервного копирования все мультимедийные файлы развлекательного характера, личные фотографии и остальное, что так дорого сердцу пользователя, но не задействовано в бизнес-процессах. Все это довольно часто можно обнаружить на общих ресурсах и в личных каталогах пользователей.

Файлы, к которым не обращались в течение длительного периода, имеет смысл перенести на отдельный съемный носитель и отправить в архив для хранения. Этот же рецепт подходит для оптимизации почтовых хранилищ, а также баз данных.

Такие простые меры позволят не только значительно разгрузить систему резервного копирования, но и помогут навести порядок на серверных ресурсах, повысив быстродействие.

Слишком долгий период между полным резервным копированием

Суть ошибки такова. В целях экономии полное резервное копирование выполняется один раз в течение длительного периода, допустим, раз в месяц. А для поддержания актуальности сохраняемых данных служит множество инкрементных или дифференциальных копий. Обычно такую странную схему создают из-за экономии средств на носители или для сокращения времени использования окна бэкапа. Все бы ничего, но при восстановлении приходится задействовать большое число инкрементных копий, что само по себе может значительно увеличить время «реанимации», а, следовательно, удлиняется пауза в работе бизнес-процессов. Некоторые системы резервного копирования, например, Acronis True Image, при повреждении одной из инкрементных копий перестают «видеть» и все последующие, созданные уже после поврежденной. Но самое страшное в этом случае – утрата единственной полной копии за требуемый период. Случись это с еженедельным бэкапом – бизнес потеряет информацию за семь дней, кроме того, возможно, удастся восстановить какие-то данные из инкрементных копий. Но в приведенном случае теряется вся работа за очень долгий период. Поэтому пользы от такого резервного копирования практически никакой.

Слишком много полных копий

Это другая крайность по сравнению с предыдущей ситуацией. Допустим, у администратора хватает и места на диске или ленте, и временного промежутка, чтобы выполнять полное резервное копирование. Он с удовольствием этот факт использует. Все бы ничего, но объем данных со временем обычно увеличивается. Растет и бизнес, появляются новые серверы в инфраструктуре, уплотняется рабочий график, все это неизменно ведет к тому, что для выполнения постоянного полного резервного копирования не хватает ни места, ни временного отрезка в «окне бэкапа». Я уже упоминал выше о необходимости постоянного мониторинга объемов информации, подлежащей сохранению. В описываемом случае объем исходных данных может быть не очень большим, но забивать носители для хранения резервных копий множеством мало отличающихся друг от друга версий вряд ли оправдано. Гораздо разумнее организовать инкрементное или дифференциальное резервное копирование, а еще лучше сделать это сразу, чтобы не изменять уже настроенную и работающую систему на ходу.

Много копий на одном носителе

Эта ошибка характерна для небольших компаний, где объем копируемых данных весьма незначительный. Возникает соблазн вместо ротации нескольких сравнительно дорогих носителей все записывать на меньшее число кассет. Проблема заключается в том, что при утрате одного, ставшего уже драгоценным носителя (например, при размагничивании, механическом повреждении, ошибках файловой системы на дисковом томе и т.д.) теряются резервные копии за весьма длительный период.

***

О резервном копировании можно говорить и писать очень долго. Нет сомнения, что эта сфера крайне важна для бизнеса. В то же время это такая область ИТ-технологий, в которой для достижения сходных целей можно использовать различные методы. Мы рассмотрели теоретические и практические основы резервного копирования, познакомились с технологическими аспектами, выработали примерный план внедрения, разобрали некоторые ошибки. Надеюсь, что этот материал поможет читателям создавать эффективные и надежные системы сохранения данных. EOF

1. Бережной А. Резервное копирование. Теория и практика. Краткое изложение. //«Системный администратор», №11, 2010 г. – С. 60-66.
2. Бережной А. Резервное копирование. Теория и практика. Краткое изложение. Часть 2. //«Системный администратор», №12, 2010 г. – С. 34-37.
3. Описание – http://ru.wikipedia.org/wiki/LTO.