

王赫 - 大数据工程师简历

个人信息

- 性别: 男
- 手机: 13263228979
- 学校: 东北石油大学
- 学历: 本科
- 岗位: 大数据开发工程师
- 邮箱: dqwanghe00@163.com
- 专业: 化学工程与工艺
- 经验: 6年

工作经历

- | | | |
|----------------|-----------------|--------------|
| • 复星集团 | 2021.02~至今 | 智能科技共享中心 |
| • 软通动力 | 2019.09~2021.01 | 平安壹账通智慧风控项目组 |
| • 北京优购文化发展有限公司 | 2016.03~2019.07 | 科技赋能-大数据组 |
| • 大庆油田工程建设有限公司 | 2010.09~2016.02 | 化建公司第三项目部 |

专业技能

- 熟练使用 Kimball 维度模型理论、OneData对数仓分层、建模, 熟悉E-R建模理论
- 熟练使用 Hive、Spark、Shell 进行数据加工, 掌握自定义UDF和常用SQL优化方法
- 熟练使用 Kylin 进行 Cube 的构建、剪枝优化, 掌握 Kylin REST API 的使用
- 熟悉 Java、Scala 语言, 熟悉 MySQL、PgSQL 的操作
- 熟悉 Hadoop、Sqoop、DataX、Kafka、Flume 的使用和基本原理
- 熟悉阿里云 DataWorks、EMR、Maxcompute 的开发流程
- 了解 Spring boot、Mybatis 等框架的基本应用
- 了解数据湖、数据虚拟化系统等相关概念
- 了解基础数据结构和算法的基本原理
- 了解DAMA数据管理体系的基础知识
- 了解 HBase、ElasticSearch、Presto

项目经历

1. 复星集团 - 集团CDP平台 - 2021.07 - 至今

- 技术架构:
 - 阿里云EMR: Hadoop、Hive、Spark、Ranger等 + OSS + Kafka + MySQL + Dataworks
- 工作内容:
 - 负责CDP数仓数据接入、分层、维度模型建模、ETL开发
 - 负责会员Id-Mapping程序的设计、开发
 - 负责标签自动计算系统大数据平台部分的设计、开发
 - 负责神策订阅程序的设计、开发
- 技术要点:

- 平台依托阿里云EMR集成Hadoop、Hive、Spark等集群组件和OSS存储系统，使用DataWorks作为统一开发接口
- 使用DataX将产业端会员及销售数据接入CDP数仓，并完成数据脱敏
- 使用HQL和Spark SQL完成数据清洗和ETL，利用连通图特性使用Spark GraphX完成会员Id-Mapping的计算
- 使用DataX实现标签管理系统和CDP平台间的数据交互，完成规则数据和元数据的同步
- 使用HQL完成规则数据的整合。编写Spark程序完成标签规则的解析、计算SQL的生成，以及会员标签的计算
- 针对复杂的标签规则，以及从隐私计算平台获取计算结果，编写Hive UDF函数
- 设计Structured Streaming程序对神策Kafka进行过滤，实现数据的实时订阅

2. 复星集团 - BFC AI智能中台 - 2021.02 - 至今

- 技术架构：
 - TiDB + MySQL + DataX + DataX Web + Spring boot
- 工作内容：
 - 0-1建设企业数据仓库
 - 负责中台数仓的架构、分层设计
 - 负责数据的主题域划分、维度模型建模，以及ETL开发
 - 负责部分数据源的数据同步
- 技术要点：
 - 使用DataX作为数据同步工具，定时从多个数据源拉取业务数据
 - 使用TiDB作为数仓的计算和存储平台，DataX Web作为任务调度工具
 - 针对不同数据源，使用SQL + Shell完成对店铺数据的整合，生成统一店铺ID与源系统ID的映射表
 - 使用SQL + DataX脚本完成维度表和事实表的数据清洗和ETL，以及DWS、ADS数据指标的计算
 - 使用Spring boot开发数据接口供其它供应商调用

3. 平安 - 风控一体化智慧分析平台 - 2019.09 - 2021.01

- 技术架构：
 - PostgreSQL + Sqoop + Hadoop + Hive + Spark + Kylin + HBase + Presto
- 工作内容：
 - 负责风控数据湖债券、股票主题域的建模、ETL开发
 - 负责智慧分析数据集群架构，Kylin与CDH、Presto的集成调研
 - 负责与平安科技沟通，实现集群部署，以及与发版、调度系统的集成
 - 负责风控数据湖到智慧分析集群的跨集群数据同步
 - 负责智慧分析Kylin Cube设计，以及构建任务的开发
- 技术要点：
 - 使用Sqoop脚本将PostgreSQL中的数据同步到风控数据湖
 - 使用HQL、Spark SQL脚本完成内外部风控数据的ETL处理
 - 利用Hadoop的distcp工具实现风控数据的跨集群增量同步
 - 通过构建Kylin Cube提高查询效率，同时也可以减少ETL工作量
 - 使用Kylin的REST API开发Shell脚本，完成STG和PRD环境Hive表的加载、删除，以及Model、Cube的创建
 - 使用Kylin的REST API开发Shell脚本，配合Linkdo调度平台，实现Cube的Segment自动构建，及对构建任务的监控
 - 使用Presto作为Kylin的下压查询引擎，应对未创建Cube的查询需求

4. 优购 - 数据分析平台 - 2016.03 - 2019.07

- 技术架构：
 - Flume + Kafka + Sqoop + Hadoop + Hive + Spark + MySQL + Azkaban + Presto
- 工作内容：
 - 参与Hadoop集群搭建及其他组件的部署
 - 负责业务数据采集
 - 负责主题域的数仓分层、建模，以及UDF编写和ETL开发工作
 - 负责改将HQL任务改造为Spark SQL任务，提高执行效率
- 技术要点：
 - 配置Flume、Kafka采集日志数据到HDFS
 - 使用Sqoop脚本实现业务数据到Hive的同步，以及ADS层数据到MySQL的同步
 - 使用HQL脚本及UDF函数完成ODS层数据的解析、清洗、脱敏等工作
 - Hive表采用ORC格式存储，减小数据体积的同时也可提高访问速度
 - 使用HQL脚本完成数仓各层的ETL工作，使用Azkaban配置ETL操作的依赖关系，进行任务调度
 - 使用Spark SQL对Hive任务进行改造，将每日的跑批任务改为Spark程序
 - 部署Presto访问Hive数据作为即席查询引擎

资格证书

- 计算机中级专业技术资格 - 数据库系统工程师
- 工信部专项技术证书 - 大数据分析师（高级）

其他信息

- 兴趣爱好
 - 音乐（ACG、古典），洞箫
 - 道家哲学，天体物理
- 自我评价
 - 为人诚恳，认真负责，敢于面对困难和挑战
 - 工作积极，对新技术有浓厚的兴趣
 - 具有良好的分析能力、自学能力、适应能力
 - 具有良好的沟通协调能力以及团队精神