

许友昌

18269009535 18269009535@163.com 杭州市拱墅区萍水街

教育背景

2015.9 ~ 2018.6 广□大学 软件工程 研究□
2011.9 ~ 2015.5 长春大学 信息与计算科学 本科

个人技能

熟练使用 Hadoop、Hive、Hbase、spark、Elasticsearch、Prometheus、Logstash、kafka、Grafana等组件，研读过 kafka、spark、Beats、prometheus部分源码，进行过相关的二次开发。掌握 java/python/shell 语言的使用。

工作经历

2017.12.05 ~ 2019.07.26 杭州大树网络技术有限公司 大数据研发工程师
参与大数据平台研发，重点负责基础数据平台板块以及ES相关业务，包括海量数据实时ETL、定制化组件开发、集群搭建与调优、日志数据与metric数据分析报表等。

2015.09 ~ 2017.08 广西大学大数据与云计算实验室 研究助理
参与广西Go平台的建设，主要负责虚拟化平台、大数据平台、云存储平台的构建与web系统的开发以及学术论文的撰写。

项目经历

2017.12.05 ~ 2019.07.26 数据仓库2.0 杭州大树网络技术有限公司

项目描述：实时同步线上数据，入库/供给到Hive数仓、TiDB、pipelineDb、ES等，以供分析挖掘、线上（历史/跨库）查询、实时计算等。

主要研发内容：

- 流式入库，线上数据实时抓取，不同存储流式入库
- 支持修改删除（Hive），源库数据修改不会造成 Hive 仓库记录重复
- 数据检查/修复，以线上数据库事务日志为基准，自动检查和修复
- 自适应调整，自动跟随源库 Schema 调整，自动感知新表

其他重要增强：Hive 事务、Debezium 定制、碎片整理、Hive 切换 Spark 引擎、Spark 支持 ORC 事务表、安全、备份、故障隔离、监控/自动化、数据质量报告 等

本人角色：参与者，负责了 kafka-connect-es 开发、ES 集群优化

2017.12.05 ~ 2019.07.26 基础数据平台 杭州大树网络技术有限公司

项目描述：日志、监控、链路追踪数据的收集、融合，以及相关的分析、挖掘和报表服务。

主要研发or选型内容：

- 采集点，埋点技术&管理、数据融合（如 traceId 嵌入日志）、采集点区分
- 传输网络，统一传输日志、监控等数据，智能缓冲和路由，灵活的数据加工
- 总线&入库，各类存储（ES、timescaleDb、HDFS）入库，支持实时处理
- 服务&应用，统一日志平台、监控平台，日志分析、告警，等
- 监控&配置，各传输&加工节点的集中配置，整套体系的监控&运维，等

本人角色：主要开发者，完成了平台构建、数据传输方案、定制化组件开发、数据格式规范、对外服务接入等

2015.09 ~ 2017.08 广西文化旅游（自然基金）广西大学、中科院软件所

项目描述：构建弹性扩展的计算集群，提供资源可控的大数据平台部署方案。

主要研发内容：

- 服务器资源管理，基于zsatck构建虚拟化管理平台，对服务器资源进行统一管理、弹性分配
- 存储服务，基于ceph构建统一式云存储服务，为不同存储类型提供统一入口
- 自动化部署，基于Ambari构建大数据自动化部署平台，结合虚拟化平台实现配置化弹性扩展集群

本人角色：主要负责人，完成方案设计、平台搭建、任务分配等

发表论文：*KPI Data Anomaly Detection Strategy For Intelligent Operation and Maintenance Under Cloud Environment*