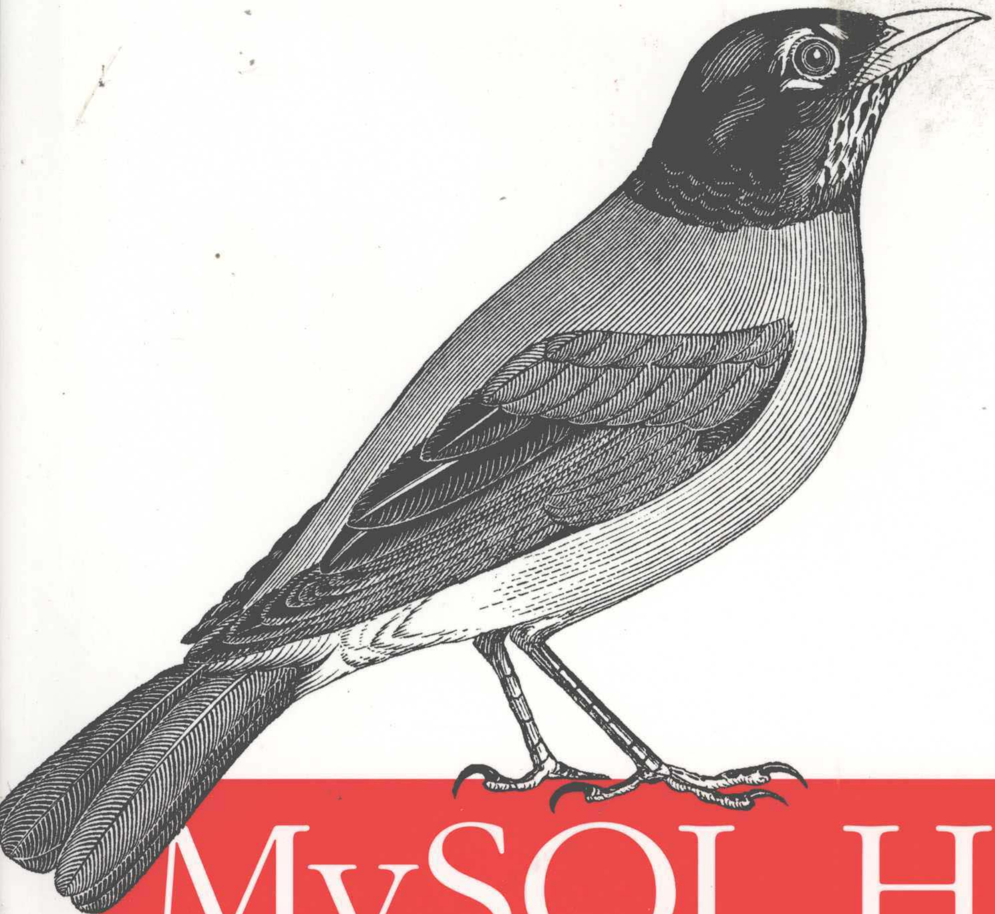
高可用性*MySQL*（影印版）



MySQL High Availability

Charles Bell, Mats Kindahl
& Lars Thalmann 著
Mark Callaghan 序

O'REILLY®
東南大學出版社

高可用性 MySQL（影印版）

MySQL High Availability

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Taipei • Tokyo

O'Reilly Media, Inc. 授权东南大学出版社出版

东南大学出版社

图书在版编目（CIP）数据

高可用性MySQL：英文/（美）贝尔（Bell, C.），（瑞典）欣达尔（Kindahl, M.），（瑞典）达勒马纳（Thalmann, L. L.）著.—影印本.—南京：东南大学出版社，2010.11

书名原文：MySQL High Availability
ISBN 978-7-5641-2525-7

I.①高… II.①贝… ②欣… ③达… III.①关系数据库－数据库管理系统，MySQL IV.①TP311.138

中国版本图书馆CIP数据核字（2010）第223172号

江苏省版权局著作权合同登记
图字：10-2010-287号

高可用性MySQL（影印版）

出版发行：东南大学出版社
地　　址：南京四牌楼2号　　邮编：210096
出 版 人：江建中
网　　址：http://www.seupress.com
电子邮件：press@seu.edu.cn
印　　刷：扬中市印刷有限公司
开　　本：787毫米×980毫米　16开本
印　　张：39印张
字　　数：764千字
版　　次：2011年1月第1版
印　　次：2011年1月第1次印刷
书　　号：ISBN 978-7-5641-2525-7
印　　数：1~1600册
定　　价：82.00元（册）

本社图书若有印装质量问题，请直接与读者服务部联系。电话（传真）：025-83792328

Foreword

A lot of research has been done on replication, but most of the resulting concepts are never put into production. In contrast, MySQL replication is widely deployed but has never been adequately explained. This book changes that. Things are explained here that were previously limited to people willing to read a lot of source code and spend a lot of time debugging it in production, including a few late-night sessions.

Replication enables you to provide highly available data services while enduring the inevitable failures. There are an amazing number of ways for things to fail, including the loss of a disk, server, or data center. Even when hardware is perfect or fully redundant, people are not. Database tables will be dropped by mistake. Applications will write incorrect data. Occasional failure is assured. But with reasonable preparation, recovery from failure can also be assured. The keys to survival are redundancy and backups. Replication in MySQL supports both.

But MySQL replication is not limited to supporting failure recovery. It is frequently used to support read scale-out. MySQL can efficiently replicate to a large number of servers. For applications that are read-mostly, this is a cost-effective strategy for supporting a large number of queries on commodity hardware.

And there are other interesting uses for MySQL replication. Online DDL is a very complex feature to implement in an relational database management system. MySQL does not support online DDL, but through the use of replication you can implement something that is frequently good enough. You can get a lot done with replication if you are willing to be creative.

Replication is one of the features that made MySQL wildly popular. It is also the feature that allows you to convert a popular MySQL prototype into a successful business-critical deployment. Like most of MySQL, replication favors simplicity and ease of use. As a consequence, it is occasionally less than perfect when running in production. This book explains what you need to know to successfully use MySQL replication. It will help you to understand how replication has been implemented, what can go wrong, how to prevent problems, and how to fix them when they crop up despite your best attempts at prevention.

MySQL replication is also a work in progress. Change, like failure, is also assured. MySQL is responding to that change and replication continues to get more efficient, more robust, and more interesting. For instance, row-based replication is new in MySQL 5.1.

While MySQL deployments come in all shapes and sizes, I care most about data services for Internet applications and am excited about the potential to replicate from MySQL to distributed storage systems like HBase and Hadoop. This will make MySQL better at sharing the data center.

I have been on teams that support important MySQL deployments at Facebook and Google. I have had the opportunity, problems, and time to learn much of what is covered in this book. The authors of this book are also experts on MySQL replication, and by reading this book you can share their expertise.

—Mark Callaghan

Preface

The authors of this book have been creating parts of MySQL and working with it for many years. Charles Bell is a senior developer working on replication and backup. His interests include all things MySQL, database theory, software engineering, and agile development practices. Dr. Mats Kindahl is the lead developer for replication and a member of the MySQL Backup and Replication team. He is the main architect and implementor of the MySQL row-based replication and has also developed the unit testing framework used by MySQL. Dr. Lars Thalmann is the development manager and technical lead of the MySQL Replication and Backup team and has designed many of the replication and backup features. He has worked with development of MySQL clustering, replication, and backup technologies.

We wrote this book to fill a gap we noticed among the many books on MySQL. There are many excellent books on MySQL, but few that concentrate on its advanced features and its applications, such as high availability, reliability, and maintainability. In this book, you will find all of these topics and more.

We also wanted to make the reading a bit more interesting by including a running narrative about a MySQL professional who encounters common requests made by his boss. In the narrative, you will meet Joel Thomas, who recently decided to take a job working for a company that has just started using MySQL. You will observe Joel as he learns his way around MySQL and tackles some of the toughest problems facing MySQL professionals. We hope you find this aspect of the book entertaining.

Audience

This book is for MySQL professionals. We expect you to have a basic background in SQL, administering MySQL, and the operating system you are running. We will try to fill in background information about replication, disaster recovery, system monitoring, and other key topics of high availability. See Chapter 1 for other books that offer useful background.

Organization of This Book

This book is written in three parts. Part I encompasses MySQL replication, including high availability and scale-out. Part II examines monitoring and performance concerns for building robust data centers. Part III examines some additional areas of MySQL, including cloud computing and MySQL clusters.

Part I, Replication

Chapter 1, *Introduction*, explains how this book can help you and gives you a context for reading it.

Chapter 2, *MySQL Replication Fundamentals*, discusses both manual and automated procedures for setting up basic replication.

Chapter 3, *The Binary Log*, explains the critical file that ties together replication and helps in disaster recovery, troubleshooting, and other administrative tasks.

Chapter 4, *Replication for High Availability*, shows a number of ways to recover from server failure, including the use of automated scripts.

Chapter 5, *MySQL Replication for Scale-Out*, shows a number of techniques and topologies for improving response time and handling large data sets.

Chapter 6, *Advanced Replication*, addresses a number of topics, such as secure data transfer and row-based replication.

Part II, Monitoring and Disaster Recovery

Chapter 7, *Getting Started with Monitoring*, presents the main operating system parameters you have to be aware of, and tools for monitoring them.

Chapter 8, *Monitoring MySQL*, presents several tools for monitoring database activity and performance.

Chapter 9, *Storage Engine Monitoring*, explains some of the parameters you need to monitor on a more detailed level, focusing on issues specific to MyISAM or InnoDB.

Chapter 10, *Replication Monitoring*, offers details about how to keep track of what masters and slaves are doing.

Chapter 11, *Replication Troubleshooting*, shows how to deal with failures and restarts, corruption, and other incidents.

Chapter 12, *Protecting Your Investment*, explains the use of backups and disaster recovery techniques.

Chapter 13, *MySQL Enterprise*, introduces a suite of tools that simplifies many of the tasks presented in earlier chapters.

Part III, High Availability Environments

Chapter 14, *Cloud Computing Solutions*, introduces the most popular cloud computing service, the Amazon.com AWS, and offers techniques for using MySQL in such virtualized environments.

Chapter 15, *MySQL Cluster*, shows how to use this tool to achieve high availability.

The Appendix, *Replication Tips and Tricks*, offers a grab bag of procedures that are useful in certain situations.

Conventions Used in This Book

The following typographical conventions are used in this book:

Plain text
: Indicates menu titles, options, and buttons.

Italic
: Indicates new terms, table and database names, URLs, email addresses, filenames, and Unix utilities.

`Constant width`
: Indicates command-line options, variables and other code elements, the contents of files, and the output from commands.

`Constant width bold`
: Shows commands or other text that should be typed literally by the user.

`Constant width italic`
: Shows text that should be replaced with user-supplied values.

This icon signifies a tip, suggestion, or general note.

This icon indicates a warning or caution.

Using Code Examples

This book is here to help you get your job done. In general, you may use the code in this book in your programs and documentation. You do not need to contact us for permission unless you're reproducing a significant portion of the code. For example, writing a program that uses several chunks of code from this book does not require permission. Selling or distributing a CD-ROM of examples from O'Reilly books *does* require permission. Answering a question by citing this book and quoting example code does not require permission. Incorporating a significant amount of example code from this book into your product's documentation *does* require permission.

We appreciate, but do not require, attribution. An attribution usually includes the title, author, publisher, and ISBN. For example: "*MySQL High Availability*, by Charles Bell, Mats Kindahl, and Lars Thalmann. Copyright 2010 Charles Bell, Mats Kindahl, and Lars Thalmann, 9780596807306."

If you feel your use of code examples falls outside fair use or the permission given above, feel free to contact us at *permissions@oreilly.com*.

We'd Like to Hear from You

Every example in this book has been tested on various platforms. The information in this book has also been verified at each step of the production process. However, mistakes and oversights can occur and we will gratefully receive details of any you find, as well as any suggestions you would like to make for future editions. You can contact the author and editors at:

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
800-998-9938 (in the United States or Canada)
707-829-0515 (international or local)
707-829-0104 (fax)

We have a web page for this book, where we list errata, examples, and any additional information. You can access this page at:

http://www.oreilly.com/catalog/9780596807306

To comment or ask technical questions about this book, send email to the following quoting the book's ISBN number (9780596807306):

bookquestions@oreilly.com

For more information about our books, conferences, Resource Centers, and the O'Reilly Network, see our website at:

http://www.oreilly.com

Safari® Books Online

Safari Books Online is an on-demand digital library that lets you easily search over 7,500 technology and creative reference books and videos to find the answers you need quickly.

With a subscription, you can read any page and watch any video from our library online. Read books on your cell phone and mobile devices. Access new titles before they are available for print, and get exclusive access to manuscripts in development and post feedback for the authors. Copy and paste code samples, organize your favorites, download chapters, bookmark key sections, create notes, print out pages, and benefit from tons of other time-saving features.

O'Reilly Media has uploaded this book to the Safari Books Online service. To have full digital access to this book and others on similar topics from O'Reilly and other publishers, sign up for free at *http://my.safaribooksonline.com*.

Acknowledgments

The authors would like to thank our technical reviewers, Mark Callaghan, Luis Soares, and Morgan Tocker. Your attention to detail and insightful suggestions were invaluable. We could not have delivered a quality book without your help.

We also want to thank our extremely talented colleagues on the MySQL replication team, including Alfranio Correia, Andrei Elkin, Zhen-Xing He, Serge Kozlov, Sven Sandberg, Luis Soares, Rafal Somla, Li-Bing Song, Ingo Strüwing, and Dao-Gang Qu for their tireless dedication to making MySQL replication the robust and powerful feature set it is today. We especially would like to thank our MySQL customer support professionals, who help us bridge the gap between our customers' needs and our own desires to improve the product. We would also like to thank the many community members who so selflessly devote time and effort to improve MySQL for everyone.

Finally, and most importantly, we would like to thank our editor, Andy Oram, who helped us shape this work, for putting up with our sometimes cerebral and sometimes over-the-top enthusiasm for all things MySQL.

Charles would like to thank his loving wife, Annette, for her patience and understanding when he was spending time away from family priorities to work on this book. You are the love of his life and his inspiration. Charles would also like to thank his many colleagues on the MySQL team at Oracle who contribute their wisdom freely to everyone on a daily basis. Finally, Charles would like to thank all of his brothers and sisters in Christ who both challenge and support him daily.

Mats would like to thank his wife, Lill, and two sons, Jon and Hannes, for their unconditional love and understanding in difficult times. You are the love of his life and he cannot imagine a life without you. Mats would also like to thank his MySQL colleagues inside and outside Sun/Oracle for all the interesting, amusing, and inspiring times together: you are truly some of the sharpest minds in the trade.

Lars would like to thank all his colleagues, current and past, who have made MySQL such an interesting place to work. In fact, it is not even a place. The distributed nature of the MySQL development team and the open-mindedness of its many dedicated developers are truly extraordinary. The MySQL community has a special spirit that makes working with MySQL an honorable task. What we have created together is remarkable. It is amazing that we started with such a small group of people and managed to build a product that services so many of the Fortune 500 companies today.

Table of Contents

Foreword xv

Preface xvii

Part I. Replication

1. Introduction 3
   - What's This Replication Stuff Anyway? 5
   - So, Backups Are Not Needed Then? 6
   - What's with All the Monitoring? 7
   - Is There Anything Else I Can Read? 8
   - Conclusion 8

2. MySQL Replication Fundamentals 11
   - Basic Steps in Replication 12
     - Configuring the Master 13
     - Configuring the Slave 15
     - Connecting the Master and Slave 15
   - A Brief Introduction to the Binary Log 17
     - What's Recorded in the Binary Log 17
     - Watching Replication in Action 18
     - The Binary Log's Structure and Content 20
     - Python Support for Managing Replication 23
     - Basic Classes and Functions 25
     - Operating System 26
     - Server Class 26
     - Server Roles 28
     - Creating New Slaves 30
     - Cloning the Master 31
     - Cloning the Slave 33

Scripting the Clone Operation 35
Performing Common Tasks with Replication 36
Reporting 37
Conclusion 43

3. **The Binary Log** **45**
Structure of the Binary Log 46
Binlog Event Structure 48
Logging Statements 50
Logging Data Manipulation Language Statements 50
Logging Data Definition Language Statements 51
Logging Queries 51
LOAD DATA INFILE Statements 57
Binary Log Filters 59
Triggers, Events, and Stored Routines 61
Stored Procedures 66
Stored Functions 69
Events 71
Special Constructions 71
Nontransactional Changes and Error Handling 72
Logging Transactions 75
Transaction Cache 76
Distributed Transaction Processing Using XA 79
Binary Log Management 81
The Binary Log and Crash Safety 82
Binlog File Rotation 83
Incidents 85
Purging the Binlog File 86
The mysqlbinlog Utility 87
Basic Usage 88
Interpreting Events 94
Binary Log Options and Variables 98
Conclusion 100

4. **Replication for High Availability** **103**
Redundancy 104
Planning 106
Slave Failures 106
Master Failures 106
Relay Failures 107
Disaster Recovery 107
Procedures 107
Hot Standby 111

Dual Masters 115
Semisynchronous Replication 124
Slave Promotion 127
Circular Replication 142
Conclusion 146

5. **MySQL Replication for Scale-Out** **147**
Scaling Out Reads, Not Writes 149
The Value of Asynchronous Replication 150
Managing the Replication Topology 152
Example of an Application-Level Load Balancer 155
Hierarchal Replication 159
Setting Up a Relay Server 160
Adding a Relay in Python 161
Specialized Slaves 162
Filtering Replication Events 163
Using Filtering to Partition Events to Slaves 165
Data Sharding 166
Shard Representation 169
Partitioning the Data 170
Balancing the Shards 171
A Sharding Example 173
Managing Consistency of Data 185
Consistency in a Nonhierarchal Deployment 185
Consistency in a Hierarchal Deployment 187
Conclusion 193

6. **Advanced Replication** **195**
Replication Architecture Basics 196
The Structure of the Relay Log 196
The Replication Threads 200
Starting and Stopping the Slave Threads 201
Running Replication over the Internet 202
Setting Up Secure Replication Using Built-in Support 204
Setting Up Secure Replication Using Stunnel 204
Finer-Grained Control over Replication 206
Information About Replication Status 206
Options for Handling Broken Connections 214
How the Slave Processes Events 215
Housekeeping in the I/O Thread 216
SQL Thread Processing 217
Slave Safety and Recovery 222
Syncing, Transactions, and Problems with Database Crashes 222

Rules for Protecting Nontransactional Statements 225
Multisource Replication 226
Row-Based Replication 229
Options for Row-Based Replication 230
Mixed-Mode Replication 231
Events for Handling Row-Based Replication 232
Event Execution 236
Events and Triggers 238
Filtering 240
Conclusion 241

Part II. Monitoring and Disaster Recovery

7. **Getting Started with Monitoring** **245**
Ways of Monitoring 246
Benefits of Monitoring 247
System Components to Monitor 247
Processor 248
Memory 249
Disk 250
Network Subsystem 251
Monitoring Solutions 252
Linux and Unix Monitoring 253
Process Activity 253
Memory Usage 259
Disk Usage 261
Network Activity 265
General System Statistics 266
Automated Monitoring with cron 268
Mac OS X Monitoring 268
System Profiler 268
Console 271
Activity Monitor 273
Microsoft Windows Monitoring 276
The Windows Experience 277
The System Health Report 278
The Event Viewer 281
The Reliability Monitor 283
The Task Manager 285
The Performance Monitor 285
Monitoring as Preventive Maintenance 288
Conclusion 288

8. **Monitoring MySQL** **291**
What Is Performance? 292
MySQL Server Monitoring 292
How MySQL Communicates Performance 293
Performance Monitoring 293
SQL Commands 294
The mysqladmin Utility 300
MySQL GUI Tools 302
MySQL Administrator 302
MySQL Query Browser 312
Server Logs 313
Third-Party Tools 316
The MySQL Benchmark Suite 318
Database Performance 319
Measuring Database Performance 320
Database Optimization Best Practices 331
Best Practices for Improving Performance 339
Everything Is Slow 340
Slow Queries 340
Slow Applications 340
Slow Replication 341
Conclusion 341

9. **Storage Engine Monitoring** **343**
MyISAM 344
Optimizing Disk Storage 344
Tuning Your Tables for Performance 345
Using the MyISAM Utilities 345
Storing a Table in Index Order 347
Compressing Tables 347
Defragmenting Tables 348
Monitoring the Key Cache 348
Preloading Key Caches 349
Using Multiple Key Caches 350
Other Parameters to Consider 351
InnoDB 352
Using the SHOW ENGINE Command 354
Using InnoDB Monitors 357
Monitoring Logfiles 359
Monitoring the Buffer Pool 360
Monitoring Tablespaces 363
Using INFORMATION_SCHEMA Tables 363
Other Parameters to Consider 365

Conclusion 366

10. Replication Monitoring 367
Getting Started 367
Server Setup 368
Inclusive and Exclusive Replication 368
Replication Threads 371
Monitoring the Master 372
Monitoring Commands for the Master 373
Master Status Variables 376
Monitoring Slaves 376
Monitoring Commands for the Slave 377
Slave Status Variables 380
Replication Monitoring with MySQL Administrator 381
Other Items to Consider 383
Networking 383
Monitor and Manage Slave Lag 383
Causes and Cures for Slave Lag 384
Conclusion 386

11. Replication Troubleshooting 387
What Can Go Wrong 388
Problems on the Master 388
Problems on the Slave 393
Advanced Replication Problems 398
Tools for Troubleshooting Replication 399
Best Practices 401
Know Your Topology 401
Check the Status of All of Your Servers 403
Check Your Logs 404
Check Your Configuration 404
Conduct Orderly Shutdowns 404
Conduct Orderly Restarts After a Failure 405
Manually Execute Failed Queries 405
Common Procedures 406
Reporting Replication Bugs 407
Conclusion 407

12. Protecting Your Investment 409
What Is Information Assurance? 410
The Three Practices of Information Assurance 410
Why Is Information Assurance Important? 411
Information Integrity, Disaster Recovery, and the Role of Backups 411