

**Модернизация аппаратно-программного комплекса обеспечения сохранности данных в связи с внедрением технологии обслуживания читательского заказа в автоматизированной системе «ИРБИС»**

**Modernizing Hardware and Software Complex of Data Safekeeping in Connection with Implementing Technology of Serving Readers' Applications in IRBIS Automated System**

**Модернізація апаратно-програмного комплексу забезпечення збереження даних у зв'язку із запровадженням технології обслуговування читачького замовлення в автоматизованій системі «ІРБІС»**

*А. И. Павлов, И. С. Баженов, Е. Б. Грешнов*

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

*Alexander Pavlov, Igor Bazhenov, and Evgeny Greshnov  
State Public Library for Science and Technology. Siberian Division  
of the Russian Academy of Sciences, Novosibirsk, Russia*

*О. І. Павлов, І. С. Баженов, Є. Б. Грешнов*

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

В докладе рассматривается изменение требований к защите данных при внедрении технологии обслуживания читательских заказов в автоматизированной системе «ИРБИС», когда поступление читательских требований происходит в режиме реального времени, и при этом стохастическим образом изменяется ряд информационных массивов, включая базу данных электронного каталога (ЭК). Соответственно, существующая регулярная система резервного копирования не обеспечивает создание репликаций в темпе поступления заказов и, следовательно, не может гарантировать требуемый уровень защиты данных. Приведен вариант решения данной проблемы посредством двухкомпонентной системы резервного копирования.

The paper covers the change in demands for data protection when implementing the technology of processing readers' orders in IRBIS Automated System. Readers' orders arrive in real time mode, so the range of information files, including the database of the e-catalog, changes stochastically. As a result, the existing regular backup system cannot ensure the production of replications at a pace with receiving applications, and cannot guarantee the necessary level of data protection. A way of solving this problem using two-component backup system is presented.

У доповіді розглянуто зміну вимог до захисту даних під час запровадження технології обслуговування читачьких замовлень в автоматизованій системі «ІРБІС», коли надходження читачьких вимог відбувається в режимі реального часу, і при цьому стохастичним чином змінюється ряд інформаційних масивів, включно з базою даних електронного каталогу (ЕК). Відповідно, існуюча регулярна система резервного копіювання не забезпечує створення реплікацій у темпі надходження замовлень і, відповідно, не може гарантувати той рівень захисту даних, який вимагається. Наведено варіант вирішення даної проблеми засобами двокомпонентної системи резервного копіювання.

С первых шагов создания автоматизированной библиотечно-информационной системы (АБИС) в ГПНТБ СО РАН особое внимание уделялось системам и технологиям сохранения данных. Тема весьма широкая и всегда актуальна для библиотек, поскольку для такого рода организаций одной из важнейших функций является хранение больших информационных массивов, особенно это характерно для крупных библиотек. В рамках этой тематики нами рассматривались различные подсистемы и аспекты, в частности: технические решения формирования информационного массива [4], надежность хранения информации в локальной компьютерной сети [6], автоматизированные архивные системы, интегрированные в АБИС [1], избирательный подход к информации при формировании резервных копий с целью оптимизации объемов архивов и скорости их формирования без ущерба надежности [2].

Естественно, что за более чем двадцатилетний период автоматизации библиотеки, свой путь развития прошли аппаратные и программные средства обеспечения АБИС. Эти изменения происходили по мере внедрения новых технологий, роста входного информационного потока, расширения локальной сети и повышения требований к производительности обработки и скорости доступа

к контенту, и, конечно, они не могли не коснуться систем, обеспечивающих надежность работы комплекса и сохранность данных [7]. Здесь стоит подчеркнуть, что все предыдущие изменения в системе резервного копирования были связаны, как и большинство модернизаций в других подсистемах АБИС, со стремительным ростом *информационного пространства*, что в большинстве случаев и становится первопричиной, определяющей развитие аппаратно-программного комплекса в целом [1,3]. Проблема библиотек это лишь частный случай общемировой проблемы сохранности информации, возникшая из-за многократного хаотичного дублирования, которое привело к тенденции опережения роста мировых информационных объемов над техническими возможностями электронных хранилищ [9].

Что касается нашей библиотеки, то в соответствие росту объема информационного ресурса, системы резервного копирования модифицировалась несколько раз, но в основном изменения ограничивались заменой устройства, формирующего архив, на более производительное (upgrade), но сохранение привычной технологии производства резервных копий не было ни случайным, ни вынужденным. Действительно, эволюция на рынке компьютерных систем в секторе систем хранения данных (СХД) вызывала дискуссии по поводу перехода на иные формы и технические решения систем архивного копирования. Эти обсуждения не прошли мимо и анализировались нами с точки зрения существующего состояния и перспектив развития конкретной АБИС ГПНТБ СО РАН [3], что и позволяло сознательно сохранять технологию резервного копирования, ограничиваясь малобюджетными техническими модификациями и программными доработками, сохраняя требуемый уровень информационной защищенности системы.

С внедрением технологии автоматизированного обслуживания читательского заказа литературы [4] требования к динамике копирования данных принципиально изменились. Произошла переоценка приоритетов среди критериев, используемых при проектировании или оценке качества функционирующей СХД [1]. Если раньше главенство отдавалось *объему*, затем *стоимости* и *масштабированию* системы, а потом только интересовали *скорости* и *удобство* копирования, то теперь определяющими стали именно *скорость и частота* репликаций данных, то есть появилась необходимость в создании дополнительной (самостоятельной, а не подменяющей существующую) системы резервного копирования.

Дело в том, что при выполнении заказа читателя на первоисточник в автоматизированном режиме, происходит постоянная (в темпе поступления заказов) модификация нескольких БД. Если – кратко, то цикл обработки заказа от формирования «требования» до возврата первоисточника на место хранения выглядит приблизительно так. В начале процесса, выбрав в электронном каталоге (ЭК) соответствующие запросу первоисточники, читатель формирует заказ, вследствие чего, выбранные библиографические записи появляются в специальной БД «Заказ» как информация для соответствующих технологических служб (фондодержателей) о поступлении запроса. Далее в случае выполнения требования изменяется статус соответствующих библиографических записей в ЭК, и, наконец, вносится отметка в БД «Читатель» о наличии у данного абонента заказанного первоисточника. Если учесть, что локальная сеть библиотеки насчитывает более ста рабочих мест для формирования требований, то станет понятно, что поступление заказов происходит постоянно в случайные моменты времени, стохастично изменяя содержимое в БД «Заказ», «Читатель» и, собственно, электронного каталога. И, наконец, после возврата выданной по заказу литературы, вновь модифицируется содержимое указанных информационных массивов, возвращая исходные статусы и состояния всех объектов заказа кроме БД «Заказ», где хранится история поступления, исполнения требований и причины возможных отказов.

В случае сбоя (разрушения или искажения информации) в этих базах до завершения цикла обработки заказа (от формирования до возврата) возникают серьезные технологические проблемы, решение которых весьма затруднительно либо просто не возможно, поскольку и сама проблема может быть обнаружена лишь при следующем обращении к данному объекту. Очевидно, что такая динамика данного технологического процесса требует и соответствующего темпа создания резервных копий и защитного дублирования данных.

Здесь необходимо разделить два принципиально отличных (по происхождению и результатам) источника возможного информационного коллапса: это, как обычно, характерные для технических систем *отказ оборудования* и, так называемый, *человеческий фактор* или ошибки оператора. Эти события весьма схожи в том, что оба оказывают негативное влияние на работу системы, в осталь-

ном они серьезно отличаются друг от друга, как по причинам их вызывающих, так и по методам минимизации потерь от воздействий каждого из этих факторов.

Для защиты от отказов оборудования существуют стандартные технические решения, повышающие надежность работы дисковых устройств путем организации так называемых RAID-систем (Redundant Array of Independent Disks) различных уровней (от RAID-1 до RAID-6 и их возможных комбинаций) [5]. Но эти технические ухищрения не могут отследить ошибочное (преднамеренное или неумышленное) действие человека-оператора, выполненное в соответствии правилам работы с точки зрения операционной системы или прикладного программного обеспечения. Очевидно, что даже при организации массива по методу RAID-1 (зеркальное дублирование информации) удаленный оператором файл не сохранится и в копии, а ошибочная запись «отразится в зеркале» также ошибочной.

Интуитивно понятно, что для восстановления такого рода информационного сбоя необходима постоянная запись ретроспективного состояния информационного массива, позволяющая вернуться (сделать откат системы) в любую временную точку «записи». Также понятно, что практически это сделать не возможно как по техническим, так и по финансовым причинам. Кроме того, сложности работы системы в условиях частого создания копий заключаются не только в потребности объемов дискового пространства или других носителей для хранения копий. Даже при относительно небольших размерах массива, работая с информацией на уровне *файлов*, процедура копирования блокирует открытые файлы на время создания копии, и, как следствие, снижается (с точки зрения пользователя) быстродействие системы. Незначительная модификация файла большого размера требует полного копирования всего файла, нерационально используя ресурсы системы.

В настоящее время применяются методы формирования резервных копий, практически, лишенные этих недостатков. Это так называемый метод *теневого копирования (shadow of copies services) «моментальных снимков (snapshot)»*, суть которого заключается в том, что в процесс копирования для создания «точек восстановления» предыдущего состояния системы включаются только блоки (кластеры), претерпевшие изменения с момента создания предыдущего моментального снимка. Таким образом процесс повторяется через определенный период времени, заданный оператором (системным администратором) при настройке системы теневого копирования.

Данный принцип создания копий может быть реализован различными способами и в различных сочетаниях с другими технологиями, аппаратными средствами и программными продуктами [8]. Например, корпорация Hewlett Packard предлагает сетевую СХД (SAN) MSA-2000 (P-2000) в качестве хранилище данных начального уровня с интегрированной поддержкой snapshot. HP MSA-2000 работает под управлением широкого набора операционных систем: 32-bit и 64-bit Windows, Red Hat and SuSE Linux, HP-UX, OpenVMS on Integrity, и может быть использована с системами виртуализации Windows Server 2008 x64 Hyper-V, VMware, and Red Hat Enterprise Linux Virtualization. Аналогичными функциями, но еще большим функционалом обладают системы хранения ряда HP LeftHand P4000 SAN позиционируемая производителем как корпоративная СХД. Обе системы способны модульно наращивать объем дискового хранилища, работать с различными носителями (SATA, SAS, FAS), подключаться к локальной сети по различным интерфейсам iSCSI или FDI на разных скоростях, и иметь различное количество каналов подключения. Все определяется бюджетом: даже MSA2000 (начальный уровень) варьирует стоимость от \$12000, а P4000 – соответственно от \$30000. Такие цены приемлемы для крайне небольшого числа крупных информационных центров, но, к счастью, и потребность в таких СХД пока не столь актуально даже для ведущих Российских библиотек. Существуют вполне приемлемые бюджетные решения на основе стандартных программных продуктах и аппаратных систем.

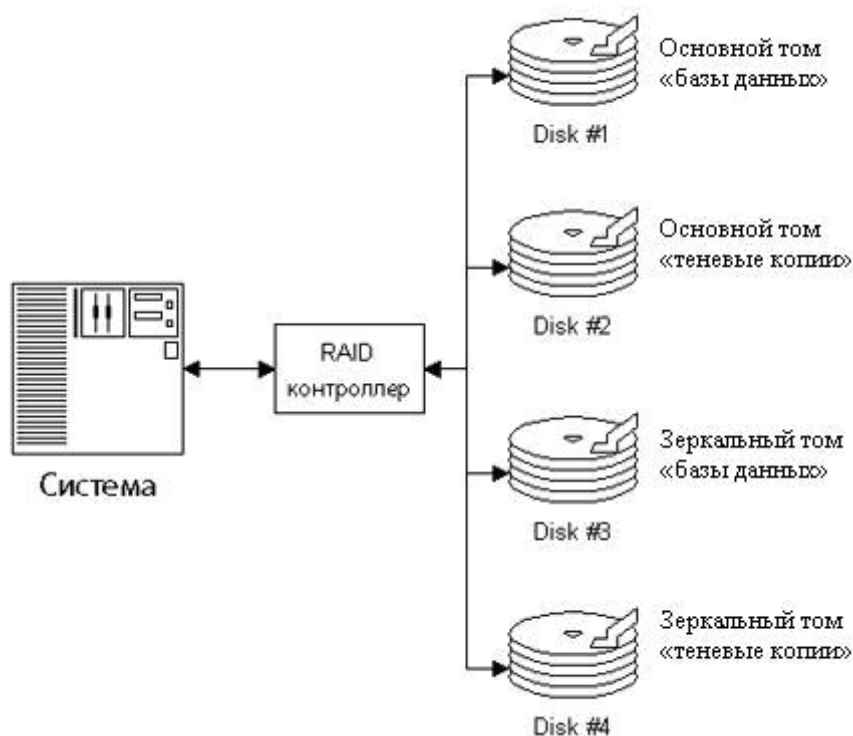
В нашем случае – ГПНТБ СО РАН – в качестве основной платформы используется сетевая операционная система Windows Server 2003 Enterprise, в составе которой присутствует стандартный сервис, выполняющий теневое копирование моментальных снимков тома VSS (*Volume Shadow of copies Service*). Этот сервис вполне соответствует нашему представлению о процедуре, обеспечивающей защиту от операторской ошибки. Как любая дополнительная задача, VSS требует определенных вычислительных ресурсов (мощностей) системы (сервера и локальной сети). Нагрузка на систему определяются объемом информационных массивов требующих, назовем это «оперативного копирования», а также интервалами между созданием «моментальных снимков» или, что-то же самое, частотой или расписанием формирования «теневых копий». Объем «снимка»

зависит от интенсивности использования информационного ресурса, в нашем случае – интенсивности поступления заказов и длительности цикла «заказ-возврат». Основные настройки работы сервиса заключаются в правильном назначении тома хранения снимков и расписания их создания. Оба параметра задаются на этапе запуска сервиса.

Вообще говоря, существуют рекомендации Microsoft по применению данного сервиса, ограничения и дополнительные сервисы для гарантированного восстановления тома информации из теневого копий, в частности при реализации VSS или подготовке точек восстановления рекомендуется делать следующее:

- Прежде чем активизировать точки восстановления VSS, выберите место хранения теневого копий с учетом рекомендаций по объему дискового пространства.
- Размещение VSS на системном или загрузочном томе – неудачное решение, так как системные файлы подвержены частым изменениям, что увеличивает число теневого копий.
- Оптимальный размер кластера для VSS – 16 Кбайт.
- Максимальное количество теневого копий ограничено числом 64, создание 65-й копии вызывает удаление первой (наиболее старой).
- При хранении теневого копий на физическом диске, отличном от диска-источника, повышается быстродействие и отказоустойчивость.
- VSS ни в коем случае не заменяет полноценное резервное копирование!

На основании вышесказанного в ГПНТБ СО РАН после внедрения автоматизированной технологии обслуживания читательского запроса на первоисточники, была разработана и установлена следующая дополнительная двухкомпонентная система защиты информации, представленная ниже:



Представленная на рисунке структура отвечает всем основным технологическим требованиям сохранности информации при автоматизированной процедуре обслуживания читательских запросов в АБИС «ИРБИС»: первая компонента RAID-1 (mirror) защищает систему от аппаратных сбоев, а теневое копирование обеспечивает защиту от ошибок в работе оператора. СХД построена с учетом рекомендаций по формированию систем теневого копирования моментальных снимков – физически разнесены рабочий том и том хранения репликаций, дополнительную страховку дает

зеркалирование тома выделенного для хранения теневого копий. Также отметим, что в конкретном случае средства позволили реализовать систему на высокоскоростных физических носителях (HDD) выполненных по технологии SCSI, характеризующиеся повышенной надежностью. Все отмеченные компоненты дают право характеризовать систему как надежно защищенную от сбоев, способных вызвать значительную потерю информации. При этом еще раз подчеркнем, что, несмотря на все свои качества, данная система является дополнением к существующей системе резервного копирования [1] и ни в коем случае не подменяет ее.

## Литература

1. Баженов И.С., Павлов А.И. Резервное копирование в АБИС [Электронный ресурс] // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы междунар. конф. «Крым 2007». – Электрон. дан. – М., 2007. – 1 электрон. опт. диск (CD-ROM). – Систем. требования: IBM PC, Windows 2000 и выше. – Загл. с этикетки диска. – N гос. регистрации 0320700790
2. Баженов И.С., Павлов А.И. Анализ информационного массива АБИС для оптимизации подсистемы резервного копирования // Науч. и техн. б-ки. – 2008. – N 1. – С. 59 – 63
3. Баженов С.Р., Павлов А.И. Проблемы сохранности электронных ресурсов в ГПНТБ СО РАН [Электронный ресурс] // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы междунар. конф. «Крым 2008». – Электрон. дан. – М., 2008. – 1 электрон. опт. диск (CD-ROM). – Систем. требования: IBM PC, Windows 2000 или выше. – Загл. с этикетки диска. – ISBN 978-5-85638-127-5. – N гос. регистрации 0320800989.
4. Баженов С.Р., Павлов А.И. Электронная технология обслуживания читателя. Начало ретроконверсии каталога // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса [Электронный ресурс]: материалы междунар. конф. «Крым 2009». – Электрон. дан. – М., 2009. – 1 электрон. опт. диск (CD ROM). – Систем. требования: IBM PC, Windows 2000 или выше. – Загл. с этикетки диска. – N гос. регистрации 0320900806
5. Воройский Ф.С. Информатика. Новый систематизированный толковый словарь-справочник (Введение в современные информационные и телекоммуникац. технологии в терминах и фактах). – 3-е изд., перераб. и доп. – М.: ФИЗМАТЛИТ, 2003. – 760 с.
6. Павлов А.И. Проблемы надежности и защиты информации в локальной компьютерной сети крупной библиотеки // Автоматизированные библиотечно-информационные системы: VI Сиб. науч. семинар с междунар. участием (1-7 июля 1996 г., г. Новосибирск): Тез. докл. и сообщ. – Новосибирск, 1996. – 64–65.
7. Павлов А.И., Баженов И.С., Кузнецов Б.Н. Архитектура аппаратно-коммуникационной части информационной системы ГПНТБ СО РАН: история, состояние и перспективы // Библиосфера. – 2008. – N 1. – С. 43–49
8. Шпунт Я. Новейшие концепции работы с информацией // Intelligent Enterprise / Корпоративные системы. – 2008. – Спецвыпуск. – С. 2–5.
9. Шрайберг Я.Л., Роль библиотек в обеспечении доступа к информации и знаниям в информационном веке. Ежегодный доклад Конференции «Крым». Год 2007. Роль электронной информации и электронных библиотек в продвижении к обществу, построенному на знаниях.