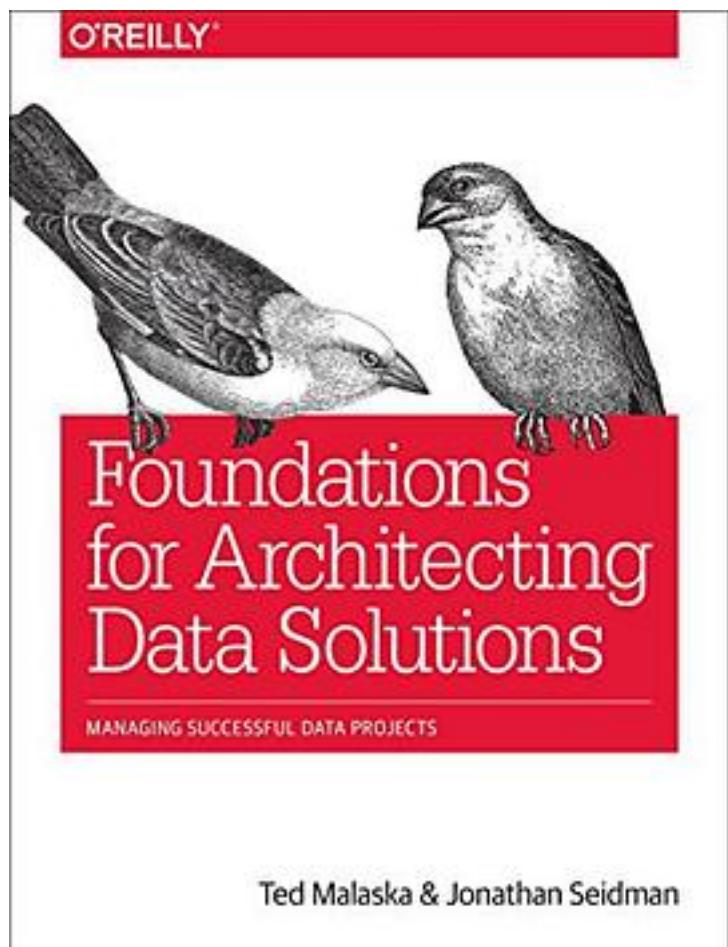


Foundations for Architecting Data Solutions



[Foundations for Architecting Data Solutions 下载链接1](#)

著者: Ted Malaska

出版者: O'Reilly

出版时间: 2018-9-30

装帧: Paperback

isbn: 9781492038740

While many companies ponder implementation details such as distributed processing engines and algorithms for data analysis, this practical book takes a much wider view of big data development, starting with initial planning and moving diligently toward

execution. Authors Ted Malaska and Jonathan Seidman guide you through the major components necessary to start, architect, and develop successful big data projects.

Everyone from CIOs and COOs to lead architects and developers will explore a variety of big data architectures and applications, from massive data pipelines to web-scale applications. Each chapter addresses a piece of the software development life cycle and identifies patterns to maximize long-term success throughout the life of your project.

Start the planning process by considering the key data project types

Use guidelines to evaluate and select data management solutions

Reduce risk related to technology, your team, and vague requirements

Explore system interface design using APIs, REST, and pub/sub systems

Choose the right distributed storage system for your big data system

Plan and implement metadata collections for your data architecture

Use data pipelines to ensure data integrity from source to final storage

Evaluate the attributes of various engines for processing the data you collect

作者介绍:

Ted Malaska is a group technical architect on the Battle.net team at Blizzard, helping support great titles like World of Warcraft, Overwatch, and HearthStone. Previously, Ted was a principal solutions architect at Cloudera, helping clients find success with the Hadoop ecosystem, and a lead architect at the Financial Industry Regulatory Authority (FINRA). He has also contributed code to Apache Flume, Apache Avro, Apache Yarn, Apache HDFS, Apache Spark, Apache Sqoop, and many more. Ted is a coauthor of *Hadoop Application Architectures*, a frequent speaker at many conferences, and a frequent blogger on data architectures.

Jonathan is a software engineer on the Cloud team at Cloudera. Prior to that, he was a solutions architect at Cloudera working with partners to integrate their solutions with Cloudera's software stack. Previously, he was a technical lead on the big data team at Orbitz Worldwide, helping to manage the Hadoop clusters for one of the most heavily trafficked sites on the internet. He's also a co-founder of the Chicago Hadoop User Group and Chicago Big Data, co-author of *Hadoop Application Architectures*, technical editor for *Hadoop in Practice*, and has spoken at a number of industry conferences on Hadoop and big data,

目录:

[Foundations for Architecting Data Solutions 下载链接1](#)

标签

软件工程

分布式

计算机

大数据

评论

似乎太简单了

在图书馆借到了, 赶紧读完. 看目录很好啊 对于第一次做system design很适合.

[Foundations for Architecting Data Solutions](#) [_下载链接1](#)

书评

[Foundations for Architecting Data Solutions](#) [_下载链接1](#)