**附件1：**

**“Hadoop大数据处理高级工程师”实战培训班课程大纲**

|  |  |  |  |
| --- | --- | --- | --- |
| **课程模块** | **课程主题** | **主要内容** | **案例和演示** |
| **模块一** | Hadoop在云计算技术的作用和地位 | * 传统大规模系统存在的问题 * Hadoop概述 * Hadoop分布式文件系统 * MapReduce工作原理 * Hadoop集群剖析 * Hadoop生态系统对一种新的解决方案的需求 * Hadoop的行业应用案例分析 * Hadoop在云计算和大数据的位置和关系 | * 数据开放,数据云服务平台(DAAS)时代 * Hadoop平台在数据云平台(DAAS)上的天然优势 * 数据云平台(DAAS 平台)组成部分 * 互联网公共数据大云(DAAS)案例 * Hadoop构建构建游戏云(Web Game Daas)平台 |
| **模块二** | Hadoop生态系统介绍和演示 | * Hadoop HDFS 和 MapReduce * Hadoop数据库之HBase * Hadoop数据仓库之Hive * Hadoop数据处理脚本Pig * Hadoop数据接口Sqoop和Flume，Scribe DataX * Hadoop工作流引擎 Oozie | * 运用Hadoop自下而上构建大规模企业数据仓库 * 暴风影音数据仓库实战解析 |
| **模块三** | Hadoop组件详解 | * Hadoop HDFS 基本结构 * Hadoop HDFS 副本存放策略 * Hadoop NameNode 详解 * HadoopSecondaryNameNode 详解 * Hadoop DataNode 详解 * Hadoop JobTracker 详解 * Hadoop TaskTracker 详解 | * Hadoop Mapper类核心代码 * Hadoop Reduce类核心代码 * Hadoop 核心代码 |
| **模块四** | Hadoop安装和部署 | * Hadoop系统模块组件概述 * Hadoop试验集群的部署结构 * Hadoop 安装依赖关系 * Hadoop 生产环境的部署结构 * Hadoop集群部署 * Hadoop 高可用配置方法 * Hadoop 集群简单测试方法 * Hadoop 集群异常Debug方法 | * Hadoop安装部署实验 * Red hat Linux基础环境搭建 * Hadoop 单机系统版本安装配置 * Hadoop 集群系统版本安装和启动配置 * 使用 Hadoop MapReduce Streaming 快速测试系统 * Hadoopcore-site，hdfs-site，mapred-site 配置详解 |
| **模块五** | Hadoop集群规划 | * Hadoop 集群内存要求 * Hadoop集群磁盘分区 * 集群和网络拓扑要求 * 集群软件的端口配置 | * 针对NameNode Jobtracker DataNode TaskTracker Hiveserver 等不同组件需求推荐服务器配置 |
| **模块六** | MapReduce 算法原理 | * Hadoop MapReduce 算法的原理和优化思想 * 灵活运用MapReduce 实现算法 | * 运用MapReduce 构建数据库算法 * Select Sort GrougBy Sum Count * Join 新进流失算法 * 使用 Y-Smart 快速转换SQL 为MapReduce 代码 |
| **模块七** | 编写MapReduce高级程序 | * 使用 Hadoop MapReduce Streaming 编程 * MapReduce流程 * 剖析一个MapReduce程序 * 基本MapReduceAPI概念 * 驱动代码 Mapper、Reducer * Hadoop流 * API 使用Eclipse进行快速开发 * 新MapReduce API * MapReduce的优化 * MapReduce的任务调度 * MapReduce编程实战 * 如何利用其他Hadoop相关技术，包括Apache Hive， Apache Pig，Sqoop和Oozie等 * 满足解决实际数据分析问题的高级Hadoop API | * Hadoop Streaming 和 Java MapReduce Api 差异。 * MapReduce 实现数据库功能 * 利用Combiners来减少中间数据 * 编写Partitioner来优化负载平衡 * 直接访问Hadoop分布式文件系统（HDFS） * Hadoop的join操作 * 辅助排序在Reducer方的合并 * 定制Writables和WritableComparables * 使用SequenceFiles和Avro文件保存二进制数据 * 创建InputFormats OutputFormats * Hadoop的二次排序 * Hadoop的海量日志分析 * 在Map方的合并 |
| **模块八** | 集成Hadoop到现有工作流  及Hadoop API深入探讨 | * 存储系统 * 利用Sqoop从关系型数据库系统中导入数据到Hadoop * 利用Flume导入实时数据到Hadoop * ToolRunner介绍、使用MRUnit进行测试 * 使用Configure和Close方法来进行Map/Reduce设置和关闭 | * 使用FuseDFS和Hadoop访问HDFS * 使用分布式缓存（Distributed Cache） * 直接访问Hadoop分布式文件系统（HDFS） * 利用Combiners来减少中间数据 * 编写Partitioner来优化负载平衡 |
| **模块九** | 使用Hive和Pig开发及技巧 | * Hive和Pig基础 * Hive的作用和原理说明 * Hadoop仓库和传统数据仓库的协作关系 * Hadoop/Hive仓库数据数据流 * Hive 部署和安装 * Hive Cli 的基本用法 * HQL基本语法 * 运用Pig 过滤用户数据 | * 使用JDBC 连接Hive进行查询和分析 * 使用正则表达式加载数据 * HQL高级语法 * 编写UDF函数 * 编写UDAF自定义函数 * 基于Hive脚本内嵌Streaming 编程 |
| **模块十** | Hbase安装和使用 | * Hbase 安装部署 * Hbase原理和结构 * Hbase 运维和管理 | * 使用Hbase+Hive 提供 OLAP SQL查询能力 * 使用Hbase+Phoenix提供 OLTP SQL能力 * 基于Hbase 的时间序列数据库 OpenTsDb 结构解析 |
| **模块十一** | Hadoop2.0 集群探索 | * Hadoop2.0 HDFS 原理 * Hadoop2.0 Yarn 原理 * Hadoop2.0 生态系统 | * 基于Hadoop2.0 构建分布式系统 |
| **模块十二** | Hadoop企业级别案例解析 | * Hadoop 结构化数据案例 * Hadoop 非结构化案例 * Hbase 数据库案例 * Hadoop 视频分析案例 | * 利用大数据分析改进交通管理 * 区域医疗大数据应用案例 * 银联大数据数据票据详单平台 * 广东移动省公司请账单系统 * 上海电信网络优化 * 某通信运营商全国用户上网记录 * 浙江台州市智能交通系统 * 移动广州详单实时查询系统 * 跨区域实时视频监控系统 |
| **模块十三** | RedHadoop 企业版本 | * 运用RedHadoop快速构建服务集群 * 运用RedHadoop DW 构建数据仓库 | * 基于RedHadoop Hive构建数据仓库平台 * 灵活运用 Hive 加速游戏数据仓库 * 基于Pig+OpenCV大规模图像人脸识别 |
| **模块十四** | Spark原理和入门 | * Spark原理；Spark的架构图；Spark运行模式介绍 * —local；—standalone；—messos；—yarn；Spark的RDD * 什么是RDD；RDD的种类；—Tranformation；—Action * Spark的存储级别；Cache介绍；Spark的容错原理 * Lineage容错；Checkpoint容错；RDD的创建 * 案例—统计单词的个数 | |

**附件2：**

**“Hadoop大数据处理高级工程师”实战培训班报名回执表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 单位名称 |  | | | | | |
| 详细地址 |  | | | 联系人 | |  |
| 电话 |  | 传真 |  | 邮箱 | |  |
| 姓名 | 性别 | 部门/职务 | 电话 | 手机 | | 邮箱 |
|  |  |  |  |  | |  |
|  |  |  |  |  | |  |
|  |  |  |  |  | |  |
|  |  |  |  |  | |  |
|  |  |  |  |  | |  |
|  |  |  |  |  | |  |
|  |  |  |  |  | |  |
|  |  |  |  |  | |  |
|  |  |  |  |  | |  |
| 户 名：软博时代（北京）科技有限公司  开 户 行：招商银行北京陶然亭支行  帐 号：110914587710501  户 名：中联软博（北京）科技有限公司  开 户 行：中国工商银行北京中关村支行  账 号：0200095609200088230 | | | | | 参会单位签字或盖章 | |

请于9月30日前发送邮件至协会秘书处，开班前一周我们将发送报到通知，届时请注意查收。

联系人：叶馨，电话：66311548，传真：020-66311543，邮箱：[gzhea@vip.163.com](mailto:gzsoa2010@vip.163.com)。