

Міністерство освіти і науки України

В.В. Пасічник, Н.Б. Шаховська

Сховища даних

НАВЧАЛЬНИЙ ПОСІБНИК

2-ге видання, стереотипне

СЕРІЯ “КОМП’ЮТИНГ”

За науковою редакцією д.т.н., професора В. В. Пасічника

Видавництво “Магнолія 2006”

Львів

УДК 004

П 23

Відтворення цієї книги або будь-якої її частини заборонено без письмової згоди видавництва. Будь-які спроби порушення авторських прав будуть переслідуватися у судовому порядку.

Гриф надано Міністерством освіти і науки України

Рецензенти:

Г.Г.Цегелик – д.ф.-м.н., професор, завідувач кафедри математичного моделювання соціально-економічних процесів Львівського Національного університету ім. Івана Франка;

М.П. Дивак – д.т.н., професор, завідувач кафедри комп'ютерних наук Тернопільського національного економічного університету;

Б.П. Русин – д.т.н., завідувач відділу ФМІ НАН України.

Пасічник В.В., Шаховська Н.Б.

П 23 Сховища даних: Навчальний посібник. – 2-ге вид., стер. – Львів: “Магнолія 2006” – 492 с.

ISBN 978-966-2025-18-7

ПП “Магнолія 2006”

Навчальний посібник написаний відповідно до освітньо-професійної програми, затвердженої Міністерством освіти і науки України наказом №444 від 07.06.2006 р. з напрямку 8.080405 «Інтелектуальні системи прийняття рішень». Згідно з тематичним планом, навчальний посібник «Сховища даних» включає предмет «Технології сховищ даних». У ньому послідовно викладені теоретичні, методологічні та прикладні аспекти проектування, побудови, застосування та супроводження сховищ даних, які розробляються з метою розв'язання аналітичних задач проблемної області.

Призначений для студентів технічних спеціальностей вищих навчальних закладів. Може бути використаний аспірантами, спеціалістами, викладачами та керівниками підприємств і організацій, які зацікавлені в поглибленні знань з проектування та супроводження сховищ даних.

УДК 332.1 (075.8)

ISBN 978-966-2025-18-7

© В.В. Пасічник, Н.Б. Шаховська
© ПП “Магнолія 2006”
© ФОП Марченко Т. В.

Зміст

Передмова наукового редактора.....	12
Список абревіатур та прийнятих скорочень.....	17
Список позначень	19
Розділ 1 Структурно-логічна схема освітньо-наукового напрямку «Інженерія даних та знань»	22
1.1. Поняття інформації, даних, знань.....	22
1.2. Структури даних.....	25
1.3. База даних	30
1.4. Сховища даних.....	31
1.5. Простори даних.....	32
1.6. Знання.....	33
1.7. База знань	34
1.8. Інженерія знань.....	35
1.9. Інтелектуальні системи.....	36
Резюме.....	38
Завдання для самостійного опрацювання	38
Тести для закріплення матеріалу	39
Розділ 2 Алгебраїчна система. Відображення логічних моделей даних у вигляді алгебраїчних систем.....	43
2.1. Поняття алгебраїчної системи.....	43
2.2. Рівні моделей даних.....	45
2.2.1. Концептуальна модель даних	45
2.2.2. Логічна модель даних.....	46
2.2.3. Фізична модель даних	57
2.2.4. Перспективи методів моделювання предметних областей.....	57
Резюме.....	58
Завдання для самостійного опрацювання	59
Тести для закріплення матеріалу	59
Розділ 3 Реляційні бази даних.....	62

3.1. Поняття реляційної бази даних	62
3.2. Поняття відношення.....	63
3.3. Поняття зв'язку даних.....	67
3.4. Поняття функції.....	69
3.5. Реляційна алгебра.....	73
3.5.1. Теоретико-множинні оператори	73
3.5.2. Спеціальні реляційні оператори.....	75
3.5.3. Поняття алгебраїчного виразу.....	78
3.6. Залежності даних у реляційних базах даних	79
3.6.1. Функціональні залежності відношень і математичне поняття функціональної залежності	79
3.6.2. Нормальні форми.....	82
3.7. Проектування схеми бази даних за допомогою ERD	85
Резюме.....	86
Завдання для самостійного опрацювання	87
Тести для закріплення матеріалу	88
<i>Розділ 4 Темпоральні бази даних.....</i>	<i>91</i>
4.1. Коротка історія.....	91
4.1.1. Ухвалення стандарту і комерційні реалізації.....	92
4.1.2. Поява області досліджень темпоральних баз даних	93
4.2. Основні поняття	94
4.2.1. Лінії часу	94
4.2.2. Інтервальне і точкове подання.....	97
4.3. Розширення реляційної моделі і СКБД.....	98
4.3.1. Створення темпоральної СКБД.....	98
4.3.2. Мова запитів до темпоральних баз даних	99
4.3.3. Подання результатів темпоральних запитів	103
4.3.4. Спеціальне значення «зараз».....	103
4.3.5. Розрідження таблиць з темпоральною підтримкою	105
4.3.6. ACID-властивості темпоральних транзакцій.....	106
4.3.7. Ефективність при роботі з темпоральною СКБД	107
4.4. Суміжні і додаткові області досліджень.....	108
4.4.1. Темпоральні бази XML-даних.....	108
4.4.2. Бази даних мультимедіа	109
4.4.3. Просторово-часові бази даних	109
4.4.4. Механізм знімків станів	110
4.4.5. Галуження ліній часу і версії.....	110

4.5. Пропозиції від виробників комерційних СКБД	111
4.5.1. TIMEDB.....	111
4.5.2. Informix TimeSeries Datablade.....	111
4.5.3. Immortal DB (прототип від Microsoft)	112
4.5.4. Технологія Oracle Flashback – крок до темпоральної СКБД	112
4.5.5. Рішення Oracle Workspace Manager – багатoversійність даних і підтримка знімків станів.....	113
Резюме.....	114
Завдання для самостійного опрацювання	114
Тести для закріплення матеріалу	114
<i>Розділ 5 Ненормалізовані бази даних</i>	<i>116</i>
5.1. Поняття ненормалізованих відношень	117
5.2. Модель Версо: реляційні операції над ненормалізованими відношеннями.....	121
5.2.1. Означення моделі Версо	121
5.2.2. Алгебра Версо.....	123
5.2.3. Зв'язок моделі Версо та реляційної моделі.....	128
5.2.4. Ефективне обчислення реляційних запитів за допомогою вибірки та проєкції Версо	130
5.3. Розділена нормальна форма для вкладених відношень	131
5.3.1. Характеристики та основні поняття розділеної нормальної форми.	131
5.3.2. Реструктурування РНФ-відношень.....	135
5.3.3. «Плоска» інтерпретація РНФ-відношень.....	137
5.3.4. Оптимізація.....	141
5.3.5. Ізоморфні атрибути	142
5.3.6. Атрибути, які можна реструктурувати	144
5.4. Порівняння моделі Версо та розділеної нормальної форми... ..	145
5.5. Огляд моделей зберігання даних	146
5.6. Зміст частково нормалізованої моделі зберігання	148
5.7. Реалізація дій з ненормалізованими відношеннями в промислових СКБД	154
5.7.1. Об'єктно-реляційні особливості та робота з вкладеними таблицями в Oracle	154
5.7.2. Реалізація вкладених таблиць засобами SQL3 та SQL4	157
5.7.3. Вкладені зв'язки в XML.....	159
5.7.4. Особливості СКБД UniVerse фірми IBM.....	160
5.7.5. Особливості реалізації ненормалізованих відношень в інших СКБД.....	161
5.8. Застосування ненормалізованих відношень	162

Резюме.....	163
Завдання для самостійного опрацювання	164
Тести для закріплення матеріалу	164
Розділ 6 Багатовимірні бази даних.....	166
6.1. Основні поняття багатовимірної моделі	167
6.2. Операції маніпулювання даними.....	169
6.2.1. Операція зрізу	169
6.2.2. Операція обертання	171
6.2.3. Операції згортки і деталізації.....	171
6.3. Операції редагування багатовимірної бази даних	172
6.3.1. Зміна значення в агрегатній точці	172
6.3.2. Видалення агрегатної точки.....	173
6.3.3. Напрями планування і верхня межа консолідації	173
6.3.4. Декомпозиція багатовимірної бази даних на паралелотопи.....	174
6.4. Залежності даних у багатовимірній моделі.....	174
6.4.1. Класифікаційні залежності.....	174
6.4.2. Залежності агрегатів у багатовимірній моделі.....	180
6.4.3. Залежності між елементами ієрархії	182
6.5. Порівняння подання даних у реляційній та багатовимірній	
моделях	186
Резюме.....	188
Завдання для самостійного опрацювання	189
Тести для закріплення матеріалу	189
Розділ 7 Сховища даних.....	191
7.1. Поняття сховища даних	191
7.1.1. Визначення та основні характеристики	191
7.1.2. Завдання, що вирішуються за допомогою сховищ даних	193
7.1.3. Модель сховища даних	194
7.1.4. Характеристика даних у сховищі.....	197
7.2. Складові сховища даних	208
7.2.1. Учасники СД.....	208
7.2.2. Програмне забезпечення.....	210
7.2.3. Параметри	211
7.2.4. Запити	212
7.3. Види ієрархії вимірів	212
7.4. Підвиди сховищ даних.....	214

7.4.1. Вітрини даних.....	214
7.4.2. Операційні сховища даних.....	217
Резюме.....	219
Завдання для самостійного опрацювання.....	220
Тести для закріплення матеріалу.....	221
<i>Розділ 8 Проектування структури сховища даних.....</i>	<i>227</i>
8.1. Реляційна модель сховища даних.....	227
8.2. Багатовимірні моделі сховища даних.....	230
8.2.1. Виміри.....	234
8.2.2. Факти.....	235
8.3. Гібридна модель сховища даних.....	236
8.4. Модель сховища даних на основі зведення даних.....	237
8.4.1. Визначення зведення даних.....	237
8.4.2. Проблеми наявних моделей сховищ даних та причини виникнення архітектури зведення даних.....	238
8.4.3. Компоненти зведення даних.....	241
8.4.4. Можливі застосування зведення даних.....	243
8.4.5. Проектування сховища даних на основі зведення даних з ієрархічним підходом.....	244
8.4.6. Проблеми, особливості застосування зведення даних.....	248
Резюме.....	249
Завдання для самостійного опрацювання.....	249
Тести для закріплення матеріалу.....	250
<i>Розділ 9 Інтеграція даних.....</i>	<i>253</i>
9.1. Проблеми, що призводять до інтеграції даних.....	253
9.2. Характеристики інтеграції даних.....	255
9.3. Історія засобів інтеграції.....	256
9.4. Методи інтеграції даних.....	256
9.4.1. Консолідація даних.....	256
9.4.2. Федералізація даних.....	259
9.4.3. Розповсюдження даних.....	260
9.4.4. Гібридний підхід.....	261
9.5. Технології інтеграції.....	262
9.6. Технологія ETL.....	263
9.6.1. Структура процесу перевантаження даних.....	263
9.6.2. Підпроцес STER.....	268

9.6.3. Підпроцес STAC	269
9.6.4. Переваги та недоліки ETL	274
9.7. Технологія ЕП	275
9.8. Технологія ЕАІ	276
9.9. Технологія ЕСМ	277
9.9.1. Поняття ЕСМ	277
9.9.2. Компоненти ЕСМ-рішення	278
9.10. Значення інтеграції для сховищ даних	282
Резюме	284
Завдання для самостійного опрацювання	285
Тести для закріплення матеріалу	285
<i>Розділ 10 Агрегація даних</i>	<i>288</i>
10.1. Поняття агрегату	288
10.2. Часткова і повна агрегація	290
10.2.1. Випадок простих вимірів	291
10.2.2. Випадок ієрархічних вимірів	293
10.3. Оператор агрегації	295
10.4. Процедури формування агрегатів	296
10.4.1. Процедура попереднього формування агрегатів	297
10.4.2. Процедура оперативного формування агрегатів	298
10.5. Агрегація розрідженого гіперкуба даних	299
Резюме	303
Завдання для самостійного опрацювання	304
Тести для закріплення матеріалу	304
<i>Розділ 11 Простори даних</i>	<i>306</i>
11.1. Причини появи просторів даних	306
11.2. Модель простору даних	308
11.3. Учасники простору даних	315
11.4. Задачі учасників простору	319
11.5. Особливості просторів даних	322
11.6. Дослідницькі проблеми	322
11.6.1. Запити в DSSP	322
11.6.2. Розкриття простору даних	322

11.6.3. Повторне використання людської праці	323
11.6.4. Зберігання й індексування просторів даних	323
11.6.5. Гарантії коректності.....	324
11.7. Перспективи просторів даних.....	324
11.7.1. Зв'язок з іншими областями	324
11.7.2. Промислові перспективи	325
Резюме.....	325
Завдання для самостійного опрацювання	326
Тести для закріплення матеріалу.....	326
<i>Розділ 12 Системи зберігання даних. Віртуалізація сховища</i>	
<i>даних.....</i>	328
12.1. Концепція зберігання даних SAN	329
12.2. Концепція зберігання даних NAS	330
12.3. Концепція зберігання даних SAS (DAS)	331
12.4. Програмні системи керування зберіганням даних	332
12.5. Технології та протоколи, що використовуються при побудові систем зберігання даних.....	334
12.5.1. Технологія Fibre Channel.....	334
12.5.2. Технологія iSCSI (Internet SCSI)	335
12.6. Віртуалізація систем зберігання даних	335
12.6.1. Поняття віртуалізації.....	335
12.6.2. Архітектура віртуалізації сховищ даних	337
12.6.3. Особливості віртуалізації	339
12.6.4. Віртуалізація на рівні сервера	340
12.6.5. Віртуалізація на рівні підсистем зберігання	341
12.6.6. Віртуалізація на рівні мережі зберігання	341
12.7. Комплексна характеристика якості віртуалізації.....	342
Резюме.....	345
Завдання для самостійного опрацювання	346
Тести для закріплення матеріалу.....	346
<i>Розділ 13 OLAP-системи</i>	349
13.1. Поняття OLAP	349
13.2. Правила та якості Кодда	352
13.3. Архітектура OLAP-систем.....	354
13.3.1. Сервер OLAP.....	355

13.3.2. Клієнт OLAP	357
13.3.3. Рівні багатомірності в OLAP	358
13.4. Програмне і апаратне забезпечення OLAP.....	359
13.4.1. Загальна характеристика програмних продуктів.....	359
13.4.2. Засоби OLAP-аналізу, пропонувані компанією Microsoft	363
13.5. OLAP-засоби і Web-технології	369
13.5.1. Internet/Intranet OLAP.....	370
13.5.2. Архітектура Web-OLAP	370
13.5.3. Підходи до реалізації Web-OLAP.....	371
13.6. Відмінності сховищ даних і облікових систем.....	374
Резюме.....	378
Завдання для самостійного опрацювання	378
Тести для закріплення матеріалу	379
<i>Розділ 14 Інтелектуальний аналіз даних.....</i>	<i>381</i>
14.1. Поняття видобування даних	381
14.2. Задачі видобування даних.....	382
14.2.1. Класифікація завдань видобування даних.....	382
14.2.2. Задачі класифікації і регресії.....	383
14.2.3. Задача пошуку асоціативних правил	385
14.2.4. Задача кластеризації.....	387
14.3. Моделі видобування даних	388
14.3.1. Передбачувані моделі.....	388
14.3.2. Описові моделі.....	388
14.4. Методи видобування даних	390
14.4.1. Базові методи	390
14.4.2. Нечітка логіка.....	390
14.4.3. Генетичні алгоритми	393
14.4.4. Нейронні мережі.....	394
14.5. Процес виявлення знань.....	395
14.5.1. Основні етапи аналізу	395
14.5.2. Підготовка початкових даних.....	397
14.6. Огляд ринку програмних продуктів видобування даних	398
14.6.1. Програмний продукт Analysis Services	399
14.6.2. Програмний продукт SAS Enterprise Miner	402
14.6.3. Система PolyAnalyst.....	406
14.6.4. Cognos і система STATISTICA Data Miner	407
14.6.5. Програмний продукт STATISTICA Data Miner.....	409

14.6.6. Програмні продукти Oracle Data Mining і Deductor	411
14.6.7. KXEN	414
Резюме.....	417
Завдання для самостійного опрацювання	417
Тести для закріплення матеріалу	418
<i>Розділ 15 Засоби створення сховищ даних.....</i>	<i>420</i>
15.1. Розвиток ринку засобів для сховищ даних	420
15.1.1. Netezza та Datallegro.....	421
15.1.2. Інструменти генерування сховища	423
15.1.3. Інструменти керування сховищем	423
15.1.4. Інструменти доступу	423
15.1.5. Пристрої для сховищ даних.....	423
15.1.6. Прогнози на майбутнє	424
15.2. Сучасні СКБД та сховища даних.....	425
15.2.1. Oracle	425
15.2.2. SQL Server 2005	429
15.2.3. Поліпшення в TRANSACT-SQL.....	433
15.2.4. Informix	434
15.2.5. Visual Warehouse	436
15.2.6. NCR: Scalable Data Warehouse	438
15.2.7. Sybase: Warehouse NOW	439
15.2.8. Продукти SAS Institute	440
Резюме.....	441
Завдання для самостійного опрацювання	441
Тести для закріплення матеріалу	441
Англомовні терміни	443
Україномовні терміни	448
<i>Предметно-іменний покажчик.....</i>	<i>477</i>
<i>Література</i>	<i>483</i>

В.В. Пасічник, Н.Б. Шаховська

Сховища даних

НАВЧАЛЬНИЙ ПОСІБНИК

2-ге видання, стереотипне

Формат 70x100/16. Папір офсетний. Гарнітура Times New Roman.
Умовн. друк. арк. 39,98. Друк цифровий.

ПП “Магнолія 2006”
м. Львів-53, 79053, Україна, тел.+380503701957
e-mail: magnol06@ukr.net

Свідоцтво про внесення суб’єкта видавничої справи
до Державного реєстру видавців, виготівників і розповсюджувачів
видавничої продукції: серія ДК № 2534 від 21.06.2006 року,
видане Державним комітетом інформаційної політики,
телебачення та радіомовлення України

Видавець Марченко Т. В.

Свідоцтво про внесення суб’єкта видавничої справи
до Державного реєстру видавців, виготовлювачів і розповсюджувачів
видавничої продукції: серія ДК № 6784 від 30.05.2019 року,
видане Державним комітетом інформаційної політики,
телебачення та радіомовлення України

Надруковано у друкарні видавця ФОП Марченко Т. В.