

## 联系方式

---

- 手机：17826800434
- Email：chenzebedee@gmail.com
- QQ/微信号：331471867

## 个人信息

---

- 陈绍迪/男/1995
- 本科/杭州电子科技大学
- 工作年限：1年
- Github：<https://github.com/ChenZebedee>
- 期望职位：大数据工程师 ETL 工程师
- 期望城市：杭州

## 工作经历

---

### 杭州玛瑙湾科技有限公司（2018年6月 ~ 至今）

#### 工作职责

1. 大数据环境搭建，维护，优化
2. 离线数据同步
3. 数据 etl 清理
4. shell 脚本编写
5. 实时同步数据处理,并输出到指定表，用于国家数据上报
6. MapReduce 开发

### 北京思特奇（2017年9月 ~ 2018年5月）

#### 工作职责

1. Hive开发，提供局方所需数据。
2. Mapreduce开发，对数据进行处理，并提供给局方
3. Shell脚本编写
4. Python脚本编写
5. Sqoop脚本编写
6. Hbase的MapReduce开发

## 项目经历

---

### 风控评分卡模型

#### 项目描述

公司为互联网金融公司，涉及小贷业务，所以风控方需要机申模型，来削减人力成本，由于目标数据需求为每个订单号对应一条数据，而原始数据中每个订单号对应着一条或多条的数据，而多条数据中还要把其中某一列转为列名，与其他列组合形成一条新的数据。最终要输出 800+ 的字段。

例：

id	column1	column2	column3
1	a	dd	ff
2	a	cc	ee
3	b	dd	jj

变成

column_1	dd_column3	cc_column3
a	ff	ee
b	jj	

技术应用：hadoop + hive + hbase + spark

1. 建模维度宽表字段整理
2. MapReduce/Spark 数据处理程序
3. HBase rowKey 设置
4. HBase 批量获取数据小程序

## canal 实时同步

### 项目描述

1. 出于安全考虑，所有数据开发均需要进行数据脱敏加密，而原先通过 shell 实现的 T+1 方案对线上 MySQL 压力较大，所以新增 canal 实时同步方案
2. 由于国家网信办要求，提供 P2P 公司的相关业务数据，需实时推送。通过 canal 实现 binlog 的数据同步，再用 spring for apache kafka 将源数据转为目标数据，之后入库

技术应用：spring boot + kafka + canal + mysql

1. canal 源码添加日志
2. kafka 搭建
3. spring boot 开发消费端程序
4. 数据脱敏功能实现
5. 上报数据消费端开发，重点在于格式转换

## 大数据运维

### 项目描述

1. 由于千万级的表进行多表关联时不能用 MySQL 很好的解决，于是就有了大数据环境搭建的工作
2. 为了缓解线上服务器压力，将月报日报等定时任务，将用大数据组件实现

3. 大数据组件的搭建、维护和优化
4. 出于安全与方便部署的考虑，采用 docker 容器技术实现集群搭建

技术应用：hadoop + hive + hbase + zookeeper + spark + sqoop + docker

1. 硬件规划
2. 搭建大数据环境
3. 性能压测
4. 性能优化，通过参数控制，以及 hadoop 自带的资源调度规则
5. hive on tez, hive on spark
6. 运营周报日报等定时任务 shell 编写

## 技能清单

---

以下均为我熟练使用的技能

- 数据库相关：MySQL
- 版本管理、文档和自动化部署工具：Git
- 编程语言：java/shell/scala
- 大数据相关：hadoop/hive/hbase/spark/kafka/zookeeper
- 容器相关：docker

## 自我介绍

---

我是一个比较喜欢折腾的人，自己在家喜欢玩玩各种东西，比如说星际矿难时，买了一台二手星际魔盒装黑群晖，用于网盘同步，自己电脑也经常在 windows 和 linux 系统之间重复重装，我也乐于助人，如果同事有任何问题，我都会积极帮助，尤其是技术上的问题，但是我更喜欢的是讨论形式的，大家来个头脑风暴，我的不足可以从中得到进步。

## 致谢

---

感谢您花时间阅读我的简历，期待能有机会和您共事。