**Hadoop是云计算的事实标准软件框架，是云计算理念、机制和商业化的具体实现，是整个云计算技术学习中公认的核心和最具有价值内容。**

**如何从企业级开发实战的角度开始，在实际企业级动手操作中深入浅出并循序渐进的掌握Hadoop是本课程的核心。**

**云计算学习者的心声：**

如何从企业级开发的角度，不断动手实际操作，循序渐进中掌握Hadoop，直到能够直接进行企业级开始，是困惑很多对云计算感兴趣的朋友的核心问题，本课程正是为解决此问题而生，学习者只需要按照一步步的跟着视频动手操作，即可完全无痛掌握Hadoop企业级开发。

同时本课程会对Hadoop的核心源码进行剖析，使学习者具备一定的修改Hadoop框架的能力，从而能够根据实际的业务情况来打造自己的框架。

**Hadoop领域4个开创先河**

**1，全程覆盖Hadoop的所有核心内容**

**2，全程注重动手实作，循序渐进中掌握Hadoop企业级实战技术**

**3，在授课的过程中会对Hadoop的核心源码进行深度剖析，使得学员具有改造Hadoop框架的能力**

**4，具备掌握Hadoop完整项目的分析、开发、部署的全过程的能力**

**---讲师:**

**王家林** :

Hadoop源码级专家，曾负责某知名公司的类Hadoop框架开发工作，专注于Hadoop一站式解决方案的提供，同时也是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，现在正负责Hadoop在搜索引擎中的研发中国电信的多次Hadoop培训，花旗银行的Hadoop内训，金立手机的Hadoop内训等，著有《云计算分布式大数据Hadoop实战高手之路---从零开始》《云计算分布式大数据Hadoop实战高手之路---高手崛起》《云计算分布式大数据Hadoop。实战高手之路---高手之巅》等；

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；

超过10本的IT畅销书作者；

|  |  |
| --- | --- |
| **Total Hadoop Professional** | |
| **培训对象** | 1，对云计算、分布式数据存储于处理、大数据等感兴趣的朋友 2，传统的数据库，例如Oracle、MaySQL、DB2等的管理人员 3，Java开发者 4，网站服务器端的开发人员 |
| **参加课程基础要求** | 对云计算有强烈的兴趣，能够看懂基本的Java语法。 |
| **培训后的目标能力** | 直接上手Hadoop工作，具备直接胜任Hadoop开发工程师的能力和系统管理员的能力。 |
| **培训技能目标** | • 彻底理解Hadoop 代表的云计算实现技术的能力  • 具备构建并驾驭Hadoop集群能力 • 具备修改Hadoop框架的能力  • 具备开发自己网盘的能力  • 具备修改HDFS具体源码实现的能力 • 从代码的角度剖析MapReduce执行的具体过程并具备开发MapReduce代码的能力 • 具备掌握Hadoop如何把HDFS文件转化为Key-Value让供Map调用的能力 • 具备掌握MapReduce内部运行和实现细节并改造MapReduce的能力  • 具体Hadoop企业管理员的实际能力 • 具备理解并通过命令行和Java两种方式操作ZooKeeper的能力  • 具备掌握HBase企业级开发和管理的能力  • 具备掌握Pig企业级开发和管理的能力  • 具备掌握Hive企业级开发和管理的能力  • 具备使用Sqoop自由转换传统关系型数据库和HDFS中数据的能力 • 具备使用Flume收集和管理分布式日志的能力  • 具备掌握Hadoop完整项目的分析、开发、部署的全过程的能力 |
| **培训职业目标** | • Hadoop工程师，能够开发任意复杂程度的Hadoop分布式应用 • Hadoop管理员，能够构建并管理Hadoop集群  • Hadoop框架源码研究和修改的能力  • Hadoop完整项目的分析、开发、部署的全过程的能力 |

**培训内容**

|  |  |  |
| --- | --- | --- |
| 时间 | 内容 | 备注 |
| 第一天 | 第1个主题：**Hadoop三问**（彻底理解Hadoop）  1、 Hadoop为什么是云计算分布式大数据的事实开源标准软件框架？  2、Hadoop的具体是如何工作？  3、Hadoop的生态架构和每个模块具体的功能是什么？  第2个主题：**Hadoop集群与管理**（具备构建并驾驭Hadoop集群能力）  1、 Hadoop集群的搭建  2、 Hadoop集群的监控  3、 Hadoop集群的管理  4、 集群下运行MapReduce程序  第3主题：**彻底掌握HDFS**(具备开发自己网盘的能力)  1、HDFS体系架构剖析  2、NameNode、DataNode、SecondaryNameNode架构  3、保证NodeName高可靠性最佳实践  4、DataNode中Block划分的原理和具体存储方式  5、修改Namenode、DataNode数据存储位置  6、使用CLI操作HDFS  7、使用Java操作HDFS  第4主题：**彻底掌握HDFS**(具备修改HDFS具体源码实现的能力)  1、RPC架构剖析  2、源码剖析Hadoop构建于RPC之上  3、源码剖析HDFS的RPC实现  4、源码剖析客户端与与NameNode的RPC通信 |  |
| 时间 | 内容 | 备注 |
| 第二天 | 第1个主题：**彻底掌握MapReduce**（从代码的角度剖析MapReduce执行的具体过程并具备开发MapReduce代码的能力）  1、MapReduce执行的经典步骤  2、wordcount运行过程解析  3、Mapper和Reducer剖析  4、自定义Writable  5、新旧API的区别以及如何使用就API  6、把MapReduce程序打包成Jar包并在命令行运行  第2个主题：**彻底掌握MapReduce**（具备掌握Hadoop如何把HDFS文件转化为Key-Value让供Map调用的能力）  1、 Hadoop是如何把HDFS文件转化为键值对的？  2、 源码剖析Hadoop读取HDFS文件并转化为键值对的过程实现  3、 源码剖析转化为键值对后供Map调用的过程实现  第3个主题：**彻底掌握MapReduce**（具备掌握MapReduce内部运行和实现细节并改造MapReduce的能力）  1、 Hadoop内置计数器及如何自定义计数器  2、 Combiner具体的作用和使用以及其使用的限制条件  3、 Partitioner的使用最佳实践  4、 Hadoop内置的排序算法剖析  5、 自定义排序算法  6、 Hadoop内置的分组算法  7、 自定义分组算法  8、 MapReduce常见场景和算法实现  第4个主题：**Hadoop集群高级实践**(具体Hadoop企业管理员的实际能力)  1、 动态增加Hadoop的Slave节点  2、 动态修改Hadoop的Replication数目  3、 使用命令管理Hadoop集群实践  4、 剖析Hadoop的安全模式  5、 日志剖析实践 |  |
| 时间 | 内容 | 备注 |
| 第三天 | 第1个主题：**实战ZooKeeper**（具备理解并通过命令行和Java两种方式操作ZooKeeper的能力）  1、ZooKeeper架构剖析及集群搭建  2、使用命令行操作ZooKeeper  3、使用Java操作ZooKeeper  第2个主题：**实战HBase**（具备掌握HBase企业级开发和管理的能力）  1、 HBase架构实现  2、 HBase的数据模型和存储模型  3、 使用CLI操作HBase  4、 使用Java操作HBase  5、 使用MapReduce代码向HBase中导入批量数据  第3个主题：**实战Pig**（具备掌握Pig企业级开发和管理的能力）  1、 Pig架构  2、 使用Pig操作数据的内幕  3、 使用Pig实例数据  第4个主题：**实战Hive**(具备掌握Hive企业级开发和管理的能力)  1、 Hive架构剖析  2、 Hive在HDFS中存储实现  3、 使用MySQL作为Hive的Metastore  4、 内部表、分区表、外部表、桶表  5、 视图  6、 自定义Hive的函数 |  |
| 时间 | 内容 | 备注 |
| 第四天 | 第1个主题：**实战Sqoop**（具备使用Sqoop自由转换传统关系型数据库和HDFS中数据的能力）  1、Sqoop架构  2、实战Sqoop把MySQL中的数据导入到HDFS中  3、实战Sqoop把HDFS中的数据导入到MySQL中  4、定义Sqoop任务  第2个主题：**实战Flume**（具备使用Flume收集和管理分布式日志的能力）  1、 Flume架构体系剖析  2、 Agent配置信息  3、 动态监控文件夹中文件的变化  4、 把数据导入到HDFS中  5、 实例监控文件夹文件的变化并把数据导入到HDFS中  第3个主题：**Hadoop系统管理进阶**（具备掌握MapReduce内部运行和实现细节并改造MapReduce的能力）  1、 Hadoop的安全模式  2、 系统监控  3、 系统维护  4、 委任节点和接触节点  5、 系统升级  6、 更多的系统管理工具实战  7、 系统管理最佳实践  第4个主题：**电信日志项目**（具备掌握Hadoop完整项目的分析、开发、部署的全过程的能力）  通过电信商搜集用户接打电话、上网等产生的日志对用户的电话和网络行为进行分析和监控，初步贯穿前面讲解的主要内容，使大家熟悉Hadoop一个完整项目的分析、开发、部署的全过程。 |  |