

Next Generation Databases

NoSQL, NewSQL, and Big Data



Guy Harrison

Apress®

Next Generation Databases

Copyright © 2015 by Guy Harrison

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

ISBN-13 (pbk): 978-1-4842-1330-8

ISBN-13 (electronic): 978-1-4842-1329-2

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Managing Director: Welmoed Spahr

Lead Editor: Jonathan Gennick

Development Editor: Douglas Pundick

Technical Reviewer: Stephane Faroult

Editorial Board: Steve Anglin, Pramila Balen, Louise Corrigan, Jim DeWolf, Jonathan Gennick,

Robert Hutchinson, Celestin Suresh John, Michelle Lowman, James Markham, Susan McDermott,

Matthew Moodie, Jeffrey Pepper, Douglas Pundick, Ben Renow-Clarke, Gwenan Spearing

Coordinating Editor: Jill Balzano

Copy Editor: Carole Berglie

Compositor: SPi Global

Indexer: SPi Global

Artist: SPi Global

Cover Designer: Anna Ishchenko

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springer.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a Delaware corporation.

For information on translations, please e-mail rights@apress.com, or visit www.apress.com.

Apress and friends of ED books may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Special Bulk Sales–eBook Licensing web page at www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this text is available to readers at www.apress.com. For detailed information about how to locate your book's source code, go to www.apress.com/source-code/.

To Catherine Maree Arnold (1981-2010)

Contents at a Glance

About the Author	xvii
About the Technical Reviewer	xix
Acknowledgments	xxi
■ Part I: Next Generation Databases.....	1
■ Chapter 1: Three Database Revolutions.....	3
■ Chapter 2: Google, Big Data, and Hadoop	21
■ Chapter 3: Sharding, Amazon, and the Birth of NoSQL.....	39
■ Chapter 4: Document Databases	53
■ Chapter 5: Tables are Not Your Friends: Graph Databases	65
■ Chapter 6: Column Databases	75
■ Chapter 7: The End of Disk? SSD and In-Memory Databases	87
■ Part II: The Gory Details.....	103
■ Chapter 8: Distributed Database Patterns	105
■ Chapter 9: Consistency Models	127
■ Chapter 10: Data Models and Storage	145
■ Chapter 11: Languages and Programming Interfaces.....	167
■ Chapter 12: Databases of the Future	191
■ Appendix A: Database Survey.....	217
Index.....	229

Contents

About the Author	xvii
About the Technical Reviewer	xix
Acknowledgments	xxi
■ Part I: Next Generation Databases.....	1
■ Chapter 1: Three Database Revolutions.....	3
Early Database Systems.....	4
The First Database Revolution	6
The Second Database Revolution.....	7
Relational theory	8
Transaction Models.....	9
The First Relational Databases.....	10
Database Wars!.....	10
Client-server Computing.....	11
Object-oriented Programming and the OODBMS.....	11
The Relational Plateau	13
The Third Database Revolution.....	13
Google and Hadoop.....	14
The Rest of the Web.....	14
Cloud Computing	15
Document Databases.....	15

■ CONTENTS

The “NewSQL”	16
The Nonrelational Explosion	16
Conclusion: One Size Doesn’t Fit All	17
Notes	18
■ Chapter 2: Google, Big Data, and Hadoop	21
The Big Data Revolution	21
Cloud, Mobile, Social, and Big Data	22
Google: Pioneer of Big Data	23
Google Hardware	23
The Google Software Stack	25
More about MapReduce	26
Hadoop: Open-Source Google Stack	27
Hadoop’s Origins.....	28
The Power of Hadoop	28
Hadoop’s Architecture	29
HBase	32
Hive.....	34
Pig.....	36
The Hadoop Ecosystem	37
Conclusion.....	37
Notes	37
■ Chapter 3: Sharding, Amazon, and the Birth of NoSQL.....	39
Scaling Web 2.0.....	39
How Web 2.0 was Won	40
The Open-source Solution	40
Sharding	41
Death by a Thousand Shards	43
CAP Theorem	43
Eventual Consistency.....	44

Amazon’s Dynamo	45
Consistent Hashing	47
Tunable Consistency	49
Dynamo and the Key-value Store Family	51
Conclusion	51
Note	51
■ Chapter 4: Document Databases	53
XML and XML Databases	54
XML Tools and Standards	54
XML Databases	55
XML Support in Relational Systems	57
JSON Document Databases	57
JSON and AJAX	57
JSON Databases	58
Data Models in Document Databases	60
Early JSON Databases	61
MemBase and CouchBase	61
MongoDB	61
JSON, JSON, Everywhere	63
Conclusion	63
■ Chapter 5: Tables are Not Your Friends: Graph Databases	65
What is a Graph?	65
RDBMS Patterns for Graphs	67
RDF and SPARQL	68
Property Graphs and Neo4j	69
Gremlin	71
Graph Database Internals	73
Graph Compute Engines	73
Conclusion	74

■ Chapter 6: Column Databases	75
Data Warehousing Schemas.....	75
The Columnar Alternative.....	77
Columnar Compression	79
Columnar Write Penalty	79
Sybase IQ, C-Store, and Vertica	81
Column Database Architectures	81
Projections.....	82
Columnar Technology in Other Databases	84
Conclusion.....	85
Note	85
■ Chapter 7: The End of Disk? SSD and In-Memory Databases	87
The End of Disk?	87
Solid State Disk	88
The Economics of Disk	89
SSD-Enabled Databases.....	90
In-Memory Databases	91
TimesTen	92
Redis.....	93
SAP HANA	95
VoltDB	97
Oracle 12c “in-Memory Database”.....	98
Berkeley Analytics Data Stack and Spark	99
Spark Architecture.....	101
Conclusion.....	102
Note	102

■ Part II: The Gory Details	103
■ Chapter 8: Distributed Database Patterns	105
Distributed Relational Databases	105
Replication.....	107
Shared Nothing and Shared Disk.....	107
Nonrelational Distributed Databases.....	110
MongoDB Sharding and Replication.....	110
Sharding	110
Sharding Mechanisms	111
Cluster Balancing	113
Replication.....	113
Write Concern and Read Preference	115
HBase	115
Tables, Regions, and RegionServers.....	116
Caching and Data Locality	117
Rowkey Ordering	118
RegionServer Splits, Balancing, and Failure.....	119
Region Replicas	119
Cassandra	119
Gossip.....	119
Consistent Hashing.....	120
Replicas	124
Snitches.....	126
Summary.....	126
■ Chapter 9: Consistency Models	127
Types of Consistency.....	127
ACID and MVCC.....	128
Global Transaction Sequence Numbers	130
Two-phase Commit.....	130
Other Levels of Consistency	130

Consistency in MongoDB.....	131
MongoDB Locking.....	131
Replica Sets and Eventual Consistency.....	132
HBase Consistency.....	132
Eventually Consistent Region Replicas.....	132
Cassandra Consistency	134
Replication Factor.....	134
Write Consistency.....	134
Read Consistency	135
Interaction between Consistency Levels	135
Hinted Handoff and Read Repair	136
Timestamps and Granularity.....	137
Vector Clocks.....	138
Lightweight Transactions.....	140
Conclusion.....	143
■ Chapter 10: Data Models and Storage.....	145
Data Models	145
Review of the Relational Model of Data.....	146
Key-value Stores	148
Data Models in BigTable and HBase.....	151
Cassandra.....	153
JSON Data Models.....	156
Storage.....	157
Typical Relational Storage Model	158
Log-structured Merge Trees	160
Secondary Indexing.....	163
Conclusion.....	166

■ Chapter 11: Languages and Programming Interfaces	167
SQL	167
NoSQL APIs	169
Riak	169
Hbase	171
MongoDB	173
Cassandra Query Language (CQL)	175
MapReduce	177
Pig	179
Directed Acyclic Graphs	181
Cascading	181
Spark	181
The Return of SQL	182
Hive	183
Impala	184
Spark SQL	185
Couchbase N1QL	185
Apache Drill	188
Other SQL on NoSQL	190
Conclusion	190
Note	190
■ Chapter 12: Databases of the Future	191
The Revolution Revisited	191
Counterrevolutionaries	192
Have We Come Full Circle?	193
An Embarrassment of Choice	194
Can We have it All?	195
Consistency Models	195
Schema	196

Database Languages	198
Storage	199
A Vision for a Converged Database.....	200
Meanwhile, Back at Oracle HQ ...	201
Oracle JSON Support.....	202
Accessing JSON via Oracle REST	204
REST Access to Oracle Tables.....	206
Oracle Graph.....	207
Oracle Sharding	208
Oracle as a Hybrid Database	210
Other Convergent Databases.....	210
Disruptive Database Technologies	211
Storage Technologies	211
Blockchain	212
Quantum Computing.....	213
Conclusion.....	214
Notes	215
■ Appendix A: Database Survey.....	217
Aerospike	217
Cassandra	218
CouchBase	219
DynamoDB.....	219
HBase	220
MarkLogic	221
MongoDB.....	221
Neo4J	222
NuoDB	223
Oracle RDBMS.....	223

Redis	224
Riak	225
SAP HANA.....	225
TimesTen	226
Vertica	227
VoltDB.....	227
Index.....	229

About the Author



Guy Harrison started working as a database developer in the mid-1980s. He has written numerous books on database development and performance optimization. He joined Quest Software (now part of Dell) in 2000, and currently leads the team that develops the Toad, Spotlight, and Shareplex product families.

Guy lives in Melbourne Australia, with his wife, a variable number of adult children, a cat, three dogs, and a giant killer rabbit.

Guy can be reached at guy.a.harrison@gmail.com at <http://guyharrison.net> and is @guyharrison on Twitter.

About the Technical Reviewer

Stéphane Faroult is a French consultant who first discovered relational databases and the SQL language 30 years ago. Stéphane joined Oracle France in its early days (after a brief spell with IBM and a period of time teaching at the University of Ottawa), and developed an interest in performance and tuning topics, on which he soon started writing training courses. After leaving Oracle in 1988, Stéphane briefly tried going straight and did a bit of operational research, but after only a year he succumbed again to the allure of relational databases. He is currently visiting faculty in the Computing and Information Science Department at Kansas State University. For his sins, Stéphane has been performing database consultancy continuously ever since; he founded RoughSea, Ltd. in 1998. In recent years, Stéphane has had a growing interest in education, which has taken various forms, including books (*The Art of SQL*, soon followed by *Refactoring SQL Applications*, both published by O'Reilly) and more recently a textbook (*SQL Success*, published by RoughSea), a series of seminars in Asia, and video tutorials (www.youtube.com/user/roughsealtd).

Acknowledgements

I'd like to thank everyone at Apress who helped in the production of this book, in particular lead editor Jonathan Gennick, coordinating editor Jill Balzano, and development editor Douglas Pundick. I'd like to especially thank Stéphane Faroult, who provided outstanding technical review feedback. Rarely have I worked with a reviewer of the caliber of Stéphane, and his comments were invaluable.

As always, thanks to my family—Jenni, Chris, Kate, Mike, and Willie—for providing the emotional support and understanding necessary to take on this project.

This book is dedicated to the memory of Catherine Maree Arnold, our beloved niece who passed away in 2010. I met Catherine when she was five years old; the first time I met her, she introduced me to the magic of Roald Dahl's *The Twits*. The last time I saw her, she explained modern DNA sequencing techniques and told me about her ambition to save species from extinction. She was one of the smartest, funniest, and kindest human beings you could hope to meet. When my first book was published in 1987, she jokingly insisted that I should dedicate it to her, so it's fitting that I dedicate this book to her memory.

—Guy Harrison
Melbourne, Australia
December 2015