

Распределенные данные

Алгоритмы работы современных систем
хранения информации

Алекс Петров



Санкт-Петербург • Москва • Минск

2024

Краткое содержание

Предисловие	12
Часть I. Подсистема хранения данных	18
Глава 1. Введение и обзор	24
Глава 2. Введение в B-деревья	43
Глава 3. Форматы файлов	62
Глава 4. Реализация B-деревьев	79
Глава 5. Обработка транзакций и восстановление	96
Глава 6. Варианты B-деревьев	129
Глава 7. Журналированное хранилище	147
Часть I. Заключение	184
Часть II. Распределенные системы	186
Глава 8. Введение и обзор	189
Глава 9. Обнаружение отказов	214
Глава 10. Выбор лидера	223
Глава 11. Репликация и согласованность	232
Глава 12. Антиэнтропия и распространение	261
Глава 13. Распределенные транзакции.....	275
Глава 14. Консенсус	298
Часть II. Заключение	333
Об авторе	336
Об обложке	336
Приложение А. Библиография	https://www.piter.com/book.phtml?978544611640

Оглавление

Предисловие к русскому изданию	11
Предисловие	12
Кому предназначена эта книга	13
Зачем мне читать эту книгу?	14
Рассматриваемые темы	14
Структура книги	15
Условные обозначения	16
Благодарности	17
От издательства	17
ЧАСТЬ I. ПОДСИСТЕМА ХРАНЕНИЯ ДАННЫХ	18
Сравнение баз данных	19
Понимание преимуществ и недостатков	22
Глава 1. Введение и обзор	24
Архитектура СУБД	25
Резидентные и дисковые СУБД	27
Колоночные и строчные СУБД	30
Файлы данных и индексные файлы	34
Буферизация, неизменяемость и упорядочение	39
Итоги	41
Дополнительная литература	42
Глава 2. Введение в B-деревья	43
Двоичные деревья поиска	43
Дисковые структуры	47
Вездесущие B-деревья	51
Итоги	61
Дополнительная литература	61
Глава 3. Форматы файлов	62
Актуальность	63

Двоичное кодирование.....	63
Основные принципы	68
Структура страницы	69
Слоттированные страницы	70
Структура ячеек	72
Объединение ячеек в слоттированные страницы	73
Управление данными переменного размера	75
Управление версиями	76
Вычисление контрольной суммы	77
Итоги	78
Дополнительная литература	78
Глава 4. Реализация В-деревьев	79
Заголовок страницы	79
Двоичный поиск	85
Распространение операций разделения и слияния	85
Перебалансировка	87
Добавление только справа	89
Сжатие	91
Очистка и обслуживание	92
Итоги	94
Дополнительная литература	95
Глава 5. Обработка транзакций и восстановление	96
Организация буферизации данных	98
Восстановление	106
Управление параллелизмом	111
Итоги	127
Дополнительная литература	127
Глава 6. Варианты В-деревьев	129
Копирование при записи	129
Абстракции для управления обновлениями	131
Ленивые В-деревья	132
FD-деревья	135
Вw-деревья	138
Кэш-независимые В-деревья	143

Итоги	145
Дополнительная литература	146
Глава 7. Журналированное хранилище	147
LSM-деревья	148
Чтение, запись и увеличение пространства	162
Подробнее о реализации	164
Неупорядоченное LSM-хранилище	171
Параллелизм в LSM-деревьях	174
Многоуровневое совмещение журналов	176
LLAMA и тщательное многоуровневое совмещение	180
Итоги	182
Дополнительная литература	182
ЧАСТЬ I. ЗАКЛЮЧЕНИЕ	184
ЧАСТЬ II. РАСПРЕДЕЛЕННЫЕ СИСТЕМЫ	186
Основные определения	187
Глава 8. Введение и обзор	189
Конкурентное выполнение	189
Общее состояние в распределенной системе	191
Абстракции распределенных систем	200
Задача двух генералов	206
Невозможность Фишера—Линча—Патерсона	207
Синхронность системы	209
Модели отказов	210
Итоги	212
Дополнительная литература	213
Глава 9. Обнаружение отказов	214
Контрольные пакеты и эхо-запросы	215
Детектор отказа с накопленным уровнем подозрительности	218
Сплетни и обнаружение отказов	219
Обратный взгляд на проблему обнаружения отказов	221
Итоги	222
Дополнительная литература	222

Глава 10. Выбор лидера	223
Алгоритм забияки.....	224
Аварийное переключение к следующему в очереди	226
Оптимизация с кандидатами и обычными узлами	227
Алгоритм с приглашениями	228
Кольцевой алгоритм	229
Итоги	230
Дополнительная литература.....	231
Глава 11. Репликация и согласованность	232
Обеспечение доступности.....	233
Печально известная теорема CAP	233
Общая память	236
Упорядочение	238
Модели согласованности	239
Модели сеансов	251
Согласованность в конечном счете	252
Настраиваемая согласованность	253
Реплики-свидетели	255
Строгая согласованность в конечном счете и структуры CRDT	256
Итоги	259
Дополнительная литература	260
Глава 12. Антиэнтропия и распространение	261
Исправление при чтении	262
Чтение с запросом хэш-суммы	264
Передача подсказки	264
Деревья Меркла	265
Битовая карта векторов версий	266
Распространение сплетен	268
Итоги	273
Дополнительная литература	274
Глава 13. Распределенные транзакции	275
Обеспечение атомарности операций	276
Двухфазная фиксация	277

Трехфазная фиксация	282
Распределенные транзакции с использованием протокола Calvin	284
Распределенные транзакции с использованием протокола Spanner	286
Секционирование базы данных	289
Распределенные транзакции с использованием библиотеки Percolator	290
Исключение координации	293
Итоги	296
Дополнительная литература	297
Глава 14. Консенсус	298
Рассылка	299
Атомарная рассылка	300
Паксос	304
Raft	320
Византийский консенсус	326
Итоги	330
Дополнительная литература	331
ЧАСТЬ II. ЗАКЛЮЧЕНИЕ	333
Дополнительная литература	334
Об авторе	336
Об обложке	336
Приложение А. Библиография	336

