

理层结果，因为它是基于更多数据的。当然，批处理层总是重新计算生日范围，所以随着时间的推移结果将越来越准确。可能你选择为生日推理使用的实现会有所变化，但你应该理解其思想。

从这些例子中你可以很明显地知道，为了便于合并，视图必须是结构化的。这对于像给定时间范围内的页面浏览量之类面向时间的查询是很自然的，但生日推理的示例特别添加了样本数量到视图中，以帮助合并。如何构建视图使得它们可以合并，是实现 Lambda 架构时必须做出的设计选项之一。

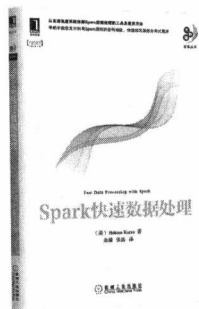
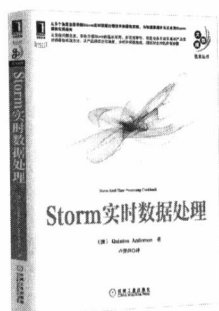
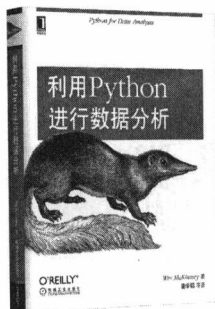
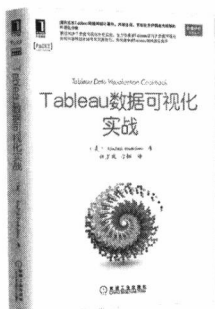
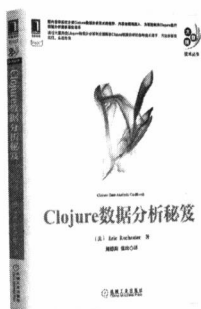
18.5 总结

Lambda 架构是从基本原理——数据问题的一般表述方式作为你已知的所有数据的函数——出发的结果，并实现了一些必需的需求，如容忍人为错误、水平可扩展性、低延迟读取和低延迟更新。我们已经探讨了 Lambda 架构，并使用了很多工具来提供核心原则的实例，比如 Hadoop、JCascalog、Kafka、Cassandra 和 Storm。我们希望大家清楚这些工具不是 Lambda 架构的必要部分。我们完全相信这些工具随着时间的推移将发生变化和发展，但 Lambda 架构的原则总是不变的。

在许多方面，Lambda 架构都超越了当前可用的工具。尽管现如今实现 Lambda 架构是非常可行的——通过本书探讨实现不同层的细节，我们尝试来说明这件事情——它当然可以更容易地实现。只有少数几个专门设计用于服务层的数据库，拥有速度层数据库将是很棒的，它可以更容易地处理视图中不再需要的过期的部分。幸运的是，构建这些工具比构建广泛的传统读/写数据库更容易，因此我们期望这些漏洞随着越来越多的人采用 Lambda 架构而得以填补。与此同时，你会发现在 Lambda 架构中，你会为不同的角色改变传统数据库的用途，并自己做一些研发使这些数据库更合用。

当第一次遇到大数据问题和大数据生态系统的工具时，人们很容易混淆和不知所措。对熟悉的关系型数据库的向往是可以理解的，关系型数据库在过去几十年里作为一个产业已经变得如此让人熟悉。希望通过学习 Lambda 架构，你已经知道了构建大数据系统可以比构建基于传统架构的系统简单得多。Lambda 架构完全解决了规范化和反规范化问题以及一些困扰传统架构的问题，并且它也有固有的容忍人为错误属性——我们认为这是不可妥协的属性。此外，它避免了基于整体读/写数据库的架构所带来的过多的复杂性。因为它基于所有数据上的方法，所以 Lambda 架构就其本性而言是通用的，这会给你带来积极处理任何数据问题的信心。

推荐阅读



■ Clojure数据分析秘笈

作者: Eric Rochester
ISBN: 978-7-111-47326-8
定价: 59.00元

■ Tableau数据可视化实战

作者: Ashutosh Nandeshwar
ISBN: 978-7-111-47283-4
定价: 39.00元

■ 利用Python进行数据分析

作者: Wes McKinney
ISBN: 978-7-111-43673-7
定价: 89.00元

■ Storm实时数据处理

作者: Quinton Anderson
ISBN: 978-7-111-46663-5
定价: 49.00元

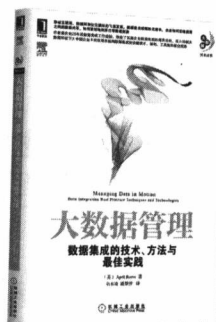
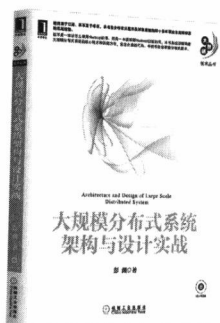
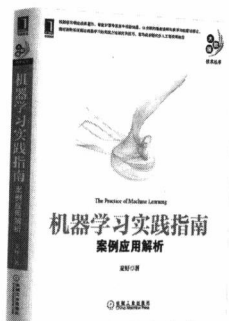
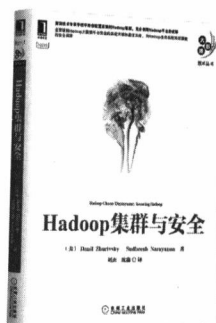
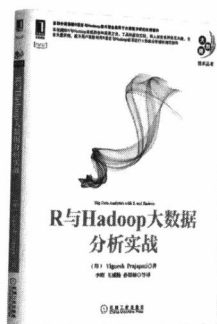
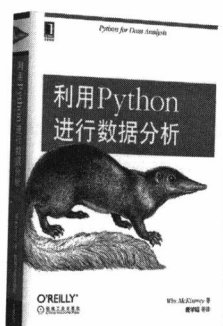
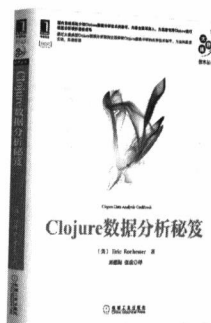
■ Splunk大数据分析

作者: Peter Zadrozny 等
ISBN: 978-7-111-46429-7
定价: 69.00元

■ Spark快速数据处理

作者: Holden Karau
ISBN: 978-7-111-46311-5
定价: 29.00元

推荐阅读



作者简介

Nathan Marz Cascalog和Storm的创始人。在2011年Twitter收购社交媒体数据分析公司BackType前，他是BackType首席工程师。在Twitter，他建立了流计算团队，提供和开发共享基础设施，为整个公司的关键实时应用提供支持。他目前是Stealth startup的创始人。

James Warren Storm8的分析架构师，精通大数据处理、机器学习和科学计算。

译者简介

马延辉 资深Hadoop技术专家，对Hadoop生态系统相关技术有着深刻的理解，在Hadoop开发和运维方面积累了丰富的经验。曾就职于阿里、Answers.com、暴风等互联网公司，从事Hadoop相关的研发和运维工作，对大数据技术的企业级落地、研发、运维和管理有着深刻的理解和丰富的实战经验。开源HBase监控工具Ella作者。现在致力于大数据技术在传统行业的落地和大数据技术的普及和推广。

向磊 前暴风影音数据平台架构师，目前在Admaster担任基础架构部架构师，惠普中国Hadoop相关课程讲师。开源项目EasyHadoop、phpHiveAdmin作者，对Hadoop及其周边生态系统的底层运维及开发、集群自动化运维、网络架构设计、集群安全、性能优化、嵌入式编程有较为深入的了解。

魏东琦 博士，长期从事软件研发工作，现就职于中国地质调查局西安地质调查中心，参加、承担过多项科研项目。现致力于地质行业与大数据技术融合的相关研究工作。

超越了个别工具或平台。任何从事大数据系统工作的人都需要阅读。

—— **Jonathan Esterhazy**, Groupon

一次全面的、样例驱动的Lambda架构之旅，由Lambda架构的发起人为您指导。

—— **Mark Fisher**, Pivotal

内含只有在经历许多大数据项目后才能获得的智慧。这是一本必须阅读的书。

—— **Pere Ferrera Bertran**, Datasalt

在批处理和近似实时处理中，简化数据管道的实际指南。

—— **Alex Holmes**, 《Hadoop实践》作者

近年来，互联网技术发展迅猛，从电子交易记录、社交网络数据分析到地震分析、分子建模，各行各业应用大数据系统的范围日益拓宽，所涉及的数据量日益“臃肿”，对处理速度的要求也日益提高，这就需要用基于硬件集群构建的架构进行存储和处理。但这种架构在提供便利的同时，也引入了大多数开发者并不熟悉的、困扰传统架构的复杂性问题。

本书教你使用一种专门设计用来获取和分析网络规模数据的架构去构建大数据系统——Lambda架构，它是一种可扩展的、易于理解的、可以被小团队用来构建和运行大数据系统的方法。除了与你分享Lambda架构的相关知识，本书还给出了相应的示例，将“理论应用于实践”，助你更好地“认识”Lambda架构，更好地将其应用到工作中。



MANNING

投稿热线: (010) 88379604

客服热线: (010) 88379426 88361066

购书热线: (010) 68326294 88379649 68995259

华章网站: www.hzbook.com

网上购书: www.china-pub.com

数字阅读: www.hzmedia.com.cn



上架指导: 计算机/大数据

ISBN 978-7-111-55294-9



9 787111 552949 >

定价: 79.00元