### 课程：Python爬虫与数据分析培训课程



业务架构概述、业务信息收集

业务角色分析-Business Role Model

0.5天

精讲、案例



业务领域建模-Business Domain Model

业务规则模型-Business Rules Model

精讲、案例



业务流程模型-Business Process Model

业务资源模型-Business Resource Model

精讲、案例



业务用例模型-Business Use Case Model

业务组件化模型

精讲、案例

0.5天

0.5天

0.5天

**【课程背景】**

本课程适合有一定Python编程经验的研发人员。本课程讲解了如何使用Python来编写网络爬虫程序, 内容包括网络爬虫简介 , 从页面中抓取数据的三种方法 , 提取缓存中的数据 ,使用多线程和多进程来进行并发抓取 , 如何抓取动态页面中的内容 , 与表单进行交互 ,处理页面中的验证码问题, 以及使用Scarpy爬虫框架来进行数据抓取 , 并在最后使用介绍的数据抓取技术对几个真实的网站进行了抓取。

**【课程时长】**

3天(18个课时)

**【培训对象】**

软件工程师， 即将从事爬虫相关技术管理的项目管理人员.

**【课程目标】**

* 掌握主流业务架构设计方法、理念，具备业务规划、设计能力

**【课程环境】**

* 操作系统: Linux
* Python版本: Python 3.x
* 集成化开发工具: Pycharm
* 虚拟环境管理器: Anaconda

课程涉及的Python 模块有 urllib、BeautifulSoup、lxml、Scrapy、Requests、Selenium等, 还有MySQL 数据库。

**【课程大纲】**

第一章：环境准备与入门 (2个课时)

* 什么是爬虫
* 为什么推荐使用Python写爬虫
* Python开发环境搭建和简介
* HTTP简介
* HTML/XML/Json简介
* MySQL/Apache安装和配置
* 爬虫框架介绍
* robots规范与爬虫原则

第二章：Python编程入门 (2个课时)

* 代码格式
* 关键字
* 基本语法
* 容器与字符串
* 函数
* 面向对象
* 文件读写
* 多线程与多进程
* 错误与异常

第三章：**业务角色分析-Business Role Model**