

**Анализ информационного массива АБИС
для оптимизации подсистемы резервного копирования**

**Analyzing ALIS Information Array
to Optimize Backup Subsystem**

**Аналіз інформаційного масиву АБИС
для оптимізації підсистеми резервного копіювання**

И. С. Баженов, А. И. Павлов

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

I. S. Bazhenov and A. I. Pavlov

*State Public Scientific and Technological Library of the Siberian Branch of the Russian Academy
of Sciences, Novosibirsk, Russia*

I. С. Баженов, А. І. Павлов

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

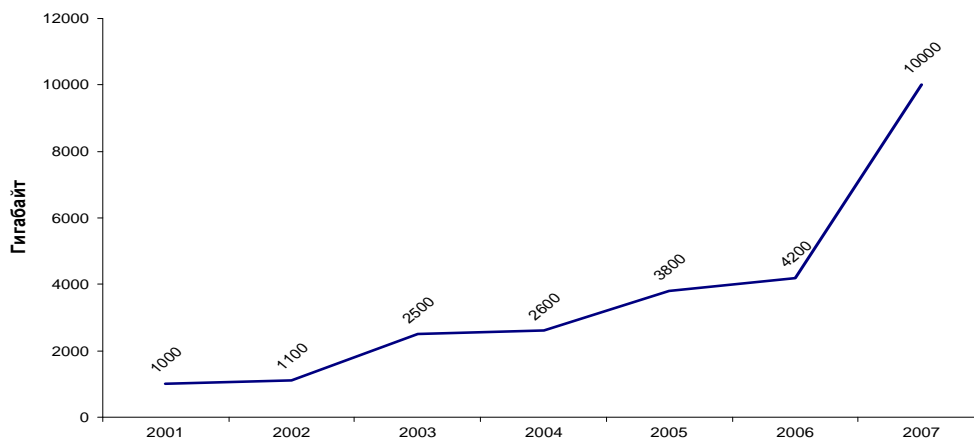
В докладе проводиться аналіз складу інформаційного ресурсу автоматизованої бібліотечно-інформаційної системи (АБИС) для оптимізації технічних рішень, що реалізують підсистему архівного копіювання даних.

The authors analyze the content of ALIS information resource seeking to optimize technological solutions for archival backup subsystem.

У доповіді проводиться аналіз складу інформаційного ресурсу автоматизованої бібліотечно-інформаційної системи (АБИС) для оптимізації технічних рішень, що реалізують підсистему архівного копіювання даних.

Процесс развития АБИС наиболее ярко проявляется в увеличении объемов хранимой информации. В свою очередь рост информационного ресурса требует своевременной поддержки соответствующего уровня системы резервного копирования, поскольку размер дискового пространства (объем хранимой информации) предъявляет соответствующие технические требования к этой системе по скорости, надежности, технологичности, возможности масштабирования и конечно стоимости. Так в ГПНТБ СО РАН по мере развития АБИС, дважды модифицировалась подсистема резервного копирования, последний раз это было сделано в 2001 г. На графике отражена динамика роста дискового массива уже за период работы последней модификации данной подсистемы.

Темпы роста дискового пространства ГПНТБ СО РАН



На начало года дисковый массив вырос более чем в четыре раза, а перспектива более чем в десять. Такие темпы роста усилительно наталкивают на мысль о необходимости очередной модернизации подсистемы архивирования, что всегда связано с достаточно серьезными финансовыми затратами. Выбранная ранее модель стримера (HP Ultrium 215) относится к линейке устройств Ultrium – это современное техническое решение, постоянно развиваемое и сохраняющее поддержку старших форматов носителей. То есть ленты, записанные на ранее используемом носителе, могут использоваться (по крайней мере, читаться) на более новых моделях устройств Ultrium, что освобождает от необходимости приобретения ленточных картриджей нового типа и перезаписи старых архивов в связи с переходом на более производительную модель стримера. Это позволяет экономить определенные денежные средства, однако стоимость новых моделей стримеров достаточно высока (от 2, 5 до 4, 5 тыс. USD), поэтому естественно, что продление жизненного цикла предыдущей модели дает положительный экономический эффект.

Продления срока службы стримера можно добиться, понизив нагрузку на него или, что то же самое, уменьшив поток сохраняемой информации. Естественно, такие действия не должны быть связаны с увеличением риска потери данных. Все должно строиться на анализе структурных компонент информационного массива АБИС с точки зрения *периодичности и доли*, вносимых в них изменений, а так же *их ценности (уникальности)*. С другой стороны по своему функционально-технологическому назначению информацию, используемую в АБИС можно разбить на три группы:

1. Операционная система (сетевая ОС)
2. Приобретаемые БД.
3. БД собственной генерации, документооборот.

Операционная система серверов

Данная информация в основном изменяется очень редко. Обычно изменениями являются установка нового устройства, регистрация новых пользователей и другие мелкие работы в системе. Однако в системе ведутся журналы событий, и основным из них является журнал, отражающий события, связанные с безопасностью (попытки входа/выхода пользователя, удаление или модификация рабочих файлов и т. д.). Этот журнал необходим для контроля над внутренней безопасностью, он заполняется очень быстро, и для сохранения его в полном объеме для дальнейшего исследования копируется ежедневно. Остальные файлы операционной системы сохраняются путем создания образа файловой системы (full backup) по необходимости (на усмотрение системного администратора). При этом восстановление операционной системы в аварийной ситуации занимает время, соизмеримое с физическим копированием файлового образа.

Приобретаемые и полнотекстовые базы данных

Этот вид информации крайне редко подвергается модификации, которая обычно заключается в поступлении нового выпуска в виде дополнительных каталогов и файлов и/или актуализацией существующего массива.

Массив сохраняется полностью, затем обновляется примерно раз в месяц (периодичность поступление новых выпусков) или по требованию (произведено большое количество изменений). При наличии лент, редком изменении и небольшом пополнении данная информация может храниться без перезаписи в течение одного года, по окончании этого срока данный массив полностью переписывается, невзирая на то, вносились какие-либо дополнения или изменения или нет.

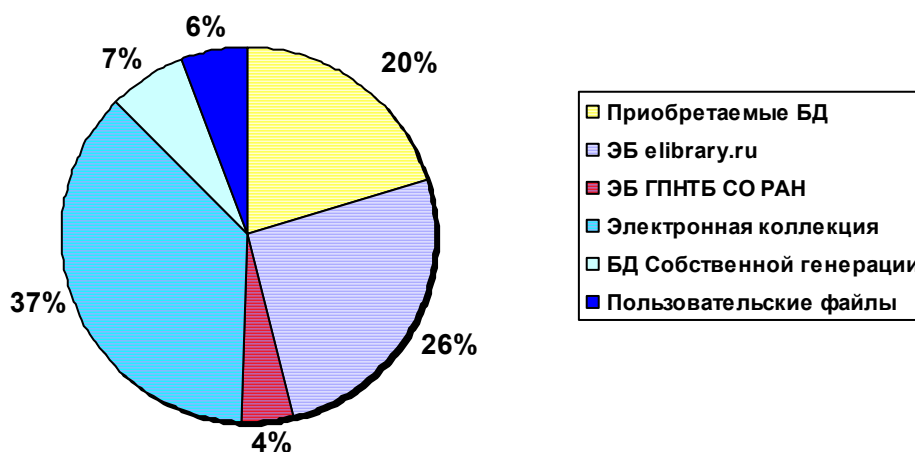
Базы данных собственной генерации и пользовательские файлы

В эту категорию входят электронные каталоги библиотеки, библиографические базы данных и т. п., а также всевозможные документы и медиафайлы общего пользования. Данная информация, меняется ежедневно, и копируется с такой же периодичностью в ночное время в режиме *обновления* на дисковый носитель отдельным томом и одновременно дублируется на магнитную ленту, но уже в режиме *накопления*. По окончании квартала процесс повторяется: начинается новый цикл с формирования *полной* копии изменяемой части информационного массива (см. диаграмму) и затем начина-

ется процесс формирования следующей квартальной резервной копии добавлением ежедневных изменений, цикл повторяется по окончании каждого квартала в течение года. Таким образом, так называемая схема «дед-отец-сын» сохраняет предыдущее накопление сроком до одного года и гарантирует постоянный охват ретроспективы на срок не менее девяти месяцев. То есть может быть восстановлена информация практически на любую дату ретроспективной глубины до года.

К этой же категории резонно отнести и **бухгалтерскую информацию**, которая состоит из баз данных «1С:Бухгалтерия» и «1С:Зарплата», различных программ подготовки отчетности (налоговой и внутренней) и разнообразные текстовые документы. Эти данные также постоянно меняются, являются уникальной информацией, которая, кроме того, характеризуется повышенной важностью и секретностью. Все файлы, кроме баз данных, относящиеся к бухгалтерской информации копируются, как и обычные файлы пользователей, хранятся также не менее девяти месяцев, но при этом дополнительно подвергаются шифрованию. Сами же бухгалтерские базы данных копируются каждые три часа во временные дисковые архивы, и раз в сутки на магнитную ленту, и также сохраняется ежедневная дисковая копия для дальнейшего скоростного восстановления в случае необходимости. Эти архивы хранятся не менее трех лет с возможностью полного восстановления на любую дату.

В целом на сегодняшний день относительное распределение объемов дискового пространства между различными информационными ресурсами ГПНТБ СО РАН выглядит следующим образом:



Как видно из диаграммы, массив с часто изменяемыми данными занимает всего 13% от общего информационного пространства (не заштрихованные сектора). При создании ежедневной резервной копии только он подвергается программному сканированию на предмет поиска модифицированных файлов, и соответствующие элементы копируются на диск (дифференциальное резервное копирование – differential backup). Практически эта величина составляет 4-10Гб, что на сегодняшний день значительно меньше одного процента общего информационного пространства АБИС ГПНТБ СО РАН.

Копирование на жесткий диск в режиме обновления происходит со значительно большей скоростью, чем на ленточный носитель. Но (в определенном временном диапазоне) скорость важна лишь для сохранения ежедневных изменений, поскольку дублируемые файлы могут блокироваться на время создания копии. Накопительный архив может вестись параллельно с рабочим режимом, практически не влияя на производительность системы. Зато в случае повреждения носителя более вероятно восстановление данных с ленточного накопителя, так как у него отсутствует файловая система, и, соответственно, таблица размещения файлов (FAT). Данные располагаются последовательно, и их структуры описания находятся в непосредственной близости по ленте. При повреждении ленты произойдет потеря лишь испорченных файлов, а случае же жесткого диска при повреждении FAT существует вероятность полной потери данных, или это повлечет за собой очень трудоемкое восстановление информации (чаще всего лишь частичное).

Все вышеописанное архивирование ведется на ленточный носитель (*картридж*) вместимостью 100 Гб (с заявленным сроком сохранности данных – 30 лет), при этом поддерживается аппаратное сжатие, за счет чего вместимость может возрасти до 300 Гб. Большему сжатию поддаются базы

данных в форматах ISIS, WinIrbis и текстовые форматы. Практически не сжимаются полные тексты в форматах PDF или TIFF. Хорошо сжимаемая информация более характерна для ежедневного и соответственно накопительного копирования, а полные тексты – для условно-постоянной части информационного ресурса. Практически, среднее значение информационного объема на единицу носителя составляет 175–180 Гб.

Существующие режимы архивирования позволяют восстановить информацию по требованию в следующих вариантах:

- восстановление удаленного или испорченного файла;
- восстановление нескольких версий одного и того же файла;
- восстановление полной копии на любую дату,
- восстановление образа операционной системы.

Таким образом, анализ информационного потока, позволяет сделать вывод, что существующее аппаратно-программный комплекс на основе недорогого стримера, способен еще достаточно долгое время обеспечивать функциональную полноту подсистемы резервного копирования, как с точки зрения пользователя, так и администратора АБИС.