

ОРГАНІЗАЦІЯ СХОВИЩА ДАНИХ У СИСТЕМІ СОЦІАЛЬНО-ЕКОНОМІЧНОГО МОНІТОРИНГУ РЕГІОНАЛЬНОГО РОЗВИТКУ

Розглядаються напрями використання концепції сховищ даних у інформаційно-аналітичній системі соціально-економічного моніторингу показників-індикаторів регіонального розвитку. Проаналізовано програмне забезпечення під інструментальні засоби для реалізації поставлених завдань.

Рішення загальної проблеми управління регіональним розвитком в Україні повинно спиратися на повне й адекватне відображення економічної і соціальної динаміки у кожній окремо визначеній сукупності одиниць системи адміністративно-територіального устрою на фоні відповідних середніх показників по країні чи області. Це передбачає необхідність цільового спостереження та порівняльного аналізу значної кількості факторів і показників при розробці та реалізації стратегії й тактики соціально-економічного розвитку регіонів. У той же час існує, як відомо, проблема організації інформаційної бази для підтримки вимірювання ефективності соціально-економічного розвитку регіонів. Аналіз підходів до вирішення цього питання показує, що пріоритетним напрямом є розробка і запровадження інформаційно-аналітичної системи соціально-економічного моніторингу показників-індикаторів розвитку регіонів (ІАС-СЕМППР), одним з основних аспектів якої має бути концепція сховищ даних.

Основна мета цього дослідження полягає у теоретичному узагальненні та обґрунтуванні методологічних і методичних засад побудови сховищ даних для забезпечення соціально-економічного моніторингу регіонального розвитку. Його завдання – обґрунтувати і визначити для інформаційно-аналітичної системи соціально-економічного моніторингу регіонального розвитку методику організації інформаційного забезпечення; зокрема методику побудови сховища даних, за допомогою якого можна б було підвищити рівень та якість інформаційно-аналітичного

забезпечення регіональних органів управління при розв'язуванні прикладних завдань.

Оскільки прийняти правильне управлінське рішення неможливо без необхідної для цього інформації, представленої, зазвичай, в кількісній формі, то в масштабах навіть органу місцевої самоврядування аналіз інформації, потрібної для ухвалення того або іншого рішення може бути досить складною задачею, оскільки крім самого аналізу, потрібно врахувати аспекти, що вимагають наявності механізмів відбору, фільтрації і інших видів попередньої обробки даних. Таким чином, будь-яка система підтримки прийняття рішень, перш за все, повинна володіти засобами відбору і представлення користувачу даних в зручній для сприйняття і аналізу формі. Як правило, одержувані при цьому масиви даних, є багатовимірними. Для опису таких масивів даних вводиться поняття багатовимірних кубів (гіперкубів, метакубів). По осях такого куба розміщуються параметри, а всередині – залежні від них дані. Уздовж кожної осі уявлені різні рівні деталізації даних. Використання такої моделі даних дозволяє підвищити ефективність роботи з ними – генерувати складні запити, створювати звіти, виділяти підмножини даних і т.д. Технологія комплексного багатовимірного аналізу даних і представлення результатів цього аналізу в зручній для використання формі, що отримала назву OLAP (On-Line Analytical Processing) ідеально вирішила б проблему багатомірності даних регіональної статистики. В більшості випадків цілком такого аналізу є сприяння прийняттю рішень, тому такого роду системи часто так і називаються – системи підтримки прийняття рішень, основні положення яких разом із концепцією сховищ даних викладені в роботі [1]. Системи OLAP можуть бути реалізовані різними способами – від найпростіших засобів аналізу даних в офісних додатках до розподілених аналітичних систем, заснованих на серверних продуктах. Проілюструємо ідею OLAP-куба на прикладі організації статистичної інформації необхідної для аналізу ефективності регіонального розвитку (рис. 1):

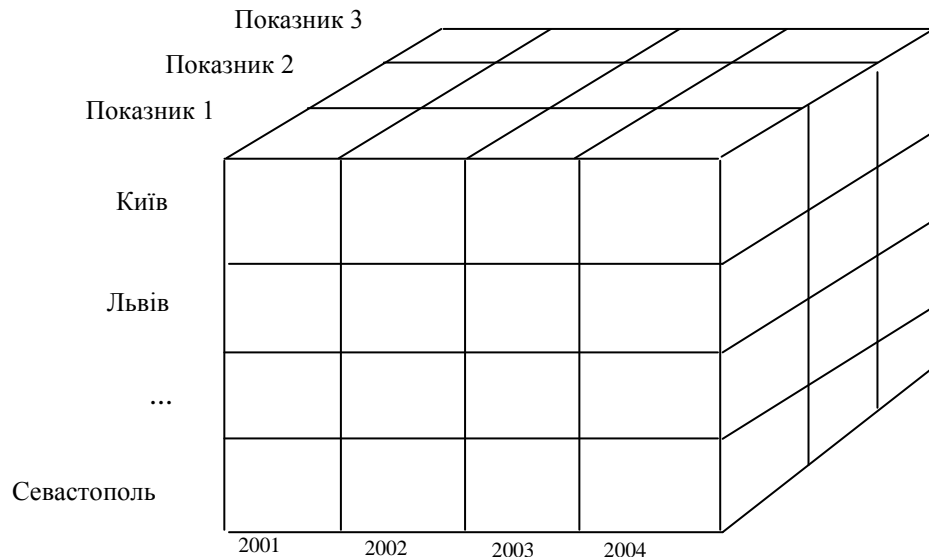


Рис. 1. Приклад організації даних з використанням концепції сховищ даних

Така модель представлення даних дозволяє одержувати потрібну інформацію, робляючи відповідні перетини (зрізи) OLAP-куба. Немає необхідності намагатися виробити геометричну інтерпретацію OLAP-куба з розмірністю більш 3-х вимірів. Дійсно, оскільки людська свідомість пристосована до сприйняття 3-х мірної дійсності, всі інші уявлення є не більше ніж абстракцією. Тим паче, що йдеться не про реальний, а про інформаційний простір, а саме поняття «багатовимірний куб» не що інше як службовий термін, що використовується для опису методу. У принципі, число вимірів може бути будь-яким. Проте, слід зазначити, що задача з великим числом вимірів, по-перше, є трудомісткою з погляду її виконання на ПК і, по-друге, її осмислення і інтерпретація результатів аналітиком може бути утрудненою і навіть приводити до помилкових рішень. Тому з методичної точки зору, складні задачі, що вимагають аналізу даних великої розмірності, слід по можливості зводити до декількох більш простих. Аналогічно, складні таблиці, що містять велику кількість полів і записів, які є важкими для читання, сприйняття і аналізу, можна розбити на декілька більш простих таблиць, що зробить роботу з ними більш зручною.

В нашому ж випадку конкретна постановка організації сховищ даних повинна бути специфікована під основні завдання ІАС-СЕМППР, зокрема такі:

- необхідність аналізу показників ефективності регіонального розвитку;

- можливість аналізу динаміки показників за будь-який часовий період;

- підвищені вимоги до можливостей динамічного аналізу, включаючи нерегламентовані звіти, різні види представлення інформації у вигляді таблиць і діаграм, агрегація даних по будь-яких розрізах, використання методів прогнозування і моделювання, гнучкі засоби формування похідних показників на основі базових, виконання аналізу за сценаріями “що-якщо” і ін.

- отримання показників на основі даних, що нагромаджуються в різних OLTP-системах, при цьому процес аналізу показників не повинен позначатися на продуктивності і без того достатньо завантажених оперативних систем. Особливістю також є те, що використовувані як джерела даних OLTP-системи, в загальному випадку функціонують на різних платформах.

- необхідність порівняння та узгодження даних, отриманих з різних оперативних систем і зовнішніх джерел.

В основі сучасного підходу до інформаційного забезпечення СППР для державних органів місцевої влади лежить ідея інтегрованого сховища даних, що забезпечує єдиний логічний погляд і доступ до інформації, розкиданої по різноманітних оперативних системах організації і поступає із зовнішніх джерел. При цьому істотно, що дані в сховищі мають динамічний характер, тобто забезпечується інтеграція не тільки різноманітних джерел, але і архівних даних, що виникають в процесі функціонування тієї або іншої оперативної системи.

Дані з оперативних систем і зовнішніх джерел піддаються різним перетворенням, узгодженню і завантажуються в централізоване сховище.

Сховище даних містить всю інформацію, необхідну для різноманітних процесів прийняття рішень, але воно не орієнтоване на виконання тих або інших прикладних функцій і з цієї точки зору не є нейтральним по відношенню до додатків. Для інформаційного забезпечення окремих функціонально замкнутих задач використовуються так звані вітрини даних, в які інформація потрапляє або з сховища даних (залежні вітрини) або безпосередньо з джерел даних, проходячи попередні перетворення (незалежні вітрини).

Вітрини даних будуються на основі реляційних або, що більш популярно, багатовимірних СУБД. Річ у тому, що для вирішення більшості задач аналізу корисними є принципи багатовимірної моделі даних і відповідні йому багатовимірні бази даних.

Загальну архітектуру системи СППР, заснованої на принципах сховищ та вітрин даних, можна описати наступною схемою (рис. 2):

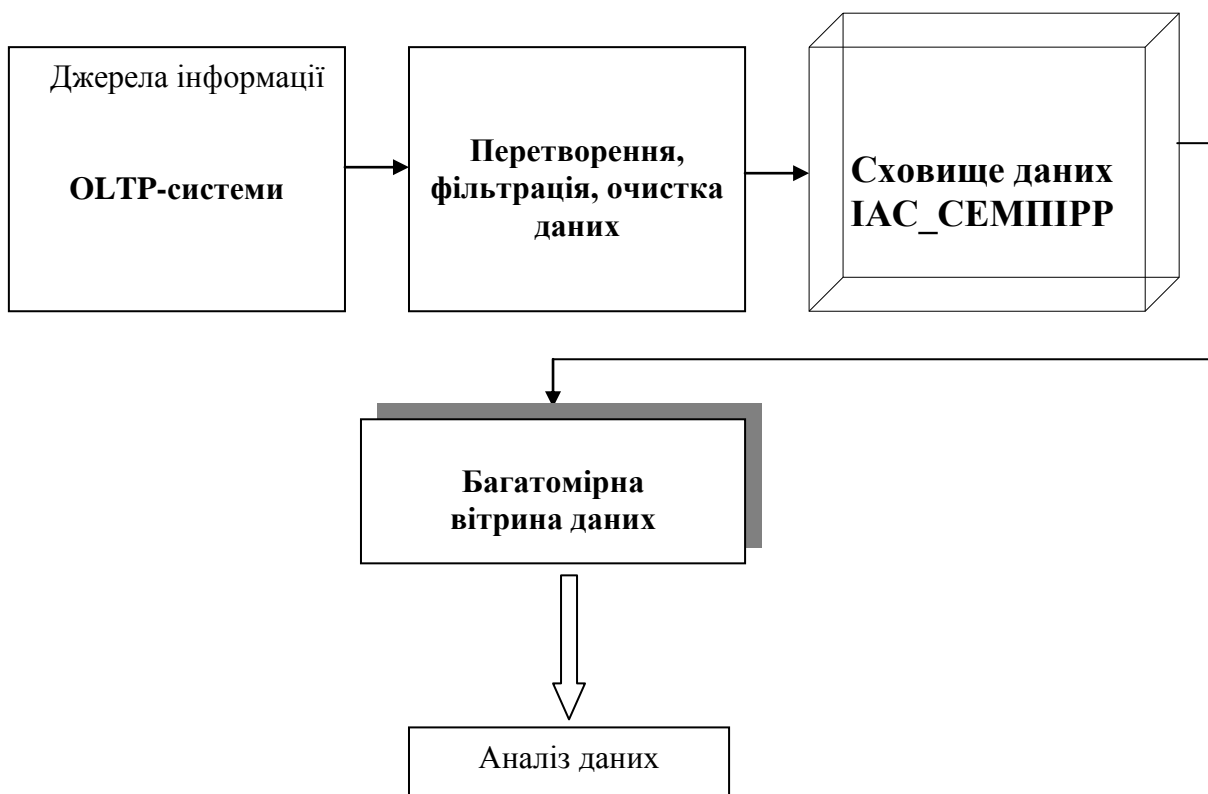


Рис. 2. Структурна схема організації СППР органів регіонального управління на базі

концепцій сховищ та вітрин даних

Щодо аналізу програмного забезпечення, необхідного для реалізації концепції сховищ даних, то на сучасному етапі є такі альтернативи:

- Microsoft SQL Server - являє собою рішення в галузі баз даних і аналізу даних для швидкого створення масштабованих рішень, зокрема Інтернет-технологій і сховищ даних. Microsoft SQL Server забезпечує масштабованість, динамічне керування даними і їх аналіз, підтримку мови XML і протоколу http. Крім того, SQL Server 2000 цілком використовує всі можливості операційної системи Windows 2000.

- IBM INFORMIX є реляційним сервером баз даних, що забезпечує високу продуктивність і масштабованість. Має додатки для інтернет, включаючи об'єктно-реляційні технології, що допомагають користувачам зберігати, шукати і керувати HTML, XML і картографічними даними та зображеннями (що дозволяє використовувати в аналізі регіонального розвитку геоінформаційні технології, накладення результатів на географічну карту тощо), а також забезпечує побудову сховищ даних і систем підтримки прийняття рішень для аналітичних і інтелектуальних додатків з використанням можливостей інтернет. Даний сервер використовує технології рівнобіжної обробки інформації для підтримки високотехнологічних додатків та побудови сховищ даних.

- Oracle – система націлена на недавно сформований ринок інтернет-додатків і відповідає самим строгим вимогам до якості обслуговування. Вона має можливості кластеризації, могутніми й економічними засобами безпеки, виключає втрати даних і дозволяє інтерактивно обмінюватися інформацією. Ядром СУБД є сервер бази даних, що поставляється в одній з чотирьох редакцій у залежності від масштабу інформаційної системи, у рамках якої передбачається його застосування.

- Sybase розроблений як для створення і підтримки традиційних OLTP і DSS додатків, так і для розвитку інтернет/інтранет систем. Однієї з

найбільш сильних сторін Sybase є підтримка розподілених транзакцій і запитів, що дозволяє об'єднати й одержати доступ до інформації, розміщеної на віддалених серверах Sybase та на інших базах даних сторонніх розроблювачів. Це СУБД, оптимізована для аналізу даних на рівні фізичного дизайну. Унікальна архітектура дозволяє обробляти незаплановані аналітичні запити в десятки-сотні разів швидше, ніж традиційні СУБД. При цьому замість надмірності даних у сховищі відбувається їхнє стискування. Sybase Warehouse Studio значно спрощує процес розробки й обслуговування сховища даних.

Якщо ж говорити про вибір певного з проаналізованих продуктів, то це потрібно робити з врахування конкретних умов впровадження системи ІАС-СЕМППР, зокрема врахувати вимоги до інших модулів системи, таких як Data Mining, Web-сервер і т. д. []. Хоча практично усі з наведених інструментальних засобів є досить гнучкими і володіють засобами інтернет-застосувань. Підсумовуючи наведене вище, зазначимо, що запропонована методика побудови та використання сховищ даних може бути застосована при розробці систем підтримки прийняття рішень для регіональних органів управління.

Література

1. W.H. Inmon, J.D. Welch, Katherine L.Glassey. Managing the Data Warehouse. Wiley Computing Publishing, 1997
2. Сахаров А.А. Принципы проектирования и использования многомерных баз данных (на примере Oracle Express Server). СУБД, N3/1996.