УДК 528.855 **doi: 10.20998/2522-9052.2020.2.11**

С. М. Андрєєв, В. А. Жилін

Національний аерокосмічний університет імені М. Є. Жуковського «ХАІ», Харків, Україна

ГЕОІНФОРМАЦІЙНА СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ НА БАЗІ СХОВИЩА ПРОСТОРОВИХ ДАНИХ ГЕОПОРТАЛУ

Анотація. Предметом дослідження є розроблення геоінформаційної системи підтримки прийняття рішень (СППР) на базі сховища просторових даних (СПД) геопорталу. **Об'єктом дослідження** є процес об'єднання та інтеграції у сховище геопорталу інформації, що містить різні типи просторових даних. **Метою роботи** є підвищення оперативності статистичного і прогнозного аналізу геоданих для ведення звітності та підтримки прийняття рішень геоінформаційних задач в режимі реального часу. **Висновки.** На підставі проведеного аналізу архітектур сучасних систем підтримки прийняття рішень запропоновано структуру, яку найдоцільніше використовувати для побудови геоінформаційної СППР на базі сховища просторових даних геопорталу. Проаналізовано основні функції, що мають виконувати підсистеми такої СППР, а саме — функції підсистеми витягання, перетворення і завантаження даних, функції підсистеми зберігання даних та підсистеми аналізу даних. Розгляд зазначених підсистем СППР проведено із детальним вивченням їх структурних та функціональних схем, аналізом існуючих технологій обробки та зберігання просторової інформації, а також сучасних методів аналізу геоданих. Разом з тим, здійснено ґрунтовний вибір системи управління базами даних (СУБД) для практичної реалізації багатовимірної бази геоданих, що виконує функції сховища просторових даних як основи СППР геопорталу. Проведено ретельний аналіз засобів Microsoft SQL Server Analysis Services (SSAS) як базової платформи для розгортання аналітичних систем Business Intelligence (BI), уніфікованої багатовимірної моделі даних Unified Dimensional Model (UDM) як ядра технології комплексного багатовимірного аналізу даних Online Analytical Processing (OLAP). Проаналізовано аспекти планування і програмування архітектури SSAS, особливості розроблення багатовимірних баз даних із застосуванням середовищ SQL BI Development Studio (BIDS) та SQL Server Management Studio (SSMS), а також специфіку застосування служби SQL Server Integration Services (SSIS) для роботи з багатовимірними базами даних. Таким чином, підготовлено підґрунтя для розроблення геоінформаційної СППР із застосуванням засобів Microsoft SQL Server багатовимірного аналізу даних для створення багатовимірного сховища просторових даних геопорталу. Врешті, запропоновано методику програмної реалізації багатовимірного сховища просторових даних геопорталу. Запропоновані підходи щодо розроблення геоінформаційної СППР на базі сховища просторових даних геопорталу забезпечують об'єднання та інтеграцію у сховище геопорталу інформації, що містить різні типи просторових даних. Програмна реалізація багатовимірного СПД геопорталу забезпечує підвищення оперативності статистичного і прогнозного аналізу геоданих для ведення звітності та підтримки прийняття рішень геоінформаційних задач в режимі реального часу.

Ключові слова: геодані; геоінформаційна система підтримки прийняття рішень; багатовимірні сховища просторових даних.

Аналіз архітектури сучасних систем підтримки прийняття рішень

До теперішнього часу у багатьох організаціях, працюючих у сфері геоінформаційних технологій, накопичені значні об'єми даних, на основі яких вирішуються різноманітні аналітичні та управлінські задачі. Проблеми зберігання великих об'ємів даних, а особливо просторово-прив'язаної інформації, стають усе більш актуальними і привертають увагу фахівців, працюючих в цій області, що привело до формування сучасного ринку технологій аналізу тематично-просторових даних щодо їх зберігання.

Робота аналітиків та керівників усіх рівнів бажано має бути організована так, щоб вони могли мати доступ до будь-якої інформації, що цікавить, користуватися зручними та простими засобами представлення і роботи з цією інформацією. Саме на досягнення цих цілей й спрямовані інформаційні системи підтримки прийняття рішень (рис. 1).

Збір та зберігання інформації, а також рішення завдань інформаційно-пошукового запиту ефективно реалізуються засобами систем управління базами даних.

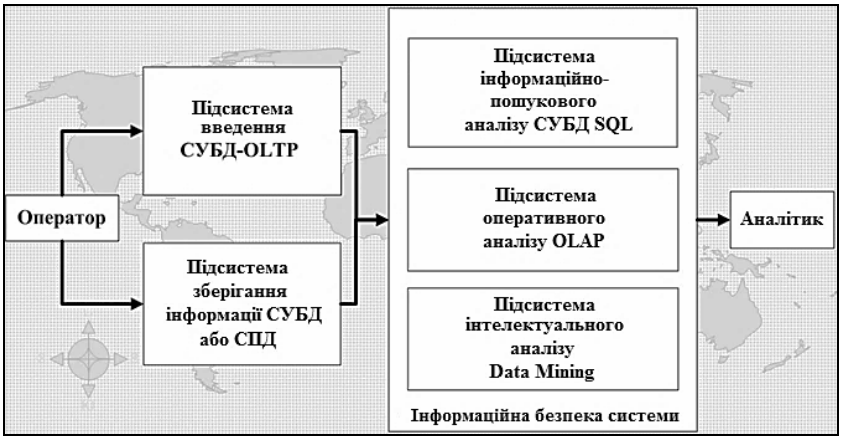

Рис. 1. Архітектура системи підтримки прийняття рішень (СППР)

Транзакційні OLTP-системи (*англ.* Online Transaction Processing, обробка транзакцій в реальному часі) – системи, призначені для введення, структурованого зберігання та обробки інформації в режимі реального часу. Вони проектуються, настроюються та оптимізуються для виконання максимальної кількості транзакцій за короткі проміжки часу. При цьому використовується фіксований набір надійних та безпечних мето-

© Андрєєв С. М., Жилін В. А., 2020

дів введення, модифікації, видалення даних і випуску оперативної звітності.

Однак, аналітичні можливості OLTP-систем значно обмежені (або взагалі відсутні), а отже OLTP-системи не підходять для повноцінного аналізу інформації в силу суперечності вимог, що пред'являються до OLTP-підсистем у складі СППР.

Для надання необхідної для прийняття рішень інформації зазвичай доводиться збирати дані з декількох транзакційних баз даних різної структури та змісту. Основна проблема при цьому полягає в неузгодженості і суперечності баз-джерел та відсутності єдиного логічного погляду на дані.

З метою об'єднання в одній системі OLTP та СППР для реалізації підсистеми надійного зберігання використовується концепція сховищ даних (СД). В основі концепції СД лежить ідея розділення даних, використовуваних для оперативної обробки та вирішення завдань аналізу, що дозволяє оптимізувати структури зберігання. Концепція СД дозволяє інтегрувати раніше роз'єднані деталізовані дані, що містяться різних зовнішніх джерелах, в єдину базу даних за рахунок їх попереднього узгодження та агрегації. Підсистема аналізу може бути побудована на основі:

– підсистеми інформаційно-пошукового аналізу на базі реляційних СУБД і статичних запитів з використанням мови SQL;

– підсистеми оперативного аналізу, що використовує технології оперативної аналітичної обробки даних OLAP в рамках концепції багатовимірного представлення даних;

– підсистеми інтелектуального аналізу, що реалізує методи і алгоритми Data Mining.

Сховище даних (Data Warehouse) — предметно-орієнтована інформаційна база даних, призначена для збору і зберігання інформації, підготовки звітів, аналізу інформації з метою підтримки прийняття рішень. Будується на базі клієнт-серверної архітектури, реляційної СУБД та утиліт підтримки прийняття рішень.

Зберігання просторових даних в ГІС-системах файловим способом перестало бути ефективним: об'єми даних постійно ростуть, з'являється необхідність змішаних запитів, спільної обробки різнорідних даних (як просторових, так і атрибутивних), а також одночасного доступу до них багатьох користувачів. У зв'язку з цим окреме зберігання просторових даних у файлах ускладнює і гальмує процес роботи з усією інформацією в цілому. Сховища просторових даних (СПД) дозволяють розв'язати цю проблему, адже забезпечують:

– цілісне зберігання інформації про об'єкт, як просторової, так і описової, в єдиній базі даних;

– можливість роботи, як з векторною, так і з растровою інформацією;

– можливість швидкого пошуку і аналізу даних одночасно за всіма характеристиками;

– підтримку просторового індексування, що дозволяє істотно підвищити продуктивність роботи з просторовими об'єктами;

– використання можливостей стандартної мови запитів, що надає можливість обробляти просторові дані на стороні сервера, збільшуючи продуктивність, зменшуючи вимоги до функціональності ГІС та надаючи єдині правила обробки даних;

– мінімізацію витрат на модернізацію існуючих систем;

– підтримку стандартів OpenGIS.

Окрім можливості працювати з єдиним джерелом інформації, аналітики повинні мати зручні засоби візуалізації даних, агрегації, пошуку тенденцій, прогнозування. Незважаючи на різноманіття аналітичної діяльності, можна виділити типові технології аналізу даних, кожній з яких відповідає певний набір інструментальних засобів. Разом зі сховищем даних ці засоби забезпечують повне рішення автоматизації аналітичної діяльності та створення СППР.

При завантаженні даних з OLTP-системи в СД відбувається дублювання даних. Проте в ході цього завантаження дані фільтруються, оскільки не усі з них мають значення для проведення процедур аналізу. У СД зберігається узагальнена інформація, яка в OLTP-системі відсутня (рис. 2).

Надмірність інформації можна звести до нуля, використовуючи віртуальне СД. У такій системі дані з OLTP-системи не копіюються в єдине сховище. Вони витягаються, перетворюються та інтегруються безпосередньо при виконанні аналітичних запитів в режимі реального часу. Перевагою віртуального СД є мінімізація об'єму даних, що зберігаються, і робота з поточними, тобто актуальними даними.

Недоліками віртуального СД можна вважати більш високий, у порівнянні з фізичним СД, час обробки запитів, необхідність постійної доступності усіх OLTP-джерел, зниження швидкодії OLTP-систем. OLTP-системи не орієнтовані на зберігання даних за тривалий період часу, в міру необхідності дані вивантажуються в архівні, тому не завжди є фізична можливість отримання повного набору даних з СД.

Скорочення витрат на проектування і розробку СД може бути досягнуте шляхом створення вітрин

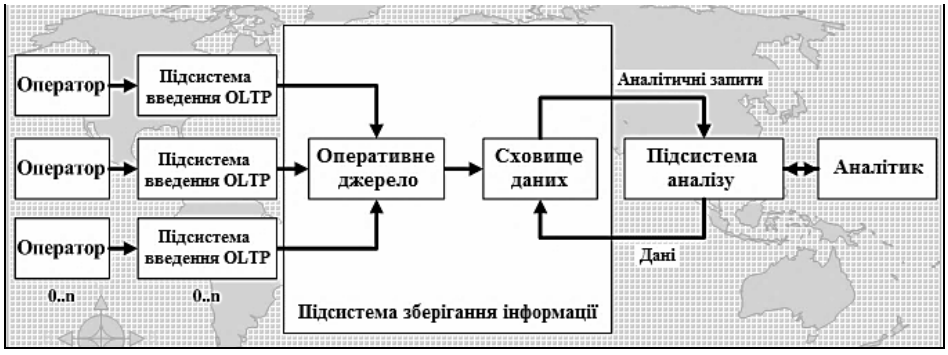

Рис. 2. Структура СППР з фізичним сховищем даних

даних (ВД). ВД — це спрощений варіант СД, що містить тільки тематично об'єднані дані (рис. 3).

ВД містять дані, орієнтовані на конкретного користувача, отже вони істотно менше фізичного СД за об'ємом і для їх реалізації потрібно менше витрат. ВД розробляються та впроваджуються набагато швидше СД, при цьому ефект від їх використання стає помітним відразу після впровадження.

Недоліками ВД є багатократне зберігання одних і тих же даних в різних ВД, а також відсутність консолідованості на рівні предметної області.

Звичайно, можливе й комплексне використання сховищ даних та вітрин даних в системах підтримки прийняття рішень. СД в цьому випадку є єдиним джерелом даних для усієї предметної області, а ВД – підмножинами даних зі сховища, організованими для представлення інформації по тематичних розділах цієї області. У тому випадку, якщо користувачеві недостатньо даних у ВД, яка для нього створювалась, він може звернутися до СД (рис. 4). Перевагами такого рішення є простота створення й наповнення ВД, оскільки наповнення походить з єдиного стандартизованого джерела очищених даних, тобто з СД, простота розширення за рахунок додавання нових ВД, а також зниження навантаження на основне СД. Недоліки полягають у надмірності, оскільки дані зберігаються і в СД, і у ВД, а також додаткових витратах на розробку СППР із СД та ВД.

Основна проблематика при проектуванні та створенні СПД в СППР полягає в такому.

1. *Інтеграція різнорідних даних*. Дані в СПД поступають з різнорідних OLTP-систем, які фізично можуть бути розташовані на різних вузлах мережі. При проектуванні і розробці СПД необхідно вирішувати задачу інтеграції різних програмних платформ зберігання.

2. *Ефективне зберігання і обробка великих об'ємів даних.* Побудова СПД припускає накопичення даних за значні періоди часу, що веде до постійного зростання об'ємів дискової пам'яті, а також зростання об'ємів оперативної пам'яті, які потрібні для обробки цих даних.

3. *Організація багаторівневих довідників метаданих*. Кінцевим користувачам СППР потрібні метадані, що описують структуру та характеристики даних, які зберігаються в СПД, а також інструменти їх візуалізації.

4. Забезпечення інформаційної безпеки СПД. Інформація про діяльність компанії, як правило, відноситься до комерційної таємниці та підлягає захисту; крім того, в СПД можуть міститися персональні дані клієнтів та співробітників, які також необхідно захищати. Для виконання цієї функції має бути розроблена політика безпеки СПД та пов'язаної з ним інфраструктури, а також реалізовані передбачені в політиці організаційні та програмно-технічні заходи щодо захисту інформації.

Просторова інформація є основною для повсякденної діяльності організацій, що надають послуги у сфері геоінформаційних технологій. Усі отримані просторові дані піддаються аналізу та оцінці, тобто досліджуються їх властивості та якість, можливість служити джерелами для вирішення завдань геоінформатики тощо.

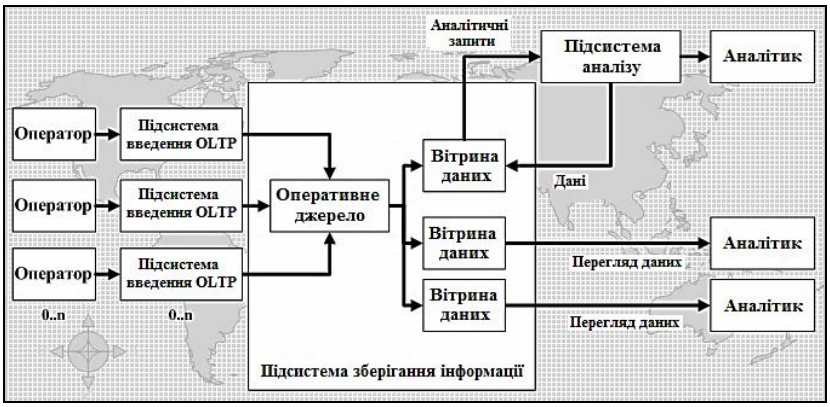

Рис. 3. Структура СППР із самостійними вітринами даних

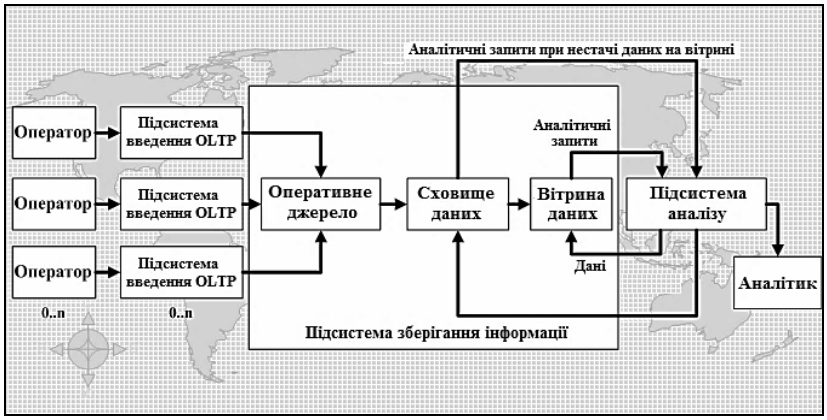

Рис. 4. Структура СППР зі сховищем даних та вітринами даних

Постановка задачі

Отже, із зазначеного вище виникає актуальна задача створення СПД, аспекти якої полягають в такому.

1. *Оцінка математичної основи* полягає в тому, щоб з'ясувати доцільність прийнятого масштабу, придатність використовуваної проекції з точки зору величини й характеру розподілу спотворень, а також можливість використання просторових даних для кількісних визначень із заданою точністю.

2. *Оцінка наукової достовірності* має встановлювати відповідність прийнятим науковим концепціям, правильність передачі реально існуючих просторових закономірностей і зв'язків, типових рис явищ.

3. *Оцінка повноти й сучасності,* передусім, торкається об'єму інформації, що знаходиться в джерелі просторової інформації. Головну роль грають два чинники: вивченість явища й призначення джерела. Від цього залежать відбір

об'єктів, що представляються, подробиці генералізації, способи графічного оформлення. Навантаження джерела може бути оцінене кількісно, наприклад, шляхом підрахунку числа об'єктів на одиницю площі. Інформативність джерела просторових даних у більшості випадків не піддається чисельній оцінці і безпосередньо залежить від цілей та методів рішення завдань. Сучасність джерела характеризується його відповідністю певній даті, періоду. Велику роль грає визначення старіння просторової інформації.

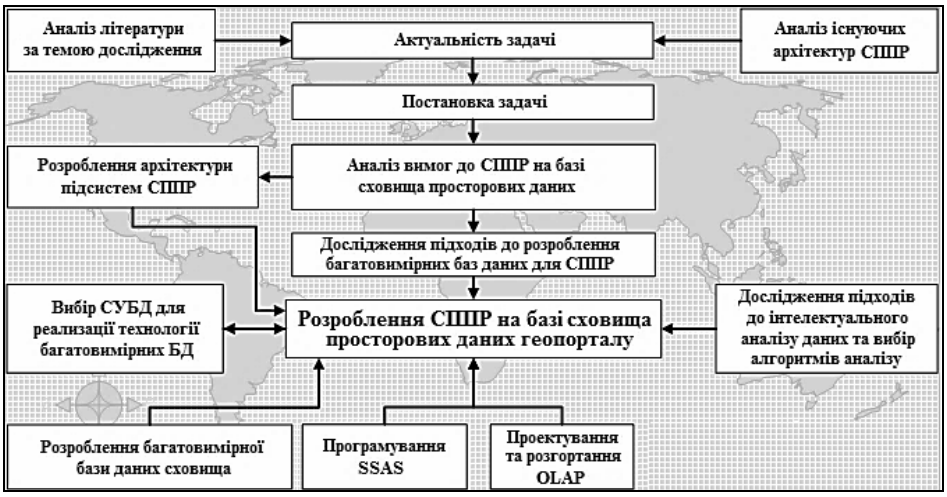

Рис. 5. Структурна схема досліджень

4. *Оцінка географічної точності* характеризує величини погрішностей, що виникають при змінюванні просторових даних (довжин, площ, кутів та інших картометричних характеристик) при переміщуванні зони спостереження по картам, планам тощо.

5. *Оцінка якості оформлення* та видання розпочинається із з'ясування розрізнюваності знаків, наочності та зрозумілості позначень тощо. З джерел даних інформація переміщується на основі певного регламенту в централізоване сховище, де здійснюється її первинна обробка. Часто виникає ситуація, коли локальні бази даних (джерела) містять однотипну за сенсом інформацію. Перед завантаженням у сховище уся ця інформація має бути погоджена, щоб забезпечити цілісність та несуперечність аналітичних даних.

6. *Тип географічної проекції.* При створенні СПД головною вимогою до даних дистанційного зондування Землі (ДДЗ) являється наявність єдиної проекції. Тому на етапі первинної обробки здійснюється переведення матеріалів ДЗЗ у цифрову форму та прив'язка їх до конкретної системи координат. У результаті формуються бази відповідних векторних матеріалів, а також бази атрибутивних даних.

Таким чином, предметом дослідження є розроблення геоінформаційної системи підтримки прийняття рішень на базі сховища просторових даних геопорталу. Об'єктом дослідження є процес об'єднання та інтеграції у сховище геопорталу інформації, що містить різні типи просторових даних. **Мета роботи** – підвищення оперативності статистичного і прогнозного аналізу геоданих для ведення звітності та підтримки прийняття рішень геоінформаційних задач в режимі реального часу. Структурну схему виконаних досліджень наведено на рис. 5.

Багатовимірний аналіз і зберігання даних в системах підтримки прийняття рішень

Науково-методичні основи багатовимірного аналізу даних. Метою інтелектуального аналізу даних (*англ.* Data Mining) є виявлення неявних закономірностей в наборах даних. Як науковий напрям він став активно розвиватися в 90-х роках XX століття, що було викликано широким поширенням технологій автоматизованої обробки інформації та накопиченням в комп'ютерних системах великих об'ємів даних. Існуючі технології дозволяли швидко знайти у базі даних потрібну інформацію, але цього у багатьох випадках було вже недостатньо. Виникла потреба пошуку взаємозв'язків між окремими подіями серед великих об'ємів даних, для чого знадобилися методи математичної статистики, теорії баз даних, теорії штучного інтелекту і знання з ряду інших областей.

Класичним вважається визначення, дане одним із засновників напряму, Григорієм П'ятецьким-Шапіро: *Data Mining* — дослідження й виявлення «машиною» (алгоритмами, засобами штучного інтелекту) в сирих даних нетривіальних, практично корисних та доступних для інтерпретації прихованих знань, які раніше не були відомі.

Аналіз літератури з цієї тематики та урахування різноманітності форм представлення даних, використовуваних алгоритмів і сфер їх застосування дають підставу для висновку, що інтелектуальний аналіз даних може проводитися, зокрема, за допомогою таких програмних продуктів:

– спеціалізованих програмних продуктів для інтелектуального аналізу;
– математичних пакетів;
– електронних таблиць (і різного роду надбудов над ними);
– засобів, інтегрованих в системи управління базами даних.

У рамках цієї роботи в першу чергу розглядатимуться засоби, інтегровані в СУБД.

В ході проведення інтелектуального аналізу даних проводиться дослідження безлічі об'єктів (чи варіантів). У більшості випадків його можна представити у вигляді таблиці, кожен рядок якої відповідає одному з варіантів, а в стовпцях містяться значення параметрів, що його характеризують.

Залежна змінна — параметр, значення якого розглядаємо як залежне від інших параметрів (незалежних змінних). Власне, цю залежність й необхідно визначити, використовуючи методи інтелектуального аналізу даних. Основні завдання інтелектуального аналізу даних полягають в такому.

1. *Завдання класифікації* полягає в тому, що для кожного варіанту визначається категорія або клас, до якого він належить. Для вирішення цього завдання необхідно, щоб безліч класів була відома заздалегідь та була б кінцевою та рахунковою.

2. *Завдання регресії* багато в чому схоже на завдання класифікації, але в ході рішення задачі регресії здійснюється пошук шаблонів для визначення числових значень параметрів, що цікавлять, при цьому ці параметри, як правило, числа з безперервних діапазонів.

3. *Завдання прогнозування* нових значень на підставі наявних значень числової послідовності (чи декількох послідовностей, між значеннями яких спостерігається кореляція). При цьому можуть враховуватися наявні тенденції (тренди), сезонність та інші чинники.

4. *Завдання кластеризації* полягає в діленні безлічі об'єктів на групи (кластери), схожі за параметрами. При цьому, на відміну від класифікації, число кластерів та їх характеристик може бути заздалегідь невідоме і визначатися в процесі побудови кластерів, виходячи зі ступеня близькості об'єднуваних об'єктів за сукупністю параметрів. Інша назва цього завдання — *сегментація*.

5. *Завдання визначення взаємозв'язків*, яке також називається *завданням пошуку асоціативних правил*, полягає у визначенні наборів об'єктів, що часто зустрічаються серед безлічі подібних наборів.

6. *Аналіз послідовностей* або *секвенційний аналіз* одними авторами розглядається як варіант попереднього завдання, іншими — виділяється окремо. Метою, в даному випадку, являється виявлення закономірностей в послідовностях подій. Подібна інформація дозволяє, наприклад, попередити збій в роботі інформаційної системи за рахунок отримання сигналу про настання події, часто передуючої збою.

7. *Аналіз відхилень* дозволяє відшукати серед безлічі подій ті, які істотно відрізняються від норми. Відхилення може сигналізувати про якусь незвичайну подію або, наприклад, про помилку введення даних оператором.

OLAP (*англ.* Online Analytical Processing) — технологія комплексного багатовимірного аналізу даних. OLAP — це ключовий компонент організації сховищ даних та СППР. Концепція OLAP була описана в 1993 році Едгаром Коддом, відомим дослідником баз даних і автором реляційної моделі бази даних. У 1995 році на основі принципів, викладених Коддом, були сформульовані такі вимоги до застосунків для багатовимірного аналізу:

– надання користувачеві результатів аналізу за прийнятний час (зазвичай, не більше 5 с), нехай навіть ціною менш детального аналізу;

– можливість здійснення будь-якого логічного і статистичного аналізу, характерного для цього застосування, та його збереження в доступному для кінцевого користувача вигляді;

– розрахований на велику кількість користувачів доступ до даних з підтримкою відповідних механізмів блокувань і засобів авторизованого доступу;

– багатовимірне концептуальне представлення даних, включаючи повну підтримку для ієрархій і множинних ієрархій (це — ключова вимога OLAP);

– можливість звертатися до будь-якої потрібної інформації незалежно від її об'єму й місця зберігання.

Основне призначення OLAP-систем — підтримка аналітичної діяльності та довільних запитів користувачів-аналітиків. Мета OLAP-аналізу — перевірка виникаючих гіпотез.

Категорії даних та інформаційні потоки у сховищі просторових даних. Усі дані в СПД поділяються на категорії (рис. 6).

Детальні дані — це дані, що переносяться безпосередньо з OLTP-підсистем. Вони відповідають елементарним подіям, які фіксуються в OLTP-системах та підрозділяються на:

– *вимірювання* – набори даних, необхідні для опису подій;

– *факти* – дані, що відбивають суть події;

– *агреговані (узагальнені) дані* — дані, що отримуються на підставі детальних шляхом підсумовування по певних вимірюваннях;

– *метадані* – дані про дані, що містяться у СПД і можуть описувати об'єкти предметної області, інформація щодо яких міститься у СПД, категорії користувачів, що використовують дані з СПД, місця та способи зберігання даних, дії, виконувані над даними, час та причини виконання різних дій над даними;

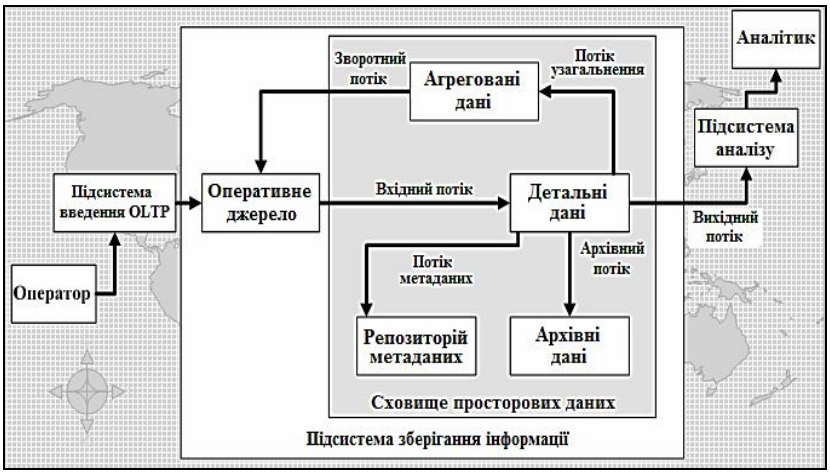

Рис. 6. Категорії даних та інформаційні потоки у СПД

Дані у СПД утворюють такі інформаційні потоки (рис. 6):

– *вхідний потік* – утворюється даними, що копіюються з OLTP-систем в СПД, при цьому дані очищуються і збагачуються шляхом додавання нових атрибутів;

– *потік узагальнення* – утворюється агрегацією детальних даних та їх збереженням у СПД;

– *архівний потік* — утворюється переміщенням детальних даних, кількість звернень до яких знизилася;

– *потік метаданих* – утворюється потоком інформації про дані, спрямовані до репозиторію даних;

– *вихідний потік* — утворюється даними, які витягуються користувачами;

– *зворотний потік* — утворюється очищеними даними, записуваними назад в OLTP-системи.

Структура та принципи функціонування системи багатовимірного аналізу даних в реальному часі. В процесі аналізу даних часто виникає необхідність побудови залежностей між різними їх параметрами, число яких може бути значним. СППР мають засоби надання користувачеві агрегованих даних для різних вибірок з початкового набору в зручному для сприйняття та аналізу вигляді. В результаті, функції агрегованих просторових даних утворюють багатовимірний нереляційний набір, який називають OLAP-кубом. Осі цього кубу містять параметри, а комірки — залежні від них агреговані дані (рис. 7). Уздовж кожної осі дані можуть бути організовані у вигляді ієрархії, яка представляє різні рівні їх деталізації.

Завдяки такій моделі даних користувачі можуть формулювати складні запити, генерувати звіти, отримувати підмножини даних. OLAP-куб може бути реалізований на основі універсальних реляційних СУБД або спеціалізованим програмним забезпеченням.

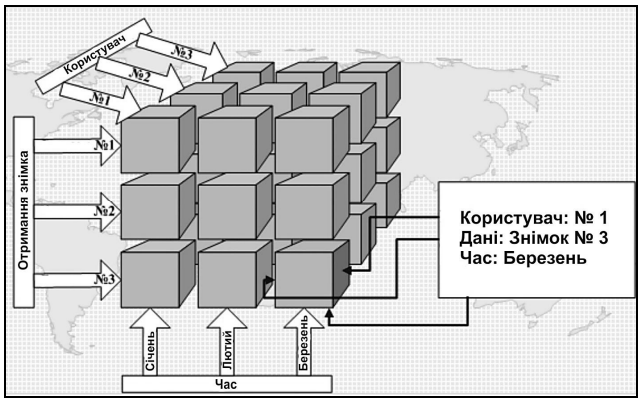

Рис. 7. Простий (тривимірний) OLAP-куб

Осі куба є вимірами, по яких відкладають параметри, що відносяться до аналізованої предметної області. На перетині осей вимірів розташовуються дані, що кількісно характеризують аналізовані факти-заходи. Над OLAP-кубом можуть виконуватися такі операції.

1. *Зріз* — формується підмножина багатовимірного масиву даних, що відповідає єдиному значенню одного або декількох елементів вимірів, які входять в цю підмножину.

2. *Обертання* — зміна розташування вимірів, представлених в звіті або на сторінці, що відображається. Наприклад, операція обертання може полягати в перестановці місцями рядків і стовпців таблиці. Крім того, обертанням куба даних є переміщення позатабличних вимірів на місце вимірів, представлених на сторінці, що відображається, й навпаки.

3. *Консолідація й деталізація* — операції, які визначають перехід за напрямом від детального представлення даних до агрегованого й навпаки, відповідно. Напрям деталізації може бути заданий як за ієрархією окремих вимірів, так і згідно з іншими стосунками, встановленими у рамках вимірів або між вимірами.

Таблиці фактів у сховищі просторових даних. Таблиця фактів є основною таблицею сховища даних. Вона містить відомості про об'єкти або події, сукупність яких надалі аналізуватиметься. Виділяють чотири типи фактів:

– факти, пов'язані з транзакціями (Transaction facts) і засновані на окремих подіях;

– факти, пов'язані з "моментальними знімками" (Snapshot facts) і засновані на стані об'єкту в певні моменти часу, наприклад, на кінець дня або місяця;

– факти, пов'язані з елементами документу (Line-item facts) і засновані на тому або іншому документі (містять детальну інформацію про елементи цього документу);

– факти, пов'язані з подіями або станом об'єкту (Event or state facts) — представляють виникнення події без подробиць про об'єкт.

Таблиця фактів містить унікальний складений ключ, що об'єднує первинні ключі таблиць вимірів. Найчастіше в цих цілях використовуються цілочисельні значення, або значення типу «дата/час», оскільки таблиця фактів може містити сотні тисяч або навіть мільйони записів, а отже зберігати в ній текстові описи, що повторюються, невигідно. При цьому як ключові, так і деякі неключові поля мають відповідати майбутнім вимірам OLAP-куба. Окрім цього, таблиця фактів містить одне або декілька числових полів, на підставі яких надалі будуть отримані агреговані дані.

Для багатовимірного аналізу придатні таблиці фактів, що містять якомога детальніші дані (тобто ті, що відповідають членам нижніх рівнів ієрархії відповідних вимірів).

У таблиці фактів немає ніяких відомостей про те, як групувати записи при обчисленні агрегованих даних — ці відомості містяться в таблицях вимірів і використовуються для побудови ієрархій у вимірах куба.

Таблиці вимірів у сховищі просторових даних. Таблиці вимірів містять незмінні або рідко змінювані дані. У переважній більшості випадків ці дані містять по одному запису для кожного члена нижнього рівня ієрархії у вимірі. Таблиці вимірів також містять як мінімум одне описове поле (зазвичай з ім'ям члена виміру) та цілочисельне ключове поле (зазвичай, це сурогатний ключ) для однозначної ідентифікації члена виміру. Якщо майбутній вимір, заснований на цій таблиці вимірів, містить ієрархію, то таблиця вимірів також може містити поля, що вказують на «батька» цього члена в цій ієрархії. Таблиця вимірів може мати додаткові атрибути членів вимірів, що містилися в початковій оперативній базі даних (наприклад, адреси й телефони клієнтів). Кожна таблиця вимірів повинна знаходитися у відношенні «один до багатьох» з таблицею фактів. Швидкість зростання таблиць вимірів має бути незначною у порівнянні зі швидкістю зростання таблиці фактів.

Архітектура СППР на базі сховища просторових даних геопорталу

Повномасштабна СППР має виконувати складні й різноманітні функції, що включають збір даних з різних джерел, їх узгодження, перетворення й завантаження у сховища, зберігання аналітичної інформації, регламентну звітність, підтримку довільних запитів, багатовимірний аналіз.

Нині існують фактичні стандарти побудови СППР, заснованих на концепції СД. Ці стандарти спираються на сучасні дослідження і загальносвітову практику створення СПД і аналітичних систем.

На підставі проведеного аналізу архітектури сучасних СППР для створення системи підтримки прийняття рішень на базі сховища просторових даних геопорталу пропонується використовувати структуру, що у загальному вигляді описується схемою з трьома виділеними шарами (рис. 8).

Дані поступають з різних OLTP-систем, із зовнішніх джерел та операційних систем відповідно до встановленого регламенту і форм звітності. Уся ця інформація перевіряється, узгоджується, перетворюється та поміщається у сховище та вітрини даних. Після цього користувачі за допомогою спеціалізованих інструментальних засобів отримують необхідну їм інформацію для побудови різних табличних і графічних представлень, прогнозування, моделювання та виконання інших аналітичних завдань.

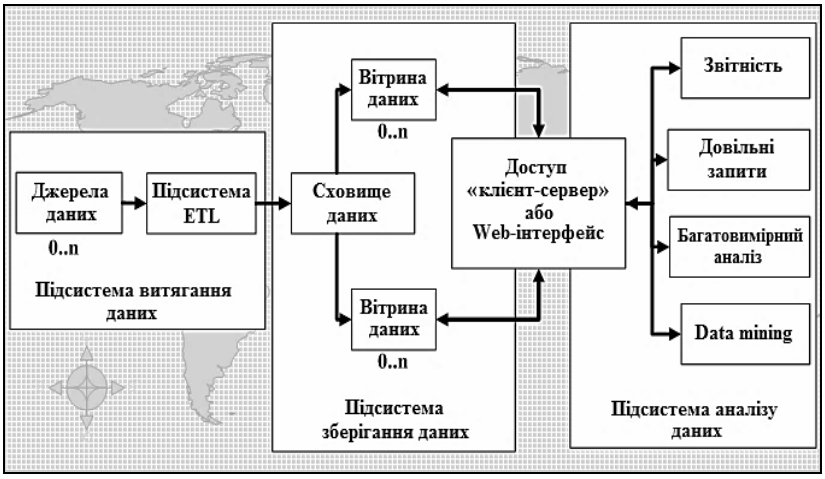

Рис. 8. Архітектура СППР на основі технології сховищ просторових даних

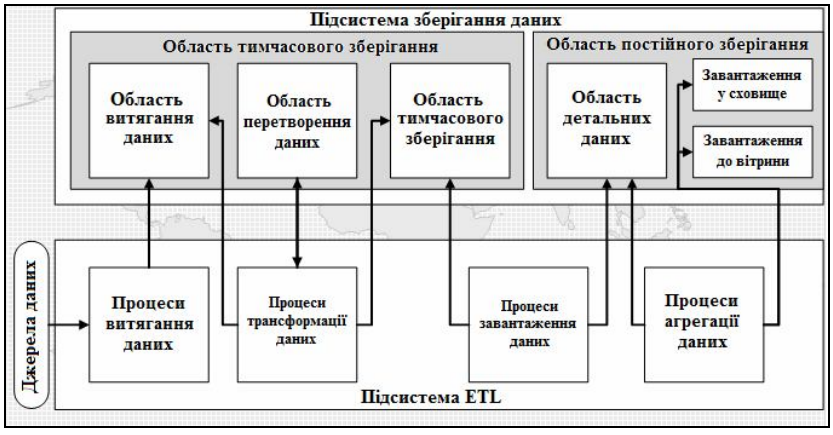

Рис. 9. Функціональна схема підсистеми ETL

Підсистема витягання, перетворення і завантаження даних. З джерел даних інформація переміщується у сховище даних. Необхідні для сховища дані не зберігаються в остаточному виді в жодній з OLTP-систем, ці дані зазвичай можна отримати з цих систем шляхом спеціальних перетворень, обчислень і агрегації. Завантаження даних з джерел у сховища має надавати такі можливості:

– витягати дані з різних баз даних і текстових файлів;

– виконувати різні типи узгодження та очищення даних;

– перетворювати дані при переміщенні їх від джерел до сховища;

– завантажувати погоджені та "очищені" дані в структури сховища.

Підсистема ETL (Extract, Transform, Load) в СПД повинна забезпечувати виконання цих функцій. Виходячи з цього, можна сформувати основні вимоги до підсистеми ETL :

– мати спроможність збирати дані від різних систем (джерел), навіть якщо одна або декілька систем в результаті збою не змогли в строк завершити свою роботу і надати необхідні дані;

– отримана інформація має бути розпізнана та перетворена відповідно до алгоритмів перетворення, а також за допомогою систем ведення звітності та метаданих;

– перетворена інформація має бути завантажена в зони тимчасового зберігання у сховищі даних та у вітрини даних, як того вимагає виробничий процес;

– засоби ETL мають забезпечувати високу пропускну спроможність для того, щоб збирати та вивантажувати постійно зростаючі об'єми даних;

– засоби ETL повинні мати потужні обчислювальні можливості та масштабованість для скорочення часу обробки даних з метою зменшення затримок в наданні даних для аналітичних робіт;

– засоби ETL мають надавати різноманітні інструменти витягання даних в різних режимах роботи — від пакетного збору для систем, некритичних до тимчасових затримок, до інкрементної обробки в режимі, близькому до реального часу.

Функціональну схему підсистеми ETL представлено на рис. 9.

Процеси витягання даних забезпечують отримання даних із джерел та зберігають ці дані в інтерфейсні таблиці області витягання даних.

Процеси перетворення (трансформації) даних витягають

дані з інтерфейсних таблиць, проводять захоплення змін, перетворення даних за певними правилами зі збереженням проміжних результатів в області перетворення даних та зберігають результат в області оперативного зберігання. Процеси перетворення даних виконують завдання наповнення таблиць області перетворення даних і області завантаження даних.

Після проведення перетворення даних вони завантажуються в область оперативного зберігання. Процеси завантаження даних проводять читання даних з області оперативного (тимчасового) зберігання, перевіряють посилальну цілісність даних та проводять їх завантаження в область детальних даних. Процеси агрегації даних проводять читання детальних даних і запис їх у СПД та вітрини даних.

Підсистема зберігання даних. Підсистема зберігання даних призначена для безпосереднього зберігання значимої, перевіреної, погодженої, несуперечливої та хронологічно цілісної інформації, яку з досить високою мірою упевненості можна вважати достовірною. СД не орієнтоване на рішення якої-небудь певної функціональної аналітичної задачі. Метою СД є забезпечення цілісності та підтримка хронології даних. У зв'язку з цим у більшості випадків для виконання певного комплексу функціонально замкнутих аналітичних завдань раціонально створювати вітрини даних. По суті вітрина є відносно невеликою, але що найважливіше, являє собою функціонально-орієнтоване СД, в якому інформація зберігається спеціальним чином, оптимізованим з точки зору вирішення конкретних аналітичних завдань певного підрозділу або групи аналітиків.

У просторових сховищах (рис. 10) первинні дані перетворюються в інформацію, придатну для використання на етапі підготовки даних. Просторова модель сховища даних містить ту ж атомарну інформацію, що й нормалізована модель, але інформація структурована по-іншому, щоб полегшити її використання й виконання запитів. Запити в процесі виконання звертаються до усе більш низького рівня деталізації без додаткового перепрограмування з боку користувачів або розробників застосунків.

Підсистема аналізу даних. Для організації доступу аналітиків до даних СПД та ВД використовуються спеціалізовані робочі місця, що підтримують необхідні технології як оперативного, так і довготривалого аналізу. Результати роботи аналітиків оформляються у вигляді звітів, графіків, рекомендацій та зберігаються як на локальному комп'ютері, так і в загальнодоступному вузлі локальної мережі.

Аналітична діяльність може бути досить різноманітною і визначатись характером вирішуваних завдань, особливостями компанії, рівнем і ступенем підготовленості аналітиків. У зв'язку з цим сучасний підхід до інструментальних засобів аналізу не обмежується використанням якоїсь однієї технології. В межах даної роботи у якості аналітичної підсистеми пропонується до використання СУБД Microsoft SQL Server, у тому числі й на підставі того, що SQL Server підтримує типи даних *geometry* та *geography*. У цих типах підтримуються методи і властивості, що дозволяють створювати, порівнювати та аналізувати просторові дані. Тип даних *geometry* використовується для планарних Евклідових даних. Він відповідає специфікації консорціуму OGC «Simple Features for QL». Тип даних *geography* використовується для зберігання еліптичних даних, відповідних координатам широти і довготи, у тому числі й даним GPS та GNSS.

В межах даної роботи не ставиться за мету виконати прив'язку до певної версії Microsoft SQL Server, адже даний продукт постійно удосконалюється, розширюючи можливості користувачів. Доцільно відзначити, наприклад, що актуальний і доступний (а у версії Express взагалі безкоштовний) на сьогодні пакет Microsoft SQL Server 2019 забезпечує інтелектуальний аналіз структурованих та неструктурованих даних, їх запити за допомогою платформи даних з найкращими в галузі показниками продуктивності та безпеки. При цьому забезпечуються переваги масштабованості та доступності для критично важливих інтелектуальних застосунків та сховищ даних. Безпека досягається наявними засобами шифрування даних під час зберігання та передачі. Взагалі, СУБД SQL Server вже більш семи років визнається найменш вразливою за результатами тестування Національного інституту стандартів і технологій США (NIST).

Засоби бізнес-аналітики SQL Server надають можливість обробляти будь-які дані та спостерігати за розвитком ситуації в режимі реального часу, отримуючи до мільонну прогнозів за секунду (як з підключенням до Internet, так і без нього).

Разом з тим, слід зауважити, що запропоновані в даній роботі програмні розробки можуть бути застосовані й з використанням більш ранніх версій SQL Server (принаймні, щонайменш навіть із релізом Microsoft SQL Server 2008).

Інтелектуальний аналіз даних за допомогою СУБД Microsoft SQL Server. Розглянемо детальніше реалізацію засобів інтелектуального аналізу даних в СУБД Microsoft SQL Server. На рис. 11 схематично

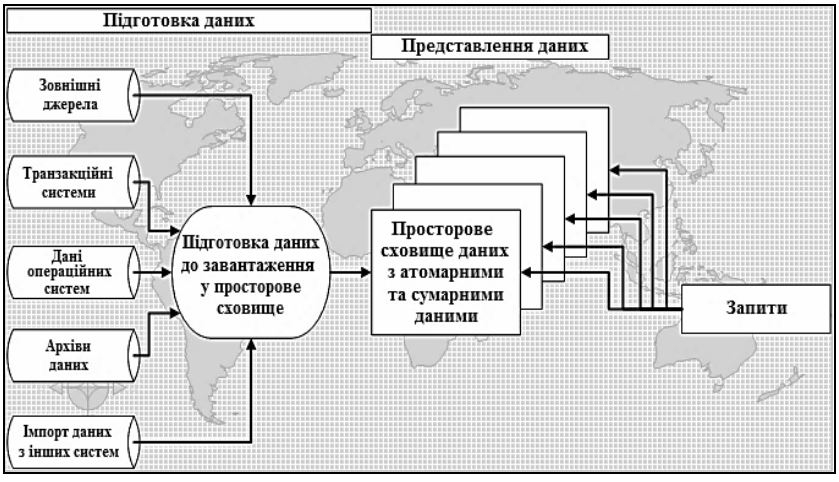

Рис. 10. Архітектура просторового сховища даних

представлено компоненти СУБД SQL Server та виділено підсистему інтелектуального аналізу даних та управління багатовимірними даними.

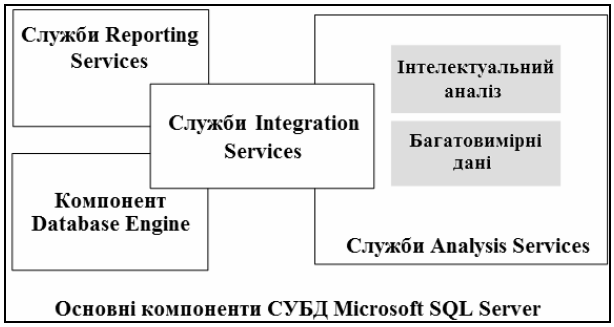

Рис. 11. Служби і компоненти СУБД Microsoft SQL Server

Служби Analysis Services надають такі функції та засоби для створення рішень по інтелектуальному аналізу даних:

– набір стандартних алгоритмів інтелектуального аналізу даних;

– конструктор інтелектуального аналізу даних, призначений для створення та перегляду моделей інтелектуального аналізу даних, управління ними та побудови прогнозів;

– мова управління інтелектуальним аналізом даних.

Структура інтелектуального аналізу даних може бути представлена як сукупність початкових даних і опису способів їх обробки. Структура містить моделі, які використовуються для аналізу даних. Одна структура може підтримувати декілька моделей. У структурі інтелектуального аналізу даних можна виділити повчальний та перевірочний набір даних, задавши процентне відношення або об'єм даних.

Модель інтелектуального аналізу даних є поєднанням самих даних, алгоритму інтелектуального аналізу даних і колекції значень параметрів та фільтрів, що управляють використанням і обробкою даних. Модель інтелектуального аналізу даних визначається на мові розширень інтелектуального аналізу даних або за допомогою майстра інтелектуального аналізу даних в середовищі Business Intelligence Development Studio (BIDS).

Алгоритм інтелектуального аналізу даних є механізмом, що створює модель інтелектуального аналізу даних. Щоб створити модель, алгоритм спочатку аналізує набір даних, здійснюючи пошук певних закономірностей і трендів. Алгоритм використовує результати цього аналізу для визначення параметрів моделі інтелектуального аналізу даних. Потім ці параметри застосовуються до усього набору даних, щоб виявити придатні до використання закономірності та отримати детальну статистику.

Нижче перераховано алгоритми інтелектуального аналізу даних, які можуть використовуватися із використанням Microsoft SQL Server:

– спрощений алгоритм Байеса – Naive Bayes;

– алгоритм дерева прийняття рішень – Decision Trees;

– алгоритм тимчасових рядів – Time Series;

– алгоритм кластеризації – Clustering;

– алгоритм кластеризації послідовностей – Microsoft Sequence Clustering;

– алгоритм взаємозв'язків – Association Rules;

– алгоритм нейронної мережі – Neural Network;

– алгоритм лінійної регресії – Linear Regression;

– алгоритм логістичної регресії – Logistic Regression.

Обробляти запити інтелектуального аналізу можна за допомогою стандартних алгоритмів, перелічених вище, а також й алгоритмів сторонніх розробників. На рис. 12 схематично представлено етапи інтелектуального аналізу даних і засоби Microsoft SQL Server, за допомогою яких вони виконуються. Вказані етапи не обов'язково будуть пройдені один за іншим. На одному з проміжних етапів може з'ясуватися, що в поточній постановці для вирішення завдання бракує даних і знадобиться знову повернутися до першого етапу.

На етапі *постановки задачі* необхідно визначити, що є метою аналізу. Вимагається вирішити низку завдань:

– з'ясувати, що саме необхідно визначити в результаті аналізу;

– зробити прогнози на підставі моделі інтелектуального аналізу даних або просто знайти змістовні закономірності та взаємозв'язки між використовуваними даними;

– з'ясувати у разі потреби у прогнозі, який атрибут набору даних необхідно спрогнозувати;

– знайти взаємозв'язки між таблицями, якщо існує декілька таблиць;

– з'ясувати, яким чином розподіляються дані і чи дають вони точне уявлення про предметну область.

Зазвичай, в процесі постановки задачі аналітик працює спільно з фахівцями в предметній області.

Рис. 12. Етапи інтелектуального аналізу даних

Етап *підготовки даних* включає визначення джерел даних для аналізу, об'єднання даних та їх очищення. Використовувані дані можуть знаходитися в різних базах і на різних серверах, можуть бути представлені у вигляді текстових файлів, електронних таблиць, існувати в різноманітних форматах.

Для об'єднання та перетворення даних використовуються можливості сховища даних. Це дозволяє істотно автоматизувати процес підготовки даних. Зібрані таким чином дані, зазвичай, потребують додаткової обробки, яка й називається очищенням. В процесі очищення при необхідності може проводитися видалення "викидів" (нехарактерних та помилкових значень), обробка відсутніх значень параметрів, чисельне перетворення (наприклад, нормалізація) тощо.

Наступним етапом є *вивчення даних*, яке дозволить зрозуміти, наскільки адекватно підготовлений набір представляє досліджувану предметну область. Тут може проводитися пошук мінімальних і максимальних значень параметрів, аналіз розподілів значень та інших статистичних характеристик, порівняння отриманих результатів з уявленнями про предметну область. Етап аналізу щодо *побудови моделей* спочатку полягає у створенні структури даних, а потім у створенні для цих структур однієї або декількох моделей. Такі моделі включають вказівки на алгоритм інтелектуального аналізу даних, його параметри, а також аналізовані дані. При визначенні моделі можуть використовуватися будь-які фільтри. Таким чином, не усі наявні в описі структури даних використовуватимуться кожною створеною для них моделлю. Модель може проходити повчання, що полягає в застосуванні вибраного алгоритму до повчального набору даних. Після цього в ній зберігаються виявлені закономірності. Часто для вирішення задачі створюється декілька моделей, заснованих на різних алгоритмах, щоб була можливість порівняти результати та вибрати найкращу.

Етап *перевірки моделі* має за мету оцінку якості роботи створеної моделі перед початком її використання. Якщо створювалося декілька моделей, то на цьому етапі робиться вибір на користь тієї, що дасть найкращий результат. При вирішенні завдань інтелектуального аналізу, що передбачають певну якість видаваного моделлю прогнозу, цю якість можна оцінити на перевірочному наборі даних, для яких відоме значення прогнозованого параметра. Точність прогнозів, що створюються моделями, можна перевірити за допомогою таких засобів, як діаграма точності прогнозів і матриця класифікації.

Таким чином, на підставі проведеного аналізу архітектур сучасних систем підтримки прийняття рішень запропоновано структуру, яку найдоцільніше використовувати для побудови СППР на базі сховища просторових даних геопорталу. Проаналізовано основні функції, що мають виконувати підсистеми такої СППР, а саме – функції підсистеми витягання, перетворення і завантаження даних, функції підсистеми зберігання даних та підсистеми аналізу даних. Здійснено ґрунтовний вибір системи управління базами даних для практичної реалізації багатовимірної бази геоданих, яка виконує функції сховища просторових даних як основи СППР геопорталу.

Отже, наступним кроком є саме програмна реалізація багатовимірної бази (сховища) просторових даних із застосуванням СУБД Microsoft SQL Server.

Технологія розроблення багатовимірного сховища просторових даних геопорталу

Багатовимірний аналіз даних засобами Microsoft SQL Server. Microsoft SQL Server Analysis Services (SSAS) є базовою платформою для розгортання аналітичних систем (Business Intelligence, BI). Основні компоненти BI-рішення Microsoft наведено в табл. 1.

Таблиця 1 – **Компоненти BI-рішення Microsoft**

Компонент	Опис компонента
SQL Server Database Engine	Масштабована, високопродуктивна СУБД, здатна зберігати великі об'єми даних, консолідації даних, що утворюються в результаті, в єдине сховище для аналізу і побудови звітів.
SQL Server Integration Services	Платформа для виконання операцій витягання, перетворення і завантаження даних, яке забезпечує заповнення СД та його синхронізацію з даними з різних джерел, що використовуються застосунками.
SQL Server Analysis Services	Забезпечує можливість побудови OLAP-рішень, включаючи можливість розрахунку ключових індикаторів продуктивності (KPI). Застосовується також для побудови data-mining-рішень, які використовують спеціалізовані алгоритми для виявлення і аналізу залежностей.
SQL Server Reporting Services	Інструментарій побудови звітів, призначений для створення, публікації і поширення деталізованих звітів, як для внутрішніх, так і для зовнішніх цілей.

Уніфікована багатовимірна модель даних служб SQL Server Analysis Services. Системи SSAS побудовані на основі уніфікованої багатовимірної моделі (Unified Dimensional Model, UDM), яка дозволяє різним типам клієнтських застосунків діставати доступ до даних, як з реляційних, так і з багатовимірних БД без використання окремих моделей для кожного типу баз даних. Основою UDM є архітектура вимірів на основі атрибутів, що надає можливість використати безліч джерел даних для створення багатовимірної моделі. Модель UDM може бути використана для створення єдиного представлення реляційних і багатовимірних даних, що включають об'єкти, аналітику, обчислення і метрики. Ця модель створює проміжний логічний рівень між фізичною реляційною базою даних, що використовується як джерело даних, і структурами куба, використовуваними для обробки призначених для користувача запитів. Модель UDM можна уявити собі як ядро системи OLAP, однією з ключових переваг якої є можливість поєднувати гнучкість і функціональне багатство традиційної реляційної моделі генерації звітів з потужними аналітичними засобами і високою продуктивністю класичної моделі системи OLAP.

За рахунок високої масштабованості служби SSAS дозволяють працювати з терабайтними базами даних і тисячами користувачів. Щоб забезпечити роботу великої кількості користувачів, уникнути конф-

ліктів при користуванні ресурсами і понизити витрати, є можливість горизонтального масштабування служб SSAS. Горизонтальне масштабування полягає в нарощуванні обчислювальних потужностей і місткості сховищ даних з метою зберігання і синхронізації декількох версій даних, але в той же час служби SSAS дозволяють організувати загальний доступ для читання інформації з однієї бази цих служб з декількох серверів, усуваючи необхідність в надмірних ресурсах. У той же час аналітичні куби служб SSAS забезпечують високошвидкісний доступ до великих об'ємів заздалегідь об'єднаних даних, а також дозволяють кінцевим користувачам отримувати дані, що цікавлять, в реальному часі.

Планування і програмування архітектури SSAS. Служби Microsoft SQL Server Analysis Services використовують як серверні, так і клієнтські компоненти для надання застосункам функцій оперативної аналітичної обробки (OLAP) та інтелектуального аналізу даних. Клієнти обмінюються даними із службами SSAS за допомогою загальнодоступного стандарту XML (для аналітики — XMLA), який є протоколом для виконання команд й отримання відповідей і надається у вигляді web-служби. Тому кожен екземпляр SSAS являється web-сервісом. Служби SSAS підтримують ядро локального куба, яке дозволяє застосункам клієнтів переглядати багатовимірні дані, що локально зберігаються. Екземпляр служб SSAS може містити декілька баз даних, а у базі даних можуть одночасно бути присутніми об'єкти OLAP і об'єкти інтелектуального аналізу даних. Застосування підключаються до вказаного екземпляра служб SSAS і до вказаної бази даних. На серверному комп'ютері може експлуатуватися декілька екземплярів служб SSAS. Кожен екземпляр служб SSAS розглядається як окремий об'єкт сервера і підключається до нього за допомогою окремого з'єднання. Кожен об'єкт сервера містить одне або декілька джерел даних, представлення джерела даних і об'єкти бази даних, а також складки і ролі безпеки. Кожен об'єкт бази даних містить декілька об'єктів-вимірів. Кожен об'єкт-вимір містить один або декілька атрибутів, що організовані у вигляді ієрархій.

З точки зору фізичної архітектури служби SSAS складаються із серверного та клієнтського компонентів. Серверний компонент служб SSAS реалізований у вигляді служби Microsoft Windows. Служби SSAS підтримують роботу декількох екземплярів на одному комп'ютері, при цьому кожен екземпляр служб SSAS реалізований як окремий екземпляр служби Windows. Клієнтська частина служб SSAS запускається як ізольована служба, і взаємодія з цією службою відбувається через XMLA з використанням протоколу HTTP або TCP.

Метою програміста SSAS є створення прикладної моделі даних. Прикладна модель визначає формат даних, в якому вони передаються аналітичним застосуванням. Основним користувачем прикладної моделі даних є клієнтський застосунок, що, власне, й представляє модель користувачеві. Прикладна модель створюється за допомогою мови багатовимірних виразів (Multidimensional Expressions, MDX), яка служить як для представлення запитів до багатовимірної бази даних, так і для опису моделі формування даних усередині неї за допомогою MDX-сценаріїв.

Існує стандартна бібліотека класів управління для служб SSAS, доступна для використання програмним способом через простір імен Microsoft.Analysis Services. Ці класи включені у файл AnalysisServices.dll і для роботи з ними слід вказати у проекті Visual Studio посилання на цю бібліотеку. Ці класи дозволяють створювати, змінювати і видаляти куби, виміри, структури інтелектуального аналізу, а також бази даних служб SSAS. Крім того, ці класи особливо корисні для автоматизації часто виконуваних завдань (наприклад, для створення нових секцій в групі заходів при появі нових даних в таблиці фактів або для повторного повчання моделі інтелектуального аналізу при появі нових даних). Розробникам зазначені класи дають можливість створювати інтерфейси адміністрування для певних наборів користувачів. Ці інтерфейси мають обмежувати доступ до об'єктів служб SSAS і дозволяти користувачам виконувати тільки певні завдання.

Розроблення багатовимірних баз даних. Середовище SQL BI Development Studio (BIDS) і середовище SQL Server Management Studio (SSMS) надають можливості для розробки рішень на основі багатовимірних баз даних. При створенні проекту автоматично створюється нове рішення, а в існуюче рішення при необхідності можна додавати інші проекти. Об'єкти, які містяться в проекті, залежать від його типу. Елементи в кожному контейнері проекту зберігаються у вигляді файлів, розташованих в теках проекту у файловій системі. Середовище SSMS розроблене для адміністрування і налаштування існуючих об'єктів в компонентах SQL Server Database Engine, Analysis Services, Integration Services і Reporting Services. Середовище BIDS призначене для розробки рішень в області аналітики, які включають функції служб Analysis Services, Integration Services і Reporting Services. Середовище SSMS надає інтегроване середовище для з'єднання з екземплярами служб SSAS, SQL Server і Reporting Services, щоб настроювати об'єкти, а також проводити адміністрування об'єктів і управляти ними в межах екземпляра служб SSAS. З використанням цих сценаріїв можна також використати середовище SSMS для створення або зміни об'єктів служб SSAS, але середовище SSMS не надає графічний інтерфейс для конструювання й визначення об'єктів.

Середовище BIDS надає інтегроване середовище для розробки аналітичних рішень. Середовище BIDS можна застосувати в проектному режимі, що використовує визначення на основі XML-об'єктів служб SSAS, тобто Integration Services і Reporting Services, що містяться в проектах і рішеннях. Використання середовища BIDS в проектному режимі означає, що зміни об'єктів служб SSAS в середовищі BIDS застосовуються до визначень об'єктів на основі XML, але не застосовуються безпосередньо до об'єкту в екземплярі служб SSAS до тих пір, поки рішення не буде розгорнуто. Середовище BIDS можна також використати в оперативному режимі, тобто безпосередньо підключатися до екземпляра служб SSAS і працювати з об'єктами існуючої бази даних. Створення проекту служб SSAS в середовищі BIDS виконується або за допомо-

гою шаблону проекту служб SSAS, або за допомогою майстра імпорту бази даних. Якщо в середовищі BIDS не завантажено жодне рішення, при створенні нового проекту служб SSAS буде автоматично створено нове рішення. Інакше новий проект служб SSAS може бути доданий до існуючого рішення або створений в новому рішенні. Для створення проекту служб SSAS в середовищі BIDS необхідно виконати такі кроки.

1. Створення проекту служб SSAS — виконується шляхом вибору шаблону проекту служб SSAS або за допомогою майстра імпорту бази даних, а також завдання імені та розміщення для нового проекту. Якщо створити проект служб SSAS, заснований на шаблоні, то він буде відкритий в середовищі BIDS, в якому можна визначити джерела даних, представлення джерел даних, кубів, вимірів, ролей та інших об'єктів служб SSAS. Можна сконструювати ці об'єкти на основі існуючих джерел даних або сформувати спеціальну реляційну структуру бази даних, засновану на кубі та конструкторі вимірів. Якщо знову створений проект служб SSAS був заснований на існуючій базі даних, метадані для цієї бази даних будуть відкриті в середовищі BIDS, в якому можна змінити метадані існуючої бази даних. Поки зміни не будуть розгорнуті, вони не вплинуть на існуючі бази даних.

2. Надання додаткових файлів, наприклад текстових файлів, що містять примітки до проекту, в теку «Різне» проекту у вікні оглядача рішень.

3. Визначення властивостей розгортання проекту, щоб задати сервер, на якому будуть розгорнуті метадані проекту та оброблювані об'єкти.

4. Збірка та розгортання рішення в екземплярі служб SSAS для тестування. При складанні рішення перевіряються визначення і залежності об'єктів, включених в проект, а також формується сценарій розгортання.

5. Перегляд і тестування розгорнутого проекту.

6. Змінення (при необхідності) визначень об'єктів і повторення у зв’язку з цим збірки та розгортання рішення.

У мові багатовимірних виразів можна звертатися із запитом до таких багатовимірних об'єктів, як куби, й витягати багатовимірні набори комірок, що містять дані куба. Базовий запит багатовимірних виразів — це інструкція SELECT, найбільш частий запит у багатовимірних виразах. Щоб отримати ґрунтовні знання про застосування багатовимірних виразів для запиту багатовимірних даних, необхідно зрозуміти, як в інструкції багатовимірних виразів SELECT визначається результуючий набір, синтаксис інструкції SELECT і як з її допомогою створювати прості запити. У багатовимірному вираженні інструкція SELECT вказує результуючий набір, що містить підмножину багатовимірних даних, яка витягається з куба. У службах SSAS сценарії багатовимірних виразів складаються з одного або декількох багатовимірних виразів або інструкцій, що заповнюють куб обчисленнями. Сценарій багатовимірних виразів визначає процес обчислень для куба і вважається частиною самого куба. Зміна сценарію багатовимірних виразів одразу змінює процес обчислень для куба. Для створення сценаріїв багатовимірних виразів можна скористатися конструктором кубів в середовищі BIDS. Взагалі, існує два типи сценаріїв багатовимірних виразів.

1. Сценарій багатовимірних виразів за умовчанням. При створенні куба служби SSAS створюють стандартний сценарій багатовимірних виразів для цього куба. У цьому сценарії визначається етап обчислення для усього куба.

2. Призначений для користувача сценарій багатовимірних виразів. Після створення куба можна додати призначені для користувача сценарії багатовимірних виразів, що розширюють характеристики обчислень куба.

Якщо з кубом не пов'язаний жоден сценарій багатовимірних виразів, куб обчислюється за сценарієм багатовимірних виразів за умовчанням. Куб має бути пов'язаний хоча б з одним сценарієм багатовимірних виразів, оскільки тільки в сценарії визначається порядок обчислення куба.

Особливості застосування служби SQL Server Integration Services для роботи з багатовимірними базами даних. Служби SQL Server Integration Services (SSIS) є платформою для побудови високопродуктивних рішень потоку операцій — витягання, перетворення і завантаження (ETL) для сховищ даних. Типові випадки застосування пакетів служб SSIS спільно із SSAS включають:

- злиття даних з різнорідних сховищ даних;
- заповнення сховищ даних та вітрин даних;
- очищення і стандартизація даних.

Дані, зазвичай, зберігаються у безлічі різних систем сховищ даних, і витягання даних з усіх джерел та їх злиття в єдиний погоджений набір даних може бути досить складним завданням. Ця ситуація може виникнути з ряду причин, а саме.

1. Зберігання архівних даних в традиційних системах сховищ даних. Ці дані можуть бути не занадто важливими для щоденних операцій, проте вони можуть мати цінність для аналізу, якому потрібні дані за тривалий період.

2. Використання різних технологій зберігання даних для зберігання операційних даних. Пакету може бути потрібно витягати дані з електронних таблиць, а також з реляційних БД перед тим, як він зможе об'єднати ці дані.

3. Дані можуть зберігатися у базах даних, що використовують різні схеми для одних і тих же даних. Пакету може бути потрібно змінити тип даних стовпця або об'єднати дані з декількох стовпців в один перед тим, як він зможе об'єднати дані.

Служби SSIS можуть підключитися до багатьох типів джерел даних, включаючи декілька джерел даних одного пакету. Служби SSIS містять компоненти джерела, що здійснюють роботу по витяганню даних з пласких файлів, робочих листів Excel, XML-документів, а також таблиць і представлень реляційних БД з джерела даних, до якого підключається пакет.

Потім дані трансформуються за допомогою перетворень, що містяться в службах SSIS. Після того, як дані перетворені в сумісні формати, вони можуть бути фізично об'єднані в один набір.

Після успішного злиття даних і застосування перетворень дані завантажуються по одному або де-

кільком призначенням. Служби SSIS містять призначення для завантаження даних у пласкі файли, необроблені файли, а також реляційні бази даних. Дані також можуть бути завантажені в набір записів, що зберігаються у пам'яті, та є доступними для інших елементів пакету. Дані в сховищах і вітринах даних оновлюються часто, а об'єм завантажуваних даних зазвичай досить великий. Служби SSIS містять завдання, яке проводить масове завантаження даних прямо з плаского файлу в таблиці та представлення SQL Server, а також компонент призначення, що виробляє масове завантаження даних у базу даних SQL Server у якості останнього кроку перетворення даних.

Пакет служб SSIS може бути налагоджений з можливістю перезапуску. Це означає, що можна перезапустити пакет із зумовленої контрольної точки. Можливість перезапуску пакету може значно економити час, особливо якщо пакет обробляє дані з великої кількості джерел. Можна використати пакети служб SSIS для завантаження у базу даних таблиць вимірів і фактів. Якщо джерело даних для таблиці виміру зберігається в декількох джерелах даних, то пакет може об'єднати дані в один набір і завантажити таблицю виміру впродовж одного процесу, замість використання окремих процесів для кожного джерела даних. Оновлення даних в сховищах і вітринах даних може стати складним завданням, оскільки обидва типи сховищ даних зазвичай містять виміри, що повільно змінюються і якими буває складно управляти за допомогою перетворення даних. Майстер вимірів, що повільно змінюються, автоматизує підтримку таких вимірів, динамічно створюючи інструкції SQL, які оновлюють та замінюють записи, оновлюють пов'язані записи, а також додають нові стовпці в таблиці.

Крім того, завдання і перетворення в пакетах служб SSIS можуть обробляти куби й виміри служб SSAS. Коли пакет оновлює таблицю у базі даних, на основі якої побудований куб, можна використати завдання і перетворення служб SSIS для автоматичної обробки куба, а також для автоматичної обробки вимірів. Автоматична обробка кубів і вимірів допомагає надавати поточні дані для користувачів обох середовищ: користувачам, які отримують дані з кубів і вимірів, а також користувачам, які отримують доступ до даних в реляційній базі даних.

Перед завантаженням даних для оперативної обробки транзакцій або у базу даних оперативної аналітичної обробки (OLAP), робочий лист Excel або файл, дані необхідно очистити і стандартизувати. Дані можуть вимагати оновлення з таких причин.

1. Дані були отримані у різних форматах. Перед тим, як дані мають бути використані, може бути потрібно перетворити їх в інший формат. Наприклад, може потрібно об'єднати ім'я і прізвище в один стовпець.

2. Дані можуть бути орендованими або придбаними. Перед тим, як дані можуть бути використані, може знадобитися їх стандартизація та очищення для відповідності прийнятним стандартам.

3. Формат даних залежить від мовного стандарту. Наприклад, дані можуть мати різні формати дати-часу, а також формати чисел. Якщо дані об'єднані з джерел різних мовних стандартів, то перед завантаженням їх необхідно привести в один формат, щоб уникнути ушкодження даних.

Служби SSIS містять вбудовані перетворення, які можна додати до пакетів для очищення і стандартизації даних, зміни регістра даних, перетворення даних в інший тип або формат, а також для створення нового стовпця даних на основі виразів. Наприклад, пакет може об'єднати стовпці, що містять імена і прізвища, в загальній стовпець повного імені, а потім перевести усі символи у верхній регістр.

Пакет служб SSIS також може провести очищення даних шляхом заміни значень в стовпцях на значення посилальної таблиці, використовуючи уточнювальні запити або нечіткі уточнювальні запити для пошуку значень в посилальній таблиці. Часто пакет спочатку застосовує уточнювальний запит і, у разі невдачі, нечіткий уточнювальний запит.

Таким чином, досліджено особливості застосування засобів Microsoft SQL Server Analysis Services (SSAS) для розроблення багатовимірного сховища просторових даних геопорталу з метою побудови на базі цього сховища просторових даних (СПД) системи підтримки прийняття рішень щодо геоінформаційних задач. В рамках цього дослідження сформульовано рекомендації стосовно розробки багатовимірних баз даних із використанням середовищ SQL BI Development Studio (BIDS) та SQL Server Management Studio (SSMS). При цьому для створення проекту служб SSAS в середовищі BIDS запропоновано алгоритм, що забезпечує стійке утворення проекту служб SSAS в середовищі BIDS або за допомогою шаблону проекту служб SSAS, або за допомогою майстра імпорту бази даних. Разом з тим, на основі аналізу особливостей застосування служби SQL Server Integration Services (SSIS) для роботи з багатовимірними базами даних сформульовано рекомендації щодо типових випадків застосування пакетів служб SSIS спільно із SSAS. Отже, оскільки визначено загальну стратегію розроблення системи підтримки прийняття рішень на базі сховища просторових даних геопорталу, навіть безвідносно до безмежних геоінформаційних завдань щодо геопорталів взагалі, цілком доцільно запропонувати методику практичної програмної реалізації саме багатовимірної бази даних геопорталів в цілому.

В даній роботі пропонується не стільки систематизація означеної дії у вигляді структурних схем та незаперечних програмних кодів, скільки практичні рекомендації, що нададуть суттєву допомогу розробникам у зазначеній сфері. Тим не менш, нижче пропонується програмна реалізація багатовимірного СПД геопорталу, яке є базою для СППР геоінформаційних задач.

Програмна реалізація багатовимірного сховища просторових даних геопорталу як бази підтримки прийняття рішень щодо геоінформаційних задач

Визначення методу з'єднання з сервером геопорталу. Програмування і налаштування джерел даних в середовищі BIDS призначене для визначення одного або декількох джерел даних для служб Microsoft SQL Server Analysis Services. Незалежно від того,

виконується робота з проектом служб Analysis Services або пряме підключення до сервера баз даних, визначити джерело даних можна на основі нового або існуючого з'єднання. Середовище SQL Server Analysis Services підтримує багато різних типів з'єднань (рис. 13).

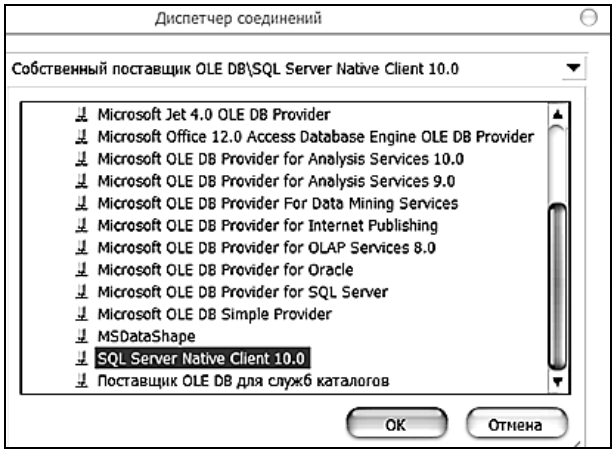

Рис. 13. Типи з'єднання з сервером середовища SSAS

У цій роботі застосовується (пропонується) з'єднання OLE DB (Object Linking and Embedding, Database) власного клієнта SQL Server. Цей набір інтерфейсів, заснованих на COM-технології, дозволяє застосункам звертатися до даних, що зберігаються в різних джерелах інформації або сховищах даних за допомогою уніфікованого доступу. Він розроблений для забезпечення високої продуктивності при підключенні до екземпляра (компонента) SQL Server Database Engine через канал OLE DB. Вибравши з'єднання, необхідно вказати дані про з'єднання, необхідні для підключення до бази даних. Це сервер або екземпляр служби, облікові дані для входу, ім'я бази даних та інші параметри, що відносяться до з'єднання. OLE DB підтримує службу перевірки достовірності екземпляра. При цьому варіанті зазвичай використовуються облікові дані процесу, що намагається отримати доступ до джерела даних задля надання доступу до них. Служби Analysis Services можна настроїти так, щоб при підключенні до базового джерела даних для виконання обробки OLAP використовувалися облікові дані запису служби, або ж задати призначений для користувача обліковий запис.

Проекти оперативної аналітичної обробки та інтелектуального аналізу даних на Microsoft SQL Server розроблені на основі логічної моделі даних пов'язаних таблиць, представлень та запитів з одного або декількох джерел даних. Ця логічна модель називається представленням джерел даних. Представлення джерела даних є об'єктом, що містить метадані з вибраних об'єктів джерела даних, включаючи зв'язки між цими об'єктами, визначені у базовому джерелі даних або в уявлених джерелах даних.

Представлення джерела даних створює кеш метаданих з джерел даних, на основі яких воно побудоване. Кешовані метадані дозволяють розробити проект служб Analysis Services без безперервного активного з'єднання з джерелом даних. Представлення джерел даних дозволяють визначати підмножину даних, що заповнюють велике сховище даних. Ще вони дозволяють визначити однорідну схему на основі різнорідних джерел даних або підмножин джерел даних. Оскільки представлення джерел даних представляють ізольовану схему, до них можна додавати усі необхідні замітки, не зачіпаючи схеми у базових джерелах даних. Представлення джерел даних містять такі елементи.

– Ім'я та опис.

– Визначення будь-якої підмножини схеми, що отримується з одного або декількох джерел даних, аж до повної схеми:

1) імена таблиць;
2) імена стовпців;
3) типи даних;
4) допустимість значень NULL;
5) довжини стовпців;
6) первинні ключі;
7) зв'язки «первинно-зовнішній ключ».

– Замітки до схеми з базових джерел даних, включаючи:

1) зрозумілі імена таблиць, представлень і стовпців;
2) іменовані запити, що повертають стовпці з одного або декількох джерел даних (що відображаються в схемі як таблиці);
3) іменовані обчислення, що повертають стовпці з джерела даних (що відображаються як стовпці таблиць або представлень);
4) логічні первинні ключі (необхідні, якщо первинний ключ у базовій таблиці не визначений або не включений в представлення чи іменований запит);
5) зв'язки «логічний первинний-зовнішній ключ» між таблицями, представленнями і іменованими запитами.

Щоб створити представлення джерела даних, в основі якого лежать декілька джерел даних, необхідно заздалегідь визначити представлення, засноване на одному джерелі даних. Це джерело даних згодом вважається первинним джерелом даних. Потім можна додати таблиці і представлення зі вторинного джерела даних. При проектуванні вимірів з атрибутами, заснованих на пов'язаних таблицях декількох джерел даних, може знадобитися визначення джерела даних Microsoft SQL Server як основного джерела даних, щоб використати його можливості обробки розподілених запитів. Вміст представлення джерела даних також відображається в конструкторі представлень джерел даних в середовищі Business Intelligence Development Studio (рис. 14). У цьому конструкторі містяться такі елементи:

– панель «Діаграма», на якій представлені в графічному виді таблиці та їх зв'язки;

– панель «Таблиці», на якій відображаються у вигляді дерева таблиці та елементи їх схем;

– панель «Організатор схем», на якій можна створювати вкладені діаграми, що дозволяють переглядати піднабори представлення джерела даних;

– панель інструментів конструктора представлень джерел даних.

Значення властивості FriendlyName для об'єкту в уявленні джерела даних можна змінити, щоб зробити його ім'я зрозумілішим. Імена цих об'єктів також можна змінювати після того, як вони були визначені. Далі буде змінено зрозуміле ім'я кожної з таблиць в

уявленні джерела даних. Це зробить зрозумілішими імена об'єктів куба і вимірів, які будуть визначені в роботі. Також в уявленні джерела даних можна привласнювати зрозумілі імена стовпцям, визначати обчислювані стовпці, сполучати таблиці і представлення, щоб зробити їх зручнішими в роботі (рис. 15).

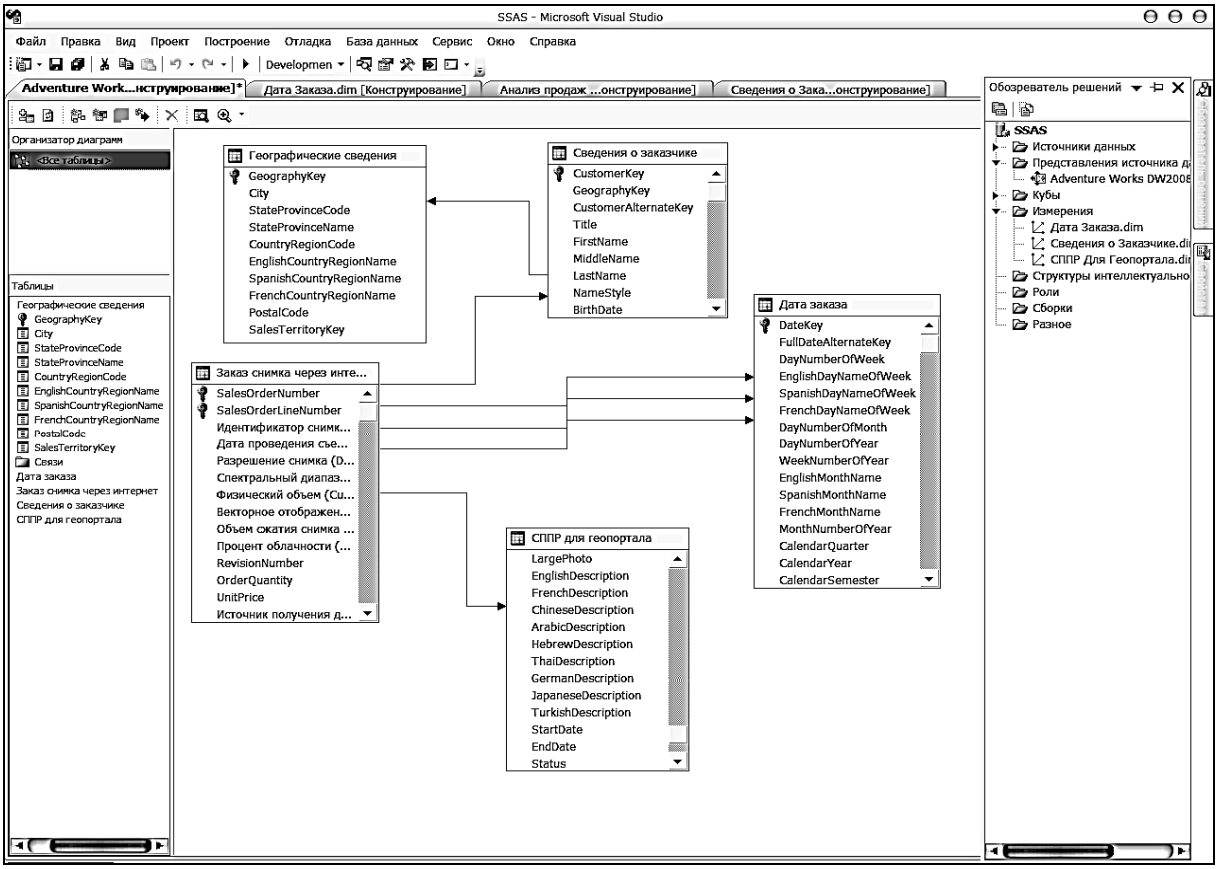

Рис. 14. Конструктор представлень джерел даних в середовищі Business Intelligence Development Studio

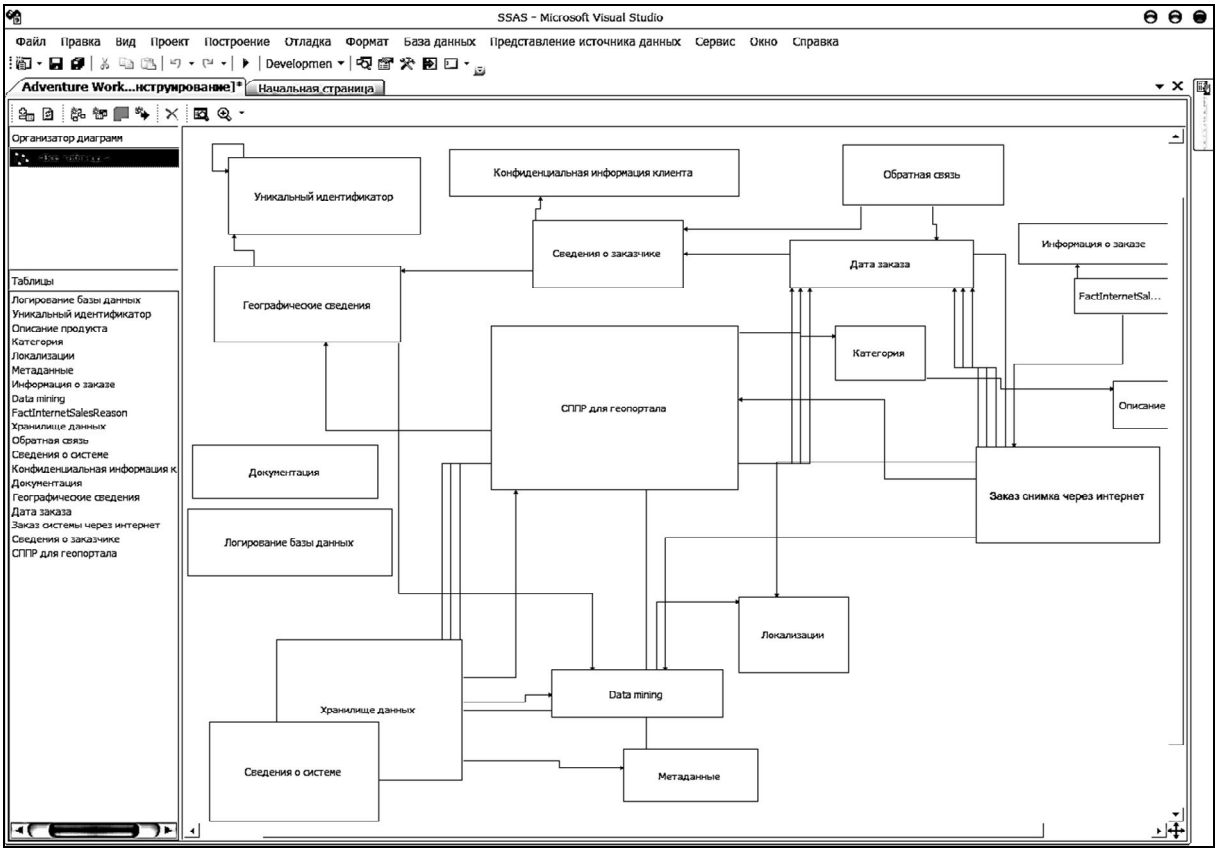

Рис. 15. Структура бази просторових даних системи підтримки прийняття рішень для геопорталу

Завдання параметрів вимірів OLAP-кубів і розгортка аналітичних служб. Куб є багатовимірною структурою, що містить виміри і заходи. Виміри визначають структуру куба, а заходи надають числові значення, що представляють інтерес для кінцевого користувача. В якості логічної структури куб дозволяє клієнтському застосуванню набувати значень у вигляді осередків куба, визначених для усіх можливих сумарних значень. Положення осередків в кубі визначаються перетином елементів вимірів. Вимір бази даних є колекцією об'єктів, що називаються атрибутами, які використовуються для надання відомостей про дані фактів в одному або декількох кубах. Ці об'єкти прив'язані до одного або до декількох стовпців в одній або декількох таблицях в уявленні джерела даних. За умовчанням ці атрибути відображаються як ієрархії атрибутів і дозволяють зрозуміти сенс даних фактів у кубі. Атрибути можуть бути організовані в призначені для користувача ієрархії, які забезпечують різні шляхи доступу до даних і допомагають користувачам при перегляді даних в кубі. Куби містять усі виміри, потрібні користувачам при аналізі даних. Екземпляр виміру бази даних в кубі називається виміром куба і відноситься до одної або декількох груп заходів в кубі. Наприклад, якщо таблиця фактів містить декілька залежних від часу фактів, то для полегшення аналізу кожного з них може бути визначений окремий вимір куба. Проте, існування тільки одного залежного від часу виміру бази даних означає необхідність існування лише однієї залежної від часу таблиці реляційної бази даних для підтримки декількох залежних від часу вимірів куба.

Для визначення вимірів, атрибутів та ієрархій бази даних і куба найпростіше скористатися майстром кубів, який дозволяє створювати виміри одночасно з визначенням куба. Майстер кубів створює виміри на основі виявлених або вказаних користувачем таблиць вимірів з представленням джерела даних, використовуваного для куба. Після цього майстер створює виміри бази даних і додає їх до нового куба, створюючи виміри куба. При створенні куба в нього можна також додати будь-які виміри, які вже існують у базі даних. Ці виміри могли бути раніше визначені майстром вимірів для іншого куба. Після того, як вимір бази даних визначений, його зміна й налаштування проводяться в конструкторі вимірів. Вимір куба може бути налагоджений за допомогою конструктора кубів (рис. 16). Програміст має можливість перегляду і редагування коду, що генерується майстром кубів. Для цього у властивостях куба необхідно вибрати функцію «Перехід до коду» (рис. 17).

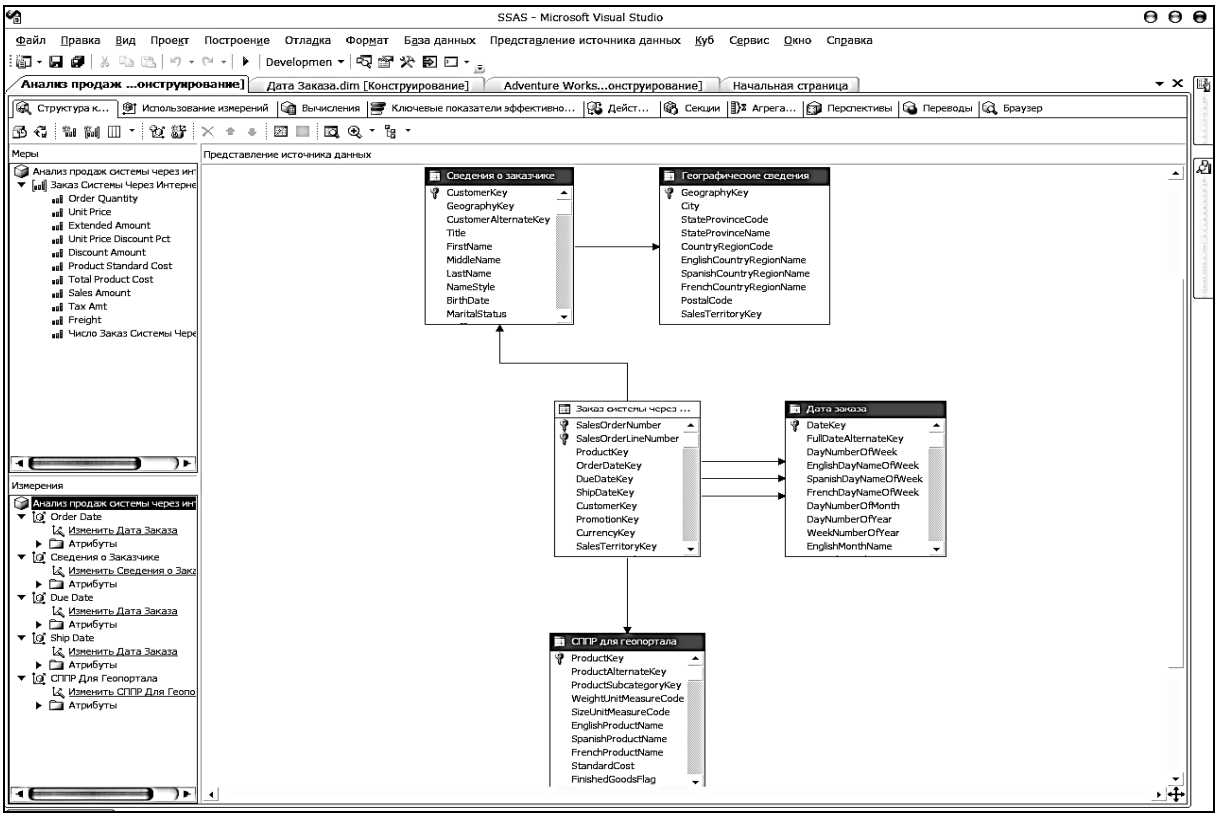

Рис. 16. Конструктор кубів

Після написання коду вимірів куба можна переходити до розгортки аналітичних служб, спираючись на XML-файл з кодогенерацією майстра кубів. В процесі розробки проекту в середовищі BI Dev Studio проект часто розгортається на сервері розробки для створення бази даних служб SSAS, визначеної проектом. Це необхідно для тестування проекту, наприклад для огляду осередків у кубі, огляду елементів виміру або перевірки формул ключових індикаторів продуктивності. Проект можна розгорнути незалежно або ж розгорнути усі проекти в рішенні. При розгортанні проекту послідовно виконуються такі дії: побудова проекту і створення вихідних файлів, що визначають базу даних служб SSAS та відповідні складові об'єкти; перевірка вибраного сервера; створення бази даних та її об'єктів на вибраному сервері.

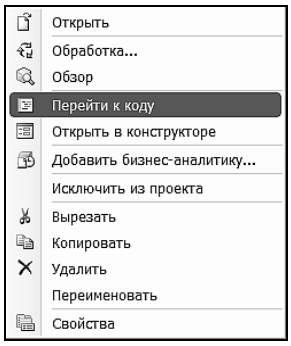

Рис. 17. Вибір функції редагування коду

Після первинного розгортання в теці <Ім'я проекту> створюється файл IncrementalSnapshot.xml. Цей файл потрібний для визначення, чи змінювалася база даних або її об'єкти на цільовому сервері поза проектом. Якщо відбувалася зміна, система запропонує переписати усі об'єкти в цільовій базі даних. Якщо усі зміни були зроблені в проекті і проект налагоджений для додаткового розгортання, на цільовому сервері будуть розгорнуті тільки зміни.

Конфігурація проекту і відповідні параметри визначають властивості розгортання, які застосовуватимуться при розгортанні проекту. У загальному проекті кожен розробник використовує свою власну конфігурацію з власними параметрами конфігурації проекту. Наприклад, кожен розробник може вказати окремий тестовий сервер для роботи окремо від інших розробників. Розгортання тестового проекту на локальному сервері віртуальної машини має такий вигляд (рис. 18). Після розгортки багатовимірного куба необхідно вказати, який зріз необхідно переглянути. Для цього слід вибрати виміри, які будуть відкладені по осях куба. На рис. 19 показаний зріз куба «Кількість замовлених знімків - Місто - Час». Після закінчення роботи з кубом, можна використати результати аналізу для створення звітів і демонстрації результатів у зручній для користувача формі. На рис. 20 показано приклад використання надбудов інтелектуального аналізу даних SQL Server для Microsoft Office щодо візуалізації результатів OLAP-аналізу.

Таким чином, отримано програмну реалізацію багатовимірного сховища просторових даних геопорталу, що забезпечує підвищення оперативності статистичного і прогнозного аналізу геоданих для ведення звітності та підтримки прийняття рішень геоінформаційних задач в режимі реального часу.

Висновки

На підставі проведеного аналізу архітектур сучасних систем підтримки прийняття рішень запропоновано структуру, яку найдоцільніше використовувати для побудови геоінформаційної СППР на базі сховища просторових даних геопорталу. Проаналізовано основні функції, що мають виконувати підсистеми такої СППР, а саме — функції підсистеми витягання, перетворення і завантаження даних, функції підсистеми зберігання даних та підсистеми аналізу даних. Розгляд зазначених підсистем СППР проведено із детальним вивченням їх структурних та функціональних схем, аналізом існуючих технологій обробки та зберігання просторової інформації, а також сучасних методів аналізу геоданих.

Разом з тим, здійснено ґрунтовний вибір системи управління базами даних для практичної реалізації багатовимірної бази геоданих, що виконує функції сховища просторових даних як основи СППР геопорталу.

Проведено ретельний аналіз засобів Microsoft SQL Server Analysis Services (SSAS) як базової платформи для розгортання аналітичних систем Business Intelligence (BI), уніфікованої багатовимірної моделі даних Unified Dimensional Model (UDM) як ядра технології комплексного багатовимірного аналізу даних Online Analytical Processing (OLAP). Проаналізовано аспекти планування і програмування архітектури SSAS, особливості розроблення багатовимірних баз даних із застосуванням середовищ SQL BI Development Studio (BIDS) та SQL Server Management Studio (SSMS), а також специфіку застосування служби SQL Server Integration Services (SSIS) для роботи з багатовимірними базами даних.

Таким чином, підготовлено підґрунтя для розроблення геоінформаційної СППР із застосуванням засобів Microsoft SQL Server багатовимірного аналізу даних для створення багатовимірного сховища просторових даних геопорталу. Врешті, запропоновано методику програмної реалізації багатовимірного сховища просторових даних геопорталу.

Запропоновані підходи щодо розроблення геоінформаційної СППР на базі сховища просторових даних геопорталу забезпечують об’єднання та інтеграцію у сховище геопорталу інформації, що містить різні типи просторових даних. Програмна реалізація багатовимірного СПД геопорталу забезпечує підвищення оперативності статистичного і прогнозного аналізу геоданих для ведення звітності та підтримки прийняття рішень геоінформаційних задач в режимі реального часу.

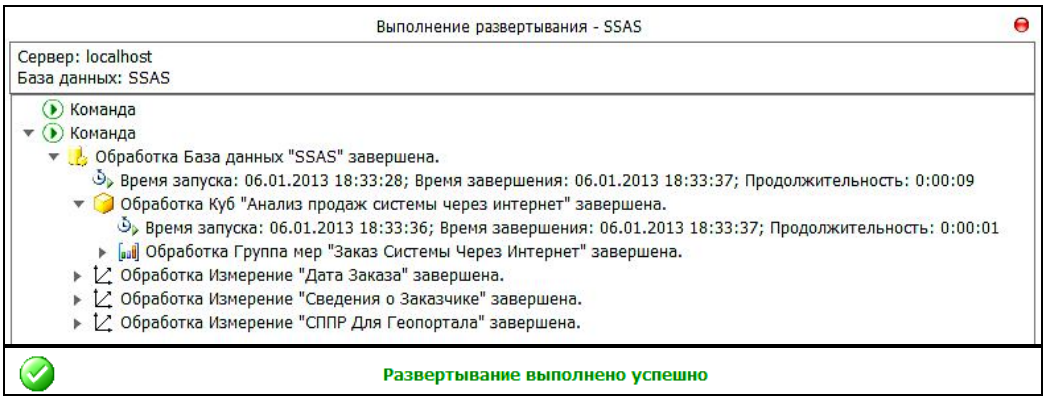

Рис. 18. Розгортка аналітичних служб на локальному сервері

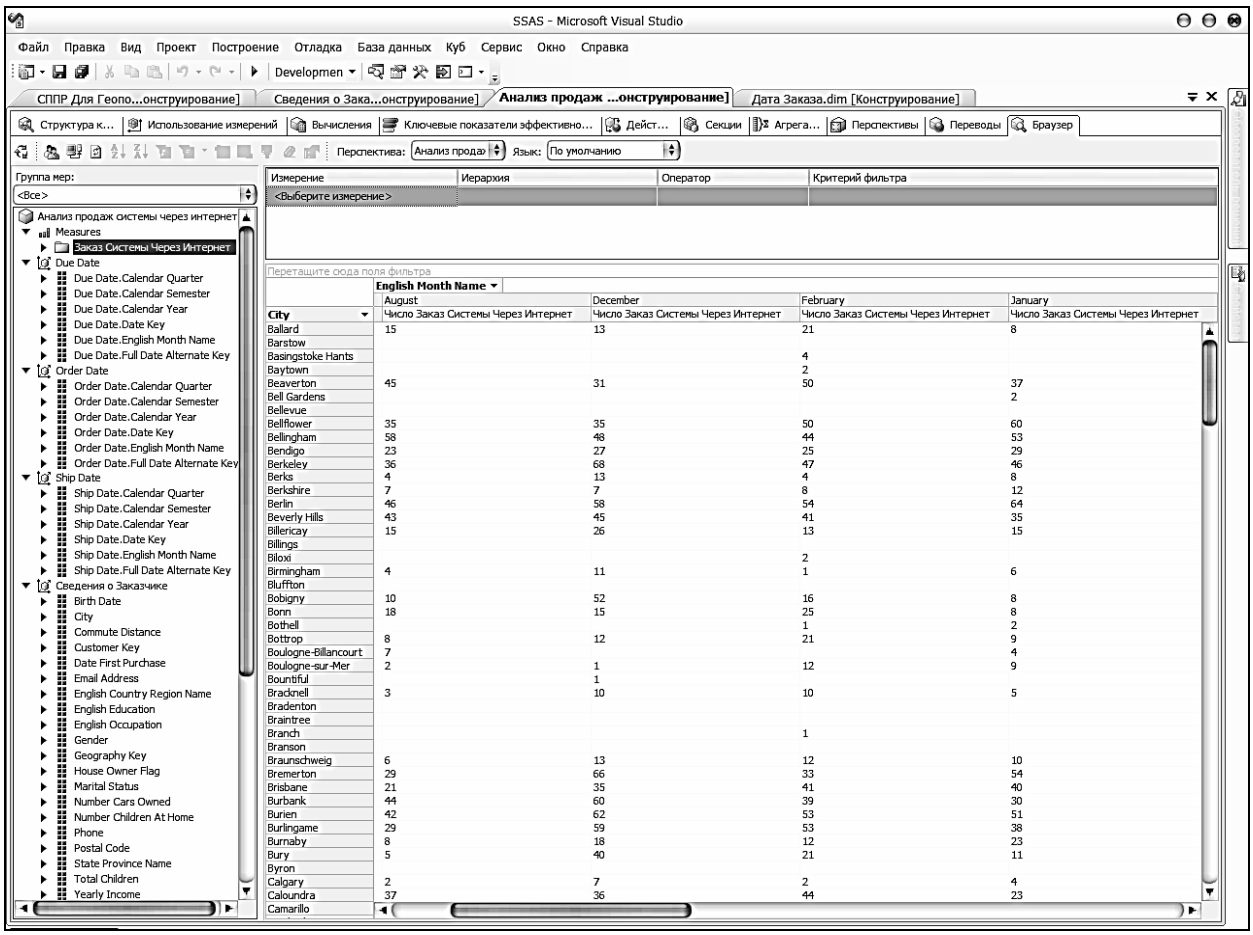

Рис. 19. Зріз куба за трьома параметрами

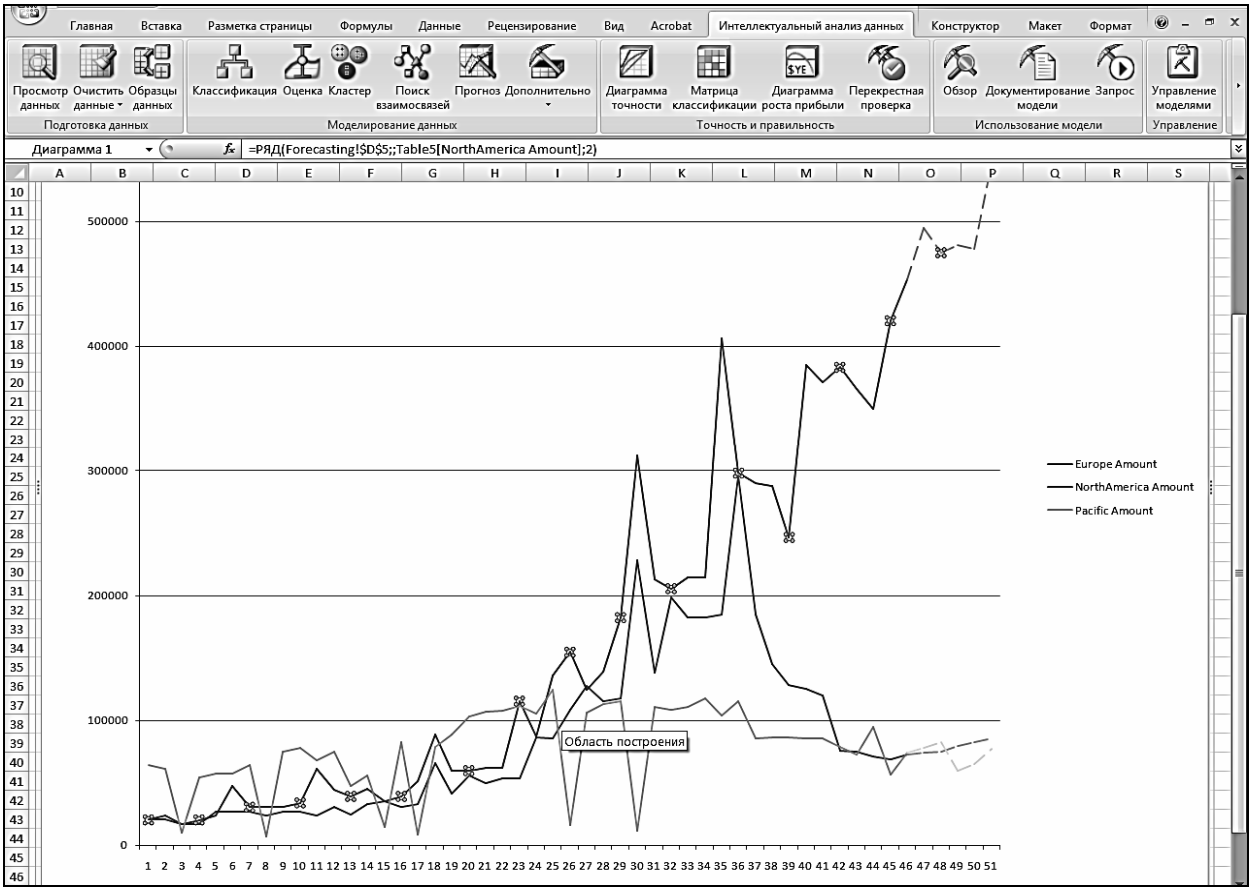

Рис. 20. Екстраполяційні оцінки замовлень знімків по статистичному часовому ряду

СПИСОК ЛІТЕРАТУРИ

1. Андрєєв С.М., Жилін В.А. Методика створення атласів історичних картографічних моделей за даними аерофотозйомки з використанням геоінформаційних технологій. *Сучасні інформаційні системи*. 2020. Т. 4, № 1. С. 45-62. DOI: http://dx.doi.org/10.20998/2522-9052.2020.1.08
2. Андрєєв С. М., Жилін В. А., Мельник А. П. Застосування анаморфозних картографічних моделей для аналізу геоданих. *Сучасні* інформаційні *системи*. 2019. Т. 3, № 3. С. 5-16. DOI: http://dx.doi.org/10.20998/2522-9052.2019.3.01
3. Андрєєв С.М., Жилін В.А. Застосування даних аерофотозйомки з безпілотних літальних апаратів для побудови 3D - моделей місцевості. *Системи управління, навігації та зв'язку*: збірник наукових праць. Полтава: Полтавський НТУ ім. Юрія Кондратюка, 2019. Вип. 1(53). С. 3-16. DOI: http://dx.doi.org/10.26906/SUNZ.2019.1.003
4. Marakas G. M. Decision support systems in the twenty-first century. Upper Saddle River, N.J.: Prentice Hall, 1999.
5. Ларичев О. И., Петровский В. А. Системы поддержки принятия решений: Современное состояние и перспективы их развития. *Итоги науки и техники. Сер. Техническая кибернетика*. Т. 21. М.: ВИНИТИ, 1987. С. 131-164
6. Bonczek R.H., Holsapple C., Whinston A.B. Foundations of Decision Support Systems. New York: Academic Press, 1981.
7. Анисимова О. Л., Зраенко Д. Ю., Комоско В. В. Хранилище пространственных объектов в составе регионального уз-ла ИПД УрФО: модель хранилища. *Пространственные данные*. 2010. № 1. С. 62.
8. Inmon W. H. Building the Data Warehouse. New-York: John Wiley & Sons, 1991. 312 p.
9. Вейцман В. М. Проектирование экономических информационных систем. Ярославль: МУБиНТ, 2002. 213 с.
10. Терелянский П. В. Системы поддержки принятия решений. Опыт проектирования. Волгоград: ВолгГТУ, 2009. 127 с.
11. Шеннон К. Работы по теории информации и кибернетике / К. Шеннон. – М.: ИИЛ, 1963. – 829 с.
12. Спирли Э. Корпоративные хранилища данных / Э. Спирли // Планирование, разработка, реализация. – М.: Вильямс, 2001. – 400 с.
13. Кузнецов С.Д. Проектирование и разработка корпоративных информационных систем / Д.С. Кузнецов. – М., 1998. – 120 с.
14. ДСТУ ISO 19101: 2002(Е) Комплекс стандартів «Географічна інформація / Геоматика» // Географічна інформація – еталонна модель. Київ, 2005.
15. Серебряков С.В. Новый подход к организации и хранению пространственных данных / В.С. Серебряков, Ю.Д. Баженова // Геодезия и картография, 2008. – Вып. 7.
16. Косяков С.В. Об интеграции ГИС и прикладных программных комплексов / В.С. Косяков, Н.В. Никольский // Материалы форума ГИС'97. М.: ГИС-Ассоциация, 1997.

REFERENCES

1. Andrieiev, S. and Zhilin, V. (2020), “Methodology of creation of atlases of historical cartographic models from data of aerophotography with the use of геоінформаційних technologies”, *Advanced Information Systems*, Vol. 4, No. 1, pp. 45-62, DOI: http://dx.doi.org/10.20998/2522-9052.2020.1.08
2. Andrieiev, S., Zhilin, V. and Melnyk, A. (2019), “The use of anamorphosis cartographic models for geodata analysis”, *Advanced Information Systems*, Vol. 3, No. 3, pp. 5-16, DOI: http://dx.doi.org/10.20998/2522-9052.2019.3.01
3. Andreev, S. and Zhilin, V. (2019), “Application of aerophotic data with unmanned aircraft for developing 3D models of terrain”, *Control, navigation and communication systems: collection of scientific works*, No. 1(53), Poltava NTU Yuri Kondratyuk, Poltava, pp. 3-16, DOI: http://dx.doi.org/10.26906/SUNZ.2019.1.003
4. Marakas, G.M. (1999), *Decision support systems in the twenty-first century*, Upper Saddle River, N.J, Prentice Hall.
5. Larichev, O.I. and Petrovsky V.A. (1987), “Decision Support Systems: The current state and prospects of their development”, *The results of science and technology*, Avg. Technical cybernetics, VINITI, Moscow, Vol. 21, pp. 131-164.
6. Bonczek, R.H., Holsapple, C. and Whinston, A.B. (1981), *Foundations of Decision Support Systems*, Acad. Press, New York.
7. Anisimova, O.L., Zraenko, D.Yu. and Komosko, V.V. (2010), ”Storage of spatial objects as a part of the regional node of IPD UrFO: model of storage”, *Spatial data*, No. 1, p. 62.
8. Inmon, W.H. (1991), *Building the Data Warehouse*, John Wiley & Sons, New-York, 312 p.
9. Weizmann, V.M. (2002), *Design of economic information systems*, MUBiNT, Yaroslavl, 213 p.
10. Terelyansky, P.V. (2009), *Decision support systems. Experience of design*, Volga State Techn. University, Volgograd, 127 p.
11. Shannon, K. (1963), *Works on information theory and cybernetics*, IIL, Moscow, 829 p.
12. Spirley, E. (2001), *Corporate data warehouses. Planning, development, implementation*, Williams, Moscow, 400 p.
13. Kuznetsov, S.D. (1998), *Design and development of corporate information systems*, Moscow, 120 p.
14. DSTU ISO 19101: 2002 (E) (2005), *Geographic Information* / Geomatics Complex, Geographic Information – Reference Model, Kyiv.
15. Serebryakov, S.V. and Bazhenov, Yu.D. (2008), “A new approach to the organization and storage of spatial data”, *Geodesy and Cartography*, Vol. 7.
16. Kosyakov, S.V. and Nikolsky N.V. (1997), “On the integration of GIS and applied software”, *Materials of the GIS'97 forum*, GIS-Association, Moscow.

Received (Надійшла) 27.02.2020
Accepted for publication (Прийнята до друку) 20.05.2020

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Андрєєв Сергій Михайлович – кандидат технічних наук, доцент, доцент кафедри геоінформаційних технологій та космічного моніторингу Землі, Національний аерокосмічний університет ім. М. Є. Жуковського «ХАІ», Харків, Україна;
Sergey Andrieiev – Candidate of Technical Sciences, Associate Professor, Associate Professor of Geoinformation Technologies and Space Monitoring of the Earth Department, National Aerospace University named after N. Ye. Zhukovskiy "Kharkiv Aviation Institute", Kharkiv, Ukraine;
e-mail: AndreevSM@gmail.com; ORCID ID: https://orcid.org/0000-0003-4256-2637

Жилін Володимир Анатолійович – кандидат технічних наук, доцент, доцент кафедри геоінформаційних технологій та космічного моніторингу Землі, Національний аерокосмічний університет ім. М. Є. Жуковського «ХАІ», Харків, Україна;
Volodymyr Zhilin – Candidate of Technical Sciences, Associate Professor, Associate Professor of Geoinformation Technologies and Space Monitoring of the Earth Department, National Aerospace University named after N. Ye. Zhukovskiy "Kharkiv Aviation Institute", Kharkiv, Ukraine;
e-mail: v.zhilin@khai.edu; ORCID ID: https://orcid.org/0000-0002-7342-3456

Геоинформационная система поддержки принятия решений на базе хранилища пространственных данных геопортала

С. М. Андреев, В. А. Жилин

Аннотация. Предметом исследования является разработка геоинформационной системы поддержки принятия решений (СППР) на базе хранилища пространственных данных (ХПД) геопортала. **Объектом исследования** является процесс объединения и интеграции в хранилище геопортала информации, содержащей различные типы пространственных данных. **Целью работы** является повышение оперативности статистического и прогнозного анализа геоданных для ведения отчетности и поддержки принятия решений геоинформационных задач в режиме реального времени. **Выводы.** На основании проведенного анализа архитектур современных систем поддержки принятия решений предложена структура, которую целесообразно использовать для построения геоинформационных СППР на базе хранилища пространственных данных геопортала. Проанализированы основные функции, которые должны выполнять подсистемы такой СППР, а именно – функции подсистемы извлечения, преобразования и загрузки данных, функции подсистемы хранения данных и подсистемы анализа данных. Рассмотрение указанных подсистем СППР проведено с детальным изучением их структурных и функциональных схем, анализом существующих технологий обработки и хранения пространственной информации, а также современных методов анализа геоданных. Вместе с тем, осуществлен обоснованный выбор системы управления базами данных (СУБД) для практической реализации многомерной базы геоданных, которая выполняет функции хранилища пространственных данных как основы СППР геопортала. Проведен тщательный анализ средств Microsoft SQL Server Analysis Services (SSAS) как базовой платформы для развертывания аналитических систем Business Intelligence (BI), унифицированной многомерной модели данных Unified Dimensional Model (UDM) как ядра технологии комплексного многомерного анализа данных Online Analytical Processing (OLAP). Проанализированы аспекты планирования и программирования архитектуры SSAS, особенности разработки многомерных баз данных с применением сред SQL BI Development Studio (BIDS) и SQL Server Management Studio (SSMS), а также специфика применения службы SQL Server Integration Services (SSIS) для работы с многомерными базами данных. Таким образом, подготовлена почва для разработки геоинформационных СППР с применением средств Microsoft SQL Server многомерного анализа данных для создания многомерного хранилища пространственных данных геопортала. Также предложена методика программной реализации многомерного хранилища пространственных данных геопортала. Предложенные подходы по разработке геоинформационных СППР на базе хранилища пространственных данных геопортала обеспечивают объединение и интеграцию в хранилище геопортала информации, содержащей различные типы пространственных данных. Программная реализация многомерного ХПД геопортала обеспечивает повышение оперативности статистического и прогнозного анализа геоданных для ведения отчетности и поддержки принятия решений геоинформационных задач в режиме реального времени.

Ключевые слова: геоданные; геоинформационная система поддержки принятия решений; многомерные хранилища пространственных данных.

Geoinformation system of decision support based on the geoportal spatial data storage

S. Andrieiev, V. Zhilin

Abstract. The subject of the study is the development of a geographic information system for decision support (DSS), based on the storage of spatial data of the geoportal. **The object of the study** is the process of combining and integrating information containing various types of spatial data into the geoportal repository. **The purpose of the work** is to increase the efficiency of statistical and predictive analysis of geodata for reporting and supporting decision-making of geographic information tasks in real time. **Conclusions.** Based on the analysis of the architectures of modern decision-making systems, a structure of geoinformation DSS's based on the spatial data storage of the geoportal is proposed. The main features and functions that are included in subsystems of such a DSS are analyzed, namely the functions of the subsystem for extracting, converting and loading data, the functions of the data storage subsystem and the data analysis subsystem. The consideration of mentioned subsystems DSS was carried out with a detailed study of their structural and functional schemes, analysis of existing technologies of processing and spatial information storage, as well as modern methods of geodata analysis. At the same time, a reasonable choice of a database management system for the practical implementation of a multidimensional geodatabase is made, which performs the functions of a spatial data storage as the basis of a DSS geoportal. A substantial analysis of Microsoft SQL Server Analysis Services (SSAS), as a base platform for the deployment of Business Intelligence (BI) analytical systems, a Unified Dimensional Model (UDM) as the core of Online Analytical Processing (OLAP) complex multidimensional data analysis technology was performed. The aspects of planning and programming the SSAS architecture, the features of developing multi-dimensional databases using SQL BI Development Studio (BIDS) and SQL Server Management Studio (SSMS), as well as the specifics of using SQL Server Integration Services (SSIS) to work with multi-dimensional databases are analyzed. Thus, the groundwork has been set for the development of geoinformation DSS using Microsoft SQL Server multidimensional data analysis tools to create a multidimensional spatial data storage of the geoportal. A methodology for software implementation of a spatial data multidimensional storage of a geoportal is also proposed. The proposed approaches to the development of geoinformation DSS based on the spatial data storage provide integration into the geoportal repository of information containing various types of spatial data. The software implementation of the multidimensional geoportal provides an efficiency level increase of statistical and predictive geodata analysis for reporting and supporting decision-making of geographic information tasks in a real time.

Keywords: geodata; geoinformation decision support system; multidimensional storage of spatial data.