

Большие данные

Принципы
и практика построения
масштабируемых систем
обработки данных
в реальном времени

Натан Марц
и Джеймс Уоррен



MANNING

Большие данные

Принципы и практика
построения масштабируемых
систем обработки данных
в реальном времени

Натан Марц
Джеймс Уоррен



Издательский дом "Вильямс"
Москва • Санкт-Петербург • Киев
2017

Оглавление

Благодарности	18
Об этой книге	21
Глава 1. Новая парадигма для больших данных	25
ЧАСТЬ I. УРОВЕНЬ ПАКЕТНОЙ ОБРАБОТКИ	53
Глава 2. Модель данных для больших данных	55
Глава 3. Иллюстрация модели данных для больших данных	77
Глава 4. Хранение данных на уровне пакетной обработки	85
Глава 5. Иллюстрация хранения данных на уровне пакетной обработки	99
Глава 6. Уровень пакетной обработки	119
Глава 7. Иллюстрация уровня пакетной обработки	151
Глава 8. Пример построения уровня пакетной обработки: архитектура и алгоритмы	181
Глава 9. Пример реализации уровня пакетной обработки	199
ЧАСТЬ II. УРОВЕНЬ ОБСЛУЖИВАНИЯ	221
Глава 10. Организация уровня обслуживания	223
Глава 11. Иллюстрация уровня обслуживания	243
ЧАСТЬ III. УРОВЕНЬ УСКОРЕНИЯ	253
Глава 12. Представления в реальном времени	255
Глава 13. Иллюстрация представлений в реальном времени	271
Глава 14. Организация очередей и обработка потоков	277
Глава 15. Иллюстрация организации очередей и обработки потоков	297
Глава 16. Микропакетная обработка потоков	309
Глава 17. Иллюстрация микропакетной обработки потоков	327
Глава 18. Лямбда-архитектура в деталях	343
Предметный указатель	363

ББК 32.973.26-018.2.75

М29

УДК 681.3.07

Издательский дом "Вильямс"

Зав. редакцией *С.Н. Тригуб*

Перевод с английского и редакция *И.В. Берштейна*

Научный консультант канд. техн. наук. *А.Я. Гладун*

По общим вопросам обращайтесь в Издательский дом "Вильямс" по адресу:
info@williamspublishing.com, <http://www.williamspublishing.com>

Марц, Натан, Уоррен, Джеймс.

М29 Большие данные: принципы и практика построения масштабируемых систем обработки данных в реальном времени. : Пер. с англ. – М. : ООО "И.Д. Вильямс", 2017. – 368 с. : ил. – Парал. тит. англ.

ISBN 978-5-8459-2075-1 (рус.)

ББК 32.973.26-018.2.75

Все названия программных продуктов являются зарегистрированными торговыми марками соответствующих фирм.

Никакая часть настоящего издания ни в каких целях не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами, будь то электронные или механические, включая фотокопирование и запись на магнитный носитель, если на это нет письменного разрешения издательства Manning Publication, Co.

Authorized translation from the English language edition published by Manning Publications Co. Copyright © 2015. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by means electronic, mechanical, photocopying, or otherwise, without prior written permission of the publisher.

Russian language edition is published by Williams Publishing House according to the Agreement with R&I Enterprises International. Copyright © 2017.

Книга напечатана согласно договору с ООО "ПРИМСНАБ".

Научно-популярное издание

Натан Марц, Джеймс Уоррен

Большие данные: принципы и практика построения масштабируемых систем обработки данных в реальном времени

Литературный редактор *И.А. Попова*

Верстка *О.В. Мишуткина*

Художественный редактор *Е.П. Дьотник*

Корректор *Л.А. Гордиенко*

Подписано в печать 10.11.2016. Формат 70х100/16

Гарнитура Times

Усл. печ. л. 29,67. Уч.-изд. л. 22,8

Доп. тираж 300 экз. Заказ № 8143

Отпечатано в АО «Первая Образцовая типография»

Филиал «Чеховский Печатный Двор»

142300, Московская область, г. Чехов, ул. Полиграфистов, д.1

ООО "И. Д. Вильямс", 127055, г. Москва, ул. Лесная, д. 43, стр. 1

ISBN 978-5-8459-2075-1 (рус.)

ISBN 978-1-617-29034-3 (англ.)

© Издательский дом "Вильямс", 2017

© Manning Publications Co, 2015