

30. *Скороход А. В.* Асимптотические методы в теории стохастических дифференциальных уравнений / А. В. Скороход. — К. : Наукова думка, 1987. — 328 с.

31. *Свешников А. А.* Прикладные методы теории случайных функций / А. А. Свешников. — Л. : Судпромгиз, 1961. — 252 с.

32. *Тихонов В. И.* Статистическая радиотехника / В.И. Тихонов. — М. : Сов. Радио, 1966. — 678 с.

33. *Хасьминский Р. З.* Устойчивость систем дифференциальных уравнений при случайных возмущениях их параметров / Р. З. Хасьминский. — М.: Наука, 1969. — 365 с.

34. *Хинчин А. Я.* Асимптотические законы теории вероятностей / А. Я. Хинчин. — М.: ОНТИ, 1936. — 98 с.

35. *Феллер В.* Введение в теорию вероятностей и ее приложения. В 2-х томах / В. Феллер. — М. : Мир, 1984. — Т. 1. — 528 с.; Т. 2. — 752 с.

36. *Чхун К.* Введение в стохастическое интегрирование: Пер. с англ. / К. Чхун, Р. Уилямс. — М. : Мир, 1987. — 152 с.

37. *Царьков Е. Ф.* Квазилинейные стохастические функционально-дифференциальные уравнения / Е. Ф. Царьков, В. К. Ясинский. — Рига: Ориентир, 1992. — 328 с.

38. *Ясинський В. К.* Задачі стійкості та стабілізації динамічних систем зі скінченною післядією / В. К. Ясинський, Є. В. Ясинський. — К. : Вид-во «ТВІМС», 2005. — 578 с.

Стаття надійшла до редакції 10.05.2011 р.

УДК: 336.71: 004.73

**Н. В. Ситник**, канд. екон. наук,

доц. кафедри ІСЕ,

**Є. А. Труш**, аспірантка кафедри ІСЕ,

ДВНЗ «КНЕУ імені Вадима Гетьмана»

## СХОВИЩЕ ДАНИХ — ДЖЕРЕЛО БАГАТОВИМІРНОГО OLAP-АНАЛІЗУ В БАНКАХ

*АНОТАЦІЯ.* У даній статті розкрито питання побудови сховищ даних для аналізу банківської діяльності. Проаналізовано основні види архітектур побудови сховищ даних та обґрунтована необхідність побудови корпоративного сховища банківської інформації. Розглянуто та проаналізовані концептуальні підходи до побудови моделі сховища даних. Визначено основні види аналізу банківської діяльності на основі сховищ даних та підходи до побудови інформаційно-аналітичної системи банку на основі OLAP.

*КЛЮЧОВІ СЛОВА.* Сховище даних, вітрина даних, OLAP-система, ETL-система, OLTP-система, аналітична задача.

**АННОТАЦИЯ.** В данной статье раскрыты вопросы построения хранилищ данных для анализа банковской деятельности. Проанализированы основные виды архитектур построения хранилищ данных та обоснована необходимость построения корпоративного хранилища банковской информации. Рассмотрены и проанализированы концептуальные подходы к построению модели хранилища данных. Определены основные виды анализа банковской информации на основе хранилищ данных та подходы к построению информационно-аналитической системы банку на основе OLAP.

**КЛЮЧЕВЫЕ СЛОВА.** Хранилище данных, витрина данных, OLAP-система, ETL-система, OLTP-система, аналитическая задача.

**ANNOTATION.** Current article describes the data warehouse (storage) creation for a bank. DW will help to analyze the banking activity. It was analyzed the main data warehouse architectures and was shown the necessity of the DW creation for the banking data. It was reviewed and analyzed conceptual approaches for the building a data warehouse model. The article presents the main types of analysis of the banking activity on the base of the data warehouse and shows the methods for creation of the information-analytical banking system on the OLAP technology base.

**KEY WORDS.** Data warehouse (storage), data mart, OLAP-system, ETL-system, OLTP-system, analytical task.

Розробка сховища даних — це один з інноваційних напрямків у розвитку інформаційних банківських технологій. Актуальність створення сховища даних для банку пояснюється тим, що існуючі автоматизовані банківські системи (АБС) складаються з фронт-офісних та бек-офісних систем і автоматизують лише задачі функціонального контролю та оперативного управління на основі інформації традиційних баз даних. Фронт-офісна частина БІС, що пов'язана з первинним обліком, обслуговуванням клієнтів та формуванням банківських документів, представлена системами оброблення транзакцій (OLTP- OnLine Transactional Processing ). Бек-офісна частина АБС — це наступне оброблення фронт-офісних даних з метою обліку банківських операцій, вона представлена інформаційними системами управління (MIS — Management Information Systems), які формують внутрібанківську звітність та звітність, яку необхідно формувати для НБУ та податкових органів. Тобто існуючі на сьогодні банківські системи, в переважній більшості автоматизують облікові операції, формують необхідну звітність, але не завжди мають засоби реалізації консолідованих аналітичних технологій. Розвиток інформаційних технологій показав, що автоматизація лише задач транзакційного класу є недостатньою з точки зору ефективності управління банківським бізнесом.

У банках існує потреба в оперативному багатоаспектному бізнес-аналізі. Для задоволення цих потреб банки застосовують нову технологію вирішення аналітичних задач, яка дістала назву OLAP (On-Line Analytical Processing). Ця технологія призначена забезпечувати аналітиків динамічним багатовимірним аналізом консолідованих даних. Виконання аналітичних запитів на традиційній базі даних нераціонально, а іноді навіть не можливе. Зручним способом зберігання даних для вирішення оперативних аналітичних задач є різновид баз даних, який носить назву сховище даних (Data Warehouse).

Сховище даних необхідно для зберігання і накопичення різнопланових даних з різних джерел за великі періоди часу, а також для швидкого доступу та пошуку релевантної запитам інформації.

Спочатку сховища даних створювалися в переважній більшості для вирішення задач класу OLAP (On-Line Analytical Processing), але враховуючи стабільність сховищ, значні обсяги накопиченої інформації, сховища даних стали перспективною платформою для інтелектуального аналізу даних (Data Mining). Основною задачею Data Mining є пошук логічних та функціональних закономірностей у накопчених даних, побудова моделей та правил, що можуть пояснити знайдені закономірності, а також можуть бути використаними при прогнозуванні. Тому на сьогодні одним із нових напрямів використання сховищ даних є інтеграція технологій OLAP і Data Mining. У роботах [7, 8] висвітлено основні проблеми такої інтеграції і введено новий термін — багатовимірний інтелектуальний аналіз — OLAP Mining або OLAM.

За визначенням Інмона сховище даних — це предметно-орієнтована, інтегрована, прив'язана до часу та незмінна сукупність даних, призначена для підтримки прийняття рішень [5].

Розрізняють такі види сховищ даних: централізоване сховище даних і кіоски чи вітрини даних. Корпоративне сховище даних чи корпоративна інформаційна фабрика (*Corporate Information Factory* — *CIF*) або більш рання назва корпоративне сховище даних (*Enterprise Data Warehouses* — *EDW*) вміщують інтегровану інформацію, зібрану із певної множини оперативних БД та зовнішніх джерел, яка характеризує всю корпорацію і необхідна для виконання консолідованого аналізу діяльності корпорації у цілому. Такі сховища охоплюють усі багаточисельні напрями діяльності корпорації і використовуються для прийняття як тактичних, так і стратегічних рішень.

Розробка корпоративного сховища даних дуже трудоемкий процес, який може становити від одного до кількох років.

Кіоски чи вітрини даних (*data marts*) це певна підмножина корпоративних даних, які характеризують конкретний аспект діяльності, наприклад, роботу якогось підрозділу. Кіоск може вміщувати як агреговані, так і первинні дані певної предметної області. Кіоск може отримувати дані з корпоративного сховища даних (залежний кіоск) чи бути незалежним і тоді джерелом поповнення його даними будуть оперативні бази даних. Розробка кіоска даних потребує значно менше часу і в середньому займає приблизно 3—4 місяці.

Корпоративне сховище даних та вітрини будуються за подібними принципами і використовують практично однакові технології.

Авторами і першими розробниками сховищ даних було запропоновано дві основні архітектури: так звана корпоративна інформаційна фабрика (Corporate Information Factory, скорочено CIF) Білла Інмона [5] і сховище даних з архітектурою шини (Data Warehouse Bus, скорочено BUS) Ральфа Кімболла (Ralph Kimball) [6].

Сховище даних з архітектурою CIF будується поетапно на базі «спірального» підходу, в основу якого покладено централізоване сховище із залежними вітринами даних. Тому іноді цей підхід називають низхідним, тобто таким при якому створення сховища виконується за принципом «зверху-вниз». Така архітектура розробляється на основі аналізу корпоративних вимог до даних. Визначившись з даними, що підлягають зберіганню в сховищі, їх спочатку вибирають з успадкованих систем-джерел. Якщо ці детальні атомарні дані не є нормалізованими, то їх приводять до 3НФ. Тобто детальні атомарні дані з успадкованих систем джерел зберігаються в реляційній моделі в нормалізованому представленні. На основі детальних атомарних даних будується просторова модель для зберігання узагальнених даних.

Сховище даних з архітектурою шини (*data-mart bus architecture with linked dimensional data marts — BUS*) Ральфа Кімболла (Ralph Kimball) будується на принципах побудови незалежних взаємопов'язаних вітрин. Перша вітрина даних будується для одного бізнес-процесу з використанням вимірів та показників, що в подальшому будуть використовуватись в інших вітринах. Наступні вітрини даних розробляються з використанням попередньо створених вимірів, що в підсумку доз-

воляє створити логічно інтегровану сокупність (шину) вітрин, яка буде виконувати роль корпоративного сховища даних. Створення сховища даних при такому підході починається із відбору даних, необхідних для бізнес-аналізу, з успадкованих систем-джерел і підготовки їх для просторового зберігання. Просторова модель при цьому може зберігати як деталізовані атомарні дані так і агрегати даних. Цей підхід до створення сховищ даних іноді називають висхідним, тобто таким, при якому створення сховища виконується за принципом «знизу-вверх».

Крім цих архітектур, запропонованих основоположниками концепції сховищ даних, на сьогодні використовується велика кількість різних архітектур, які описані в роботах [1, 3].

Згідно [4] формалізовано сховище даних можна описати наступним чином:

$$DW = \langle DB, rf, RF, rm, RM, func \rangle,$$

де  $DB$  — множина баз даних — джерел сховища даних;

$rf$  — множина відношень фактів;

$RF$  — схема  $rf$ ;

$rm$  — множина відношень метаданих;

$RM$  — схема  $rm$ ;

$func$  — множина процедур (рішень).

Важливим при побудові сховища даних банку вирішення питання його концептуальної архітектури. Вибір архітектури побудови сховища для банку обумовлюється організацією оброблення даних у банку: децентралізоване (розподілене) та централізоване оброблення даних.

Є два можливих варіанти децентралізованої технології: повна децентралізація та децентралізація на рівні філії банку. *Повна децентралізація* полягає в тому, що у кожному з організаційних підрозділів банку установлюється окрема незалежна копія АБС, яка працює зі своєю автономною базою даних. Обмін інформацією виконується підсумковими, консолідованими даними по завершенню банківського дня електронними чи паперовими документами. *Децентралізація на рівні філії* полягає в тому, що у кожній з філій установлюється окрема незалежна копія АБС, яка працює зі своєю автономною базою даних, а всі її безбалансові відділення працюють з нею в on-line-режимі. Системи організації інформаційної технології з децентралізацією на рівні філії є найбільш поширеними в Україні.

*Централізована система* працює з єдиною базою даних у головному офісі банку, забезпечуючи режим on-line доступу до даних всіх інших підрозділів банку. В єдиній базі даних зберігається вся нормативно-довідкова інформація, а також дані про всіх клієнтів та про всі операції, що забезпечує відсутність дублювання та протирічливості даних, спрощує процедури їх адміністрування та надає можливість співставлення даних та контролю за операціями, які виконані одним клієнтом у різних підрозділах банку. Тобто централізоване оброблення даних ліквідує «прив'язку» клієнта до певної дирекції чи філії банку, надаючи йому можливість виконання операцій у будь-якому офісі банку. Централізація надає переваги при виконанні операцій та їх бухгалтерському проведенні в режимі реального часу.

На перший погляд, для банків з централізованим обробленням даних підходить централізована архітектура, тобто корпоративне банківське сховище з залежними вітринами, а для банків з децентралізованим обробленням даних — сховище даних з архітектурою шини.

Не дивлячись на те, що централізація банківських інформаційних систем вимагає значних капіталовкладень на придбання централізованої АБС, обладнання, каналів зв'язку, створення нових технологій безперебійної роботи системи, дана тенденція відображає стратегічний напрямок розвитку інформаційних технологій у банківському бізнесу, яка буде переважати в майбутньому. Крім того, враховуючи щоденну консолідацію та централізоване формування банківського балансу і звітності на рівні головного офісу банку, вважаємо за доцільне для банків з децентралізованим обробленням формувати також централізоване корпоративне банківське сховище. Це дозволить централізовано аналізувати діяльність головного банку та його філій. Крім того такий підхід не буде потребувати внесення радикальних змін у сховищі даних при переході банку на централізоване оброблення даних. Аналіз показав, що на сьогодні українські банки використовують тематичні вітрини даних для аналізу окремих аспектів банківської діяльності, тому за цих умов корпоративне банківське сховище краще будувати з архітектурою шини.

Дуже часто в банках використовується кілька облікових OLTP-систем, які виступають джерелом для сховища даних. Кількість таких систем залежить від стану інформатизації банку та політики щодо впровадження IT-технологій.

Взаємозв'язок сховищ даних з OLTP-системами показано на рис. 1.

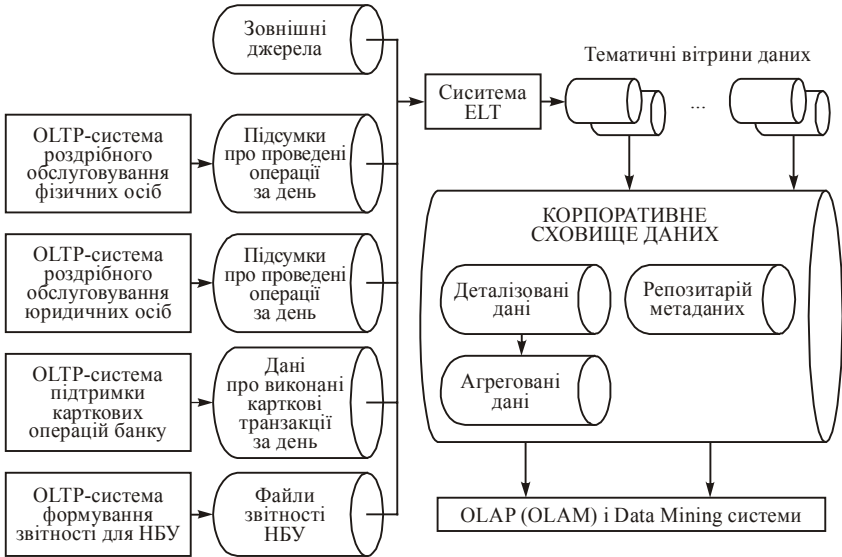


Рис. 1. Взаємозв'язок сховищ даних з OLTP-системами

Сховище даних містить деталізовані дані та агреговані, тобто узагальнені дані. Ступінь деталізації та узагальнення даних, визначається потребами банківських аналітиків. Обов'язковою компонентою сховища даних є репозиторій метаданих, який складається з бази метаданих і браузера для їх перегляду. Метаданими називають бізнес-інформацію, що описує елементи сховища даних, бізнес-правила, процеси та джерела даних.

Ключовою компонентою при побудові сховища даних є ETL-система (Extraction, Transformation, Loading), що виконує процедури відбору, перетворення та завантаження даних до сховища. В першу чергу ETL-системою виконується відбір з баз даних OLTP-систем інформації, що необхідна для бізнес-аналізу. Враховуючи, що дані в сховище надходять з різних джерел, де вони можуть мати різні імена, формати, одиниці вимірювання і способи кодування, тому перш ніж виконати їх завантаження, дані перевіряються на коректність, очищаються від помилок, приводяться до одного єдиного способу кодування, виду та формату, в необхідній мірі узагальнюються і агрегуються. З цього моменту дані представляються користувачеві у вигляді єдиного інформаційного простору, які набагато простіше аналізувати.

Для здійснення того чи іншого виду банківського аналізу за допомогою OLAP-технології доцільно виділити окремі тематичні вітрини даних. У вітринах зберігається необхідна інформація для певного виду аналізу, наприклад для аналізу кредитного портфелю, причому дуже часто у вітрині може бути відсутні дані нижнього рівня, тобто деталізована інформація, а лише агреговані дані.

Важливим моментом створення банківського корпоративного сховища є побудова базової моделі сховища, тобто визначення основних бізнес-цілей для яких створюється сховище, переліку інформаційних об'єктів (сутностей) та зв'язків між ними. У практиці впровадження банківських сховищ відомі дві концепції до побудови моделі сховища даних [2]: перший варіант оснований на Головній книзі банку, коли в основу моделі покладено облікова схема «Рахунки + Клієнти + Проведення», другий варіант — побудова сховища на базі банківських угод (модель угод), коли за основу береться наступна схема «Угода—Рахунки—Операція» чи «Угода—Операція—Проведення». При використанні моделі угод кожній угоді відповідають певні операції (видача, погашення, нарахування), вид (вимога чи зобов'язання), сума, фінансовий інструмент і дата виконання, а також рахунки. Модель угод добре зарекомендувала себе в системі RS-DataHouses для ресурсних угод: кредити, депозити, міжбанківські кредити (залучені та розміщені), кредитні лінії, угоди з цінними паперами і репо, Forex-угоди і SWAP, гарантії і акредитиви. Необхідно зауважити, що концепція сховища, яка побудована на обліковій схемі більш підходить для банків, що використовують вітчизняні АБС, друга концепція на базі угод — для банків, що експлуатують АБС західних розробників.

Дані сховища використовуються для аналізу фінансової діяльності банку, яка охоплює задачі управління активами, пасивами, ліквідністю, дохідністю, співвідношенням між активами і пасивами, власним капіталом, кредитним портфелем, портфелем цінних паперів, ризиком (валютним, відсотковим, операційним, ризиком позабалансових операцій тощо).

Кожна з перерахованих задач аналізу фінансової діяльності характеризується своїми методами здійснення аналізу, і в той же час вони є взаємопов'язаними. В основному, всі аналітичні банківські задачі зводяться до наступних напрямів аналізу:



1. Аналіз бруто-показників банку — величини активів, пасивів, власного капіталу, прибутків, кредитів тощо. Часто оцінка здійснюється на основі співставлення власних показників з аналогічними показниками інших банків.

2. Аналіз ресурсної бази за обсягами, структурою та основними тенденціями розвитку складових. При цьому здійснюється класифікація окремих статей ресурсів банку, розрахунок та вивчення динаміки структурних показників.

3. Аналіз активів банку за обсягами, структурою та основними тенденціями розвитку складових. Активи банку класифікуються за окремими статтями, розраховується та вивчається динаміка структурних показників.

4. Аналіз ліквідності банку на основі розрахунку фінансових коефіцієнтів та їх порівняння з критеріальним рівнем.

5. Аналіз та визначення ступеня збалансованості активів і пасивів за строками та сумами, ГЕП-аналіз, спред-аналіз.

6. Аналіз дохідності банку на основі аналізу даних балансу та звітів про прибутки та збитки. Розраховуються якісні та кількісні показники, що характеризують дохідність банку, ефективність використання активів, структуру доходів та витрат банку.

Крім вище перерахованих видів аналізу, здійснюється аналіз окремих видів діяльності: аналіз кредитного портфелю, портфелю цінних паперів, кредитоспроможності клієнтів, достатності власного капіталу, відсоткової маржі, прибутковості окремих операцій та підрозділів, показників ліквідності тощо.

Розробка сховищ даних потребує їх ретельного проектування та вибору способу представлення даних на логічному рівні. Питання проектування сховищ даних детально викладені в роботі [9]. Враховуючи те, що переважна більшість банківських OLTP-систем реалізована під управлінням реляційних СКБД (Oracle, SQLServer, Sybase, DB2), то для банківських установ підходять традиційні просторові моделі (dimensional model) сховищ даних «зірка» (star schema) та «сніжинка» (snowflake schema), що підтримуються цими системами.

Згідно вимог Basel II, банку необхідно накопичувати статистику не менше ніж за сім років, тому наявність сховища є необхідною умовою функціонування сучасного банку. Західні банки приступили до створення сховищ оперативної інформації, що надає можливість у режимі реального часу відстежувати шахрайські трансакції, проводити моніторинг платіжних операцій для боротьби з відмиванням «брудних» грошей.

На основі сховищ даних у банках розробляються інформаційно-аналітичні системи, основними типовими компонентами яких є:

- банківські OLTP-системи, як основне джерело інформації для сховища даних;
- ETL-засоби відбору, перетворення узгодження та транспортування даних до сховища;
- репозитарій для зберігання моделей даних і метаданих;
- інструментальні засоби для реалізації OLAP -запитів;
- інструментальні засоби для реалізації інтелектуальних запитів.

Для побудови інформаційно-аналітичної системи банку краще використовувати інструменти IBM Cognos і SAP Business Objects, які признані лідерами ринку OLAP. Для бізнес-аналізу (Business Intelligence, BI) і використання в банківських інформаційно-аналітичних системах також можна використовувати розробки наступних компаній: Actuate, Arcplan, Brio, Computer Associates, Crystal, Hummingbird, Hyperion, Informatica, Information Builders, Microsoft, MicroStrategy, Oracle, Peoplesoft, ProClarity, SAP, SAS, Siebel.

### **Література**

1. *Асадуллаев С.* «Архитектуры хранилищ данных — I». — [http://www.ibm.com/developerworks/ru/library/sabir/axd\\_1/index.html](http://www.ibm.com/developerworks/ru/library/sabir/axd_1/index.html).
2. *Михеев А., Савкин Г.* Становление модели RS-DataHouse // Журнал RS-Club. — № 4. — 2006. — С. 64—68.
3. Корпоративные хранилища данных. Планирование, разработка, реализация. Том. 1: Пер. с англ. — М.: Вильямс, 2001
4. *Медиковський М.О., Шаховська Н.Б.* Формалізація операцій над джерелами даних у просторі даних.[http://www.vstu.vinnica.ua/~oeipt/files/index\\_18.files/6.pdf](http://www.vstu.vinnica.ua/~oeipt/files/index_18.files/6.pdf)
5. *W. H. Inmon*, Building the Data Warehouse, QED/Wiley, 1991 .
6. Data-Warehousing-Kimball-Model-vs-Inmon-Model <http://www.scribd.com/doc/15487492/>
7. *Нар J.* OLAP Mining An Integration of OLAP with Data Mining // Proc. IFIP Conf jn Data Semantics Switzerland. — 1997
8. *Parsaye K.* OLAP and Data Mining: Bridging the Gap// Database Programming and Design. — 1997. — № 2.
9. *Ситник Н. В.* Проектування баз і сховищ даних: Навч. посібник. — К.: КНЕУ, 2004. — 348 с.

Стаття надійшла до редакції 12.05.2011 р.