

ЄДИНА МЕТОДИКА СТВОРЕННЯ ІНФОРМАЦІЙНОГО СХОВИЩА

Розглядаються проблеми створення сховищ даних, єдиної методики проектування та побудови інформаційних сховищ. Надані рекомендації щодо концептуальних проблем створення інформаційних сховищ.

Consideration is given to the problems of data warehouses creation, united method of planning and construction of data warehouses. Following recommendations as to the conceptual problems of creation of data warehouses are presented.

Сьогодні відбувається бурхливий розвиток прогресивних інформаційних технологій. Усе частіше підприємства і фірми зіштовхуються з проблемою вибору тієї системи, що буде вирішувати завдання аналізу на підставі накопиченої інформації і бути системою підтримки прийняття рішень директора чи керівника.

Наприкінці 80-х років багато розроблювачів вважали, що інформаційні проблеми організацій відразу ж вирішуються, як тільки з'явиться деяка стратегічна система управління базами даних, що зможе охопити всі підрозділи служби, а дані зосередяться в рамках єдиної системи і, отже, будуть легко доступні. В даний час існують приклади, де в організаціях експлуатується кілька баз даних і для успішної роботи з інформацією вимагаються не просто різнотипні бази даних, а різні покоління систем управління базами даних. Стала очевидною необхідність "ізолювати" бізнес компанії, правильніше, людей, що займаються цим бізнесом, від технологічних особливостей баз даних, надати користувачам єдиний погляд на корпоративну інформацію незалежно від того, де вона фізично зберігається. Це стимулювало появу визначеного ряду інформаційних систем, зокрема, технології інформаційних сховищ (сховищ даних). Основна мета інформаційного сховища - створення єдиного логічного представлення даних, що містяться в різнотипних базах даних, чи, іншими словами, єдиної моделі даних.

У більшості випадків для завдань аналізу даних виділяється сховище даних. Великі організації оперують величезною кількістю даних, необхідних для здійснення ефективного управління. Це може бути досягнуте на основі накопичення інформації та можливості використання її для вирішення аналітичних задач, розв'язуваних керівництвом з використанням систем підтримки ухвалення рішення.

Відповідно, виникла проблема створення інформаційного сховища за методикою. Не існує на даний момент єдиної методології проектування і побудови сховища.

Аналіз досліджень в цій області показує, що ця проблема є актуальною на сьогоднішній день, і не була вирішена раніше.

Основним завданням дослідження є розробка такої методики, яку можуть застосовувати при побудові сховища в різних предметних областях.

До сьогоднішнього моменту виникла проблема, яка не була освітлена ні в літературних джерелах, ні в джерелах мережі Internet, що призводить до необхідності вирішення завдання.

Використання інформаційних сховищ як напрямок інформаційних технологій з'явилося зовсім недавно і вже став актуальним. Виникають проблем при побудові сховища.

Варто визначитися, а чому буде використовуватися інформаційне сховище, а не база даних. Відповідно до класичного визначення Б.Інмона, виділені визначені критерії відмінності інформаційного сховища від бази даних [1,с.120]. Після того, як визначається вибір використання інформаційного сховища, а не бази даних, необхідно приступити до проектування даного інформаційного сховища. А виходить, необхідно визначити виміри, атрибути інформаційного сховища.

Тобто при розв'язанні даної проблеми зформулюємо цілі:

- необхідно враховувати, що знадобиться більш детальна інформація, тоді для цих цілей додатково буде використовуватися база даних. Однак база даних – це своєрідне «тимчасове сховище», тобто тут дані можуть змінюватися і віддалятися, що не відповідає визначенню інформаційного сховища.

- для проектування, побудови і підтримки сховища даних необхідно вирішити ряд завдань доступу і вибірки даних, їхньої попередньої обробки, очищення і виправлення від некоректних значень, завантаження даних у сховище у відповідності зі спроектованою моделлю для цілей аналізу і, нарешті, ціла група завдань, пов'язана з побудовою автоматизованого середовища поповнення і керування сховищем.

У зв'язку з цим виникають питання проектування і побудови інформаційних сховищ.

Перш ніж приступити до побудови інформаційного сховища, так само як і до іншого глобального проекту, ми повинні детально продумати всі елементи проектування. У результаті роботи інформаційно-аналітичної системи у випадку необдуманих проблем можуть виникнути труднощі, наслідки яких необоротні. Якщо тільки задуматися, що інформаційно-аналітична система міністерства надзвичайних ситуацій дасть помилку і тим самим вплине на ухвалення рішення людини, що відповідає за результат події, скільки доль можуть бути травмовані, не говорячи вже про людські втрати, фізичний стан цих людей. Переваги інформаційного сховища над базами даних підштовхують нас саме до проектування і побудови інформаційного сховища.

Найпершою проблемою є проблема проектування. Не можна починати технічну розробку, не маючи ретельно проробленого проекту. Якщо починати з вирішення найбільш очевидних завдань, не звертаючи уваги на потенційно існуючі, то така система буде безупинно знаходитися в стадії розробки і переробки.

Однією зі стадій проектування інформаційного сховища повинний бути аналіз вимог компанії. Для цього на основі експертних запитів необхідно виявити всі актуальні і потенційні потреби компанії, що повинні задовольнятися спроектованою інформаційною системою, зрозуміти які потоки даних існують усередині компанії, оцінити обсяги інформації, що повинні підтримуватися й оброблятися інформаційною системою. Для того, щоб інформаційне сховище відповідало призначеним завданням і цілям, воно повинно бути організоване і структуроване.

Інформаційне сховище може бути різного масштабу. Так, якщо це комерційна фірма, інформаційне сховище може бути використане в головному офісі і призначатися тільки для аналізу і управління керівництвом цієї фірми. А може бути і так, що

інформаційне сховище буде використано в різних дочірніх підприємствах цієї фірми (як у межах одного міста чи країни, так і за їхніми межами). Варто враховувати, що виходячи з того, на який масштаб орієнтується інформаційне сховище, воно може відрізнятися не тільки ознаками, атрибутами, вимірами, але також структурою, рівнями.

Розглянемо приклад, що підтверджує необхідність розробки єдиної методології для створення інформаційних сховищ.

Найбільш важливі рішення, із прийняттям яких зіштовхуються люди, визначаються тим, який рівень ризику прийнятний, як знизити дію ризиків і можливі втрати, що неприпустимі. Поряд з різними способами управління існує можливість розгляду управління ризиками як елемент системи підтримки ухвалення рішення. Успіх у управлінні буде забезпечений, якщо буде налагоджений своєчасний доступ до максимально повного й одночасно з цим достовірної інформації, що безпосередньо позначається на швидкості і якості прийнятих рішень. На підставі систем підтримки прийняття рішень дані, застосовувані для аналізу, необхідно виділити в окремі бази даних – сховища чи дані інформаційні сховища. Тут дуже важливо правильно змодельовати сховище даних, орієнтуючись на предметну область. Однією з областей, наприклад, є управління ризиками.

У результаті збору і збереження даних здійснюється побудова моделі інформаційного сховища, застосованого для аналізу.

Моделювання предметної області вкрай важливо для таких застосувань як розробка інформаційного сховища - у сховищі даних звичайно концентруються різноманітні логіко-математичні моделі і структури даних. В інформаційному сховищі один раз завантажені дані теоретично ніколи не міняються. Стосовно них можливі тільки дві операції: початкове завантаження і читання (доступ). В інформаційному сховищі зберігається деяка підсумкова інформація, що у базах даних взагалі відсутня, тому що під час завантаження в інформаційному сховищі запису сортуються, очищаються від зайвої інформації і приводяться до єдиного формату. Після обробки дані представлені інакше.

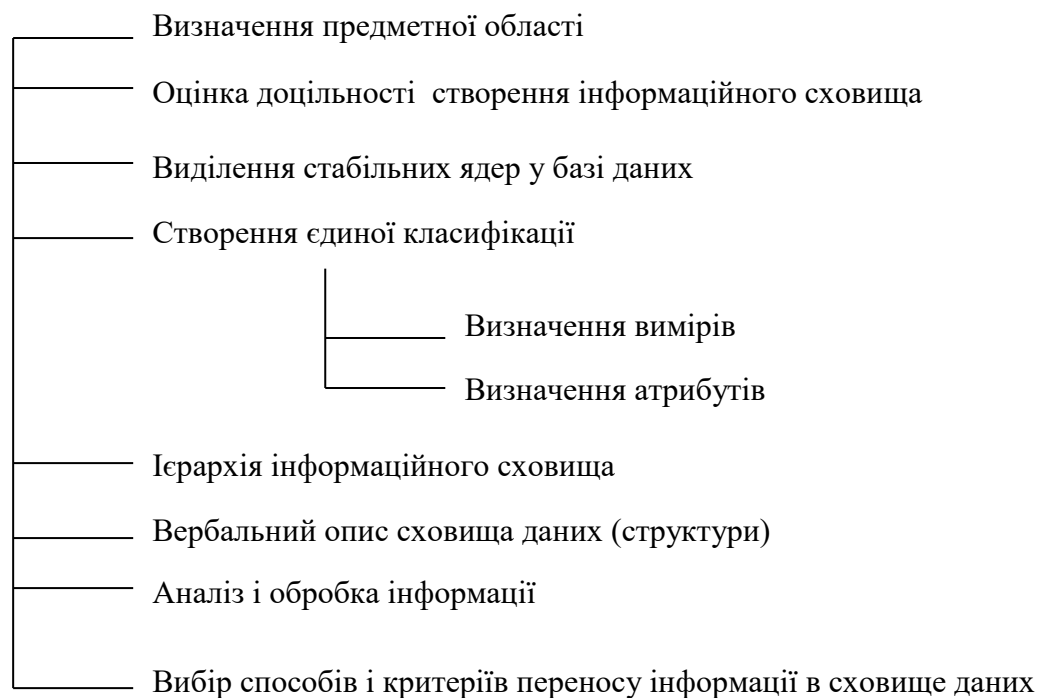
Структура об'єктів, предметної області визначають структуру інформаційного сховища. Знання існують як відношення між даними, як іменованій зв'язок окремих записів у сховище даних. При цьому необхідно, щоб при зміні структури предметної області цей зв'язок міг легко змінюватися і за адресацією зв'язків даних і за типами зв'язків між ними. Ми повинні змінювати поточну структуру сховища даних для фіксації знання, як поточного відношення між даними. Аналіз таких зв'язків подає інформацію, не представлену в явному вигляді - логічні інтелектуальні висновки.

От це і не дозволяє скільки-небудь ефективно робити існуючі інформаційні сховища. Їхня структура є предметом початкового вибору проектувальника. Усі сучасні сховища не можуть ефективно змінювати свою структуру. Стабільна структура зв'язків скорочує життєвий цикл бази, виключає звертання до неї із заздалегідь не передбаченими запитами і призведе, у кінцевому рахунку, до створення обмежених фактографічних сховищ даних, як би їх і не намагалися в рекламних цілях представляти інформаційними сховищами.

Інформація в базах даних створюється групами авторів під одним центром управління, але так буває не завжди. Коли використовується інформація з додатків, створених різними авторами, що не погодять один з одним свої розробки, виникає необхідність приведення даних з різних об'єктів даних до загального вигляду.

Раніше починалася спроба концептуального моделювання об'єктів інформаційних технологій, що авторами відносяться до класу баз даних [2,с. 77-81]. При цьому формулювалися, в основному, концептуальні припущення, що дозволяли моделювати семантику баз даних. Припущення, сформульовані різними авторами, дуже відрізняються, але в той же час у більшості визнається, що будь-які сучасні бази даних повинні задовольняти визначеному загальновизнаному набору властивостей [3,с.122]. Так як інформація з баз даних переноситься в сховище даних, єдине представлення в авторів повинне бути й у відношенні структури інформаційного сховища.

На підставі усього вище викладеного очевидно, що в даний час виникла задача створення єдиної методології проектування і побудови інформаційних сховищ, тому що не існує узгодження створення інформаційних сховищ у різних областях діяльності. На мал.1 пропонується один з варіантів вирішення даного питання.



Мал.1 Варіант методології створення інформаційного сховища.

Даний малюнок представляє рішення проблеми єдиної методики створення інформаційних сховищ. Перше, з чого починається проектування будь-якої бази даних, у конкретному випадку сховища даних, є визначення предметної області. При чому необхідно упевнитися в доцільності побудови інформаційного сховища, а не бази даних, чи задовольняють вимогам створення сховища даних цілі, завдання і функції компанії.

Для однієї і тієї ж предметної області може бути побудоване різноманіття інформаційних моделей. Вони можуть відрізнятися використанням математичним апаратом, структурою, схемами відносин, атрибутами, відносинами й іншими характеристиками. Вибір конкретної інформаційної моделі визначається припущеннями, сформульованими як на етапі проектування сховища даних для виділеної предметної області, так і на етапі розвитку. У процесі рішення завдань можуть виникати нові класи завдань, що змінюють концептуальні припущення.

Бази даних є основними компонентами інформаційних моделей предметних областей. Інформаційні моделі відбивають концептуальну сутність об'єктів предметної області [4,с.175]. Одержання нової інформації про предметну область приводить до появи нових записів, появою нових результатів застосування. З даних визначеної моделі можуть бути виділені залежності між атрибутами, записами.

Наступним етапом при створенні інформаційного сховища є виділення стабільних ядер у базах даних. У базі даних можуть міститися не тільки динамічні дані, але і статичні — постійна незмінна інформація, якій властива ознака стабільності. Статичні дані, що переносяться в інформаційне сховище, характеризуються тим, що вони зберігаються там стільки часу, скільки буде існувати саме сховище [5,с.128]. Характер представлення залежностей визначається інструментальними засобами, розроблювальними на рівні додатків бази даних. Незалежно від форми представлення даних, передбачається, що розвинута база даних містить основний базис даних інформаційної моделі предметної області. На етапі переходу бази даних у стаціонарний стан одержуємо для об'єкта в предметній області закінчений інформаційний опис. У цьому полягає одне з основних завдань створення баз даних.

Якщо виходити з предметної області ризикових ситуацій в утворенні, то прикладом може бути ступінь ризикової ситуації. У конкретному осередку інформаційного сховища містяться дані про втрати і збитки, фактори, що вплинули на ризикову ситуацію, кількість фінансово порушених (потерпілих) людей і т.д., де ступінь ризикової ситуації — величина постійна, а отже і статична. Ця інформація формує стабільне ядро інформаційного сховища. У зв'язку з цим необхідно визначити критерії існування стабільних ядер у базах даних.

Аналіз і структурування інформації в інформаційному сховищі легко здійснюється на основі створення класифікації і дозволяє розглядати інформаційне сховище як безліч стабільних ядер баз даних. Створення і впровадження єдиної системи термінів і класифікації понять забезпечить підвищення ефективності накопичення знань, взаємодії з іншими підсистемами, що прискорить і здешевить упровадження нових методів і програмних засобів проектування інформаційного сховища.

Для того, щоб інформація була структурована в сховище даних, вона на основі класифікації повинна бути представлена у вигляді гіперкуба. Єдина класифікація призведе до вирішення проблеми визначення атрибутів і вимірів, а також багатомірність і багаторівневність сховища даних

Ієрархія інформаційного сховища здійснюється на основі створеної класифікації. Залежно від цього зважається багатомірність і багаторівневність інформаційного сховища.

Проектування сховища даних закінчується вербальним описом структури. Після цього приступають до аналізу й обробки інформації, що буде перенесена в інформаційне сховище.

Вибір способів і критеріїв переносу інформації з баз даних в інформаційне сховище є заключним етапом при створенні сховища даних.

Створення єдиної методології побудови інформаційних сховищ становить інтерес у розвитку нових інформаційних технологій. Різноманітність предметних галузей привело нас саме до створення єдиної методики побудови інформаційних сховищ. Проблема актуальна на сьогоднішній момент не тільки на місцевому і обласному рівні, але і на регіональному в масштабах країни.

Список літератури

1. Корнеев В.В., Гареев А.Ф., Васютин С.В. Базы данных. Интеллектуальная обработка информации. – М.: Ноледж, 2000. – 120 с.
2. Гриф М.Г. Модель представления знаний для проектирования процессов и систем / Гриф М.Г., Козак Д.А. // Управляющие системы и машины. – 1995. – №3. – С. 77-81.
3. Малахов Є.В. Вопросы организации иерархических информационных хранилищ // Перспективи. – 1997. – № 1. – С. 122 – 123.
4. Шлеер С., Меллор С. Объектно-ориентированный анализ: моделирование мира в состояниях. – К.: Диалектика, 1993. – 175 с.
5. Гуляев А.И. Временные ряды в динамических базах данных. – М.: Радио и связь, 1989. – 128с.