

## **Проблеми створення інформаційних сховищ на основі стаціонарних компонентів баз даних та управління метаданими**

Першими по-справжньому ефективними, з погляду співвідношення витрат і отриманих вигод, типів інтелектуального програмного забезпечення стали сховища даних. Вони дозволили отримати перший практично цінний результат від зусиль, тривалий час направлених на створення архітектурних принципів побудови. Сховище даних призначено для забезпечення інформацією аналітиків і керівників в різних аспектах - не підготовлених наперед запитів. Одна з основних задач сховищ даних надати аналітикам і керівникам доступ до злагодженої картини господарської діяльності.

У результаті дослідження в області інформаційних сховищ з'ясувалося, що не існує розробок і немає спроб у створенні єдиного підходу до керування метаданими в різних предметних областях. У попередній роботі була розроблена єдина методика створення інформаційних сховищ [1].

У зв'язку з цим виникла задача безпосередньо переносу стаціонарних атрибутів, компонентів в інформаційне сховище з баз даних різних предметних областей, а внаслідок цього і робота над метаданими.

На даний момент не існує прикладів застосування операцій над метаданими після переносу стабільних компонентів різних баз даних в інформаційне сховище.

Відповідно, виникла проблема створення інформаційного сховища, яке здійснюється на основі стаціонарних компонентів баз даних, та керування метаданими.

Основною задачею дослідження є перенос інформації з баз даних до інформаційних сховищ та застосування операцій реляційної алгебри до метаданих, які можуть застосовувати при проектуванні сховища в різних предметних областях.

До сьогоднішнього моменту виникла проблема, яка не була освітлена ні в літературних джерелах, ні в повній мірі в джерелах мережі Internet, що приводить до необхідності рішення задачі.

Використання інформаційних сховищ як напрямку інформаційних технологій з'явилося зовсім недавно і вже є актуальним. Виникає маса проблем при створенні сховища.

Традиційні підходи до створення інформаційних сховищ ґрунтуються виключно на моделюванні статичного представлення реального світу. З погляду моделювання часу сховища даних принципово відрізняються від оперативних систем. Моделі сховищ даних інтенсивно використовують тимчасові відмітки [2].

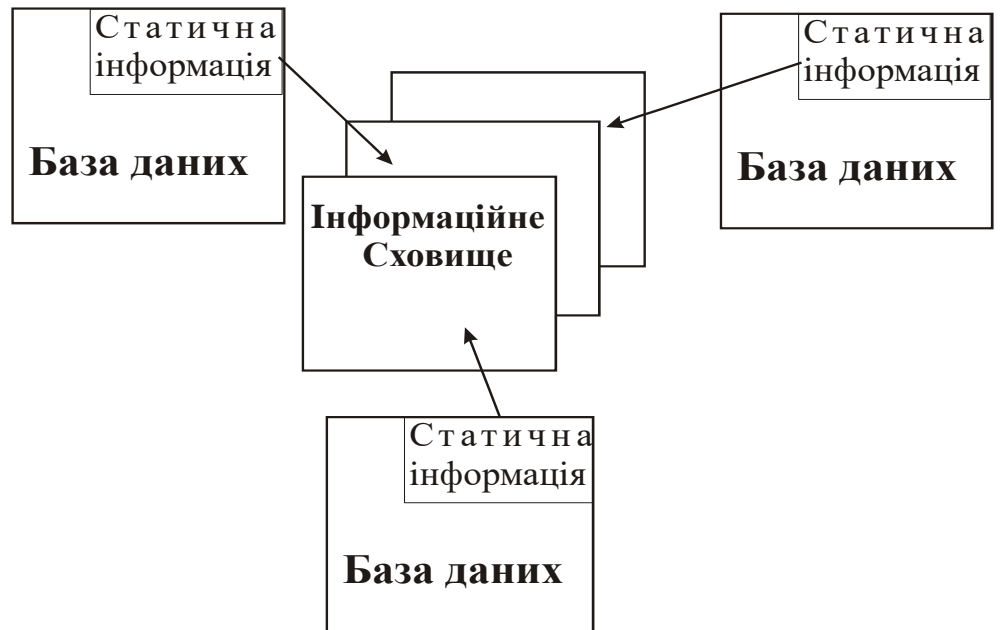
На мал.1 представлений загальний вид створення інформаційного сховища. Як видно з малюнка, інформаційне сховище створюється шляхом переносу статичної інформації з баз даних. Протягом визначеного проміжку часу в базах даних формується статична інформація. Тобто та інформація, що після закінчення тривалого терміну не змінюється. І в даному випадку, має місце виділяти дану інформацію в деяку статичну базу даних. Для цих цілей безпосередньо служить інформаційне сховище, що буде накопичувати стабільні дані. Тим самим, інформаційне сховище є граничним станом стаціонарних компонентів баз даних.

Виходячи з вище викладеного припущення, у базах даних чітко повинна просліджуватися інформація для подальших операцій з нею. Але це буде здійснюватися в єдиній базі – інформаційному сховищі. Дана ідея має місце у випадку, якщо на підприємстві виникають задачі, зв'язані з аналізом інформації.

Тоді, виникає і ряд операцій стосовно інформаційного сховища. Дані, що надходять з баз даних, як правило, стабільні та різномірні. На рівні фізичної реалізації інформаційного сховища необхідно привести дані до єдиного формату.

Фізична реалізація сховищ даних може в значній мірі розрізнятися. Звичайно сховище даних містить частини, які обслуговують поточні і історичні багатомірні аналітичні дані для підтримки ухвалення рішень[3]. Різні за часом зрізи можуть використовувати різні технології для зберігання і доступу до даних. Поточні дані, що деталізуються, можуть бути спроектовані як окремі куби, об'єднуючі не тільки поточні,

але і історичні дані. Історичні дані також можуть бути оформлені у вигляді окремих кубів або партицій більш загальних кубів. У тому випадку, коли вони оформлені у вигляді окремих кубів, як правило, створюються віртуальні куби, тобто куби, які не мають власних даних, а спираються виключно на дані інших - реальних кубів, з яких вони полягають.



Мал.1 Створення інформаційних сховищ на основі стаціонарних компонент баз даних

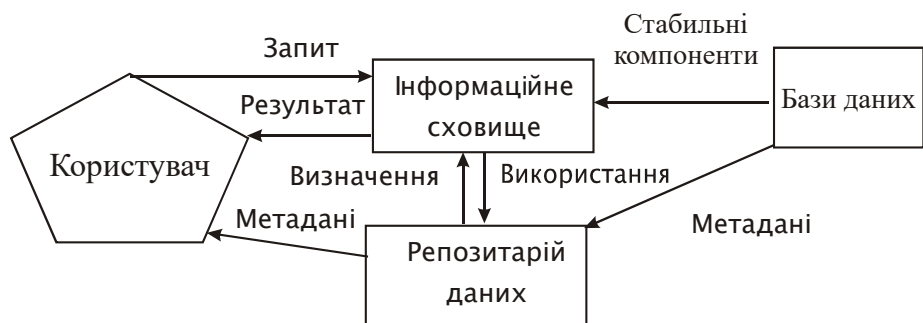
Відповідно до визначення архітектури інформаційного сховища, що приводиться Ериком Сперлі, дані з оперативних систем і зовнішніх джерел надходять у сховище даних[4]. Як вище зазначено стає питання управління метаданими. Однак, тоді потрібно визначитися зі збереженням метаданих, до яких планується застосування операцій реляційної алгебри. Опис метаданих повинно десь зберігати.

Це питання вирішено наступним образом: опис семантичних структур передається в репозитарій метаданих - базу даних, що зберігає опис семантичних структур із указівкою технічних підсистем їх підтримуючих.

Тоді користувач буде формувати запит на підставі інформації, отриманої з репозитарія метаданих. Запити направляються безпосередньо в сховище, де виконуються вибірки даних і формуються результати запитів. Здійснення запитів буде відбуватися в кілька етапів:

- користувач дає запит в інформаційне сховище;
- інформаційне сховище містить дані з оперативних систем (запис даних у деяке проміжне сховище, де вони проходять узгодження й очищення);
- сховище даних одержує інформацію про метадані у репозитарії даних;
- користувач одержує інформацію про метадані у репозитарії даних;
- виконується запит.

У такій послідовності явно позначений процес перенесення даних з оперативних систем і зовнішніх джерел у сховище даних, використовуючі метаінформацію (див.мал.2).



Мал.2 Схема формування запитів

Треба визначитися, що в процесі формування метаданих кожному елементу ієрархії присвоюється унікальний код, що включає код "батьківського вузла". Такий код дозволяє реалізувати роботу з укрупненими значеннями параметрів.

Необхідні назви таблиць і полів будуть міститися в метайнформації, і відтворюються при проходженні ієрархії відповідно до коду параметра.

Нарешті, після переносу даних з баз даних до інформаційних сховищ, для управління метаданими потрібні підсистеми аналізу накопиченої інформації. Це питання ускладнюється, якщо становиться питання застосування цих підсистем до різних предметних областей. З їх допомогою можна здійснювати вибірки, виконувати над ними теоретико-множинні операції і кількісний аналіз, використовувати реляційну алгебру над таблицями. При цьому передбачається отримати нові знання про невідомі поки закономірності і залежності.

Докладніше зупинимося на процесі формування метаданих.

В потоці інформації необхідно дуже швидко орієнтуватися. І тут наявні паперові каталоги, списки, довідки, отримані з технічних носіїв, вже не допомагають. По-перше, інформація дуже швидко змінюється, масиви поповнюються, коректуються і перераховуються. По-друге, об'єм такої інформації великий. По-третє, довідкова інформація - це основа для переходу на безпаперову технологію обробки інформації про стан середовища. Для вибірки даних потрібні атрибути метаданих, яких іноді немає в початкових масивах даних. Таким чином, для успішного пошуку високоякісних даних необхідні різноманітні метадані - відомості про масиви даних, формати їх обміну, програмні засоби їх обробки, організація, що збирають і бережуть дані і інші.

Коли пошук даних здійснюють користувачі, що працюють в одній предметній області, то їм, частіше всього, достатньо одного атрибута. Складніша ситуація у осіб, що працюють в різних предметних областях. Тут потрібні більш повні і докладні метадані не тільки про самий масив даних, але і з яких платформ дані отримані, в рамках яких проєктів вони створені, з використанням яких програм. Виникає маса питань по управлінню метаданими.

Метадані виникають вже у момент спостережень, а далі вони узагальнюються, доповнюються новими атрибутами.

Основними об'єктами метаданих є:

- джерела даних;
- інформаційні ресурси;
- інформація для управління даними (формати, програмні засоби, кодифікатори, моделі);
- відомості про дані (прогнози).

Характерною межею метаданих є їх невисока частота змін відносно потреб фірми, задач та функцій, необхідних для прийняття рішення. Таке положення зв'язано з тим, що більшість об'єктів мало змінюється в часі (це можуть бути роки, десятки років).

Проте, важливо прослідити зміну окремих об'єктів в часі, тому до складу атрибутів вводиться час (даний момент розглядався більш детально у попередніх працях). Основні дані, необхідні для створення баз метаданих, містяться в різного роду технічній документації, тому вони вимагають пошуку, впорядкування і формалізації. Процес створення метаданих включає наступні операції: вибір характеристик, розробка структури бази даних, формалізація документів і занесення даних на носій.

Метадані володіють рядом специфічних особливостей [5]:

- одноразове введення інформації при первинному завантаженні метаданих з подальшим внесенням змін і багатократне її використання протягом достатньо тривалого проміжку часу;

- відносно мала активність оновлення інформації, як по частоті, так і за об'ємом коректування;

- відділення в часі, за джерелами, виконавцям процесу оновлення від процесу використання, як наслідок, необхідність забезпечення сумісності інформації, що підлягає редагуванню;

- наявність чітких ознак класифікації і групування інформації;

- необхідність централізації загальних відомостей про дані і децентралізацію локальних, детальних відомостей про дані.

Склад і точність опису характеристик об'єктів метаданих залежить від рівня управління даними (локальний, регіональний, національний, міжнародний), масштабу системи, етапу обробки даних.

Метадані повинні містити форматовану інформацію про інформаційні ресурси: імена параметрів, коди, властивості, формати, адреси зберігання і т.п. - все, що необхідне інформаційним технологіям, щоб знайти інформаційні ресурси і правильно їх обробити. Для зберігання і забезпечення доступу до метаданих необхідно розробити схему бази метаданих, що включає опис структур таблиць з метаданими і дозволяюча без дублювання організувати зберігання метаданих.

Задачі, поставлені відносно метаданих:

- зберігати всі об'єкти предметної області в єдиному наборі таблиць БД;

- динамічно описувати зміну структури зберігання залежно від часу;

- підтримувати ієрархічну структуру типів об'єктів;

- задавати складні характеристики атрибутів, наприклад, динамічне зберігання;

- відстежувати зв'язки між об'єктами в часі.

Управління таблицями в інформаційному сховищі приводить нас до думки управління метаданими засобами, так званої реляційної алгебри. За допомогою реляційної алгебри можливо застосувати операції над метаданими, можливості роботи з метаінформацією.

Що стосується метаданих інформаційного сховища, то є сенс застосувати наступні операції:

- об'єднання;

- перетин;

- віднімання;

- декартове множення;

Тобто можливо управляти об'єктами із змінним вектором атрибутів метаданих за допомогою реляційної алгебри, використовуючі ці операції. Але треба чітко висліджувати процес переносу стабільної інформації до сховища даних.

Результатом данної роботи є представлення переносу стабільних атрибутів з баз даних різних предметних областей, а також запропонований підхід до керування метаданими, ідеєю якого є застосування операцій реляційної алгебри.

Дослідження в області використання операцій над метаданими інформаційних сховищ становить інтерес у розвитку нових інформаційних технологій. Різноманітність предметних галузей привело нас саме до мети застосування операцій реляційної алгебри над метаданими інформаційних сховищ.

Концептуальні проблеми створення інформаційних сховищ представляють інтерес на різних підприємствах, де виникла необхідність перенесення інформації в інформаційне сховище. Проблема актуальна на сьогоднішній момент не тільки на місцевому і обласному рівні, але і на регіональному в масштабах країни.

#### Список літератури

1. Чернишова Т.В. Єдина методика створення інформаційного сховища // Вісник соціально-економічних досліджень. – 2004. – №18 . – С. 404-408.
2. Шлеер С., Меллор С. Объектно-ориентированный анализ: моделирование мира в состояниях. - К.: Диалектика, 1993.
3. Малахов С.В. Вопросы организации иерархических информационных хранилищ // Перспективи. – 1997. – № 1. – С.122 – 123.
4. Гуляев А.И. Временные ряды в динамических базах данных. – М.: Радио и связь, 1989. – 128с.
5. Meta Data & knowledge Management: Meta Data Repository Myths, David Marco, DM Review, 2002