Hadoop 2.X

HDFS源码剖析

徐鹏 著

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

本书以 Hadoop 2.6.0 源码为基础，深入剖析了 HDFS 2.X 中各个模块的实现细节，包括 RPC 框架实现、Namenode 实现、Datanode 实现以及 HDFS 客户端实现等。本书一共有 5 章，其中第 1 章从总体上介绍了 HDFS 的组件、概念以及典型的流程，同时详细介绍了 HDFS 各个组件间 RPC 接口的定义。第 2 章介绍了 Hadoop RPC 框架的实现，Hadoop RPC 是 HDFS 各个组件间通信所依赖的底层框架，可以理解为 HDFS 的神经系统。第 3~5 章分别介绍了 Namenode、Datanode 以及 HDFS 客户端这三个组件的实现细节，同时穿插介绍了 HDFS 2.X 的新特性，例如 Namenode HA、Federation Namenode 等。

阅读本书可以帮助读者从架构设计与源码实现角度了解 HDFS 2.X，同时还能学习 HDFS 2.X 框架中优秀的设计思想、设计模式、Java 语言技巧以及编程规范等。这些对于读者全面提高自己的技术水平有很大的帮助。

图书在版编目（CIP）数据

Hadoop 2.X HDFS 源码剖析 / 徐鹏著. —北京：电子工业出版社，2016.3
ISBN 978-7-121-28155-6

Ⅰ. ①H… Ⅱ. ①徐… Ⅲ. ①分布式文件系统—研究 Ⅳ. ①TP316

中国版本图书馆 CIP 数据核字（2016）第 027311 号

策划编辑：张春雨
责任编辑：葛 娜
印 刷：三河市双峰印刷装订有限公司
装 订：三河市双峰印刷装订有限公司
出版发行：电子工业出版社
北京市海淀区万寿路 173 信箱 邮编：100036
开 本：787×980 1/16 印张：35.25 字数：879 千字
版 次：2016 年 3 月第 1 版
印 次：2016 年 3 月第 1 次印刷
定 价：108.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

前　言

今天 Hadoop 已经成为大数据处理中十分重要的平台，一个以 Hadoop 为基础的活跃的开源生态圈已经逐步形成，Hadoop 的应用也由最初的互联网行业发展到金融行业、电信行业、IT 设备商以及数以万计的中小企业。Hadoop 的 HDFS 组件（Hadoop 分布式文件系统）以及 MapReduce 组件分别为上层框架提供了分布式存储和计算的能力。

HDFS 作为 Hadoop 中解决分布式存储的基础组件，最早是根据 GFS（Google File System）论文的概念模型来设计实现的。然而，随着 HDFS 上层框架的丰富以及应用场景的扩展，用户对 HDFS 的功能、性能、稳定性、扩展性以及可靠性的要求越来越高，HDFS 2.X 版本也就应运而生。相对于 HDFS 1.X，HDFS 2.X 版本提出了很多振奋人心的新特性，如 Namenode HA、Federation Namenode、集中式缓存、快照等。但令人惋惜的是，至今还没有一本能够深入剖析 HDFS 2.X 内部实现细节，以及介绍 HDFS 2.X 新特性的书籍。本书的出现填补了上述空白，它是国内第一本深入剖析 HDFS 2.X 源码实现的书籍。

本书以 Hadoop 2.6.0 源码为基础，深入剖析了 HDFS 2.X 中各个模块的实现细节，包括 RPC 框架实现、Namenode 实现、Datanode 实现以及 HDFS 客户端实现等。阅读本书可以帮助读者从架构设计与源码实现角度了解 HDFS 2.X，同时还能学习 HDFS 2.X 框架中优秀的设计思想、设计模式、Java 语言技巧以及编程规范等。这些对于读者全面提高自己的技术水平有很大的帮助。

如何阅读本书

由于篇幅原因，本书并没有介绍 HDFS 实现中的一些基础知识，例如 Java NIO、动态代理、protobuf 等。而是直接切入源码分析 HDFS 的设计与实现，同时介绍了一些经典的设计模式、Java 语言技巧在 HDFS 实现中的应用。希望读者在阅读本书之前，先搭建好源码环境，并了解相应的基础知识，这样学习效果会更好。

本书一共有 5 章，相互之间的联系比较紧密，有联系的小节都有注释标注，读者可以根据注释跳跃阅读。

第 1 章是 HDFS 概述，从总体上介绍了 HDFS 的组件、概念以及典型的流程，同时详细介绍了 HDFS 各个组件间 RPC 接口的定义。由于 HDFS 流程大都比较复杂，往往涉及多个组件的配合，读者在阅读后续的具体章节时，可以查阅本章内容获取某个流程的总体描述，以

及 RPC 接口的具体定义。

第 2 章介绍了 Hadoop RPC 框架的实现，Hadoop RPC 是 HDFS 各个组件间通信所依赖的底层框架，可以理解为 HDFS 的神经系统。通过阅读本章，读者可以学习到一个典型的分布式 RPC 框架的实现细节，在本章中会介绍较多的设计模式，编程模型以及语言技巧，请读者注重积累。

第 3~5 章分别介绍了 Namenode、Datanode 以及 HDFS 客户端这三个组件的实现细节，同时穿插介绍了 HDFS 2.X 的新特性，例如 Namenode HA 就放在 Namenode 章介绍，而 Federation Namenode 新特性对 Datanode 的修改比较多，所以就放在 Datanode 章介绍。

读者在阅读本书的过程中，如果发现有任何不当之处，烦请您将意见和建议发往邮箱 xupeng.bupt@gmail.com，不胜感激。

本书代码

本书分析的代码版本为 Hadoop 2.6.0，书中部分较长的代码做了省略，完整的代码请从官网 http://hadoop.apache.org 下载。

致谢

感谢互联网，感谢开源软件，感谢 Hadoop 社区，感谢本书引用文献的所有原作者，是你们为 Hadoop 爱好者打开了一扇大门。

感谢电子工业出版社博文视点的张春雨老师，是您的信任使得这本书的出版成为可能。同时还要感谢许多我不知道名字的编辑为本书最终出版所做的付出和努力。

感谢丁雷在百忙之中抽出时间对本书提出许多建设性意见，同时感谢左谱军、张德阳、闫飞翔以及张涛对本书的审阅。

2015 年写这本书时正是自己很困难的一段时期，我很感恩有许多朋友在生活、工作上给我帮助以及包容。感谢吴佳宁、刘文博，你们一直是第一个伸出援手的哥们。感谢郑晓彤、袁玮、贺子昂、李强、和紫东、刘丹、何一舟为我引荐机会，谢谢你们。

特别感谢远见，书的撰写过程是如此漫长，是你在这段时间里把自信、阳光和快乐传播给我，让我更加积极、勇敢和有信心。没有你，这本书永远无法完成。

最后感谢我的父母和妹妹，谢谢你们默默为我做出的牺牲和付出，你们永远是我前进的动力。

目　录