

УДК 02:004:[027.021:061.12(470)]
ББК 78.023.3+78.347.41(2Рос)

СОВРЕМЕННАЯ ТЕХНИЧЕСКАЯ БАЗА КРУПНОЙ БИБЛИОТЕКИ

© А. И. Павлов, 2014

*Государственная публичная научно-техническая библиотека
Сибирского отделения Российской академии наук
630200, г. Новосибирск, ул. Восход, 15*

Описана история развития и результаты многолетней работы автоматизированной библиотечно-информационной системы (АБИС) ГПНТБ СО РАН, а также конкретные технические решения различных технологических уровней. Приведены результаты эксплуатации централизованной системы автоматизации библиотечно-информационных процессов в научных центрах СО РАН как технологии, наиболее четко отражающей перспективы развития АБИС ГПНТБ СО РАН.

Ключевые слова: серверные платформы, система защиты данных, резервное копирование, коммуникации, система автоматизации библиотек «ИРБИС-64», автоматизированная библиотечно-информационная система, электронный каталог, базы данных, центр коллективного пользования, удаленный доступ, Web-технологии, сервер TCP/IP, протокол Z39.50, канал VPN.

The history of the development and the results of many years work of the automated library information systems (ALIS) in SPSTL SB RAS, as well as specific technical solutions on various technological levels are described. Operation results of the centralized system of library and information processes automation in the scientific centers of the SB RAS as technologies that are most clearly reflect the prospects of developing ALIS in SPSTL SB RAS are presented.

Key words: server platforms, data protection system, backup copying, communications, library automation system «IRBIS-64», automated library-information system, electronic catalog, databases, center for collective use, remote access, Web technologies, TCP/IP server, the Protocol Z39.50, VPN channel.

Уже с середины 1980-х гг. в Государственной публичной научно-технической библиотеке Сибирского отделения Российской академии наук (ГПНТБ СО РАН) велись работы по автоматизации библиотечных технологических процессов на базе существующего комплекса технических и программных средств, формируя исходный вариант автоматизированной библиотечно-информационной системы (АБИС). В то время аппаратная часть системы состояла из терминального комплекса на базе мини-ЭВМ (СМ-2420) и небольшого числа локальных автоматизированных рабочих мест (АРМ), которые обеспечивали в основном ввод (так называемые устройства подготовки данных на флоппи-дисках или магнитной ленте) для дальнейшего переноса информации в центральную машину.

На пике своего развития описываемый комплекс имел следующие технические характеристики: 16-разрядный центральный процессор с тактовой частотой 10 МГц, оперативная память объемом 4 МБ и максимальное дисковое пространство 100 МБ. Этот хост-компьютер обеспечивал одновременную работу до 30 терминальных рабочих мест, в качестве которых использовались алфавитно-цифровые

монокромные дисплеи с максимальной скоростью асинхронного обмена 9600 бит/с, а практически передача данных происходила вдвое медленней (4800 бит/с). Более подробное описание комплекса приведено в [1].

Система имела связь с большой ЭВМ (ЕС-ЭВМ) в Вычислительном центре (ВЦ) Академгородка Сибирского отделения Академии наук СССР, через которую осуществлялся выход в глобальные сети для выполнения информационных запросов в пакетном режиме. Библиотека связывалась с ВЦ по выделенной телефонной линии через модем с максимальной скоростью передачи 1200 (практически 600) бит/с.

1980-е и особенно начало 1990-х гг. характеризовались активным проникновением на российский рынок персональных компьютеров (ПК) и объединением их в локальные вычислительные сети (ЛВС). Библиотека в меру финансовых возможностей развивалась в этом направлении. В 1989 г. был сформирован участок автоматизированной обработки входного потока литературы. Эта первая ЛВС ГПНТБ СО РАН объединяла восемь бездисковых ПК класса XT на основе архитектуры ArcNet (скорость передачи 2,5 Мбит/с), работающих под управлением

сетевой операционной системы (ОС) Novell Netware v2.15, установленной на серверной платформе с процессором i286 (АТ).

Со временем топология и структура сети становились другими, росло число и потенциал серверных платформ, увеличивались и меняли архитектуру дисковые массивы, наращивалось количество и улучшались технические характеристики сетевых рабочих станций и технологических АРМ.

Оценить эволюцию аппаратной части АБИС можно, сравнив характеристики уже упомянутого первого варианта локальной сети с актуальными параметрами аппаратно-программного комплекса библиотеки. Технические показатели первого варианта сети:

- сервер РС АТ i286 с тактовой частотой 16 МГц, оперативной память 24 МБ, дисковой памятью 60–120 МБ;
- шинная топология по витой паре (одна пара) со скоростью передачи 2,5 Мбит/с (ArcNet);
- восемь бездисковых станций РС-ХТ (практически терминальных).

Уровень аппаратной базы и программный комплекс сетевых сервисных приложений, используемых в настоящее время как непосредственно для АБИС, так и для автоматизации других технологий в ГПНТБ СО РАН, представлен в табл. 1.

Очевидно, что характеристики сервера начального сетевого решения по всем параметрам как минимум на порядок уступают техническим характеристикам среднего АРМ из числа присутствующих сейчас в АБИС ГПНТБ СО РАН. Общее количество сетевых рабочих мест на сегодняшний день составляет порядка 400 машин различной конфигурации и широкого спектра технических характеристик, начиная от терминалов узконаправленных на работу с ЭК для формирования заказа книг (около 80 шт.), заканчивая многоядерными ПК, близкими по характеристикам к серверным платформам типа HOME или CAMELOT (см. табл. 1). Таких машин, естественно, меньше (около 25 шт.). Подавляющее большинство (примерно 270 шт.) – универсальные ПК средней производительности. Все основные рабочие места подключены по высокоскоростным каналам

Т а б л и ц а 1

Качественный состав локальной компьютерной сети ГПНТБ СО РАН (серверная часть)

Сервер / хост-компьютер	Технические характеристики	Назначение	Сетевая ОС
CAMELOT	Процессор: Pentium D, LGA775, Dual Core 3 ГГц Оперативная память: 4 ГБ Дисковая память: 1,2 ТБ	Z39.50 (ZooPARK) Mail-server (почтовый) FTP-сервер	UNIX Free BSD
HOME	Процессор: intel Core i7-920 2,66 ГГц/8 МБ Socket 1366 Оперативная память: 12 ГБ Дисковая память: RAID-10 500 ГБ	«1С» (бухгалтерский учет), бухгалтерская отчетность	Windows 2008 Enterprise
MAIN	HP Proliant DL580 G5: Процессор: Xeon Quad Core E7330 2,40 ГГц (четыре) Оперативная память: 48 ГБ Дисковая память: 2*30 ГБ (зеркало), 146 ГБ (IRBIS64)	Главный сервер локальной сети, автосохранение данных на стримере, сетевые сценарии, обновления антивирусных баз данных, сервер-«читатель», «сетевые ресурсы ГПНТБ СО РАН», Wins-сервер, IRBIS64-сервер, ресурс Users, терминальный сервер для Linux-клиентов, шлюз на SAN	Windows 2003x64 Enterprise
MAIN-2	HP Proliant DL580 G7: Процессор: Xeon Quad Core E7-4830 2,73 ГГц (четыре) Оперативная память: 256 ГБ Дисковая память: 2*250 ГБ (зеркало)	Сервер библиотечной корпорации, Web-server (Home page и поисковая система), File-server (собственные базы данных)	Windows 2008 R2x64 Enterprise
SAN	Promise VTrak M500i – дисковый массив (модуль сети хранения данных): Общий объем: 24 ТБ Интерфейс: iSCSI-SATA Dual Ethernet 1000 Base-T	File-server: цифровые коллекции, приобретаемые базы данных, патенты России, изобретения стран мира, ресурс Users	Multi-system

до 1000 Мб/с, а часть ПК (мало значительная для основных технологий) может подключаться по беспроводным каналам (Wi-Fi). В основном это личные компьютеры (ноутбуки, смартфоны, нетбуки и т. п. устройства) и временные подключения АРМ, связанные с перестановками, ремонтом и т. п. процессам жизнедеятельности Библиотеки. Качественный состав и технические возможности коммуникационного оборудования приведены в таб. 2.

Более того, сравнивать технические параметры этих комплексов некорректно, а чаще и невозможно: за прошедшее время изменились стандарты и принципы построения внутренней архитектуры, как самих компьютеров, так и локальных сетей, протоколов, топологий и коммуникационное оборудование. Появились новые принципиально иные информационные технологии. При этом в современной конфигурации, как и в изначальной структуре аппаратной части АБИС (вычислительного комплекса на базе хост-ЭВМ СМ-2420), присутствуют все те же функциональные и технологические подсистемы и уровни. Это объясняется тем, что каждая инфор-

мационная система (ИС) в сущности направлена на сбор (создание), обработку, хранение и передачу информации.

Информационные технологии (от *англ.* information technology, IT) – широкий класс дисциплин и областей деятельности, относящихся к технологиям создания, сохранения, управления обработки данных, в том числе с применением вычислительной техники. Как следует из определения, реализация – суть ИС.

Для АБИС как частного случая информационной системы основные задачи тем более остаются относительно неизменными в процессе развития системы, поскольку прежними остаются цели, на достижение которых направлена работа системы в целом. Состав задач может быть более детальным или, напротив, состоять из укрупненных блоков, все зависит от контекста рассмотрения ИС. Чтобы иметь представление о принципах и направлениях развития АБИС в ГПНТБ СО РАН на протяжении четверти века, достаточно детализации задач, представленной в табл. 3.

Т а б л и ц а 2

Качественный состав локальной компьютерной сети ГПНТБ СО РАН (коммуникационная часть)

Коммуникационное оборудование	Коммутатор (switch) 10/100/1000 Mbps		Беспроводные (Wi-Fi)
	управляемые	неуправляемые	
Количество устройств (подключений)	20 шт. (всего около 480 каналов)	15 шт. (всего 270 каналов)	6 шт. (около 100)

Т а б л и ц а 3

Распределение технологических задач по административным уровням АБИС

Уровни администрирования	Технологии работы с данными		
	Доступ	Управление	Хранение
Организационный	Составление расписаний работы сервера, уровни доступа по категориям пользователей, правовые аспекты, платные услуги и т. п.	Контроль проведения регламентных работ по актуализации данных обновления программных приложений; обеспечение индивидуальной ответственности за корректность информации	Выработка требований по периодичности проведения архивации данных и испытательных и профилактических работ
Программный	Пользовательский интерфейс, коммуникационные протоколы, установка драйверов сетевых устройств, защита от несанкционированного доступа	Обеспечения процедур пополнения и коррекции (исправление и удаление записей) и конвертирование данных в различные форматы	Автоматизация процесса архивирования данных, процедур аварийного завершения и последующего восстановления информации, анти-вирусная защита
Технический (аппаратный)	Представление результатов на различных физических носителях, каналы связи, выбор коммуникационного оборудования, топология сети и т. д.	Реализация аппаратной части серверных платформ и рабочих станций	Выбор дисковых массивов и интеграция в аппаратный комплекс технических средств архивирования; аппаратное дублирование (избыточность)

Таблица 3 представляет собой матрицу, каждый элемент (ячейка) которой отражает набор задач, находящийся в компетенции определенного уровня (группы), занимающейся внедрением и развитием АБИС (по горизонтали) и относящейся к конкретной технологии работы с информационными данными согласно приведенному выше определению ИС (по вертикали). В контексте данной статьи основной объект – третий технический (аппаратный) уровень задач. Примерно в той же степени будет рассмотрен программный уровень. Организационных задач мы практически не коснемся.

Часто в автоматизированных ИС аппаратный и программный уровни рассматривают как единый (аппаратно-программный). Однако в нашем случае все-таки их стоит разделить, так как зачастую аппаратная база оставляет возможность широкого выбора программных средств, равно как и конкретное программное обеспечение (ПО) может предполагать различные аппаратные реализации. Следовательно, в случае слабой корреляции аппаратного и программного уровней (ввиду их неодинаковых компетенций) степень ответственности каждого из уровней в отношении данного решения будет также сильно различаться.

Серверный уровень

Используемые сегодня в ГПНТБ СО РАН серверные платформы, задействованные в АБИС, не являются самыми передовыми, однако их уже нельзя относить к серверам начального уровня. Это достаточно производительные платформы корпоративного ранга, которые целиком и полностью справляются со своими технологическими функциями. Их внушительный вычислительный потенциал поддерживается на должном уровне за счет правильного администрирования в сочетании с периодической модернизацией (upgrade) аппаратной части.

Выше подчеркивалось, что данной работе приоритет отдается описанию аппаратной структуры АБИС, однако программой части также предполагалось уделить определенное внимание. В частности отметим, что выбор ОС и прикладного ПО проводился комплексно, исходя из конкретных финансово-экономических, кадровых и технологических условий текущего состояния Библиотеки. Так, в середине 90-х гг. прошлого века на смену сетевой ОС NetWare (Novell) и библиотечно-ориентированной информационно-поисковой системе ISIS пришли, соответственно, ОС Windows (Microsoft) и система автоматизации библиотек (САБ) ИРБИС (разработчик ГПНТБ России). Данная ОС и САБ изначально ориентированы на совместную работу по автоматизации технологических процессов библиотеки и для создания информационного ресурса БД «Электронный каталог».

Дальнейшее развитие АБИС проходило и происходит в этом ключе: меняются версии ОС, появляются другие возможности и учитываются новые технологические потребности АБИС при разработке очередных версий САБ ИРБИС и, соответственно, модернизируется аппаратная часть комплекса. Современные характеристики данного сегмента указаны в табл. 1 и в сочетании с рис. 1 дают наглядное представление об общей структуре аппаратно-программного комплекса АБИС ГПНТБ СО РАН.

До сих пор из всей схемы, изображенной на рис. 1, основное внимание было уделено блоку «ИРБИС-сервер». Это естественно, поскольку этот блок соответствовал приоритетному историческому направлению развития АБИС. Сейчас подсистема «ИРБИС-сервер» решает все технологические задачи библиотеки в автоматизированном режиме (комплектование, обработка первоисточника, учет фонда, книговыдача, обслуживание абонентов и т. д.). Это внутренняя и наиболее отработанная часть системы, тем не менее и здесь не все полностью закончено, как в разработке, так и во внедрении. Процесс развития продолжается: совершенствуются существующие АРМ, добавляются новые функции, но теперь это не требует больших и скорых проектов и прежнего финансирования. Фактически определенная законченность этапа автоматизации технологии и роль ГПНТБ СО РАН как центральной библиотеки СО РАН привели в 2012 г. к совместному решению Библиотеки и Института вычислительной техники (ИВТ) СО РАН разработать систему коллективного обслуживания сети библиотек СО РАН. Тогда в рамках Программы РАН № 15 «Информационные, управляющие и интеллектуальные технологии и системы» от ГПНТБ СО РАН был заявлен, поддержан и в течение года реализовывался проект «Создание прототипа единого центра автоматизации библиотечно-информационных процессов СО РАН».

Предусмотренная в рамках проекта организация центра коллективного пользования (ЦКП) САБ велась на базе и, так сказать, по образу и подобию АБИС, созданной в ГПНТБ СО РАН. Серверная платформа ЦКП выбрана по аналогии с основной платформой «ИРБИС-сервер» и в технологической программной части почти полностью ей тождественна. В централизованной системе, как и в ГПНТБ СО РАН, использован вариант САБ в реализации ИРБИС-64, т. е. тот, на котором реально опробованы, доработаны и функционируют в промышленном режиме все типовые технологические звенья АБИС, включая администрирование, комплектование, систематизацию, каталогизацию, книговыдачу и читательский поиск. Причем поиск может выполняться как в режиме локальной сети, так и удаленно (Web-ИРБИС), а также и в мультибазовом режиме (Web-ИРБИС с компонентой Z39.50).

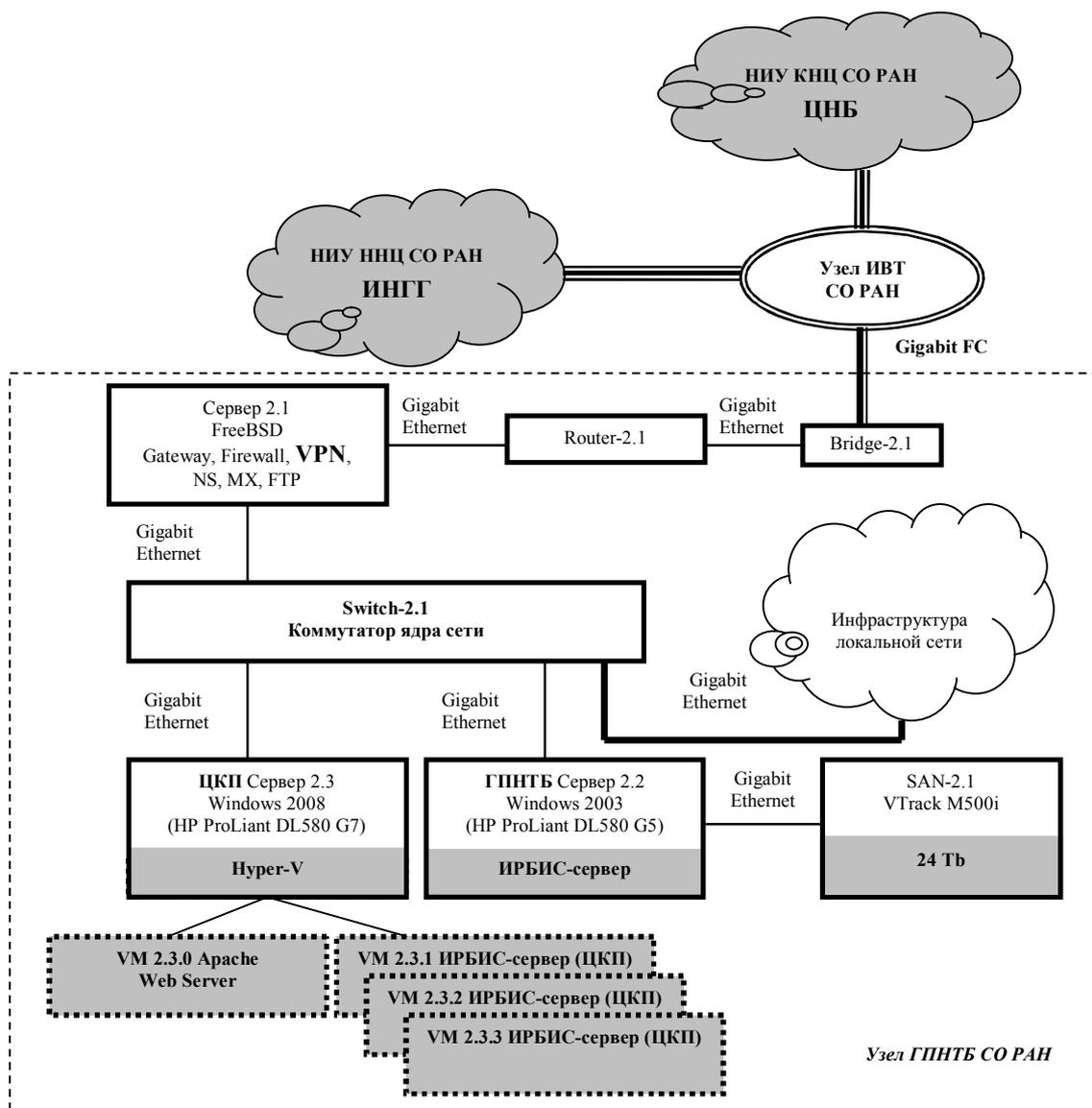


Рис. 1. Блок-схема серверного узла ГПНТБ СО РАН в общей схеме коммуникаций

В основе всех процессов – взаимосвязанное функционирование пяти типов АРМ: «Комплекатор», «Каталогизатор», «Читатель», «Книговыдача», «Администратор». Данный перечень охватывает весь требуемый спектр задач и способен удовлетворить все технологические запросы библиотек – потенциальных корпоративных членов ЦКП. Исходя из представленных пяти типов АРМ, согласно собственным потребностям, библиотеки могут сформировать свой конкретный набор.

Все сказанное выше отображено на блок-схеме рис. 1. «Сервер ЦКП» и «ИРБИС-сервер» (ГПНТБ) демонстрируют явную схожесть: они используют однотипные платформы Hewlett Packard серии ProLiant DL580, только немного отличающиеся по техническим характеристикам, что естественно для разных поколений (см. табл. 1).

Такой подход к аппаратной реализации ядра централизованной САБ (ЦСАБ) позволил сконцентрироваться на решении задач менее отработанного на тот момент коммуникативного уровня (см. табл. 3): выборе и проверке работы сетевых протоколов, режимов доступа, степени информационной защищенности, надежности передачи данных и т. д.

Верхний уровень в пунктирном прямоугольнике (см. рис. 1) демонстрирует внутрисетевые коммуникации. Вынесенные за пределы пунктира объекты – это удаленные пользователи (библиотеки институтов Новосибирского и Красноярского научных центров) и провайдерский коммуникационный узел ИВТ СО РАН, обеспечивающий все внешние (глобальные, межрегиональные, корпоративные) связи.

Доступ к информационным ресурсам

Центральная научная библиотека Красноярского научного центра (ЦНБ КНЦ) СО РАН в качестве удаленного клиента участвовала в эксперименте, направленном на выявление проблем эксплуатации АБИС под управлением ИРБИС-64 в режиме дистанционного доступа и оценки качества коммуникаций, стабильности работы, ресурсного запаса серверной части, а также сложности административной и временной разобщенности. ЦНБ КНЦ СО РАН была выбрана на самом начальном и, следовательно, весьма сложном этапе эксплуатации прототипа далеко не случайно.

Опыт работы в системе ИРБИС, наличие достаточной технической базы и общий хороший уровень подготовки персонала предполагали (и это подтвердилось в ходе работ) создание условий, при которых, как и при выборе серверного ядра ЦСАБ, можно будет сосредоточиться на решении задач удаленного доступа к ЦКП и доработать технологические АРМ ИРБИС при функционировании в этом режиме. Такой подход позволил достаточно быстро и эффективно провести начальный этап проекта [5].

Создание ЦСАБ СО РАН планировалось (и было желательно) провести на базе существующей структуры коммуникаций (как внешних, так и внутренних). Внутрисетевую структуру ГПНТБ СО РАН подробно описали А. И. Павлов, И. С. Баженов, Б. Н. Кузнецов [1] все намеченные ими перспективы реализованы и прошли достаточную проверку временем. Подключение дополнительного числа абонентов не должно было серьезно повлиять на общий трафик локальной сети библиотеки. Что касается внешних коммуникаций, они полностью находятся в ведении провайдера СО РАН (ИВТ), являются частью структуры распределенной информационной системы федерального уровня, для которой еще более критичны любые модификации.

Поэтому сутью первого этапа проекта стала оценка соответствия задействованных коммуникаций существующему уровню комфортности работы в АБИС. То есть по ощущениям оператора работа в удаленном режиме должна мало отличаться от работы в локальной сети института по времени реакции системы, надежности передачи данных, частоты обрыва, времени восстановления связи и т. п.

После опробования ряда вариантов подключения к ЦКП абонентов, как Новосибирского научного центра, так и КНЦ, был выбран и в конечном итоге принят в качестве рабочего и вполне отвечающего требованиям так называемый протокол виртуальной частной сети (VPN). Протокол VPN обеспечивает более защищенный режим передачи данных по сравнению с базовым TCP/IP протоколом, принятым в Интернете.

Итак, мы рассмотрели значительный временной и технологически важный этап перехода ГПНТБ СО РАН от традиционной крупной библиотеки с огромным фондом к библиотеке с практически полностью автоматизированной технологией. Последующая организация на базе собственной АБИС Централизованной САБ однозначно соответствует более высокому технологическому уровню. Результаты этого этапа показали определенную готовность системы к корпоративному расширению в рамках СО РАН, как по количеству участников, так и по их географическому положению.

Тем не менее подчеркнем, что простое расширение системы возможно лишь в небольшом масштабе. На начальном этапе тестирования прототипа системы по VPN-технологии каждый владелец сети (НИУ СО РАН) может со своего рабочего места иметь доступ к серверу ЦКП в ГПНТБ СО РАН, где установлены TCP/IP-сервер, RDP remote desktop protocol – удаленный рабочий стол, электронный каталог и прочие БД [6]. Это отработанный вариант реализации, когда, по сути, осуществляется доступ к физическому серверу и установленным на нем пользовательским приложениям, и все информационные массивы предоставляются одному–двум, в общем, весьма ограниченному числу пользователей (организаций), как в обычной локальной сети. Поэтому большинство нештатных ситуаций (сбои в работе системы, перезагрузка сервера, установка или обновление ПО, резервное копирование и т. п.) решаются относительно просто – путем оповещения пользователей, предупреждений, либо выполнение процедур переносится в нерабочее для удаленного клиента время.

Если же, согласно планам расширения корпорации, на одном сервере (в одной ОС) начнут работать несколько организаций из разных часовых поясов или с разным режимом работы, то влияние вышеперечисленных ситуаций может серьезно затруднить администрирование сервера ЦКП. Даже незначительные сбои способны скомпрометировать саму идею ЦСАБ. Поэтому, учитывая перспективу дальнейшего развития централизованной системы, в Библиотеке уже прорабатываются решения возможных проблем путем модернизации аппаратно-программной базы серверной компоненты.

Технические характеристики сервера ЦКП – платформа MAIN-2 (см. табл. 1) – результат выполненной в прошлом году за счет собственного бюджета модификации сервера по увеличению количества центральных процессоров и расширению оперативной памяти. Модернизация увеличила более чем вдвое общую производительность сервера. Это в сочетании с ранее выбранным системным ПО (ОС Windows 2008 R2x64 Enterprise), предусматривающем функцию виртуализации Hyper-V, по сути, реализует на платформе ЦКП популярную

сейчас «облачную» технологию (см. затемненные элементы на рис. 1).

Технология виртуальных машин (VM) позволяет:

- более рационально использовать аппаратные ресурсы серверов (на одном физическом сервере можно запустить несколько виртуальных серверов с несовместимыми задачами или приложениями);
- запускать специфические задачи на отдельном виртуальном сервере;
- значительно снизить сложность администрирования (удобное управление VM с хост-сервера);
- повысить надежность за счет кластеризации хостов (распределение сложных и важных VM на нескольких физических серверах) и более легкого резервного копирования и восстановления VM;
- поддерживать виртуальные сети (виртуальных коммутаторов) с большим количеством сетевых адаптеров (позволяет легко коммутировать различные сети на одной VM);
- выполнять практически незаметный для пользователя перенос VM с одного физического сервера на другой.

Пока корпоративных пользователей немного, конфликтов интересов и серьезных проблем между членами корпорации (организациями) не возникает. Тем не менее работа по виртуализации необходима для приобретения опыта, который, очевидно, будет полезен при развитии всей системы в целом. Вышеперечисленные атрибуты виртуализации как нельзя лучше соответствуют планам расширения ЦСАБ СО РАН, но могут быть использованы и во многих других подсистемах АБИС. Поэтому в рамках эксперимента с прошлого года клиенты корпорации работают на отдельных виртуальных «ИРБИС-серверах». Структурное разбиение на VM по организациям СО РАН, обеспечивающее независимую работу при возможных сбойных ситуациях, см. на рис. 1 (нижний уровень).

С другой стороны, может показаться, что локализация абонентов на разных VM создаст трудности из-за разделения ресурсов, которые должны находиться в общем доступе для корпоративных пользователей. На самом деле это не является большой проблемой, так как каждой VM можно предоставить доступ практически ко всей аппаратной конфигурации серверной платформы. Сложность взаимодействия модулей ИРБИС, размещенных на разных серверах, частично позволит обойти объединение VM в частную сеть (VPN) на уровне файловых ресурсов. Это исключает необходимость дублирования большей части данных, например, для обеспечения работы Web-ИРБИС, межбиблиотечного абонемента и удаленной регистрации пользователей.

Последний не рассмотренный пока элемент серверного уровня – дисковый массив, а точнее, сеть хранения данных (*Storage Area Network, SAN*). Для

аппаратной части АБИС, особенно для крупной библиотеки, характерен приоритет требования к емкости электронного хранилища по сравнению с вычислительной мощностью серверной части, хотя, и то, и другое не маловажно, особенно для ЦКП и организации VM. Стремительный, лавинообразный рост информационного потока – общемировая тенденция, которая в полной мере и едва ли не в первую очередь касается библиотек как общественного института, наиболее подготовленного к упорядоченному хранению информации.

По мнению экспертов компании IDC¹ (уже на 2007 г.) темпы роста производства систем хранения данных, которые, как принято считать, соответствуют так называемому закону Мура (*G. Moore*)², отстают от темпов роста информационного потока на столько, что к 2010 г. (по их прогнозам) это должно было привести к тому, что общее мировое хранилище данных окажется в состоянии разместить лишь 60–70% всемирного информационного массива. Возможно, так оно и случилось. Конечно, ГПНТБ СО РАН не может изменить мировую тенденцию, но адекватно отреагировать на заявления экспертов библиотека старалась и целенаправленно инвестировала определенную часть бюджета на развитие своего информационного хранилища.

Очевидна следующая закономерность: удвоение объемов информационного хранилища примерно каждые два года (рис. 2).

Приблизительно та же тенденция была и на начальных этапах – до первого терабайта, что как раз согласуется с уже упомянутым законом Мура. Рост производства устройств хранения повышает их доступность (например, ценовую, что весьма значимо для библиотек). В общем, и этот немаловажный критерий, и общее понимание значимости структуры аппаратной части АБИС позволили не допустить на протяжении всего периода автоматизации серьезного дефицита в размещении информационных массивов библиотеки.

Анализируя текущее состояние подсистемы электронного хранения (см. рис. 2), нельзя назвать ее «узким местом». На сегодняшний день общий объем серверного дискового пространства превышает 24 Тбайта – величина значительная даже для крупной организации или корпоративной компьютерной сети. Тем не менее потребность в дальнейшем росте есть и предусмотрена в планах развития.

¹ International Data Corporation (IDC) – крупная аналитическая фирма, специализирующаяся на исследованиях рынка ИТ, подразделение компании International Data Group. По данным IDC на компанию работают более тысячи аналитиков в 110 странах мира.

² Закон Мура – эмпирическое наблюдение, сделанное Г. Муром – одним из основоположников корпорации Intel. Основное положение закона Мура: число транзисторов на кристалле микропроцессора удваивается каждые два года.

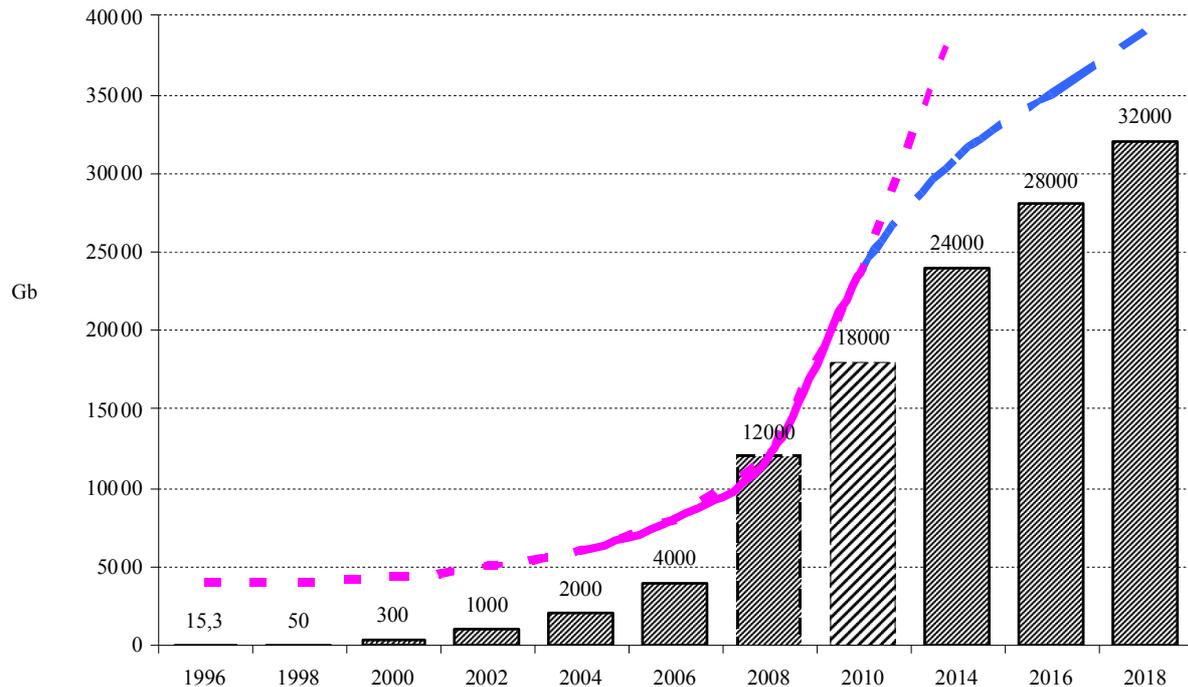


Рис. 2. Динамика роста информационного хранилища ГПНТБ СО РАН в 1996–2018 гг.

С другой стороны, в последние годы рост объемов дискового пространства в ГПНТБ СО РАН несколько уменьшился. Это можно объяснить рядом причин, начиная от глобальных, отмечаемых аналитиками мирового информационного развития, заканчивая банальными, например, финансовые (будем надеяться, временные) затруднения. Если согласиться с прогнозами экспертов IDC, тенденция к опережению все возрастающим информационным потоком темпов производства электронных хранилищ естественно снизит общий информационный объем из-за самой невозможности сохранять все полностью, проще говоря, часть информации обречена исчезнуть. Об этом давно и много говорят, обсуждая такой параметр, как важность (полезность или значимость) информации.

Пытаясь объяснить стремительный рост информационного потока, эксперты приводят ряд доводов, например: простоту производства информации (широчайший спектр и доступность средств записи) и при этом повышение ее стоимости (информация стала реальным товаром в современном обществе); слабую уверенность собственников в надежности хранения данных на цифровых носителях; до сих пор еще часто встречающиеся, недостаточно производительные каналы связи; беспорядочное, многократное копирование одних и тех же данных как следствие большинства приведенных выше причин.

Обеспечение сохранности информации – задача непростая и очень важная для библиотек, распола-

гающих большими информационными массивами (как традиционными, так и электронными) и, как правило, слабым финансированием, особенно на цели развития современной технологической базы. В докладе генерального директора ГПНТБ России Я. Л. Шрайберга на Международном библиотечном форуме в Крыму 2007 г. со ссылкой на упомянутые выше исследования компании IDC четко выделена проблема резервного копирования данных, и, достаточно последовательно и конкретно, обозначен подход к созданию цифровых хранилищ и архивов. По словам автора: «...еще до того, как будут определены виды и объемы электронного контента, важно определиться с технологией и устройствами хранения информации. Ответить на вопросы: Что хранить; Как хранить? Где хранить, в том числе и на каких носителях? Сколько резервных копий нужно? Что хранится временно, что постоянно? Что вообще можно не хранить?»

На том же форуме в унисон прозвучали два доклада, сделанные специалистами ГПНТБ СО РАН по этой тематике: о технологии резервного копирования и методах избирательного архивирования данных, применяемых в Библиотеке [2, 3].

Система обеспечения сохранности данных

Интуитивно ясно, что как бы не были реализованы технологические процессы доступа и управления данными, если созданный контент не защищен надежно от разрушения или потери, то смысл работы

ИС в целом сводится к нулю, превращается в процесс без гарантии результата. Поэтому с самого начала промышленной эксплуатации АБИС в режиме локальной сети важной задачей было обеспечить надежность работы системы. Печальный опыт на начальном этапе работ, когда применялись аппаратные и программные методы по защите данных, но не были выполнены дополнительные, как тогда казалось, второстепенные мероприятия по обеспечению надежной работы, перевел задачу в разряд главной. Случай лишний раз подтвердил, что в вопросе надежности нет мелочей. Использование любых динамических методов защиты информации даже от аппаратных сбоев носителей и применение алгоритмов исправления ошибок (RAID-технологии) не должно умилять значения традиционных методов защиты данных, как то: резервное копирование, стабилизация электропитания и т. д.

Последующее развитие технической части АБИС в первую очередь было направлено: на обеспечение требуемых параметров электрической сети, защитное заземление и отключение, разделение осветительной, бытовой и компьютерной электрических сетей. В дальнейшем стало нормой применение источников бесперебойного питания на особо важных и ответственных технологических участках локальной сети (серверы, узловые концентраторы, устройства внешних коммуникаций и т. п.). Затраты на решение этих задач оправдываются однозначно, поскольку в большинстве случаев здания библиотек изначально малоприспособлены для использования компьютерной техники. Энергоемкие устройства (лифты, конвейеры, вентиляция и т. д.) создают серьезные проблемы в электропитании, способные привести не только к сбоям в работе вычислительной техники, но и физически вывести из строя дорогостоящее оборудование. В настоящее время надежность (бесперебойная работа) приобрела особое значение еще и потому, что информационный ресурс крупной библиотеки должен быть постоянно представлен в сети Интернет. Системе необходимо обеспечивать круглосуточный режим работы, достаточный уровень защиты от несанк-

ционированного доступа, быстрое устранение неисправностей и т. п. Рассмотренное выше технологическое решение ЦСАБ по мере развития будет лишь повышать требования и актуальность данной подсистемы.

Описанная И. С. Баженовым технология защиты данных при возможных аппаратных сбоях [2] была организована как автоматизированная система архивации информационного ресурса на ленточный носитель. Устройство создания резервных копий – стример (модель Ultrium-215 – первое поколение стримеров производства Hewlett Packard, работающих по технологии LTO) обеспечивал запись на ленточные картриджи емкостью от 100 до 200 (режим сжатия) Гб. Формируемый таким образом архив гарантировал относительно быстрое (на то время процесс длился максимально трое суток) восстановление, как информационных ресурсов и пользовательских данных, так и системной информации.

За прошедший с тех пор период эта часть общей системы защиты данных принципиально мало изменилась, если не сказать – осталась прежней. Это говорит о том, что система надежна, технологична и работоспособна, изменения коснулись лишь модернизации аппаратной части и, соответственно, перехода к новой версии ПО, поставляемого производителем вместе с устройством. Замена аппаратной части была проста, естественна и логична: с появлением даже не регулярных и не значительных сбоев в работе стримера, он как физически и морально уже устаревший, был заменен моделью третьего поколения той же линейки стримеров – HP Ultrium-920.

Модель Ultrium-920 обладает значительно лучшими характеристиками по всем основным параметрам, но при этом совместима с предыдущим устройством по интерфейсу и по чтению используемых раньше носителей (табл. 4). Последнее качество особенно ценно, поскольку чтение ранее сделанных архивов без перезаписи на новые носители исключает необходимость наличия второго устройства (старого типа, но в исправном состоянии), экономит и время, и финансы.

Т а б л и ц а 4

Технические характеристики и аппаратная совместимость стримерных моделей Ultrium (трех поколений LTO)

Тип носителя	Емкость ленты, Гб		Скорость чтения, Мбит/с		LTO 1 HP Ultrium-215		LTO 2		LTO 3 HP Ultrium-920	
	норма	сжатие	норма	сжатие	Чтение	Запись	Чтение	Запись	Чтение	Запись
LTO 1 HP Ultrium-215	100	200	20	40	Да	Да	Да	Да	Да	Нет
LTO 2	200	400	40	80	Нет	Нет	Да	Да	Да	Да
LTO 3 HP Ultrium-920	400	800	80	160	Нет	Нет	Нет	Нет	Да	Да

Относительно функционального расширения системы защиты информации необходимо подчеркнуть, что за последние 10–12 лет развития АБИС ГПНТБ СО РАН произошли значительные количественные перемены, естественно вызвавшие качественные изменения в системе и технологии. Большинство этих изменений уже отмечались выше, осталось сделать ряд уточнений, чтобы подчеркнуть актуальность внедрения дополнительных систем защиты информации.

Во-первых, это значительный рост информационного массива (см. рис. 2), что потребовало применения методов, способных сократить время создания резервных копий. В основном проблему решило избирательное копирование [3] и только что рассмотренная модернизация ленточного устройства хранения. Также для оперативного восстановления была организована процедура создания дополнительной копии на дисковом носителе, хранящейся так называемый дифференциальный архив, т. е. файлы, модифицированные за текущий рабочий день, естественно, по его окончании.

Во-вторых, после внедрения технологии автоматизированного обслуживания читателей (последний крупный технологический процесс АРМ «Книговыдача»), в работе АБИС появились элементы и процессы, характерные для системы реального времени. До сих пор резервное копирование касалось статичных или условно статичных массивов данных. В связи с чем была внедрена система так называемых моментальных снимков (snapshot), способная с заданной периодичностью до нескольких минут фиксировать изменения информации, формируя «точки возврата» к актуальному на момент записи состоянию информационного массива.

В настоящее время мы имеем возможность возвращаться к состояниям системы с интервалом в 30 мин. и глубиной ретроспективы до полутора суток. Если от технических сбойных ситуаций защищает RAID-система дисковых массивов, то технология «моментальных снимков» способна до определенной степени снизить риск ошибок, вызванных «человеческим фактором» – оператором, обслуживающим читательские запросы на всей технологической цепи выполнения заказа [4].

Итак, мы рассказали на примере ГПНТБ СО РАН о технических средствах и технологических решениях задач, связанных с автоматизацией библиотечно-информационных процессов доступа, управления и хранения данных на различных уровнях администрирования АБИС (аппаратном, программном и организационном). Кроме того, привели варианты конкретных, созданных и используемых в Библиотеке аппаратно-программных комплексов, формирующих основу технологического обеспечения АБИС. С учетом всего сказанного, логично акцентировать в заключении, что современная круп-

ная библиотека в обязательном порядке должна включать в свою техническую базу следующие компоненты:

- развитую высокоскоростную локальную компьютерную сеть,
- средства коммуникации с внешним миром через глобальные сети (Интернет),
- достаточный резерв электронного хранилища,
- надежную систему резервного копирования,
- систему защиты информации на всех перечисленных подсистемах.

Так получилось, что все описанные ранее технические решения рассматривались с позиции обслуживания абонентов библиотеки, а с точки зрения формирования своего собственного контента речь шла только о БД «Электронный каталог». На самом деле в ГПНТБ СО РАН существует масса направлений по созданию БД собственной генерации, как библиографических, так и полнотекстовых, а также музейных графических коллекций и т. п. Для этих целей Библиотека имеет развитые специализированные рабочие места с очень высокими техническими характеристиками и располагает широким спектром периферийного оборудования: от ординарного до достаточно уникального, например, проекционные формата А2 и высокопроизводительные поточные сканеры формата А3.

Данное оборудование позволяет создавать коллекции первоисточников, в частности старых, редких и рукописных изданий, а также имидж-каталоги, что весьма эффективно повышает уровень обслуживания читателей, поскольку формирование «Электронного каталога» из ретроспективной части силами библиотеки по существующей технологии согласно подсчетам специалистов растягивается на десятилетия. И даже в этом направлении ретро-конверсии имидж-каталог имеет большое значение, обеспечивая как из локальной сети, так и с удаленных рабочих мест доступ к электронным образам каталожных карточек.

В заключении хочется добавить, что, несмотря на весь арсенал технических средств, эффективная работа библиотек может быть построена лишь на основе корпоративного взаимодействия. Иногда потребность в таком взаимодействии в большей степени испытывают крупные библиотеки, поскольку накопленный за время своего существования фонд на традиционных носителях, даже при наличии относительно неплохого технического обеспечения, труднее перевести в формат современных носителей. Эти тенденции отражены в деятельности современных Российских библиотечных и информационных ассоциаций и сообществ. Корпоративный подход используется, например, при создании «Сводного каталога библиотек России» (ассоциация ЛИБНЕТ), отражается в участии Ассоциации региональных библиотечных консорциумов (АРБИКОН)

в проекте некоммерческого партнерства по созданию сводного каталога периодики библиотек России «МАРС». Корпоративный подход очевиден и в формировании электронного контента «Президентской библиотеки им. Б. Н. Ельцина» посредством привлечения к участию ведущих библиотек России – РГБ и РНБ, ассоциации ЛИБНЕТ. Эта тенденция давно присутствует и в деятельности ГПНТБ СО РАН, которая выступает как организатор и участник многих корпоративных проектов России и регионе.

Литература

1. Павлов А. И., Баженов И. С., Кузнецов Б. Н. Архитектура аппаратно-коммуникационной части информационной системы ГПНТБ СО РАН: история, состояние и перспективы // Библиосфера. – 2008. – № 1. – С. 43–49.
2. Павлов А. И. Резервное копирование в АБИС // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса : материалы междунар. конф. «Крым 2007». – М., 2007. – URL: <http://www.gpntb.ru/win/inter-events/crimea2007/cd/66.pdf> (дата обращения: 21.04.2014).
3. Баженов И. С., Павлов А. И. Анализ информационного массива АБИС для оптимизации подсистемы резервного копирования // Науч. и техн. б-ки. – 2008. – № 1. – С. 59–63.
4. Павлов А. И., Баженов И. С., Грешинов Е. Б. Модернизация аппаратно-программного комплекса обеспечения сохранности данных в связи с внедрением технологии обслуживания читательского заказа в системе ИРБИС // Науч. и техн. б-ки. – 2010. – № 11. – С. 89–96.
5. Развитие программно-технологического комплекса информационно-библиотечной среды СО РАН 2010–2012 (заключительный) : отчет о НИР : IV.31.2.2 / Гос. публич. науч.-техн. б-ка Сиб. отд-ния Рос. акад. наук ; рук. работы Б. С. Елепов ; исполн. С. Р. Баженов [и др.]. – Новосибирск, 2013. – 133 с. – № ГР 01201058522. – Инв. № 02201357333.
6. Баженов С. Р. Опыт эксплуатации прототипа Центра автоматизации библиотечно-информационных процессов СО РАН в экспериментальном режиме обслуживания Центральной научной библиотеки Красноярского научного центра / С. Р. Баженов, Б. Н. Кузнецов, А. И. Павлов, О. А. Рогозникова // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса : материалы междунар. конф. «Крым 2013». – М., 2013. – URL: <http://www.gpntb.ru/win/inter-events/crimea2013/disk/101.pdf> (дата обращения: 21.04.2014).

Материал поступил в редакцию 29.05.2014 г.

Сведения об авторе: Павлов Александр Иванович – заведующий отделом компьютерной и множительной техники, тел.: (383) 266-50-89, e-mail: super@spsl.nsc.ru