

РОЗРОБКА ТА ДОСЛІДЖЕННЯ ДЕНОРМАЛІЗОВАНОЇ МОДЕЛІ ДАНИХ ДЛЯ АВТОМАТИЗОВАНОЇ СИСТЕМИ МОНІТОРИНГУ ПАРАМЕТРІВ ІОНОСФЕРИ

Розглянуто проблему підвищення ефективності обробки та збереження даних в автоматизованій системі моніторингу параметрів іоносфери. Проведено аналіз існуючих структур даних з урахуванням їх специфіки. Запропоновано підхід на основі використання денормалізованої реляційної моделі даних, доцільність якого підтверджена експериментальним шляхом. Виконані тести показали збільшення продуктивності обробки даних в 1.7 - 2.1 рази

Ключові слова: модель даних, денормалізація, система моніторингу, параметри іоносфери

Рассмотрена проблема повышения эффективности обработки и хранения данных в автоматизированной системе мониторинга параметров ионосферы. Проведен анализ существующих структур данных с учетом их специфики. Предложен подход на основе использования денормализованной реляционной модели данных, целесообразность которого подтверждена экспериментальным путем. Выполненные тесты показали увеличение производительности обработки данных в 1.7 – 2.1 раза

Ключевые слова: модель данных, денормализация, система мониторинга, параметры ионосферы

В. Є. Сокол

Старший викладач, PhD*

E-mail: vlad.sokol@gmail.com

М. В. Ткачук

Професор, доктор технічних наук*

E-mail: tka@kpi.kharkov.ua

А. Г. Шаповалов

Кафедра автоматизованих систем управління*

*Національний технічний університет

«Харківський політехнічний інститут»

вул. Фрунзе, 21, м. Харків, Україна, 61002

E-mail: andreshapovalov@gmail.com

1. Вступ. Актуальність проблеми та мета дослідження

В останні роки внаслідок все більшого освоєння і практичного використання навколосезного простору різко зросла потреба в прогнозах космічної погоди (space weather): сукупності явищ, що відбуваються у верхніх шарах земної атмосфери і навколосезному космічному просторі [1]. Сьогодні, як і раніше, при вирішенні цієї проблеми значну роль відіграє радіозв'язок, із активним використанням супутникових систем, радіонавігації, глобальної систем позиціонування (GPS). Характеристики розповсюдження радіохвиль безпосереднім чином пов'язані зі станом іоносфери і тому точність прогнозів космічної погоди, в першу чергу, залежить від якості даних щодо стану іоносфери Землі.

Одним з найкращих методів дослідження іоносфери є метод некогерентного розсіювання радіохвиль (НРР) [2].

Він дозволяє з високою точністю по всьому висотному діапазону оцінити параметри іоносферної плазми: електронну концентрацію, температуру електронів та іонів, газовий склад і т.д. Важливою задачею в методі НРР є підвищення ефективності обробки відповідних даних з метою отримання максимально можливого об'єму корисної інформації. Для вирі-

шення цих завдань створюються спеціальні автоматизовані системи моніторингу параметрів іоносфери (АСМШ), до складу яких входять бази даних (БД), що мають забезпечувати певні показники якості обробки даних в таких системах. Харківський радар НРР, який входить до інфраструктури Інституту іоносфери НАН та МОНМС України (далі - Інститут іоносфери) є єдиним і найбільш інформативним джерелом даних щодо основних параметрів іоносферної плазми в середніх широтах Центральноєвропейського регіону [1]. Він дозволяє отримувати наступні параметри іоносфери: концентрацію електронів, температури електронів та іонів, вертикальну складову швидкості перенесення плазми, а також відомості про відносне іонному складі. Діапазон досліджуваних висот становить 100 - 1500 км.

Для Інституту іоносфери дуже важливим є питання організації обміну накопиченими даними з міжнародною науковою спільнотою, зокрема з міжнародною інформаційною системою Madrigal [3], яка містить БД наземних вимірювань параметрів верхньої атмосфери Землі та іоносфери, що створена у відповідності з програмою Coupling, Energetics and Dynamics of Atmospheric Regions (CEDAR) [4]. На даний час така взаємодія технічно ускладнена через те, що немає сучасної інформаційної технології накопи-

чення та подальшої обробки даних, які отримуються з Харківського радару НРР.

Тому проблематика та мета цієї статті, а саме: розробка та дослідження спеціального виду моделі даних, яка має забезпечити підвищення ефективності функціонування БД у складі АСМПІ Інституту іоносфери, є актуальною науково-технічною проблемою.

2. Аналіз існуючих структур даних

В результаті роботи радару НРР за добу отримують 1440 файлів трьох різних форматів [1]: S, SNew, V, кожен з яких має свою структуру та розмір.

Файл S поділяється на три основних блоки: заголовок, власне дані та службова інформація. Заголовок містить дату створення файлу та номер сеансу роботи радару. Блок даних поділяється на 3 сегменти: сегмент із даними про потужність сигналу, сегмент з значеннями автокореляційної функції (АКФ) та сегмент з даними «Нуль» аналогово-цифрового перетворення (АЦП). Сегмент потужності сигналу містить дані, які отримуються з 680 рівней висоти. Сегмент АКФ містить дані з 170 рівней висоти, при цьому кожна висота АКФ відповідає кожній 4-й висоті потужності сигналу. З кожною з висот пов'язаний масив даних, який містить 18 значень АКФ. Сегмент «Нуль» АЦП містить дані з 680 висот, а сегмент із службовою інформацією містить спеціальні маркери, які записуються в файл після первинної обробки даних.

Файл формату SNew містить 2 основних блоки [1]: заголовок та власне дані. Заголовок має ту ж структуру, як і у файлі S. В свою чергу, блок даних поділяється на 3 сегменти: АКФ, «Нуль» АЦП та потужність по короткому імпульсу АЦП. Сегмент дани АКФ поділяється на 4 масиви (канали), кожен з яких містить дані з 680 рівней висоти. З кожною з висот пов'язаний масив даних, який містить 19 значень АКФ. Сегмент «Нуль» АЦП також має 4 канали, між першими двома та останніми розташовується блок потужності по короткому імпульсу. Блок потужності по короткому імпульсу містить 2 канали, кожен з яких має дані з 680 висот.

Файл формату V поділяється на 3 основних блоки [1]: заголовок, власне дані та службова інформація. Блоки заголовку та службової інформації мають те ж саме призначення, що й у файлі S формату. Блок даних поділяється на 2 сегменти: «Додаткова потужність» та «Нуль» АЦП. Сегмент «Додаткова потужність» містить дані з 230 висот, кожна висота, в свою чергу, містить масиви даних з 2-х каналів. Кожен канал містить масив з 10 значень. Блок «Нуль» АЦП містить дані з 230 висот.

Таким чином, навіть попередній аналіз трьох існуючих форматів файлів даних НРР дозволяє зробити висновок, що їх можна описати за допомогою відповідних XML-схем, а потім конвертувати в структуру інтегрованої БД.

Розробка відповідної технології є окремою задачею, вирішення якої виходить за межі цього дослідження, і в подальшому в даній статті розглядається саме проблема побудови ефективної моделі даних для інтегрованої БД.

3. Розробка денормалізованої моделі даних для БД АСМПІ

На підставі проведеного аналізу структури існуючих файлів даних радару НРР (див. попередній розділ) та беручи до уваги комплекс функціональних задач їх подальшої обробки [5], для створення інтегрованої БД АСМПІ доцільно обрати добре відому реляційну модель даних [6 – 8]. В той же час, маючи на меті вирішення основної задачі дослідження: підвищення ефективності обробки даних в системі, на відміну від традиційного підходу до розробки реляційних БД, який передбачає проведення процедури нормалізації схеми реляційної БД до третьої нормальної форми (3NF) [6], було зроблено припущення щодо того, що застосування денормалізованої моделі БД, тобто представлення її у другій нормальної формі (2NF) [6], дозволить скоротити час виконання основних запитів на обробку даних [7 – 8].

Ці альтернативні схеми БД в нотації IDEF1X [6] наведені на рис. 1 та 2 відповідно. Розглянемо більш детально саме першу з них, для того щоб представити переваги запропонованого підходу.

Сутність header (рис. 1) містить інформацію про окремий файл дани НРР, яка є загальною для усіх форматів файлів. Вона складається з: унікального ідентифікатора файлу (id), дати створення (date), номеру сеансу роботи (session), рівня некогерентного розсіяння при довгому імпульсі (nr0), рівня некогерентного розсіяння при короткому імпульсі (nr1), типу файлу (file_type) та додаткової інформації (not_used_data_arr), яка необхідна для приведення файлу до відповідного розміру.

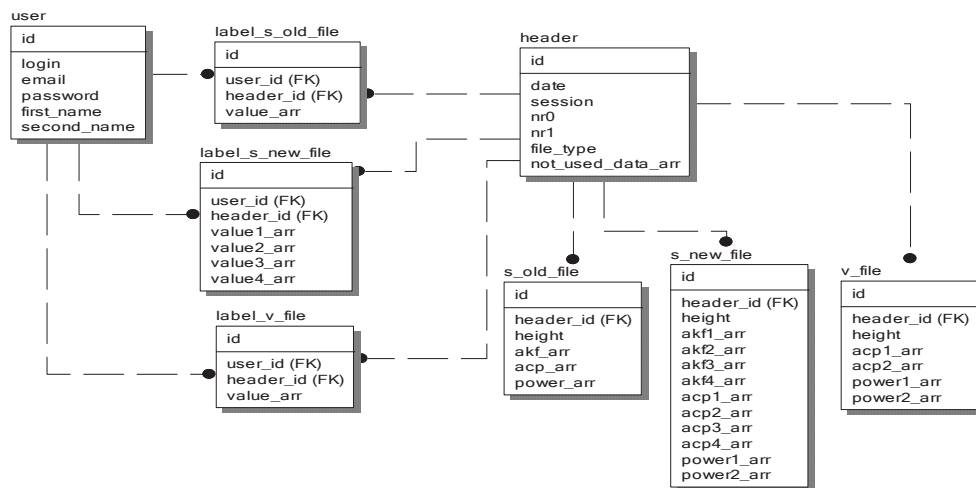


Рис. 1. Нормалізована схема БД (3NF)

Сутність s_old_file містить дані, що отримані з файлу формату S.

Вона складається з наступних атрибутів (field): унікального ідентифікатора запису (id), унікального ідентифікатора файлу (header_id), номеру рівня висоти (height), масиву значень АКФ (acf_arr), масиву значень АЦП (acp_arr), масиву значень потужності (power_arr).

Сутність s_new_file містить дані, що отримані з файлу формату SNew.

Вона складається з наступних атрибутів: унікального ідентифікатора запису (id), унікального ідентифікатора файлу (header_id), номеру рівня висоти (height), масиву значень АКФ першого каналу (acf1_arr), масиву значень АКФ другого каналу (acf2_arr), масиву значень АКФ третього каналу (acf3_arr), масиву значень АКФ четвертого каналу (acf4_arr), масиву значень АЦП першого каналу (acp1_arr), масиву значень АЦП другого каналу (acp2_arr), масиву значень АЦП третього каналу (acp3_arr), масиву значень АЦП четвертого каналу (acp4_arr), масиву значень потужності першого каналу (power1_arr), масиву значень потужності другого каналу (power2_arr).

Сутність v_file містить дані, що отримані з файлу формату V.

Вона складається з наступних атрибутів: унікального ідентифікатора запису (id), унікального ідентифікатора файлу (header_id), номеру рівня висоти (height), масиву значень АКФ першого каналу (acf1_arr), масиву значень АКФ другого каналу (acf2_arr), масиву значень АЦП першого каналу (acp1_arr), масиву значень АЦП другого каналу (acp2_arr), масиву значень потужності першого каналу (power1_arr), масиву значень потужності другого каналу (power2_arr).

Сутність label_s_old_file містить мітки для кожної висоти по одному каналу для файлу формату S.

Вона складається з атрибутів: унікального ідентифікатора запису (id), унікального ідентифікатора файлу (header_id) та масиву спеціальних міток (value_arr).

Сутність label_s_new_file містить мітки для кожної висоти по кожному з чотирьох каналів для файлу формату SNew.

Вона складається з наступних атрибутів: унікального ідентифікатора запису (id), унікального ідентифікатора файлу (header_id), масиву міток (value_arr1) для першого каналу, масиву міток (value_arr2) для другого каналу і т. д.

Для побудови денормалізованої схеми БД (рис. 2) врахуємо ту обставину, що в АСМПІ в основному будуть виконуватися запити на отримання значень затримок АКФ, АЦП та потужності для сеансів роботи радару НРР за певні проміжки часу, а в алгоритмах аналітичної обробки даних переважно використовується інформація про всі висотних ділянках [5]. Це дає можливість зберігати в БД двомірний масив даних для кожної висоти у вигляді певної послідовності байтів.

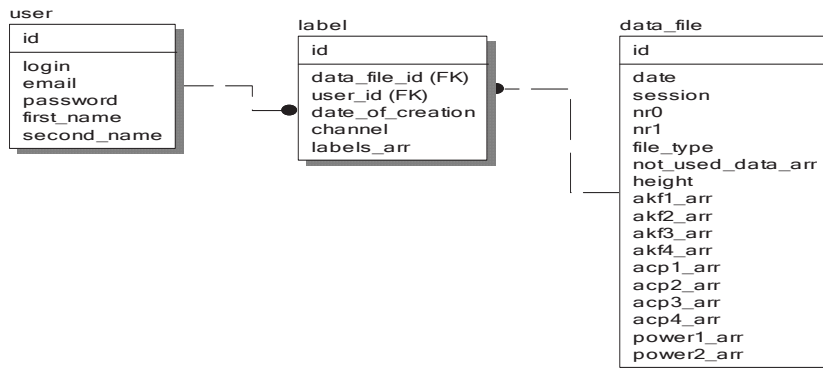


Рис. 2. Денормалізована схема БД (2NF)

В кінцевому рахунку, такий вигляд моделі даних для інтегрованої БД у складі АСМПІ дозволить суттєво зменшити загальний час на конвертацію цих параметрів для подальшого застосування алгоритмів обробки даних.

4. Експериментальне дослідження запропонованої моделі даних

Для перевірки ефективності запропонованого підходу обидві схеми БД були програмно реалізовані з використанням таких технологій як EMS SQL Manager 2007 [8] та NetBeans 7.0 [9], а для проведення обчислювальних експериментів до цих БД були імпортовані дані з 100 файлів формату Snew (розмір одного файлу складав при цьому приб. 200 Кб).

У такий спосіб були отримані наступні експериментальні дані.

4.1. Експеримент №1

За допомогою програми EMS SQL Manager 2007 [10] визначається, скільки часу виконується простий SQL-запит на вибірку даних (SELECT). Для цього:

- створюється запит на вибірку даних із імпортованих файлів;
- фіксується час, за який виконався цей запит.

4.2. Експеримент №2

За допомогою технології NetBeans 7.0 визначається, скільки часу займає отримання даних з БД для подальшого виконання процедури відновлення даних після фільтрації [5].

Для цього:

- в програмі, що обробляє процедуру відновлення даних встановлюються точки для виміру відповідних моментів часу;
- створюється запит на вибірку даних з імпортованих файлів;
- визначається загальний час отримання даних.

4.3. Експеримент №3

За допомогою технології NetBeans 7.0 визначається, скільки часу виконується саме процедура відновлення даних [5] після процедури їх фільтрації. Для цього:

- безпосередньо в процедурі відновлення даних встановлюються точки для виміру відповідних моментів часу;
- створюється запит на вибірку даних з імпортова-

них файлів;

- виконується процедура відновлення даних після фільтрації;

- визначається загальний час виконання процедури відновлення даних.

Результати проведених експериментів наведені в табл. 1.

Таблиця 1

Результати тестування БД, які створені на основі нормалізованої та денормалізованої моделі даних

№ експерименту	Нормалізована схема БД	Денормалізована схема БД
1	10,764 с	5,163 с
2	32,356 с	15,270 с
3	61,012 с	34,095 с

Таким чином, проведені експерименти свідчать, що, загалом, запити до денормалізованої схеми БД виконуються приблизно в 1,7 – 2,1 разів швидше, ніж до її нормалізованої версії.

5. Висновки

В результаті аналізу предметної області було розглянуто основні проблеми, що пов'язані з організацією процесів накопичення та обробки даних в автоматизованій системі моніторингу параметрів іоносфери (АСМПІ). Для підвищення ефективності функціонування відповідної БД цієї системи була розроблена її денормалізована реляційна модель даних, що враховує особливості алгоритмів обробки параметрів іоносфери. Шляхом проведення обчислювальних програмних експериментів встановлено, що цей підхід дозволяє суттєво підвищити швидкість обробки даних в АСМПІ.

В подальшій роботі планується розширити існуючу модель даних за рахунок збереження таких показників стану іоносфери як: кількість іонів, електронів і т.п.; та деякої додаткової інформації про стан радару в різні моменти його функціонування, оскільки це є дуже важливим при дослідженні іоносфери за певний період часу. Також планується провести аналіз швидкодії виконання запитів при збільшенні розміру БД до значення 80 Гб, що відповідає фактичному об'єму даних, що які були накопичені в АСМПІ за весь період її експлуатації.

Література

1. Інститут іоносфери НАН та МОНМС України [Електронний ресурс] – Режим доступу: \WWW/ URL: <http://www.iion.org.ua/> – 15.07.2013 р.
2. Альперт, Я. Л. Распространение электромагнитных волн и ионосфера [Текст] / Я. Л. Альперт. – М. : "Наука", 1972. – 564с.
3. Проект Madrigal [Електронний ресурс] – Режим доступу: \WWW/ URL: <http://www.openmadrigal.org/> – 15.07.2013 р.
4. Програма CEDAR [Електронний ресурс] / Національний науковий фонд США. – Режим доступу: \WWW/ URL: <http://www.nsf.gov/pubs/2005/nsf05554/nsf05554.htm> – 15.07.2013 р.
5. Богомаз, А. В. Этапы обработки данных радара некогерентного рассеяния [Текст] / А. В. Богомаз // Конференция молодых учёных «Дистанционное радиозондирование ионосферы (ИОН-2012)». – Харьков, Украина, 18 – 20 апреля 2012 г.
6. Коннолли, Т. Базы данных. Проектирование, реализация и сопровождение. Теория и практика [Текст] / Т. Коннолли. – М. : Издательский дом "Вильямс", 2003. – 1440 с.
7. Корнеев, В. В. Базы данных. Интеллектуальная обработка информации. / В. В. Корнеев, А. Ф. Гараев. – М. : Издательство "Нолидж", 2000 – 450 с.
8. Энсор, Д. Oracle. Проектирование баз данных / Д. Энсор, Й. Стивенсон. – М. : Издательство "ЛОРИ", 2003 – 624с.
9. NetBeans IDE [Електронний ресурс] – Режим доступу: \WWW/ URL: <https://netbeans.org/features/index.html/> – 15.07.2013 р.
10. EMS SQL Manager [Електронний ресурс] – Режим доступу: \WWW/ URL: <http://www.sqlmanager.net/products/mysql/manager/> – 15.07.2013 р.