

可视化组件Flink任务开发 (Volt)

大纲

- 背景
- 平台目标
- 技术架构
- 平台总结
- 未来工作
- 结束

背景

- 开发实时任务门槛高
- 持续集成迭代效率低
- 运维成本高有单点

平台目标

- 开发门槛低
- 持续集成迭代快
- 运维友好

围绕目标，Volt平台集成了什么功能？

时间线

2018年7月
实时项目启动

2018年8月
图模式任务上线

2018年11月
8业务30+任务

2018年12月
协同开发、Flink独立jar



图模式翻译框架
实时组件库
插件SDK
Flink On Yarn
Flink state机制

实时开发web UI
Kafka升级0.11
图模式任务运行
指标展示Grafana
运行历史

保活监控报警
任务视图修改并发度
接入广告、推荐、交易
接入中间件、分销、智能营销
推荐Flink独立集群下线
广告推荐storm任务下线
风控Flink集群下线

多人协同权限管控
多集群提交
SQL组件支持sql语法
实时流表schema可视化
支持Flink独立Jar提交
接入风控实时任务

时间线

2019年2月
移动办公

2019年3月
新增业务组件
50+线上任务

2019年7月
Flink多版本

2019年8月
纯SQL模式
调试改进

移动端适配
首页类目导航
实验一键复制
实验模版创建

HDFS写组件
去重过滤组件
Hades读组件
Kafka多集群

解耦Flink版本
支持event time
支持state TTL
Flink配置透传

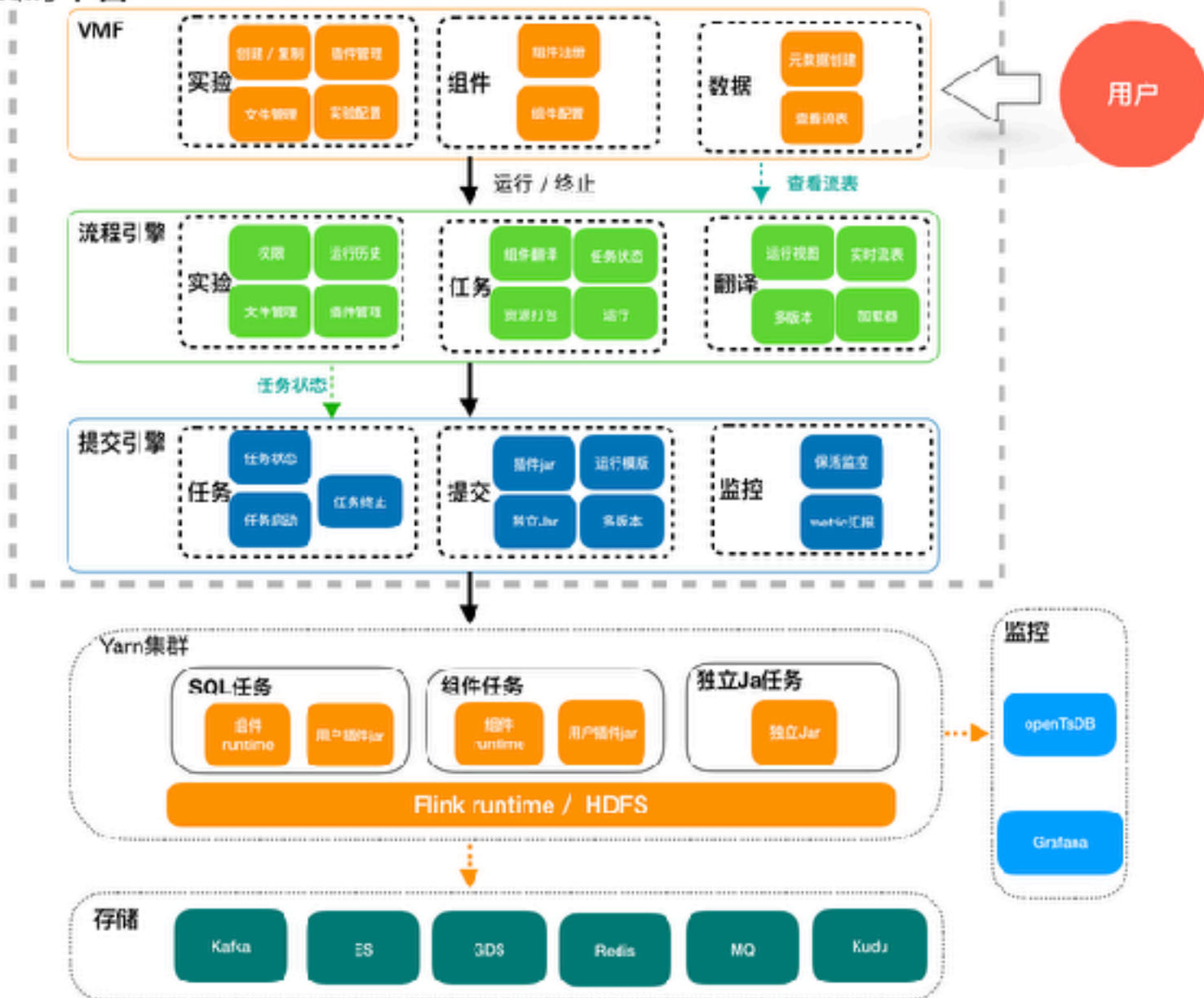
纯SQL开发
子图运行

技术实现

- 平台架构
- Flink任务的组件化构建

平台架构

实时平台

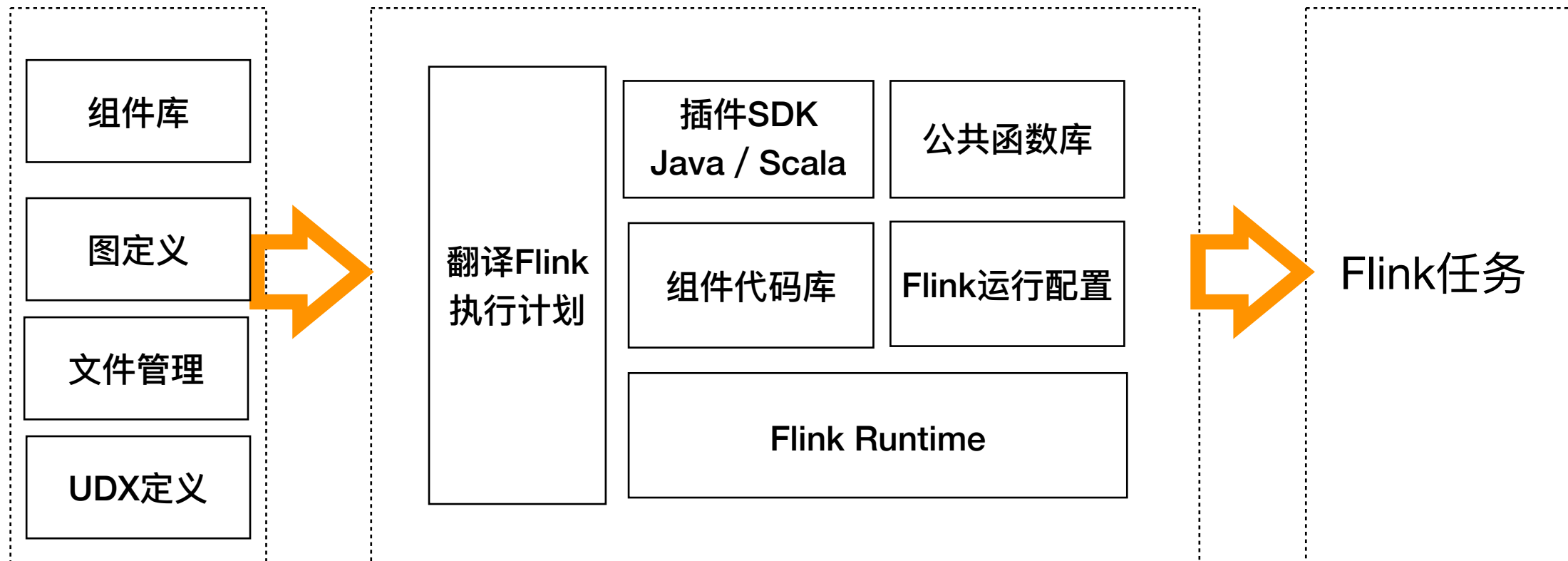


Flink任务的组件化构建

Volt: 开发

Vinci: Flink任务翻译

Yarn: 运行



如何低成本做到持续集成高效率？

开发介绍

- 实时场景
- 首页导航
- 开发三大模式
- 指标监控

首页导航

- 日常环境：<http://volt.daily.vdian.net/>



全局模版

← → ↺ 不安全 | volt.daily.vdian.net



jlyiping

我创建的项目

测试

实时测试任务

二级分组

离线算法任务

我管理的项目

私有模板

全局模板

实时任务流程(506)

UDTF例子

jlyiping

2018-12-12 11:01:42

...

实时任务流程(636)

日常kafka读取和打印的例子

jlyiping

2019-01-10 16:44:18

...

实时任务流程(655)

纯sqi例子

jlyiping

2019-08-06 14:50:18

...

实时任务流程(659)

独立Jar例子

jlyiping

2019-08-06 16:13:23

...

组件和任务开发界面

浏览器地址: vdt.daily.wdian.net/project?id=505

UDTF例子(jiyping)

100% 保存 复制

组件

数据源组件

- UT/H5日志
- kafkaJSON
- kafka自定义
- was数据源
- wdianm数据源

特征交换组件

- sql组件
- filter
- select
- UnionAll
- customUDF
- OverUDAG
- TurnUDAG
- SlideUDAG

任务流图

```
graph TD; A[UT/H5日志1921] --> B[select1920]; B --> C[filter1918]; C --> D[customUDTF1917]; D --> E[printSink1918];
```

帮助文档

过滤规则

```
buyer_id.isNotNull &&  
buyer_id.like("%8%")
```

查看schema

保存 运行历史 节点执行 任务视图 配置

开发模式

- 图模式（组件任务）
- 独立Jar
- 纯SQL

图模式

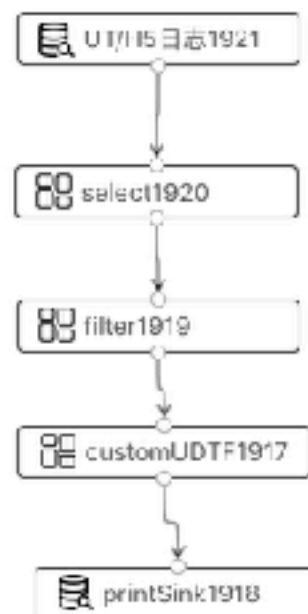
- 任务链接: <http://volt.daily.vdian.net/project?id=505>
- LogParser代码示例: https://gitlab.vdian.net/recommend/vinci-plugin-sdk/blob/yiping_test/src/main/java/com/vdian/vinciplugin/example/OneFieldLogParser.java
- UDTF代码示例: https://gitlab.vdian.net/recommend/vinci-plugin-sdk/blob/yiping_test/src/main/java/com/vdian/vinciplugin/example/TwoWords.java

UDTF例子(jiyiping)

[帮助文档](#)

组件配置

查看schema



运行配置

← → ↺ ⚠ 不安全 | volt.daily.vdian.net/project?id=505 ☆ 🔥 24 📄 📉 JD 👤 ⋮

< UDTF例子(jtyiping) - 100% 保存 复制 >

组件 数据源组件库

管理 UT/H5日志

元数据 kafkaJSON

kafka自定义

vss数据源

vdianmq数据源

特征实践组件库

sql组件

filter

select

UnionAll

customUDF

实时任务配置

运维配置 ☒ 保活监控 ☐ 自动重启

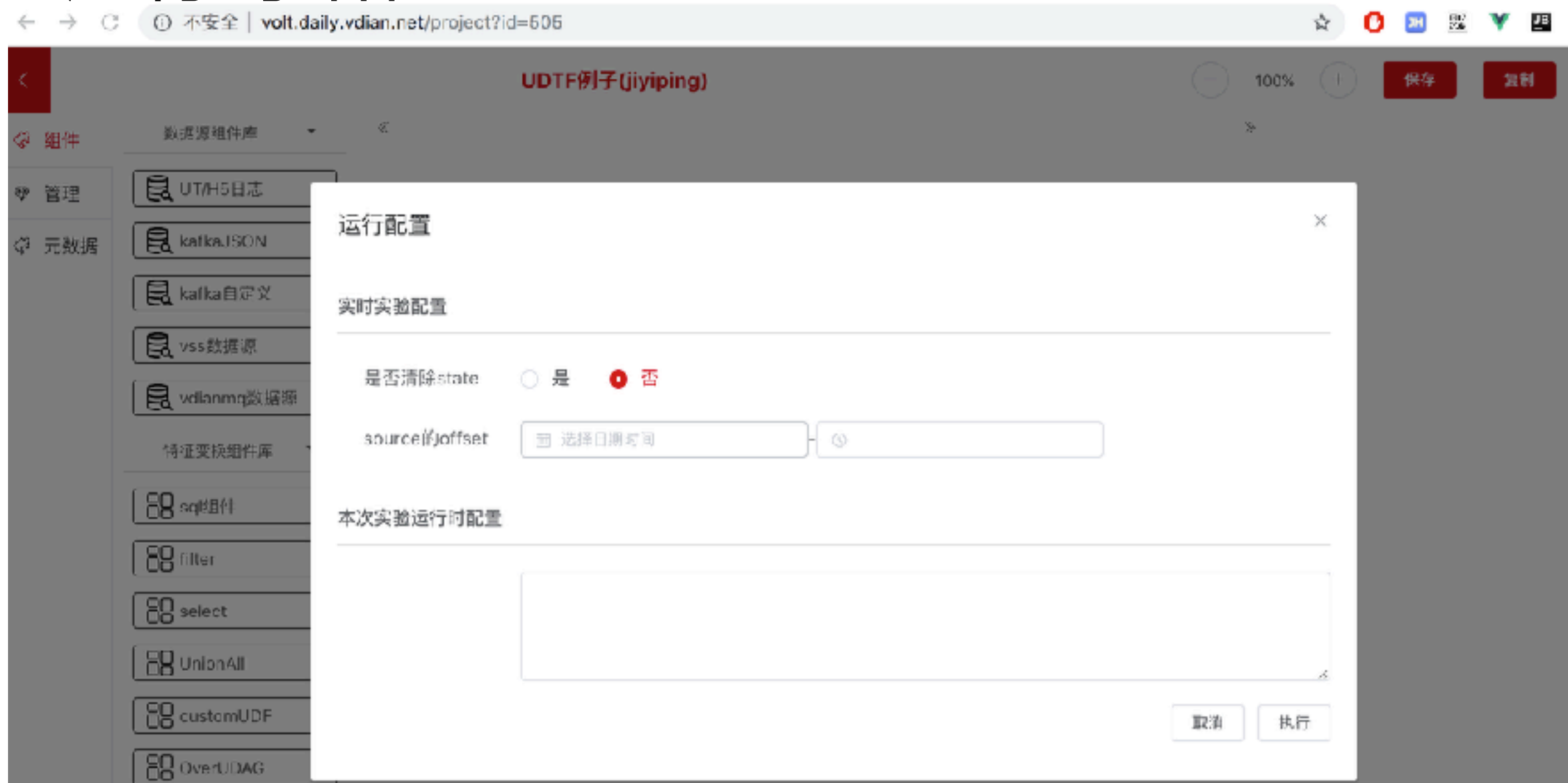
管理员 王启明 张智

常用参数

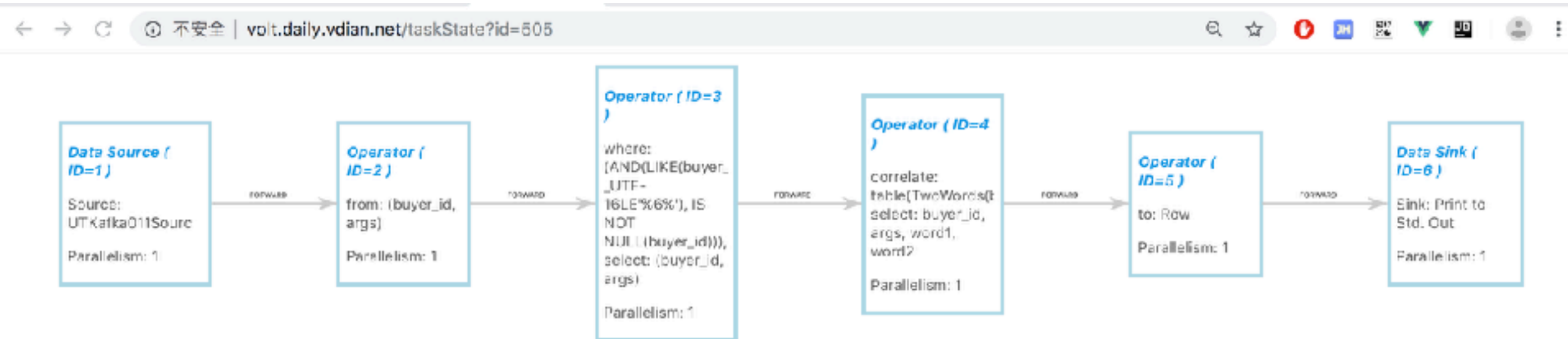
```
needCheckpoint=false
defaultParallelism=2
flink.taskmanager.memory=8192
flinkHome=flink-1.5.1_yoga
user.queue=vinci
mainJarName=vinci-runtime-3.1.1-SNAPSHOT-shaded.jar
stateTtlHour=1
esClusterName=bigdata_es_public
```

取消 确认

运行对话框



任务视图



查看Schema

→ 不安全 | volt.daily.vdian.net/project?id=505

UDTF例子(jiyiping)

100% 保存 复制

查看schema

输入表名 filter1919

字段名称	字段类型
buyer_id	String
args	String

输出表名 customUDTF1917

字段名称	字段类型
buyer_id	String
args	String
word1	String
word2	String

组件配置

查看schema

运行历史

← → ↺ 不安全 volt.daily.vdian.net/project?id=606 ☆ 100% 保存 复制

UDTF例子(jiyiping)

组件 数据源组件库

- 管理 UT/H5日志
- 元数据 kafkaJSON
- kafka自定义
- vss数据源
- vdianmq数据

特征变换组件库

- sql组件
- filter
- select
- UnionAll
- customLIDE

运行历史

实例id	执行状态	提交者	开始时间	结束时间	yarn状态	操作
5170	执行失败	jiyiping	2019-08-06 15:10:04	2019-08-06 15:10:07	无	提交日志
5168	任务取消	jiyiping	2019-08-06 14:54:50	2019-08-06 15:09:59	FINISHED	提交日志 Dashboard Yarn地址
5146	任务取消	jiyiping	2019-08-05 16:30:18	2019-08-05 16:47:43	FINISHED	提交日志 Dashboard

独立Jar

- 例子： <http://volt.vdian.net/project?id=1015>
- 优点： 已有Flink任务Jar， 迁移成本低。
- 缺点： 实现逻辑黑箱不可见。任务视图/查看schema不可用。



lodi广告特征(张津玮)



100%



保存

复制



组件

管理

元数据

customUDTF

redisQuery

queryGds

去重过滤

queryHades

数据输出组件库

写kafka

emptySink

printSink

GDSSink

esSink

redisSink

hdfsAppendSink

拓展组件库

Flink-jar组件

«

»

Flink-jar组件10162

帮助文档

主jar名

Varian-1.0-20190729.060312-116-jar.wi

入口类名

com.weidian.vc.varian.StreamingCenter

flink启动参数

lodi_feature

终止

运行历史

任务视图

配置

纯SQL

- 例子: <http://volt.daily.vdian.net/project?id=655>
- 优点: 适合关系型数据计算, SQL语法简洁易学
- 缺点: 非结构化计算和机器学习算法难写SQL。性能调优依赖Flink SQL内部调优



纯sql例子(jiyiping)



100%



保存

复制



组件

数据源组件库

管理

特征变换组件库

元数据

数据输出组件库

拓展组件库

Flink-jar

Flink-sql

帮助文档

编辑

Flink-sql 2582
(纯sql测试例子)

运行

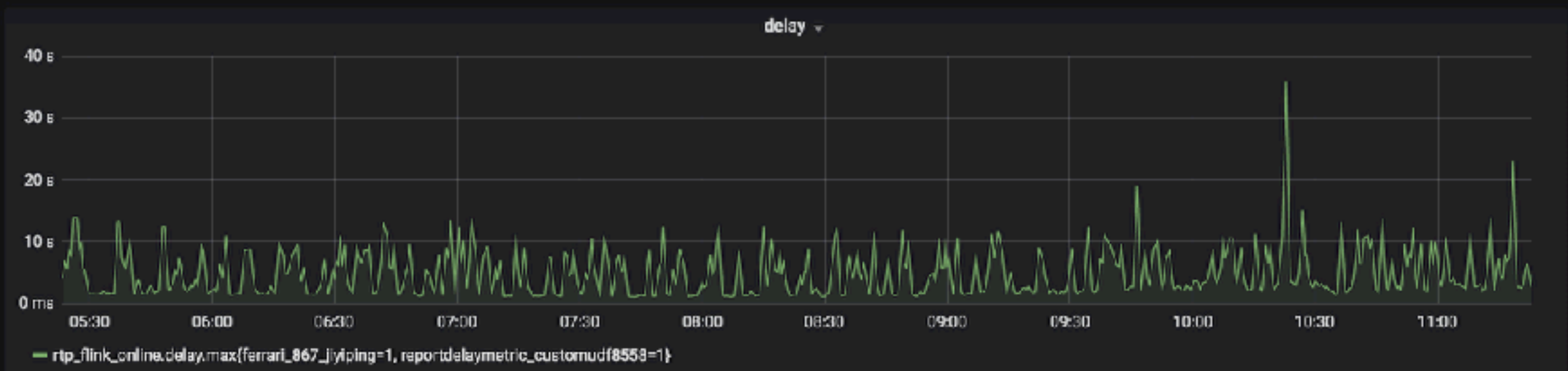
运行历史

任务视图



配置

指标监控

- 示例: http://bigdata-grafana.vdian.net/d/BaXpM_7Zk/rtp_flink_online_867-cai-ni-xi-huan-tian-bao-biao?panelId=4&fullscreen&orgId=3
- volt平台的metric前缀: rtp_flink_online



Graph General **Metrics** Axes Legend Display Alert Time range ✕

 Data Source OpenTSDB ▾  Query Inspector

⊞	A	Metric	rtp_flink_online.delay.max	Aggregator	max ▾	Alias:	🔒	series alias	≡	👁	🗑
		Down sample	interval ⓘ	Aggregator	avg ▾	Fill	none ▾	Disable downsampling <input type="checkbox"/>			
→		Filters	🔒 ⓘ	ferrari_867_jiyiping = literal_or(1), groupBy = true ✂ ✕		reportdelaymetric_customudf8553 = literal_or(1), groupBy = true ✂ ✕		+			
		Tags	🔒 ⓘ	+							
		Rate	<input type="checkbox"/>								

↑ job 名字

↑ 节点名字

后续的规划有哪些？

现有实时场景

- 广告：实时报表、作弊用户实时识别
- 推荐：用户实时特征
- 搜索：实时query质量、实时曝光点击率
- 智能营销：上新push实时策略
- 中间件：trace日志实时聚合
- 风控：实时风控
- 交易：交易链路实时计算

平台集成的功能

- 组件库、插件机制 -> 提取公共功能，业务逻辑自定义
- 界面运行/终止、运行历史 -> 运维去单点，历史追溯
- 任务视图、运行配置 -> 性能调优、配置透传
- 移动办公、监控、自动重启 -> 减少人工，随时响应
- 公共模版、一键复制、管理员 -> 权限控制，协同开发

用户需求是不停的

不变的目标

- 降开发成本、提高迭代效率、运维友好

几大方向

- 纯SQL开发的便捷
- 组件的自定义和快速创建
- 实时计算结合机器学习

FAQ

- 企业微信群：搜索“Volt实时平台开发答疑群”
- 插件工程：https://gitlab.vdian.net/recommend/vinci-plugin-sdk/tree/yiping_test
- Grafana地址：<http://bigdata-grafana.vdian.net/?orgId=3>