



# Jinliang

## 个人技能

---

- 掌握Python、SQL编程语言，能够熟练运用HQL，SparkSQL对数据进行查询、统计等操作
- 掌握Spark及其组件SparkCore、SparkSQL、Spark Structured Streaming的使用，能够使用SparkSQL结合Structured Streaming进行离线和实时的数据分析
- 掌握Hadoop相关组件及原理，熟悉HDFS存储、MR计算及Yarn资源调度，了解Zookeeper工作原理，并能借助Zookeeper实现集群高可用，使用Presto进行计算
- 熟练使用Flume日志进行采集，掌握Kafka基本原理，使用Kafka消息中间件存储并消费数据，并与Spark整合使用进行数据的实时处理
- 熟悉ElasticSearch(ES)分布式全文检索及倒排索引原理，并使用ES与Hive、Spark集成进行开发
- 能够使用Kettle、Sqoop进行数据采集，使用FineBI进行可视化，使用DolphinScheduler(DS)和Oozie进行任务和工作流调度
- 了解HBase框架原理，了解Linux系统，掌握常用命令，可以编写常用Shell脚本
- 了解Flink实时处理框架，Pytorch深度学习框架，Python爬虫，Lisp语言
- 英语六级(CET-6)，良好的听说读写能力

## 工作经验

---

大展信息科技（深圳）有限公司

大数据开发工程师

2020.11-2022.09

## 项目经历

---

惠买电商数据平台二期

2022.03-2022.09

**项目架构：** Hive + Spark + Presto + Hadoop + ES + Flume + Kafka + Sqoop + Oozie

**项目描述：** 该项目是惠买电商数据平台二期项目，实现惠买电商平台的用户画像系统。使用Spark计算引擎计算相关标签，基于DS调度 workflow、通过Sqoop和Flume进行离线和实时数据的导入，使用ElasticSearch和MySQL进行数据存储，统一由Presto将实时埋点和日志分析结果和离线标签分析结果提供给应用方。通过标签系统准确把握用户消费习惯，为公司科学运营决策提供数据支持。

**个人职责：**

- 通过Sqoop将离线数据导入到ES并映射到Hive表，并使用Oozie调度
- 使用Flume将Nginx日志和用户埋点日志采集到Kafka
- 通过获取ES中的数据，结合MySQL获取标签规则数据，使用SparkSQL完成规则匹配标签开发，如年龄段，性别，职业等
- 使用SparkSQL完成多个统计类标签开发，如用户消费行为、支付方式的统计预测分析等
- 使用Spark Structured Streaming对用户埋点日志进行ETL处理，并且进行统计分析

### 惠买电商数据平台一期

2021.09-2022.03

**项目架构：** CM + Hive + Presto + HDFS + Sqoop + Oozie + Yarn

**项目描述：** 该项目是惠买电商数据平台一期项目，实现惠买电商平台的离线数仓。使用Hive+Presto计算引擎计算离线指标，使用Oozie调度 workflow，Sqoop进行数据的导入与导出；使用Hue来进行日常操作，使用CM进行部署和运维。有销售主题、商品主题、用户主题、促销主题、客服主题等，准确把握新零售业务的运营状况，为公司科学运营决策提供数据依据。

#### 个人职责：

- 通过Sqoop收集用户、商品、订单、支付等系统的业务数据，将数据导入到Hive集群中
- 使用拉链表解决保存历史数据，基于Oozie进行任务的调度工作
- 负责销售主题和商品主题各层的建模与分析计算工作
- 负责平台的维护和实现调度，处理各种异常场景，包括优化配置、数据倾斜、内存溢出等

### 正迅保险大数据分析系统

2020.11-2021.07

**项目架构：** HDFS + Yarn + Hive + Spark + Sqoop + DS

**项目描述：** 该项目是基于正迅保险的意外终身寿险重构的大数据分析平台。基于CM平台实现精算系统相关的计算工作；利用Sqoop从OLTP将保险数据导入到HDFS上；使用Hive进行元数据管理，并且底层使用SparkSQL进行计算，将计算结果通过SparkSQL导出到MySQL中给BI和投保界面使用；整个流程通过DS进行调度工作。通过大数据平台为保费、保险准备金和现金价值等的计算提供更加高效的解决方案，为公司业务经营提供数据支持。

#### 个人职责：

- 负责使用Sqoop将数据进行采集，整合到Hive上
- 使用SparkSQL结合自定义UDF函数完成保险因子和后续保费的迭代计算
- 负责行业指标分析，并且将结果导出到MySQL，使用DS完成任务调度工作
- 对Spark程序进行优化工作，如shuffle调优和缓存持久化等

## 教育经历

广州大学	硕士	2018.09-2021.05
广州大学	本科	2014.09-2018.06

## 个人评价

- 善于思考，逻辑清晰，有较强的逻辑思维能力及良好的学习工作能力
- 对于新事物有很强的好奇心，勇于接受挑战，积极进取，对于事务发展有较强嗅觉
- 责任感强，踏实肯钻，积极配合上司的工作