

*M&rrtelona Access to Yoter Plerrret-Size Dette?*

*The Definitive Guide*

*O'REILLY” Lars George*



Iolalu-o-o:

*HBase The Definitive Guide*

S CCT LL TT LLLL LL LLLLLLLLSS BBCTlC MMLLLLLL CCLLLLLLLLLL L

virtually colko, amount of claia. This look shows you low Åpich-IIB) (r fulfil (ir (= s h= #r &#

LLLLLLCCCCCLLL T T LLS LLLLLLLL CTCCCCCCCS 0LLLCS CLLLLS C !!!!!!wsärkl iiiilli \*\*\* ¢!!uiiiris, whik: \*ri\*uring thāl wri!

CCLL LL LSLTCCCLL CCCLLLL CCCCCTTS CCCLCC CCCCLLCC LL asking posited outstions about Hikio. This look provides CCCCCCTTT CCCCCCS LLCCLLLLCC CCC C CCCTCCCm CCCC CCCCCCCCCS

Liiiiiılxix.: gar pl:Inning tri piiri i inila Izi:i:iiri: right way,

Making Data Work

Bostover how tight integration with Hadoop makes \*rilatiility with HEA: ::Hir:

Distribute Lrge dáta $i-t, irrri$$ariiriexpensiwer!u\*t\*ri\* commodity servers

Aroes: HBase with native Java clients, or with gateway servers providing REST, Avro, or Thrift APIs

(Gętęłęłąłş tiri HBążę';ärrhitėėturę,Ęn:¡¡¡¡¡¡¡¡¡ thë##qrążję #rr:#, Writ#), b):#und (# (miðr:

Integritë HBălţi: With Hidririp'; MIRĖidurë friirreyriskfrir mħassively parallësizēti data förstēšsirgjöł:

Learn how to tune cloisters, designschémas opy tables, Importbulk data, decommission rodes, and many other tasks

г+scources at qг+\*\*\*сапodata.

**aTC CB LLL BeeBB BCCBBBBe e eCCCeS t rata trils, irid těchritiliige; thi turn lig dala**

Iпta &плгt dғыізісп5. Ғтні ігіпгтыtніп ш.гын

**Լ: ::I.EE**

:E -1- - 1 -1

|li

**ĽAH HH.G:**

固

晶晶亨臀品\*[

*"Distributed systems are*

*osen coochabes. Lars George dissills practical engineering from the toporosthese fysics, gir'inglésriri excepţionalgstideso HBase in the process."*

—Josh Patterson org

Lärs Isroorgo, an Hoo: onIniilor, Works as 3 ohith his Aralıileri Hir L kılırları Hrı ~цppst Histikostрі ипсl HBust:

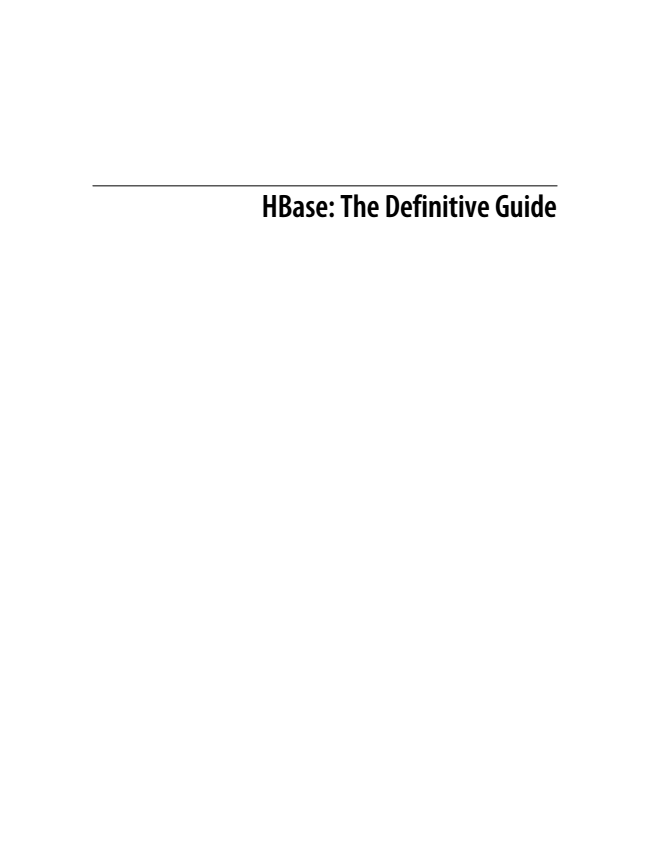
Ilirsigh lethristal suppriri, #rg (#k, årld If Airling Нr has sрн кеп лі чыг\*\*\*\* Н:ixixxxp 1 хет і ігснірі пихніпр:х вигнІ Інтgr cтинi татні злигін нFYI: JEMI in Hallaxels.

ኧD]

Twitter: oreillymedia farebbök tömstreilly

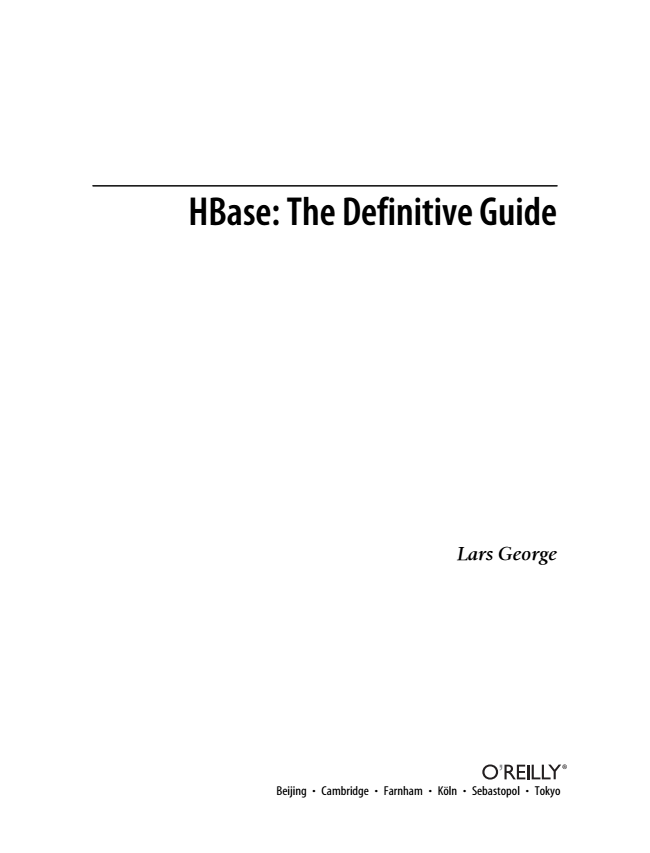
O'REILLY”

oreilly.com



**HBase: The Definitive Guide**

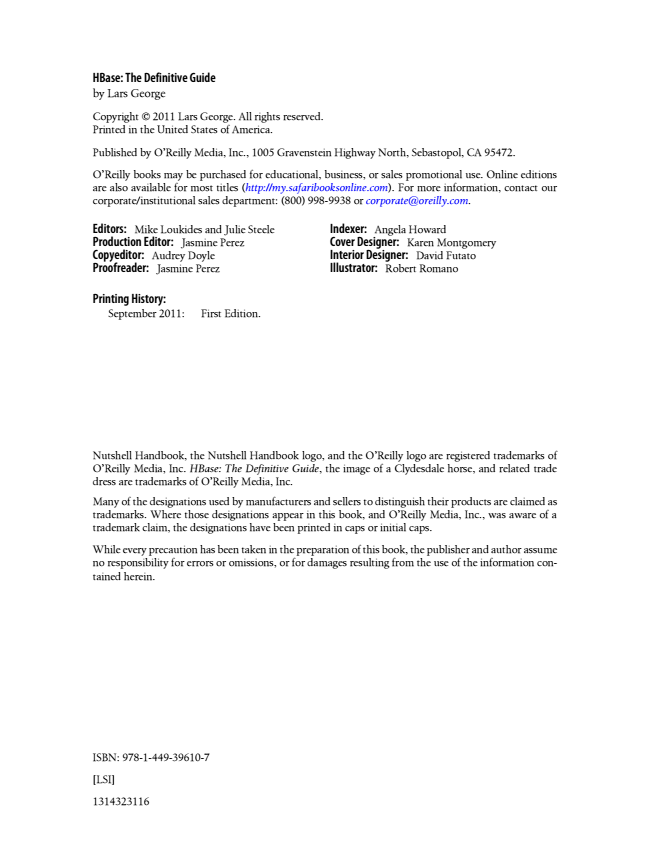
****

****

**HBase: The Definitive Guide**

***Lars George***

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo



**HBase: The Definitive Guide by Lars George**

Copyright © 2011 Lars George. All rights reserved. Printed in the United States of America.

Published by O’Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O’Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (http://my.safaribooksonline.com). For more information, contact our corporate/institutional sales department: (800) 998-9938 or corporate@oreilly.com.

**Editors: Production Copyeditor: Proofreader: Mike Editor: Audrey Jasmine Loukides Jasmine Doyle Perez**

and Perez

Julie Steele

**Indexer: Cover Interior Illustrator: Designer: Designer:**

Angela Robert Karen Howard

David Romano

Montgomery

Futato

**Printing History:**

September 2011: First Edition.

Nutshell Handbook, the Nutshell Handbook logo, and the O’Reilly logo are registered trademarks of O’Reilly Media, Inc. HBase: The Definitive Guide, the image of a Clydesdale horse, and related trade dress are trademarks of O’Reilly Media, Inc.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and O’Reilly Media, Inc., was aware of a trademark claim, the designations have been printed in caps or initial caps.

While every precaution has been taken in the preparation of this book, the publisher and author assume no responsibility for errors or omissions, or for damages resulting from the use of the information con- tained herein.

ISBN: 978-1-449-39610-7

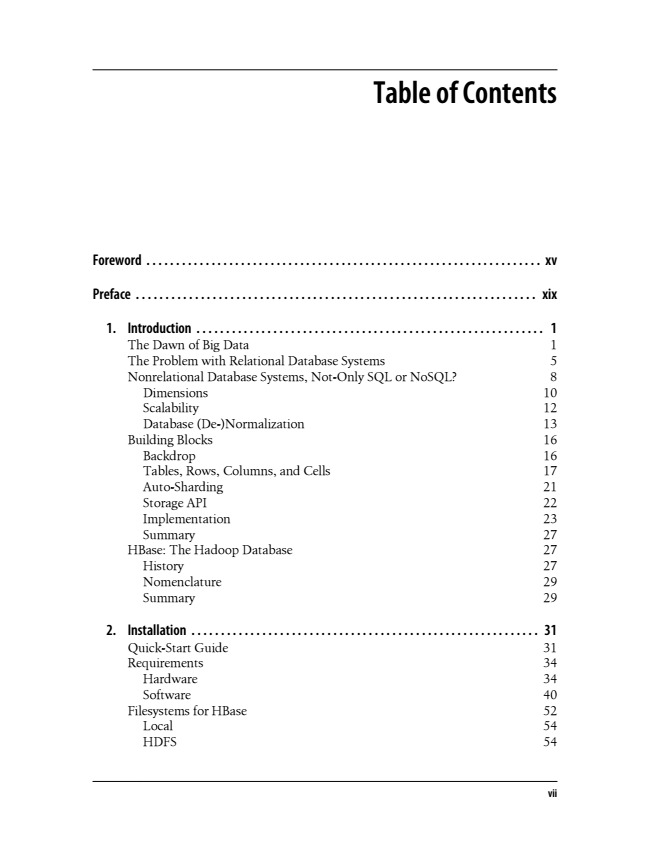
[LSI]

1314323116



*For my wife Katja, my daughter Laura, and son Leon. I love you!*

**

**

**Table of Contents**

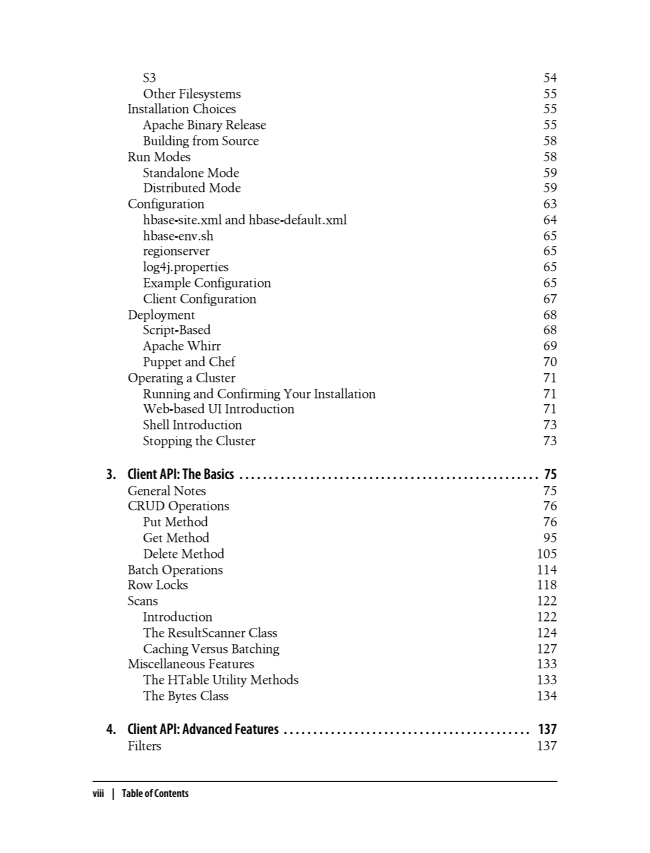
**Foreword . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . xv**

**Preface . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . xix**

1. Introduction ........................................................... 1 The Dawn of Big Data 1 The Problem with Relational Database Systems 5 Nonrelational Database Systems, Not-Only SQL or NoSQL? 8 Dimensions 10 Scalability 12 Database (De-)Normalization 13 Building Blocks 16 Backdrop 16 Tables, Rows, Columns, and Cells 17 Auto-Sharding 21 Storage API 22 Implementation 23 Summary 27 HBase: The Hadoop Database 27 History 27 Nomenclature 29 Summary 29

2. Installation ........................................................... 31 Quick-Start Guide 31 Requirements 34 Hardware 34 Software 40 Filesystems for HBase 52 Local 54 HDFS 54

**vii**

****

S3 54 Other Filesystems 55 Installation Choices 55 Apache Binary Release 55 Building from Source 58 Run Modes 58 Standalone Mode 59 Distributed Mode 59 Configuration 63 hbase-site.xml and hbase-default.xml 64 hbase-env.sh 65 regionserver 65 log4j.properties 65 Example Configuration 65 Client Configuration 67 Deployment 68 Script-Based 68 Apache Whirr 69 Puppet and Chef 70 Operating a Cluster 71 Running and Confirming Your Installation 71 Web-based UI Introduction 71 Shell Introduction 73 Stopping the Cluster 73

3. Client API: The Basics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 75 General Notes 75 CRUD Operations 76 Put Method 76 Get Method 95 Delete Method 105 Batch Operations 114 Row Locks 118 Scans 122 Introduction 122 The ResultScanner Class 124 Caching Versus Batching 127 Miscellaneous Features 133 The HTable Utility Methods 133 The Bytes Class 134

**4. Client API: Advanced Features . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 137 Filters 137**

**viii | Table of Contents**