

8.2. Лабораторна робота №2 «Створення і використання OLAP-кубів».

Завдання до лабораторної роботи.

1. Створити OLAP-куб і отримати крос-таблицю для варіанту сховища даних, розробленого в попередній лабораторній роботі відповідно до заданого варіанту.

2. Виконати всі операції з крос-таблицею, відповідно до методики, представленої в розділі «Порядок виконання роботи. Частина 1».

3. Здійснити агрегацію даних відповідно до методики, представленої в розділі «Порядок виконання роботи. Частина 2».

4. Виконати фільтрацію даних крос-таблиці відповідно до методики, представленої в розділі «Порядок виконання роботи. Частина 3».

Теоретична частина.

Куб є одним з поширених методів комплексного багатовимірного аналізу даних, що отримали назву *OLAP (On-Line Analyzing Process)*. У його основі лежить представлення даних у вигляді багатовимірних кубів, названих також *OLAP-кубами* або *гіперкубами*. По осях багатовимірної системи координат відкладаються ті або інші параметри бізнес-процесу, що аналізується. Наприклад, для продажів це може бути товар, регіон, тип покупця. Зазвичай як один з вимірів використовується час. По осях (вимірам) багатовимірної системи координат знаходяться дані, які кількісно

характеризують процес - факти. Це можуть бути об'єми продажів в штуках або в грошовому вираженні, залишки на складі, витрати, суми і тому подібне. Користувач, що аналізує інформацію, може виконувати перетин куба по різних напрямках, отримувати звідні (наприклад, по роках) або, навпаки, детальні (по тижнях) дані і здійснювати інші операції необхідні для ефективного аналізу.

Результатом візуалізації даних, організованих в OLAP-куб є *крос-таблиця* - зручний засіб візуалізації багатовимірних даних і здобуття необхідних форм звітів. Крос-таблиця будується на основі багатовимірного представлення у вигляді OLAP-куба і містить виміри і факти, визначені при побудові куба. Основною особливістю крос-таблиці є те, що її *структура не є жорстко визначеною*. Маніпулюючи за допомогою миші заголовками вимірів, користувач може добитися, аби крос-таблиця виглядала найбільш інформативно. У системі Deductor Studio передбачена можливість будувати на основі крос-таблиці крос-діаграму, основна особливість якої полягає в тому, що вона автоматично перебудовуватиметься відповідно до будь-яких змін крос-таблиці.

Порядок виконання роботи.

ЧАСТИНА 1. Побудова крос-таблиць.

1. Розміщення вимірів. Крос-таблиця є розміщенням багатовимірних даних на площині у вигляді звідної таблиці. Отже, перш, ніж будувати цю таблицю необхідно вказати виміри і факти. Наприклад, виміри – це «Місяць, Найменування товару, Група товару», а факт – «Кількість проданого товару». Виміри можуть бути розміщені в рядках і стовпцях крос-таблиці. При запуску «*Мастер настройки отображения*» спочатку весь список вибраних вимірів відображується у вікні «*Доступные измерения*». Натискуючи кнопки «>» та «<>» справа і знизу від цього вікна можна розмішувати вибрані виміри в рядках і стовпцях таблиці (рис. 8.7).

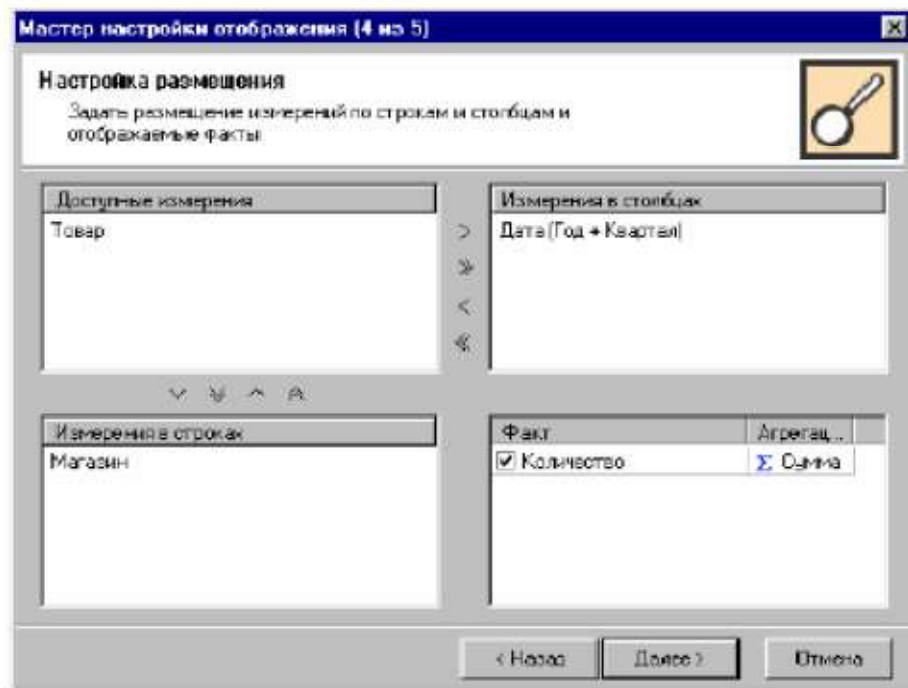


Рис. 8.7. Розміщення вимірів в таблиці.

Побудована крос-таблиця для раніше розглянутого прикладу виглядатиме таким чином (рис. 8.8).

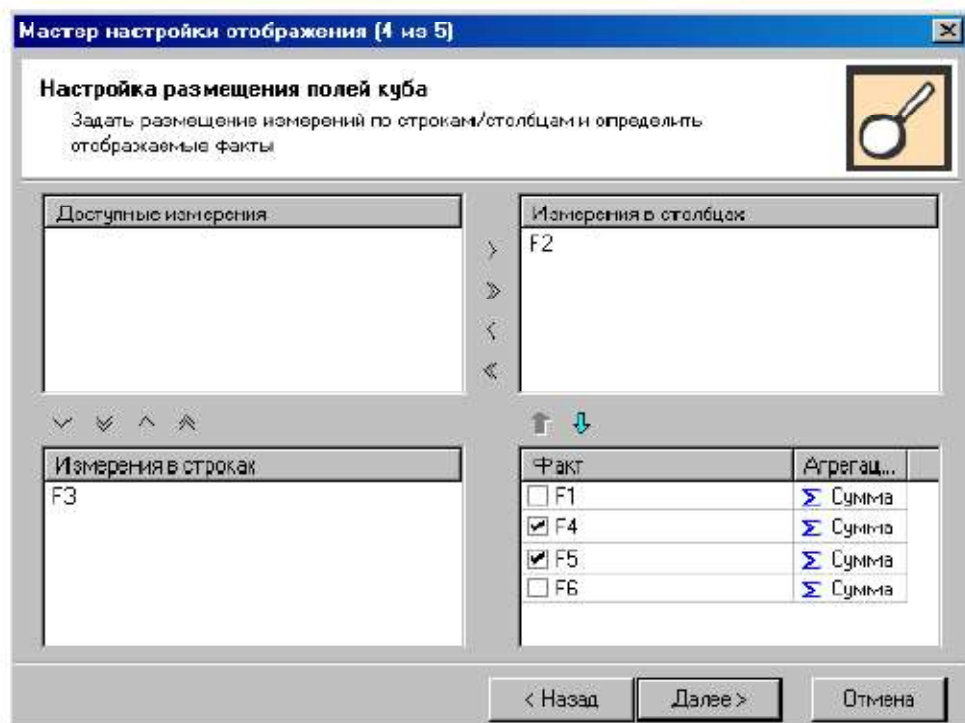


Рис. 8.8. Побудування крос-таблиці.

Тут можна так само вибрати, які факти відображувати в крос-таблиці на пересіченні вимірів і яку функцію застосовувати при їх агрегації (об'єднанні). У даному прикладі факт «Кількість» і «Сума до оплати» підсумовуватиметься. Побудована крос-таблиця для даного прикладу виглядатиме таким чином (рис. 8.9).

		F2		01.03.05		10.03.05		11.03.05	
F3	<>	Σ F4	Σ F5	Σ F4	Σ F5	Σ F4	Σ F5	Σ F4	Σ F5
<		0,00	0,00						
Клиент		0,00	0,00						
ОАО "ОСК"				1 000,00	7 500,00			900,00	9 000,00
ООО "ДСК"						1 500,00	15 000,00		
Итого									


Рис. 8.9. Крос-таблиця для наведеного прикладу.

2. Виміри в крос-таблиці зображаються спеціальними полями. Сині поля показують виміри, що беруть участь в побудові таблиці. Зеленими полями відображуються приховані виміри, що не беруть участь в побудові таблиці. Є можливість перебудувувати таблицю за допомогою миші «на льоту». Зробити це можна, якщо перетягувати поля із заголовками вимірів. Приведемо різні варіанти зміни таблиці в такий спосіб.

- Зробити вимір, що бере участь в побудові таблиці прихованим. Для цього потрібно перетягнути поле із заголовком виміру в рядок з прихованими вимірами або в полі F2, що розкривається, прибрати відмітки. Наприклад, нас цікавлять продажі лише за 12.03.05 і 15.03.05. Результат

		F2		12.03.05		15.03.05	
F3	<>	Σ F4	Σ F5	Σ F4	Σ F5	Σ F4	Σ F5
ОАО "ОСК"		900,00	9 000,00				
ООО "ДСК"				2 000,00	20 000,00		

- Зробити прихований вимір таким, що бере участь в побудові таблиці. Його можна додати до вимірів в рядках або стовпцях за допомогою перетягання мишею.

- Поміняти два виміри місцями. Змінювати розташування вимірів можна, використовуючи *операцію транспонування таблиці*. В результаті транспонування дані, що раніше відображались в рядках, відобразатимуться в стовпцях, а дані в стовпцях перетворяться в рядки. Транспонування у багатьох випадках дозволяє оперативно зробити таблицю зручнішою для сприйняття. Аби застосувати операцію транспонування слід скористатися кнопкою  «Транспонировать таблицу» на панелі інструментів.

У наведених вище прикладах крос-таблиця будується по всіх значеннях вимірів. Проте інколи виникає необхідність побудувати таблицю в розрізі лише деяких значень, наприклад, за першу і третю декади. Включати або виключати значення вимірів в таблицю можна, натискає на трикутник в полі заголовка виміру. Наприклад, якщо натискувати на трикутник в полі заголовка виміру «Дата» відкриється список його значень.

ЧАСТИНА 2. Способи агрегації і відображення фактів.

1. Передбачено декілька способів об'єднання фактів в крос-таблиці:

- сума – обчислюється сума об'єднаних фактів;
- мінімум – серед всіх об'єднаних фактів в таблиці відображається лише мінімальний;
 - максимум - серед всіх об'єднаних фактів в таблиці відображається лише максимальний;
 - середнє – обчислюється середнє значення об'єднаних фактів;
 - кількість – в крос-таблиці відображатиметься кількість об'єднаних фактів.

2. Для зміни способу агрегації фактів потрібно викликати вікно «*Настройка размещения*», натискає кнопку «*Изменение размещения*» на панелі

інструментів. Також існує можливість відображати факти не лише їх значеннями, але і у відсотках по рядках або стовпцях крос-таблиці. Для застосування такого способу відображення потрібно натискувати кнопку «*Отобразить факты как*» на панелі інструментів.

У крос-таблиці факти відображаються з двома знаками після коми і вирівняні по правому краю чарунки. Є можливість змінити форматування елементів таблиці, викликавши вікно «*Настройка форматов отображения данных*». Викликати його можна, натискує кнопку «*Настройка отображения фактов*» (рис. 8.10).

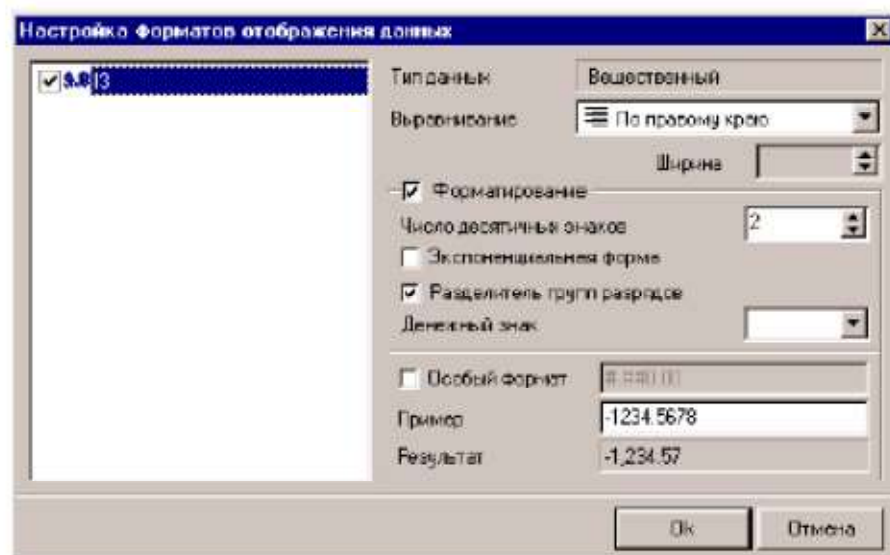


Рис. 8.10. Вікно «*Настройка форматов отображения данных*».

Призначення полів наступне:


- «*Выравнивание*» – визначає вирівнювання значень у чарунках. Може набувати значень: «*По левому краю*», «*По центру*», «*По правому краю*»;
- «*Форматирование*» – застосувати особливе форматування або залишити все як є;
- «*Число десятичных знаков*» – визначає число знаків після коми;
- «*Экспоненциальная форма*» – якщо прапорець встановлений, то число відобразатиметься в експоненціальній формі. Наприклад, число 153.47 виглядатиме як 1.5347E+2;

- «Разделитель групп разрядов» – відображати або не відображати роздільник розрядів. Тобто число може виглядати так 1289 або так 1,289.

ЧАСТИНА 3. Селектор – фільтрація даних крос-таблиці.

1. Селектор є потужним засобом фільтрації даних в крос-таблиці. Фільтрація може виконуватися двома способами:

- по значеннях фактів;
- по значеннях вимірів, шляхом безпосереднього вибору значень із списку, або відбору їх по умові. Фільтрація виконується окремо по кожному виміру.

Аби приступити до роботи з селектором досить на панелі інструментів натискувати на кнопку  «Селектор», після чого буде відкрито вікно селектора. Зліва відображуються всі виміри крос-таблиці і поле «Факты», що означає фільтрацію по фактах. Справа знаходяться елементи:

- «Измерение». Фільтрація передбачає, що в таблиці залишиться лише частина значень деякого виміру. Це поле якраз і задає вимір, значення якого будуть відфільтровані.
- «Факт». У крос-таблиці може міститися один та більше фактів. Фільтрація відбуватиметься по значенням вибраного тут факту.
- «Агрегация». Можна вибрати функцію агрегації, відповідно до якої слід виконати відбір записів. В результаті будуть вибрані лише ті записи, агреговані значення яких задовольняють вибраній умові.
- «Условие». Умова відбору записів по значенням вибраного факту.

Поле умова може приймати наступні значення:

- «Первые N». Значення виміру сортуються в порядку убутання факту і вибираються перші N значень вимірів. Таким чином, можна, наприклад, знаходити лідерів продажів – перші 10 товарів, що найбільше продаються, або перші 5 найбільше вдалих днів.

- *«Последние N»*. Значення виміру сортуються в порядку убавання факту і вибираються останні N значень вимірів. Наприклад, 10 найменш популярних товарів.

- *«Доля от общего»*. Значення виміру сортуються в порядку убавання факту. У цій послідовності вибирається стільки перших значень виміру, щоб вони в сумі давали задану долю від загальної суми. Наприклад, можна відібрати клієнтів, що приносять 80% прибутки, – група А по ABC класифікації.

- *«Диапазон»*. Результатом відбору будуть записи, для яких значення відповідного факту лежить в заданому діапазоні.

- *«Більше»*. Будуть відібрані записи, значення відповідного факту, для яких буде більше вказаного значення.

- *«Меньше»*. Будуть відібрані записи, значення відповідного факту, для яких буде менше вказаного значення.

- *«Равно»*. Будуть відібрані записи, значення відповідного факту, для яких буде рівно вказаному значенню.

Наведемо приклад. Хай нам потрібно визначити товари, що мають найбільший попит. Вхідна крос-таблиця містить 15 товарів (рис. 8.11).

Товар	Σ Количества
Товар 1	92.00
Товар 10	167.00
Товар 11	110.00
Товар 12	133.00
Товар 13	162.00
Товар 14	145.00
Товар 15	54.00
Товар 2	131.00
Товар 3	125.00
Товар 4	120.00
Товар 5	145.00
Товар 6	163.00
Товар 7	193.00
Товар 8	147.00
Товар 9	154.00
Итого	2,070.00

Рис. 8.11. Вхідна крос-таблиця товарів.

Застосуємо до неї селектор (рис. 8.12).

Измерение: Товар

Факт: Количество

Агрегация:

- Сумма
- Минимум
- Максимум
- Среднее
- Количество

Условие: Первые N

Значение: 5

Рис. 8.12. Селекторный анализ данных.

В результаті отримаємо 5 найбільш популярних товарів. Таку вибірку можна отримати по будь-якому факту. У даному прикладі – це кількість. Тому ми отримали товари, що найбільше продавалися. Якщо відфільтрувати по націнці, то отримаємо найбільш прибутковий товар

2. Крос-діаграма. Крос-діаграма є діаграмою заданого типу, побудованою на основі крос-таблиці. Основна відмінність крос-діаграми від звичайної діаграми в тому, що вона однозначно відповідає поточному стану крос-таблиці і при будь-яких її змінах змінюється відповідно. Наведемо приклад крос-діаграми для наступної крос-таблиці (рис. 8.13).

	Дата (Год + Квартал)				
Магазин	2004-Q1	2004-Q2	2004-Q3	2004-Q4	Итого
Магазин 1	384.00	326.00	256.00	187.00	1.133.00
Магазин 2	251.00	255.00	267.00	164.00	937.00
Итого	615.00	581.00	523.00	351.00	2.070.00

Рис. 8.13. Крос-таблица работы магазинов.

Крос-діаграма для неї має наступний вигляд (рис. 8.14).

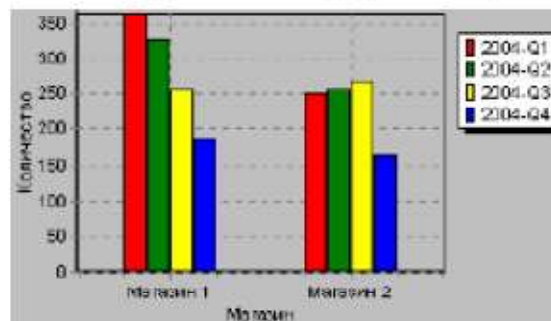


Рис. 8.14. Крос-діаграма роботи магазинів.

На цій діаграмі можна спостерігати поквартальну тенденцію продажів в різних магазинах. До крос-діаграми, так само як і до крос-таблиці можна застосовувати транспонування. Результат транспонування приведеної вище діаграми буде наступний (рис. 8.15).

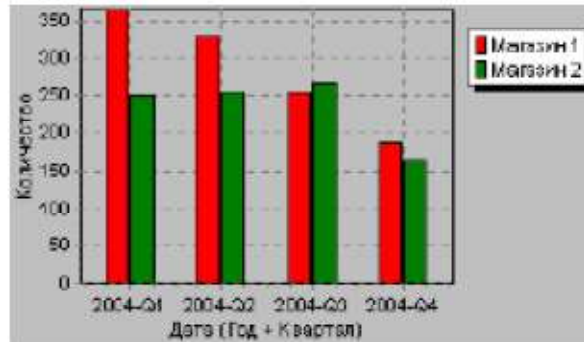






Рис. 8.15. Транспортування крос-діаграми.

Крос-діаграма будується по одному з фактів крос-таблиці. Вибрати факт, що цікавить, можна, натискує кнопку  на панелі інструментів.

При побудові діаграми вводяться обмеження числа серій і числа точок в кожній серії. Дане обмеження викликоне, з одного боку, великими обчислювальними витратами при побудові діаграми, а з іншої - складністю сприйняття великих діаграм. Кнопка  на панелі інструментів матиме синій колір, якщо обмеження не перевищені, і червоний  інакше. Це попереджає про те, що на крос-діаграмі користувач бачить не всю інформацію. Клацання по кнопці виводить вікно «Сведения об ограничениях», в якому представлена інформація:

- скільки серій і точок, фактично відображається на крос-діаграмі;
- максимально можливе число серій і точок (за умовчанням 50 і 100 відповідно) для крос-діаграми;
- кількість серій і точок, мінімально необхідне для того, щоб діаграма відображалася повністю

Таким чином, якщо фактична кількість серій і точок відповідає мінімально необхідному для повного відображення крос-діаграми, то

відображується вся інформація. Крос-діаграма володіє цікавою можливістю – побудовою тренду. У багатьох випадках тренд дозволяє побачити тенденції, які зазвичай приховані із-за великого розкиду значень, наявності відхилень, не типових для процесу, що відображається, і так далі. Лінія тренду отримується шляхом згладжування рядів даних, на основі яких побудована крос-діаграма, за допомогою виділення і відсікання великих відхилень, які в більшості випадків заважають оцінити загальний характер процесу. Кнопка  «Тренд» на панелі інструментів дозволяє включити відображення тренду. Як тільки режим відображення кнопки буде включений, на панелі інструментів вікна крос-діаграми стануть доступні налаштування «Гладкість лінії» і «Міра огрублення». Гладкість лінії визначає міру згладжування вхідного ряду значень, а міра огрублення визначає «масштаб» деталей, що відсіваються. Чим вище значення цих налаштувань, тим більше гладкою буде лінія тренду. Комбінуючи ці налаштування можна добитися найкращого результату. Проте слід врахувати, що для процесів, які швидко змінюються, ці значення мають бути більше, а для повільних процесів - менше. Якщо задати їх дуже великими, то вхідний процес буде згладжений і огрублений в такій мірі, що буде втрачена і корисна інформація, а лінія тренду звородніє в сходинок або пряму лінію.