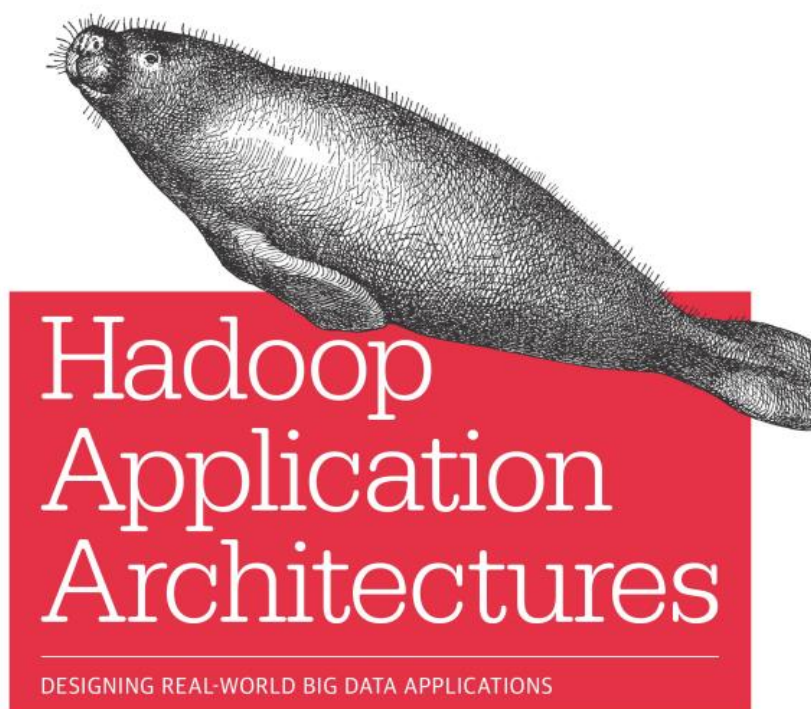


[Hadoop Application Architectures\[PDF\]](#)

Hadoop Application Architectures - Designing Real-World Big Data Applications由 O'Reilly 于2015年7月出版，共364页。

O'REILLY®

Copyrighted Material



Mark Grover, Ted Malaska,
Jonathan Seidman & Gwen Shapira

Copyrighted Material

如果想及时了

解Spark、Hadoop或者Hbase相关的文章，欢迎关注微信公共帐号：iteblog_hadoop

目录

- Chapter 1 Data Modeling in Hadoop
- Chapter 2 Data Movement
- Chapter 3 Processing Data in Hadoop
- Chapter 4 Common Hadoop Processing Patterns
- Chapter 5 Graph Processing on Hadoop

Chapter 6 Orchestration

Chapter 7 Near-Real-Time Processing with Hadoop

Chapter 8 Clickstream Analysis

Chapter 9 Fraud Detection

Chapter 10 Data Warehouse

通过本书可以学到以下知识

- Factors to consider when using Hadoop to store and model data
- Best practices for moving data in and out of the system
- Data processing frameworks, including MapReduce, Spark, and Hive
- Common Hadoop processing patterns, such as removing duplicate records and using windowing analytics
- Giraph, GraphX, and other tools for large graph processing on Hadoop
- Using workflow orchestration and scheduling tools such as Apache Oozie
- Near-real-time stream processing with Apache Storm, Apache Spark Streaming, and Apache Flume
- Architecture examples for clickstream analysis, fraud detection, and data warehousing

下载地址

本博客文章除特别声明，全部都是原创！

原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。

本文链接: **【】** ()