**甘世玉**

女 5年工作经验 1990年11月 已婚

现居住于上海-杨浦区 本科 户口江西-九江  
手机：188888888888 E-mail：[ganshiyu@badou.com](mailto:ganshiyu@badou.com)

**求职意向**

大数据开发、数据仓库

**自我评价**

自学能力强，善于思考，吃苦耐劳，有良好的沟通能力，善于与他人相处，富有团队合作精神，爱运动。

**专业技能**

1、熟练linux、shell、pyhton、java、scala  
2、精通oracle、mysql、teradata、sqlserver数据库、sql的开发和执行计划的跟踪与优化  
3、精通hadoop、mapreduce、hive、hbase、spark、kafka、hue、oozie、impala  
5、精通数据仓库的模型设计、数据转换、抽取和调度的设计

6、熟练NLP、TF-ITF、LCS、中文分词

7、熟悉CB、CF推荐算法和Kmeans、朴素贝叶斯、逻辑回归

**工作经历**

2017.5 – 至今：XX**快递**

职位：大数据开发

汇报上级：数据架构师

企业规模：1000人以上（其中技术部门300人）

企业介绍：XX物流有限公司创立于2009年，总部位于上海青浦，是一家提供全国性快递服务的规模型快递企业

工作描述：

1. 对各大业务系统、日志系统，结构化和非结构化数据进行实时与离线的采集
2. 开发spark任务对数据进行清洗、计算、转换和推送，报表的开发以及业务系统的交互
3. 拉链表、维度表、事实表和宽表等模型设计与开发
4. 审核和优化spark任务，减少任务执行时间以及spark任务资源合理的分配与设置
5. 对海量数据列式存储，以及kudu表和hive表分区存储与优化，数据倾斜的处理
6. 运用NLP、TF-ITF、LCS等算法对文本进行相似度的匹配与归类

2015.8 – 2017.5：**XX XX快递**

职位：大数据开发

汇报上级：部门经理

企业规模：1000人以上（其中技术部门250人）

企业介绍：XX快递创于1993年是一家以经营快递为主的国内合资（民营）企业

工作描述：

1. 主要负责数据仓库的设计、报表的开发，etl工具数据抽取、转换和加载，存储过程、mr和sql的开发以及调度
2. 主要运用python、shell、sql、hive、sqoop完成日常的工作
3. 对转换、sql、存储过程和mr任务进行优化，以及指导和审核开发人员的sql
4. 数据库的存储优化，表的行列混式压缩并提高数据库查询性能、分区设计和分区自动化管理
5. 数据分析和统计

2014.2 – 2015.7：**XX XX信息科技有限公司**

职位：数据库开发

汇报上级：项目经理

企业规模：200人

企业介绍：XX XX信息科技有限公司是中国教育信息化领域的重要软件供应商和系统集成商

工作描述：

oracle数据库的安装与配置以及数据库的设计，数据接口的设计与开发，存储过程、方法和触发器的编写，数据库的备份、数据库的优化和数据库的日常维护 linux服务器的部署与配置

**项目经验**

**2017.5 – 至今：XXX数据分析系统V2.0 XX快递**

项目简介：

该项目是随着公司快递业务的增长，数据规模的增大，致力于给公司和网点人员提供可视化的数据监控和分析，提高快递的操作时效，降低公司成本，提高公司和网点的效益。该系统集成PC和APP，包括报表分析，成本收入预测，驾驶舱，品质管理，实时数据查询，大屏数据看板，数据监控等智能一体化数据分析系统，汇集公司所有业务系统的数据 。

个人职责：

1. 主要负责hadoop平台的拉链表、事实表和宽表的模型设计
2. 负责源系统数据实时与离线采集，接入hadoop平台
3. 开发数据采集统一框架，生成可动态配置源数据采集
4. 运用spark scala代码和sparksql的开发任务进行离线数据和实时数据的转换、清洗、计算。持久化到kudu的ONLINE实时层，hive 的ODS、EDW和DM离线层的
5. 基于python，针对各个job血缘关系，进行任务依赖的自动设置
6. 运用NLP、TF-ITF、LCS等算法对物品、订单和地址等文本数据进行相似度匹配和指标的归类
7. 统一管理和开发公共模型框架。审核spark代码和优化大于平均任务时间50%的job，以及指导前端开发sql和报表数据库的执行计划跟踪

**2015.8 – 2017.5：XX数据分析系统 XX快递**

项目简介：

该项目由oracle一体机和hive搭建成离线数据仓库。进行业务统计，数据监控、数据分析，快递时效和操作时效等，进而提高快递和操作时效、优化转运中心线路等BI商业智能系统

个人职责：

1. 负责数据抽取、转换和调度，模型报表的设计与开发，数据库的设计和存储优化

2、主要运用sqoop、kettle和加载文件的等技术对数据进行离线增量以及全量的采集

3、主要用存储过程、python开发的mr和sql进行离线的数据计算、清洗和转换，维护计算任务程序

4、设计任务调度框架与调度shell的代码的开发，可对任务的依赖、包括层与层的关系、任务并发数、重跑和实时的监控邮件，灵活的任务调度处理

**2016.1 – 2017.5：车辆结算系统 XX快递**

项目简介：

该项目自动实现全网网络车，网络加班车，外租车系统自动化结算，做到及时结算账单，及时下账，减少财务结算人员GPS监控对比工作，提高账单的准确率

个人职责：

1. 负责用sqoop同步车管宝系统mysql数据到hive

2、负责同步前端费用管理员设置的费率，利用mr生成账单费率拉链

3、运用mr和sql进行账单的离线的计算、开发与维护

**2014.3 – 2015.7：同济大学本研一体化教务系统 XXX XXX科技有限公司**

项目简介：

该项目整合同济大学本科生和研究生教务系统的需求功能，完成对学籍、选课、排课、考试和毕业审核的等的多功能系统

责任描述：

1、负责该项目数据库管理与开发，linux数据库服务器和应用服务器的部署与安装，负载的搭建  
2、本研老系统与新系统的之间的数据同步与过滤  
3、完成财务系统、人事系统和数据仓库数据交换的接口设计与编写，提交用户需求的各种数据报表工作  
4、监控数据库，处理解决数据库的日常问题，对数据库和数据表进行调优

**教育背景**

2009.9 – 2013.7      南昌大学科学技术学院          软件工程 统招本科

主修课程：数据库、数据结构、java、c++、高等数学、线性代数等