TURING 图灵程序设计丛书

MANNING

Hadoop in Action

Hadoop 实战

[美] Chuck Lam 著
韩冀中 译

- 纵情享受海量数据之美
- 揭开云计算的神秘面纱
- 深入分析，追本溯源

人民邮电出版社
POSTS & TELECOM PRESS

TURING 图灵程序设计丛书

Hadoop in Action

Hadoop实战

[美] Chuck Lam 著
韩冀中 译

人民邮电出版社
北京

图书在版编目（CIP）数据

Hadoop实战 / (美) 拉姆 (Lam,C.) 著 ; 韩冀中译.
-- 北京 : 人民邮电出版社, 2011.10
（图灵程序设计丛书）
书名原文: Hadoop in Action
ISBN 978-7-115-26448-0

Ⅰ. ①H… Ⅱ. ①拉… ②韩… Ⅲ. ①数据处理－应用软件－网络编程 Ⅳ. ①TP274

中国版本图书馆CIP数据核字(2011)第191474号

内 容 提 要

作为云计算所青睐的分布式架构，Hadoop是一个用Java语言实现的软件框架，在由大量计算机组成的集群中运行海量数据的分布式计算，是谷歌实现云计算的重要基石。本书分为3个部分，深入浅出地介绍了Hadoop框架、编写和运行Hadoop数据处理程序所需的实践技能及Hadoop之外更大的生态系统。

本书适合需要处理大量离线数据的云计算程序员、架构师和项目经理阅读参考。

图灵程序设计丛书

Hadoop实战

◆ 著 [美] Chuck Lam
译 韩冀中
责任编辑 卢秀丽

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
邮编 100061 电子邮件 315@ptpress.com.cn
网址 http://www.ptpress.com.cn
三河市潮河印业有限公司印刷

◆ 开本：800×1000 1/16
印张：16.75
字数：417千字 2011年10月第1版
印数：1－5 000册 2011年10月河北第1次印刷

著作权合同登记号 图字：01-2011-0806号

ISBN 978-7-115-26448-0

定价：59.00元

读者服务热线：(010)51095186转604 印装质量热线：(010)67129223

反盗版热线：(010)67171154

版 权 声 明

前　言

我很长时间里都痴迷于数据。当我还是一名电气工程本科生的时候，数字信号处理就对我产生了极大的吸引力。我发现音乐、视频、照片和很多其他的东西都可以被视为数据。数据的计算不断带来并加强了这些感性的体验。我当时认为那是最酷的事情。

随着时间的推移，我继续为数据所展现的崭新视野而欣喜。近几年社交数据和大数据崭露头角。特别是大数据，它对我而言是一个智力挑战。早先我已经学会了从统计学角度来观察数据，新的数据类型“只是”需要新的数学方法。这并不简单，但至少我已经得到过训练，了解它们所需的资源也非常丰富。另一方面，大数据涉及系统级的创新和新的编程方法。我从未得到过这样的训练，更重要的是，不只我一个人如此。有关在实践中处理大数据的知识在一定程度上是一种魔法。许多用于扩展数据处理的工具和技术都是如此，包括缓存（例如memcached）、复制及分片，当然还有MapReduce/Hadoop。近几年，我的时间都花在不断地学习这些技术上。

从个人经历看，学习这些技术最大的障碍出现在学习过程的中段。开始时，很容易找到引导性的博客和演示文稿，它们会教你如何做一个“Hello World”的示例。当足够熟悉之后，你就会知道如何在邮件列表中提问，在大小会议中邂逅专家，甚至自己阅读源代码。但在这中间存在一个巨大的知识落差，你的胃口更大了，但又不太清楚下一步该问什么问题。对Hadoop这种最新的技术而言尤为如此。需要一个有组织的说明，将你从开始的“Hello World”引领到可以从容地在实践中应用Hadoop。这就是我希望本书所做到的。幸好我发现了Manning出版社的In Action系列丛书，它们正与此目标相吻合，而且出版社有一群优秀的编辑帮助我达成目标。

我非常享受写作这本书的时光，希望它能为你开启畅游Hadoop的旅途。

致　谢

很多人为这本书提供了灵感并做出了奉献。首先我要感谢James Warren。他将分析引入RockYou，我们一起在公司上下灌输了Hadoop的使用。我从他身上学到了很多东西，他甚至还为初稿出谋划策。

我很幸运有很多人为我提供了Web 2.0行业以外的有趣案例。为此我要感谢罗治国、徐萌、孙少陵、Ken MacInnis、Ryan Rawson、Vuk Ercegovac、Rajasekar Krishnamurthy、Sriram Raghavan、Frederick Reiss、Eugene Shekita、Sandeep Tata、Shivakumar Vaithyanathan以及Huaiyu Zhu。

我也要感谢这本书的许多审阅者。他们为早期的初稿提供了有价值的反馈意见。特别是Paul O'Rorke，他虽是技术审阅人，但却提出了许多超出其职责的中肯建议，告知我如何让这份手稿更为出色。我期待有朝一日能够看到由他自己写的书。我也很享受与Jonathan Cao的长期交谈。他在数据库和大规模系统上的专业知识，为我理解Hadoop的功能提供了广阔的视野。

其他审阅者在此期间对草稿做了大量的反复阅读，多谢他们宝贵的意见，他们是：Paul Stusiak、Philipp K. Janert、Amin Mohammed-Coleman、John S. Griffin、Marco Ughetti、Rick Wagner、Kenneth DeLong、Josh Patterson、Srini Penchikala、Costantino Cerbo、Steve Loughran、Ara Abrahamian、Ben Hall、Andrew Siemer、Robert Hanson、Keith Kim、Sopan Shewale、Marion Sturtevant、Chris Chandler、Eric Raymond以及Jeroen Benckhuijsen。

我很幸运与Manning出版社很出色的一群人工作。特别感谢Troy Mott让我开始本书的撰写工作，并有足够耐心等待我把它完成。也多亏Tara Walsh、Karen Tegtmeyer、Marjan Bace、Mary Piergies、Cynthia Kane、Steven Hong、Rachel Schroeder、Katie Tennant以及Maureen Spencer。他们的支持是了不起的。我想象不出比这更好的工作团队。

不用说，所有对Hadoop及其生态系统做出贡献的人都值得称赞。Doug Cutting发起了它，Yahoo颇具远见地最早支持它。Cloudera现在将Hadoop推广给更多的企业用户。加入成长中的Hadoop社区令人兴奋不已。

最后但重要的是，我要感谢所有的朋友、家人和同事在我编写这本书时给我的支持。

作 者 在 线

购买本书可以免费访问 Manning 出版社的内部论坛，在那里可以对这本书进行评论、提出技术问题，并从作者和其他用户那里获得帮助。你可以通过网址 www.manning. com/HadoopinAction 访问并注册论坛。在注册之后，该页面会为你提供如何进入论坛、可以获得的帮助以及论坛的行为规则等信息。

Manning 出版社承诺在读者之间，以及读者和作者之间建立有意义的对话平台。这种承诺并不包含作者的参与，作者在论坛上所作的贡献依然是自愿的（且是无偿的）。我们建议你尝试向作者问一些有挑战性的问题，免得让他兴趣索然！

只要这本书在出版，作者在线论坛和先前的讨论文档都可通过出版商的网站进行访问。

关于本书

Hadoop是一个开源框架，它遵循谷歌的方法实现了MapReduce算法，用以查询在互联网上分布的数据集。这个定义自然会导致一个明显的问题：什么是map（映射），为什么它们需要被reduce（归约）？使用传统机制分析和查询大规模数据集会非常困难，当查询自身很复杂时尤为如此。实际上，MapReduce算法将查询操作和数据集都分解为组件——这就是映射。在查询中被映射的组件可以被同时处理（即归约）从而快速地返回结果。

这本书教会读者如何使用Hadoop并编写MapReduce程序。目标读者为不得不离线处理大量数据的程序员、架构师和项目经理。这本书引导读者去获得Hadoop的一个副本、在集群中安装并编写数据分析程序。

在书的开篇，为了让Hadoop和MapReduce的基本理念更易于掌握，本书在Hadoop的默认安装上运行了几个易于理解的任务，例如文档正文中词频变化的分析。然后在使用Hadoop开发MapReduce应用的过程中，探究其基本概念，包括仔细观察框架的组成、Hadoop在各种数据分析中的使用以及Hadoop实战中的大量实例。

MapReduce是一个在概念和实现上都很复杂的想法，要了解运行Hadoop的方方面面对于用户而言是一个挑战。本书除带给你Hadoop的运行机理之外，还会教你在MapReduce框架下写出有意义的程序。

本书假定读者基本掌握了Java，因为大多数代码示例是用Java写的。熟悉基本的统计学概念（如直方图和相关）将有助于读者理解更高级的数据处理示例。

路线图

本书将12章划分为3个部分。

第一部分的3章介绍了Hadoop的框架，涵盖我们理解并使用Hadoop所需的基础知识。这些章节描述了构成一个Hadoop集群的硬件组件，以及建立一个可运行系统的安装及配置方法。第一部分还从高层描述了MapReduce框架，并让你能编写和运行第一个MapReduce程序。

第二部分包含5章，给出编写和运行Hadoop数据处理程序所需的实践技能。在这些章节中，我们将探讨使用Hadoop分析专利数据集的各种实例，包括Bloom filter这样的先进算法。我们还将给出对生产环境下使用Hadoop极其有用的编程和管理技术。

第三部分被称为“Hadoop也疯狂”，包含这本书的最后4章，将探讨Hadoop之外更大的生态系统。云服务提供了创建Hadoop集群的另一种方案，可以替代那种由自己购买并拥有硬件集群的

方式。许多附加产品包在MapReduce之上提供了更高级别的编程抽象。最后，我们会看到几个用Hadoop解决实际业务问题的案例。

附录包含HDFS命令的列表及其说明和使用方法。

编码约定及代码下载

所有列表或文本中的源代码都是用固定宽度字体与普通文本相区别的。许多代码清单中都给出了代码注释，重要的概念被突出地显示。有时，随代码清单还会给出由数字符号相连的注释。

本书的示例代码可从Manning出版社的网站www.manning.com/HadoopinAction上下载。

关于封面图

本书封面上的图为“一个来自达尔马提亚Kistanja的年轻人”。该图取自克罗地亚19世纪中叶传统服饰影集的一个副本，作者为尼古拉·阿尔塞诺维奇，由Ethnographic博物馆在2003年于克罗地亚的斯普利特出版。该图得自于一位乐于助人的斯普利特Ethnographic博物馆馆员，这个博物馆位于该城镇在中世纪罗马时的核心位置，是公元304年左右罗马皇帝戴克里先的宫殿遗址。这本书包含来自克罗地亚不同地域的颜色精美的插图，附有服饰和日常生活的说明。

Kistanja是一个小镇，位于克罗地亚的布科维卡地区。它坐落在达尔马提亚北部，有悠久的罗马和威尼斯的历史。在克罗地亚，“mamok”一词是指单身汉、花花公子或求婚者（一个在求爱年龄的年轻男人），在封面上的这个年轻人看起来干净利落，很明显正穿着他最好的衣服，小巧玲珑的白色亚麻衬衫，色彩鲜艳的绣花背心，这样的衣服他只有在去教堂和节日时才会穿——或者是去约会一位年轻女士。

过去200年间，着装和生活方式已经发生变化，曾经如此丰富的地域多样性已渐渐消失了。现在，各大洲的居民已经很难分辨，更遑论分隔只有几英里的不同村庄或城镇的人。也许我们用文化差异换来了一个更丰富的个人生活——必然是更为多样和快节奏的技术生活。

Manning出版社取材此类古老书籍中的插图，用两个世纪前丰富多样的地域生活来制作书的封面，用以庆贺计算机行业的创造性和主动性。

目　录

Part 1 第一部分

Hadoop——一种分布式编程框架

本书第一部分所介绍的是理解与使用 Hadoop 的基础。首先描述 Hadoop 集群的硬件构成及系统安装与配置，然后从高层次阐述 MapReduce 框架，并让你的第一个 MapReduce 程序运行起来。

本部分内容

第1章 Hadoop简介

本章内容

- 编写可扩展、分布式的数据密集型程序的基础知识
- 理解Hadoop和MapReduce
- 编写和运行一个基本的MapReduce程序

今天，我们正被数据所包围。人们上载视频、用手机照相、发短信给朋友、更新Facebook、网上留言及点击广告等，这使得机器产生和保留了越来越多的数据。你甚至此时此刻可能正在自己的电脑屏幕上阅读本书的电子版，并且可以确定的是，你在书店购买本书的记录已经被存为数据。[①]

数据的指数级增长首先向谷歌、雅虎、亚马逊和微软等这些处于市场领导地位的公司提出了挑战。它们需要遍历TB级和PB级数据来发现哪些网站更受欢迎，哪些书有需求，哪种广告吸引人。现有工具正变得无力处理如此大的数据集。谷歌率先推出了MapReduce，这是个用来应对其数据处理需求的系统。这个系统引起了广泛的关注，因为许多其他的企业同样面临数据膨胀的挑战，并且不是每个人都能够为自己重新量身订制一个专有的工具。Doug Cutting看到了机会并且领导开发了一个开源版本的MapReduce，称为Hadoop。随后，雅虎等公司纷纷响应，为其提供支持。今天，Hadoop已经成为许多互联网公司基础计算平台的一个核心部分，如雅虎、Facebook、LinkedIn和Twitter。许多传统的行业，如传媒业和电信业，也正在开始采用这个系统。我们将在第12章的案例中介绍《纽约时报》、中国移动和IBM等公司如何使用Hadoop。

Hadoop及大规模分布式数据处理，正在迅速成为许多程序员的一项重要技能。关系数据库、网络和安全这些在几十年前被认为是程序员可选技能的知识，今天已经成为一个高效程序员的必修课。同样，基本理解分布式数据处理将很快成为每个程序员的工具箱中不可或缺的一部分。斯坦福和卡内基–梅隆等一流的大学已经开始将Hadoop引入他们的计算机科学课程。这本书将会帮助你，一名执业的程序员，快速掌握Hadoop并用它来处理你的数据集。

---

① 当然，你读的是本正版书，对吗？

本章正式介绍Hadoop，找出它在分布式系统和数据处理系统方面的定位，并概述MapReduce编程模型。我们基于现有工具实现一个简单的单词统计示例，来彰显大型数据处理的挑战。然后，在使用Hadoop实现该示例之后，你会深刻体会Hadoop的简洁明了。我们还将讨论Hadoop的历史以及人们对MapReduce范式的一些观点。不过，让我先简单介绍一下为什么我写这本书，以及它为什么对你有用。

1.1 为什么写《Hadoop 实战》

实话实说，我第一次接触Hadoop即被其强大的能力所吸引，但随后在编写基本例程时却经历了一段令人沮丧的过程。虽然Hadoop官方网站上的文档相当全面，但是为简单的疑问找到直截了当的解答却并不总是那么容易。

写作本书的目的就是要解决这个问题。我不会关注过多的细节，相反，我提供的信息会有助于你快速创建可用代码，并会涉及在实践中最常遇到的更高级的话题。

1.2 什么是 Hadoop

按照正式的定义，Hadoop是一个开源的框架，可编写和运行分布式应用处理大规模数据。分布式计算是一个宽泛并且不断变化的领域，但Hadoop与众不同之处在于以下几点。

- 方便——Hadoop运行在由一般商用机器构成的大型集群上，或者如亚马逊弹性计算云（EC2）等云计算服务之上。
- 健壮——Hadoop致力于在一般商用硬件上运行，其架构假设硬件会频繁地出现失效。它可以从容地处理大多数此类故障。
- 可扩展——Hadoop通过增加集群节点，可以线性地扩展以处理更大的数据集。
- 简单——Hadoop允许用户快速编写出高效的并行代码。

Hadoop的方便和简单让其在编写和运行大型分布式程序方面占尽优势。即使是在校的大学生也可以快速、廉价地建立自己的Hadoop集群。另一方面，它的健壮性和可扩展性又使它胜任雅虎和Facebook最严苛的工作。这些特性使Hadoop在学术界和工业界都大受欢迎。

图1-1解释了如何与Hadoop集群交互。Hadoop集群是在同一地点用网络互连的一组通用机器。数据存储和处理都发生在这个机器“云”中[①]。不同的用户可以从独立的客户端提交计算“作业”到Hadoop，这些客户端可以是远离Hadoop集群的个人台式机。

并非所有分布式系统的构建都如图1-1所示的一样。下面，我们简要介绍一下其他的分布式系统，以便更好地展现Hadoop所依据的设计理念。

① 虽非绝对必要，但通常在一个Hadoop集群中的机器都是相对同构的x86 Linux服务器。而且它们几乎总是位于同一个数据中心，并通常在同一组机架里。

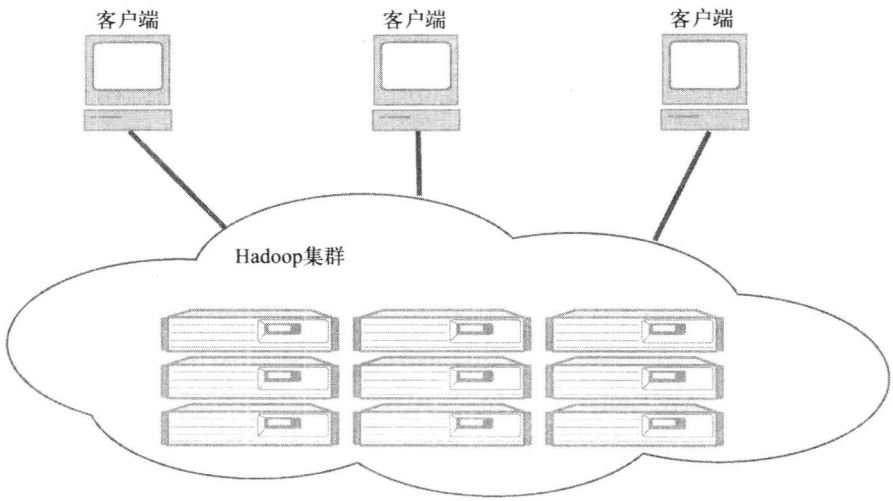

图1-1　一个Hadoop集群拥有许多并行的计算机，用以存储与处理大规模数据集。客户端计算机发送作业到计算云并获得结果

1.3　了解分布式系统和Hadoop

摩尔定律在过去几十年间对我们都是适用的，但解决大规模计算问题却不能单纯依赖于制造越来越大型的服务器。有一种替代方案已经获得普及，即把许多低端/商用的机器组织在一起，形成一个功能专一的分布式系统。

为了理解盛行的分布式系统（俗称向外扩展）与大型单机服务器（俗称向上扩展）之间的对比，需要考虑现有I/O技术的性价比。对于一个有4个I/O通道的高端机，即使每个通道的吞吐量各为100 MB/sec，读取4 TB的数据集也需要3个小时！而利用Hadoop，同样的数据集会被划分为较小的块（通常为64 MB），通过Hadoop分布式文件系统（HDFS）分布在集群内多台机器上。使用适度的复制，集群可以并行读取数据，进而提供很高的吞吐量。而这样一组通用机器比一台高端服务器更加便宜！

前面的解释充分展示了Hadoop相对于单机系统的效率。现在让我们将Hadoop与其他分布式系统架构进行比较。一个众所周知的方法是SETI @ home，它利用世界各地的屏保来协助寻找外星生命。在SETI @ home，一台中央服务器存储来自太空的无线电信号，并在网上发布给世界各地的客户端台式机去寻找异常的迹象。这种方法将数据移动到计算即将发生的地方（桌面屏保）。经过计算后，再将返回的数据结果存储起来。

Hadoop在对待数据的理念上与SETI@home等机制不同。SETI @ home需要客户端和服务器之间重复地传输数据。这虽能很好地适应计算密集型的工作，但处理数据密集型任务时，由于数据规模太大，数据搬移变得十分困难。Hadoop强调把代码向数据迁移，而不是相反。参考图1-1，我们看到Hadoop的集群内部既包含数据又包含计算环境。客户端仅需发送待执行的MapReduce程序，而这些程序一般都很小（通常为几千字节）。更重要的是，代码向数据迁移的理念被应用