

HDFS文件命令



本附录列出了用于管理文件的HDFS命令。它们的形式为`hadoop fs -cmd <args>`，其中`cmd`为特定的文件命令，`<args>`是一个数目可变的参数。

表A-1为命令的用法。方括号（[]）中的参数是可选的，省略号（...）意味着这些可选参数是可重复的。FILE为文件名，而PATH既可以是文件名，也可以是目录名。SRC和DST分别特指源和目的路径名。LOCALSRC和LOCALDST进一步要求位于本地文件系统之上。

表A-1 命令及其用法

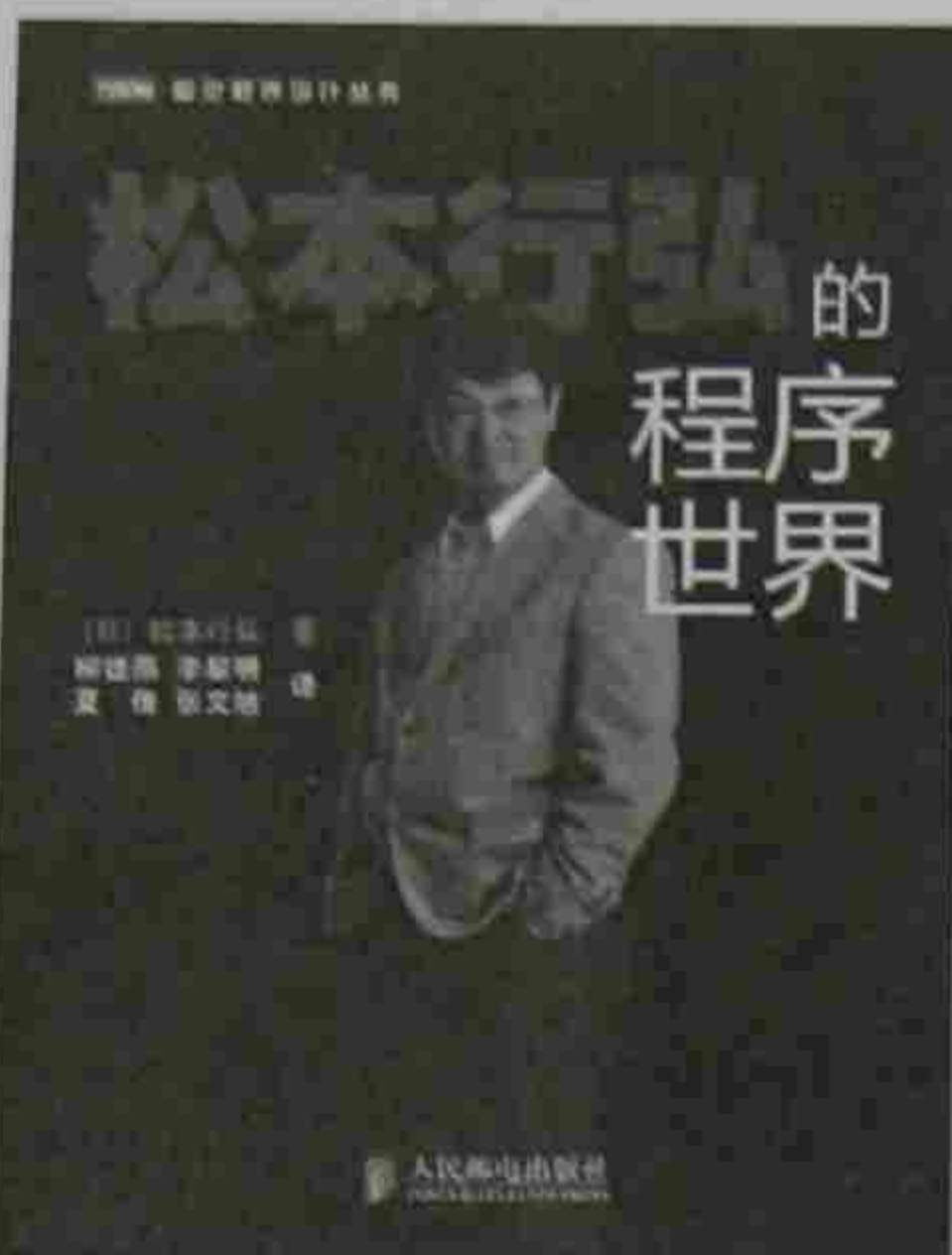
命 令	用法与描述
cat	<code>hadoop fs -cat FILE [FILE ...]</code> 显示文件内容。若要读取压缩文件，应该使用 <code>text</code> 命令
chgrp	<code>hadoop fs -chgrp [-R] GROUP PATH [PATH ...]</code> 变更文件和目录的群组。选项 <code>-R</code> 递归地执行变更。用户必须为文件的所有者或超级用户。关于HDFS文件权限系统的更多背景信息，请参见8.3节
chmod	<code>hadoop fs -chmod [-R] MODE[,MODE ...] PATH [PATH ...]</code> 变更文件和目录的访问权限。类似于Unix的对应命令，MODE可以为一个3位八进制数，或 <code>{augo}+/-{rwxX}</code> 。选项 <code>-R</code> 递归地执行变更。用户必须为文件的所有者或超级用户。关于HDFS文件权限系统的更多背景信息，请参见8.3节
chown	<code>hadoop fs -chown [-R] [OWNER][:[GROUP]] PATH [PATH...]</code> 变更文件和目录的所有者。选项 <code>-R</code> 递归地执行变更。用户必须为超级用户。关于HDFS文件权限系统的更多背景信息，请参见8.3节
copyFromLocal	<code>hadoop fs -copyFromLocal LOCALSRC [LOCALSRC ...] DST</code> 等同于 <code>put</code> （从本地文件系统中复制文件）
copyToLocal	<code>hadoop fs -copyToLocal [-ignorecrc] [-crc] SRC [SRC ...] LOCALDST</code> 等同于 <code>get</code> （将文件复制到本地文件系统中）
count	<code>hadoop fs -count [-q] PATH [PATH ...]</code> 显示由PATH确定的子目录个数、文件个数、使用字节个数，以及所有的文件/目录名。选项 <code>-q</code> 显示额度信息
cp	<code>hadoop fs -cp SRC [SRC ...] DST</code> 将文件从源复制到目的。如果指定了多个源文件，目的端必须为一个目录
du	<code>hadoop fs -du PATH [PATH ...]</code> 显示文件大小。如果PATH是一个目录，会显示该目录中每个文件的大小。文件名用完整的URI协议前缀表示。请注意虽然 <code>du</code> 反映了磁盘使用情况，但不能望文生义，因为真实的磁盘使用情况依赖于块大小和副本系数

(续)

命 令	用法与描述
dus	<pre>hadoop fs -dus PATH [PATH ...]</pre> <p>类似于du, 但是当作用于目录时, dus会显示文件大小之和, 而非逐个显示每个文件的大小</p>
expunge	<pre>hadoop fs -expunge</pre> <p>清空回收站。如果打开回收站属性, 当文件被删除时, 它首先会移动到临时目录.Trash/中。只有超过用户设置的延迟之后, 文件才会被从.Trash/目录中永久删除。而expunge命令强制删除.Trash/目录中的所有文件。请注意只要文件仍在.Trash/目录中, 它就可以被恢复到原始目录中</p>
get	<pre>hadoop fs -get [-ignorecrc] [-crc] SRC [SRC ...] LOCALDST</pre> <p>将文件复制到本地文件系统。如果指定了多个源文件, 本地目的端必须为一个目录。如果LOCALDST被置为-, 文件被复制到stdout</p> <p>HDFS计算每个文件中每个块的校验。一个文件的校验被独立存储在一个隐藏文件中。如果从HDFS读取一个文件, 在这个隐藏文件中的校验被用于验证文件的完整性。对于get命令, 选项-crc会复制这个校验文件。选项-ignorecrc则会在复制时跳过校验检查</p>
getmerge	<pre>hadoop fs -getmerge SRC [SRC ...] LOCALDST [addnl]</pre> <p>获取由SRC指定的所有文件, 将它们合并为单个文件, 并写入到本地文件系统LOCALDST。选项adnl将在每个文件的结尾加入一个换行符</p>
help	<pre>hadoop fs -help [CMD]</pre> <p>显示命令CMD的用法信息。如果不输入CMD, 它会显示所有命令的用法信息</p>
ls	<pre>hadoop fs -ls PATH [PATH ...]</pre> <p>列出文件和目录。每个入口点会显示文件名、权限、所有者、组、大小和修改时间。文件入口点还会显示它们的副本系数</p>
lsr	<pre>hadoop fs -lsr PATH [PATH ...]</pre> <p>ls的递归版本</p>
mkdir	<pre>hadoop fs -mkdir PATH [PATH ...]</pre> <p>创建目录。会创建路径中所有缺失的父目录 (类似Unix的mkdir -p)</p>
moveFromLocal	<pre>hadoop fs -moveFromLocal LOCALSRC [LOCALSRC ...] DST</pre> <p>类似于put, 只是本地的源在成功复制到HDFS上之后会被删除</p>
moveToLocal	<pre>hadoop fs -moveToLocal [-crc] SRC [SRC ...] LOCALDST</pre> <p>显示一条“not implemented yet”消息</p>
mv	<pre>hadoop fs -mv SRC [SRC ...] DST</pre> <p>将文件从源移动到目的。如果指定多个源文件, 目的端必须为一个目录。不允许跨文件系统的移动</p>
put	<pre>hadoop fs -put LOCALSRC [LOCALSRC ...] DST</pre> <p>从本地文件系统中复制文件或目录到目标文件系统。如果LOCALSRC被置为-, 则输入为stdin且DST必须为文件</p>
rm	<pre>hadoop fs -rm PATH [PATH ...]</pre> <p>删除文件和空目录</p>
rmr	<pre>hadoop fs -rmr PATH [PATH ...]</pre> <p>rm的递归版本</p>
setrep	<pre>hadoop fs -setrep [-R] [-w] REP PATH [PATH ...]</pre> <p>改变文件的目标副本系数, 放入REP中。选项-R将递归地改变PATH指定目录中所有文件的目标副本系数。副本系数需要用一定的时间才能达到目标值。选项-w将等待副本系数与目标值相匹配</p>

(续)

命 令	用法与描述
stat	<pre>hadoop fs -stat [FORMAT] PATH [PATH ...]</pre> <p>显示文件中的“统计”信息。FORMAT字符串完全被打印出来，但会按以下设定的格式进行替换</p> <ul style="list-style-type: none"> %b 以数据块为单位的文件大小 %F 根据文件类型为字符串“directory”或“regular file” %n 文件名 %o 块大小 %r 副本 %y 以yyyy-MM-dd HH:mm:ss格式显示的UTC时间 %Y 自1970年1月1日 (UTC) 起计的毫秒数
tail	<pre>hadoop fs -tail [-f] FILE</pre> <p>显示FILE中最后1KB数据</p>
test	<pre>hadoop fs -test [-ezd] PATH</pre> <p>对PATH进行如下类型的检查。</p> <ul style="list-style-type: none"> -e PATH是否存在。如果PATH存在，返回0。 -z 文件是否为空。如果长度为0，返回0。 -d 是否为目录。如果PATH为目录，返回0
text	<pre>hadoop fs -text FILE [FILE ...]</pre> <p>显示文件的文本内容。当文件为文本文件时，等同于cat。文件为压缩格式 (gzip 以及Hadoop的二进制序列文件格式) 时，会先解压缩</p>
touchz	<pre>hadoop fs -touchz FILE [FILE ...]</pre> <p>创建长度为0的文件。如果文件已经存在且长度非0，则报错</p>



松本行弘的程序世界

书号: 978-7-115-25507-5
作者: まつもとゆきひろ (松本行弘)
译者: 柳德燕 李黎明 夏倩 张文旭
定价: 75.00 元

“Ruby 之父”松本行弘教你大师级的程序思考方法!

松本行弘 (matz, Yukihiro Matsumoto) 1965 年出生。岛取县米子市出身。筑波大学第三学院资讯学系毕业。于 1993 年开始开发面向对象的编程语言 Ruby, 于 1995 年发表。目前为网络应用通讯研究所 (NaCI) 研究员, 并兼任乐天技术研究所的研究员。以 Matz 的昵称为人所知。

本书是探索程序设计思想和方法的经典之作。作者从全局的角度, 利用大量的程序示例及图表, 深刻阐述了 Ruby 编程语言的设计理念, 并以独特的视角对与编程相关的各种技术进行了考察。阅读本书不仅可以深入了解编程世界各个要素之间的关系, 而且能够学到大师级的程序思考方法。

www.ChinaDBA.net 中国DBA超级论坛



软件项目成功之道

书号: 978-7-115-25965-3
定价: 39.00 元



软件研发之道: 微软开发团队的经验法则

书号: 978-7-115-25260-9
定价: 45.00 元



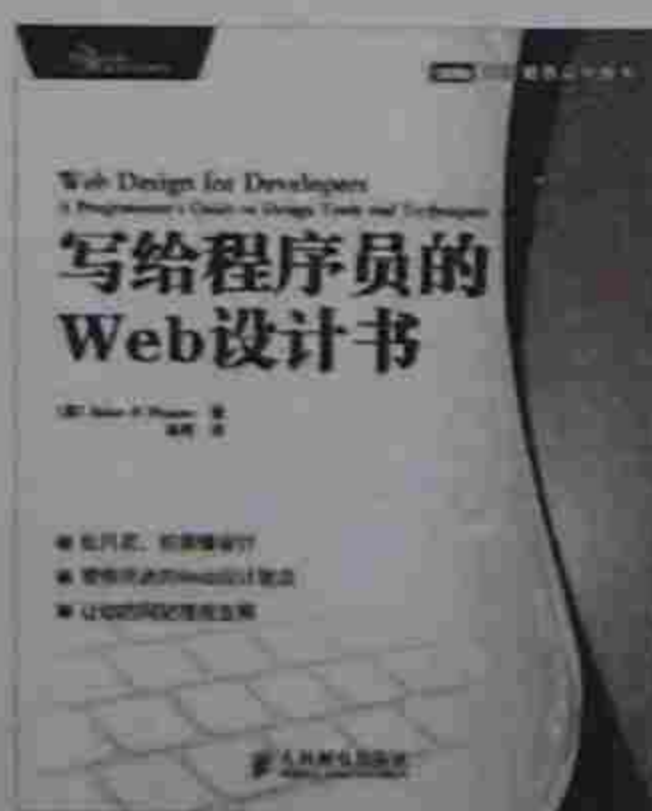
浮现式设计: 专业软件开发的演进本质

书号: 978-7-115-25978-3
定价: 59.00 元



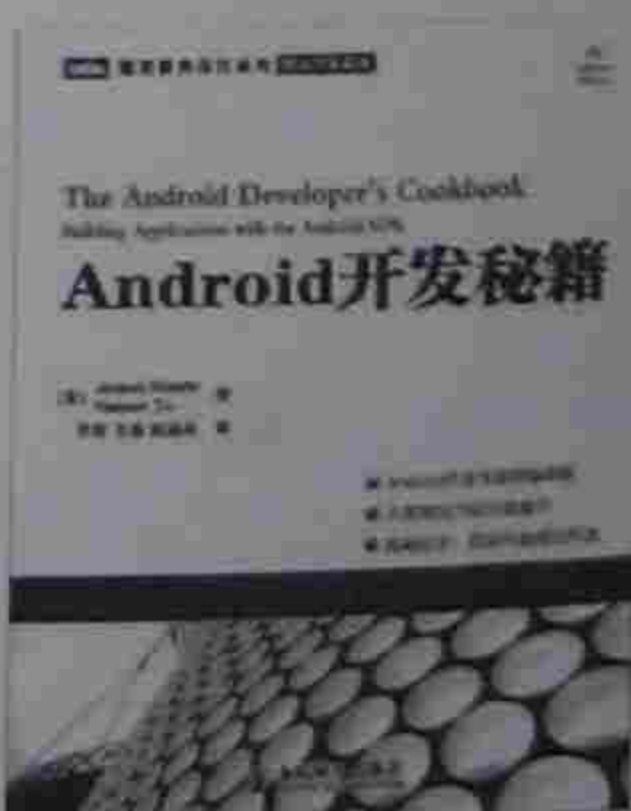
未公开的 Oracle 数据库秘密

书号: 978-7-115-26016-1
定价: 69.00 元



写给程序员的 web 设计书

书号: 978-7-115-25911-0
定价: 59.00 元



Android 开发秘籍

书号: 978-7-115-25718-5
定价: 49.00 元

Hadoop in Action

Hadoop 实战

“本书是初学者的指路明灯，是高级用户的洞察力之源。”

——Philipp K. Janert, Principal Value公司

“为你全面阐释Hadoop的内容、成因和运行机理。”

——Paul Stusiak, Falcon 技术公司

“将Hadoop阐释清楚的最佳图书！”

——Rick Wagner, Acxiom公司

“全面覆盖Hadoop！他书无而本书有。”

——John S. Griffin, Overstock.com

“本书是对Hadoop和MapReduce的极佳介绍。”

——Kenneth DeLong, BabyCenter公司

Apache Hadoop是一个NoSQL应用程序框架，在分布式集群中运行，它适合于处理大数据集。如果需从数据中分析信息，那么Hadoop是你的最佳选择。

本书是一本深受读者好评的专著，旨在教会你如何以MapReduce方式编写程序，其中包含MapReduce编程中的最佳实践及设计模式。书中内容由浅入深，以几个简单的例子开始，继而转向Hadoop在较为复杂的数据分析中的应用。此外，还介绍了Streaming API 及Pig和Hive等工具。

 MANNING

图灵社区: www.ituring.com.cn

反馈/投稿/推荐信箱: contact@turingbook.com

热线: (010)51095186转604

分类建议 计算机/数据库/Hadoop

人民邮电出版社网址: www.ptpress.com.cn



ISBN 978-7-115-26448-0



9 787115 264480 >

ISBN 978-7-115-26448-0

定价: 59.00元