

УДК 330.1(075.8)

О.В. ТЕЛЕПНЄВ

Харківський національний університет ім. В.Н. Каразіна, Україна

СТРУКТУРА НАКОПИЧУВАЧА ДАНИХ ПОДІЙНО-ОРІЄНТОВАНОЇ ІНФОРМАЦІЙНО-АНАЛІТИЧНОЇ СИСТЕМИ ПІДТРИМКИ ЖИТТЄВОГО ЦИКЛУ СКЛАДНИХ ВИРОБІВ

Поставлено задачу розробки структури накопичувача даних подійно-орієнтованої інформаційно-аналітичної системи підтримки життєвого циклу складних виробів (ІАС ЖЦ). Розглянуто загальні теоретичні основи технології створення й застосування накопичувачів даних (Data Warehouse). Запропонована структура накопичувача даних ІАС ЖЦ та окреслені особливості її елементів.

накопичувач даних (Data Warehouse), метадані, багатовимірне моделювання (Dimensional Modeling), життєвий цикл продукту, подія, подійно-орієнтована інформаційно-аналітична система підтримки життєвого циклу складних виробів

Вступ

Подійно-орієнтована інформаційно-аналітична система підтримки життєвого циклу складних виробів (ІАС ЖЦ) допомагає вирішувати багато задач, що встають перед співробітниками організацій, які беруть участь у процесах життєвого циклу виробів. Вона дозволяє консолідувати різноманітні дані про виріб на основі подійно-орієнтованого підходу. Ці дані розміщуються у накопичувачі даних системи, що є однією з її головних частин. На даний час не визначена структура накопичувача даних для ІАС ЖЦ. Цьому питанню й *присвячена* дана робота.

Дослідження проводиться на межі наступних предметних областей: інформаційна підтримка життєвого циклу виробів (CALS – Continuous Acquisition and Life Cycle Support – безперервні збирання (інформації) й підтримка життєвого циклу), створення накопичувачів даних (Data Warehouse) для підтримки аналітичних задач.

У межах концепції CALS розробляються системи класу PDM (Product Data Management – управління даними про продукт) та їхній розвиток – системи PLM (Product Life Cycle Management – управління життєвим циклом продуктів). Розробники цих систем прагнуть об'єднати у єдиному інформаційному

просторі усі дані про вироби. PDM-системи (PLM-системи) забезпечують управління накопиченням, обробкою документів (даних) різного формату, що продукуються під час життєвого циклу продуктів. Проте орієнтованість цих систем на роботу саме з документами, різномірність самих даних значно ускладнюють аналіз поточного стану продуктів.

Для вирішення таких задач необхідне створення інформаційно-аналітичної системи, що з самого початку орієнтована на аналіз даних про події. Важливим елементом цієї системи є накопичувач даних, в якому зберігається така інформація.

Накопичувач даних (Data Warehouse) є основою інформаційно-аналітичної системи. Тому значну увагу слід приділяти його структурі (метаданим) та наповненню (даним).

При створенні баз даних для аналітичних систем проводиться денормалізація. Наприклад, сутності “Персона”, “Виріб”, “Об’єкт” створюваної бази даних не відповідають другій нормальній формі. Більш того, оскільки у перелік задач системи не входить аналіз інформації окремо щодо імені чи по батькові персони, то у одному атрибуті “Персони” об’єднуються ім’я, по батькові та прізвище, чим порушуються вимоги першої нормальної форми. Такі “вільності” стали можливими завдяки автома-

тичному процесу заповнення накопичувача даними, що пройшли попередню очистку.

Базовими поняттями багатовимірного моделювання є факти та виміри [1].

Факт – об'єднання пов'язаних даних, що репрезентує елемент бізнесу, бізнес-транзакцію чи подію, що можуть бути застосовані у аналізі бізнесу чи бізнес-процесів. У накопичувачі таблиці фактів є центральними таблицями, що містять дані. Важливим параметром фактів є адитивність, що визначає можливість підсумовування факту вздовж заданого виміру. Факти бувають:

- адитивні – підсумовуються вздовж усіх вимірів на будь-яких рівнях ієрархії;
- напівадитивні – підсумовуються лише вздовж окремих вимірів;
- неадитивні – не можна підсумовувати.

Виміри – набір елементів одного типу, який визначає точку погляду на факт. Зазвичай, елементи виміру мають нечислові значення.

У логічній структурі накопичувача використовуються схеми “зірка”, “сузір'я”, “сніжинка”.

Ідея схеми зірки (star schema) полягає в тому, що є таблиці для кожного виміру, а усі факти містяться у одній таблиці, що індексується множинним ключем, складеним із ключів окремих вимірів. Кожен промінь схеми зірки задає, у термінології Кодда, напрямок консолідації даних за відповідним виміром.

У складних задачах з багаторівневими вимірами має сенс звернутися до розширень схеми зірки – схеми сузір'я (fact constellation schema) і схеми сніжинки (snowflake schema).

У цих випадках окремі таблиці фактів створюються для можливих сполучень рівнів узагальнення різних вимірів. Це дозволяє домогтися кращої продуктивності, але часто призводить до надмірності даних і до значних ускладнень у структурі бази даних, у якій виявляється величезна кількість таблиць фактів.

Якщо багатовимірна модель реалізується у вигляді реляційної бази даних, варто створювати довгі й “вузькі” таблиці фактів і порівняно невеликі та “широкі” таблиці вимірів. Таблиці фактів містять чисельні значення осередків гіперкуба, а інші таблиці визначають утримуючий їх багатомірний базис вимірів. Частина інформації можна одержувати за допомогою динамічної агрегації даних, що розподілені у незіркоподібних нормалізованих структурах, хоча при цьому варто пам'ятати, що запити, які включають агрегацію, при високонормалізованій структурі БД можуть виконуватися досить повільно.

На цей час розроблено загальні теоретичні основи технології Data Warehouse. Проте значний інтерес представляє розробка для предметних областей додатків, що базуються на цій технології.

Задачею статті є розробка структури накопичувача даних подійно-орієнтованої інформаційно-аналітичної системи підтримки життєвого циклу складних продуктів.

Результати досліджень

Предметом розгляду ІАС ЖЦ є “Подія”, що характеризується часом здійснення та меткою адекватності події.

Вона розглядається з наступних точок зору (вимірів):

- “Продукт” – будь-який продукт (деталь, вузол, агрегат, система, документ, програмний продукт), створений чи використаний всередині організації, що потрапив у сферу її ділових інтересів; розгляд подій, пов'язаних з “Продуктом”, має інтерес для користувачів системи;
- “Вид події” – вимір, що містить перелік подій, які можуть траплятися з виробом під час його життєвого циклу та є суттєвими для організації;
- “Об'єкт” – місце, де відбувалася подія;
- “Персона” – особа-виконавець дій, пов'язаних з подією.

До структури бази даних накопичувача також входить допоміжна інформація – коментарі до подій у цифровому або символічному вигляді.

Основною відмінністю структури цього накопичувача від типової для OLAP є відсутність агрегатів, що дозволяє використовувати для проектування універсальні засоби, незважаючи на наявність у PowerDesigner та Allfusion Erwin Data Modeler спеціалізованих засобів для проектування “кубів” [2].

Структура накопичувача представляє собою зрізку (рис. 1).

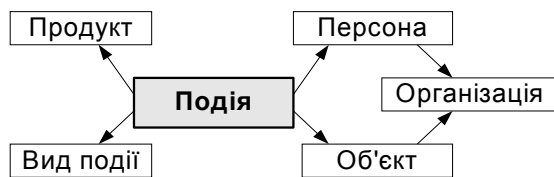


Рис. 1. Структура накопичувача даних на рівні фактів та вимірів

Атрибутами факту “Подія”, окрім часу здійснення та признака актуальності, є зовнішні ключі – посилення на таблиці вимірів (рис. 2).

“Подія”
ІД Події
Дата події
ІД Виду події
ІД Об'єкта
ІД Продукту
ІД Продукту більшого рівня ієрархії
ІД Документа
Актуальність події

Рис. 2. Структура таблиці факту “Подія”

“Виріб більшого рівня ієрархії” вводиться для фіксації подій, що пов'язані з включенням/виключенням зі складу продукту компонентів (“Продуктів”).

У продуктах складної ієрархічної структури “Продукт більшого рівня ієрархії” може, у свою чергу, бути складовою частиною.

Виконання операцій з “Продуктом” регламентується “Документами” (кресленнями, специфікаціями, програмами для приладів з ЧПУ, тощо). Однак ці “Документи” також є й продуктами організації.

Отже, три поля таблиці факту “Подія” (“ІД Продукту”, “ІД Продукту більшого рівня ієрархії”, “ІД Документа”) посиляються на три таблиці, що мають однакову структуру та зміст, який відіграє різні ролі у різних подіях. Випадки, коли вимір відіграє різні ролі, досить типові для багатовимірних моделей [3, 4]. Зважаючи на значний обсяг цих таблиць, доцільно використовувати лише один фізичний екземпляр таблиці “Виріб”. Інші виміри реалізуються з використанням представлень (view).

Важливим елементом структури накопичувача є вимір “Вид події”, що задає перелік подій, доступних для аналізу. При створенні цього переліку слід балансувати між необхідністю більш точної класифікації подій та обміркованими обмеженнями кількості їхніх видів. З одного боку виділення окремих підтипів подій дозволяє більш точно описувати конкретну подію. Однак велика кількість видів подій може ускладнити як введення даних, так і їхній аналіз. Перелік подій є специфічним для кожної окремої сфери діяльності.

У якості первісного ключа для таблиці “Подія” використовується сурогатний ключ – ідентифікатор події (“ІД Події”). Це поле автоматично заповнюється засобами СУБД та використовується при посиленні на окремі екземпляри сутності “Подія”.

Введення до таблиці “Подія” параметра “Актуальність події”, що має значення „істина” або „хибність”, викликано тим, що подія може втрачати свою актуальність внаслідок здійснення іншої події.

Набір атрибутів таблиці “Продукт” повинен бути достатнім для опису і виробів, і документів, і компонентів програмного забезпечення.

Атрибут “Умовне позначення та серійний номер” дозволяє однозначно ідентифікувати об’єкт серед множини, попередньо обмеженої атрибутами “Вид продукту” й “Повна назва продукту”. Цей атрибут не є атомарним (порушення першої нормальної форми), зважаючи на недоцільність розділення цих параметрів у конкретній задачі аналізу.

Умовне позначення є досить природнім параметром для всіх трьох видів “продукту”. Проте не кожний з них має серійний номер. Вирішенням цієї проблеми є генерація сурогатного номеру для “роздрібних” виробів, що найчастіше поступають партіями. Основою для нього може стати дата отримання партії виробів та умовний номер елемента в партії. Для версії документа, або програмного продукту серійним номером може виступати номер версії.

Вимір “Об’єкт” містить інформацію, що у залежності від контексту може бути пов’язана:

- з обробляючим обладнанням (“Обробка виробу” – токарний верстат);
- випробувальним обладнанням (“Перевірка працездатності” – холодильна камера);
- об’єктом у замовника (підстанція, електровагон), на якому змонтовано виріб.

Вимір “Об’єкт” пов’язано з таблицею “Організація”.

Вимір “Персона” включає в себе інформацію про фізичну особу (прізвище, ім’я та по батькові) та службу (“Пост персони” й посилання на “Організацію”).

Тут, як і у вимірі “Об’єкт”, робиться крок у сторону нормалізації, обумовлений спільним використанням інформації про організацію.

Ще одним елементом структури накопичувача є текстові та числові дані про подію. Формально вони не входять до структури гіперкубу, що використовується при багатомірному аналізі подій, але доповнюють відомості, які він містить. Деяку складність у проектування цих елементів структури накопичувача вносить те, що з однією подією можуть бути

пов’язані декілька параметрів. Таких подій, з якими пов’язана велика кількість числових параметрів, дуже багато. Прикладами можуть служити завдання характеристик виробу, результати замірів. Кількість цих параметрів може коливатися від 0 до кількох десятків (наприклад, під час збору метрологічних характеристик). Це, в якійсь мірі, ускладнює моделювання сутності, що містить інформацію про подію.

Для того, щоб позбутися її, створені таблиці “Числова інформація” та “Текстова інформація”, що посилаються на таблицю “Подія” та містять числову, або текстову інформацію, пов’язану з конкретною подією.

Висновок

У цій публікації розглянуто структуру накопичувача даних ІАС ЖЦ. Не менш важливим питанням є наповнення накопичувача даними. Йому буде присвячена наступна робота.

Література

1. Ballard Ch., Herreman D., Schau D., Bell R., Kim E., Valencic A. Data Modeling Techniques for Data Warehousing. – IBM ITSO, 1998. – 199 p.
2. Маклаков С.В. ВРwin и ERwin. CASE-средства для разработки информационных систем. – 2-е изд., испр. и дополн. – М.: ДИАЛОГ-МИФИ, 2001 – 304 с.
3. Kimball R. Data Warehouse Role Models – [Електр. ресурс]. – Режим доступу: <http://www.dbmsmag.com>.
4. Конолли Т., Белт К., Строчан А. Базы данных. Проектирование, реализация и сопровождение. – М.: Изд. дом «Вильямс», 2000. – 1120 с.

Надійшла до редакції 25.04.05

Рецензент: д-р техн. наук, проф. Г.Н. Желткевич, Харківський національний університет ім. В.Н. Каразіна.