|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **姓名** | 李澍楠 | **性别** | 男 | **年龄** | 27 | **从业年限** | 4年 |
| **毕业院校** | 大庆师范学院 | | | **学历** | 本科 | **专业** | 自动化（电子） |
| **一、专业技能描述** | | | | | | | |
| 1.熟练掌握hadoop生态圈系统CDH组件，如Hive，Hbase，Spark，zookeeper  2.熟练掌握流批处理框架flink运行时工作原理，设计开发测试流式数据处理算子。  3.熟练使用java，python开发语言，springMVC，springboot, mybatis-plus, flask等主流框架。  4.熟悉各种数据库和搜索引擎的使用，如Mysql。ES，Solr，clickhouse.  5.熟练使用消息中间件：redis，kafka，MQ等  6熟练运用Docker，gitlab，Jenkins等集成kubernetes实现代码CI/CD部署。  7.熟悉一些典型的算法，例：关联规则，聚合，逻辑递归，神经网络。 | | | | | | | |
| **二、主要工作经历** | | | | | | | |
| 2020.05-至今 北京九章云极科技有限公司 中级研发工程师  2019.02-2020.05 文思海辉金融 大数据服务工程师  2017.08-2019.01 明码（上海）生物科技有限公司 大数据开发工程师 | | | | | | | |
| **三、主要项目经验** | | | | | | | |
| **2020.12-至今 联邦学习 中级研发工程师Ⅰ**  项目描述：帮助企业客户保证数据隐私安全及合法合规的基础上，实现共同建模，提升AI模型的效果,孵化出分布式机器学习平台即联邦学习,基于联邦学习框架Fate，根据多参与方数据分布的不同，灵活运用横向，纵向，联邦迁移实现数据建模。  项目职责：  1，负责海量数据抽取，处理，存储。以及高并发场景下交易查询优化。完成投产部署。  2，负责系统前后端接口设计，开发，调试，优化，与fate原生代码交互，算法运行。  3，负责系统流式功能算子的开发调试，优化高并发，低延迟场景下功能算子，妈祖tps等指标。  4，负责跨组之间的项目沟通，进度协调，编写部署投产文档模板，实现稳定投产部署。  5，负责调研并调试fate原生算法支持，多host下，统计分析（均值，方差，中位数），机器学习（逻辑回归，SecureBoost），神经网络（DNN）等。  6，负责原生fate工作流和模型仓库管理模块开发，调试。如工作流管理，以及模型预测，  7，负责日志模块管理，包括计算日志入库，联合任务凭证管理，存证发现追踪等。  **2020.5-2020.12 浦发RT二期联机查询（大数据应用项目）中级研发工程师Ⅰ**  项目描述：实现浦发银行，老系统海量数据的迭代更新，做到新兴技术对主流业务的逻辑加工及功能扩展，完成新一代系统改革创新，加强银行数据业务处理系统的健壮性，稳定性，高可扩展性等特点。埋点解决业务逻辑处理细节。  项目职责：  1. 负责梳理业务加工逻辑，设计并制定相应查询方案。  2. 利用spring，cxf，hbase，ES，clickhouse等技术栈开发优化联机查询业务。  3. 通过接收ESB系统请求报文，结合技术栈解析请求报文并及时做出响应，最终封装响应报文输出返回。  4. 优化解析报文后做出响应的时间，测试并排查技术间适配性问题。  5，负责海量数据抽取，处理，存储。以及高并发场景下交易查询优化。完成投产部署。  **2019.12-2020.4太保集团客户关系图谱平台（图谱模型项目）大数据开发工程师**  项目描述：搭建机器学习平台，通过平台提供机器学习平台在线建模功能。做到数据接入、建模分析、模型训练、模型生产化一站化实现，快速的发掘大数据及人工智能在保险行业的场景及价值，赋能到业务生产的各个环节。  项目职责：  1. 基于hive获取用户特征行为数据同步到solr和hbase  2. 清晰分析数据，进行ETL操作  3. 拉取具体使用数据，调用模型预测，跑批出数据模型。  4. 输出结果。  **2019.02-2019.12 国债数据条线数据服务支撑**  **项目描述：根据新一代系统开发需求规范，**着重处理企业内非结构化、半结构化以及海量的结构化数据。依托大数据处理、分析、挖掘手段，为管理分析类应用提供数据支持。  **开发环境与技术：**  开源组件：  Zookeeper,Kafka,hbase, Hadoop,Hive,Spark,Mysql ,nginx等  **本人职责：**  **1、**基于Hadoop平台的系统管理，代码开发优化，以及开发文档的编写。  2、利用统一调度工具Control-M开发配置大数据平台的工作任务调度。  3、利用ETL工具informatica开发配置优化数据流管道。  4、编写并调优Sql，解决大数据平台和数据仓库不同层级之间的数据对接。  **2017.08 - 2019.02 微基因项目优化体系**  **项目描述：**基于基因大数据的系统开发、维护、部署上线，保证医疗基因数据（VCF）高效读写、海量存储、实时分析等系统操作的稳定运行。构建基因医疗大数据的系统体系。  **开发环境与技术：**  开源组件：  Zookeeper,Kafka,Hadoop,Hive,Spqrk,Mysql ,nginx等  **本人职责：**  1、将oss康码基因数据（VCF）抽取到Hadoop平台的hdfs待预处理操作。  2、转译VCF文件内容，提取相关字段数据，数据排序以及数据统计等相关数据清洗。  3、针对不同层不同阶段的代码优化，提高运行效率。  4、利用hbase的builkload对hdfs上转译后的基因数据做hbase的高效上传。  5、结合Kafka、Spark和HBASE特点，利用Azkaban调度工具，共同实现数据的海量存储，高效查询。  **2018.02 - 2019.1 来因部落日志分析系统**  **项目描述：**主要是将APP端的服务日志，服务器端的加载日志，做到报错问题的定向分析，以及用户行为的动态分析。  **开发环境与技术：**  开源组件：Zookeeper,Kafka,Elasticsearch,logstash,kibana,hbase,等。  **本人职责：**  1、部署开发调试ELK日志分析系统，将日志分为多个层，用户层、故障层、模块指标层等等。  2、在elasticsearch中创建不同的索引存储不同的日志层。  3、定时监控故障层的日志数据，反馈报告故障发生失败的原因，及时调整优化。  4、利用Kibana生成界面对用户层的数据进行折线图，扇形图，柱状图等，图形胡的展示。  5、针对不同模块指标，制定相应模块的调整方案。  **2017.08-2018.04 CI/CD集成自动化部署**  **项目描述**： 主要是提升开发效率，实现开发，测试，上线发布等自动化流程体系，减少中间环节人工误操作带来的不确定性的发布风险。  **开发环境与技术：**  开源组件：  Jenkins,docker,gitlab,maven,shell,kong等。  **个人职责:**  1、利用docker容器，gitlab，Jenkins配置代码集成自动化部署（CI/CD）。  2、针对前端和后台的不同代码配置相应的dockerfile，docker-compose文件，将项目封装成docker容器，并实现打包部署上线。  3、配置gitlab-webhook连接Jenkins实现项目的自动触发自动部署。  4、编写pipline脚本，规范gitlab秘钥，各个项目全局凭据，提高项目署效率。  5、利用kong网关配置前端H5项目的访问域名，实现kong网关的灵活运用。 | | | | | | | |