

高可用性的HDFS



[高可用性的HDFS 下载链接1](#)

著者:文艾

出版者:清华大学出版社

出版时间:2012-5

装帧:平装

isbn:9787302282587

《高可用性的HDFS:Hadoop分布式文件系统深度实践》专注于Hadoop分布式文件系统（HDFS）的主流HA解决方案，内容包括：HDFS元数据解析、Hadoop元数据备份方案、Hadoop Backup Node方案、AvatarNode解决方案以及最新的HA解决方案Cloudrea HA Name Node等。其中有关Backup Node方案及AvatarNode方案的内容是该书重点，尤其是对AvatarNode方案从运行机制到异常处理方案的步骤进行了详尽介绍，同时还总结了各种异常情况下AvatarNode的各种处理方案。

《高可用性的HDFS:Hadoop分布式文件系统深度实践》从代码入手并结合情景分析、

案例解说对HDFS的元数据以及主流的HDFS HA解决方案的运行机制进行了深入剖析，力求使读者在解决问题时做到心中有数，不仅知其然还知其所以然。

《高可用性的HDFS:Hadoop分布式文件系统深度实践》光盘包含本书部分操作的视频教程以及所有源代码、脚本等开发文件。

《高可用性的HDFS:Hadoop分布式文件系统深度实践》读者主要为云计算相关领域的研发人员、云计算系统管理维护人员，也适合作为高校研究生和高年级本科生的专业课辅助教材。

作者介绍:

目录: 目录

第1章 HDFS HA及解决方案 1

1.1 HDFS系统架构 2

1.2 HA定义 3

1.3 HDFS HA原因分析及应对措施 4

1.3.1 可靠性 4

1.3.2 可维护性 5

1.4 现有HDFS HA解决方案 5

1.4.1 Hadoop的元数据备份方案 6

1.4.2 Hadoop的SecondaryNameNode方案 7

1.4.3 Hadoop的Checkpoint ode方案 7

1.4.4 Hadoop的BackupNode方案 8

1.4.5 DRDB方案 9

1.4.6 FaceBook的AvatarNode方案 10

1.5 方案优缺点比较 10

第2章 HDFS元数据解析 13

2.1 概述 14

2.2 内存元数据结构 14

2.2.1 INode 15

2.2.2 Block 16

2.2.3 BlockInfo和DatanodeDescriptor 17

2.2.4 小结 17

2.2.5 代码分析——元数据结构 18

2.3 磁盘元数据文件 24

2.4 Format情景分析 27

2.5 元数据应用场景分析 45

第3章 Hadoop的元数据备份方案 47

3.1 运行机制分析 48

3.1.1 NameNode启动加载元数据情景分析 50

3.1.2 元数据更新及日志写入情景分析 64

3.1.3 Checkpoint过程情景分析 73

3.1.4 元数据可靠性机制 109

3.1.5 元数据一致性机制 110

3.2 使用说明 110

第4章 Hadoop的Backup Node方案 113

4.1 Backup Node概述 114

4.1.1 系统架构 115

4.1.2 使用原则 115

4.1.3 优缺点 116

- 4.2 运行机制分析 116
 - 4.2.1 启动流程 117
 - 4.2.2 元数据操作情景分析 141
 - 4.2.3 日志池 (journal spool) 机制 151
 - 4.2.4 故障切换机制 156
- 4.3 实验方案说明 158
- 4.4 构建实验环境 158
 - 4.4.1 网络拓扑 159
 - 4.4.2 系统安装及配置 160
 - 4.4.3 安装JDK 170
 - 4.4.4 虚拟机集群架设 171
 - 4.4.5 NameNode安装及配置 173
 - 4.4.6 Backup Node安装及配置 173
 - 4.4.7 Data Node安装及配置 174
 - 4.4.8 Clients安装及配置 175
- 4.5 异常解决方案 175
 - 4.5.1 异常情况分析 175
 - 4.5.2 NameNode配置 175
 - 4.5.3 Backup Node配置 182
 - 4.5.4 Data Node配置 185
 - 4.5.5 NameNode宕机切换实验 189
 - 4.5.6 NameNode宕机读写测试 196
- 第5章 AvatarNode运行机制 205
 - 5.1 方案说明 206
 - 5.1.1 系统架构 206
 - 5.1.2 思路分析 208
 - 5.1.3 性能数据 209
 - 5.2 元数据分析 209
 - 5.2.1 类FSNamesystem 210
 - 5.2.2 类FSDirectory 210
 - 5.2.3 AvatarNode的磁盘元数据文件 211
 - 5.3 AvatarNode Primary启动过程 211
 - 5.4 AvatarNode Standby启动过程 217
 - 5.4.1 AvatarNode的构造方法 217
 - 5.4.2 Standby线程的run()方法 218
 - 5.4.3 Ingest线程的run()方法 220
 - 5.4.4 Ingest线程的ingestFSEdits ()方法 220
 - 5.4.5 Standby线程的doCheckpoint()方法 221
 - 5.5 用户操作情景分析 223
 - 5.5.1 创建目录情景分析 223
 - 5.5.2 创建文件情景分析 231
 - 5.6 AvatarNode Standby故障切换过程 240
 - 5.7 元数据一致性保证机制 242
 - 5.7.1 元数据目录树信息 242
 - 5.7.2 Data Node与Block数据块映射信息 243
 - 5.8 Block更新同步问题 246
 - 5.8.1 问题描述 246
 - 5.8.2 结论 246
 - 5.8.3 源码分析 246
- 第6章 AvatarNode使用 253
 - 6.1 方案说明 254
 - 6.1.1 网络拓扑 254
 - 6.1.2 操作系统安装及配置 255
 - 6.2 使用Avatar打补丁版本 255

- 6.2.1 Hadoop源码联机Build 256
- 6.2.2 Hadoop源码本地Build 262
- 6.2.3 NFS服务器构建 264
- 6.2.4 Avatar分发与部署 267
- 6.2.5 Primary (namenode0) 节点配置 269
- 6.2.7 Data Node节点配置 276
- 6.2.8 Client节点配置 278
- 6.2.9 创建目录 279
- 6.2.10 挂载NFS 280
- 6.2.11 启动Ucarp 280
- 6.2.12 格式化 281
- 6.2.13 系统启动 281
- 6.2.14 检查 282
- 6.2.15 NameNode失效切换写文件实验 283
- 6.2.16 NameNode失效切换读文件实验 291
- 6.3 Avatar FaceBook版本的使用 294
- 6.3.1 Hadoop FaceBook版本安装 294
- 6.3.2 节点配置 295
- 6.3.3 启动HDFS 300
- 6.3.4 NameNode失效切换 302
- 第7章 AvatarNode异常解决方案 305
- 7.1 测试环境 306
- 7.2 Primary失效 306
- 7.2.1 解决方案 306
- 7.2.2 写操作实验步骤 307
- 7.2.3 改进写操作机制 313
- 7.2.4 读操作实验步骤 313
- 7.2.5 小结 317
- 7.3 Standby失效 317
- 7.4 NFS失效 (数据未损坏) 317
- 7.4.1 解决方案 317
- 7.4.2 写操作实验步骤 318
- 7.4.3 读操作实验步骤 320
- 7.4.4 小结 322
- 7.5 NFS失效 (数据已损坏) 323
- 7.5.1 解决方案 323
- 7.5.2 写操作实验步骤 324
- 7.5.3 读操作实验步骤 327
- 7.5.4 小结 330
- 7.6 Primary先失效, NFS后失效 (数据未损坏) 331
- 7.6.1 解决方案 331
- 7.6.2 写操作实验步骤 331
- 7.6.3 读操作实验步骤 333
- 7.6.4 小结 334
- 7.7 Primary先失效 (数据未损坏), NFS后失效 (数据损坏) 335
- 7.7.1 解决方案 335
- 7.7.2 写操作实验步骤 335
- 7.7.3 读操作实验步骤 338
- 7.7.4 小结 339
- 7.8 NFS先失效 (数据未损坏), Primary后失效 340
- 7.8.1 解决方案 340
- 7.8.2 写操作实验步骤 340
- 7.8.3 读操作实验步骤 342
- 7.8.4 小结 343

7.9 NFS先失效（数据损坏），Primary后失效（数据损坏）	344
7.9.1 解决方案	344
7.9.2 写操作实验步骤	344
7.9.3 读操作实验步骤	346
7.9.4 小结	348
7.10 实验结论	348
第8章 Cloudera HA NameNode使用	349
8.1 HA NameNode说明	350
8.2 CDH4B1版本HDFS集群配置	351
8.2.1 虚拟机安装	351
8.2.2 nn1配置	351
8.2.3 dn1~dn3配置	355
8.2.4 HDFS集群构建	358
8.3 HA NameNode配置	361
8.3.1 nn1配置	361
8.3.2 其他节点配置	365
8.4 HA NameNode使用	367
8.4.1 启动HA HDFS集群	367
8.4.2 第1次failover	368
8.4.3 模拟写操作	368
8.4.4 模拟Active Name Node失效，第2次failover	369
8.3.5 模拟新的Standby NameNode加入	370
8.5 小结	371
• • • • •	(收起)

[高可用性的HDFS 下载链接1](#)

标签

hadoop

HDFS

分布式

存储

计算机

计算机科学

数据库

放下

评论

这是个啥JB啊都能出书

照着源码简单解读了几种HDFS备份方式，简单看了下，没什么意思（2012.12.12京东购）

最有价值的是图，代码，少了太多的上下文了

内容其实不多，有很多都是环境的安装和测试，不过对于我这个门外汉来说，可以快速了解一下HADOOP的高可用机制，3分。

这个方向比较小，受众太窄，而且变化太快了，一两年后可能技术就完全不同了，写个博客挺好的，不知道为什么单独成书了。从内容上看，基本是贴代码加一些分析，架构图类图状态图之类的比较少，看起来比较费劲，不如自己看代码来的方便。

主要是讲几种高可用的NameNode部署，关于原理讲了一点点，为什么源码解释就只是贴代码呢？

难道这就是传说的源码分析...,一坨一坨的java代码。

看了20%，首先版本太老了，其次作者后边分析了太多的非主流版本也都消失了。分析的过程也不够有体系。

比较全面，深入

[高可用性的HDFS_下载链接1](#)

书评

首先说说这本书把，我觉得是非常烂的，一堆一堆的java源代码，里面的操作过程也是机械式的。难道这就是传说中的源码分析？
就像window的文件系统是fat32,linux的是ext4一样，这里的hdfs只是可以存储在多台机器上，一个作为主用，其他作为备份，主机坏了，某个备份点可以自动切...

[高可用性的HDFS_下载链接1](#)