

Оглавление

О курсе	13
На кого ориентирован курс	13
Какие знания будут получены	13
Структура курса	14
Программные средства, используемые в курсе	14
Благодарности	15
Часть I. От теории к практике	17
Глава 1. Введение	19
1.1. Базы данных и СУБД	19
1.2. Требования к СУБД	20
1.3. Разделение данных и программ	23
1.4. Языки запросов	26
1.5. Целостность и согласованность	26
1.6. Отказоустойчивость	28
1.7. Безопасность и разграничение доступа	29
1.8. Производительность	29
1.9. Создание приложений, взаимодействующих с базой данных	33
1.10. Итоги главы	34
1.11. Контрольные вопросы	35
Глава 2. Теоретические основы БД	37
2.1. Модели данных	37
2.1.1. Идентификация и изменяемость	38
2.1.2. Навигация и поиск по значениям	40
2.1.3. Объекты и коллекции объектов	41
2.1.4. Свойства моделей данных	41
2.2. Реляционная модель данных	42
2.2.1. Основные понятия реляционной модели данных	43
2.2.2. Реляционная алгебра	47
2.2.3. Другие языки запросов	54
2.2.4. Особенности реляционной модели данных	56
2.2.5. Нормальные формы	57
2.2.6. Практические варианты реляционной модели данных	61

2.3.	Средства концептуального моделирования	63
2.3.1.	Модель данных «сущность — связь»	64
2.3.2.	Концептуальные объектные модели	70
2.4.	Объектные и объектно-реляционные модели данных	71
2.5.	Другие модели данных	73
2.5.1.	Слабоструктурированные модели данных	73
2.5.2.	Модели для представления знаний	74
2.5.3.	Ключ — значение	74
2.5.4.	Устаревшие модели данных	75
2.6.	Примеры проектирования схемы в модели «сущность — связь»	75
2.7.	Библиографические комментарии	81
2.8.	Упражнения	83
Глава 3. Знакомимся с базой данных		85
3.1.	Установка базы данных	85
3.2.	Подключение к серверу базы данных	85
3.3.	Простой клиент: psql	87
3.4.	Итоги главы	90
3.5.	Упражнения	90
Глава 4. Введение в SQL		91
4.1.	Назначение языка SQL	91
4.2.	Быстрый старт	92
4.2.1.	Простые типы данных	92
4.2.2.	Основные конструкции и синтаксис	95
4.2.3.	Описание данных: отношения	95
4.2.4.	Заполнение таблиц	99
4.2.5.	Чтение данных	101
4.2.6.	Модификация данных	103
4.3.	Запросы	104
4.3.1.	Фильтрация и проекция	105
4.3.2.	Произведение и соединение	106
4.3.3.	Псевдонимы для таблиц	111
4.3.4.	Вложенные подзапросы	112
4.3.5.	Упорядочивание результата	116
4.3.6.	Агрегирование и группировка	117
4.3.7.	Теоретико-множественные операции	119
4.3.8.	Вывод результатов после модификации данных	121
4.3.9.	Последовательности	122
4.3.10.	Представления	124

4.4. Структуры хранения	126
4.5. Логическая организация данных	132
4.6. Итоги главы	135
4.7. Упражнения	135
Глава 5. Управление доступом в базах данных	139
5.1. Модели защиты и разграничения доступа	139
5.2. Пользователи и роли в СУБД	141
5.3. Объекты и привилегии	143
5.4. Итоги главы	145
5.5. Упражнения	145
Глава 6. Транзакции и согласованность базы данных	147
6.1. Определение и основные требования к транзакциям	148
6.2. Аномалии конкурентного выполнения	150
6.3. Восстановимость	153
6.4. Диспетчеры и протоколы	154
6.5. Использование транзакций в приложениях	155
6.6. Уровни изоляции	158
6.7. Точки сохранения	161
6.8. Долговечность	162
6.9. Итоги главы	163
6.10. Упражнения	164
Глава 7. Разработка приложений СУБД	167
7.1. Проектирование схемы базы данных	169
7.2. Объектно-реляционная потеря соответствия	172
7.3. Использование каркасов объектно-реляционных отображений	174
7.3.1. Наследование	175
7.3.2. Запросы	179
7.3.3. Когда применять каркасы?	179
7.4. Кеширование данных	180
7.5. Взаимодействие с базой данных	183
7.5.1. Параметры запросов	183
7.5.2. Унифицированные средства взаимодействия	185
7.5.3. Интерфейс PostgreSQL для приложений	186
7.6. Некоторые общие задачи	187
7.6.1. Ограничение доступа к данным	187
7.6.2. Поддержка многоязычности	189
7.7. Настройка	192

7.8. Проектирование декларативных запросов	194
7.9. Итоги главы	195
7.10. Упражнения	196
Глава 8. Расширения реляционной модели	197
8.1. Ограниченность реализаций SQL	197
8.2. Реализация объектных расширений в PostgreSQL	200
8.2.1. Наследование	200
8.2.2. Определение типов данных	201
8.2.3. Домены	202
8.2.4. Коллекции	202
8.2.5. Указатели	203
8.3. Функции	204
8.4. Слабоструктурированные данные: JSON	205
8.5. Слабоструктурированные данные: XML	209
8.6. Активные базы данных	213
8.7. Итоги главы	218
8.8. Упражнения	218
Глава 9. Разновидности СУБД	221
9.1. Классы приложений БД	221
9.2. Структуры хранения	223
9.3. Архитектуры связи с приложениями	224
9.4. Оборудование	226
9.4.1. Носители данных	226
9.4.2. Вычислительные ресурсы	228
9.5. Хранилища данных	230
9.5.1. Агрегатно-ориентированные базы данных	232
9.5.2. Базы данных на основе графов	233
9.6. Выбор СУБД для построения информационных систем	233
9.7. Итоги главы и первой части	236
9.8. Упражнения	237
Часть II. От практики к мастерству	239
Глава 10. Архитектура СУБД	241
10.1. Интерфейс приложений	242
10.2. Обеспечение согласованности и отказоустойчивости	243
10.3. Выполнение запросов	244

10.4. Организация хранения данных	246
10.5. Управление процессами и оперативной памятью	248
10.6. Параллельные и распределенные базы данных	249
10.7. Расширения и расширяемость	251
10.8. Безопасность	252
10.9. Итоги главы	252
10.10. Упражнения	252
Глава 11. Структуры хранения и основные алгоритмы СУБД	255
11.1. Хранение объектов логического уровня	255
11.1.1. Размещение коллекций объектов	256
11.1.2. Размещение данных на страницах	260
11.1.3. Хранение больших объектов	263
11.1.4. Строки или колонки?	264
11.2. Индексы	265
11.2.1. Одномерные индексы	267
11.2.2. Пространственные индексы	275
11.2.3. Инвертированные индексные структуры	280
11.2.4. Разреженные индексы	282
11.2.5. Сигнатурные индексы	282
11.2.6. Особенности реализации индексов в PostgreSQL	284
11.3. Выполнение алгебраических операций	286
11.3.1. Алгебраические операции и алгоритмы	286
11.3.2. Операции выборки данных	287
11.3.3. Сортировка	289
11.3.4. Алгоритм вложенных циклов	291
11.3.5. Алгоритм соединения на основе сортировки и слияния	294
11.3.6. Соединение на основе хеширования	297
11.3.7. Многопоточное соединение	299
11.4. Итоги главы и библиографические комментарии	300
11.5. Упражнения	301
Глава 12. Выполнение и оптимизация запросов	303
12.1. Стадии обработки запроса	303
12.2. Подготовка и выполнение	306
12.3. Оптимизация запросов	308
12.3.1. Задача оптимизации	308
12.3.2. Сокращение пространства планов	310
12.3.3. Алгоритмы оптимизации	311

12.4.	Модели стоимости	321
12.4.1.	Функции и модели стоимости	321
12.4.2.	Модели стоимости для алгоритмов бинарных операций	322
12.4.3.	Оценки селективности	325
12.4.4.	Статистические характеристики данных	326
12.5.	Другие подходы к оптимизации запросов	328
12.5.1.	Адаптивное выполнение запросов	329
12.5.2.	Параметрическая оптимизация	332
12.5.3.	Семантическая оптимизация	333
12.5.4.	Многокритериальная оптимизация	333
12.6.	Итоги главы	334
12.7.	Упражнения	334
Глава 13.	Управление транзакциями	337
13.1.	Критерии корректности конкурентного выполнения	338
13.1.1.	Формальные модели корректности	338
13.1.2.	Изоляция мгновенных снимков	347
13.1.3.	Расписания с множественными версиями данных	351
13.1.4.	Восстановимость	354
13.1.5.	Дополнительные свойства классов расписаний	357
13.2.	Диспетчеры и протоколы	357
13.2.1.	Требования и критерии оценки	358
13.2.2.	Блокировки	360
13.2.3.	Двухфазные протоколы, использующие блокировки	362
13.2.4.	Тупики	364
13.2.5.	Другие протоколы на основе блокирования	366
13.2.6.	Протокол на основе меток времени	367
13.2.7.	Реализации протокола SI	369
13.2.8.	Многоверсионные протоколы	370
13.2.9.	Блокировки или метки времени?	372
13.3.	Ослабленные критерии корректности: уровни изоляции в SQL	372
13.4.	Итоги главы	375
13.5.	Упражнения	375
Глава 14.	Надежность баз данных	377
14.1.	Восстановление после отказов	377
14.2.	Отказы сервера баз данных	378
14.2.1.	Журнал транзакций	378
14.2.2.	Рестарт сервера	382
14.2.3.	Контрольные точки	384

14.3.	Разрушение носителя	385
14.3.1.	Экспорт и импорт	387
14.3.2.	Копирование с восстановлением по журналам	387
14.3.3.	Резервные серверы баз данных	390
14.4.	Итоги главы	391
14.5.	Упражнения	391
Глава 15.	Дополнительные возможности SQL	393
15.1.	Дополнительные средства SQL	393
15.1.1.	Общие табличные выражения	393
15.1.2.	Рекурсивные запросы	397
15.1.3.	Аналитические и оконные функции	401
15.2.	Избыточные структуры хранения	405
15.2.1.	Материализованные представления	405
15.2.2.	Индексы	408
15.3.	Итоги главы	416
15.4.	Упражнения	417
Глава 16.	Функции и процедуры в базе данных	419
16.1.	Хранимые подпрограммы	419
16.2.	Процедурный язык PL/pgSQL	426
16.2.1.	Структурные конструкции языка PL/pgSQL	427
16.2.2.	Работа с объектами базы данных	430
16.2.3.	Динамический SQL	434
16.2.4.	Обработка исключительных ситуаций	436
16.3.	Функции и процедуры на языке SQL	440
16.4.	Итоги главы	442
16.5.	Упражнения	442
Глава 17.	Расширяемость PostgreSQL	443
17.1.	Пользовательские агрегаты	443
17.2.	Типы данных, операторы и классы операторов	446
17.3.	Индексы	450
17.4.	Другие инструменты расширения	452
17.4.1.	Модули расширения	453
17.4.2.	Обертки сторонних данных	454
17.4.3.	Подключение новых процедурных языков	455
17.5.	Итоги главы	455
17.6.	Упражнения	456

Глава 18. Полнотекстовый поиск	457
18.1. Модели информационного поиска	457
18.1.1. Предварительная обработка текста	459
18.1.2. Булева модель информационного поиска	459
18.1.3. Векторные модели информационного поиска	462
18.2. Средства полнотекстового поиска в PostgreSQL	465
18.3. Поддержка нечеткого поиска в PostgreSQL	467
18.3.1. Триграммный поиск	467
18.3.2. Фонетический поиск	469
18.4. Итоги главы	470
18.5. Упражнения	471
Глава 19. Безопасность данных	473
19.1. Безопасность и разграничение доступа	473
19.2. Основные понятия и модели	474
19.3. Особенности ролей в PostgreSQL	475
19.4. Привилегии	476
19.5. Права доступа при выполнении хранимых функций	477
19.6. Разграничение доступа на уровне строк таблиц	479
19.7. Регистрация событий и изменений	483
19.8. Итоги главы	484
19.9. Упражнения	484
Глава 20. Администрирование баз данных	487
20.1. Планирование конфигурации сервисов хранения данных	489
20.2. Безопасность и разграничение доступа	492
20.3. Конфигурация баз данных	492
20.4. Мониторинг баз данных	494
20.5. Настройка производительности	497
20.5.1. Настройка серверов баз данных	500
20.5.2. Настройка схемы базы данных	503
20.5.3. Настройка запросов	507
20.5.4. Целостная настройка приложений	509
20.6. Надежность и доступность	509
20.7. Техническое обслуживание базы данных	512
20.8. Итоги главы	513
20.9. Упражнения	513
Глава 21. Репликация баз данных	515
21.1. Множественные копии данных	515

21.2.	Согласованность реплик	516
21.3.	Согласованность, доступность, разделение сети	519
21.4.	Поддержка единой логической копии	520
21.5.	Симметричные протоколы синхронизации реплик	521
21.6.	Репликация главной копии	522
21.7.	Резервные серверы базы данных	525
21.8.	Репликация в системе PostgreSQL	526
21.9.	Итоги главы	528
21.10.	Упражнения	529
Глава 22.	Параллельные и распределенные СУБД	531
22.1.	Архитектуры параллельной и распределенной обработки	531
22.2.	Параллельные серверы баз данных	534
22.2.1.	Конфигурации оборудования	534
22.2.2.	Гранулярность параллелизма	535
22.2.3.	Размещение данных	537
22.2.4.	Параллельные алгоритмы для бинарных операций	538
22.2.5.	Параллелизм между операциями	541
22.2.6.	Не все так просто	542
22.2.7.	Параллельные запросы в PostgreSQL	543
22.3.	Выполнение запросов в распределенных СУБД	545
22.3.1.	Конфигурации распределенных баз данных	545
22.3.2.	Организация доступа к удаленным данным	546
22.3.3.	Подготовка и выполнение запросов	549
22.4.	Согласованность в распределенных системах	551
22.4.1.	Распределенные транзакции	551
22.4.2.	Протоколы управления транзакциями	552
22.4.3.	Завершение распределенных транзакций	554
22.5.	Итоги главы	556
22.6.	Упражнения	556
Заключение		559
Список литературы		563
Предметный указатель		569

О курсе

На кого ориентирован курс

Курс рассчитан на студентов классических и технических университетов и других вузов, имеющих базовую подготовку по программированию и продолжающих специализироваться в областях, близких к программированию.

Какие знания будут получены

В курсе подробно рассматриваются основные понятия, устройство и принципы работы СУБД, а также технологии (архитектура, алгоритмы, структуры данных), лежащие в их основе.

Прослушавшие курс получают уверенные знания и практические навыки по следующим вопросам:

- устройство и принципы работы СУБД;
- проектирование баз данных;
- работа с SQL — составление и оптимизация запросов;
- разработка серверных приложений;
- использование различных типов индексов;
- обработка транзакций и одновременный доступ;
- основы эксплуатации баз данных;
- обеспечение надежности хранения, отказоустойчивости и высокой доступности;
- принципы организации и работы параллельных и распределенных СУБД;
- работа со слабоструктурированными данными (JSON, XML).

Такая подготовка позволит на старших курсах (в магистратуре) специализироваться на разработке и настройке приложений баз данных либо в областях проектирования и разработки СУБД.

Структура курса

Курс состоит из двух частей.

Первая часть рассчитана на студентов младших курсов бакалавриата. В ней рассматриваются основные сведения о базах данных и системах управления базами данных: реляционная модель данных, язык SQL, обработка транзакций.

Вторая часть курса может читаться на последних курсах бакалавриата или для студентов магистратуры. Во второй части подробно рассматриваются технологии, лежащие в основе функционирования СУБД, а также современные направления и тенденции развития СУБД, основные аспекты их практического применения. При этом некоторые темы, рассмотренные в первой части, изучаются повторно на более глубоком уровне.

Курс в основном касается классических реляционных и объектно-реляционных СУБД, но затрагивает также тематику неклассических СУБД.

Практические занятия не только помогают закрепить пройденный на лекциях материал. Они содержат много дополнительной информации, закрепляющей и расширяющей знания, изложенные в теоретической части. В качестве СУБД для практических занятий используется PostgreSQL.

Как первая, так и вторая части курса могут быть выделены в самостоятельные курсы. Отдельные разделы курса могут быть скомбинированы так, чтобы получить более практическую или более фундаментальную направленность либо адаптировать курс к конкретному учебному плану вуза.

Программные средства, используемые в курсе

Для эффективного освоения материала курса и для выполнения упражнений необходимо установить на компьютере ряд программных продуктов. Набор этих продуктов может зависеть от используемой операционной системы и от других обстоятельств, но в любом случае понадобятся:

- система управления базами данных PostgreSQL;
- демонстрационная база данных, которая используется в большинстве примеров;
- текстовый редактор для подготовки запросов на языке SQL.

Установка PostgreSQL и демобазы рассматривается в главе 3.

Для выполнения упражнений по созданию и редактированию моделей данных может потребоваться инструмент для редактирования диаграмм.

Для разработки приложений на императивных языках программирования (C, C++, Java, Python и др.) потребуются соответствующие среды разработки, однако такие упражнения не включены в состав этого курса.

При работе с PostgreSQL можно использовать ряд приложений, предоставляющих графические интерфейсы для работы с базами данных. Многие из этих приложений могут работать с различными СУБД. Как правило, в таких системах ограничены возможности работы с особенностями любой конкретной СУБД. При освоении материала этого курса целесообразнее использовать средства самой системы PostgreSQL или другие программы, спроектированные специально для работы с PostgreSQL.

Благодарности

Подготовка этого курса была бы невозможна без активной поддержки со стороны компании Postgres Professional и ее руководства, в частности О. Бартунова и И. Панченко. Качество материала существенно улучшилось благодаря усилиям Е. Рогова, взявшего на себя огромный труд по редактированию курса.

Главы 7 и 9 написаны совместно Е. Горшковой и Б. Новиковым, глава 20 — совместно Н. Графеевой и Б. Новиковым. В подготовке упражнений принимали участие В. Бусаров, К. Секереш, Г. Шалыгина и Е. Михайлова.

Часть I

**От теории
к практике**

Глава 1

Введение

Компьютеры используются повсеместно: невозможно найти предприятие или учреждение, которые не применяли бы их для решения производственных или управленческих задач. Подобные высказывания не слишком заметны в средствах массовой информации, потому что они уже давно не являются новостью. Профессионалы, однако, понимают, что на самом деле важны не компьютеры, а информационные системы, которые на них работают, а в центре любой информационной системы находятся данные.

Эта книга о том, как хранить данные, обеспечивать их корректность и сохранность и как их обрабатывать эффективно.

1.1. Базы данных и СУБД

Появление и относительно широкое распространение в начале 1960-х гг. запоминающих устройств достаточно большой емкости с возможностью доступа к произвольным участкам памяти — магнитных дисков — открыло широкие возможности для создания сложных структур долговременно хранимых данных. Высокая скорость обновления небольших объемов данных (доли секунды) создала условия для создания приложений, способных функционировать в режиме оперативной работы (online). В отличие от систем предшествующих поколений время ответа стало измеряться не сутками, а секундами или долями секунды.

Эти возможности, однако, привели к существенному усложнению кода приложений и, как следствие, к удорожанию их разработки и снижению надежности. В связи с этим появилась идея централизации функций управления данными, которая привела к появлению систем, предоставляющих приложениям услуги по обработке данных. Такие системы получили название систем управления базами данных (СУБД).

Поскольку СУБД используются многими приложениями, ожидается, что они могут обеспечивать более высокие значения эксплуатационных характеристик, таких как надежность хранения и эффективность обработки, недостижимые при индивидуальной разработке средств управления данными для каждого приложения.

Важно отметить, что многие особенности и характеристики, присущие ранним системам управления базами данных, связаны с требованиями тех областей применения и классов приложений, которые были наиболее актуальны в то время. В первую очередь это приложения, работающие в режиме оперативной обработки (online transaction processing, OLTP) в банковской и финансовой сферах.

Прежде чем обсуждать, каким образом эти области применения повлияли на характеристики СУБД, уточним значение некоторых терминов, которые будут использоваться в дальнейшем.

Система управления базами данных (СУБД) — это программный комплекс, обеспечивающий централизованное хранение данных и предоставляющий приложениям услуги по обработке данных.

Совокупность данных, хранимых под управлением СУБД, называется *базой данных*. Оригинальное английское словосочетание data base дословно переводится как «основание, состоящее из данных». В русском словосочетании «база данных» этот смысл несколько искажается. На самом деле это — фундамент, на котором строятся приложения и который состоит из данных. Действительно, данные (а следовательно, база данных) являются очень существенной частью практически любой информационной системы.

Система управления базами данных, находящаяся в фазе выполнения, связанная с некоторой конкретной базой данных и готовая выполнять запросы на обработку этой базы данных, называется *экземпляром (instance)* или *сервером базы данных*. На самом деле экземпляр и сервер — разные понятия: один сервер баз данных может управлять несколькими экземплярами баз данных, однако это различие станет важным только начиная с главы 5.

1.2. Требования к СУБД

Ранние системы управления данными очень сильно различались как по своей внутренней организации, так и по предоставляемым возможностям. Потребо-

важилось несколько лет, для того чтобы определить, каковы основные функции систем управления базами данных и какие требования следует предъявлять к таким системам.

Основные требования к системам управления базами данных были сформулированы в документе, опубликованном в 1971 г. комитетом по системам и языкам обработки данных (CODASYL) [28], русский перевод которого издан в 1975 г. [64]. Основой для этих требований послужил анализ систем, применявшихся в период подготовки отчета, и особенностей прикладных областей, в которых эти системы использовались.

В дальнейшем круг областей применения СУБД непрерывно расширялся, появлялись новые системы и уходили старые, однако многие из этих требований остались актуальными и сегодня, и большинство современных СУБД в той или иной форме реализует значительную часть этих требований. Однако далеко не все классы приложений, в которых используются современные системы, предъявляют те же требования к обработке данных, поэтому и системы реализуются иначе.

Приложения, относящиеся к классу оперативной обработки (OLTP), характеризуются тем, что:

- каждое выполнение приложения занимает мало времени (в идеале — не больше долей секунды);
- данные используются совместно многими приложениями;
- при каждом выполнении приложение использует ничтожную долю общего объема хранимых данных, и обычно количество используемых данных не зависит от общего объема базы.

Важно также отметить, что процессы обработки данных и структуры данных в тех областях, в которых использовались ранние СУБД, фактически были формализованы задолго до появления электронных вычислительных систем. Так, правила бухгалтерского учета в основном сложились в XIV веке и мало изменились в последующем. Возможно, это привело к тому, что СУБД, как правило, ориентированы на обработку структурированных данных.

Перечислим основные требования к системам управления базами данных.

Разделение программ и данных. Описание структуры данных должно быть отделено от кода приложений, и система должна допускать независимое

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru