

С.С. КОЗЛОВ, аспирант, НТУ “ХПИ”, Харьков

ДАнные И СУБД ДЛя БАЗы ДАНных ХАРьКОВСКОГО РАДАРА НЕКОГЕРЕНТНОГО РАССЕЯНИЯ

Розглянуті дані та вироблена їх класифікація, організація бази даних, зроблено вибір системи управління базою даних.

Рассмотрены данные и произведена их классификация, организация базы данных, произведён выбор системы управления базой данных.

Information is considered and their classification, organization of database, is produced, the choice of the database management system is produced.

Введение. Увеличивающийся объём ионосферных данных, получаемых в Институте ионосферы с помощью радара некогерентного рассеяния и автоматической ионосферной станции “Базис”, привёл к необходимости создания базы данных (БД). Создание БД позволит улучшить информационное обеспечение, повысить быстродействие процесса обработки ионосферных данных измерительного комплекса, появляется возможность с любого рабочего места организации получить доступ к информационному ресурсу, за счёт подключения локальной сети к Internet или создания внутрикорпоративной сети на базе Intranet.

Предложена классификация данных, которые целесообразно хранить в БД Института ионосферы. Рассмотрена структура данных и возможность их оптимального размещения в БД. Предложена система управления БД (СУБД) для БД харьковского радара некогерентного рассеяния и автоматической ионосферной станции “Базис”.

Анализ литературы. Проведённый анализ показал, что данные Института ионосферы, которые предполагается хранить в базе, можно разделить на три группы.

– исходные данные, которые формируются системой первичной обработки радара НР [1] и блоком цифрового преобразования автоматической ионосферной станции “Базис”:

– служебные данные, которые необходимы для корректной обработки исходных данных.

– выходные данные, которые представляют собой высотные и временные распределения значений таких геофизических параметров.

Существующие подходы к созданию модели данных. Одним из этапов проектирования БД является даталогическое проектирование, представляющее собой создание логической структуры данных [2].

Одними из первых таких моделей стали иерархические, сетевые и модели на основе инвертированных списков. Однако сложности их практического использования, увеличение объёма и структурной

сложности хранимых данных привели к широкому распространению реляционных СУБД [4].

Реляционная модель состоит из трёх частей, описывающих разные аспекты реляционного подхода: структурной части, манипуляционной части и целостной части. Преимуществом реляционных БД является наличие мощного и в тоже время простого языка SQL – Structured Query Language, позволяющего выполнять многие необходимые пользователям операции.

Одним из недостатков реляционной СУБД является отсутствие в ней сложных типов данных (таких, как рисунки, графики, диаграммы и т. п.) и абстрактных типов, определяемых пользователем.

Попытка встроить семантику в сам механизм управления БД подтолкнула к созданию объектно-ориентированной модели, в основу которой положены принципы, характерные для объектно-ориентированного программирования: инкапсуляция, полиморфизм и наследование.

Преимуществом объектно-реляционных СУБД является сохранение преемственности с системами предыдущего поколения, а также возможность применения уже зарекомендовавших себя методов хранения информации в реляционных СУБД, возможность использования новых объектно-ориентированных механизмов, повышающих эффективность хранения и обработки данных.

Постановка задачи. В ходе анализа возможности оптимизации системы хранения и обработки радиофизических данных был выявлен существенный недостаток, суть которого заключается в несовершенстве хранения, а именно записи больших объёмов информации (до 10 Гб/сут) после первичной обработки на CD-диски. Современная система хранения и обработки требует преобразований своей структуры. Предложено провести анализ данных, получаемых в Институте ионосферы, и рассмотреть возможность создания электронной БД, которая позволит существенно снизить время на обработку информации.

Анализ входных и выходных данных. В целом, структура базы данных радара должна учитывать особенности использования метода НР. Она предназначена для хранения радиотехнической информации, получаемой в процессе геофизических экспериментов, хранения измеренных высотно-временных зависимостей АКФ и спектров сигнала НР, вычисленных высотно-временных вариаций параметров ионосферы и результатов их анализа. Анализ входной информации показал, что помимо файлов, получаемых с корреляторов, и файлов ионограмм, в БД необходимо хранить и служебную информацию, содержащую метки к данным файлов корреляторов, причины проведения экспериментов, техническое состояние аппаратуры. Таким образом, данные целесообразно разделить на группы [3].

Исходные данные, которые получаются на выходе системы первичной обработки. В настоящее время производится корреляционная обработка на трёх персональных компьютерах. Форматы выходных файлов каждой из систем сходны. Файлы состоят из заголовка (рис. 1), который содержит дату и время, число зондирующих импульсов, а значит и развёрток дальности, из которого можно рассчитать время накопления, и накопленных автокорреляционных функций (АКФ) некогерентного рассеянного сигнала, высотных профилей мощности сигнала НР, значений “нуля” АЦП вдоль развёртки дальности.

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

Рис. 1. Структура файла четырёхканального коррелятора

Исходные файлы формируются в течение одной минуты, записываются в двоичном коде и имеют приблизительные размеры 18, 21 и 218 Кб. К исходным данным также относятся ионограммы, получаемые с помощью автоматической ионосферной станции “Базис”. Ионограммы формируются в течение 15 мин, представляют не типизированные файлы и имеют размер около 12 Мб.

Из результатов ориентировочного расчёта следует, что существующий объём исходной информации, накопленной за период с 1990 по 2010 гг., составляет приблизительно 70 Гб, а объём поступающей первичной информации в год приблизительно около 8 Гб.

Под служебными данными будем понимать такие данные, которые необходимы для корректной обработки первичной информации. Три вида служебных данных, которые следует хранить в БД Института ионосферы:

1) Метки отбракованных данных. Каждый сеанс содержит информацию со 170 либо с 680 высотных участков. Удаление данных (создание меток) проводится либо вручную, либо в автоматизированном режиме. В любом случае, необходимо иметь различные варианты массивов меток. При создании меток вручную часто руководствуются субъективными критериями, и результаты отбраковки разными операторами будут различны. Результаты автоматизированного режима также будут различаться из-за разных алгоритмов и параметров фильтрации. Поэтому необходимо кроме меток хранить информацию о том, кто проводил фильтрацию, каким программным обеспечением и с какими параметрами.

2) Журналы радиопередающего и радиоприёмного устройств, а также журнал работы АИС “Базис”. В этих журналах содержится информация о режимах работы, помеховой обстановке, внештатных ситуациях, результатах экспресс-обработки.

3) Графики измерений с указанием причины их проведения (плановые измерения или измерения, связанные с затмением Солнца или стартами ракет), связанные данные (время, в которое происходило затмение, максимальное покрытие диска, время старта и характеристики ракетносителя, расстояние до места старта), а также известные значения геофизических параметров (например, уровень солнечной активности, состояние магнитного поля) во время измерений. Такая информация необходима для сопоставления выходных данных различных дней, сезонов и лет.

Под выходными данными понимаются данные, которые используются в дальнейшем для геофизических расчётов. Минимальный набор геофизических параметров:

- температура ионов $T_i(h, t)$;
- температура электронов $T_e(h, t)$;
- концентрация электронов $N(h, t)$;
- относительное содержание ионов кислорода $[O^+]/[M]$;

- относительное содержание ионов водорода $[H^+]/[N]$;
- относительное содержание ионов гелия $[He^+]/[N]$;
- относительное содержание молекулярных ионов $[M^+]/[N]$;
- вертикальная составляющая скорости переноса плазмы $V_z(h, t)$.

База данных. При проведении расчётов параметров ионосферной плазмы обычно накапливаются большие объёмы информации. Эта информация хранится в виде разрозненных файлов, которые содержат сведения обо всех экспериментальных результатах [5]. Основная задача базы данных заключается в унификации представления разнотипной информации, требуемой для различных алгоритмов и устранении ее дублирования. Она должна, во-первых, производить систематическое накопление и хранение практически неограниченных объёмов как экспериментальных, так и других данных.

Во-вторых, база данных должна дать возможность проводить регулярные и оперативные уточнения характеристик изучаемых процессов. В частности, получать однородные выборки по интересующим пользователя параметрам с целью проверки выдвигаемых статистических гипотез.

На диаграмме развёртывания (рис. 2) показана упрощённая структура обмена потоками информации.

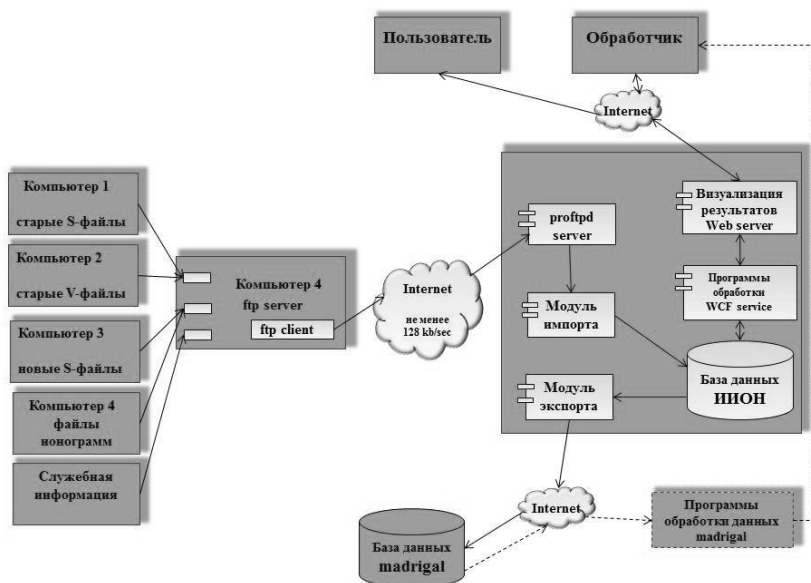


Рис. 2. Диаграмма развёртывания

Очевидно, что из-за территориальной разобщённости измерительного и исследовательского комплексов, необходима оперативная передача исходной информации. Это возможно реализовать с помощью сети Internet со скоростью не менее 128 Кб/с. Предложено формирование пакетов исходных данных с последующей их передачей через ftp-server. При передаче исходных данных в Харьков предусмотрено шифрование пакетов и проверка целостности. Далее исходная информация импортируется в базу данных. Генерализация файлов на основе заголовка, т.е. обобщение информации по времени, значительно упростит поиск по различным параметрам и ускорит доступ к ним. Это также реализуется за счёт преобразования различных частей файла, содержащих информацию об одном параметре, в одномерные массивы. Предполагается модуль экспорта, который автоматически формирует пакеты установленного формата для Международной базы данных радаров НР, а также модули визуализации исходных и выходных данных.

Выбор системы управления базой данных. Все СУБД могут быть поделены на настольные и промышленные. Настольные СУБД, такие как Access, FoxPro предназначены для создания либо автономных информационных систем (ИС), либо ИС файл-серверного типа. Промышленные СУБД, такие как Oracle, MS SQL Server, Postgress и др. предназначены для построения клиент-серверных информационных систем. СУБД, как правило, предоставляет разработчику язык программирования, который включает в себя специализированный язык управления базами данных. Для наиболее распространённых баз данных реляционного типа таким языком является язык SQL.

Использование СУБД при построении информационных систем призвано реализовать физическую и логическую независимость прикладного программирования от данных. Физическая независимость от данных заключается в том, что работа программного обеспечения ИС не будет зависеть от изменений, которые могут происходить на внутреннем, физическом уровне. Эти изменения могут заключаться, например, в том, что будет изменена файловая система или же в том, что изменится структура тех файлов, которые составляют базу данных [6].

Логическая независимость прикладного программирования от данных при использовании СУБД в трёхуровневой структуре доступа к данным заключается, прежде всего, в том, что добавление новых элементов (например, добавление нового столбца в таблицу) в структуру данных никак не влияет на функционирование программного обеспечения. Учитывая основные требования, предъявляемые к СУБД для данного объекта:

- поддержка больших объёмов данных (база/таблица/запись);
- поддержка индексирования данных;
- простота лицензирования, бесплатность, открытость исходных кодов, поддержка сообществом;

- поддержка механизма партиционирования (разбиения таблицы на подтаблицы на основе даты);
- поддержка необходимых типов данных (например, массивов, бинарных файлов для ионограмм);
- наличие инструментов для работы с БД (администрирования, конфигурирования, резервного копирования);
- компрессия данных.

Была выбрана PostgreSQL – свободная объектно-реляционная система управления базами данных.

Выводы. Таким образом, в результате проведённых исследований получены следующие результаты.

- 1) Проанализированы структуры файлов исходной информации.
- 2) Проведена оценка объёма исходной информации.
- 3) Рассмотрена возможность импорта исходной радиофизической и геофизической информации в БД и типизация исходных файлов в БД.
- 4) Представлена возможность нормализации данных.
- 5) Сформулированы основные требования к СУБД и произведён её выбор.
- 6) С использованием преимуществ СУБД было реализовано партиционирование и индексирование данных.

Список литературы: 1. Пуляев В.А., Дзюбанов Д.А., Домнин И.Ф. Определение параметров ионосферы методом некогерентного рассеяния радиоволн: монография – Х.: Учебник НТУ “ХПИ”, 2011. – 240 с. 2. Гарсия-Молина Г., Ульма Дж., Уидом Дж. Системы баз данных. Полный курс. – М.: Вильямс, 2008. – 1088 с. 3. Лысенко В.Н., Кононенко А.Ф., Черняк Ю.В. Корреляционная обработка сигнала некогерентного рассеяния // Вестник Национального технического университета “Харьковский политехнический институт”. Тематический выпуск “Радиофизика и ионосфера”. – 2004. – № 23. – С. 49 – 62. 4. Чекалов А.П. Базы данных: от проектирования до разработки приложений. – СПб.: Питер, 2006. – 384 с. 5. Рыбанов А.А. Инструментальные средства автоматизированного проектирования баз данных. – Волгоград: ВолГТУ, 2007. – 96 с. 6. Семенов В.А., Морозов С.В., Порох С.А. Стратегии объектно-реляционного отображения: систематизация и анализ на основе паттернов // Труды Института Системного Программирования РАН. – 2004. – Т. 8, часть 2. – С. 53 – 92.

Поступила в редакцию 12.10.2011