《大数据》

图书基本信息

书名：《大数据》

13位ISBN编号：9787121202650

10位ISBN编号：7121202654

出版时间：2013-5

出版社：电子工业出版社

作者：周宝曜，刘伟，范承工

页数：320

《大数据》

内容概要

本书从实际技术解决方案出发，提出了大数据技术四层架构，即基础设施层、管理层、分析层、应用层。在此基础上，全面剖析了当前大数据领域中的主流技术，并配以行业应用实例和一线研发人员的独到见解。力求使读者能够通过阅读此书，全面了解当前大数据技术动态和发展趋势，并可针对自己面临的大数据问题找到可行的解决方案。

《大数据》

作者简介

周宝曜，博士，现任EMC中国研究院数据科学实验塞主任，资深主任研究员。新加坡南洋理工大学计算机工程博士，清华大学学士和硕士。主要从事大数据相关的蕾理集构和分析挖掘算法的研究。曾就职于惠普中国研究院和IBM中国研究院。已发表国际学术论文30余篇，并拥有4项美国专利和多项中国专利。 刘伟，博士，现任EMC北京研发中心总经理、EMC中国研究院院长。此前蕾任思辩大中华区高级副总裁、中国研发中心高级总监。惠蕾中国研究院院长等职务。现担任CCF高级会员，中国电子学会云计算专家委员会委员，中国计算机学会大数据专家委员会委员等社会职务。美国科罗拉多州立大掌电子工程与计算机科学博士，南开大学电子学及物理学本科及硕士。 范承工，博士，是VMware中国研发中心和EMC中国卓越研发集团的创始人，现任VMware全球高级副总裁，负责领导VMware应用存储与数据产品的全球研发团队。加州理工学院电子工程硕士和博士，纽约古柏联舍学院电子工程学士。国际知名分布式系统和网络存储技术专家。曾是Rainfinity公司联台创始人之一，任首席技术官。

《大数据》

书籍目录

《大数据》

《大数据》

章节摘录

版权页： 插图： 3.4.6对大数据处理的意义 相对于普通的应用，大数据的分析与处理对网络有着更高的要求。要求涉及从带宽到延时，从吞吐率到负载均衡，以及可靠性、服务质量控制等方方面面。同时随着越来越多的大数据应用部署到云计算平台中，对虚拟网络的管理需求就越来越高。首先，网络接入设备虚拟化的发展，在保证多租户服务模式的前提下，还能同时兼顾高性能与低延时、低CPU占用率。其次，接入层的虚拟化保证了虚拟机在整个网络中的可见性，使得基于虚拟机粒度（或大数据应用粒度）的服务质量控制成为可能。覆盖网络的虚拟化，一方面使得大数据应用能够得到有效的网络隔离，更好地保证了数据通信的安全；另一方面也使得应用的动态迁移更加便捷，保证了应用的性能和可靠性。软件定义的网络更是从全局的视角来重新管理和规划网络资源，使得整体的网络资源利用率得到优化利用。总而言之，网络虚拟化技术通过对性能、可靠性和资源优化利用的贡献，间接提高了大数据系统的可靠性和运行效率。 3.5基础架构的安全：云环境中面临的新的安全挑战 云计算在安全方面引入了怎样的新问题呢？那么首先得回答云计算的本质是什么？所谓云计算，宏观来讲就是按需利用“ 远端 ”的资源（诸如CPU，内存，网络，存储等）进行“ 计算 ”，以达到和“ 本地 ”计算等价甚至更好的目的，诸如省钱，管理便捷，可扩展性强，绿色环保等。然而天下没有免费的午餐，云计算在数据安全方面引入了很多新问题。譬如在云计算基础架构服务层（Infrastructure as a Service，IaaS），主要有以下两类问题：①新的安全问题，诸如信任问题（特指租客和云服务商之间），多租客之间的资源隔离问题；②对已有的安全攻击，IaaS是否更容易被攻击？或者存在新的技术方法去避免这些攻击。 安全问题中的信任和隔离问题，源于云计算的新模型。由于资源使用者和管理者角色的分离，衍生IaaS的使用者和IaaS提供者之间的信任问题。我们把云资源的使用者，称为云租户。比如一个小型公司租赁了Amazon上的EC2服（主要指虚拟机），并在EC2上搭建了一个网站，那么这个公司就是AmazonEC2的租户，而使用网站的用户只是这个小公司的客户。由于资源不由租客完全控制，那么租客就有疑问：怎么确定租赁的资源仅仅为我所用，而不被其他租客或者云管理员非法使用，导致数据的丢失或者泄露。

《大数据》

编辑推荐

《大数据》

精彩短评

1、搞了本二手书给自己扫下盲
2、书本身内容很好，书页质量也可以，但是寄来的书是破的，都裂开了，哼
3、作为了解大数据的全景还是不错的。
4、本书适合刚接触大数据的读者
5、简单入门书，还算比较详细
6、相当不错啊，希望以后出更多的好书
7、呕心沥血啊
8、标准作品吧，结构清晰。
9、我参与写的~
10、比某名不副实的同类热门书籍要好
11、有广度有深度 Reference
12、感觉是学术论文拼凑起来的一本书。章节和章节之间连贯性不强，而且里面对某些技术的介绍还有重复。完全不考虑读者的学习曲线。都在讲些很空洞的东西，学术性的东西。感觉是本来就懂大数据的人可以拿来当个字典用。如果之前对大数据不熟，那看了等于白看。

《大数据》

精彩书评

1、一本用通俗的方式、多角度阐释“ 大数据 ”的技术普及读物；涵盖大数据应用的前沿技术，还有许多精彩应用案例哦。值得推荐！摘录：“ 大数据并不是一个准确的术语；相反，它是对各种数据永不休的积聚的一种表征。它用以描述那些呈指数级增长，并且因太大、太原始或非结构化程度太高而无法使用关系数据库方法进行分析的数据集。 ”“ 大数据需要数据科学，大数据的下一步是数据科学，要做到的不仅是存储和管理，更包括预测式的分析（比如，如果这样做，会发生什么？）。 ”
2、本书适合刚接触大数据的朋友了解大数据的相关背景和在这个领域所涉及到的相关技术和分工，类似于科研论文中的综述，写得通俗易懂，着重在广度而不是深度。对于这个行业的从业人员，可以了解自己的工作是在大数据架构中的位置。
3、刚开始介绍的大数据的出现，图文并茂；然后说大数据相关的技术架构（四层--基础架构层->数据存储管理层->数据分析层->数据应用层）；针对这四层自底而上，介绍了相关的开源技术，应该说是对这些开源工具的简单介绍和应用场景，具体的知识点(理论)还需要参考其他书籍；最后说了些案例，具体还是需要自己去实践呀。。。。总的来说有点泛泛而谈，不过了解总体框架，对于一个初学者来说，挺好的一本书，推荐（先泛而精，可能才知道如何搞大数据吧,）。 from 大数据粗学者
。

《大数据》

章节试读

1、《大数据》的笔记-第4页

通过使用高速（Velocity)的采集，发现和、或分析，从超大容量（Volume）的多样（Variety)数据中经济地提取价值（Value).

大数据快速增长的部分原因归功于智能设备的普及，比如传感器和医疗设备，以及智能建筑，比如大楼和桥梁。此外，非结构化信息，比如文件，电子邮件和视频，将占到未来10年新生数据的90%。非结构化信息的增长部分归功于高宽带数据的增长，比如视频。用户手中的手机和移动设备室数据量爆炸的一个重要原因。

《大数据》

版权说明