关于举办大数据Hadoop与Spark架构应用实战

培训班的通知

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 北京  6月  26-29 | 上海  7月  28-31 | 宁波  8月  26-29 | 南京  9月  23-26 | 杭州  10月  28-31 | 上海  11月  26-29 | 福州  12月  17-20 | 北京  2017年1月  13-16 |

**各有关单位：**

根据党中央和国务院“互联网+”行动计划战略布局，落实国务院促进大数据发展行动纲要，响应工业和信息化部培养大数据技术高端人才的号召,中国信息化培训中心特推出了大数据平台Hadoop与Spark架构技术及应用实战课程培训班。通过专业的大数据Hadoop与Spark技术架构体系与业界真实案例来全面提升大数据项目高管、大数据平台架构师，以及大数据开发工程师与大数据应用设计人员的专业水平，旨在培养专业的大数据Hadoop与Spark技术架构专家，培养大数据技术和应用创新型人才，促进大数据技术在各行业内部及跨行业进行实施应用，以及企事业单位的大数据项目开发和落地，并利用大数据提升竞争力优势。现将相关事宜通知如下：

1. **培训特色**
2. 课程培训业界最流行、应用最广泛的Hadoop与Spark大数据技术体系。强化大数据平台的分布式集群架构和核心关键技术实现、大数据应用项目开发和大数据集群运维实践、以及Hadoop与Spark大数据项目应用开发与调优的全过程沙盘模拟实战。
3. 通过一个完整的大数据开发项目及一组实际项目训练案例，完全覆盖Hadoop与Spark生态系统平台的应用开发与运维实践。课堂实践项目以项目小组的形式进行沙盘实操练习，重点强化理解Hadoop与Spark大数据项目各个阶段的工作重点，同时掌握作为大数据项目管理者的基本技术与业务素养。
4. 本课程的授课师资都是有着多年在一线从事Hadoop与Spark大数据项目的资深讲师，采用原理技术剖析和实战案例相结合的方式开展互动教学、强化以建立大数据项目解决方案为主体的应用开发、技术讨论与交流咨询，在学习的同时促进讲师学员之间的交流，让每个学员都能在课程培训过程中学到实实在在的大数据技术知识体系，以及大数据技术应用实战技能，具备实际大数据应用项目的动手开发实践与运维管理部署能力。授课过程中，根据学员需求，增设交流环节，可将具体工作中遇到的实际问题展开讨论，讲师会根据学员的实际情况微调授课内容，由讲师带着全部学员积极讨论，并给出一定的时间让学员上台发言，现场剖析问题的症结，规划出可行的解决方案。
5. **培训目标**

**通过本次课程的培训，学员可达到如下目标：**

* 1. 深刻理解在“互联网+”时代下大数据的产生背景、发展历程和演化趋势；
  2. 了解业界市场需求和国内外最新的大数据技术潮流，洞察大数据的潜在价值，；
  3. 理解大数据项目解决方案及业界大数据应用案例，从而为企业在大数据项目中的技术选型及技术架构设计提供决策参考；
  4. 全面掌握业界最流行的Hadoop与Spark大数据技术体系；
  5. 掌握大数据采集技术；
  6. 掌握大数据分布式存储技术；
  7. 掌握NoSQL与NewSQL分布式数据库技术；
  8. 掌握大数据仓库与统计机器学习技术；
  9. 掌握大数据分析挖掘与商业智能（BI）技术；
  10. 掌握大数据离线处理技术；
  11. 掌握Storm流式大数据处理技术；
  12. 掌握基于内存计算的大数据实时处理技术；
  13. 掌握大数据管理技术的原理知识和应用实战；
  14. 深入理解大数据平台技术架构和使用场景；
  15. 娴熟地运用Hadoop与Spark大数据技术体系规划解决方案满足实际项目需求；
  16. 掌握如何部署符合生产环境要求的Hadoop大数据集群；
  17. 熟练地掌握基于Hadoop与Spark大数据平台进行应用程序开发、集群运维管理和性能调优技巧。

1. **课程大纲**

|  |  |  |
| --- | --- | --- |
| **时间** | **课程模块** | **课程内容** |
| **第一天**  **上午** | **大数据技术基础** | 1. 大数据的产生背景、发展历程 2. 大数据和云计算的关系 3. 大数据应用需求以及潜在价值分析 4. 业界最新的大数据技术发展态势与应用趋势 5. 大数据项目的技术选型与架构设计 6. “互联网+”时代下的电子商务、制造业、零售批发业、电信运营商、互联网金融业、网上银行、电子政务、移动互联网、教育信息化等行业应用实践与应用案例剖析 |
| **业界主流的大数据技术产品与项目解决方案** | 1. 国内外主流的大数据解决方案介绍 2. 当前大数据解决方案与传统数据库方案的剖析比较 3. Apache大数据平台方案剖析 4. CDH大数据平台方案剖析 5. HDP大数据平台方案剖析 6. 开源的大数据生态系统平台剖析 |
| **Hadoop大数据平台剖析** | 1. Hadoop的发展历程以及产业界的实际应用介绍 2. Hadoop大数据平台架构 3. 基于Hadoop平台的PB级大数据存储管理与分析处理的工作原理与机制 4. Hadoop的核心组件剖析 |
| **第一天**  **下午** | **大数据分布式存储系统原理及其应用实践** | 1. 分布式文件系统HDFS的简介 2. HDFS系统的主从式平台架构和工作原理 3. HDFS核心组件技术讲解 4. 基于HDFS的大型存储系统应用开发实战 5. HDFS集群的安装、部署、配置与性能优化实践 6. HDFS与Linux NFS3交互技术以及本地化部署应用实践 7. 分布式键值存储系统的平台架构、核心技术以及应用开发 8. PB级大数据存储项目的案例分析 |
| **大数据MapReduce与Yarn并行处理平台** | 1. MapReduce并行计算模型 2. MapReduce作业执行与调度技术 3. 第二代大数据计算框架Yarn的工作原理以及DAG并行执行机制 4. MapReduce应用开发环境的部署，以及大数据并行处理应用程序开发 5. MapReduce高级编程技巧与性能优化实践 6. MapReduce与Yarn大数据分析处理案例分析 |
| **Hadoop应用实践操作训练** | 1. 部署与配置HDFS，熟练操作HDFS SHELL，HDFS与NFS操作，以及HDFS API开发实践 2. 部署与配置MapReduce与Yarn及其开发实践 3. Hadoop的Linux二次开发环境部署与配置 |
| **第二天**  **上午** | **HBase分布式数据库管理系统** | 1. NoSQL数据库与NewSQL数据库技术介绍，及其在半结构化和非结构化大数据方面的应用实践 2. HBase分布式数据库简介、数据模型以及工作原理 3. HBase分布式数据库集群的平台架构和关键技术剖析 4. HBase应用项目开发技巧，以及客户端开发实战 5. HBase表设计与数据操作以及数据库管理API调用 6. HBase集群的安装部署与配置优化 7. ZooKeeper分布式协调服务系统的工作原理、平台架构、集群部署与配置应用实战 8. HBase集群的运维与监控管理 |
| **HBase半结构化数据管理应用实践操作训练** | 1. 部署与配置HBase集群以及HBase的性能优化 2. 部署与配置ZooKeeper分布式集群 3. 构建HBase开发环境 4. HBase数据库操作及项目实践 |
| **第二天**  **下午** | **Hive大型数据仓库集群平台及其应用实践** | 1. 基于Hadoop的大型分布式数据仓库基础知识，HIVE在行业中的数据仓库应用案例 2. Hive大数据仓库简介以及应用介绍 3. Hive数据仓库集群的平台体系结构、核心技术剖析 4. Hive Server的工作原理、机制与应用 5. Hive数据仓库集群的安装部署与配置优化 6. Hive应用开发技巧 7. Hive SQL剖析与应用实践 8. Hive数据仓库表与表分区、表操作、数据导入导出、客户端操作技巧 9. Hive数据仓库报表设计 10. Hive JDBC与ODBC的工作原理与实现机制 11. Hive HWI、CLI客户端操作以及UDF应用实践 |
| **Mahout大数据分析挖掘平台及其应用实践** | 1. Mahout集群的安装部署与配置优化 2. Mahout实现客户分析，广告分析，日志分析，规律预测，关联分析，定向推荐等应用程序的开发与应用实战 3. Mahout性能优化与分析挖掘算法参数的优化技巧 |
| **Hive数据仓库与Mahout数据挖掘平台的应用实践操作训练** | 1. 部署与配置HIVE集群，以及HIVE性能调优 2. 构建HIVE开发环境 3. HIVE数据仓库操作及项目实践 4. 实现Mahout与Hadoop HBase的应用集成，实现日志数据分析挖掘项目的应用实践 |
| **第三天**  **上午** | **Spark大数据实时处理平台剖析** | 1. Spark的发展历程以及业界的实际应用介绍 2. Spark实时大数据处理平台架构 3. Spark RDD内存弹性分布式数据集的工作原理与机制 4. Spark的核心组件剖析 5. 基于Spark的实时数据仓库与实时分析挖掘处理在行业中的应用实践案例 |
| **基于Spark的实时数据仓库和实时数据分析挖掘处理平台的实现机制，以及SparkSQL，Spark Streaming，MLib，GraphX，SparkR的应用实践** | 1. 内存计算模型和实时处理技术介绍 2. Spark中各个分布式组件的处理框架及工作原理 3. Spark SQL实时数据仓库的实现原理机制及应用实践 4. Spark Streaming流式数据实时处理机制及应用实践 5. Spark MLib实时机器学习算法应用实践与案例应用 6. Spark GraphX实时图数据处理应用实践与社交网络分析应用案例 7. SparkR的实现原理与应用实践 8. Spark组件的应用编程开发实战 9. Spark与Hadoop的集成解决方案实践 |
| **Spark平台与各个组件的实践操作训练** | 1. 部署与配置Spark集群，以及Spark性能调优 2. 构建Spark开发环境 3. Spark程序运行以及操作 4. Spark SQL应用操作实训 5. Spark Streaming应用操作实训 6. Spark MLib应用操作实训 7. Spark GraphX应用操作实训 8. SparkR应用操作实训 9. Spark与HBase集成数据分析实验实训 |
| **第三天下午** | **Storm流式数据处理平台架构及其应用实践** | 1. Storm流式处理系统的平台架构和工作原理 2. Storm关键技术剖析 3. Storm集群安装部署与配置优化 4. Storm日志流数据分析项目应用实战 5. Storm和Hadoop，Spark的应用集成项目实践 |
| **大数据智能化ETL操作工具以及Hadoop集群运维监控工具平台应用** | 1. Hadoop与DBMS之间数据交互工具的应用 2. Sqoop导入导出数据的工作原理，以及Sqoop集群安装部署与配置 3. Kettle集群的平台架构、核心技术工作原理以及应用案例 4. Kettle大数据ETL工具的部署与配置，以及应用实战 5. 利用Sqoop实现MySQL与Hadoop集群之间的数据导入导出交互程序 6. Hadoop大数据运维监控管理系统HUE平台的安装部署与应用配置 7. Hadoop运维管理监控系统Ambari平台的安装部署与应用配置 8. Hadoop集群运维系统Ganglia, Nagios的安装部署与应用配置 |
| **大数据分布式采集与分布式消息订阅系统及其应用实践（可选）** | 1. Flume-NG数据采集系统的数据流模型、平台架构、集群部署与配置应用实战 2. Kafka分布式消息订阅系统的应用介绍、平台架构、集群部署与配置应用实战 |
| **内存数据库管理系统及其应用实践（可选）** | 1. Impala实时查询系统平台架构、核心关键技术剖析 2. Impala实时查询系统的部署与应用开发实践 3. Redis内存数据库集群架构以及核心技术剖析 4. Redis集群的部署与应用开发实战与案例分析 |
| **Cassandra数据管理系统应用实践（可选）** | 1. Cassandra集群的平台架构以及核心关键技术 2. Cassandra一致性哈希算法与数据对象分布策略 3. Cassandra集群的安装部署与配置优化 4. Cassandra应用开发实战与案例分析 |
| **大数据项目应用完整实践与咨询讨论** | 1. 根据讲师布置的实际应用案例，开展大数据完整项目部署设计和应用开发实践 2. 大数据项目的需求分析、应用实施以及解决方案分享咨询与交流讨论 |

1. **授课师资**

**钟老师** 现任职于中科院某研究所，高级工程师，副高职称，项目组负责人，博士毕业于中国科学院计算技术研究所，获工学博士学位（计算机系统结构方向）。中培教育的大数据、云计算、移动互联网系列课程建设与教学专家。近六年来带领团队主要从事大数据与云计算技术项目的研发与IT项目管理工作。钟老师有着多年的企业内训和公开课培训讲师经历，主要讲授大数据平台技术、云计算、移动互联网、电子商务、IT信息软件项目管理、企业信息化规划与管理、IT战略规划与企业架构、数据中心主机规划与IDC系统运营等企业实战类培训课程。钟老师将原理技术剖析和应用实战相结合的授课风格受到广大公开课学员和企业内训学员的欢迎。

**蒋老师** 清华大学博士，云计算专家 熟悉主流的云计算平台，并有商业与开源云计算平台的实践经验,对云计算关键技术有深刻了解和实践经验，如分布式系统、虚拟化、分布式文件系统、云存储等，参与并领导多个大型云计算项目。对大数据关键技术有深刻了解和实践经验，如NoSQL数据库、大数据处理、Hadoop、Hive、HBase、Spark等。

**钱老师**  大数据专家。在电信、电力、金融行业从事Java开发和架构设计的工作；资深云计算研发工程师。作为项目的主要成员和负责人参与并领导完成了多个大型复杂项目，并成功应用于行业解决方案，如海量数据匹配系统、电力行业实时数据采集分析系统等。设计并实现了实时索引系统-云搜，成功应用与某国企知识库系统。并可应用与互联网行业的搜索等应用。完成多个云计算解决方案的架构，涉及到金融行业海量数据分析与数据处理系统、海量日志分析系统、电力用电信息统计系统等，获得业界认可。

1. **培训费用**

5800 元/人（含培训费、考试费、证书费、资料费、午餐） 食宿统一安排，费用自理。

本课程由中国信息化培训中心颁发《大数据开发高级架构师》证书，证书查询： www.zpedu.org （中国信息化培训中心）。

请学员带一寸彩照二张（背面注明姓名），身份证复印件一张。

1. **大数据Hadoop与Spark架构应用报名回执**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 单位名称(开发票名称) | |  | | | | | |
| 通讯地址 | |  | | | 邮 编 |  | |
| 联 系 人 | |  | 电 话 |  | 传 真 |  | |
| 学员姓名 | 职务 | 性别 | 邮 箱 | | 联系电话及手机 | 培训地点 | 是否住宿 |
|  |  |  |  | |  |  |  |
|  |  |  |  | |  |  |  |
|  |  |  |  | |  |  |  |
|  |  |  |  | |  |  |  |
|  |  |  |  | |  |  |  |
|  |  |  |  | |  |  |  |
| 汇 款  方 式 | 户 名：北京中培伟业管理咨询有限公司  开户行：北京农村商业银行卢沟桥支行营业部  帐 号：0203010103000033172 | | | | 学员签字或盖章 | | |
| 户 名：中培创成（北京）科技有限公司  开户行：建设银行北京生命园支行  帐 号：11001048600052506592 | | | |
| 户 名：北京高培伟业管理咨询有限公司  开 户 行：中国工商银行股份有限公司北京菜市口支行  帐 号：0200001809200152881 | | | |

1. **联系方式**

**联系人：**承澎

**电 话：**13121772782

**QQ号码：**764753253

**Email：** 764753253@qq.com [rpzxn@163.com](mailto:rpzxn@163.com)

**网 址：**www.zpedu.org

二**○**一六年一月一日