

УДК 004.45
ББК 32.973
С80

Стин ван М., Таненбаум Э. С.

С80 Распределенные системы / пер. с англ. В. А. Яроцкого. – М.: ДМК Пресс, 2021. – 584 с.: ил.

ISBN 978-5-97060-708-4

В третьем издании классического труда профессоров Амстердамского университета Эндрю Таненбаума и Мартена ван Стина обсуждаются принципы и парадигмы распределенных систем.

Кроме обширного теоретического материала в книге приведен код на языке Python (размещен на сайте dmkpress.com), демонстрирующий использование полученных знаний на практике.

В числе рассматриваемых тем: основные характеристики распределенных систем; архитектуры программных компонентов, входящих в систему; процессы и коммуникации; присваивание имен; согласованность и репликация; отказоустойчивость и безопасность.

Для опытных разработчиков, занимающихся распределенными системами, а также студентов профильных вузов.

УДК 004.45
ББК 32.973

Title of English-language edition Distributed Systems, 3rd edition., published by Maarten van Steen. Russian language edition copyright © 2021 by DMK Press. All rights reserved.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

ISBN 978-15-430573-8-6 (англ.)

© Maarten van Steen
and Andrew S. Tanenbaum, 2017

ISBN 978-5-97060-708-4 (рус.)

© Оформление, издание, перевод,
ДМК Пресс, 2021

Содержание

Предисловие	14
От издательства	15
Глава 1. Введение	16
1.1. Что такое распределенная система?	17
Характеристика 1: совокупность автономных вычислительных элементов	17
Характеристика 2: единая цельная система	19
Промежуточное программное обеспечение и распределенные системы	20
1.2. Цели дизайна	22
Поддержка совместного использования ресурсов	22
Создание прозрачных распределений	23
Типы прозрачности распределений	23
Степень прозрачности распределения	26
Открытость	27
Функциональная совместимость, компоновка и расширяемость	28
Отделение политики от механизма	29
Масштабирование	30
Размерность масштабируемости	31
Техника масштабирования	36
Ловушки	39
1.3. Типы распределенных систем	40
Высокопроизводительные распределенные вычисления	41
Кластерные вычисления	42
Сетевые вычисления	44
Облачные вычисления	46
Распределенные информационные системы	50
Распределенная обработка транзакций	50
Интеграция корпоративных приложений	53
Распространенные системы	56
Повсеместно распространенные вычислительные системы	56
Мобильные вычислительные системы	59
Сенсорные сети	63
1.4. Резюме	68
Глава 2. Архитектуры	70
2.1. Архитектурные стили	71
Многоуровневая архитектура	72
Многоуровневые протоколы связи	73

Уровни приложений	75
Объектно-ориентированные и сервис-ориентированные архитектуры	77
Ресурсные архитектуры	79
Архитектура публикация-подписка	82
2.2. Организация промежуточного программного обеспечения	87
Упаковщики	87
Перехватчики	89
Модифицируемое промежуточное ПО	90
2.3. Системная архитектура	91
Централизованные организации	92
Простая архитектура клиент-сервер	92
Многоуровневая архитектура	93
Децентрализованные организации: одноранговые системы	96
Структурированные одноранговые системы	97
Неструктурированные одноранговые системы	100
Иерархически организованные одноранговые сети	103
Гибридные архитектуры	105
Системы пограничных серверов	106
Совместные распределенные системы	107
2.4. Примеры архитектур	110
Сетевая файловая система NFS	110
Веб	114
Простые веб-системы	114
Многоуровневые архитектуры	116
2.5. Резюме	117
Глава 3. Процессы	119
3.1. Потоки	120
Введение в потоки	120
Использование потоков в нераспределенных системах	122
Реализация потоков	125
Потоки в распределенных системах	128
Многопоточные клиенты	128
Многопоточные серверы	130
3.2. Виртуализация	132
Принцип виртуализации	133
Виртуализация и распределенные системы	133
Типы виртуализации	135
Применение виртуальных машин в распределенных системах	139
3.3. Клиенты	141
Сетевые пользовательские интерфейсы	141
Пример: система X Window	141
Сетевые вычисления для тонких клиентов	143
Клиентское программное обеспечение для прозрачности распространения	144
3.4. Серверы	146
Общие вопросы дизайна	146

Параллельный сервер против итеративного сервера	146
Связь с сервером: конечные точки.....	146
Прерывание сервера.....	148
Серверы без сохранения состояния против серверов, сохраняющих состояние	148
Объектные серверы	150
Пример: веб-сервер Apache	157
Кластеры серверов.....	159
Локальные кластеры	159
Общая организация.....	159
Глобальные кластеры	163
Пример использования: PlanetLab.....	167
V-серверы.....	169
3.5. Миграция кода	170
Причины переноса кода.....	171
Миграция в гетерогенных системах	176
3.6. Резюме	179
Глава 4. Коммуникации	182
4.1. Основы	183
Многоуровневые протоколы	183
Эталонная модель OSI.....	183
Протоколы промежуточного программного обеспечения.....	189
Типы коммуникаций.....	190
4.2. Удаленный вызов процедуры.....	192
Основная операция RPC.....	193
Передача параметров.....	198
Поддержка приложений на основе RPC.....	202
Генерация заглушки	202
Языковая поддержка	203
Вариации RPC	205
Асинхронный RPC	205
Многоадресный RPC.....	206
Пример: распределенная вычислительная среда RPC.....	207
Введение в распределенную вычислительную среду DCE	208
Цели DCE RPC.....	208
Написание клиента и сервера	209
Привязка клиента к серверу	211
Выполнение RPC.....	212
4.3. Коммуникации, ориентированные на сообщения	212
Простой временный обмен сообщениями с сокетами	213
Расширенный переходный обмен сообщениями	218
Использование шаблонов обмена сообщениями: ZeroMQ.....	218
Интерфейс передачи сообщений (MPI)	223
Постоянная связь, ориентированная на сообщения.....	226
Модель очереди сообщений.....	226
Общая архитектура системы очереди сообщений.....	228

Брокеры сообщений	230
Пример: система очереди сообщений IBM WebSphere	233
Обзор	233
Каналы.....	234
Передача сообщений.....	235
Управление оверлейными сетями.....	237
Пример: расширенный протокол очереди сообщений (AMQP).....	238
Основы	239
AMQP связи.....	239
AMQP обмена сообщениями.....	241
4.4. Многоадресная связь	242
Многоадресная рассылка на уровне дерева приложений	242
Проблемы с производительностью в оверлеях	243
Многоадресная передача сообщений на основе лавинной маршрутизации	246
Распространение данных по принципу сплетни	250
Модели распространения информации	250
Удаление данных.....	254
4.5. Резюме	255
Глава 5. Присваивание имен.....	257
5.1. Имена, идентификаторы и адреса.....	258
5.2. Бесструктурное (плоское) наименование	261
Простые решения	261
Широковещание	262
Прямые указатели	263
Методы домашнего местоположения	265
Распределенные хеш-таблицы	267
Общий механизм.....	267
Иерархические методы	271
5.3. Структурированное наименование	276
Пространства имен.....	277
Разрешение имени	279
Механизм закрытия	280
Связывание и монтаж	281
Реализация пространства имен.....	284
Распределение пространства имен.....	285
Реализация разрешения имен.....	287
Пример: система доменных имен.....	292
Пространство имен DNS	292
Реализация DNS.....	294
Пример: сетевая файловая система	298
5.4. Наименование на основе атрибутов.....	303
Службы каталогов.....	304
Иерархические реализации: протокол LDAP.....	305
Децентрализованные реализации	308
Использование распределенного индекса	309

Пространственные кривые	310
5.5. Резюме	315
Глава 6. Координация	317
6.1. Синхронизация часов	318
Физические часы	319
Алгоритмы синхронизации часов	323
Сетевой временной протокол	325
Алгоритм Беркли	326
Синхронизация часов в беспроводных сетях	327
6.2. Логические часы	330
Логические часы Лампорта	331
Пример: полностью упорядоченная многоадресная рассылка	333
Векторные часы	337
6.3. Взаимное исключение	342
Обзор	342
Централизованный алгоритм	343
Распределенный алгоритм	344
Алгоритм кольцо токенов	346
Децентрализованный алгоритм	347
6.4. Алгоритмы выбора	350
Алгоритм хулигана	351
Кольцевой алгоритм	352
Выборы в беспроводной среде	353
Выборы в масштабных системах	356
6.5. Системы локации	357
GPS: система глобального позиционирования	357
Когда GPS не выбор	359
Логическое позиционирование узлов	360
Централизованное позиционирование	361
Децентрализованное позиционирование	363
6.6. Сопоставление распределенных событий	364
Централизованные реализации	364
6.7. Координация на основе сплетен	370
Объединение	370
Служба одноранговой выборки	372
Структура оверлея, основанная на сплетнях	373
6.8. Резюме	374
Глава 7. Согласованность и репликация	377
7.1. Введение	378
Причины репликации	378
Репликация как метод масштабирования	379
7.2. Модели согласованности, ориентированные на данные	381
Непрерывное согласование	382
О понятии конит	383

Согласованный порядок операций	386
Последовательная согласованность	386
Причинная согласованность	391
Группирование операций	393
Согласованность и когерентность	395
Конечная согласованность	395
7.3. Модели согласованности, ориентированные на клиента	398
Монотонные чтения	400
Монотонные записи	402
Чтение собственных записей	404
Запись следует за чтением	405
7.4. Управление репликами	406
Поиск лучшего местоположения сервера	406
Репликация и размещение контента	408
Постоянные реплики	409
Реплики, инициированные сервером	409
Реплики, инициированные клиентом	411
Распространение контента	412
Состояние против операции	412
Протоколы извлечения и проталкивания	413
Одноадресная и многоадресная рассылка	416
Управление реплицированными объектами	417
7.5. Согласованность протоколов	420
Непрерывная последовательность	420
Ограничивающее числовое отклонение	420
Граничные отклонения устаревания	422
Ограничение отклонений порядка	422
Первичные протоколы	423
Протоколы удаленной записи	423
Протоколы локальной записи	424
Протоколы реплицируемой записи	426
Активная репликация	426
Протоколы на основе кворума	426
Протоколы кеширования	428
Реализация согласованности, ориентированной на клиента	432
7.6. Пример: кеширование и репликация в сети	434
7.7. Резюме	445
Глава 8. Отказоустойчивость	448
8.1. Введение в отказоустойчивость	449
Базовые концепции	449
Модели отказов	452
Маскировка отказов посредством избыточности	456
8.2. Устойчивость процесса	458
Устойчивость групповых процессов	458
Организация групп	458
Управление членством	459

Маскировка и репликация отказа	460
Консенсус в неисправных системах со сбоями	461
Пример: Paxos.....	463
Основная идея Paxos	464
Понимание Paxos.....	468
Консенсус в неисправных системах с произвольными отказами.....	475
Почему 3k процессов недостаточно	476
Почему достаточно 3k + 1 процессов	477
Пример: практическая византийская отказоустойчивость	481
Некоторые ограничения по реализации отказоустойчивости.....	484
Относительно достижения консенсуса	484
Согласованность, доступность и разделение.....	486
Обнаружение отказов.....	487
8.3. Надежная связь клиент-сервер	489
Двухточечная связь	490
Семантика RPC при наличии отказов	490
Клиент не может найти сервер.....	490
Потерянные сообщения запроса	491
Сбой сервера	491
Потерянные ответные сообщения	494
Клиент неисправен.....	495
8.4. Надежное групповое общение	496
Атомарная многоадресная рассылка	503
Виртуальная синхронность	503
Порядок сообщений	505
8.5. Распределенная фиксация	509
8.6. Восстановление	517
Введение.....	517
Контрольная точка.....	520
Скоординированная контрольная точка	521
Независимая контрольная точка.....	521
Регистрация сообщений.....	523
Вычисления, ориентированные на восстановление	526
8.7. Резюме.....	526
Глава 9. Безопасность	529
9.1. Введение	530
Угрозы безопасности, политики и механизмы	530
Проблемы дизайна	532
Контроль.....	532
Уровни механизмов безопасности.....	533
Распределение механизмов безопасности.....	535
Простота	536
Криптография	537
9.2. Безопасные каналы.....	541
Аутентификация	541
Аутентификация на основе общего секретного ключа.....	542

Аутентификация с использованием центра распределения ключей	545
Аутентификация с использованием криптографии с открытым ключом	548
Целостность и конфиденциальность сообщений	549
Цифровые подписи	549
Сессионные ключи	551
Безопасное групповое общение	553
Конфиденциальное групповое общение	553
Безопасные реплицированные серверы	553
Пример: система Kerberos	556
9.3. Контроль доступа	558
Общие вопросы управления доступом	558
Матрица контроля доступа	559
Брандмауэры	562
Безопасный мобильный код	564
Отказ в обслуживании	568
9.4. Безопасное наименование	570
9.5. Управление безопасностью	571
Управление ключами	571
Установка ключей	572
Распространение ключей	573
Безопасное управление группами	575
Управление авторизацией	577
Возможности и атрибуты	577
Делегирование (прав)	580
9.6. Резюме	582