

# 李海成



联系电话: (+86) 15652322066

电子邮箱: [lihaicheng\\_work@163.com](mailto:lihaicheng_work@163.com)

出生年月: 1992 年 06 月

最高学历: 硕士 (2019 年 06 月毕业)

籍贯: 山东省烟台市

求职意向: 数据工程师 (ETL)

## 教育背景:

---

2016 年至今	中国科学院软件研究所	硕士	计算机技术
2012 年至 2016 年	山东大学 (211 . 985 重点大学)	学士	计算机科学与技术

## 个人介绍:

---

学习能力强, 吃苦耐劳, 富有钻研精神, 服从意识强, 团队意识强

熟练掌握 java 语言, 熟悉 linux 系统, , 熟悉 python 语言, 熟悉 sql 语言

熟练使用 HDFS、MapReduce、Hive、HBase、Flume 等框架

对 Hadoop、Spark 等大数据开源框架有强烈的学习热情

## 实习经历:

### 真融宝 数据实习生 (大数据风控部)

工作: 公司有贷款业务, 对于每笔贷款, 需要根据贷款人的大量信息, 评估偿还能力, 我负责开发软件, 用于特征提取与计算。

收获: 熟悉企业开发流程, 熟悉数据开发流程, 巩固 python 编程。

## 项目经历:

### 移动端信息安全攻防

描述: 发现 Android API 的设计漏洞, 实现概念验证攻击进行隐私数据窃取

工作: 实验中发现了关于 Camera2 的 Android API 存在一些危险性, 实现了一款恶意软件, 通过 service 和 activity 配合, 利用反射重现了谷歌修复过的漏洞, 用于在 camera 使用

前、中、后随时抢占摄像头，拍摄用户隐私信息。实现爬虫爬取大量 apk 文件，研究分析。与谷歌交流并获得肯定回复，完成论文并向 A 区会议 ICSE 投稿。

收获：巩固 java 编程技能，学习了 Android 恶意软件设计及缓解技术。学习 python 爬虫设计。

收视率统计

描述：利用机顶盒数据，统计某个节目的收视率

工作：搭建起 HDFS 集群，利用 Flume 做数据采集并汇总到 HDFS 上，编写 MapReduce 程序进行数据清洗，将半结构化数据整理成规则的数据格式，单独搭建 Hive 客户端，做进一步统计，最终获得了每个节目在分钟、小时、天三个时间维度下的收视率，并存储于 Hive 中。

收获：掌握了 Hadoop 中 HDFS、YARN、Zookeeper 集群的搭建及架构原理，学会使用 Flume 做日志采集与聚合，学会使用 Map Reduce 进行数据清洗，学会 Hive 关于表的操作，分区的操作。

网站爬虫

描述：爬取视频网站电视剧信息并加以存储

工作：编写 java 爬虫，爬取优酷网关于电视剧的信息，如剧名、导演、收视人数等，搭建 HBase 集群用于存储数据。

收获：掌握了一系列爬虫相关的开源库的用法，如利用 HttpClient 做页面下载，HTMLCleaner+Xpath 做页面解析，掌握了应对反扒虫的技巧，学会使用 HBase 在各种场景下的建表策略，包括 rowkey 的设计原则和宽表、窄表的选择，以及对 HBase 数据的增删改查操作。

获奖情况：

“蓝桥杯”大赛 java 软件开发组	山东省省级三等奖
数学建模比赛	山东大学校级二等奖
科研立项	山东大学校级二等奖
挑战杯比赛	山东省省级三等奖
研究生奖学金	学业奖学金二等奖