
此教程来自于王家林免费发布的 3 本 Hadoop 教程：云计算分布式大数据 Hadoop 实战高手之路（共 3 本书）：

1， 王家林编写的“云计算分布式大数据 Hadoop 实战高手之路---从零开始”带领您无痛入门 Hadoop 并能够处理 Hadoop 工程师的日常编程工作，进入云计算大数据的美好世界。

2， 王家林编写的“云计算分布式大数据 Hadoop 实战高手之路---高手崛起”通过数个案例实战和 Hadoop 高级主题的动手操作带领您直达 Hadoop 高手境界。

3， 王家林编写的“云计算分布式大数据 Hadoop 实战高手之路---高手之巅”通过当今主流的 Hadoop 商业使用方法和最成功的 Hadoop 大型案例让您直达高手之巅，从此一览众山小。

王家林简介：

Android 架构师、高级工程师、咨询顾问、培训专家；

通晓 Android、HTML5、Hadoop，迷恋英语播音和健美；

致力于 Android、HTML5、Hadoop 的软、硬、云整合的一站式解决方案；

国内最早（2007 年）从事于 Android 系统移植、软硬整合、框架修改、应用程序软件开发以及 Android 系统测试和应用软件测试的技术专家和技术创业人员之一。

HTML5 技术领域的最早实践者（2009 年）之一,成功为多个机构实现多款自定义 HTML5 浏览器，参与某知名的 HTML5 浏览器研发；

云计算分布式大数据处理的最早实践者之一，Hadoop 的狂热爱好者，不断的在实践中用 Hadoop 解决不同领域的大数据的高效处理和存储；

超过 10 本的 IT 畅销书作者；

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel： 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains

---

智者说，要想最快的进步，主要有两点：
1，向第一名学习，向有结果的人学习；
2，采用持续的、大量的、有决心的行动；

想在云计算分布式大数据时代游刃有余，您可参加
王家林亲授的上海 7 月 6-7 日云计算分布式大数据 Hadoop 深入浅出案例驱动实战：

课程信息：
http://www.cnblogs.com/guoshiandroid/archive/2013/06/06/3122665.html

这是一个全程实作的公开课，由浅入深，循序渐进，萃取出实际开发中最常用、最实用的内容并以深入浅出的方式把难点化于无形之中。

课程主题：云计算分布式大数据 Hadoop 深入浅出案例驱动实战
开课时间：2013 年 7 月 6 号-7 号（周六、周日）（9:00-12:00,13:30-17:00）
上课地点：上海

王家林的联系电话：18610086859

新浪微博:http://weibo.com/ilovepains

QQ：1740415547

Hadoop 讨论 QQ 群：312494188

Weixin：wangjialinandroid

官方博客：http://www.cnblogs.com/guoshiandroid/

报名：
请填写以下项目，发邮件到 18610086859@126.com ，为保证每个学员的听课质量，我们采用小班授课制度，请提前报名，我们将尽快告诉您是否还有座位为您预留，以及其他后续细节。

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains

所在单位名称：
姓名：
部门/职务：
通讯地址及邮编：
电话：
手机：
E-mail：
QQ：

名额有限，先到先得，按听课证号顺序入座，报满则停止。

需要在训练现场对自己的项目进行剖析的学员，可以提前把自己面临的 Hadoop 问题发布到王家林的邮箱 18610086859@126.com ，家林会在课堂上作为案例现场解决。

这一讲主要使用 HDFS 命令行工具操作 Hadoop 分布式集群初体验：

Step 1：使用 HSFS 命令向 Hadoop 分布式集群存放一个大文件；

Step 2：删除文件并用两份副本在 HDFS 上存放数据；

现在开始动手做！

Step 1：使用 HSFS 命令向 Hadoop 分布式集群存放一个大文件；

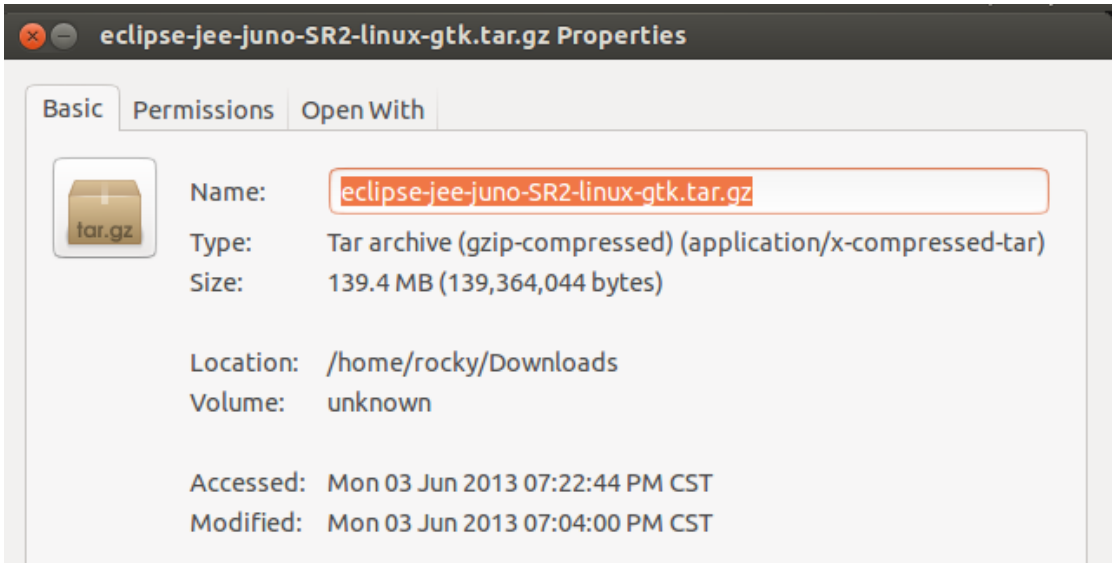
我们曾下载过 Eclipse 这个 IDE，如下所示：

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains



此时可以看出这个文件的大小是 139.4M，这个大小超过了 Hadoop 默认文件块的大小 64M，因此这个文件讲会被分块存储。

下面我们使用命令行工具，在 hadoop.main 这台机器上把这个文件存入整个分布式文件系统：

```
rocky@hadoop:/usr/local/hadoop/bin$ hadoop fs -put /home/rocky/Downloads/eclipse
-jee-juno-SR2-linux-gtk.tar.gz /
rocky@hadoop:/usr/local/hadoop/bin$
```

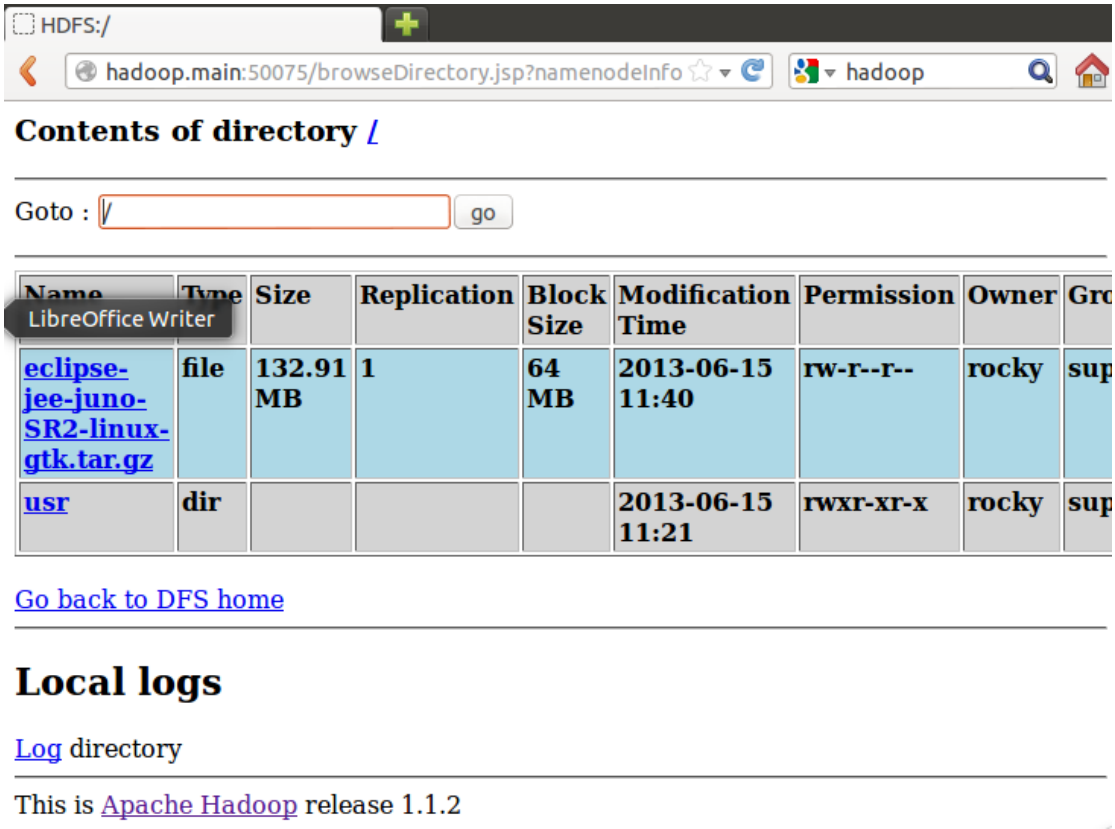
通过上述命令就把 Eclipse 传递到了 hdfs 文件系统的根目录下，此时我们通过 Web 控制台查看一下文件系统：

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains



可以看到文件成功从 Ubuntu 本地发布到了 Hadoop 分步式文件系统上！

点击文件我们可以看到该文件分块的信息：

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains

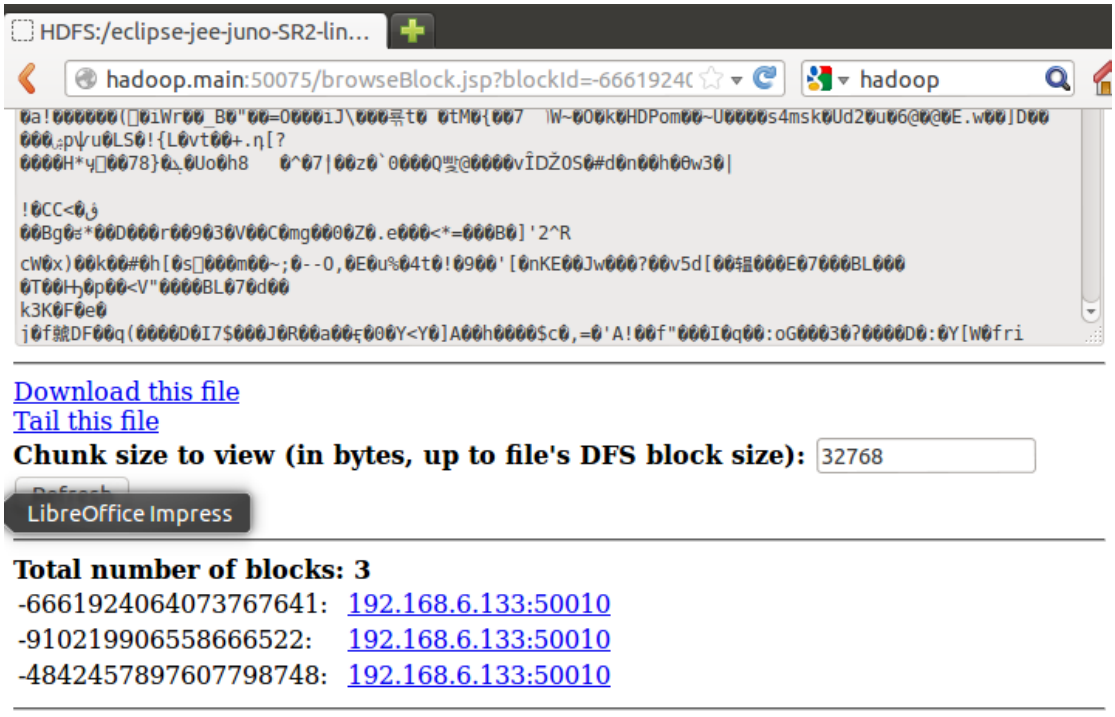
在分块信息下面有具体的分块数据：

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains



可以看到数据被分成了三个块，但是大家同时会注意到这个快在同一台主机上， 按照前面的配置，这些数据应该是分布在 不同的主机节点上的，为何会出现现在的情况呢？其实，这不是一个问题，因为我们的节点比较少，Hadoop 是根据自己的算法调整把这个文件存放在了一个节点上。

Step 2：删除文件并用两份副本在 HDFS 上存放数据；

当我们现在有两个 DataNode 的情况下，可以把 replication 配置伪 2

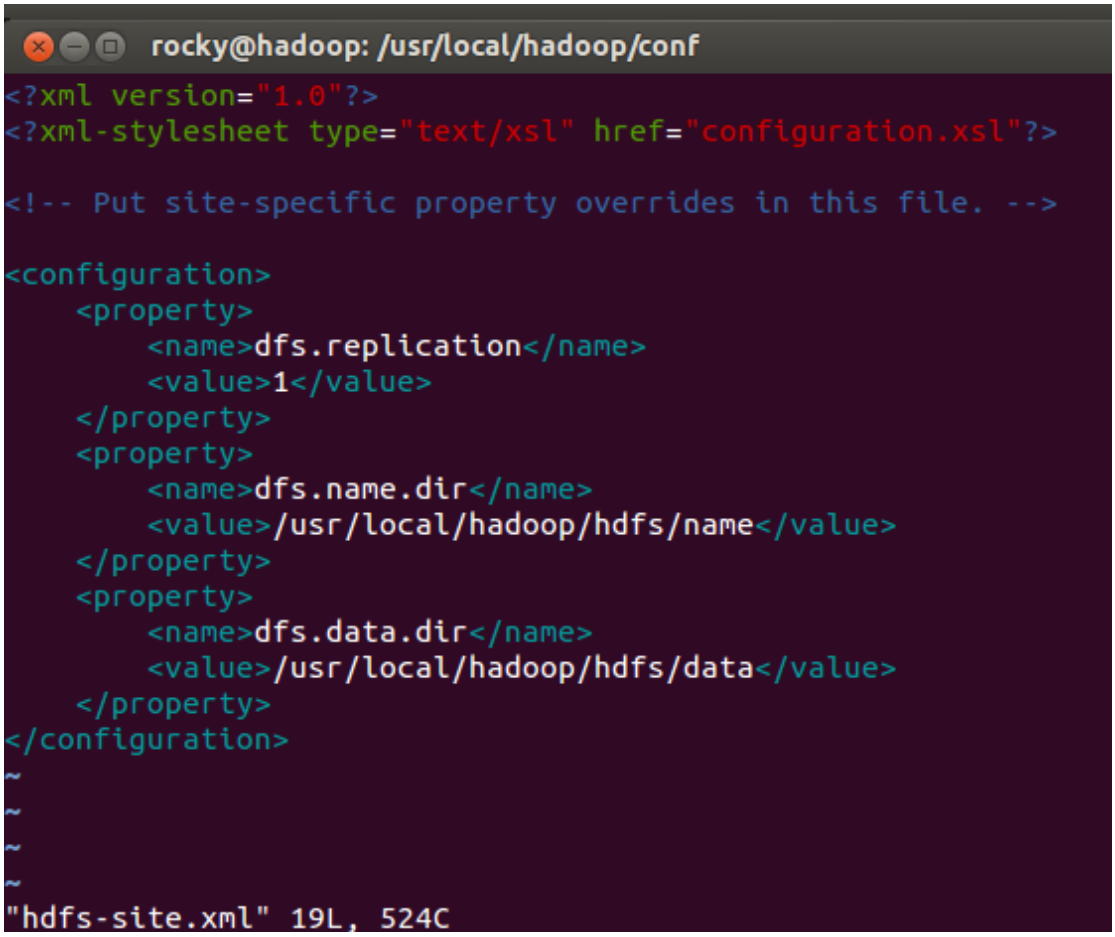
进入文件：

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains



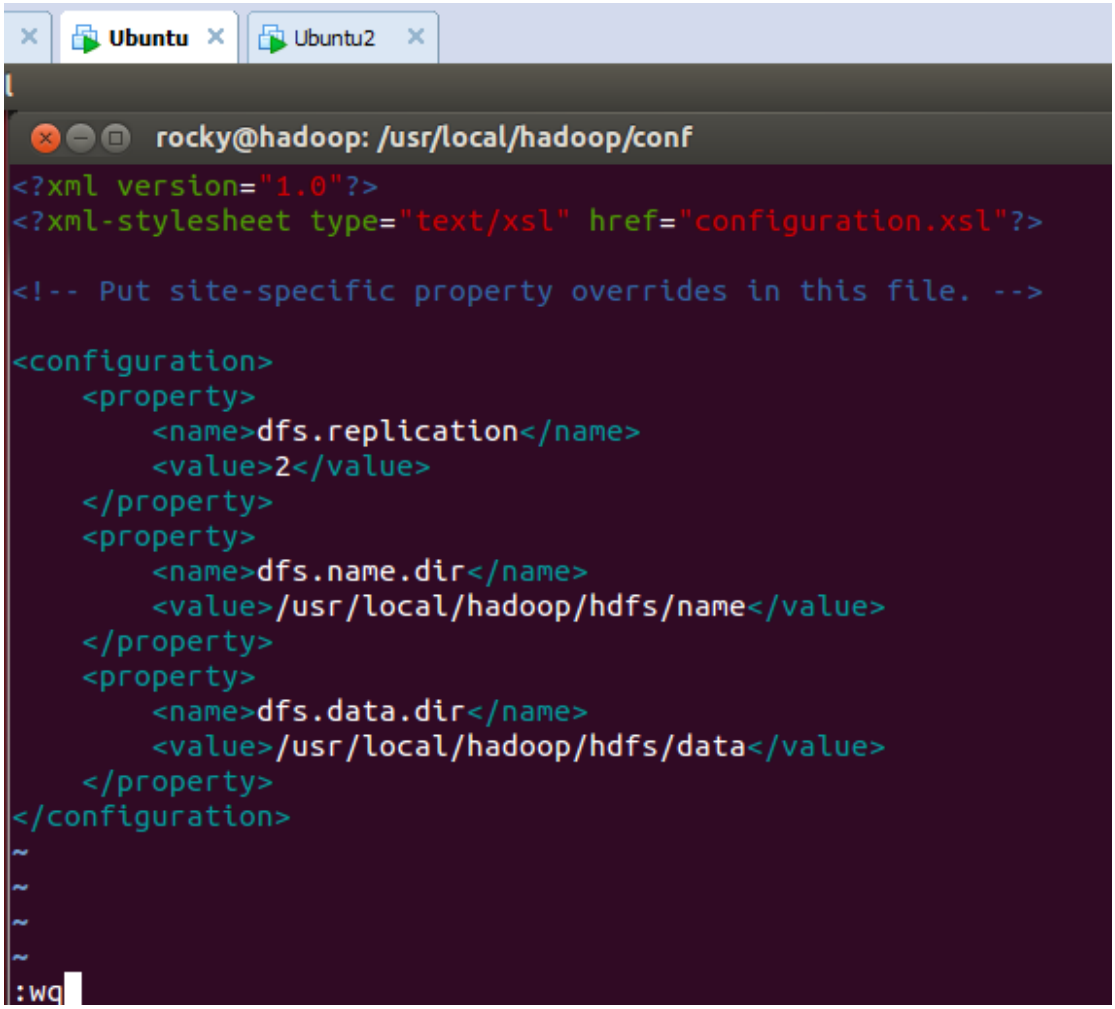
王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains

修改后：



保存退出：

停止 Hadoop 集群并重新启动：

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains

```
rocky@hadoop:/usr/local/hadoop/bin$ stop-all.sh
stopping jobtracker
hadoop.main: stopping tasktracker
hadoop.slave: stopping tasktracker
stopping namenode
hadoop.slave: stopping datanode
hadoop.main: stopping datanode
hadoop.main: stopping secondarynamenode
rocky@hadoop:/usr/local/hadoop/bin$ start-all.sh
starting namenode, logging to /usr/local/hadoop/libexec/../logs/hadoop-rocky-nam
enode-hadoop.main.out
hadoop.slave: starting datanode, logging to /usr/local/hadoop/libexec/../logs/ha
doop-rocky-datanode-hadoop.slave.out
hadoop.main: starting datanode, logging to /usr/local/hadoop/libexec/../logs/had
oop-rocky-datanode-hadoop.main.out
hadoop.main: starting secondarynamenode, logging to /usr/local/hadoop/libexec/..
/logs/hadoop-rocky-secondarynamenode-hadoop.main.out
starting jobtracker, logging to /usr/local/hadoop/libexec/../logs/hadoop-rocky-j
obtracker-hadoop.main.out
hadoop.slave: starting tasktracker, logging to /usr/local/hadoop/libexec/../logs
/hadoop-rocky-tasktracker-hadoop.slave.out
hadoop.main: starting tasktracker, logging to /usr/local/hadoop/libexec/../logs/
hadoop-rocky-tasktracker-hadoop.main.out
rocky@hadoop:/usr/local/hadoop/bin$
```

接下来把刚刚上传到 HDFS 的 Eclipse 删除掉：

```
rocky@hadoop:/usr/local/hadoop/bin$ hadoop fs -rm /eclipse-jee-juno-SR2-linux-gt
k.tar.gz
Deleted hdfs://hadoop.main:9000/eclipse-jee-juno-SR2-linux-gtk.tar.gz
rocky@hadoop:/usr/local/hadoop/bin$
```

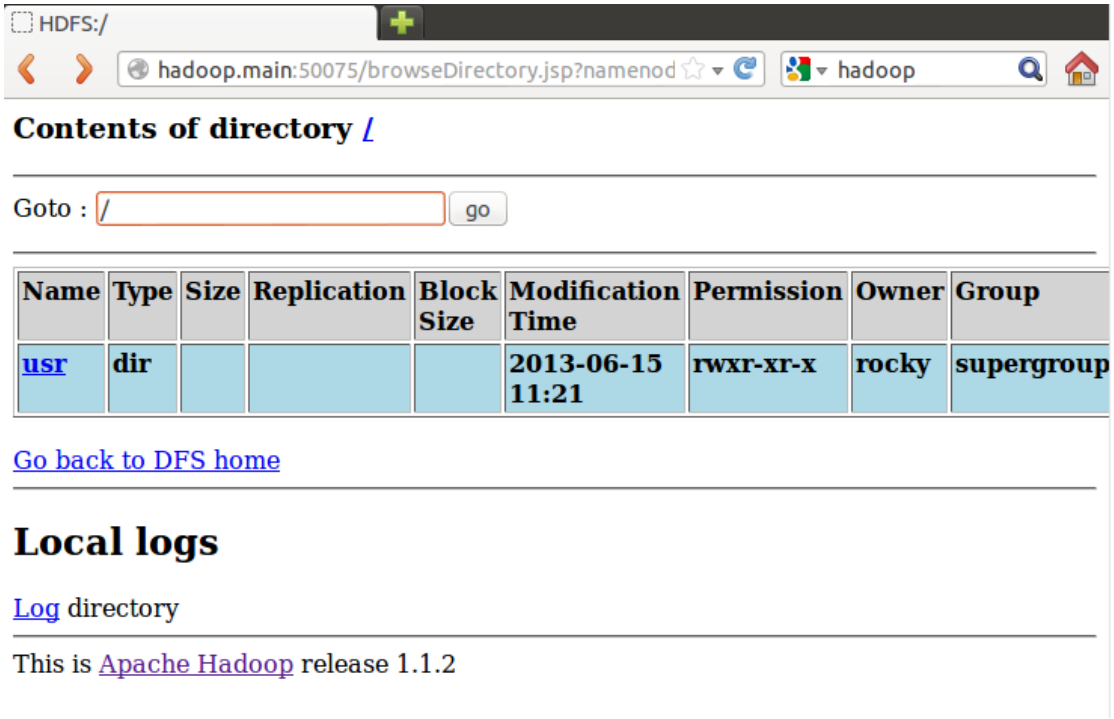
此时 HDFS 中的 Eclipse 就被删除掉了：

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains



接下来我们再次把文件从本地上传给 HDFS:

```
rocky@hadoop:/usr/local/hadoop/bin$ hadoop fs -put /home/rocky/Downloads/eclipse
-jee-juno-SR2-linux-gtk.tar.gz /
rocky@hadoop:/usr/local/hadoop/bin$
```

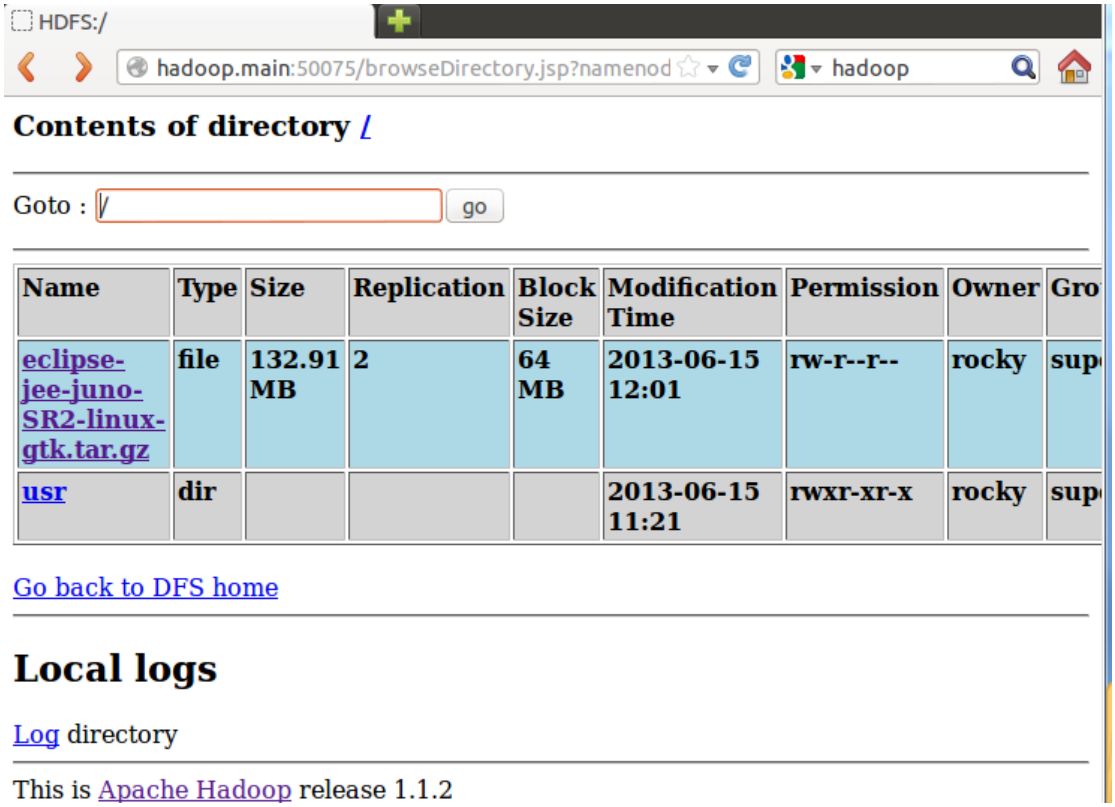
上传成功，打开 Web 控制台：

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains



可以看到文件上传成功，此时我们点击该文件：

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains

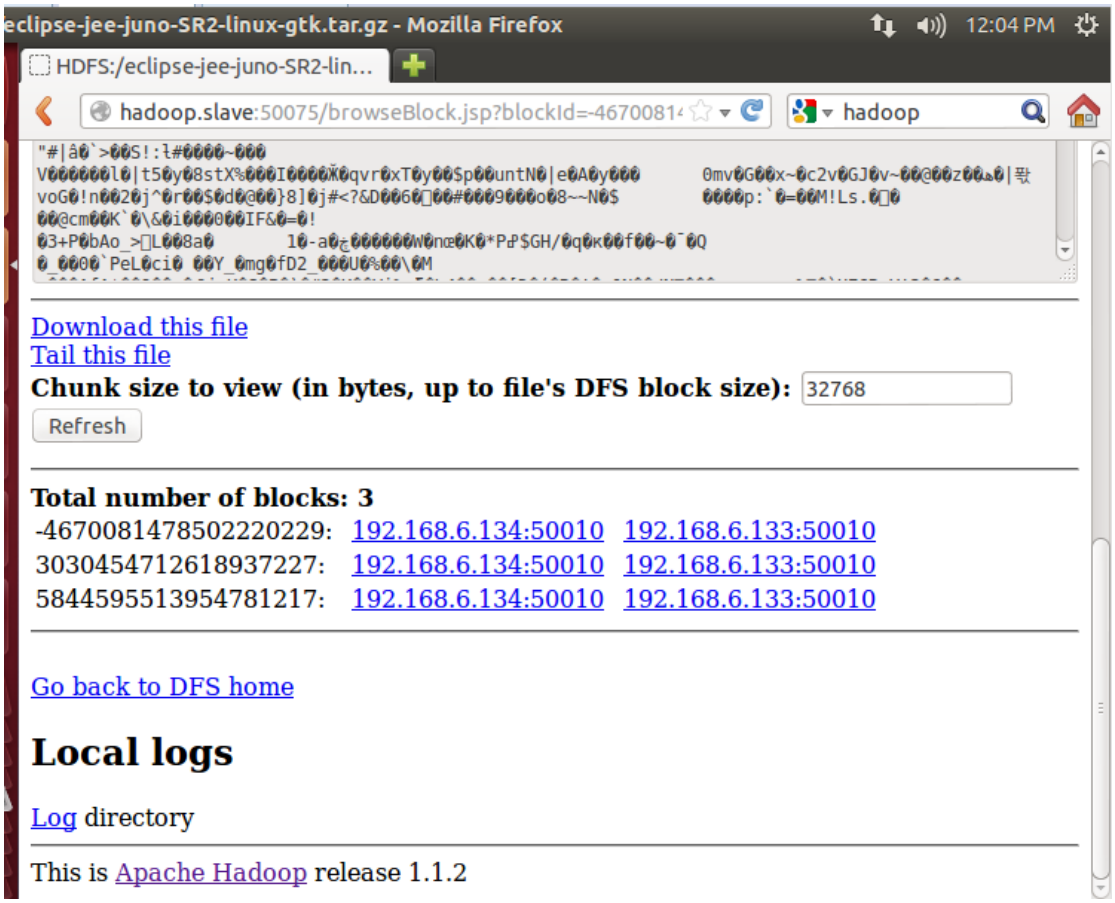
在这个页面的下面我们看到了数据存放的信息：

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains



可以看出每份数据都有两个副本，分别存放在 hadoop.main 和 hadoop.slave 这两天机器上。

对 HDFS 操作的初体验到此结束。

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel: 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains

---

王家林一站式全系列云计算大数据 Hadoop&Android&HTML5&iOS&Linux 训练课程第三个版本（20130606）：

王家林：

Android 架构师、高级工程师、咨询顾问、培训专家；

通晓 Android、HTML5、Hadoop，迷恋英语播音和健美；

致力于 Android、HTML5、Hadoop 的软、硬、云整合的一站式解决方案；

国内最早（2007 年）从事于 Android 系统移植、软硬整合、框架修改、应用程序软件开发以及 Android 系统测试和应用软件测试的技术专家和技术创业人员之一。

HTML5 技术领域的最早实践者（2009 年）之一,成功为多个机构实现多款自定义 HTML5 浏览器，参与某知名的 HTML5 浏览器研发；

云计算分布式大数据处理的最早实践者之一，Hadoop 的狂热爱好者，不断的在实践中用 Hadoop 解决不同领域的大数据的高效处理和存储；

超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com

Tel: 18610086859

QQ：1740415547

QQ 群：312494188

Weixin：wangjialinandroid

Weibo:http://weibo.com/ilovepains

Blog:http://blog.sina.com.cn/ilovepains

王家林完整的云计算大数据分布式训练课程：

王家林一站式全系列云计算大数据 Hadoop&Android&HTML5&iOS&Linux 训练课程第三个版本：
http://www.cnblogs.com/guoshiandroid/archive/2013/06/06/3122798.html

---

CourseID	课程名称	开课类型	课程时长
CH101	亚马逊云服务开发实战	公开课/企业内训	6 小时
CH102	云计算分布式大数据 Hadoop 企业级开发动手实战	公开课/企业内训	24 小时
CH103	云计算分布式大数据 Hadoop 入门经典	公开课/企业内训	12 小时
CH104	云计算分布式大数据 Hadoop 深入浅出案例驱动实战	公开课/企业内训	18 小时
CH105	云计算分布式大数据 Hadoop 深入浅出案例驱动实战（4 天版本）	公开课/企业内训	24 小时
CH106	王家林的云计算分布式大数据 Hadoop 数据库管理员最佳实践	公开课/企业内训	18 小时
CH107	云计算实战：Hadoop 开发全程代码实战（面向软件工程师、数据库工程师、网络后台开发人员等）	公开课/企业内训	12 小时
CH108	云计算实战：Hadoop 数据库管理员实战（面向数据库管理员、系统管理员等）	公开课/企业内训	18 小时
CH109	Hadoop 深入浅出开发实战	企业内训	24 小时
CH110	王家林的云计算实战：Hadoop 大数据处理之生态系统和成功案例（面向 CIO、CTO、DBA、架构师等）	企业内训	6 小时

王家林完整的 Android 训练课程：

CourseID	课程名称	开课类型	课程时长
AF101	Android 系统完整训练：开发搭载 Android 系统的产品	公开课/企业内训	24 小时

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel： 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains

AF102	Android 框架深入浅出： HAL&Framework &Native Service &Android Service 架构设计与实战开发	公开课/企业内训	18 小时
AF103	Android 软硬整合框架精髓实战	公开课/企业内训	18 小时
AF104	Android Framework 系统整合与维护	公开课/企业内训	12 小时
AF105	Android 4.x porting：移植技术与实战训练	公开课/企业内训	12 小时
AF106	Android Binder IPC Subsystem	企业内训	12 小时
AF107	Android UI & View Subsystem 架构与设计解析	企业内训	12 小时
AF108	Android ActivityManager 架构与设计解析	企业内训	12 小时
AF109	Android WindowManager 架构与设计解析	企业内训	12 小时
AF110	Application Launching & Launcher Design	企业内训	12 小时
AD201	Android 平台应用开发最佳实践	公开课/企业内训	24 小时
AD202	精通移动互联网下 Android 应用程序开发实战	公开课/企业内训	18 小时
AT301	云时代 Android 应用测试最佳实践	公开课/企业内训	18 小时
AT302	Android 系统测试最佳实践	公开课/企业内训	18 小时
AF111	Android 架构及实战技术（深入浅出级别）（为手机厂定制）	公开课/企业内训	18 小时
AF112	Android 架构及实战技术（专家级）	公开课/企业内训	18 小时
AD203	Android 应用程序开发实战深入浅出	公开课/企业内训	18 小时

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel： 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains

王家林一站式全系列云计算大数据 Hadoop&Android&HTML5&iOS&Linux 训练课程第三个版本：
http://www.cnblogs.com/guoshiandroid/archive/2013/06/06/3122798.html

---

王家林完整的 HTML5 训练课程：

CourseID	课程名称	开课类型	课程时长
WB101	面向 Web Cloud 的 HTML5 App 开发实战：Browser&HTML5&CSS3&PhoneGap&jQuery Mobile& WebSocket&Node.js	公开课/企业内训	12 小时
WB102	HTML5 端云整合：智能端应用与云端服务整合开发实战	公开课/企业内训	18 小时
WB103	Node.js 与云端服务开发	企业内训	12 小时
WB104	云端数据库入门：NoSQL 与 Open API 实战	企业内训	12 小时
WB105	HTML5 端云整合：HTML5 彻底研究与开发实战	公开课/企业内训	24 小时

王家林完整的无线终端 Android&iOS&Linux 测试训练课程

CourseID	课程名称	开课类型	课程时长
TD101	Android 测试最佳实践	公开课/企业内训	12 小时
TD102	移动互联网云计算时代的智能终端测试实战课程	公开课/企业内训	24 小时
TD103	iOS 测试最佳实践	公开课/企业内训	12 小时
TD104	实战测试驱动开发在嵌入式系统中的应用	公开课/企业内训	12 小时

王家林：通晓 Android、HTML5、Hadoop，迷恋英语播音和健美，超过 10 本的 IT 畅销书作者；

Email: 18610086859@126.com Tel： 18610086859 QQ：1740415547 QQ 群：312494188

微信：wangjialinandroid

Weibo: http://weibo.com/ilovepains Blog: http://blog.sina.com.cn/ilovepains