

РОБОТА ЗІ СХОВИЩЕМ ДАНИХ В АНАЛІТИЧНІЙ ПЛАТФОРМІ DEDUCTOR**Петрова О.О.,****Харківський національний університет міського господарства ім. О.М. Бекетова,****Мірошнікова К.А.,****Харківський національний університет будівництва та архітектури**

***Анотація.** У статті використано технологію сучасного комп'ютерного аналізу даних для створення логічної та фізичної структури сховища даних мережі будівельних магазинів з використанням аналітичної платформи Deductor, в якій сконцентровані сучасні методи вилучення, маніпулювання, аналізу та візуалізації даних.*

***Ключові слова:** сховища даних, інтелектуальний аналіз даних, аналітична платформа, консолідація.*

Вступ

Кожний супермаркет упродовж своєї роботи накопичує велику кількість зібраних даних, в яких знаходиться в сирому вигляді дуже корисна інформація, використання якої забезпечить ефективність та прибутковість бізнесу. Сучасні інструментальні засоби, такі як аналітичні платформи, дозволяють обробляти цифрову інформацію та знаходити в ній укладені знання, які можуть бути використані при ухваленні бізнесових рішень.

Аналітична платформа – це спеціалізоване програмне рішення (або набір рішень), які містять у собі всі інструменти для вилучення закономірностей із сирих даних: засоби консолідації інформації в єдиному джерелі (сховище даних), вилучення, перетворення, трансформації даних, алгоритми Data Mining, засоби візуалізації.

Аналітична платформа Deductor є основою для створення закінчених прикладних рішень у галузі аналізу даних. Реалізовані в Deductor технології дозволяють на базі єдиної архітектури пройти всі етапи побудови аналітичної системи від створення сховища даних до автоматичного підбору моделей і візуалізації отриманих результатів [1–2].

Аналітична платформа Deductor відноситься до Business Intelligence – систем (BI-system) [1]. Основними функціями BI-технології ведення «інтелектуального» бізнесу є: введення даних (завантаження); зберігання даних (сховище даних); аналіз даних (OLAP, Data Mining).

Аналіз публікацій

У книзі «The Data Warehouse Toolkit» Р. Кімболл (R. Kimball) запропонував підхід

до побудови сховищ даних, також відомий як підхід «знизу вгору». Це «Багатовимірне сховище даних» [3].

Другим основним підходом до побудови сховищ даних є підхід Інмона, відомий як підхід «зверху вниз». Це «Реляційне сховище даних». Б. Інмон визначив поняття сховища даних як «предметно орієнтовані, інтегровані, незмінні, що підтримують хронологію, набори даних, які організовані для цілей підтримки управління» [4].

В роботі [5] автор описує підхід до проектування сховищ даних на основі структурованості даних і запитів користувачів, вводить поняття мультибазового сховища даних, будує його математичну модель та пропонує двофазовий алгоритм проектування мультибазових сховищ даних із використанням генетичного алгоритму.

В роботах [6–8] виділено два напрями застосування систем Data Mining: як масового продукту і як інструменту для проведення унікальних досліджень, та чотири основні сфери застосування технології Data Mining: наука, бізнес, роздрібна торгівля і Web-напряма.

Як інструментальний засіб для розробки сховища обрано аналітичну платформу Deductor, в якій реалізовано методики аналізу Data Mining, Knowledge Discovery in Databases та OLAP. Опис технологічної платформи для створення закінчених аналітичних рішень Deductor Academic базується на технічній документації, яка знаходиться у відкритому доступі на офіційному сайті компанії BaseGroup Labs і є для некомерційного навчального використання безкоштовною і призначеною тільки для освітніх цілей [1].

Мета і постановка завдання

Метою дослідження є створення бази даних за допомогою аналітичної платформи Deductor та завантаження до створеної бази даних інформації різного типу.

Для досягнення заданої мети дослідження вирішувалися наступні завдання:

- 1) вивчити предметну область та структуру отримані знання;
- 2) виконати аналіз існуючих структур баз даних;
- 3) розробити структури таблиць для розробки логічної та фізичної структури сховища даних;
- 4) виконати консолідацію інформації в єдиному джерелі для наступного інтелектуального аналізу даних.

У вказаній роботі автори розробили сховище даних мережі будівельних магазинів, яке дозволило відображати неочевидні регулярності в даних, складові так званих прихованих знань. Як об'єкт дослідження обрано початкові дані трьох магазинів будівельної мережі, що розташовані в різних містах України [9].

Розробка сховища даних

В основі концепції сховищ даних лежить інтеграція раніше роз'єднаних деталізованих даних, що містяться в архівах, накопичуваних у традиційних системах транзакційної обробки, що надходять із зовнішніх джерел, в єдину базу даних.

Deductor Warehouse – багатовимірне сховище даних, що акумулює всю необхідну для аналізу предметної області інформацію. Вся інформація у сховищі зберігається у структурах типу «сніжинка», де в центрі розташовані таблиці фактів, а «променями» є виміри,

причому вимір може посилатися на інші виміри.

Така архітектура сховища є найбільш адекватною задачам аналізу даних, тому що аналітик на практиці оперує багатовимірними поняттями. Кожна «сніжинка» називається процесом і описує певну дію, як у нашому прикладі, продаж товару. Як виміри були використані прості списки [1].

Deductor Warehouse реалізує універсальне багатовимірне зберігання, тобто може містити множину процесів з різною кількістю вимірів і фактів. Налаштування процесів, задання вимірів, атрибутів і фактів може здійснюватися за допомогою редактора метаданих, вбудованого в Deductor Studio [1].

Авторами розроблено проект, який зберігає в собі: послідовності обробки даних (сценарії), налаштовані візуалізатори, змінні проекту і службову інформацію. За допомогою Майстрів імпорту, експорту та обробки було сформовано сценарій, який складається з вузлів. Майстер підключень забезпечив створення налаштувань підключень до джерел і приймачів даних. Майстер візуалізації налаштував візуалізатори для конкретного вузла у табличному, графічному та описовому виглядах.

Розробка сховища даних виконувалася у версії Deductor Academic, в якій підтримуються тільки два джерела та приймача даних: Deductor Warehouse та текстові файли з розподільником.

Автори розробили в MS Excel таблиці: Код та найменування товарів (рис. 1); Код та групи товарів (рис. 2); Відділи, які відображають магазини будівельної мережі трьох міст України (рис. 3).

	A	B
1	Група.Код	Найменування групи
2	77	ГАЗОСИЛИКАТНЫЕ БЛОКИ
3	81	ПАЗОГРЕБНЕВЫЕ БЛОКИ
4	20	КИРПИЧ
5	44	ПЛЕНКА
6	84	СЕТКИ МЕТАЛЛИЧЕСКИЕ
7	37	ПАРО-ГИДРОИЗОЛЯЦИОННЫЕ ПЛЕНКИ
8	69	МАСТИКА

Рис. 1. Групи товарів (фрагмент)

I27		
A	B	C
1	Товар.Код	Найменування товару
2	599	Блоки "ЭКО" из ячеистого бетона D-500 (600x200x300мм)
3	464	Блоки "ЭКО" из ячеистого бетона D-500 (600x250x100мм)
4	576	Блоки "ЭКО" из ячеистого бетона D-500 (600x250x125мм)
5	761	Блоки "ЭКО" из ячеистого бетона D-500 (600x250x200мм)
6	586	Блоки "ЭКО" из ячеистого бетона D-500 (600x250x300мм)
7	104	Плита гипсовая пазогребневая полнотелая KNAUF 667*500*100мм
8	408	Плита гипсовая пазогребневая гидрофобизированная полнотелая KNAUF 667*500*100мм
9	938	Плита гипсовая пазогребневая полнотелая Волма 667*500*100мм
10	855	Плита гипсовая пазогребневая гидрофобизированная полнотелая Волма 667*500*100мм
11	724	Плита гипсовая пазогребневая полнотелая Formap 667*500*80мм
12	796	Кирпич одинарный лицевой ВАНИЛЬ Магма 250*120*65мм
13	246	Кирпич одинарный лицевой КРАСНЫЙ Магма 250*120*65мм
14	446	Кирпич одинарный лицевой ШОКОЛАД Магма 250*120*65мм
15	131	Кирпич полуторный лицевой ВАНИЛЬ Магма 250*120*88мм
16	122	Кирпич полуторный лицевой КРАСНЫЙ Магма 250*120*88мм

Рис. 2. Товари (фрагмент)

G21				
A	B	C	D	
1	Код відділу	Назва відділу		
2	1	Епіцентр Харків		
3	2	Епіцентр Київ		
4	3	Епіцентр Полтава		
5				

Рис. 3. Відділи

Рис. 4 являє собою опис процесу продажів у трьох магазинах мережі «Епіцентр» у різних містах країни. При такій структурі сховища можна припустити, що унікальність

точки у просторі визначається сукупністю вимірів: Дата, Код Товару, Код відділу, Час покупки [1].

N24							
A	B	C	D	E	F	G	H
1	Дата	Код відділу	Код товару	Час покупки	Кількість	Ціна	Сума
2	22.05.16	1	464	13	15	65,33	979,95
3	13.03.16	2	761	12	2	134,55	269,1
4	14.03.16	3	408	15	56	966,60	54130
5	15.03.16	2	938	16	89	569,31	50669
6	13.03.16	2	724	19	100	785,65	78565
7	17.03.16	3	796	18	103	23,22	2391,7
8	22.05.16	1	246	13	56	19,36	1084,2
9	19.03.16	2	446	15	12	24,52	294,24

Рис. 4. Продажі (фрагмент)

Таблиці, створені в MS Excel, були перетворені у структуровані текстові файли із роздільниками для зберігання початкових даних із зазначенням призначення, тобто порядку використання поля набору даних, отриманого в результаті імпорту стовпця (поля), у процесі подальшої обробки імпортованих даних.

У навчальній версії як джерело даних було обрано сховище даних Deductor Warehouse з указанням параметрів бази даних, в якій буде створено фізичну та логічну структуру сховища даних. Після створення порожнього сховища, в якому немає жодного об'єкта (процесу, виміру, факту), було відображено спроектовану базу даних будівельної мережі у сховище за допомогою редактора метаданих. На першому кроці були створені виміри та додано до кожного виміру, окрім вимірів Дата і Час, текстові атрибути з подальшою реалізацією ієрархії вимірів, як наведено на рис. 5.

На наступному кроці створення пустого сховища, після того, як всі виміри та посилання було зроблено, виконано процес формування «сніжинки» з назвою «Продажі». Для цього були додані чотири існуючих виміри та два факти (рис. 6).

Після проектування структури і метаданих другий етап було присвячено наповненню сховища даними із зовнішніх структурованих джерел.

Початковими даними для сховища даних були використані чотири текстових файли, з указанням, в який з вимірів буде завантажуватися інформація. На рис. 7 наведено визначення варіантів агрегації атрибутів та фактів.

У результаті виконання всіх дій було:

- створено і наповнено сховище даних;
- написано сценарій завантаження (поповнення) даних із джерела і сховища даних.

Отриманий сценарій завантаження прив'язаний не до даних безпосередньо, а до їх структури, тобто в ньому змодельовано послідовність дій, яку потрібно виконувати для завантаження даних у сховище даних: імена файлів-джерел, відповідність полів тощо. Один раз створений сценарій згодом застосовується для поповнення сховища даними. Для подальшої обробки даних зі створеного сховища було обрано процес «Продажі» з визначеннями, які виміри та атрибути будуть імпортуватися для створення звітів.

Результати вилучення даних з Deductor Warehouse у вигляді таблиці та куба наведено на рис. 8–9.

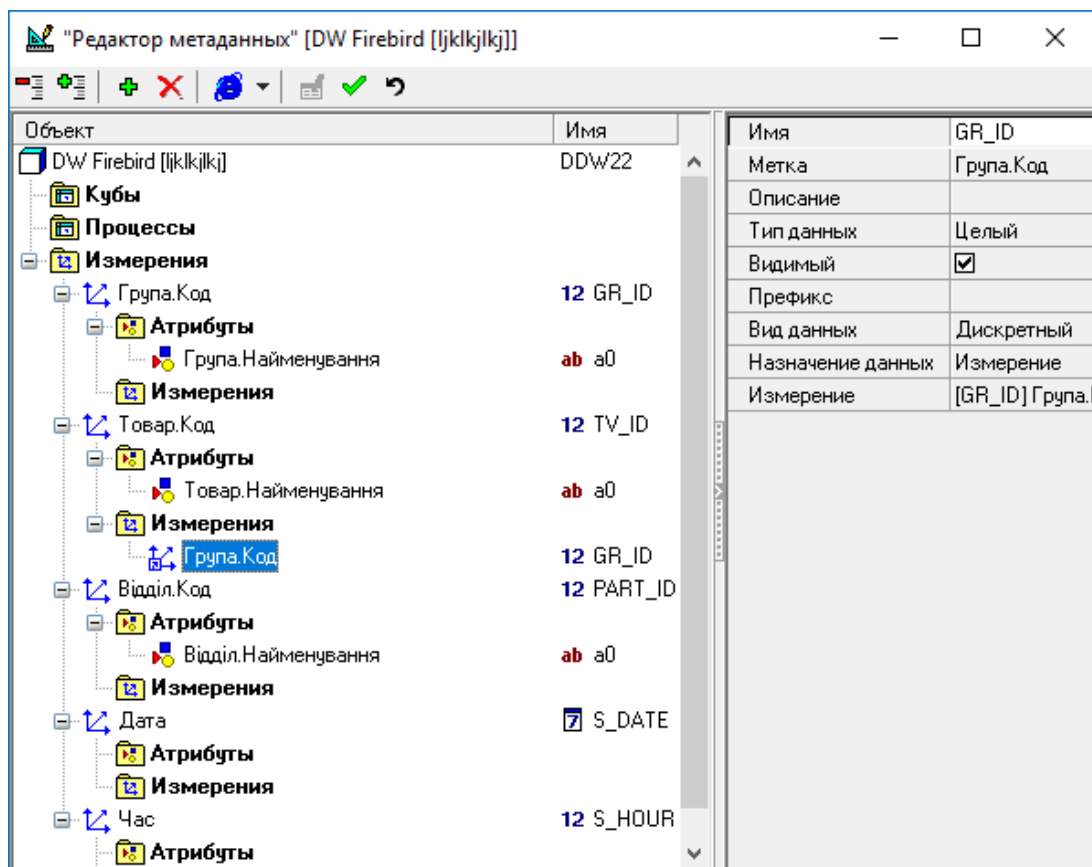


Рис. 5. Реалізація ієрархії вимірів у структурі метаданих сховища

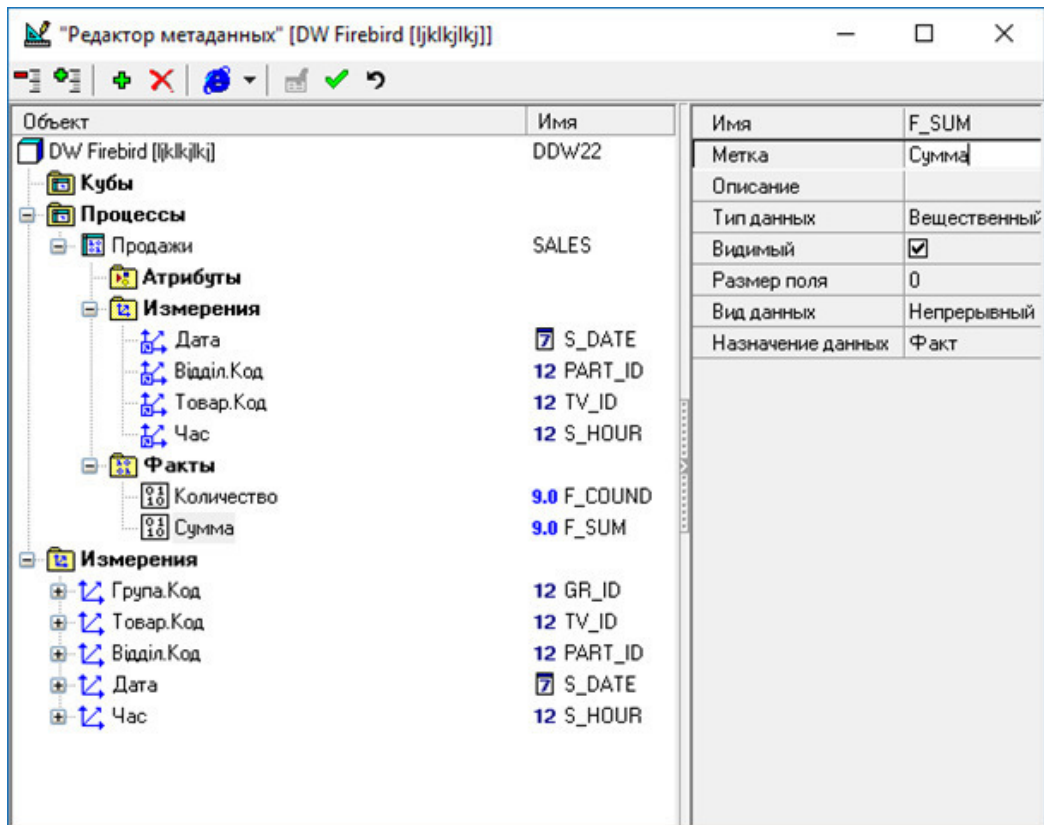


Рис. 6. Процес формування «сніжинки» з назвою «Продажі»

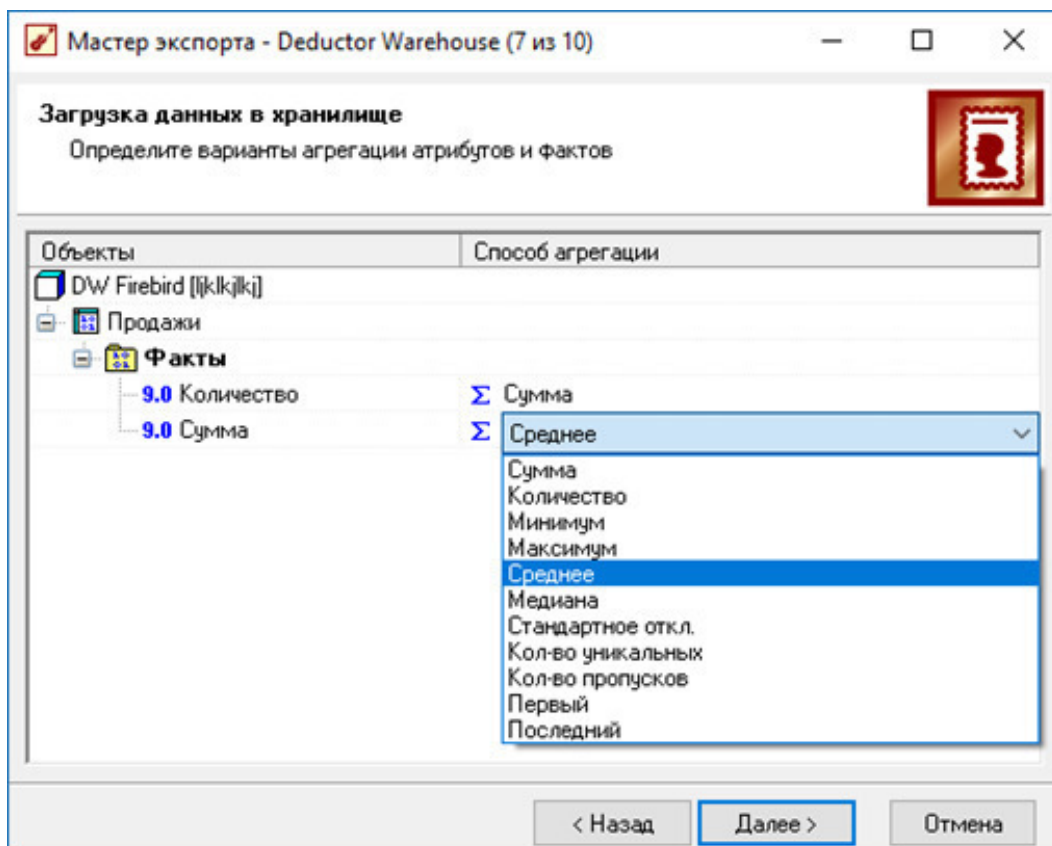


Рис. 7. Визначення варіантів агрегації атрибутів та фактів

Дата	Відділ.Код	Відділ.Код	Товар.Код
		Відділ.Найменування	
13.03.2016	2	Епіцентр Київ	724
28.03.2016	2	Епіцентр Київ	576
31.03.2016	1	Епіцентр Харків	918
02.05.2016	1	Епіцентр Харків	417
15.05.2016	1	Епіцентр Харків	528
22.05.2016	2	Епіцентр Київ	431
22.05.2016	2	Епіцентр Київ	681
05.06.2016	1	Епіцентр Харків	550
13.03.2016	2	Епіцентр Київ	761
15.03.2016	2	Епіцентр Київ	108
21.03.2016	2	Епіцентр Київ	376
06.04.2016	1	Епіцентр Харків	446
26.04.2016	1	Епіцентр Харків	450
20.05.2016	3	Епіцентр Полтава	763
14.06.2016	3	Епіцентр Полтава	257
18.06.2016	3	Епіцентр Полтава	453
19.06.2016	2	Епіцентр Київ	633
14.03.2016	3	Епіцентр Полтава	408
19.03.2016	2	Епіцентр Київ	446
20.03.2016	3	Епіцентр Полтава	122
28.03.2016	2	Епіцентр Київ	796
28.03.2016	3	Епіцентр Полтава	729
10.04.2016	1	Епіцентр Харків	108
22.05.2016	3	Епіцентр Полтава	502

Рис. 8. Подання звіту в табличному вигляді (фрагмент)

Відділ.Код	Товар.Код	Група.Код	Дата																
			13.03.2016		14.03.2016		15.03.2016		17.03.2016		19.03.2016		20.03.2016						
Σ	Количес	#	Количес	Σ	Количес	#	Количес	Σ	Количес	#	Количес	Σ	Количес	#	Количес	Σ	Количес	#	
1																			
2	102,00	2			135,00	2			12,00	1									
3			56,00	1			103,00	1							45,00				
Итого:	102,00	2	56,00	1	135,00	2	103,00	1	12,00	1	45,00								

Рис. 9. Представлення інформації про продажі в різних магазинах будівельної мережі в розрізі міста (відділи) та з розбиттям на дати у вигляді двовимірної таблиці в режимі візуалізації Куб

Висновки

Використано поглиблений метод аналізу даних для створення сховища даних, яке містить відомості, що є корисними у процесі управління мережею магазинів і використання якого надає конкурентні переваги.

Розроблене сховище даних і гнучкі механізми передобробки, очищення, завантаження, візуалізації даних забезпечують швидкість виконання запитів користувача, здійснення аналітичної обробки та отримання аналітичної звітності для керівників та

аналітиків комерційних підприємств у їх повсякденній діяльності.

Користувач може отримати відповідь на поставлене питання і переглянути результати найбільш зручним для нього способом.

Подальше заповнення сховища різнорідними даними надасть переваги в обробці даних та прогнозуванні діяльності мережі будівельних магазинів.

Література

1. BaseGroup Labs. Технологии анализа данных [Электронный ресурс]. – Режим доступа: <http://www.basegroup.ru>.
2. Аналитическая платформа Deductor [Электронный ресурс]. – Режим доступа: <http://deductor.com.ua/>.
3. Kimball Ralph. The data warehouse toolkit: the complete guide to dimensional modeling. / Ralph Kimball. – Wiley, 2002 – 436 p.
4. W. H. Inmon. Corporate Information Factory Components [Электронный ресурс]. / W. H. Inmon – Inmon Data Systems. – Режим доступа: <http://www.inmoncif.com/view/26>.
5. Яцишин А.Ю. Проектирование мультибазовых сховищ даних на основі двофазного алгоритму / А.Ю. Яцишин // Вісник НТУУ. – Вип. 55. – С. 23–27.
6. Чубукова И.А. Data Mining: учебное пособие / И.А. Чубукова. – М.: Интернет–университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. – 382 с.
7. Киселев М. Средства добычи знаний в бизнесе и финансах / Киселев М., Соломатин Е. // Открытые системы. – № 4. – 1997. – С. 41–44.
8. Методы и модели анализа данных OLAP и Data Mining / Барсегян Ф., Куприянов М., Степаненко В., Холод И. – СПб: БХВ–Петербург, 2008. – 456 с.
9. Интернет–магазин «Стройберг» [Электронный ресурс]. – Режим доступа: <http://www.stroyberg.ru/>
10. Новожилова М.В. Практикум. Интеллектуальный анализ данных в аналитичній платформі Deductor: навчальний посібник. / М.В. Новожилова, О.О. Петрова, О.І. Чуб. – Харків: ХНУБА, 2014. – 158 с.
5. Yatsishin A.Yu. Proektuvannya mul'tybazovykh skhovyshch danykh na osnovi dvofaznogo alhorytmu [Designing multi-database data warehouses based on a two-phase algorithm.] / A.Yu. Yatsishin // Visnyk NTUU. – Vyp. 55. – S. 23–27.
6. Chubukova I.A. Data Mining: uchebnoye posobiye [Data Mining: a study guide.] / I.A. Chubukova. – M.: Internet–universitet informatsionnykh tekhnologiy: BINOM: Laboratoriya znaniy, 2006. – 382 s.
7. Kiselev M. Sredstva dobychi znaniy v biznese i finansakh [Means of mining knowledge in business and finance.] / Kiselev M., Solomatine E. // Otkrytyye sistemy. – № 4. – 1997. – S. 41–44.
8. Metody i modeli analiza dannykh OLAP i Data Mining [Methods and models of data analysis OLAP and DataMining.] / [Barseghyan F., Kupriyanov M., Stepanenev V., Holod I.] – SPb: BKHV–Peterburg, 2008. – 456 s.
9. Internet–magazin «Stroyberg». Available at: <http://www.stroyberg.ru/>
10. Novozhilova M.V. Praktikum. Intelektual'nyy analiz danykh v analitychniy platformi Deductor. Navchal'nyy posibnyk [Workshop. Intellectual analysis of data in the analytical platform Deductor.] / M.V. Novozhilova, O.O. Petrova, O.I. Chub. – Kharkiv: KHNUBA, 2014. – 158 s.

Petrova Olena Oleksandrivna, Candidate of Technical Sciences, Associate Professor of the Department of Applied Mathematics and Information Technologies, (098) 849-90-76, Petrifurs@gmail.com, O.M. Beketov National University of Urban Economy in Kharkiv,

Miroshnikova Katerina Antonivna, student, (097) 204-73-07, kate_mir@ukr.net, Kharkiv national university of civil engineering and architecture.

Working with the data warehouses in analytical platform deductor

Petrova O.O., Miroshnikova K.A.

Problem. Each supermarket collects a large amount of data during its work, in which there is a very useful information in a raw form, the use of which will ensure the efficiency and profitability of the business. **Goal.** The purpose of the study is to create a database using the analytical platform Deductor and download various types of information to the created database. **Methodology.** An in-depth data analysis method is used to create a data warehouse, which contains information that is useful in the management of the network of stores and the use of which provides a competitive advantage. The Deductor

References

1. BaseGroup Labs. Tekhnologii analiza dannykh [Technologies of data analysis.] Available at: <http://www.basegroup.ru>.
2. Analiticheskaya platforma Deductor [Analytical platform Deductor.] Available at: <http://deductor.com.ua/>.
3. Kimball Ralph. The data warehouse toolkit: the complete guide to dimensional modeling. / Ralph Kimball. – Wiley, 2002 – 436 p.
4. W. H. Inmon. Corporate Information Factory Components / W. H. Inmon – Inmon Data Systems – Available at: <http://www.inmoncif.com/>

analytical platform, which implements Data Mining, Knowledge Discovery in Databases and OLAP analytics, has been selected as a tool for the development of the data warehouses.

Results. *The result is creating the logical and physical structure of the data warehouses network of building shops using the Deductor tool, which concentrates modern methods of data extraction, manipulation, analysis and visualization. **Originality.** Further filling of the data warehouse with diverse data will provide advantages in data processing and forecasting of the network of construction stores. The user can get the answer to the question and see the results in the most convenient way. **Practical value.** The developed data storage and flexible tools for pre-processing, cleaning, downloading, data visualization provide seamless performance of user queries, analytical processing and obtaining analytical reporting for managers and analysts of commercial enterprises in their daily activities.*

Key words: *data storage, intellectual data analysis, analytical platform, consolidation.*

Работа с хранилищем данных в аналитической платформе Deductor

Аннотация. *В статье использовано технологию современного компьютерного анализа данных для создания логической и физической структуры хранилища данных сети строительных магазинов с использованием аналитической платформы Deductor, в которой сконцентрированы современные методы извлечения, манипулирования, анализа и визуализации данных.*

Ключевые слова: *хранилище данных, интеллектуальный анализ данных, аналитическая платформа, консолидация.*

Петрова Елена Александровна, к.т.н., доцент кафедры прикладной математики и информационных технологий,
(098) 849-90-76, Petrfurs@gmail.com,
Харьковский национальный университет городского хозяйства им. А.Н. Бекетова,

Мирошникова Екатерина Антоновна, студент,
(097) 204-73-07, kate_mir@ukr.net,
Харьковский национальный университет строительства и архитектуры.