

杨华磊 - Java&大数据开发工程师

杨华磊 | 统招本科 | 山东师范大学 | 亦庄 | 17600505796(微信同号) | yyqq188@foxmail.com

个人简述

10年开发经验,能针对业务找准技术要点并积极解决技术难点。善于找到需求中的通用技术点,并封装为工具。善于研究或预研技术上的问题,并积极分享,根据业务要求设计技术架构,并开发框架,分拆任务模块。

个人网站 :<https://yanghl.top>

博客: <https://blog.csdn.net/yyqq188>

github: <https://github.com/yyqq188>

工作经历

- 2020.3-至今 中电金信 高级设计开发工程师
- 2016.11-2020.3 亚信科技CTC数据应用开发部 数据技术经理
- 2015.2-2016.11 北京数字新思科技有限公司 大数据工程师

工具&项目

基于新华实时数仓,以sparksql的方式写流式作业的脚手架

sparkstreaming hbase java

github地址: https://github.com/yyqq188/spark_sql_helper

工具描述:

此工具解决了之前实时业务指标开发中,**开发效率低,维护难度大的痛点**,原来项目基于java 或 scala来开发实时业务指标,开发周期长,代码冗余且维护难度越来越大,该工具以sql的形式开发实时指标,并以json文件的形式来部署实时任务的实时作业脚手架工具. 如果要开发新的算子,利用脚手架工具,只需要实现脚手架的transform接口并实现其中的方法,就可以在json 文件中指定并调用.日常的实时任务开发就转为sql开发,进而可以实现流批一体开发.该脚手架工具利用阿里开源的datax库,并参考了apache的seatunnel项目的实现思路.

通用FlinkSQL数据开发平台

flink springboot hibernate vue codemirrorjs

工具描述:

结合实际的工作场景, 开发此基于flink的 sql开发平台, 目标用户为业务人员。可以通过写flinksql一站式开发实时程序并能监控job的执行情况。产品以后台管理平台的形式开发, 包括配置管理, 新建任务, 任务列表, 日志管理, 系统配置管理, 告警配置, 用户配置五个模块的功能。前后端分离架构, 前端基于vue框架开发,后端基于springboot,结合flink rest api接口开发任务和集群监控的后台逻辑, 通过codemirrorjs收集前端sql语句并分离出DDL DML语句, 配合前端的任务设置, 生成job执行命令。该项目是采用前后端分离的方式开发的, 由于技术实现上有前端vue后端springboot数据上flink的技术, 技术综合性高, 难点在数据层如何将sql以flink的job的形式提交到yarn上来执行。

页面展示:

配置管理

日志管理

系统管理

报警管理

用户管理

新建任务

* 任务名称

任务名称

* 运行模式

运行模式

* flink运行配置

Flink运行配置

* Checkpoint信息

Checkpoint信息

* 三方jar地址(自定义udf, 连接器等jar地址, 多个用换行)

SQL

点击上传

只能上传sql文件(以.sql结尾)

配置管理

日志管理

运行日志

系统管理

报警管理

用户管理

运行日志

请输入内容

Q

#	jid	name	state	startTime	executionMode	restartStrategy	jobParallelism	endTime	duration	lastModification	tasksTotal	tasksCreated	tasksScheduled	tasksDeploying	tasksRunning	tasksFinished	tasksCanceling	tasksCancelled
1	d74d4d2374f6a2af6903339e9556a0f0	Socket Window WordCount	FAILED	1611541516861	PIPELINE	Cluster level default restart strategy	1	1611541816962	300101	1611541816962	2	0	0	0	0	0	0	1
2	4a3b8bbd979f44b1136e7b618b774fe8	Socket Window WordCount	RUNNING	1611540270291	PIPELINE	Cluster level default restart strategy	1	-1	23374851	1611540271277	2	0	0	0	2	0	0	0

共 2 条 5条/页 < 1 > 前往 1 页

系统管理

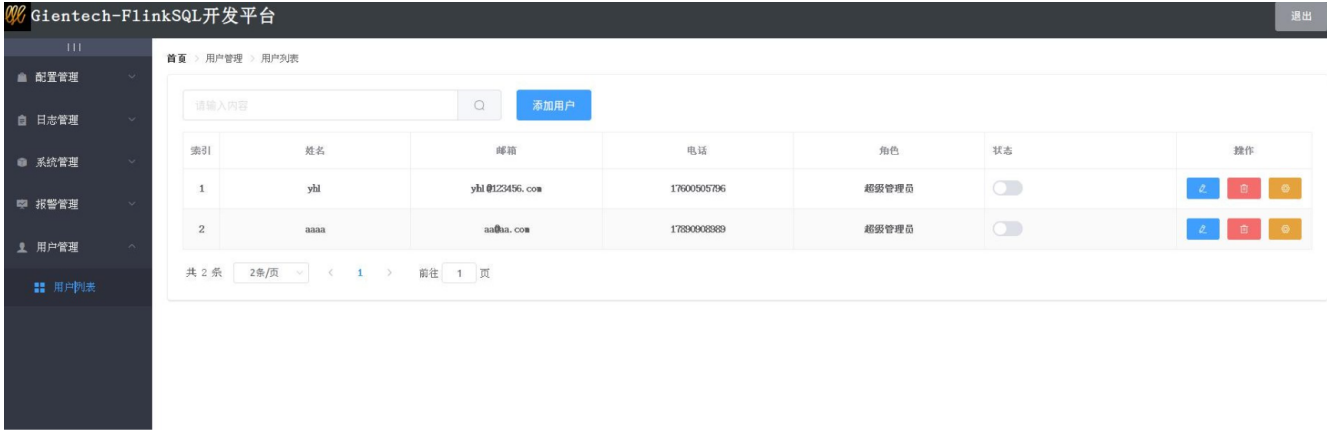
系统设置

请输入内容

Q

添加系统设置

#	名称	键值	描述	操作
2	flink_home	/opt/module/flink-1.12.1/	flink客户端目录(必填)	添加删除



基于flink的同步Gaussdb和DWS数据工具开发

flink kafka dws

工作描述:

需求沟通,需求分析,拆解技术点并具体开发实施

项目描述:

该工具实现消费DSG和OGG格式的kafka数据,并可分别同步到Gaussdb和DWS数据库中.微批同步,可以调节微批的时间跨度以及微批的数据量大小,提高同步效率.采用配置文件的方式,增加同步的新表,配置新表的字段信息和主键信息,无主键也可以同步,无需开发同步逻辑.

实时数据即席查询平台 2020

flink clickhouse python

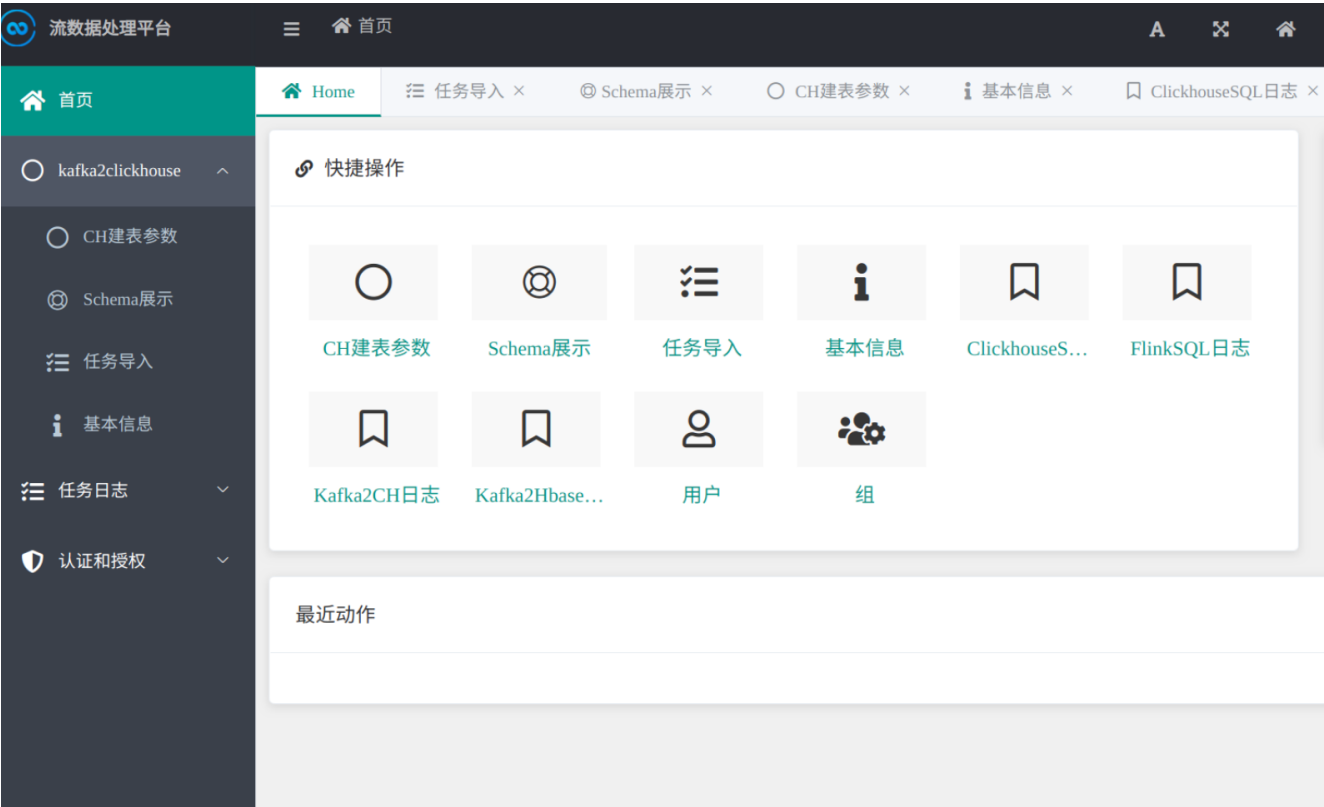
github地址:<https://github.com/yyqq188/IStreamBigDataProcessingPlatform>

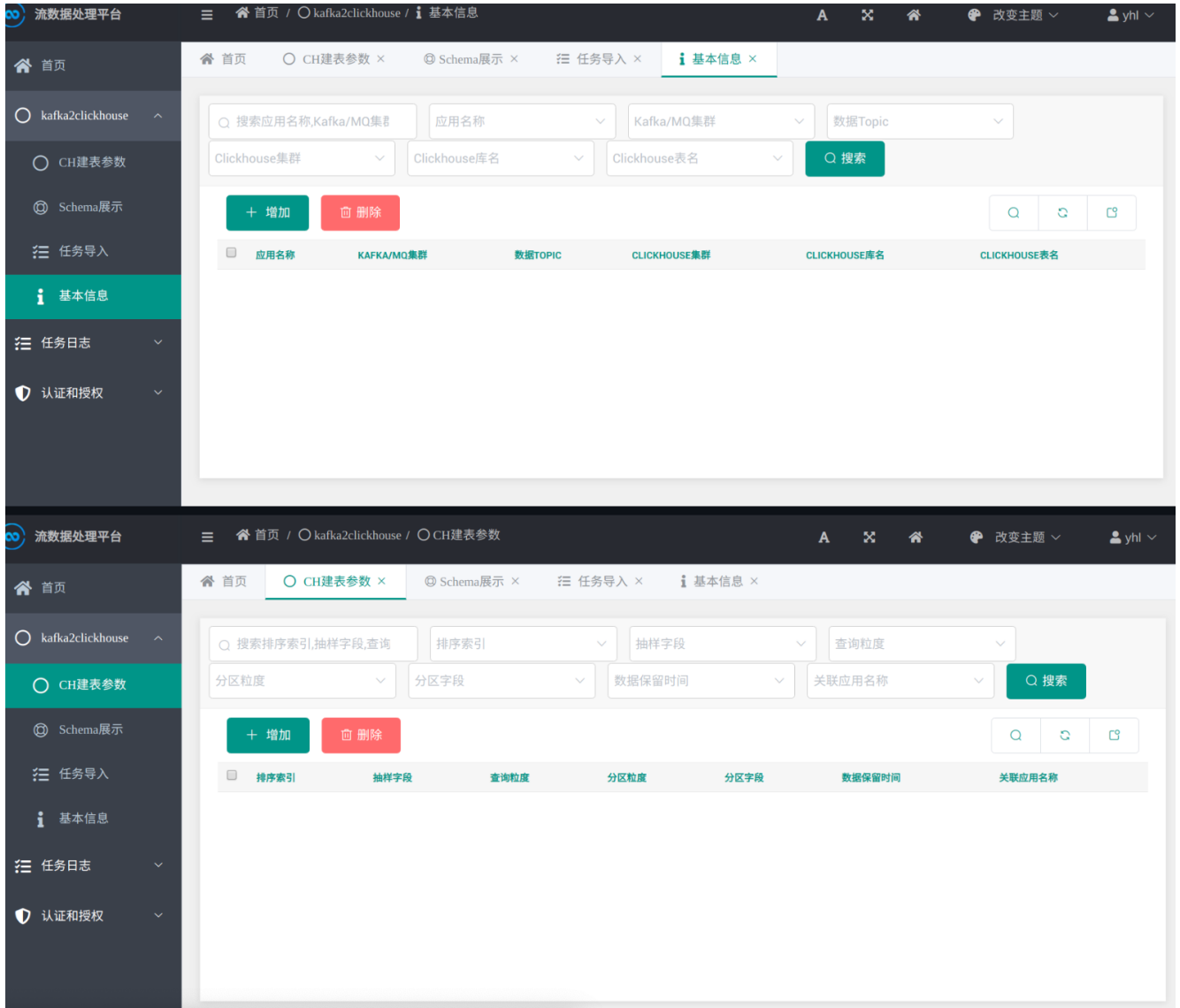
开发背景:

为了满足电信总部对实时数据处理的需求,在完成中国电信全国5G开户数实时统计的基础上,开发实时数据处理平台,

工具描述:

平台主要分三大模块，实时数据导入（hbase/clickhouse），实时计算(flinkSQL)，即席查询(clickhouseSQL)。业务人员只需要在页面上按需配置导入规则，生成的参数信息会存储在mysql 并被实时检测出生成导入程序并检测启动。实时计算和即席查询利用SQL语句描述业务逻辑。该产品可减少开发人员日常工作量，并能更快的响应业务需求。





政企爬虫工具 2018.10-2019.5

docker python scrapy

github地址:https://github.com/yyqq188/spider_tool_common

dockerhub地址:yyqq188/spider_tool_common

开发背景:

浙江电信政企需要爬取国家，省，市，县各级的公共采购网或招投标网上的招投标信息，由于需要采集的网站数量很多，为了简化爬虫程序的开发工作量，参考开源爬虫框架scrapy开发该工具

工具描述:

把招投标相关的通用逻辑加入到代码生成中， 分别开发通用版和招投标版的爬虫工具，并封装成pip包，并开发人员只需要下载对应工具的docker镜像， 并写好规范的json配置文件，就可以生成初始化好的爬虫程序，可以即可启动爬取或在此基础上进行修改后再爬取。**降低爬虫代码开发工作量,并将开发和部署统一起来,简化维护步骤**

dockerhub页面

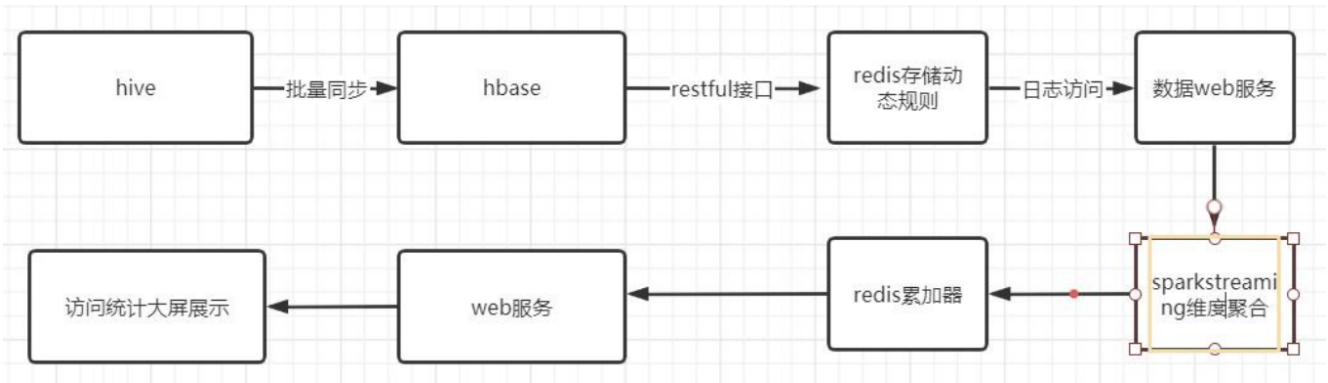
yyqq188 / spider_tool_common Updated 2 years ago		Not Scanned	☆ 0	↓ 34	Public
TAG	OS	PULLED	PUSHED		
v2		3 months ago	2 years ago		
v1		6 months ago	2 years ago		
latest		3 months ago	2 years ago		

中国电信营销服务数据调用实时展示平台 2017

sparkstreaming kafka flume hbase hive redis

项目描述:

后台利用hive从中间表跑出结果宽表并批量打入hbase，利用redis做动态规则和用户权限设置，开发restful接口服务，设计restful接口规范。配置flume拦截器，将接口服务日志利用flume分别导入到kafka和hdfs。开发sparkstreaming程序根据业务指标分不同维度实时统计指标并封装为json存入redis，供前端实时调取展示.流程如下



电商受众监控平台

java kafka hbase spark hive

项目描述

该监控平台对目前主流的电商平台（京东，一号店，国美，苏宁等）在客户公司现有的销量，市场份额占比，地域分布，商品库等各种指标进行实时以及离线监控，使用hdfs，hive作为底层存储，使用分布式的kafka消息中间件对实时数据进行动态存储，使用sparkstreaming，spark进行各个维度指标实时统计和批处理统计。

主要工作是对当时利用路由采集的log日志，进行维度分析统计，供BI人员进行分析。

职责描述

数据处理代码编写

与产品经理配合，梳理统计口径

工作奖惩

2018-2019 亚信科技股份有限公司**优秀员工**

外派杭州期间,在浙江电信政企项目中及时解决未预估到的技术问题,从而保证项目进度正常完成.