

# 《Hadoop 2.X HDFS源码剖析》

## 图书基本信息

书名：《Hadoop 2.X HDFS源码剖析》

13位ISBN编号：9787121281554

出版时间：2016-3

作者：徐鹏

页数：516

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：[www.tushu111.com](http://www.tushu111.com)

# 《Hadoop 2.X HDFS源码剖析》

## 内容概要

《Hadoop 2.X HDFS源码剖析》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。《Hadoop 2.X HDFS源码剖析》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《Hadoop 2.X HDFS源码剖析》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。

# 《Hadoop 2.X HDFS源码剖析》

## 作者简介

徐鹏 2005-2012年 在北京邮电大学完成本科以及硕士的学习，目前就职于今日头条担任基础架构研发工程师。长期关注大数据处理、分布式系统的研究以及Hadoop相关技术的应用与开发。

## 书籍目录

### 第1章 HDFS

1

#### 1.1 HDFS概述

1

##### 1.1.1 HDFS体系结构

1

##### 1.1.2 HDFS基本概念

2

#### 1.2 HDFS通信协议

4

##### 1.2.1 Hadoop RPC接口

4

##### 1.2.2 流式接口

20

#### 1.3 HDFS主要流程

22

##### 1.3.1 HDFS客户端读流程

22

##### 1.3.2 HDFS客户端写流程

24

##### 1.3.3 HDFS客户端追加写流程

25

##### 1.3.4 Datanode启动、心跳以及执行名字节点指令流程

26

##### 1.3.5 HA切换流程

27

### 第2章 Hadoop RPC

29

#### 2.1 概述

29

##### 2.1.1 RPC框架概述

29

##### 2.1.2 Hadoop RPC框架概述

30

#### 2.2 Hadoop RPC的使用

36

##### 2.2.1 Hadoop RPC使用概述

36

##### 2.2.2 定义RPC协议

40

##### 2.2.3 客户端获取Proxy对象

45

##### 2.2.4 服务器获取Server对象

54

#### 2.3 Hadoop RPC实现

63

##### 2.3.1 RPC类实现

63	
2.3.2	Client类实现
64	
2.3.3	Server类实现
76	
第3章	Namenode ( 名字节点 )
88	
3.1	文件系统目录树
88	
3.1.1	INode相关类
89	
3.1.2	Feature相关类
102	
3.1.3	FSEditLog类
117	
3.1.4	FSImage类
138	
3.1.5	FSDirectory类
158	
3.2	数据块管理
162	
3.2.1	Block、Replica、BlocksMap
162	
3.2.2	数据块副本状态
167	
3.2.3	BlockManager类 ( done )
177	
3.3	数据节点管理
211	
3.3.1	DatanodeDescriptor
212	
3.3.2	DatanodeStorageInfo
214	
3.3.3	DatanodeManager
217	
3.4	租约管理
233	
3.4.1	LeaseManager.Lease
233	
3.4.2	LeaseManager
234	
3.5	缓存管理
246	
3.5.1	缓存概念
247	
3.5.2	缓存管理命令
247	
3.5.3	HDFS集中式缓存架构
247	

3.5.4 CacheManager类实现	248
3.5.5 CacheReplicationMonitor	250
3.6 ClientProtocol实现	251
3.6.1 创建文件	251
3.6.2 追加写文件	254
3.6.3 创建新的数据块	257
3.6.4 放弃数据块	265
3.6.5 关闭文件	266
3.7 Namenode的启动和停止	268
3.7.1 安全模式	268
3.7.2 HDFS High Availability	276
3.7.3 名字节点的启动	301
3.7.4 名字节点的停止	306
第4章 Datanode (数据节点)	307
4.1 Datanode逻辑结构	307
4.1.1 HDFS 1.X架构	307
4.1.2 HDFS Federation	308
4.1.3 Datanode逻辑结构	310
4.2 Datanode存储	312
4.2.1 Datanode升级机制	312
4.2.2 Datanode磁盘存储结构	315
4.2.3 DataStorage实现	317
4.3 文件系统数据集	334
4.3.1 Datanode上数据块副本的状态	335
4.3.2 BlockPoolSlice实现	

335	
4.3.3	FsVolumeImpl实现
342	
4.3.4	FsVolumeList实现
345	
4.3.5	FsDatasetImpl实现
348	
4.4	BlockPoolManager
375	
4.4.1	BPServiceActor实现
376	
4.4.2	BPOfferService实现
389	
4.4.3	BlockPoolManager实现
396	
4.5	流式接口
398	
4.5.1	DataTransferProtocol定义
398	
4.5.2	Sender和Receiver
399	
4.5.3	DataXceiverServer
403	
4.5.4	DataXceiver
406	
4.5.5	读数据
408	
4.5.6	写数据 ( done )
423	
4.5.7	数据块替换、数据块拷贝和读数据块校验
437	
4.5.8	短路读操作
437	
4.6	数据块扫描器
437	
4.6.1	DataBlockScanner实现
438	
4.6.2	BlockPoolSliceScanner实现
439	
4.7	DirectoryScanner
442	
4.8	DataNode类的实现
443	
4.8.1	DataNode的启动
444	
4.8.2	DataNode的关闭
446	
第5章	HDFS客户端
447	

5.1 DFSCient实现	447
5.1.1 构造方法	448
5.1.2 关闭方法	449
5.1.3 文件系统管理与配置方法	450
5.1.4 HDFS文件与目录操作方法	451
5.1.5 HDFS文件读写方法	452
5.2 文件读操作与输入流	452
5.2.1 打开文件	452
5.2.2 读操作——DFSInputStream实现	461
5.3 文件短路读操作	481
5.3.1 短路读共享内存	482
5.3.2 DataTransferProtocol	484
5.3.3 DFSCient短路读操作流程	488
5.3.4 Datanode短路读操作流程	509
5.4 文件写操作与输出流	512
5.4.1 创建文件	512
5.4.2 写操作——DFSOutputStream实现	516
5.4.3 追加写操作	543
5.4.4 租约相关	546
5.4.5 关闭输出流	548
5.5 HDFS常用工具	549
5.5.1 FsShell实现	550
5.5.2 DFSAdmin实现	552

## 《Hadoop 2.X HDFS源码剖析》

### 精彩短评

- 1、深入浅出，能够清晰了解hadoop原理。
- 2、迅速翻阅中，主要目的还是了解原理，不太关心细枝末节的实现。。。千万要改掉“一言不合就撸源码”的习惯啊!!!
- 3、作者从HDFS的主要流程，谈到了Hadoop RPC的具体调用细节，再从NameNode和DataNode的两个角度讲两者之间的交互，很细致，读完之后，确实觉得对自己现在的开发工作有帮助。
- 4、2.x的一些新特性都包括了
- 5、深入浅出，很棒！
- 6、相当不错的书，深入分析了hadoop 2.x HDFS源码，绝对值得一读。
- 7、对Hadoop源码的讲解深入浅出，非常适合想学习Hadoop源码的小伙伴们，墙裂推荐！
- 8、很好的一本书，讲解清晰，强烈推荐！
- 9、赞！终于有一本书可以彻底解答我对hadoop的疑惑了。作者深入浅出，既有对源码落尽严谨的分析又有自己实际项目经验的传授。收获很大！
- 10、书很厚，讲的很清楚，看了一部分，好几次有种恍然大悟的感觉。。导师推荐的，一定要啃下来！
- 11、好书强推，讲解清晰，深入浅出，一定要读！
- 12、深入浅出，赞一个
- 13、相当用心的作品！
- 14、作者对hadoop理解很到位，书不错，讲得很清楚，学习Hadoop的同学不要错过，很值得一读！

## 版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:[www.tushu111.com](http://www.tushu111.com)