

# Резервное копирование в АБИС

## Making Backup Copies Using the Automated Library Information System

### Резервне копіювання в АБІС

*А. И. Павлов*

*Государственная публичная научно-техническая библиотека СО РАН, Новосибирск, Россия*

*Alexander Pavlov*

*State Public Library for Science and Technology,  
Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia*

*О. І. Павлов*

*Державна публічна науково-технічна бібліотека СВ РАН, Новосибірськ, Росія*

Рассматриваются подходы к оценке технических решений по реализации подсистемы резервного копирования данных в автоматизированной библиотечно-информационной системе (АБИС).

Under consideration are approaches to the evaluation of technical solutions applied to the implementation of the backup copying subsystem of the automated library information system.

Розглядаються підходи до оцінювання технічних рішень з реалізації підсистеми резервного копіювання даних в автоматизованій бібліотечно-інформаційній системі (АБІС).

Рост информационного потока в последние годы во многом связан с качественным изменением состава передаваемых данных: кроме традиционных текстовых файлов, высокоскоростные каналы связи переносят огромное количество графической и мультимедийной информации. Это стало возможно в результате значительного подъема в технологии и производстве элементной базы микропроцессорных систем, одним из отражений которого стало появление в самом начале века нового поколения устройств хранения данных на жестких магнитных дисках – serial ATA<sup>1</sup> (SATA). Тогда это потеснило на рынке компьютерных систем SCSI<sup>2</sup>-устройства, традиционно используемые для построения серверных дисковых массивов [1]. На сегодняшний день развитие этого направления продолжается нарастающими темпами, что дает возможность даже средней библиотеке реализовать АБИС с относительно крупными дисковыми пространствами и достаточно адекватно реагировать на постоянный рост информационного потока, а это, в свою очередь, создает реальные предпосылки эволюции библиотек от традиционных **архивных хранилищ** к современным **информационным центрам**.

С технической точки зрения построение больших дисковых массивов для работы с данными на базе недорогих, но менее надежных компонентах (в сравнении со SCSI) требует дополнительных мер по защите информации от потерь. Вопрос сохранности данных никогда не оставался в стороне при проектировании и развитии АБИС в ГПНТБ СО РАН, однако прежде все ограничивалось описанием технологической процедуры [2] и констатацией аппаратных средств для реализации традиционного технического решения [1], и это было естественно, поскольку еще три, четыре года назад вопросы выбора носителя и устройства хранения практически не вызывали дискуссий среди технических разработчиков АБИС. Постоянно усиливающаяся экспансия SATA в секторе устройств дисковой памяти рынка компьютерных систем не только повысила актуальность резервного копирования, но и привела к тому, что сами SATA-устройства стали альтернативой решения данной технической проблемы. В настоящее время все чаще обсуждается вопрос принципиальной реализации системы резервного копирования, поскольку теперь наряду с традиционно используемыми для этих целей стримерами, все чаще предлагаются новые решения резервирования – на

<sup>1</sup> ATA (Advanced Technology Attachment) – серия интерфейсов и протоколов, используемых для организации доступа к жестким дискам и устройствам памяти на сменных носителях (CD, DVD, MOD и т.п.) в компьютерах.

<sup>2</sup> SCSI (Small Computer System Interface) – серия интерфейсов и протоколов, используемых для подключения дисковых массивов и отдельных периферийных устройств.

жёстких дисках. Развитие идет в обоих направлениях достаточно интенсивно, что приводит к многообразию устройств и решений резервного копирования с различными техническими характеристиками, предоставляя потребителю широкий выбор, но и делая эту задачу более сложной.

В целом тема резервного копирования чрезвычайно обширная, и в данной работе коснемся ее в той части, которая наиболее присуща средним и относительно крупным библиотекам. Ранее [1] были определены три основные технологические задачи администрирования данных: *доступ*, *управление* и *хранение*, и предложены способы оценки их технической реализации, тогда первым двум задачам было уделено достаточно внимания [1,3]. Аналогично целью данной работы является количественные оценки эффективности в отношении систем резервного копирования, на основании тех же критериев:

- надежность,
- стоимость,
- масштабируемость,
- быстродействие,
- функциональность,

как это было сделано в отношении других технических подсистем АБИС. Хотя это не принципиально, критерии перечислены в порядке их значимости, в этом порядке и рассмотрим, что конкретно будет пониматься под этими определениями для системы резервного копирования.

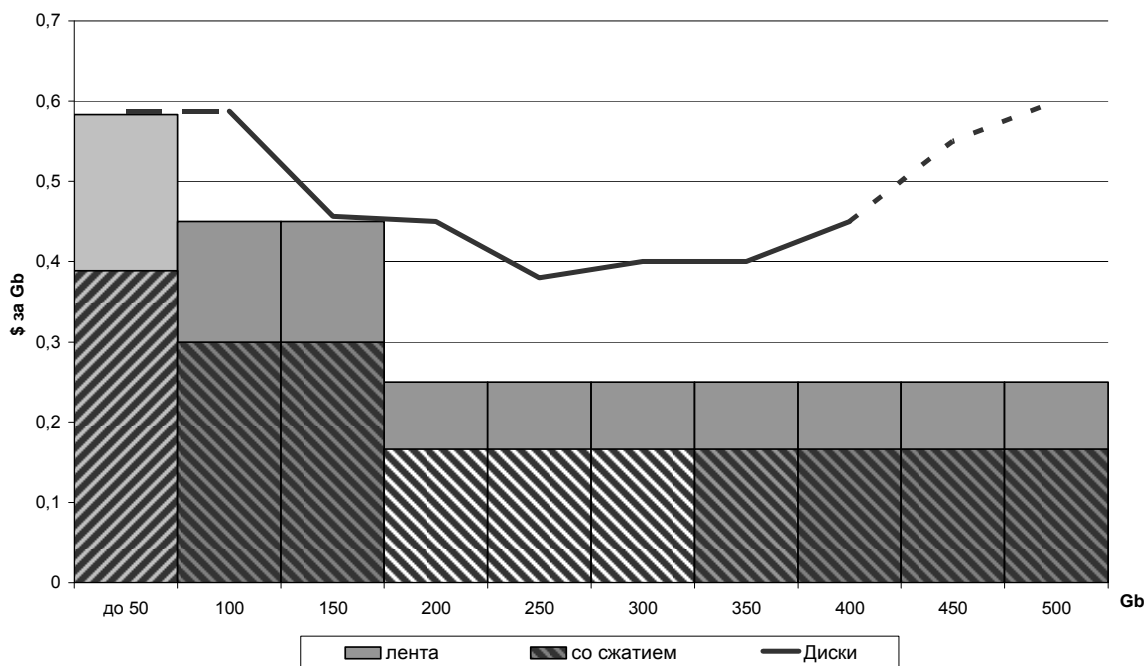
На верхней ступени (наиболее значимой) находится *надежность*, что естественно для системы обеспечения сохранности информационных данных. Чаще всего приводится следующая техническая характеристика на подобные устройства – это время наработки на отказ или среднее время между отказами (MTBF – Mean Time Between Failures), которую можно использовать в качестве количественного эквивалента данного критерия. Существуют и другие характеристики, например: рабочий цикл (Duty Cycle). И хотя его численное значение часто приводится и для дисков, и для лент, однако для каждого типа устройств в этот параметр вкладывается разный физический смысл, поэтому эту характеристику резонно использовать, сравнивая устройства между собой внутри одного класса. Практически, современные технологии производства устройств хранения информации достигли того уровня, при котором (теоретически) моральное старение аппаратных средств наступает гораздо раньше их физического износа. Так значения MTBF для устройств выбранного класса, составляют от 250 тысяч до миллиона часов и более, при этом как следствие *закона Мура*<sup>3</sup> каждые два года происходит двукратное улучшение основных технических характеристик микропроцессорных устройств. То же самое можно сказать об уровне надежности носителей хранимых данных: постоянно совершенствуются современные материалы, применяемые при их изготовлении, повышая одновременно и плотность записи, и сроки эксплуатации. И сами устройства используют различные технологии и методы записи информации, например, избыточность кодов, продольный и поперечный контроль, чтение после записи и т. п., которые позволяют восстанавливать данные даже при сбоях или механических повреждениях носителя. Дисковые накопители всегда имеют преимущество по этому параметру по сравнению со стримерами.

На втором месте – *стоимость*. Чтобы адекватно сравнивать различные технические устройства и физические носители, в качестве численного эквивалента этому параметру выберем стоимость единицы хранения определяемую как отношение цены носителя (картриджа или диска) к его информационной емкости. Для определенности цена выражена в долларах США, а объем – в гигабайтах (Гб). Приведенный график, зависимости этой характеристики от информационной емкости носителя, демонстрирует преимущество ленточных картриджей над дисковыми носителями по этому критерию почти во всем диапазоне размерностей. При этом стоит отметить, что даже небольшое отличие в стоимости единицы хранения (Гб) может привести к значительному суммарному экономическому эффекту, так как общий объем информационного ресурса для выбранного класса АБИС порою достигает нескольких тысяч единиц. Более того, система априори предполагает рост общего дискового пространства, а, следовательно, и усиления этого экономического эффекта в будущем.

---

<sup>3</sup> Доктор Гордон Мур (G. Moore) – один из основателей корпорации Intel по его утверждению: *уплотнение транзисторов на микросхемах позволяло регулярно (примерно раз в 2 года) удваивать производительность.*

**Изменение относительной стоимости в зависимости от информационного объема единицы носителя**



На графике можно выделить три характерных области: первая, где стоимость единицы хранения на стримере и диске, практически, равны, средняя область говорит о незначительных расхождениях в этой характеристике, и есть третья область, где удельная стоимость для лент сохраняет свое значение, а для дисковых носителей данный параметр уже начинает расти. Исходя из этого, можно утверждать, что в первой области предпочтительнее применение дисковых систем, а второй и особенно третий участок, напротив, говорят об экономическом преимуществе решений на основе стримера.

На третье место в данном случае помещена оценка *быстродействия*. В качестве численного эквивалента для этого критерия естественно выбрать скорость обмена данными (transfer rate) между устройством резервного копирования и компьютером (точнее серверным дисковым массивом). В общем случае цикл записи и чтения складывается из поиска, чтения, передачи по интерфейсу и записи. Это усредненное время для дисков (в нашем случае) определяется пропускной способностью интерфейса (SATA), для стримеров как устройств последовательного доступа это время рассчитывается с учетом среднего времени поиска (позиционирования). По этому параметру стримеры всегда уступали дисковым носителям (в одном ценовом классе), но часто это может быть их единственным минусом. При этом абсолютные значения скорости передачи данных у ленточных накопителей постоянно растут, в большинстве своем, обеспечивая допустимое время копирования. Резервное копирование в традиционной технологии предполагает, если не полную остановку системы для пользователя, то некоторый дискомфорт в работе: задержки в реакции на запрос, блокировку (временная недоступность) всех или части данных и т. п., поэтому, чем короче будет процедура записи архива, тем технологичнее выглядит сама система. Вообще резервное копирование можно разделить на полное (full backup<sup>4</sup>), нарастающее (incremental backup<sup>5</sup>) и дифференциальное (differential backup<sup>6</sup>). Полное копирование выполняется достаточно редко – это касается наиболее стабильной части информации (операционная система, приобретенные базы данных и т. п.), операция долговременная, выполняется в отсутствие пользователей, обычно во время регламентных работ. Часто (ежедневно) изменяемая информация в таких системах как АБИС составляет

<sup>4</sup> <http://www.tech-faq.com/lang/ru/full-backup.shtml>

<sup>5</sup> <http://www.tech-faq.com/lang/ru/incremental-backup.shtml>

<sup>6</sup> <http://www.tech-faq.com/lang/ru/differential-backup.shtml>

небольшие объемы (от 0,5 до 2% общего размера массива). При чем, этот процент с ростом общего объема имеет тенденцию к уменьшению, и такое копирование выполняется ежедневно обычно в ночное время, что минимизирует неудобства пользователя. Накопительный архив создается из дневных резервных копий на протяжении периода времени в соответствии установленному организацией регламенту (неделя, декада, месяц и т. д.)

Под **масштабируемость** для резервного копирования будем понимать способность подсистемы в перспективе решать проблемы, связанные с ростом информационных массивов, то есть инвестиции на создание системы архивного копирования существующих объемов информации должны быть рассчитаны на достаточный рост массива данных, желательно, со снижением стоимости единицы хранения. Из этого видно, что данный параметр очень близко связан с ценовым критерием, однако в каждом конкретном случае может быть по-разному оценен. Вообще стримеры всегда имели более выигрышную позицию по этому критерию за счет и цены, и простоты технического решения, поскольку являются устройствами со сменным носителем и изначально проектируемые как устройства резервного копирования.

**Функциональность** трудно сопоставить с каким-либо числовым эквивалентом, в данном случае будем понимать под этим существующий уровень или степень автоматизации процесса архивирования, что имеет описательный характер для класса рассматриваемых систем. Поскольку магнитные ленты традиционно использовались для архивного хранения, стримеры находятся несколько впереди по отношению к дисковым системам. Процедуры резервирования и восстановления понятны, практически всё здесь стандартизировано, а программное обеспечение обладает богатым набором функций. Стример легко устанавливается в систему, при этом обычно используется интерфейс SCSI. Оператору (в классе определенном нами систем) остается своевременно менять компактные картриджи, которые удобно хранить в сейфе или перенести в другое место, поскольку они не так, как жесткие диски чувствительны к механическим воздействиям.

Ниже в таблице сведены все характеристики для различных величин рабочих информационных массивов. То, что оценка системы резервного копирования не может быть сделана без учета общего информационного объема АБИС интуитивно понятно и, собственно, не требует каких-то особых пояснений или доказательств, принимаем эту зависимость как основную посылку (**постулат**).

Таблица

Критерий оценки	Менее 1 Тб		1-2 Тб		3-5 Тб		Свыше 5 Тб	
	НМД	НМЛ	НМД	НМЛ	НМД	НМЛ	НМД	НМЛ
Надежность (MTBF тыс. час.)	<b>600</b>	<b>250</b>	<b>600</b>	<b>250</b>	<b>600</b>	<b>300</b>	1000	400
Условная стоимость одного гигабайта (\$)	<b>0.6-0.45</b>	<b>0.6-0.45</b>	<b>0.45-0.4</b>	<b>0.45-0.2</b>	<b>0.4</b>	<b>0.2</b>	0.6-0.45	0.25
Быстродействие (трансфер Mb/s)	<b>155</b>	<b>7,5</b>	<b>155</b>	<b>7.5-30</b>	<b>155-300</b>	<b>30-80</b>	300-600	80-160
Масштабируемость	+	+	±	+	±	+	+	+
Функциональность	+	+	-	+	±	+	+	+

- плохая,  
± удовлетворительная,  
+ хорошая.

Кроме того, можно предложить еще ряд эмпирических зависимостей, пользуясь которыми в совокупности с определенными выше критериями, можно рассчитать диапазон требуемого объема носителя и на основании его оценить проектируемую или существующую систему резервного копирования.

1. Системы резервного копирования в любом случае довольно дорогостоящие, поэтому должны рассчитываться не менее чем на пятилетний срок эксплуатации.

2. Рост информационного пространства библиотеки является результатом развития и автоматизации внутренних технологических процессов, и динамика этого роста хорошо согласуется с упомянутым **законом Мура**.
3. Ежедневное<sup>7</sup> (differential backup) копирование не превышает 1-2% общей величины информационного массива, и как следствие, примерно **10%** от той же величины будет определять информационный размер носителя, подходящего для ведения incremental backup.

Данные закономерности позволяют сузить область выбора технических средств для реализации системы резервного копирования, следующим образом: зная существующий размер дискового пространства, по закону Мура экстраполируем его величину на заданный период (5-6 лет), затем определяем размер носителя (5-10% общего объема на перспективу 3-5 лет) и на основании этой величины выбираем в доступном ценовом диапазоне ту или иную систему (см. график).

Отметим, что АБИС рассматриваемого класса соответствует именно центральной части графика и, соответственно, таблицы тоже. Это вызывает необходимость более тщательно рассматривать все критерии, включая и неформализованные характеристики, поскольку в данном диапазоне выбор может быть сделан как в пользу традиционных стримерных, так и новых дисковых решений. В этом ключе стоит отметить, что в отличие от дисков, которые в традиционном варианте используются практически без дополнительных затрат, системы со сменными носителями требуют приобретения самого устройства (стримера), цена которого достаточно высока и может быть компенсирована за счет относительно низкой стоимости ленточного носителя. Нетрудно просчитать экономию, зная существующий объем дискового массива (информационного пространства), выбранный регламент на создание резервных копий и разницу в стоимости единицы хранения.

Обзор некоторых источников [4,5] и собственный опыт разработки и эксплуатации системы резервного копирования в ГПНТБ СО РАН позволяют утверждать, что применение стримера на сегодняшний день может быть эффективно в случаях когда:

- хранение связано с транспортировкой архивных данных (перевозкой);
- очень критичное отношение геометрическим объемам (для хранения);
- оперируют с большими массивами информации (3-5 Тб);
- роль стримеров в основном сводится к созданию полных копий (full backup).

Первые три пункта уже рассматривались и пояснялись выше. По четвертому пункту можно подчеркнуть следующее: хорошо зарекомендовала себя гибридная схема резервного копирования, когда стример используется для создания архивов и накопительных копий, а ежедневное копирование (изменений) ведется на магнитный диск. В этом случае максимально используются достоинства обеих систем: быстрое копирование дневных изменений, которое тем более выполняется в ночное время, практически, остается незаметным для пользователя, как и накопительное копирование, которое теперь может быть выполнено в любое время. Сохраняется относительно малый объем архива. К единственному недостатку можно отнести необходимость создания (или использования) не стандартизованного для дисковых систем программного обеспечения, определяющего необходимую для дифференциального копирования информацию (см. таблицу).

Говоря о перспективах технического резервного копирования для систем выбранного класса, можно отметить в первую очередь то, что при сохранении тенденции развития обоих направлений (стримеров и магнитных дисков), становится все более неэффективно экономически организовывать подсистему резервного копирования на стримере, если общее дисковое пространство на настоящий момент менее 5 Тб. При этом резервную копию (differential, а возможно и incremental backup) резоннее создавать на дисковых устройствах с «горячей» заменой (hot swap), а реализацию полных копий правильнее с экономической точки зрения организовать на корпоративной основе использования стримера. Во-вторых, если сохранятся динамика развития дисковых носителей, а снижение цен на стримеры и соответственно ленточные носители не пойдет опережающими темпами, то будущее за системами резервного копирования на основе жестких магнитных дисков. В-третьих, стоит отметить, что процесс вытеснения традиционных систем резервного копирования может еще более ускорить появление на рынке компьютерной техники оптических устройств и носителей повышенной (1 и более Тб) емкости и достаточно высокого (теоретически 960 Мб/с)

<sup>7</sup> Согласно выбранному на предприятии регламенту.

быстродействия [6,7]. При таких технических параметрах в случае массового производства (и соответственно приемлемой стоимости) это решение станет реальной альтернативой стримерам и в секторе архивного копирования (full backup), где пока их позиции наиболее сильны.

### Литература

1. Павлов А. И. Технические решения формирования информационного пространства крупной библиотеки // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса [Электронный ресурс]: 11-я Междунар. конф. «Крым 2004». – Судак, 2004. – 1 электрон. опт. диск (CD-ROM)
2. Баженов С. Р. Использование электронных информационных ресурсов в крупной библиотеке и проблемы обеспечения их сохранности // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса [Электронный ресурс]: 11-я Междунар. конф. «Крым 2004». – Судак, 2004. – 1 электрон. опт. диск (CD-ROM)
3. Павлов А. И. Тенденции развития материально-технической базы библиотеки в условиях современных информационных технологий // Электронные ресурсы региона: проблемы создания и взаимоииспользования: Материалы науч. -практ. конф. (Новосибирск, 25-28 окт. 2004 г.). – Новосибирск, 2005.– С. 221–229.
4. Чеканов Д. Системы резервирования данных: четыре альтернативы стримерам // Tom's hardware guide, [www.thg.ru](http://www.thg.ru) [Электронный ресурс] режим доступа: [http://www.thg.ru/storage/20030226/tape\\_drives-13.html](http://www.thg.ru/storage/20030226/tape_drives-13.html) – 26. 02. 2006
5. Залужный Д. Системы хранения данных на магнитных лентах // Тематический портал о компьютерах, [www.hardvision.ru](http://www.hardvision.ru) [Электронный ресурс] режим доступа: [http://www.hardvision.ru/to\\_articles/act\\_desc/id\\_256/?dir=storage&doc=hp\\_tape](http://www.hardvision.ru/to_articles/act_desc/id_256/?dir=storage&doc=hp_tape) – 03.05.2005.
6. Володин В. Первые голографические диски //Техинфо [Электронный ресурс] режим доступа: [http://stolica.ru/techinfo/article/2006\\_03/31\\_01.htm](http://stolica.ru/techinfo/article/2006_03/31_01.htm) – 31. 03. 2006.
7. Лишин Л., Лишин И. Архивы. Критерии выбора // Официальный сайт «Издательство 625» [Электронный ресурс] режим доступа: <http://rus.625-net.ru/625/2006/08/tehnol02.htm> – №8 2006.