

## ВЫБОР СТРУКТУРЫ БАЗЫ ДАННЫХ ДЛЯ ХРАНЕНИЯ ИНФОРМАЦИИ ХАРЬКОВСКОГО РАДАРА НЕКОГЕРЕНТНОГО РАССЕЙЯНИЯ

### Введение

Все увеличивающийся объем ионосферной информации, получаемый в обсерватории Института ионосферы (г. Харьков) [1] в процессе функционирования радара некогерентного рассеяния (НР) и ионосферной автоматической станции (АИС), обуславливает необходимость создания специализированной базы ионосферных данных (БИД), полученных над украинским регионом. База будет способствовать улучшению информационного обеспечения, повышению быстродействия процесса обработки ионосферной информации, даст возможность пользователям иметь доступ к информационным ресурсам через локальную сеть, сеть Internet или через внутрикорпоративную сеть на базе Intranet.

В настоящее время ионосферная информация с данного исследовательского радиолокационного комплекса записывается на CD- и DVD-диски. При этом имеет место проблема преобразования и упорядочивания информационных массивов значительной продолжительности. Для дальнейшей оптимизации систем хранения и дальнейшей обработки данных требуется учет сложности процедур преобразования информации, накапливаемой в большей части в on-line режиме. Так как в результате дистанционного радиозондирования ионосферы образуются большие информационные объемы (до 10 ГБайт/сутки), то для организации их оптимального размещения в БИД с целью долгосрочного хранения необходимы классификация получаемых результатов и оптимизация их структуры. Также требуется ввод в действие системы управления базой с учетом специфики геофизической обработки и визуального анализа экспериментальных данных [2].

Таким образом, целью исследований является анализ данных, получаемых в Институте ионосферы, и рассмотрение возможности организации электронной БИД, которая позволила бы существенно оптимизировать процедуры систематизации результатов, записи и последующего представления ионосферной информации.

### Анализ состояния вопроса

Рассмотрим общие подходы, связанные с созданием подобной системы хранения данных.

Одним из этапов проектирования любых баз данных (БД) является даталогическое проектирование, представляющее собой создание логической информационной структуры [3]. Первыми такими системами стали иерархические, сетевые модели, а также модели на основе инвертированных списков. Однако сложности их практического использования, увеличенные объемы и структурная сложность хранимых данных привели к появлению и широкому распространению новых разработок – реляционных баз хранения информации [4].

Реляционная модель состоит из трех составляющих, относящихся к разным аспектам реляционного подхода: структурной части, а также манипуляционной и целостной частей. Преимуществом реляционных БД является наличие мощного и в то же время простого языка SQL – Structured Query Language, позволяющего пользователям выполнять многие необходимые операции.

Одним из недостатков реляционной базы невозможность хранения в ней сложных типов информации (в виде рисунков, графиков, диаграмм и т. п.), а также информации абстрактного типа, который бы определялся самим пользователем.

Попытка встроить семантику в механизм управления БД подтолкнула исследователей к созданию объектно-ориентированной модели, в основу которой были заложены принципы,

характерные для объектно-ориентированного программирования: инкапсуляция, полиморфизм и наследование. Преимуществом таких объектно-реляционных СУБД является сохранение преемственности с системами предыдущего поколения, а также возможность применения уже зарекомендовавших себя методов хранения информации, возможность использования новых объектно-ориентированных механизмов, повышающих эффективность хранения и представления данных.

### **Анализ ионосферной информации**

Как было сказано выше, если рассматривать функционирование базы данных применительно к специфике работы радара НР, в данном случае ее структура должна учитывать особенности метода ионосферного радиозондирования. Так, в базе должны храниться [2] исходная радиофизическая информация, получаемая в процессе проведения экспериментов, рассчитанные высотно-временные зависимости автокорреляционных функций (АКФ) и спектров сигнала НР, вычисленные параметры ионосферы, результаты их анализа и последующей геофизической интерпретации. Помимо компьютерных файлов, образуемых на выходах корреляторов, и файлов, содержащих ионограммы, БИД должна хранить и служебную информацию, содержащую комментарии к файлам, сведения об условиях проведения экспериментов, техническом состоянии аппаратуры и др.

Проведенный анализ показал, что получаемые в результате зондирования ионосферы данные, которые предполагается хранить в БИС, по содержанию условно можно разделить на три основные группы:

- исходные радиолокационные данные, которые системами (корреляторами) первичной обработки формируются в виде статистических характеристик сигнала рассеяния, поступающего на вход радиоприемного устройства радара НР [1], и блоком цифрового преобразования отраженного сигнала формируются на выходе автоматической ионосферной станции [5];

- выходные данные, которые являются результатом радиофизической обработки уже статистических характеристик и представляют собой рассчитанные высотно-временные распределения значений ряда геофизических параметров ионосферной среды;

- служебная информация о состоянии аппаратуры, режимах ее функционирования, алгоритмах обработки данных и т.д.

Подробнее рассмотрим структуру *исходных данных*, которые образуются на выходе системы первичной обработки. Эти данные представляют собой набор *S*-файлов, содержащих результат корреляционной обработки сигнала рассеяния. Внутренность каждого файла (рис. 1) состоит из поля заголовка, отражающего дату и время проведения эксперимента, номер сеанса и количество излученных зондирующих импульсов, определяющих период статистического накопления результатов. Поле данных содержит рассчитанные вдоль развертки дальности 680 значения АКФ сигнала НР по 19 ординат в каждой, причем полученных одновременно с выходов четырех канальных приемников. Кроме того, в файле содержится высотный профиль мощности сигнала НР, а также высотный профиль усредненных значений напряжения этого сигнала, позволяющий контролировать уход “нуля” АЦП. Имеется также набор служебных меток – результат анализа данных на достоверность, проведенного программой фильтрации отражений от летательных объектов.

В исходных данных имеет место еще один набор информации – в виде *V*- и *C*-файлов, формируемых в чуть отличающемся формате [6] при функционировании двух корреляционных систем предыдущего поколения.

Все исходные файлы в процессе статистического (от развертки к развертке) усреднения информации формируются, как правило, в течение одной минуты, записываются в двоичном коде и имеют объемы от 0,02 до 0,2 МБайт.

Заголовок			word	день		
			word	месяц		
			word	год		
			word	часы		
			word	минуты		
			word	секунды		
			word	номер сеанса		
			word	nr0		
			word	nr1		
			word	резерв		
Данные	1	1	1	dword	АКФ сигнала НР (4 канала)	
			...			
		19	dword			
		...				
		680	1	dword		
			...			
	19		dword			
	2	--				
	3	--				
	4	--				
	1	1		dword	«Ноль» АЦП (2 канала)	
			...			
		680	dword			
		2	--			
		1	1	dword	Мощность сигнала НР (2 канала)	
			...			
680	dword					
2	--					
3	--			«Ноль» АЦП (2 канала)		
4	--					
Служебная информация	1	dword			Метки	
	...					
	680	dword				

Рис. 1. Структура файла данных для случая работы четырехканального коррелятора

К исходным данным также относятся ионограммы, получаемые с помощью автоматической ионосферной станции. С использованием своего формата [5] они образуются каждые 15 мин и представляют собой файлы объемами по 12 МБайт.

Под *выходными данными* понимаются результаты расчета параметров среды [1], которые являются основой для дальнейшего геофизического анализа. Набор высотного распределения ионосферных параметров, представленных внутри этих файлов объемами по 0.1 Мбайт для сеансов длительностью 15 мин следующий:

- температура ионов  $T_i(h)$ ;
- температура электронов  $T_e(h)$ ;
- концентрация электронов  $N_e(h)$ ;
- относительное содержание ионов кислорода  $O^+(h)$ ;
- относительное содержание ионов водорода  $H^+(h)$ ;
- относительное содержание ионов гелия  $He^+(h)$ ;
- относительное содержание молекулярных ионов  $M^+(h)$ ;
- радиальная составляющая скорости переноса плазмы  $V_r(h)$ .

Под *служебными данными* будем подразумевать сведения, поступающие из журналов операторной записи особенностей работы радиопередающего и радиоприемного устройств радара НР, а также сведения из журнала работы АИС. С этих журналов переносится информация о режимах аппаратурной работы и алгоритмах обработки, о помеховой обстановке, внештатных ситуациях, результатах экспресс-анализа. Также переносится указание условий проведения экспериментов (плановые измерения или измерения, связанные с затмением Солнца, со стартами ракет, прочее). Присутствуют и связывающие эксперимент данные (время затмения, максимальное покрытие диска, время старта и характеристики ракетносителя, расстояние до места старта), а также значения геофизических параметров (уровень солнечной активности, состояние магнитного поля Земли и др.). Подобная информация нужна при сопоставлении выходных данных для различных дней, сезонов и лет.

### **База данных**

Таким образом, при проведении расчетов параметров ионосферной плазмы накапливаются большие объемы информации. Из результатов анализа следует, что существующий объем архивной информации, накопленный за период 1990 – 2011 гг., составляет приблизительно 70 *ГБайт*, причем плотность поступающих первичных данных достигла 8 *ГБайт/год*.

Основная задача базы данных заключается в унификации представления разнотипной ионосферной информации и устранении ее дублирования. То есть база должна, во-первых, производить периодическое накопление и хранение практически неограниченных объемов экспериментальных данных и сопровождающую их информацию. Во-вторых, она должна позволять осуществлять регулярные и оперативные расчеты и уточнения характеристик изучаемых ионосферных процессов. Как один из примеров, например, потребуются однородные высотно-временные выборки по ряду интересующих пользователя параметров с целью проверки выдвигаемых статистических гипотез.

На диаграмме развертывания (рис. 2) показана структурная реализация предлагаемых процедур передачи информации с использованием технических и программных средств, выполняющих роль организаторов движения потоков ионосферной информации в процессе ее накопления и анализа.

Генерализация файлов на основе заголовка, т.е. обобщение информации по времени, значительно упростит поиск по различным параметрам и ускорит доступ к ней. Для этого необходимо реализовать преобразование различных частей файла, содержащих информацию об одном параметре, в одномерные массивы. Предлагается процедура формирования пакетов исходных данных с последующей их передачей через ftp-server. Передача исходных данных в Харьков на основной proftpd-server предусматривает также и шифрование пакетов с последующей проверкой их целостности.

Предполагается функционирование модуля экспорта, который автоматически формирует пакеты установленного формата, в том числе и для Международной базы данных Madrigal (США, Колорадо), а также модулей обработки и визуализации исходных и выходных данных через WCF- и Web-servers.

Следует отметить, что имеет место некоторая сложность бесперебойного функционирования такой структуры, которая вызвана частичной территориальной разобщенностью радиолокационных систем Института ионосферы по отношению к системам последующей обработки и геофизического анализа. Когда возникает необходимость в оперативной передаче исходной информации при полевых условиях работы (окрестности г. Змиева, где отсутствуют ретрансляционные станции), не всегда удается обеспечить без сбоев функционирование сети Internet даже при скорости в 128 *КБайт/с*.

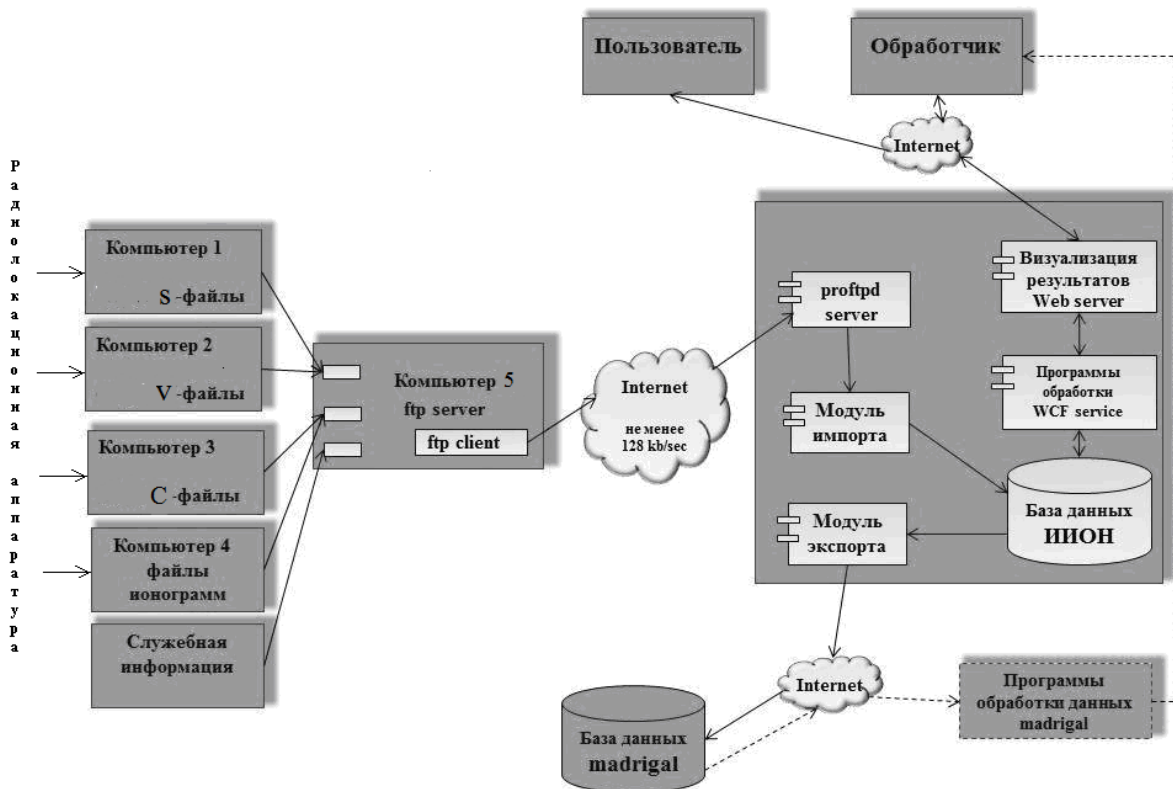


Рис. 2. Диаграмма развертывания ионосферной информации

### Выбор системы управления базой данных

Все СУБД могут быть поделены на настольные и промышленные. Настольные СУБД, такие как Access и FoxPro, предназначены для создания либо автономных информационных систем, либо систем файл-серверного типа. Большой интерес представляют промышленные СУБД, такие как Oracle, MS SQL Server, Postgress и другие, которые предназначены для построения клиент-серверных информационных систем и, как правило, предоставляют разработчику язык программирования, который включает в себя специализированный язык управления базами данных. Для наиболее распространенных баз данных реляционного типа таким языком является SQL.

В нашем случае использование СУБД при построении информационной ионосферной системы призвано реализовать физическую и логическую независимость прикладного программирования от хранимых данных. Физическая независимость от данных заключается в том, что работа программного обеспечения информационных систем не будет зависеть от изменений, которые могут происходить на внутреннем, физическом уровне. Эти изменения, например, могут заключаться в том, что дальнейшая модернизация корреляторов вызовет усовершенствование структуры исходной файловой системы, или же смену наполнения выходных файлов базы данных. Достоинство же гибкого варианта управления позволит вводить новые элементы в структуру данных (к примеру, добавлять информацию о новом ионосферном параметре в виде еще одного столбца таблицы), никак не нарушая при этом функционирование программного обеспечения.

Основные требования, предъявляемые к СУБД для данного объекта (набор радиофизической и геофизической информации по результатам работы радара НР), предполагают наличие следующих возможностей:

- поддержка больших объемов ионосферных данных;
- поддержка индексирования данных;

- простота лицензирования, бесплатность, открытость исходных кодов, поддержка научным сообществом;
  - поддержка механизма партиционирования (процедур разбиения информационных таблиц на подтаблицы на основе, например, дат проведения экспериментов);
  - поддержка необходимых типов данных (массивов, бинарных файлов, графического вида ионограмм и пр.);
  - наличие инструментов для работы с БИД (администрирования, конфигурирования, резервного копирования);
  - компрессия данных для минимизации информационных объемов.
- Исходя из таких требований, в качестве наиболее оптимальной была выбрана система PostgreSQL, представляющая собой свободную объектно-реляционную программу управления базами данных.

### **Выводы**

В результате проведенных исследований по выбору электронной базы данных для хранения сведений о параметрах ионосферы, рассчитываемых при работе радиолокационных систем, использующих метод некогерентного рассеяния радиоволн, получены следующие результаты:

1. Проведен анализ объемов и внутренних структур файлов, содержащих исходные и модифицируемые результаты расчетов в процессе функционирования устройств обработки ионосферной информации,
2. Рассмотрена возможность преобразования и импорта радиофизической и геофизической информации в базу данных с целью типизации ее структуры.
3. Сформулированы основные требования к системе управления базой данных и произведен ее выбор.

**Список литературы:** 1. Пуляев, В.А., Дзюбанов, Д.А., Домнин, И.Ф. Определение параметров ионосферы методом некогерентного рассеяния радиоволн. – Харьков : НТУ “ХПИ”, 2011. – 240 с. 2. Пуляев, В.А. Программное обеспечение автоматизированной системы радара некогерентного рассеяния // Вестн. НТУ “ХПИ” : сб. науч. тр. – Харьков : НТУ “ХПИ”, 2003. – № 26. – С. 91–94. 3. Гарсия-Молина, Г., Ульма, Дж., Уидом, Дж. Системы баз данных. Полный курс. – М. : Вильямс, 2008. – 1088 с. 4. Чекалов, А.П. Базы данных: от проектирования до разработки приложений. – СПб. : Питер, 2006. – 384 с. 5. Богомаз, А.В., Козлов, С.С., Пуляев, В.А. Данные для базы Института ионосферы // Конф. молодых ученых «Дистанционное радиозондирование ионосферы (ИОН-2011)» (Харьков, Украина, 12 – 15 апреля 2011 г.) : сб. тезисов. – 2011. – С. 47. 6. Богомаз, А.В. Особенности процедуры накопления ионосферных данных, полученных с помощью многоканального корелятора // Програма XIX Міжнар. наук.-практ. конф. «Інформаційні технології: наука, техніка, технологія, освіта, здоров’я». Секція 17. Навколоземний простір. Радіофізика та іоносфера. – 2011. – С. 184. 7. Семенов, В.А., Морозов, С.В., Порох, С.А. Стратегии объектно-реляционного отображения: систематизация и анализ на основе паттернов // Тр. Ин-та Системного Программирования РАН. – 2004. – Т. 8, ч. 2. – С. 53 – 92.

*Национальный технический университет «ХПИ»*

*Поступила в редколлегию 02.04.2012*