# 简历-ETL开发工程师

|  |
| --- |
| * 个人信息 |

姓名：喻汉生 性别：男

专业：计算机科学与技术（本科） 工作经验：2年

邮箱：[15617577080@163.com](mailto:15617577080@163.com) 手机：15617577080

|  |
| --- |
| * 技术生态圈 |

* 熟悉使用Flume完成数据采集、分流等功能，理解Flume的组成架构及工作原理。能够搭建多种拓扑结构的数据采集系统，使用Ganglia完成数据监控，并进行传输优化。
* 熟悉Hadoop大数据生态体系，能够搭建高可用的Hadoop生态集群。
* 熟悉HDFS组成架构、存储原理、数据的读写流程，及优化HDFS工作性能。
* 熟悉MapReduce工作机制与运行原理，用Tez替换计算引擎，提升Hive计算查询速度。
* 熟悉Mysql关系型数据库，建SQL索引优化性能。
* 熟练编写sql语句，完成复杂的sql需求。
* 熟悉数据仓库的设计、搭建、表关系的模型构建，使用Hive 实现海量结构化日志的分层管理与查询分析。
* 熟悉常用的Linux的Shell命令，能在Linux系统下搭建开发环境，及主机性能分析。
* 熟悉Zabbix监控各个主机性能，并完成主机性能报警及程序监控。
* 独立搭建并使用Jenkins,Git，Maven私服等持续化集成部署软件。
* 了解JavaEE框架(SSM)并能够简单搭建Web可视化工程。
* 了解SpringBoot框架，完成数据接口的发布。

|  |
| --- |
| * 工作经验园 |

公司名称：大地零一（深圳）科技有限公司 职位：大数据研发工程师

* 参与项目：
* 日志埋点分析系统（数仓项目）
* 中企经营分析系统（BI报表）
* 主要职责：
* 参与并讨论公司架构选型，技术选取，平台功能规划与设计。
* 搭建公司数仓服务平台，日常维护平台软件，优化性能。
* 负责公司各个技术点的调研，协助同事做技术支持。
* 负责数仓数据的接入，清洗，业务主题建模。
* 结合业务需求设计数据分析场景，并形成后台实现方案。
* 独立完成业务数据建模，指标分析，并将结果转化为运营指标。
* 优化Kettle，Mysql,Flume ,Kafka,Hadoop参数，实现系统调优。
* 负责公司集群的维护，监控，异常处理及报警，实现服务器集群正常运行。
* 学习新技术，提高整个平台的计算能力和效率。

|  |
| --- |
| * 项目经验区 |

* 日志埋点分析系统（数仓） 2018.8 – 2019.7

开发环境：IntelliJ IDEA+Jdk1.8+Maven+Git

运行环境：Flume-1.7.0+Kafka-0.11+Hadoop-2.7.2+Hive-1.2.1+Mysql5.6+可视化

项目描述：主要对“E肚仔”App用户行为数据和业务数据库的数据进行存储和分析统计。针对各种用户，商品等多维度去统计分析，根据分析结果优化App。改变不同人群、不同地区的销售策略，帮助提升公司营业额以及市场占有率。

工作职责：

* 参与讨论数据仓库需求分析、分层设计。
* 将用户行为日志数据和业务数据采集至数仓数据原始层。
* 实现产品，运营陆续提出的需求，包括但不限于留存用户、连续活跃用户。
* 持续维护数仓健康状态，不断优化数据仓库性能。

技术要点：

* 数仓结构分析与设计
* 数仓构建为四层:数据原始层（ods）、数据明细层（dwd）、数据服务层（dws）、数据应用层（ads），保证了数据间解耦、分层处理、便于管理。
* 数据原始层（ods）：加载用户行为日志和业务数据，保持数据原貌。
* 数据明细层（dwd）：使用自定义UDF函数对用户行为日志进行解析，并过滤脏数据，对敏感数据加密，对用户重要信息脱敏，对表降维处理。
* 数据服务层（dws）：对数据轻度聚合，并将查询频度较高的表联合起来制作宽表，一表多用，解决大部分需求，提高工作效率。
* 数据应用层（ads）：对dws层的数据从不同维度聚合，结果写入Mysql。
* 数仓优化
* 初期使用Tez引擎，运行提升速度两倍。
* 数仓中通过数据建模理论，对数仓中的数据采用多冗余，低范式构建星型模型，减少关联查询，提高查询效率。
* Hive元数据管理：基于Keepalived实现Mysql HA，每日凌晨后做元数据备份。
* 合理设置小表阈值，在资源允许的情况下尽量使用MapJoin以避免Shuffle。
* 避免小文件，在Map操作之前合并小文件，在任务结束时合并小文件

工作成果：通过对商品各指标、用户行为主题、下单流程各指标进行分析统计，对 app 进行优化，针对不同人群，不同区域改变销售策略，帮助公司获得了更大营收和市场占率。

* 中企经营分析系统（BI报表） 2019.9 – 至今

运行环境： Kettle-7.1+Mysql-5.6+PowerBI

项目描述：通过接入“中企动力”的业务数据，对数据进行分析统计,实现不同维度的日报，周报，月报开发，为中企公司的报表提供数据支持，让产品和运营能够快速做出决策，反映公司业绩和营收额，供PM及管理人员掌握产品销售情况。

工作职责：

* 参与讨论系统的需求分析，架构搭建。
* 独立完成业务数据到数仓ST层的数据导入。
* 独立完成日报、周报、月报开发，包括但不限于销售、产品、在职人数等指标。
* 维护数仓系统稳定运行，优化数仓性能。

技术要点：

* 数仓为5层:（数据缓冲层）st、数据原始层（ods）、数据明细层（dw）、数据集市层（dm）、维度数据层（dim）
* 维度数据层（dim）：存放各种需要的维度数据，例如日期，产品，区域等维度。
* 数据缓冲层（st）：抽取业务方数据，保持数据原貌，只做缓冲，导完即清除。
* 数据原始层（ods）：抽取st层数据，对特殊数据进行格式要求处理，加上时间戳。
* 数据明细层（dw）：根据指标需求，通过对ods层的数据统计、分析、计算保存明细数据，比如一些比率值，这层就要把分子分母全部保存。
* 数据集市层（dm）：通过对dw层的数据按需处理，对外提供服务。
* 将Kettle的文件资源库迁移到数据库资源库，便于管理。
* 将编写好的各种报表Job使用定时任务，按需执行，任务失败报警策略。
* Kettle组件的优化，一个Job内各个组件的执行顺序控制优化。

工作成果：完成中企运营及产品提出的需求，按需展示报表，实时提供数据可靠性及准确性。

|  |
| --- |
| * 自我评价 |

* 勤奋认真，做事热情积极，较强的责任心。
* 环境适应能力强，能很快适应新的工作环境。
* 具备团队合作意识，善于协调团队中的人际关系。
* 能够虚心接受他人的意见，三省吾身。
* 始终积极向上，热爱学习。