

*В.М. Шаймарданов*

**СОЗДАНИЕ ПРОГРАММНО-АППАРАТНОГО КОМПЛЕКСА  
ДОЛГОВРЕМЕННОГО ХРАНЕНИЯ И ИСПОЛЬЗОВАНИЯ  
ГИДРОМЕТЕОРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ**

*V.M. Shaimardanov*

**CREATION SOFTWARE-HARDWARE COMPLEX  
OF LONG-TERM STORAGE  
AND USE OF THE HYDROMETEOROLOGICAL INFORMATION**

*В статье рассматриваются вопросы выбора технических и программных средств обеспечения долговременного хранения и использования для обслуживания гидрометеорологических данных на базе современных роботизированных библиотек. Представлена принципиальная схема расстановки технических средств с целью обеспечения надежного бессрочного хранения информации. Рассмотрены средства оцифровки бумажных документов и обслуживания потребителей печатными данными.*

*Ключевые слова: бессрочное хранение, программно-аппаратные средства архива, роботизированные библиотеки.*

*The article addresses the issues of choosing hardware and software for long-term storage and the use in managing hydrometeorological data based on state-of-the-art robotic libraries. A schematic diagram of the layout of technical means is presented to ensure reliable and permanent storage of information. The tools for digitizing paper documents and serving the consumer with printed data are considered.*

*Keywords: long-term storage, software-hardware of archiving, robotic libraries.*

## **Введение**

В Росгидромете создан уникальный Государственный фонд данных о состоянии окружающей природной среды (Госфонд). В нем сосредоточены документы и данные наблюдений за состоянием окружающей природной среды по территории России, зарубежным территориям и акватории Мирового океана в результате деятельности станций и постов гидрометеорологической сети, морских судов и других организаций.

Долговременное хранение данных Фонда на машиночитаемых носителях осуществляется централизованно во ВНИИГМИ-МЦД. Уже сегодня Росгидромету необходимо сохранить более 5 Пб данных. К настоящему времени в фонде накоплено и хранятся эти данные на полудюймовых магнитных лентах, картриджах, CD-ROM, микрофильмах и листовых материалах. Имеется большое количество данных, полученных в рамках международных проектов и обменов на различных технических носителях (магнитооптические диски, магнитные картриджи и др.).

Сложившаяся за последние десятилетия система архивного хранения не обеспечивала приемлемого времени доступа к данным – реальное время досту-

па к отдельным файлам достигало нескольких суток из-за несоответствия технических средств и методов хранения современному уровню развития систем архивации данных.

Большая часть фонда хранится, в основном, на устаревших носителях – полдюймовых магнитных лентах, большинство из которых находится под угрозой физической деградации. Некоторая часть данных перенесена на картриджи IBM 3480, которые также являются устаревшим носителем, и компакт-диски, которые нельзя рассматривать как носители для долговременного хранения. Поэтому задача обеспечения бессрочного и безопасного хранения данных, их систематизации, описание для получения сведений о наличии, поиска и обеспечения эффективного доступа к данным с такими объемами является очень актуальной и остро необходимой.

Следует отметить, что образовался значительный разрыв между современными ИТ технологиями и уровнем обслуживания Госфонда, что могло привести к невосполнимой потере уникальной информации и, фактически, утрате всего научного наследия Росгидромета. Используемые технические средства, технологии и решения по информационному обслуживанию Госфонда являются морально и физически устаревшими.

Развитие Фонда данных является важной составной частью «Стратегии деятельности в области гидрометеорологии и смежных с ней областях на период до 2030 г. (с учетом аспектов изменения климата)».

Стратегией определены следующие задачи развития Фонда данных:

1. Обеспечение бессрочной и надежной сохранности архивного фонда на технических носителях, оперативного и целенаправленного доступа к данным, решение проблемы централизованного управления процессами ведения фонда.

2. Обеспечение надежности долговременного хранения и доступа к гидрометеорологической информации и информации о загрязнении окружающей среды на традиционных бумажных носителях.

3. Развитие надежных систем обслуживания потребителей неоперативной (режимной) гидрометеорологической, гелиогеофизической информацией, информацией о загрязнении окружающей среды, о состоянии и изменениях климатической системы.

Достижения развития комплекса Архивной системы Гидрометслужбы должны обеспечить:

- необратимое закрепление позитивных сдвигов в результате завершения проекта «Модернизация и техническое перевооружение учреждений и организаций Росгидромета»;

- ликвидацию отставания от аналогичных комплексов развитых зарубежных стран;

- объемы архивных систем Государственного фонда данных, уровень развития средств связи, достаточные для восприятия лавинообразно нарастающих потоков данных о состоянии окружающей среды;

- возможности модернизации в периоды до 2030 г. на уровне, соответствующем и превышающем мировой уровень и учитывающем передовые технологии этих периодов.

Реализация модернизации архивной системы обеспечит, наряду с бессрочным хранением данных, движение архивных данных от базовых архивов до продукции, направляемой пользователям, включая поддержку трех уровней данных (базовые архивы, производные массивы или архивы, базы данных) и, соответственно, трех основных подсистем архивной системы:

*Подсистема хранения*, обеспечивающая хранение базовых и производных архивов и манипулирующая данными, в основном, на уровне файлов.

*Подсистема обработки*, обеспечивающая формирование слоя производных массивов (архивов) и манипулирующая данными как на уровне файлов, так и на уровне их содержимого.

*Подсистема обслуживания*, обеспечивающая формирование необходимых для обслуживания производных массивов и баз данных и манипулирующая данными на уровне, характерном для СУБД и различных систем анализа данных. Производные массивы могут создаваться как для непосредственного обслуживания потребителей или решения научно-исследовательских задач, так и в качестве промежуточного звена – для загрузки в базы данных. Производные массивы широкого назначения могут иметь ценность, не меньшую, чем исходные базовые архивы. В этом случае они сами становятся объектами архивного хранения и называются производными архивами.

В модернизированной архивной системе этим трем подсистемам соответствуют три *функционально-технологических* блока:

- базовый архив;
- оперативный архив;
- обмен и обслуживание информацией.

Для обеспечения сохранности информации в настоящее время, во ВНИИГМИ-МЦД реализован технологический комплекс перезаписи данных с МЛ ЕС на картриджи IBM 3480. Переписанные данные с МЛ комплектуются в новую архивную единицу хранения в среднем объеме до 170 МБ информации (9–10 МЛ) и записываются на картриджи формата IBM 3480 для включения в последующем на новые технические носители бессрочного хранения.

### **Основные технические решения**

При выборе технических средств, необходимых для переоснащения, был проведен анализ надежности и долговечности хранения на различных технических носителях, а также имеющихся и используемых в основных мировых центрах данных, хранящих большой объем информации. На основе анализа в качестве технического носителя выбрана магнитная лента, а с учетом обеспечения сохранности данных, их объемов и автоматизации работы принята Роботизированная ленточная библиотека.

Среди существующих в настоящее время ленточных библиотек, предлагаемых различными производителями (IBM, HP, SUN, Quantum), при прочих равных условиях нами выбрана библиотека фирмы IBM TS3500. Именно данная библиотека удовлетворяет требованиям по емкости, масштабируемости, а также, что не менее важно, учитывался опыт сервисного обслуживания. К тому же, после приобретения SUN компанией Oracle новым руководством было поставлено под сомнение дальнейшее развитие собственных ленточных устройств. Обращает внимание так же тот факт, что уже с 16 мая 2006 г. IBM установлен мировой рекорд по плотности записи данных на магнитную ленту: – 6,67 млрд бит на квадратный дюйм. Это более чем в 10 раз превосходит показатели самых популярных сегодня стандартных устройств хранения на магнитной ленте; картридж стандартных габаритов сможет вмещать до 8 ТБ информации без сжатия.

Для эффективного управления библиотеками выбран сервер IBM z9BC. Выбор, в первую очередь, обусловлен существующими до настоящего времени технологиями для работы и архивными данными на МЛ. Технологии и программные средства считывания, обработки и контроля архивных данных развернуты по базе IBM 4381, что и определило необходимость использования ОС z/VM для наследования существующих решений. ОС z/VM поддерживает драйверы, необходимые для подключения лентопротяжных устройств МЛ и подсистемы картриджами IBM 3480, а также позволяет перенести все технологии и программные средства без дополнительной переделки.

Кроме МЛ и картриджами IBM 3480 необходимо обеспечить перенос информации и с других устаревших и разнородных носителей, как-то:

- стримерные ленты различных типов;
- дискеты и компакт-диски;
- магнитооптические диски и др.

Кроме этого, учитывались также другие факторы, основными из которых, являются:

- опыт работы на ЭВМ IBM;
- управление роботизированной библиотекой в других мировых центрах данных, например, в NCDC (США), центру аналогичному нашему;
- возможность подключения НМЛ для считывания большого объема данных, записанных в свое время на ЕС ЭВМ.

Архитектура IBM z9 BC позволяет создать несколько независимых виртуальных машин (LPар). В каждой LPар установлена ОС z/VM и созданы от одной до нескольких виртуальных машин (в зависимости от количества задач). На большинстве виртуальных машин установлена ОС zLinux.

Такой подход обусловлен необходимостью изолировать различные процессы, запущенные и выполняемые на сервере отдельно. В серверах компании IBM System Z, к которой относятся z9 BC это реализовано наиболее удачно – все процессы почти «физически» разделены. Что позволяет надежно и эффективно разворачивать различные технологии в рамках одного сервера.

**В качестве дисковой системы хранения используется IBM DS8300** общей емкостью 100Тб. Этот выбор обусловлен желанием минимизировать затраты на сопровождение и внедрение (один производитель с сервером), избежать возможных трудностей на стыках, которые могут возникать между устройствами различных производителей. Кроме того, на тот момент IBM DS8300 имела наилучшие в отрасли технические показатели и возможности, например скорость передачи данных между устройствами по оптическим кабелям до 4ГБ.

Конфигурация позволяет разбить систему хранения на две разнозначные подсистемы хранения в одном устройстве, в нашем случае, в пропорции 25 % и 75 %. Такое разделение позволяет в рамках одного дискового массива создать разделы, данные которых изолированы друг от друга. Один раздел (25 % от общего объема) используется для работы с данными сервера IBM z9 и имеет соответствующую файловую структуру. Другой раздел (75 % от общего объема) имеет файловую структуру для работы с открытыми ОС (Windows, Linux)

### **Блок оцифровки бумажных носителей**

Наряду с хранением данных на различных технических носителях, во ВНИИГМИ-МЦД накоплен большой фонд данных на бумажных носителях. Для повышения доступности этих данных стоит задача скорейшего сканирования, по возможности всех бумажных носителей.

Для решения этой задачи реализована начальная стадия проекта по созданию блока оцифровки бумажных носителей. В настоящее время приобретено три сканера:

- Эллар Сканмакс – высокопроизводительный цветной потоковый сканер, для расшитых документов формата А3.
- Два сканера Эллар Планскан – книжные, планетарные сканеры, для сшитых документов формата А2, один черно-белый, второй цветной.

Развернут участок сканирования, где осуществляется перевод данных в формат PDF. После формирования PDF-файл передается в технологию хранения оцифрованных документов, созданную на базе IBM Content Manager On Demand.

### **Блок выпуска печатной продукции**

Стоит также отметить, что в рамках переоснащения развернут блок выпуска документации на бумажных носителях, основные задачи которого – выпуск бюллетеней ВМО, ежемесячников, ежегодников и атласов, а также оперативная печать.

В зависимости от тиража печати целесообразно использовать два типа печатных машин:

- классическая офсетная печатная машина Namada – для черно-белой или полноцветной неоперативной печати тиражом более 300 экз. Использование Namada предполагает полный технологический цикл по подготовке материалов к печати, используемый на всех офсетных машинах.

- HP Indigo 5000 – цифровая офсетная печатная машина для черно-белой и полноцветной оперативной печати или при печати до 300 экз.

Кроме печатных машин в блок выпуска документации включен комплекс послепечатной обработки:

- одноножевая бумагорезательная машина Wohlenberg;
- листоподборочно-брошюровальная линия Duplo с возможностью фальцовки и сшивания проволокой;

Таким образом, с учетом изложенного выше, разработан вариант структуры архивной системы Росгидромета [3], который состоит из:

- двух роботизированных ленточных библиотек по 4 Пб каждая;
- сервера управления архивной системы IBM z9, BC с 3-мя процессорами;
- дискового массива IBM DS8300 объемом более 100 Тб;
- оборудования для создания сети хранения данных (SAN);
- конвертеров для подключения устаревших лентопротяжных устройств и подсистем на картриджах;
- печатающих полиграфических устройств и пр.

На рис. 1 представлено оборудование, вошедшее в переоснащение архивной системы.

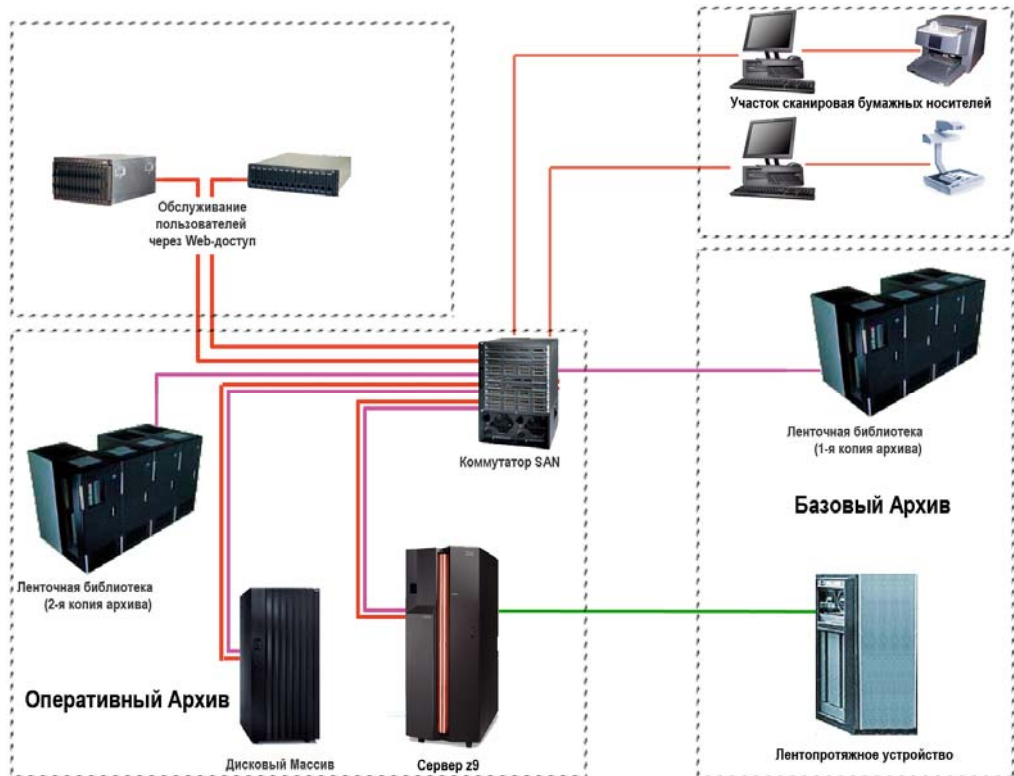


Рис. 1. Результат переоснащения архивной системы

С целью повышения катастрофоустойчивости архивной системы принято решение о размещении двух роботизированных библиотек в различных зданиях, которые между собой территориально разнесены. Такое решение позволит повысить сохранность данных в случае стихийных бедствий, потери электропитания на одной из площадок, аппаратных сбоев.

### **Комплекс программных средств**

Для повышения эффективности использования оборудования и управления данными, а также для разработок новых технологий обслуживания потребителей гидрометеорологической информацией с учетом всех аспектов было выбрано и поставлено общесистемное и доработано прикладное программное обеспечение, часть из которого развернута на базе IBM z9, другая – на серверах под управлением ОС Windows. Основные ОС IBM z/VM 5.3 (5.4) и zLinux RedHat развернуты на сервере IBM z9.

В состав общесистемного входит следующее основное программное обеспечение:

- операционная система VM/ESA релиз 1;
- языки программирования PL/1, Assembler;
- среда разработки Rexx.

Прикладное программное обеспечение функционирует в режиме VM в среде CMS и предназначено для решения следующих задач:

- чтение магнитных лент в формате SL, NL OS360 и в специальном формате помехозащищенного кодирования;
- контроль полноты и качества перезаписи;
- перезапись данных на картриджи формата IBM 3480;
- специальное чтение картриджей формата IBM 3480 для передачи их в программную среду MS Windows.

В состав Прикладного программного обеспечения входят следующие программы:

- средства терминального доступа к Центральному управляющему серверу и средства файлового обмена по FTP;
- комплекс программ управления данными Аисори (функционирует в среде Windows);
- чтения и восстановления поврежденных данных с магнитных лент с плотностью записи 800BPI, 1600BPI, 6250BPI и контроля перезаписи, написанные на языках Assembler, Rexx, PL/1 в среде CMS VM/ESA;
- обслуживания диспетчерских файлов процесса перезаписи и записи данных на картриджи IBM 3480, написанные на языках Assembler, Rexx, PL/1 в среде CMS VM/ESA;
- чтения данных с картриджей HDS 7480, написанные на языках Assembler, Rexx, PL/1 в среде CMS VM/ESA;

- выполнения перекодировки текстовых данных из формата ДКОИ в формат ASCII и обратно с возможностью создания своих таблиц перекодировки при обмене данными Центрального управляющего сервера с Рабочими станциями терминального доступа по FTP.

Для управления данными архива, обеспечения их перемещения между системами хранения и создания резервных копий используется IBM Tivoli Storage Manager. Это ПО применяется для управления информацией в сложных распределенных информационных корпоративных вычислительных средах, имеет собственную базу метаданных хранимой информации, позволяет автоматизировать функции резервного копирования и восстановления данных.

Таким образом, установлены следующие технические и программные средства:

- сервер IBM System z9 BC;
- дисковый массив IBM System Storage DS8300 – объем более 100 Тб;
- ленточные библиотеки IBM System Storage 3500 Tape Libraries – суммарный объем более 8 Пб;
- сканеры для оцифровки бумажных носителей;
- цифровая офсетная машина;
- комплексы до печатной и послепечатной обработки;
- шелко-трафаретный комплекс;
- ОС IBM z/VM 5.3 and 5.4 и zLinux;
- ПО управления данными IBM Tivoli Storage Manager;
- ПО индексации данных IBM Content Manager OnDemand;
- СУБД Oracle, SAS, ГИС Bentley, Сервер приложений.

### **Заключение**

В результате переоснащения достигнуто следующее:

- сокращено время доступа к данным от суток до минут;
- повышена доступность данных, в том числе, удаленная;
- осуществлена консолидация данных с различных носителей в единую среду и упрощен доступ к ним;
- обеспечено бессрочное и надежное хранение;
- реализована возможность обслуживания информационной продукцией (ежемесячники, ежегодники, атласы и пр.), занесенной на бумагу.

### **Литература**

1. Данные производителя, IBM, [www.ibm.com](http://www.ibm.com)
2. Шаймарданов М.З., Веселов В.М., Стерин А.М., Шаймарданов В.М. Концепция модернизации архивной системы Мирового Метеорологического Центра // Труды ВНИИГМИ – МЦД, 2007, вып. 172, с. 3-37.
3. Шаймарданов М.З., Шаймарданов В.М. Единый государственный фонд данных // Труды 6-го Всеросс. метеорологич. съезда, 2011, С.-Петербург, с. 74-85.