

个人简历

个人信息

姓 名： 吕艺恩 性 别： 男
出生年月： 1993 年 12 月 籍 贯： 陕西省咸阳市
毕业院校： 郑州大学 (211) 学 历： 一本
专 业： 电子信息工程
联系电话： 13220130998 电子邮件： lvvien@qq.com



求职意向

目标职能： 数据开发工程师 **目标地点：** 北京

工作经历

单位名称： 58 到家
职位名称： 高级大数据工程师
工作时间： 2017 年 9 月-至今
参与项目： 1、58 到家数据仓库建设
2、到家平台反作弊实时流数据

单位名称： 北京捷通华声科技股份有限公司
职位名称： 大数据开发工程师
工作时间： 2016 年 3 月-2017 年 8 月
参与项目： 1、灵云大数据开放统计平台
2、国家电网 95598 智能语音质检分析
3、绿瘦集团智能语音质检分析

单位名称： 中兴软创科技股份有限公司
职位名称： 数据开发工程师
工作时间： 2015 年 6 月-2016 年 3 月
2014 年 10 月-2015 年 5 月 (实习)
参与项目： 1、联通数据服务云
2、信息安全大数据分析及预警平台

专业技能

大数据技能：

- 1 掌握 MR,Spark 计算框架,深刻理解其底层运行机制;
- 2 深刻理解 hdfs 分布式文件系统存储结构和高可用原理;
- 3 熟练掌握 scala 编程,能熟练运用 scala 进行 spark RDD 编程;
- 4 掌握 Spark streaming,Storm 流式处理框架,熟练掌握流式编程模型及其流式数据开发;
- 5 熟悉 flume,kafka 等日志收集分发框架的使用,与 storm,spark 进行整合进行数据实时处理;
- 6 熟练运用 hive 数据仓库工具,熟练掌握基于 hive sql 的 ETL,进行数据仓库建设;
- 7 深刻理解数据仓库维度建模理论,掌握数据仓库的分主题分层设计方法
- 8 熟悉 hbase 数据库的使用及其编程,深入理解 Hbase 的存储原理和存储架构;
- 9 熟悉 kylin,elasticsearch,kibana,Sqoop,zookeeper,azkaban 等大数据相关组件的使用;

基础技能：

- 1 精通 Java 编程,深刻理解面向对象设计思想,熟练使用 IO 流操作、集合框架、多线程等技术。

熟悉反射、注解、动态代理等 Java 高级技术；

- 2 掌握常用的数据结构和算法；
- 3 熟悉常用的 java 设计模式（如单例、装饰、工厂等设计模式）；
- 4 熟练掌握 scala 编程,熟悉 python 脚本语言，有一定 python 开发经验；
- 5 熟悉 linux 操作系统及常用的 shell 命令，能够编写简单的 shell 脚本；
- 6 熟悉 JVM 内存模型及垃圾回收机制；

数据库技能：

- 1 熟练掌握 Oracle、MySQL 关系数据库及其 sql 编写；
- 2 熟练掌握 redis 内存数据库的使用；

JavaWeb 技能：

- 1 熟悉使用 Spring, SpringMVC, Mybatis, Hibernate 等 web 相关技术；

项目经历

项目一：58 到家数据仓库

软件架构： Hadoop+Spark+Hive+ Kafka+Flume+Sqoop+Kylin

项目描述：

数据仓库基于内部的可视化数据开发平台建设，主要承担面向数据分析人员进行快速数据提取，数据的可视化报表支持，面向决策部门进行业务决策支持，反映业务并驱动业务增长，业务流程优化数据支持，日常运营数据支持，统一数据口径，提供公司整体数据支持

职责描述：

- 1 负责 58 到家平台业务的数据仓库基础模型设计与优化，持续模型迭代开发
 - 2 收集数据需求业务数据调研，沟通制定相关流量数据埋点规范并接入采用 spark 离线清洗入库
 - 3 完成数据模型 ETL 工作，解决 ETL 中的性能问题
 - 4 完成面向业务的各种指标报表、数据提取与 OLAP 分析数据支持
 - 5 响应业务方的数据需求，提供面向业务的离线与实时的数据支持
-

项目二：到家平台反作弊实时流数据

软件架构： Spark streaming + Kafka+Flume+Redis

项目描述：

该项目针对到家平台下单流程进行是否作弊，根据规则判定，对作弊行为进行拦截

职责描述：

- 1 前期实时数据需求沟通
 - 2 编写 spark streaming 任务将接入业务库 binlog 数据与下单环节后端埋点进行实时清洗，向策略同学提供反作弊实时数据流
 - 3 后期为反作弊系统效果评估建仓，统计计算反作弊相关指标
-

项目三：灵云大数据开放统计平台

软件架构： Spark+Hbase+Elasticsearch+Mysql+sqoop+hive+flume+azkaban

项目描述：

该项目主要对灵云人工智能公有云平台和 SDK 端相关人工智能能力如 ASR、TTS、NRU、OCR 等十种人工智能能力的使用中产生的数据的实时与离线统计分析。通过统计分析各应用、用户、能力、终端、请求数据及语料等相关数据，提取统计相关业务指标，并形成相关报表提供给运营部门和决策者制定有效的策略方案。提供公有云平台的日常运行日志的统计分析，提供历史语料数据存储与检索导出供算法研究部

门用于模型训练优化。

职责描述：

- 1 语料数据收集与处理方式重构，采用 spark 任务将原始语料与识别结果解析入库到 hbase
 - 2 云服务请求日志数据离线清洗到 hive 表中
 - 3 与运营平台、能力平台等进行平台数据对接
 - 4 使用 hive 进行业务数据建仓，云平台数据 etl，实现运营等其他部门的指标报表需求，提供后台报表页面离线数据支持
-

项目四：国家电网 95598 智能语音质检分析

软件架构：Kafka+Storm+Redis+Spark+Hive+Elasticsearch+Hbase+Oracle

项目描述：

该项目主要实现对国家电网 95598 热线服务进行智能质量检测与相关业务分析，对话务员的服务态度、服务质量和水平服务的智能监督与评测，辅助人工进行话务质检。采用 ASR 技术将通话语音转换为文本，同时对海量的通话文本数据进行中文分词、词频统计、关键词提取、话务的时间地域维度分布、来电原因、客户的联系方式住址、情绪判定、主题分析、业务关注点、热词统计、是否重复来电等信息进行挖掘提取与分析处理，实现国家电网话务数据的可视化展现。实时质检实现在线实时识别来电客户的来电原因等，向话务员推送相关业务的标准话术与业务，辅助提取客户相关信息提升客服人员的效率，为质检班长实现在线进行语音质检工作的能力。

职责描述：

- 1 参与前期项目分析，数据采集、离线和实时处理部分功能模块设计，编写相关设计文档
 - 2 storm 流式数据处理任务开发，进行规则匹配进行业务术语实时推送
 - 3 spark 离线话务数据清洗、批量进行有关标签标注，根据相关业务规则进行离线质检打分
 - 4 采用 hive 进行数据的抽取、统计与维度分析，建设话务系统的离线数据仓库，为相关报表页面提供数据支持
 - 5 将话务文本及相关数据建立 ES 索引进行全文检索,并结合 hbase 建立二级索引，加速历史语料的检索与实时读取
-

项目五：绿瘦集团智能语音质检分析

软件架构：Flume+Spark+Hive+Elasticsearch+Hbase+MySQL

项目描述：

该项目主要实现对绿瘦集团热线服务进行智能质量检测与业务分析，辅助人工进行话务质检。采用 ASR 技术将通话语音转换为文本，对识别文本进行数据清洗统计，同时采集相关话务的随路数据，进行离线统计分析处理。通过对来电用户的通话内容利用挖掘算法进行文本分析，对来电咨询客户进行用户画像（是否结婚、抽烟、肥胖、职业...）。将相关数据部分索引至 ES 提供数据检索。

职责描述：

- 1 参与前期项目分析，数据采集功能模块设计，离线统计分析模块设计，编写相关设计文档
 - 2 随路数据收集工作，利用 flume 将随路数据收集到 hdfs 中
 - 3 采用 spark RDD 进行离线话务数据清洗，对通话文本进行规则匹配，对用户画像相关属性进行标签标注，完成离线数据批量处理
 - 4 利用 hive 进行数据抽取、维度分析与指标计算，建设离线数据仓库，配置 sqoop 将相关统计计算指标同步至业务数据库支持离线数据报表展示
 - 5 将用户标签、有关质检数据、部分文本分析结果等添加 ES 索引提供分布式快速检索
-

项目六：中国移动信息安全大数据分析及预警平台

软件架构：Hadoop+Kafka+Spark+Redis+Hive+Mysql

项目描述：

该平台针对信息安全各项治理指标进行统一采集、统计分析和数据挖掘，对全国各地的信息安全态势进行整体分析，实现信息安全治理的整体情况和发展趋势进行整体情况感知以及进行有效预警。通过与集管平台、策略运营分析平台、互联网监控平台等平台数据的采集并统计监控各平台数据服务情况。通过离线计算对全国电信诈骗、不良信息、信息及网络安全等数据多维度统计、通过流式处理实现实时垃圾信息拦截、投诉等实时展示以及对数据采集监控告警。

职责描述：

- 1 编写离线数据收集脚本，采用 flume 将离线批量数据收集到 hdfs 中
- 2 编写 spark 任务对采集的离线数据文件进行数据预处理，生成规范数据文件写入 hdfs
- 3 将离线 hdfs 数据 load 到 hive 表中基于维度建模方式开发仓库模型表，根据业务需求进行指标数据报表开发，并同步 mysql 中供前端展示
- 4 参与编写 spark streaming 实时流任务，对诈骗电话、垃圾短信、不良信息以及各平台采集数据等部分实时指标数据统计

项目七：联通数据服务云

软件架构：Spring+SpringMVC+Mybatis+Mysql+Jquery

项目描述：

该系统承担对联通云门户提供一个可视化管理平台，实现全国联通分公司实现统一线上业务管理功能，对用户统一标签库、共享模型库、共享分析库等相关标签与数据挖掘模型维护，结合后台大数据应用对用户数据离线计算，面向各业务部门及省份用户提供数据需求、数据服务、模型应用、用户标签、行业报告等各类大数据产品及应用，建立企业内的数据产品应用支撑。

职责描述：

- 1 进行统一标签库模块部分功能开发（包括前端页面、后台业务逻辑、数据接口层）
- 2 数据分析报告展示及导出等业务功能开发
- 3 项目文档的编写

自我评价

- 1、参与过多个项目的数据开发，熟悉常见数据处理流程，能快速融入团队；
- 2、学习新事物的能力较强，能较快的掌握一门新技术；
- 3、对工作尽职尽责，乐于从事有挑战性的工作；