



## Cloudera Administrator Training for Apache Hadoop(CCA131)

尊敬的 \_\_\_\_\_ 先生/女士，您好！

Cloudera Administrator Training for Apache Hadoop(CCA131) 将于2017年12月在北京召开。

### 会议内容

#### Cloudera Apache Hadoop 管理员培训

##### Cloudera Apache Hadoop 培训及认证帮助您进阶大数据知识及技能

Cloudera 大学为期4 天的Hadoop 管理员培训将帮助学员综合理解并全 面掌握通过Cloudera Manager对Hadoop机群的运维管理，包括：安装，配置、 负载平衡及性能调优等。通过该培训，Hadoop 系统管理员将能准备好应对实际运维中遇到的挑战。

“Cloudera 的 Hadoop 管理员培训让我快速并全面地掌握了我在解决客户所面临的大数据及云平台挑战时所需要的技能和知识，大量节省了时间！”

——Canonical

### 培训内容

通过讲师在课堂上的讲解，以及实操练习，学员将学习以下内容：

- Cloudera Manager 管理机群的特性，譬如日志汇总、配置管理、报告、
- 报警及服务管理。
- YARN、MapReduce、Spark 及 HDFS 的工作原理。
- 如何为你的机群选取合适的硬件和架构。
- 如何将 Hadoop 机群和企业已有的系统进行无缝集成。
- 如何使用 Flume 进行实时数据采集以及如何使用 Sqoop 在 RDBMS 和 Hadoop 机群之间进行数据导入导出。
- 如何配置公平调度器为 Hadoop 上的多用户提供服务级别保障。
- 产品环境中 Hadoop 机群的最佳运维实践。
- Hadoop 机群排错、诊断问题和性能调优。

### 培训对象及学员基础

“面向系统管理员和 IT 经理，需具备 Linux 经验，无需 Apache Hadoop 基础。

### 认证

“结束本课程培训后，我们建议学员准备并注册参加 Cloudera 认证CCA Administrator管理员考试（CCA131）。通过并获取该证书是向公司及客户证明个人在Hadoop领域的技术和专长的有力依据。

## 会议日程

### 课程介绍

#### Apache Hadoop 介绍

- Hadoop 动机
- 基本概念
- Hadoop 核心部件

#### Hadoop 机群安装

- 机群管理方案
- Cloudera Manager 特性
- Cloudera manager 安装
- Hadoop (CDH) 安装

#### Hadoop 分布式文件系统 (HDFS)

- HDFS 特性
- 读写文件
- NameNode 内存考虑
- HDFS 安全简介
- HDFS Web UI
- 使用 HDFS Shell

#### YARN 上的 MapReduce 和 Spark

- 计算平台在 Hadoop 里扮演的角色
- YARN : 机群资源管理器
- MapReduce 概念
- Apache Spark 概念
- Yarn 上的计算平台
- YARN Web UI 及 Shell
- YARN 应用运行日志

#### Hadoop 配置及服务运行日志

- Cloudera Manager 配置管理机制

- 定位配置参数及进行配置变更
- 管理角色实例及添加服务
- 配置 HDFS 服务
- 配置 Hadoop 服务运行日志
- 配置 YARN 服务

#### HDFS 导入数据

- 使用 Flume 从外部数据源实时导入数据
- 使用 Sqoop 从关系数据库导入数据
- REST 接口
- 导入数据的最佳实践

#### Hadoop 机群规划

- 规划考虑因素
- 硬件选择
- 虚拟化选项
- 网络因素
- 节点配置

#### Hive , Impala 及 Pig 的安装及配置

- Hive
- Impala
- Pig

#### Hadoop 客户端及 Hue

- 什么是 Hadoop 客户端
- 安装及配置 Hadoop 客户端
- 安装及配置 Hue
- 使用 Hue 进行身份验证及授权

#### 机群高级配置

- 高级配置参数
- Hadoop 端口配置
- HDFS 机柜感知配置

- HDFS 高可靠性配置

## Hadoop 安全

- Hadoop 安全的重要性
- Hadoop 安全性概念
- Kerberos 简介
- 使用 Kerberos 保护Hadoop 机群
- 其他安全特性

## 资源管理

- 使用静态服务池配置 Linux cgroup
- 公平调度器
- 配置动态资源池
- YARN 内存及 CPU 设置
- Impala 查询调度

## 机群维护

- 检查 HDFS 状态
- 机群间复制数据
- 添加 / 移除机群节点
- 机群数据负载平衡
- 目录快照
- 机群升级

## 机群监控及排错

- Cloudera Manager 监控特性
- 监控 Hadoop 机群
- Hadoop 机群排错
- 常见配置不当问题

## 总结

### 会议门票

8500元/人次，包含一次对应的考试（如果来参加培训的人不考试，仅参加培训的价格为6500/人次）

费用包含：教材、实验手册、虚拟机、税票费用（8500是含考试的）

教室设施：投影、WiFi、排插、饮水（三餐及住宿请自理）

PS:上课时需携带笔记本电脑，虚拟机及课件由讲师发放

