



谭泽斌

## 求职意向

大数据开发工程师

## 联系方式

手机: 18735474916

邮箱: tanzebin@foxmail.com

博客: [www.tanzebin.xyz](http://www.tanzebin.xyz)

## 个人信息

出生年月: 1993.02

籍贯: 山西朔州

## 英语

CET6, CET4

## 荣誉证书

国家励志奖学金

北京邮电大学学业奖学金一等

重庆邮电大学优秀学生奖学金二等

全国计算机等级考试四级



## 教育背景

2017- 至今	北京邮电大学	硕士	电子科学与技术
2012- 2016	重庆邮电大学	本科	通信工程



## 专业技能

- 熟悉 Java, 了解 Scala、Python, 掌握常用的数据结构和计算机网络的基础知识;
- 了解 Spark, 掌握 Spark RDD, Spark SQL, Spark Streaming 的基本编程应用;
- 了解 Hadoop, 理解 HDFS 分布式文件存储系统原理
- 了解 Linux, 掌握常用的 Linux 命令;
- 了解 Flume 等日志收集, 掌握 Flume 和 Kafka 的对接;
- 对于大数据生态系统, 了解 Zookeeper, Hive, HBase, Sqoop;



## 项目经历

## ➤ 基于 Spark Streaming 日志实时流分析

系统环境: Spark + Hadoop + HBase + Flume + Zookeeper + Kafka

项目目标: (1) 通过日志统计网站上不同课程访问的数量;  
(2) 统计从其他网站引流过来的课程的访问量。项目要点: (1) 编写 python 脚本产生模拟数据, 定时执行脚本, 将日志数据写到文本文件;  
(2) 配置 Flume 采集数据, 将日志数据 sink 到 Kafka;  
(3) Spark Streaming 实时消费 Kafka 的数据, 进行数据清洗和分析, 然后把分析结果存到 HBase。

## ➤ 基于 Spark 的离线电影推荐系统

系统环境: Spark + Hadoop + Hive + MySQL

项目目标: 根据不同的用户 ID 为其推荐合适的电影

项目要点: (1) 把原始数据存入 HDFS, 然后经过结构化导入到 Hive 表;  
(2) 使用 Spark SQL 把数据切分成训练集和验证集;  
(3) 使用 Spark mllib 的 ALS 算法生成训练模型;  
(4) 使用模型, 产生推荐结果。

## 实习经历

## ➤ 中国农业银行软件研发中心

## 项目 (一): 负责一个 Web 打印机控制系统的设计和实现。

(1) 该系统基于 B/S 架构, 客户浏览器访问服务器时, 该打印系统能够自动安装, 过滤掉虚拟打印机, 直接将打印作业输出到物理打印机。

(2) 基于 PDFBox 库, 和 Java Web Start 技术, 通过 Servlet 动态生成 jnlp 文件, 实现打印系统的自动安装和打印参数的传递;

## 项目 (二): PDF 文档行间距水印编码的加密和基于图像识别的水印提取

(1) 通过 iText 精确定位每行文字的坐标, 根据二进制的水印比特流适当调整某些行的位置, 将水印信息嵌入;

(2) 提取水印主要用到了 OpenCV, 对文档扫描件图像进行中值滤波, 二值化, 然后进行水平投影, 分析行间距, 提取水印信息。