

云帆大数据学院

---

# [项目实战高薪班]



专注于大数据行业高端实用技术培训，为在职人员提供企业一线技术交流、学习平台，是在职人员晋升、转行的最佳选择。

创英才典范学院 造百年在线教育  
[www.yftech.com](http://www.yftech.com)

# 《企业级 Hadoop 2. x 项目实战课程》

## 目录

目录.....	2
1. 课程概况.....	3
2. 授课对象.....	4
3. 云帆优势.....	4
4. 培训效果.....	4
5. 学习目标.....	8
6. 实力承诺.....	9
7. 就业服务.....	10
8. 讲师介绍.....	10
9. 课程主线脉络.....	11
9.1、以企业实际项目为课程导向.....	11
9.2、SQL on Hadoop 企业大数据应用.....	11
9.3、数据深入分析的机器学习与图形计算.....	12
9.4、实时的流式计算.....	12
10. 课程内容安排.....	12
10.1、大数据平台基础环境.....	12
10.2、【文件日志分析】项目.....	14
10.3、【数据实时查询】项目.....	16
10.4、【机器学习与实时计算】项目.....	17
10.5、企业级大数据平台 CDH 5.3. x.....	18
10.6、项目总结与面试.....	19
11. 赠送内容.....	20
12. 学费情况.....	20
13. 开课情况.....	20
14. 报名流程.....	21
15. 疑问解答.....	21

# 云帆大数据学院

---

有没有一套课程是针对多年开发经验的人员，可以深入学习大数据一线技术的课程？而不是仅仅浮于表面，简单了解 Hadoop2.x 周边生态框架的课程？

云帆大数据针对有多年开发经验的人员推出的进阶课程《项目实战课程》，以企业实际项目为课程导向，从业务需求、技术调研，到框架选型、版本确定，以及最后的功能开发、大数据平台建设等方面全面学习。

本套课程是以企业实际项目为课程导向，所有的知识都是经过多年一线企业实际运用提取总结而成，学完后真正具备企业开发所需要的动手能力和架构级别的知识能力；

**云帆大数据学院首创：“保姆式手把手教学”+“教授式点拨教学”相结合的教学方式，让在职人员轻松转行大数据！**

## 1. 课程概况

本课程目标是培养目前在云计算领域非常紧缺的大数据技术方向的开发、运维人才，课程涉及到大数据技术中绝大部分知识，即夯实了学员的基本功底，又通过结合公司真实的项目提高学员动手能力和解决问题的能力，完全可以让开发人员达到有 1-2 年大数据工作经验的水平实力，学习能力强的或有 IT 工作经验的甚至可以达到 2 年以上大数据的水平，进入公司后可以快速地投入到工作当中，提高学员在企业中竞争力。

## 2. 授课对象

本课程适合以下人员学习：

- 1、计算机相关专业本科以上学历，并有至少一年软件开发经验
- 2、有一定的英语基础，能阅读简单的英文技术文档
- 3、学习能力强，有上进心，渴望从事大数据云计算行业
- 4、至少熟练使用 Java 编程语言，对计算机网络、数据结构有一定基础

## 3. 云帆优势

国内最早的在线教育学院之一，一直专注于在职人员高端技术学习和培训；授课老师均有多年一线研发、企业技术培训经验；云帆采用讲师合伙人制，每位授课老师都是在创业授课，为学员贡献出全部的精力；欢迎每位符合要求的人员加入云帆大数据，共创美好未来。

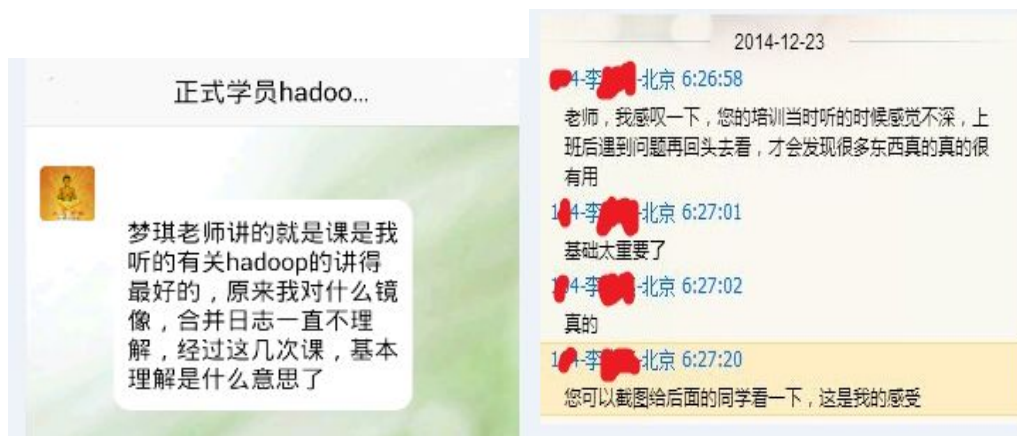
## 4. 培训效果

花钱培训的目的无非是能快速上手大数据工作，换一份高薪的工作；由于大部分学员都是在职的，学习时间是相当紧张和宝贵的；

如何能在当前混乱的培训市场上，快速找到一家真正有实力的培训学校学到真正企业一线开发技术，是比较困难的。云帆大数据特地推出《企业级 Hadoop 2.x 项目实战课程》，讲解全是企业一线开发知识，全面覆盖工作中用到的知识点，全是干货、不搞噱头，课程如何

# 云帆大数据学院

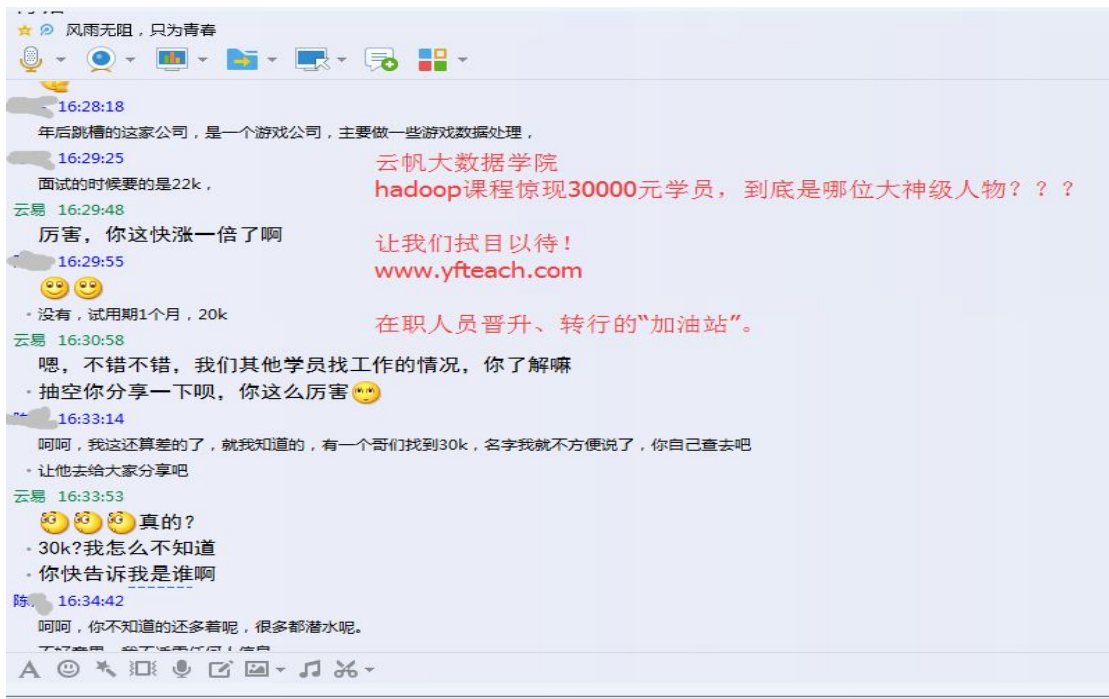
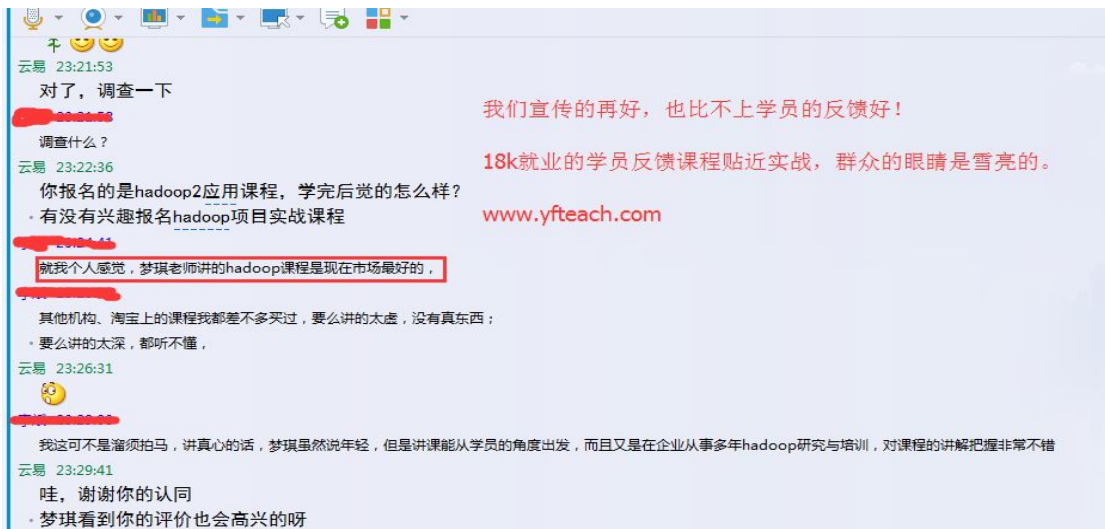
看看学员的评价：



创英才典范学院 造百年在线教育

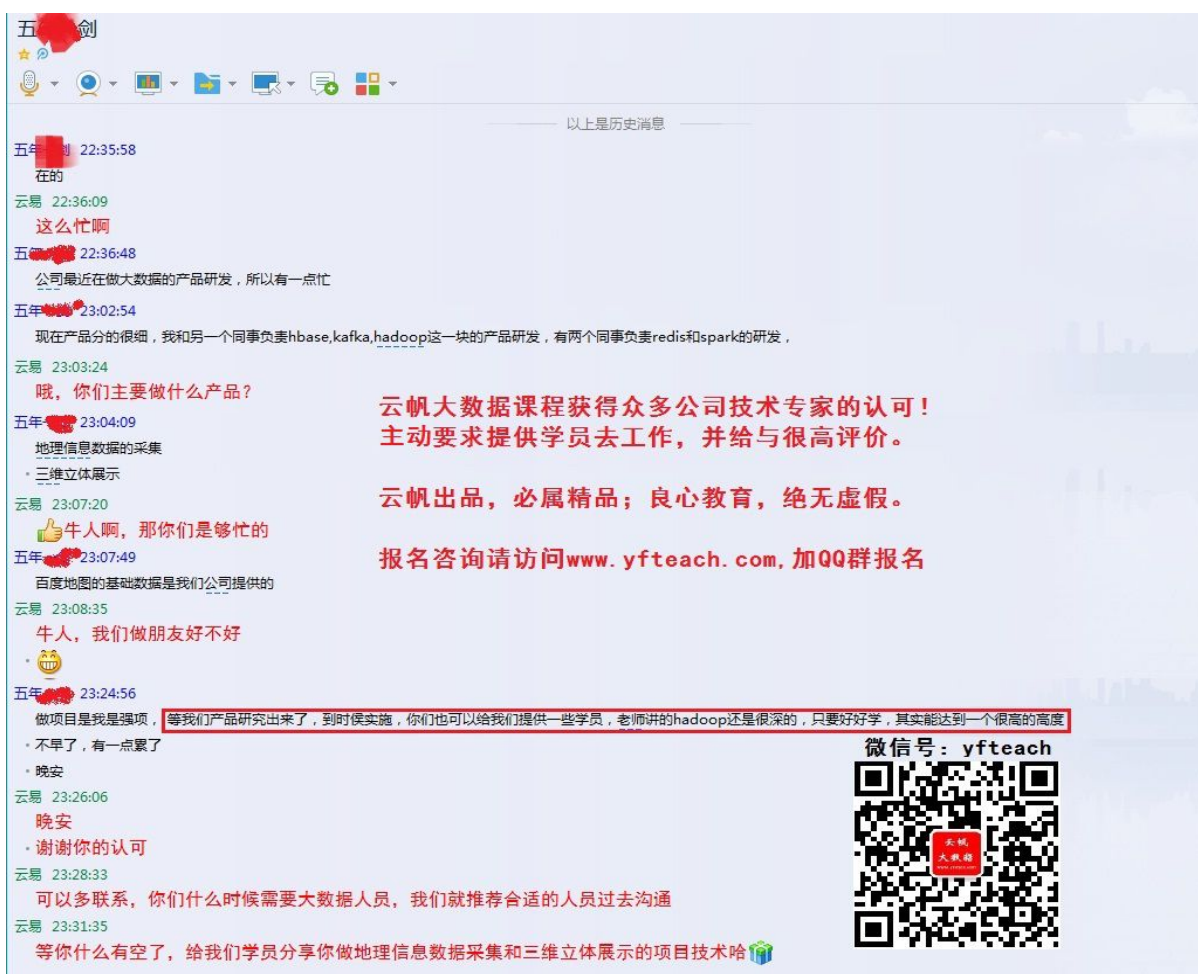
www.yfteach.com

# 云帆大数据学院





# 云帆大数据学院



**我们一直在努力，只为给您实现梦想的实力。**

## 5. 学习目标

本套课程是以企业实际开发技术为基础进行讲解，所有的知识都是经过多年一线企业实际运用提取总结而成，让你真正具备企业开发所需要的能力和知识；

- 1、 培养目前企业中云计算大数据方向的实际开发项目型人才。
- 2、 以企业大数据项目为驱动，培养技术框架研发、动手能力和解决问题的能力；
- 3、 让有编程基础的人达到 2 年以上大数据真正企业实际工作经验水平；
- 4、 进入公司后，快速地投入到工作当中，提高企业在工作中的竞争力；
- 5、 胜任大数据方面机器学习、图形计算、实时计算开发和平台搭建管理；

学习完本套课程即可从事云计算的大数据开发方向又可从事服务器运维方向，可以胜任大数据数据挖掘、大数据实时处理、虚拟化平台搭建和开发，构建公有云和私有云，并在云上部署和开发应用等工作。



## 6. 实力承诺

### 实力承诺

(所述言论, 承担法律责任)

- 1、 **小班化教学**, 严格控制每班的人数, 给大家充分交流、认识; 让你不仅学到知识更能扩展人脉;
- 2、 一次报名, 终身学习; 只要课程有更新, 就可以免费学习; 由于市场技术不断更新, 课程也一直增加新内容, 正式学员**终身免费**学习;
- 3、 国内首创**口碑退款制度**, 如有老学员反馈课程质量有问题, 三天内免费退学退款;
- 4、 国内首家在线课程**分期付款**学习, 让你零担心、零压力的深入学习与自由选择; 你选择我们, 我们选择信任你;
- 5、 国内首套以企业**实际项目**为导向的实战课程, 全面系统深入的讲解 Hadoop 项目每个环节将每个框架、知识点都融合到实际项目中, 绝非单独讲解 hadoop 生态框架, 全程实时在线讲解; 从业务需求、技术调研, 到框架选型、版本确定, 以及最后的功能开发、大数据平台建设; 带你深入浅出学习整个项目开发。
- 6、 截至目前, 云帆大数据学院**正式学员年龄平均 28 岁以上, 平均工作年限 5 年以上**; 以他们的**实力与眼光**选择云帆大数据学习, 你值得选择。

# 云帆大数据学院

---

**真正承诺是来自实力的保障，选择云帆大数据就是选择高薪。**

云帆大数据推出的国内第一套以企业完整实际项目为教学课程，是你进入大数据领域的捷径，学完本课程相当于你拥有**至少两年**以上实战经验水平！

## 7. 就业服务

- 1、老师一对一解答遇到的问题，并针对其以后工作中的问题给与指导；
- 2、一对一辅导简历、面试，直到找到高薪工作
- 3、与知名猎头企业合作，优先推荐面试。
- 4、和众多知名企业 HR、CTO 合作，直接进入企业面试。

**云帆大数据学院提供一线实战研发技术课程+就业推荐服务+就业后工作指导。**

## 8. 讲师介绍

**梦琪** 资深 Hadoop 技术实践者和研究者，对 Hadoop 技术有非常深刻的认识和理解，有着丰富的实践经验。2010 年开始接触 Hadoop/MapReduce，进行调研和实际应用开发；曾主导项目，通过整合 Hadoop 的各种组件用以分析省、市交通卡口数据，为构建智能城市中的智慧交通提供保障。在数据存储、数据分析、数据挖掘、机器学习方面有多年开发经验，有自己独到的见解，致力于让

# 云帆大数据学院

---

Hadoop 大数据应用更简单。参加过众多大型项目，包括《Xxx 市日志审计核查系统》、《Xxx 省交通卡口数据分析系统》、《广东 Xxx 市公安局稽查布控系统》、《Xxx 市公安局卡口系统》、《EAI 服务总线系统系统》、《信息管理与整合系统》等建设、组织开发工作。

云帆大数据学院提供一线实战研发技术课程+就业推荐服务+就业后工作指导服务；欢迎报名学习云帆大数据其他课程：**《企业级 Hadoop2. x 应用开发课程》、《企业级 Spark 1. x 应用开发课程》**；详情请加入 QQ 群：338025410，咨询课程顾问。

## 9、课程主线脉络

### 9.1、以企业实际项目为课程导向

从业务需求、技术调研，到框架选型、版本确定，最后进行功能开发、大数据平台建设；

### 9.2、SQL on Hadoop 企业大数据应用

- 1) 以 Hive 为核心的数据 ETL 和分析
- 2) HBase 与 Phoenix 结合的 Hadoop 准实时数据库存储于查询
- 3) 基于 Google Dremel 思想的 Cloudera Impala 实时 SQL 数据查询系统

## 9.3、数据深入分析的机器学习与图形计算

- 1) 基于 MapReduce 之上的机器学习 Mahout (分类、聚类、推荐)
- 2) 基于 Spark 之上机器学习库 MLlib 和 GraphX 图形计算

## 9.4、实时的流式计算

## 10、课程内容安排

### 10.1、大数据平台基础环境

此阶段课程价格：599 元

#### 1、实战课程开课

- 1) 课程大纲讲解
- 2) 目前企业大数据 Hadoop 方面招聘要求
- 3) 讲师学员互动交流

#### 2、企业大数据应用场景

- 1) 海量数据存储和处理
- 2) 大数据方面的四大应用场景

#### 3、企业级 Hadoop 项目研发

- 1) Hadoop 版本选项
- 2) Hadoop 生态系统框架选择

- 3) 以日志分析为原型的大数据平台架构
- 4) 以海量数据实时存储查询为原型的大数据平台架构
- 5) 以实时计算、机器学习为主的推荐大数据平台架构

## 4、企业大数据平台基础架构

- 1) 大数据平台规模
- 2) 分布式数据存储 HDFS 企业架构
- 3) 分布式资源管理 YARN 企业架构
- 4) 高可用分布式协作框架 Zookeeper 架构

## 5、部署 Hadoop 2.6.x 企业级环境

- 1) Hadoop 2.6.x 伪分布式安装、测试、说明
- 2) Hadoop 2.6.x 分布式集群安装、部署、测试
- 3) Zookeeper 3.4.5 体系结构、分布式安装部署、用户操作与功能调优
- 4) 分布式文件系统 HDFS HA 原理、安装、部署、测试
- 5) 分布式资源管理框架 YARN RM HA 原理、部署
- 6) Hadoop 2.6.x 集群环境性能调优

## 6、CDH 和 HDP 版本大数据平台讲解

- 1) Cloudera 的 CDH 4 和 CDH 5 以及集群管理工具 CM
- 2) Hortonworks 的 HDP 2.2 说明

3) 中小型企业实际的 Hadoop 版本比较与选型

## 10.2、【文件日志分析】项目

此阶段课程价格：2199 元

### 1、实际项目业务

- 1) 【日志】类型业务分析
- 2) 日志业务分析讲解

### 2、MapReduce 数据离线处理

- 1) 回顾 MapReduce 编程和执行过程
- 2) 依据业务编写实例程序
- 3) MapReduce 监控与调优细节

### 3、使用 SQL 进行数据分析 Hive

- 1) 什么是 Hive、Hive 体系结构、Hive 安装和负载均衡搭建
- 2) Hive 表的创建、数据的加载、CLI 操作
- 3) Hive 数据定义和高级查询
- 4) 使用 Hive 分析日志（结合 MapReduce）
- 5) Hive 的内置函数和自定义 UDF 和 UDAF 实战
- 6) 实际项目中 Hive 使用的调优



## 4、文件数据收集工具 **Flume**

- 1) Flume 用途、体系结构、安装部署
- 2) 编写 Flume Agent 配置信息
- 3) 数据导入到 HDFS 中

## 5、Hadoop 与 RDBMS 数据相互转换工具 **Sqoop**

- 1) 安装配置 Sqoop
- 2) 使用 Sqoop 把数据从 mysql 导入到 HDFS 中
- 3) 使用 Sqoop 把数据从 HDFS 导出到 mysql 中
- 4) Sqoop 其他功能使用 and 实际项目应用场景

## 6、工作流调度框架 **Oozie**

- 1) Oozie 功能、编译、安装
- 2) 使用 Workflow 调用 Mapreduce 和 Hive
- 3) 定时调用 **Workflow**
- 4) 基于 dataset 调用 Workflow-基于文件事件

## 7、回顾总结

- 1) 基于企业【日志】模型数据分析完整架构图
- 2) 大数据平台建设、运维、调优等说明
- 3) 大数据平台优化方案

## 10.3、【数据实时查询】项目

此阶段课程价格：2999 元

### 1、企业海量数据查询

- 1) 项目需求功能描述
- 2) 数据采集

### 2、面向列的分布式实时数据库 HBase

- 1) HBase 定义、数据模型、物理模型
- 2) HBase 安装、Shell 命令使用、WEB UI 监控
- 3) HBase JAVA API 使用、与 MR 集成
- 4) HBase 表的设计、Rowkey 设计和查询过滤器
- 5) HBase 创建表时的高级选项（如预创分区、压缩）
- 6) HBase 数据的导入导出和迁移、读写性能调优

### 3、基于 HBase 之上的 SQL 框架

- 1) 与 Hive 集成
- 2) Phoenix 框架
  - ①功能、用途
  - ②安装、部署、测试
  - ③二级索引、序列、视图等

## 4、 Impala 实时 SQL 数据查询系统

- 1) Impala 的体系架构、与 Hive 关系
- 2) Impala 支持的数据类型，SQL 命令，内置函数，查询接口
- 3) Impala 对几种主要的 SQL 查询模式

## 5、项目架构

- 1) 具体业务讲解（设计表、索引表）
- 2) 数据实时采集与容灾备份
- 3) 项目架构设计优化、HBase 源码查看

## 10.4、【机器学习与实时计算】项目

此阶段课程价格：2599 元

### 1、深入数据挖掘

- 1) 数据深入分析（机器学习和图形计算）
- 2) 基于 MR 的 Mahout 和基于 Spark MLlib
- 3) Spark 之上 GraphX 图形计算应用

### 2、机器学习 Mahout

- 1) Mahout 编译、环境安装部署、测试
- 2) 三类学习库（分类、聚类、推荐）
- 3) 聚类算法 K-means 和 canopy、查看分析源码

创英才典范学院 造百年在线教育

[www.yfteach.com](http://www.yfteach.com)

4) 贝叶斯分类算法、应用场景

5) 基于用户和项目推荐算法

## 3、实时计算 Storm

1) Storm 的基本概念、Storm 的应用场景、Storm 和 Hadoop 的对比

2) Storm 集群搭建、Storm 配置文件配置项讲解

3) Storm 常用组件和编程 API, 使用 Maven 构建项目, 进行 Topology 程序编写

4) 消息队列 Kafka、与 Storm 集成

## 4、项目应用

1) 电商实时推荐业务讲解

2) Storm 实时 ETL 应用讲解

## 10.5、企业级大数据平台 CDH 5.3.x

此阶段课程价格: 1699 元

## 1、实时内存计算 Spark

1) Spark 应用场景、与 MapReduce、Storm 的比较和优势

2) Spark 概念和编程模型、Spark 原理

3) Spark Standalone 集群安装部署

4) Spark 企业项目应用举例

## 2、回顾大数据处理框架

- 1) 数据存储的实质
- 2) 数据处理工具（导入导出、数据分析）

## 3、大数据平台核心管理

- 1) HDFS 数据备份和数据迁移
- 2) 集群的升级本质和升级过程

## 4、CDH 5.3.x 安装

- 1) CDH 4 和 CDH 5 版本说明
- 2) 大数据平台管理工具 CM5.x 离线安装
- 3) 使用 CM5 安装部署配置监控 CDH 5
- 4) 如何使用 CM5 管理大数据平台、CM5 功能介绍演示
- 5) Cloudera Hue 安装、使用、说明

## 10.6、项目总结与面试

- 1) 企业级大数据三大项目类型
- 2) 面试简历编写（从招聘需求入手编写简历）
- 3) 项目面试讲解
- 4) HDFS、HBase、MapReduce、Storm、Spark 相关面试讲解

**注：由于新技术不断出现和成熟应用，课程中也会随着增加新内容。**

# 云帆大数据学院

---

## 11、赠送内容

赠送价值 400 元的《Storm 入门课程》;

**赠送价值 599 元的国内首套《企业级 Hadoop 2. x 面试课程》;**

赠送《企业级 Hadoop 2. x 入门课程》;

优先享受就业推荐内容

## 12、学费情况

《企业级 Hadoop 2. x **项目实战**课程》**原价 16999 元，优惠价 8999 元;**

**每期内容都更新，每期都会涨价，把握机会哦。。**

## 13、开课情况

### 上课规划

花钱培训的目的无非是能快速上手大数据工作、或换一份高薪的工作；由于大部分学员都是在职的，学习时间是相当紧张和宝贵的，我们的课程规划如下：

- 1、通过互联网实时视频软件（如 QQ 群视频、YY 群视频等），进行实时在线授课
- 2、上课周期为一百二十天，总共约十八周，每周三次课，每次课二个小时，共六十次课左右
- 3、每周一会将本周课表通过邮件方式发给大家，如有时间变动尽量提前半天通知



# 云帆大数据学院

---

4、授课方式，实时在线上课、课后视频辅导、讲义笔记代码复习和课后讨论答疑交流

## 腾讯课堂

云帆大数据学院成立于 2013 年，已入驻腾讯课堂，精彩公开课不断，上课地址：<http://yfteach.ke.qq.com>

## 14、报名流程

**报名流程：** 咨询课程顾问→填写报名表→支付宝缴费→付款核实→进入正式学员群→获取学习资料→正式学习

官方唯一支付宝：[yunfanwl@sina.com](mailto:yunfanwl@sina.com)

（新支付宝 [yfteach@sina.com](mailto:yfteach@sina.com) 将于 2015 年 8 月 1 日启用）

其他付款方式，请咨询课程顾问或是致电：**18518270792**

## 15、疑问解答

### 1、 担心网上学习没有氛围、没有自制力怎么办？

大家对于网上学习的担心是孤独感和缺乏目标造成的。在云帆大数据学习群中，全班同学在正式学员群中交流讨论、分享学习成果，让你感觉你的学习并不孤独；

老师会控制大家的考勤和学习进度，了解每位同学的学习情况，对任何一名同学都不抛弃、不放弃，带着大家一起向着未来那份满意的工作冲锋！

## 2、和自学有什么区别？

一个人看书、看视频学习的问题有：一个人孤独的学习，很难坚持下来；一个简单的问题自己搞了几天都搞不定，浪费了很多时间，如果有人指导的话可能一分钟就搞定了；不知道自己学的怎么样，感觉自己“都学过了”，可能又感觉“什么都不会”；求职时，技术固然很重要，但是简历、面试更重要，自学时没人指导就业，很难找到满意的工作。

云帆大数据为同学们营造大家一起学习的氛围，并且对同学进行严格的跟踪管理；老师一天超过 12 小时在线为同学们答疑；老师会对每位同学的学习情况进行跟踪，确保大家学的踏踏实实；毕业前，老师会专门讲解“如何写简历、如何投简历、如何面试、如何谈薪资避免贱卖”，并对每位同学的就业全程进行指导；

**云帆大数据学院与国内知名猎头公司战略合作，针对有跳槽意向的学员，优先推荐使用。**

## 3、学习中遇到问题怎么办？

老师从早上 9:00 到晚上 23:30 都在线，可以通过 QQ、微信、电话等随时问题，老师会及时解答问题。有问题直接语音提问，免去打字的麻烦；代码有问题直接用手机拍下来发给老师；遇到复杂的问题，老师还会通过 QQ 远程协助等方式进行解答。

## 4、上课时间是怎么样的？

每周两到三次课程，每次课程 2-3 个小时，20:30-23:00 上课，有时会相应的延长一段时间；课程表每周一会发送到各位正式学员邮

# 云帆大数据学院

---

箱。如果临时有事请假，也可以拿到当天的课程视频、资料；下次课程的时候在一起在线学习。

## 5、对网络带宽要求是怎么样的？

只要流畅的观看优酷等的视频，一般就没问题。为了解决大学的网络到晚上较慢的问题，我们还提供网盘视频下载，您可以下载下来再学习。对于直播课程也会提供录像供大家下载。

## 6、课程需要学多长时间？

在线课程项目班为 4 个月左右，视频学习的同学根据学习时间的不同，吸收能力不同，这个时间各有长短。

## 7、如果第一次没有学会，后面能免费学习吗？

一次报名，终身学习，只要同课程内容有更新，都可以免费学习的；

## 7、参加学习后可以推荐工作么？

云帆大数据与业界多家公司有合作，每期结束后，我们将免费推荐优秀的学员给业界优秀的企业；同时，我们与业界著名猎头公司建立了合作关系，我们将遵从学员的意见，推荐优秀的毕业学员给猎头公司；

## 关于云帆大数据学院 hadoop 学习说明

**其一**，现在一般使用做大数据存储和处理的，使用 Hadoop 框架的公司，项目都是比较的，或者是一个大项目中的一个小部分关于大数据查询与数据分析的。由此的话，涉及到的东西非常多，并不是仅仅数据存储和写写 MR 或者 HiveSQL 那么简单，抽出来的话，也是非常

创英才典范学院 造百年在线教育

[www.yfteach.com](http://www.yfteach.com)

困难的；

**其二**，就目前来说，做大数据存储和数据处理，基础知识非常的重要，不能仅仅说可以做，就好了比如数据的存储这块，就需要了解 HDFS 存储结构原理思想，只有这样的话，才能对具体实际项目进行综合考虑，看如何去做，一般大公司，都是需要自己进行修改源代码进行调整的，之所以这样，Hadoop 框架发展才这么快，很多问题被大家发现，也在逐步完善之中

**其三**，对于员工进入公司云计算和大数据的，一般公司都是由自己的需求的，而不是像 JEE 那样应用系统，整体思路差不多就可以做了，而是需要具体分析，研讨，进行研发，去做的。如果对 Hadoop 框架原理都不知道，仅仅说我会用，难道就是简单的调用 PAI，编写代码吗？ 这肯定是不可以的

**其四**，网上的资料确实很多，但是也很杂，最好的资料是官方的 Reference ，可是真正看的人有多少呢。又有多少人 真正在自己学习 hadoop 的时候，坚持下来，去从网上查资料呢，坚持学习下来呢。至少从我在企业中培训的人员来看，很多人缺少积极性的；换个角度，网上去所搜资料 ， 需要时间、精力啊