

# 基于Flink的流计算平台

梁永锋 (天姥)

# 目录

01

流计算开发运维痛点

02

基于Flink的流计算平台

01

# 流计算开发运维痛点

开发

调优

运维

01

# 流计算开发运维痛点

## 任务需要底层API开发

- 环境配置复杂
- 理解引擎接口逻辑
- Java、Scala等偏底层语言
- 拷包运行任务，一致性



01

# 流计算开发运维痛点

## 任务逻辑调试

- 造数据
- UT、IT
- 远程debug
- 防止生产环境污染
- 结果数据对比



## 上下游数据预览

- 多种存储客户端
- 可视化方式各有不同
- 如何关联排查问题
- 数据安全



01

# 流计算开发运维痛点

## 任务指标曲线

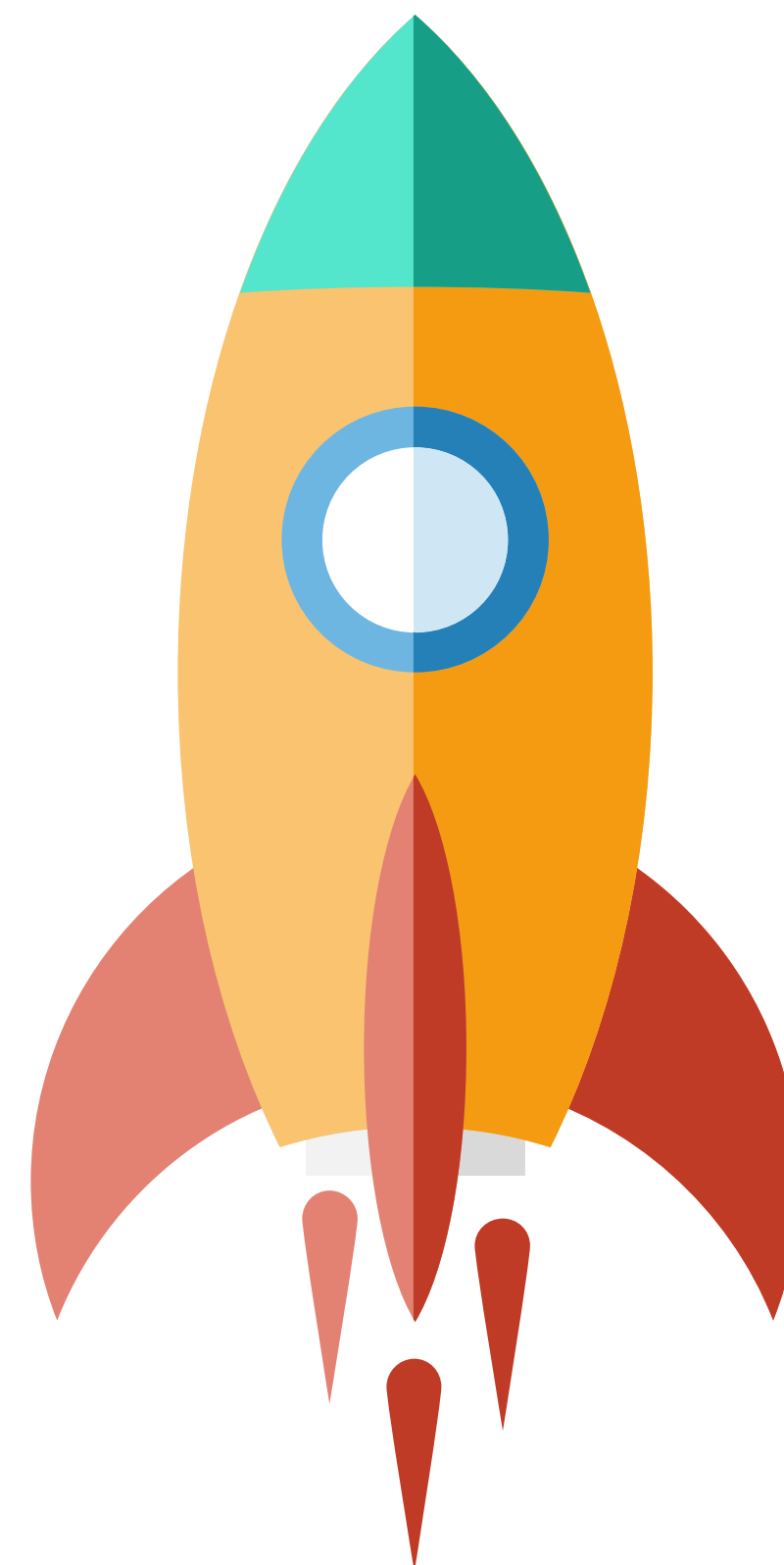
- 任务大盘
- 物理参数
- 逻辑指标
- 任务健康度
- failover
- checkpoint

01

# 流计算开发运维痛点

## 性能调优

- 上下游批量读写
- 资源配置
- 反压点
- 数据倾斜





01

# 流计算开发运维痛点

## 监控报警

- 延时，没数据
- 数据波动
- failover
- . . .



02

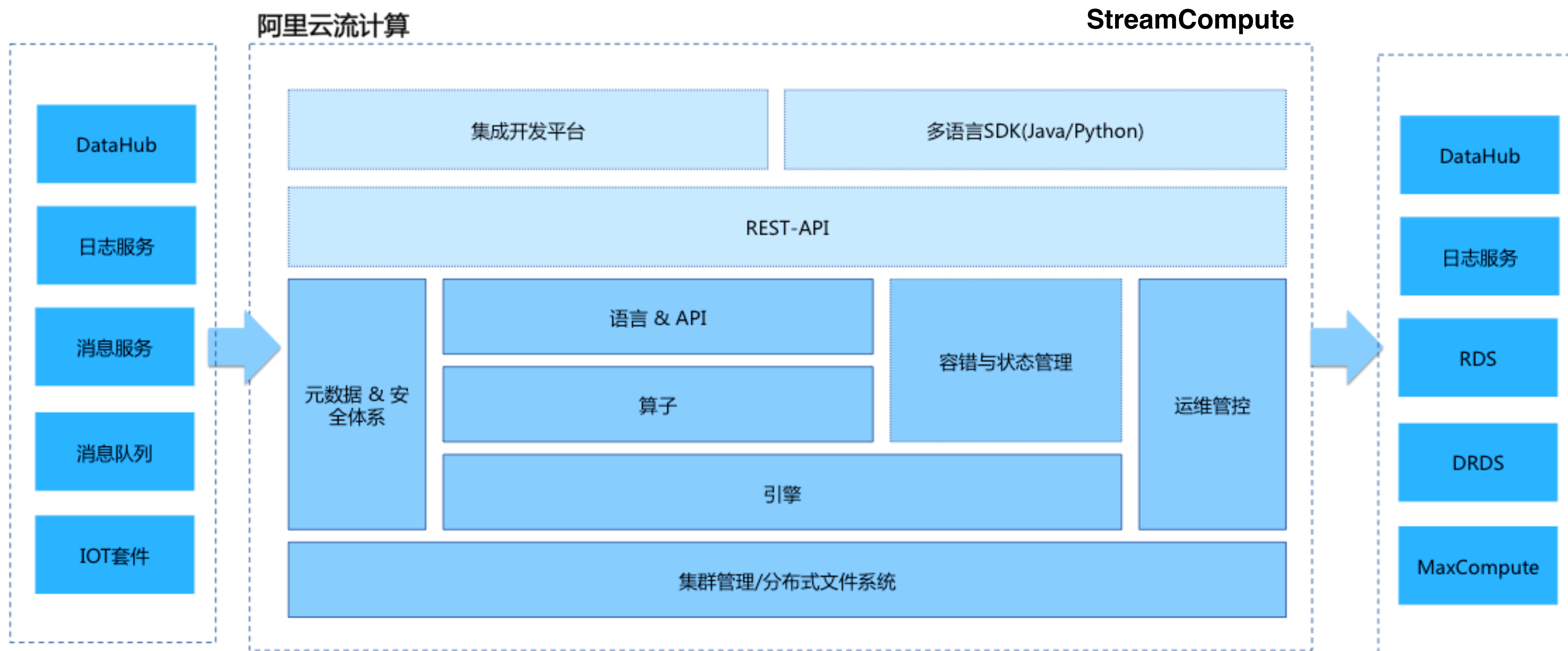
# 基于Flink的流计算平台

StreamCompute

AliCloud

一站式平台

# 基于Flink的流计算平台







## 双11实时大屏

- 交易峰值: 30+ 万笔/s
- 支付峰值: 20+ 万笔/s
- 日志峰值: 数亿条/s





02

## 基于Flink的流计算平台



Apache Flink

+



Alibaba's Improvements

=



Alibaba Blink



Alibaba Blink

+



Productization

=



StreamCompute



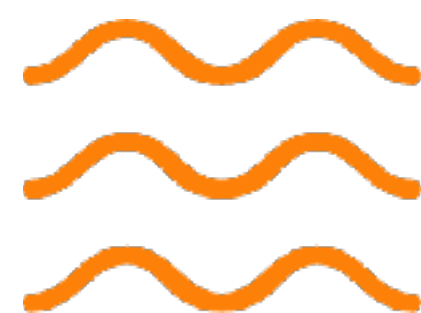


02

## 基于Flink的流计算平台



UDF/UDTF/UDAF



Stream JOIN, etc.



Retraction



Window AGG

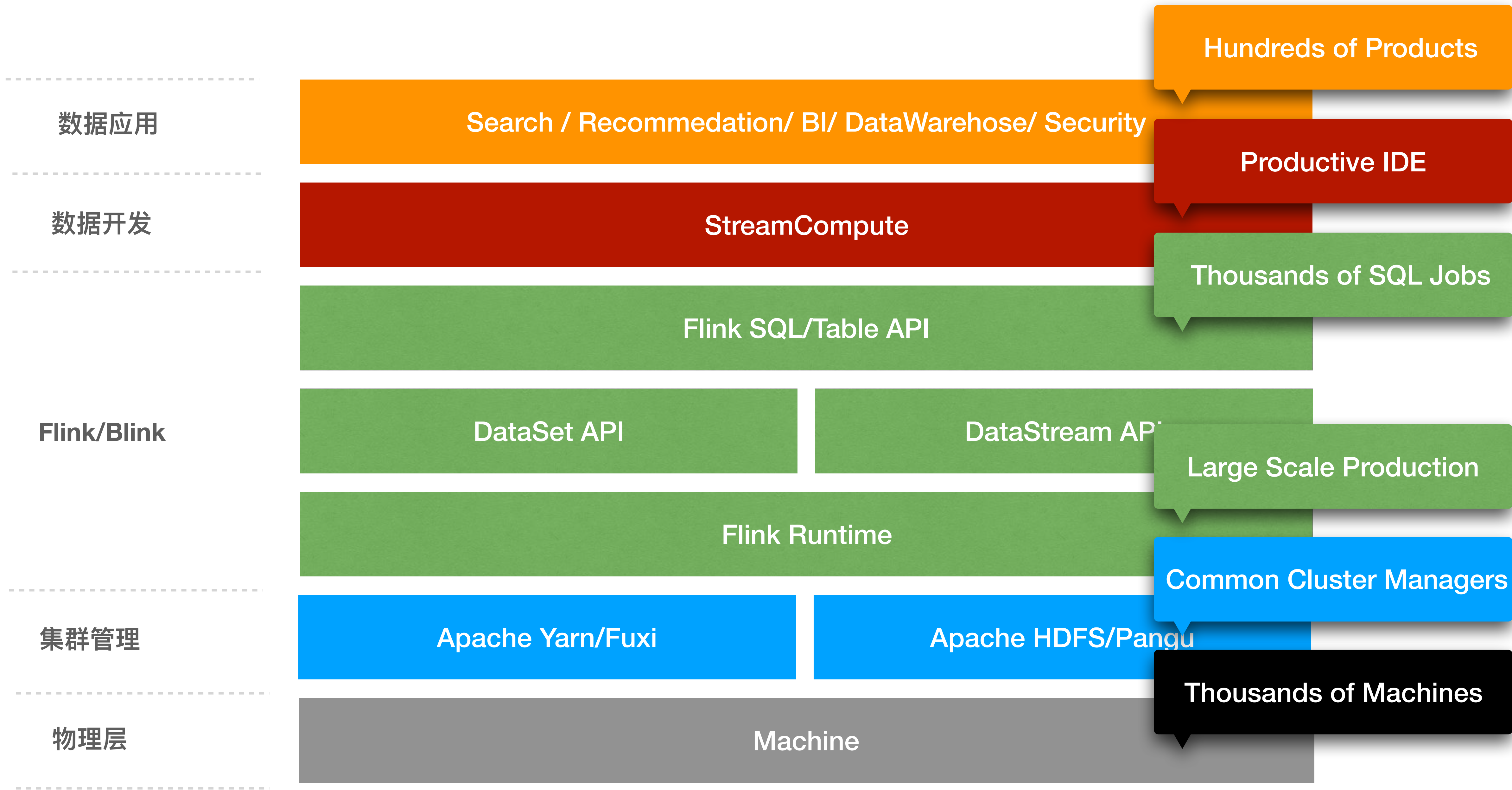


DML: INSERT etc.



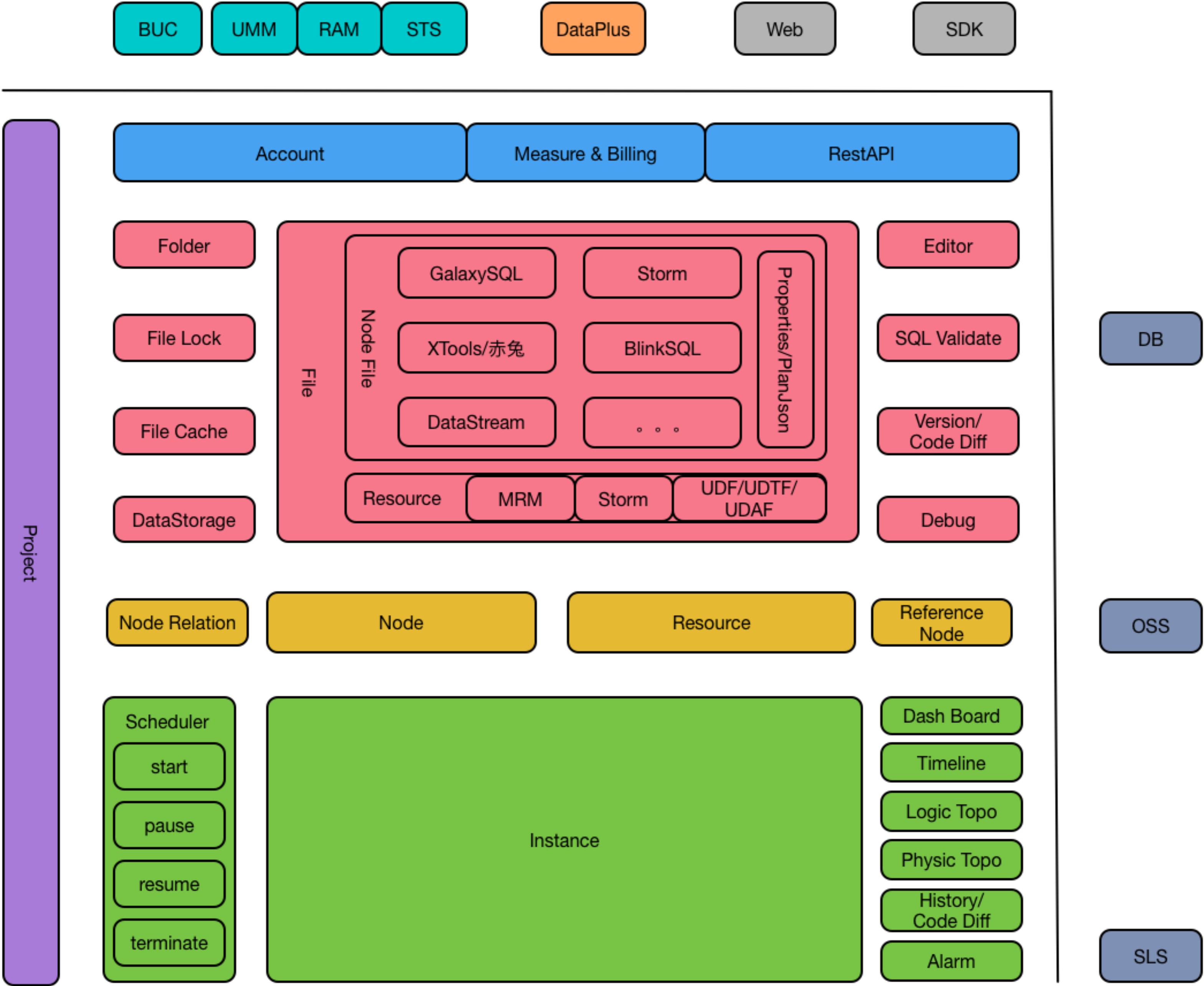
DDL

SQL改造

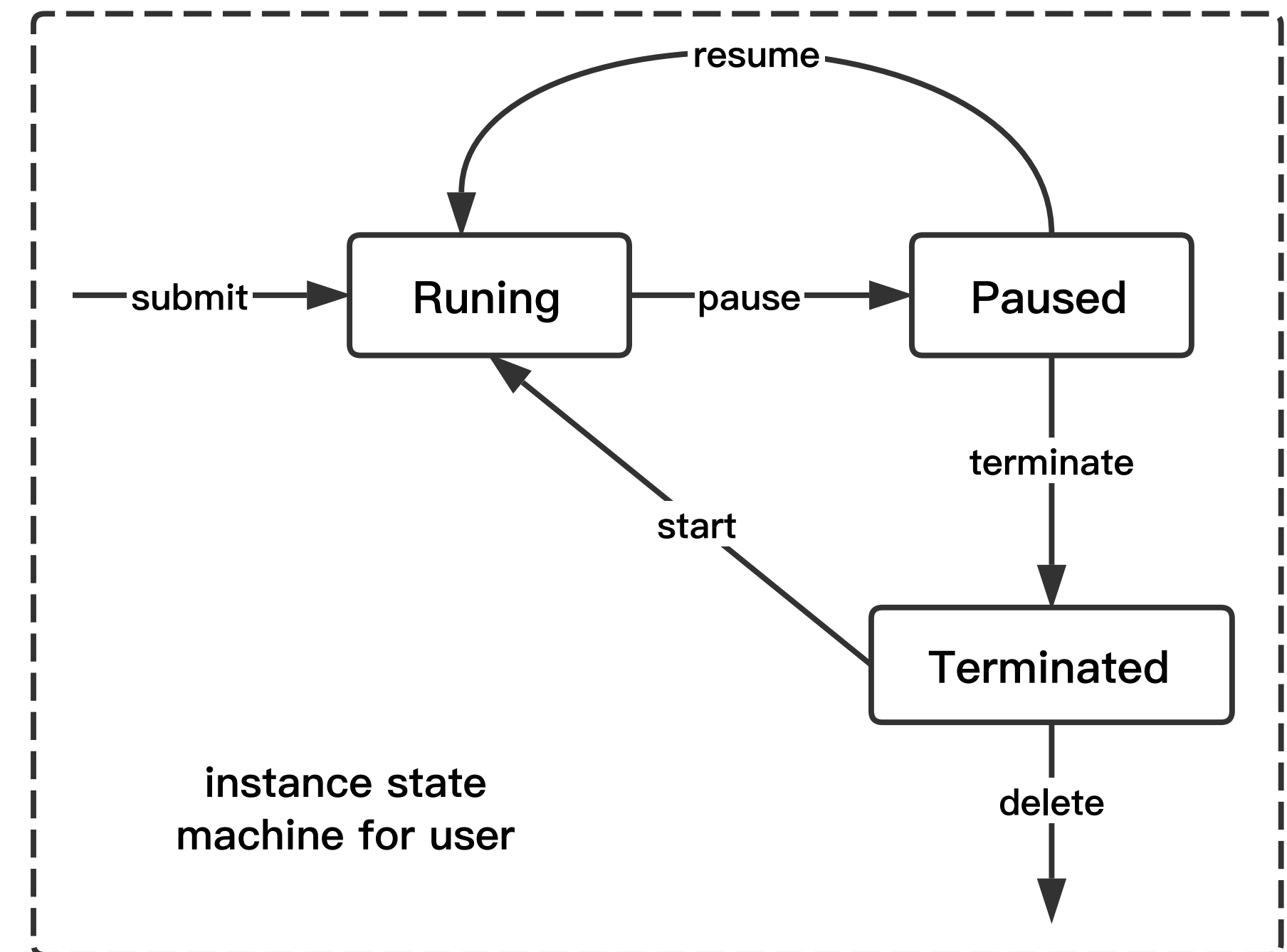
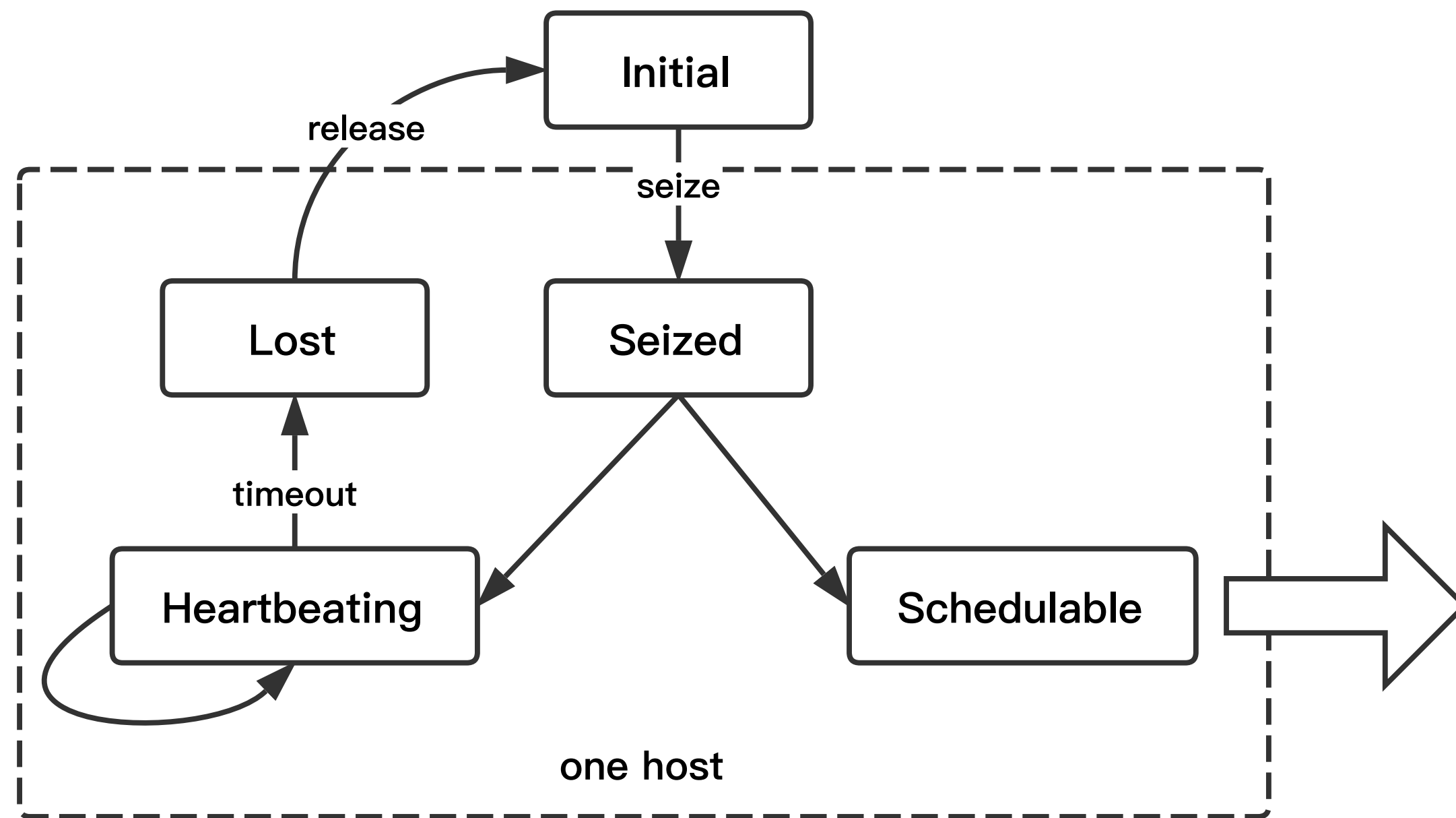


# 基于Blink的流计算平台

- namespace: project
- 阿里云账号权限体系
- 一站式
  - 数据探查
  - 数据开发
  - 数据运维
  - 性能调优
  - 监控报警
- 轻量化

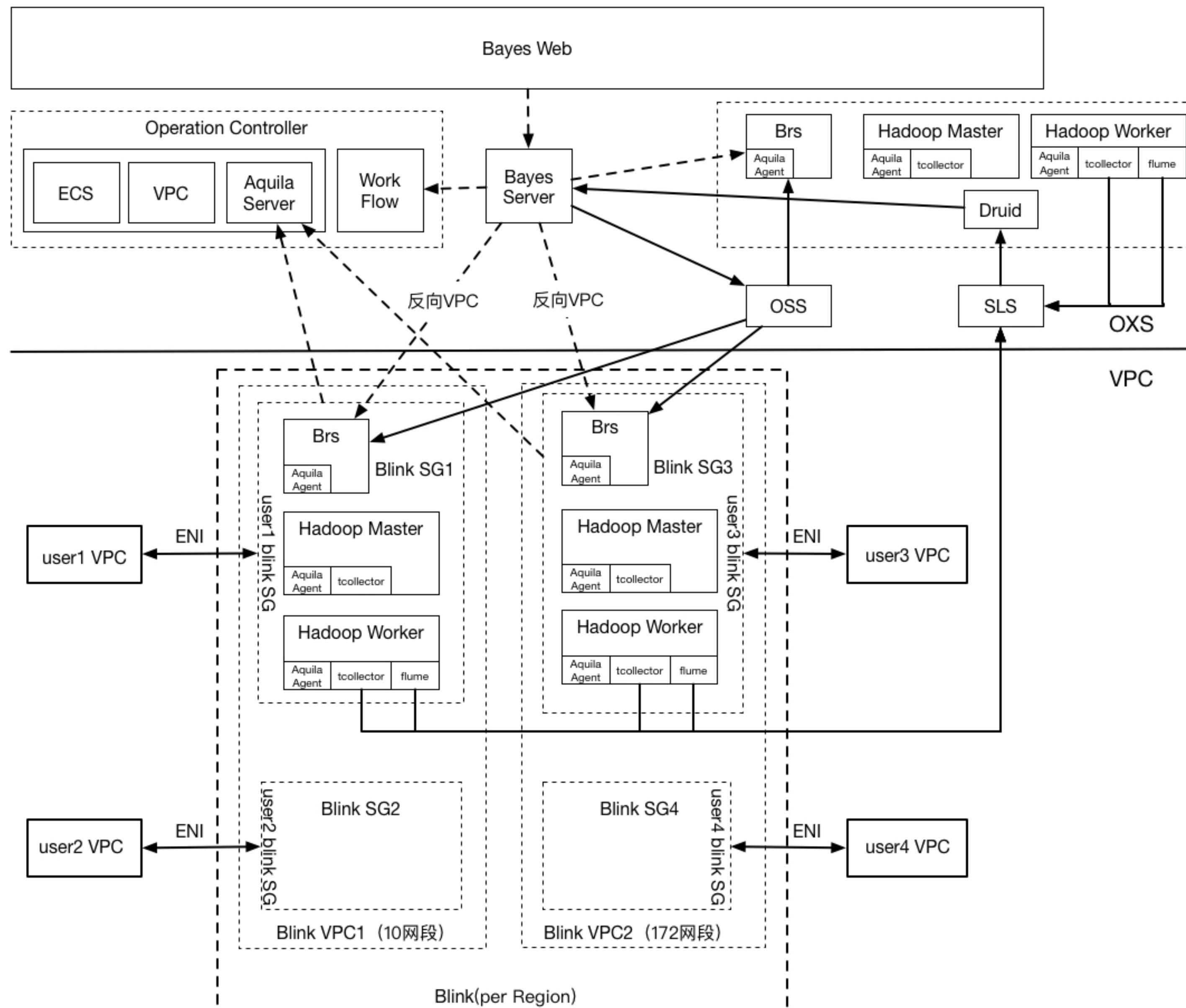


# 基于Flink的流计算平台



```
<!-- 抢占待运行instance ! -->
<update id="seizeInstance"...>

<!-- 给抢占的且运行的instance发心跳 ! -->
<update id="heartbeatInstances"...>
```





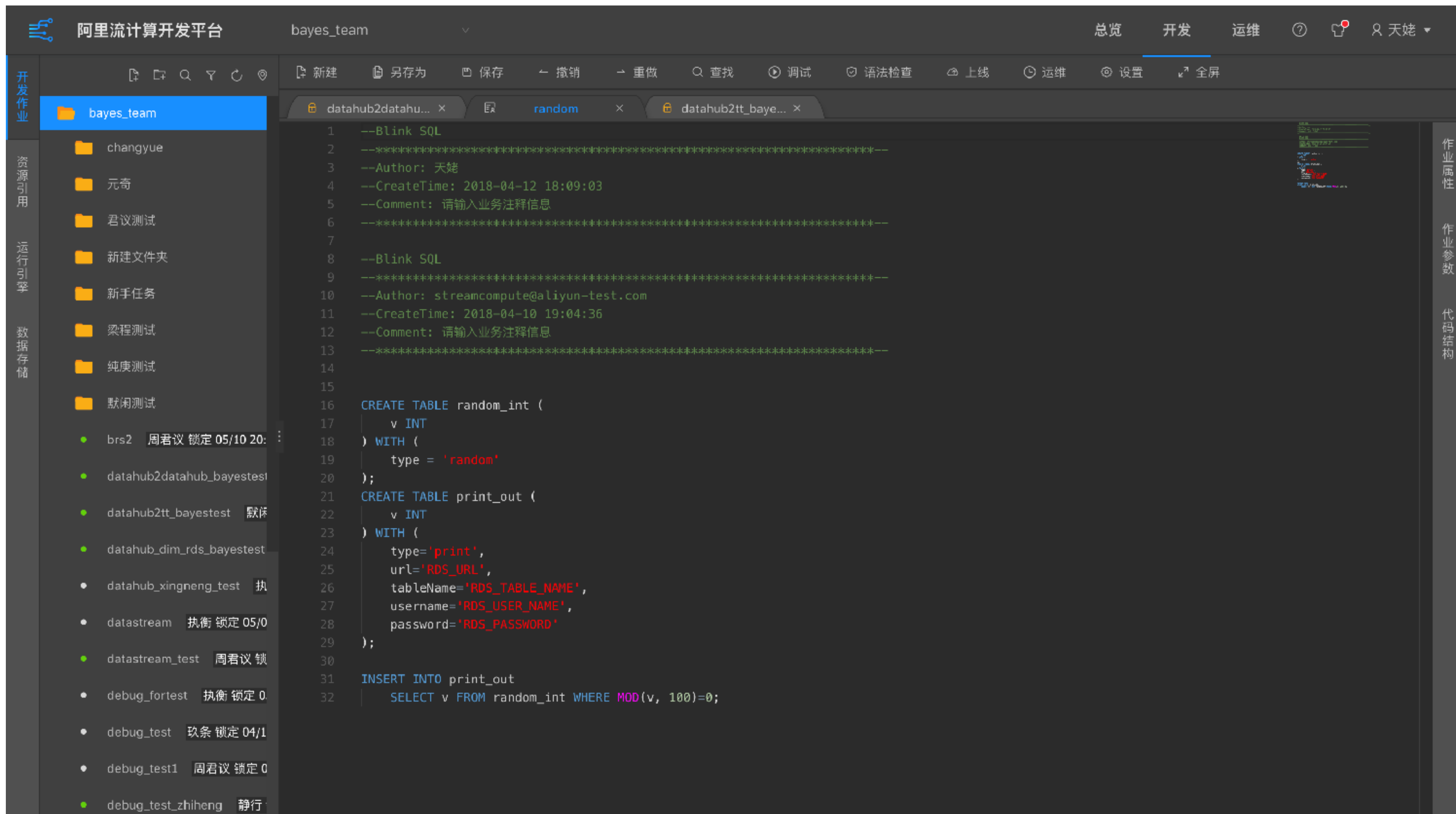




使用SQL+UDX解决底层API问题

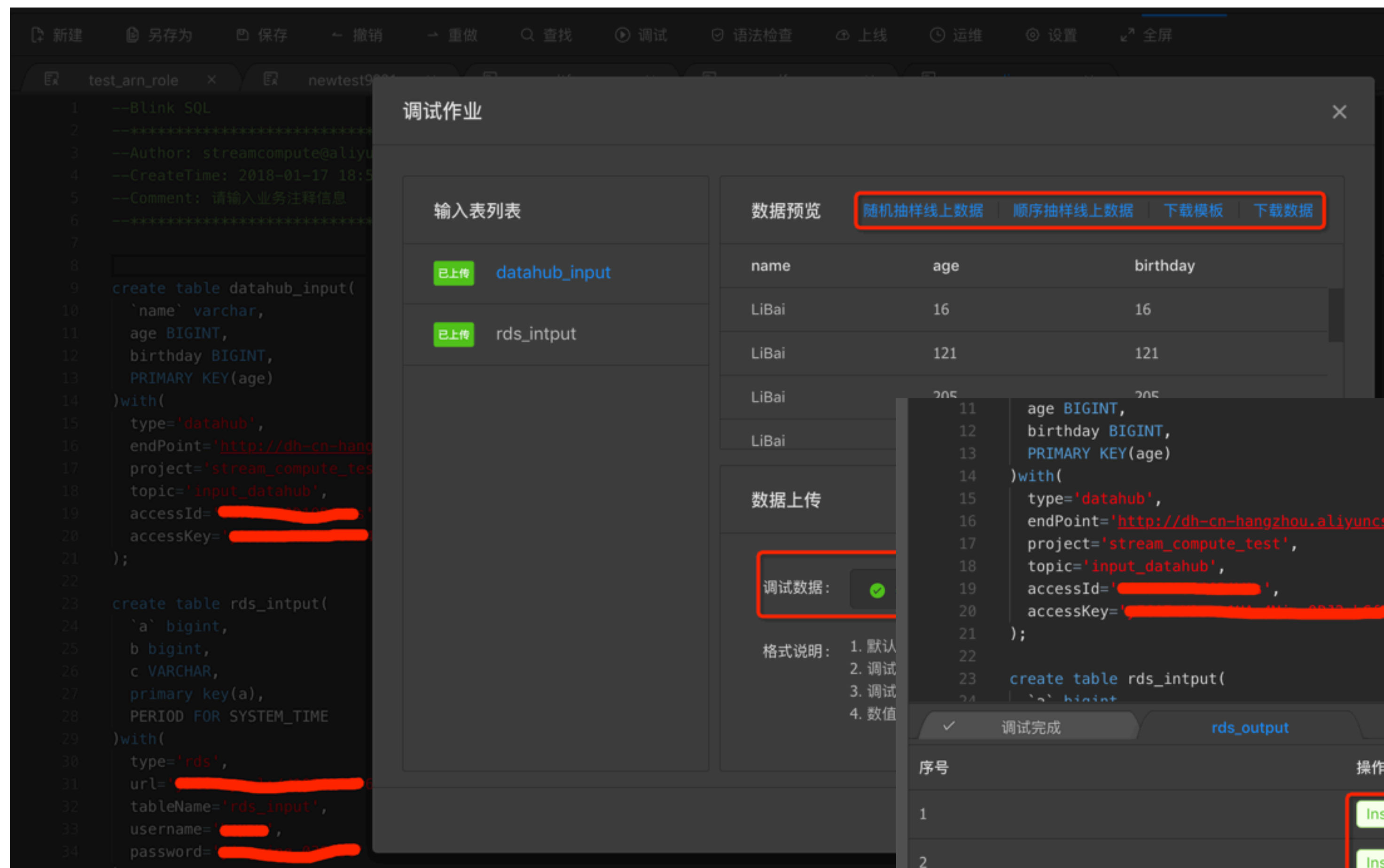
```
129 CREATE VIEW tmp_gift_info_view_m1 as
130 select a.pf,(case when c.typeId is null then 16 when c.typeId='' then 16 else c.typeId) as typeId,
131 a.`user_id`,a.user_nick,
132 a.gift_cost,a.gift_count,a.dateline,DATE_FORMAT(a.dayid,'yyyyMMdd') as dayid
133 from(
134 select d.platform as pf,g.cate_id,g.room_id,g.room_nick,g.`user_id`, g.user_nick,
135 g.price as gift_cost,g.gift_count ,g.`datetime` dateline,g.`datetime` as dayid
136 from gift_v1 g
137 left join ods_wb_platform_i_tbl FOR SYSTEM_TIME AS OF PROCTIME() AS d
138 on g.plat=d.alias
139 ) as a
140 LEFT join ods_wb_categories_i_tbl FOR SYSTEM_TIME AS OF PROCTIME() AS c
141 on a.cate_id=c.cateId and a.pf=c.platform
142 ;
143
144 --1m
145 CREATE VIEW tmp_gift_info_view_m2 as
146 select cast(pf as int) pf,cast(typeId as int) typeId,1 as `anchorType`,room_id,room_nick,
147 cast(TUMBLE_START(dateline, interval '1' MINUTE) as timeStamp) as start_time,
148 cast(TUMBLE_END(dateline, interval '1' MINUTE) as timeStamp) AS end_time,
149 dayid
150 from
151 tmp_gift_info_view_m1 as b
152 GROUP BY TUMBLE(dateline, interval '1' MINUTE),pf,typeId,room_id,room_nick,dayid
153 ;
154
155 --5m
156 CREATE VIEW tmp_gift_info_view_5m as
157 select cast(pf as int) pf,cast(typeId as int) typeId,1 as `anchorType`,room_id,room_nick,
158 cast(HOP_START(dateline, interval '1' MINUTE, interval '5' MINUTE) as timeStamp) as start_time,
159 cast(HOP_END(dateline, interval '1' MINUTE, interval '5' MINUTE) as timeStamp) AS end_time,
160 dayid
161 from
162 tmp_gift_info_view_m1 as b
163 GROUP BY Hop(dateline, interval '1' MINUTE, interval '5' MINUTE),pf,typeId,room_id,room_nick,dayid
164 ;
165
```

# 基于Flink的流计算平台





使用local debug解决  
造数据、调试难问题

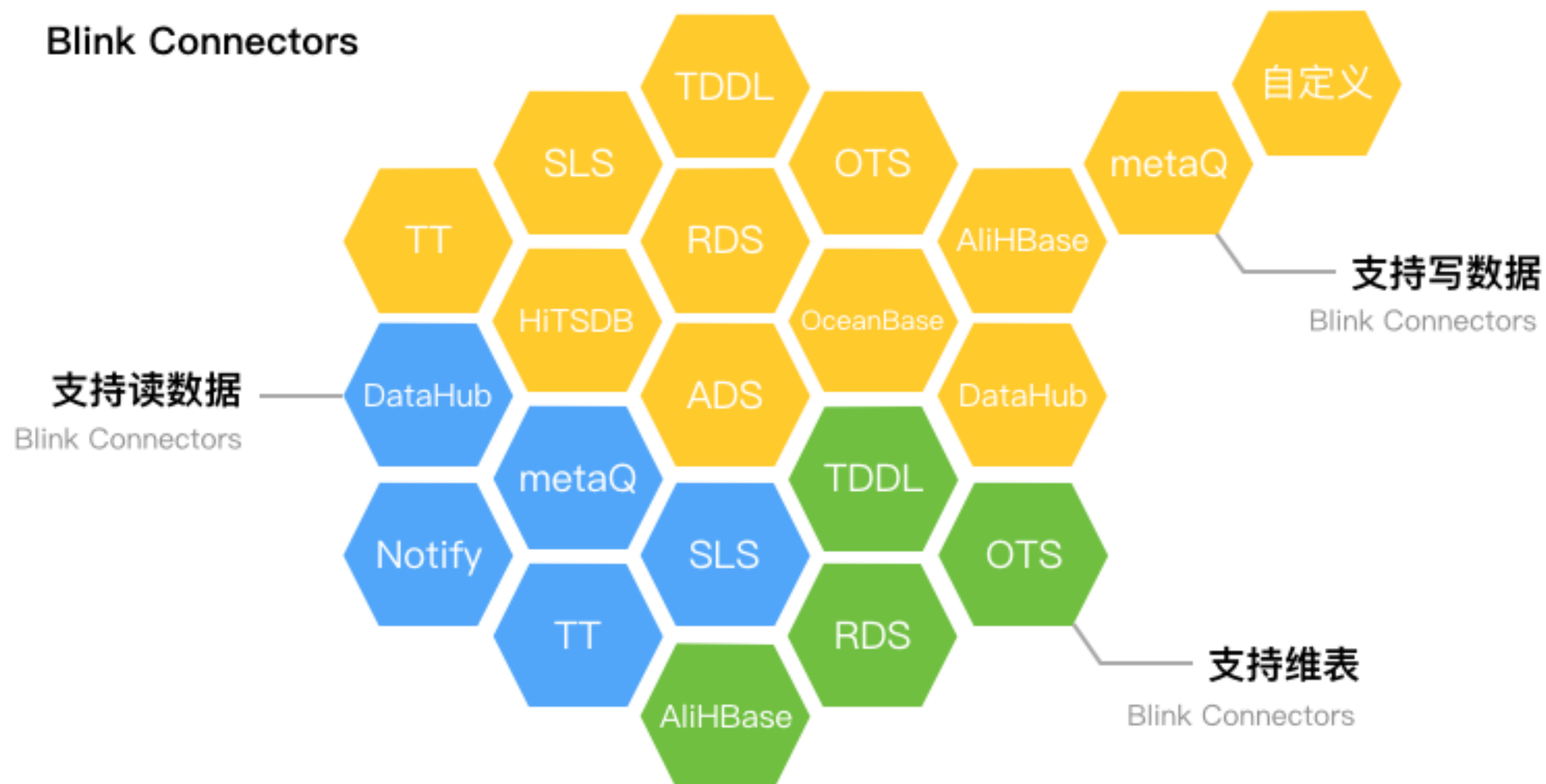


11	age BIGINT,
12	birthday BIGINT,
13	PRIMARY KEY(age)
14	)with(
15	type='datahub',
16	endPoint='http://dh-cn-hangzhou.aliyuncs.com',
17	project='stream_compute_test',
18	topic='input_datahub',
19	accessId='[redacted]',
20	accessKey='[redacted]',
21	);
22	
23	create table rds_input(
24	'a' bigint,
25	b bigint,
26	c VARCHAR,
27	primary key(a),
28	PERIOD FOR SYSTEM_TIME
29	)with(
30	type='rds',
31	url='[redacted]',
32	tableName='rds_input',
33	username='[redacted]',
34	password='[redacted]'

序号	操作	id	name
1	Insert	16	LiBai
2	Insert	121	LiBai
3	Insert	16	LiBai

# 基于Flink的流计算平台



丰富的connector

使用“数据存储”打通各类型存储系统

23ageBIGINT,24PRIMARY KEY(id)

数据表详情

作为输入表引用

作为结果表引用

数据抽样

存储信息

名称: window\_input | 存储类型: DataHub | 创建时间: 2017-12-01 09:46:19 | shard数量: 1 | 生命周期: 3 天

数据预览

Shard ID	System Time	a (STRING)	b (STRING)
0	2018-04-11 16:00:09	静夜思	

数据抽样

抽样结果

\* 指定时间 : 2018-04-11 16:02:45

\* Shard ID : 0

\* 抽样数量 : 10

\* 分隔符 : ,

抽样数据

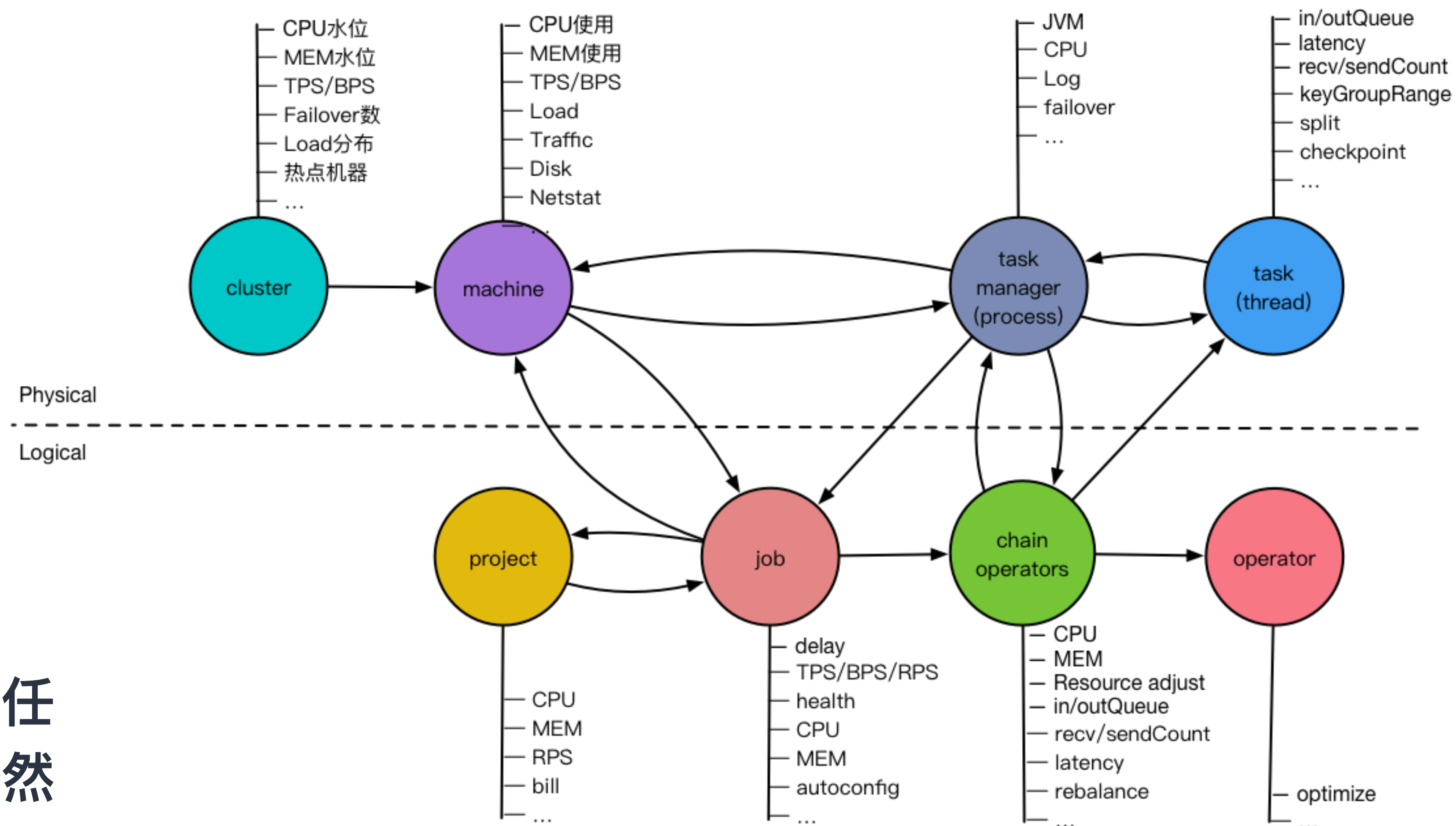
下载数据

抽样结果

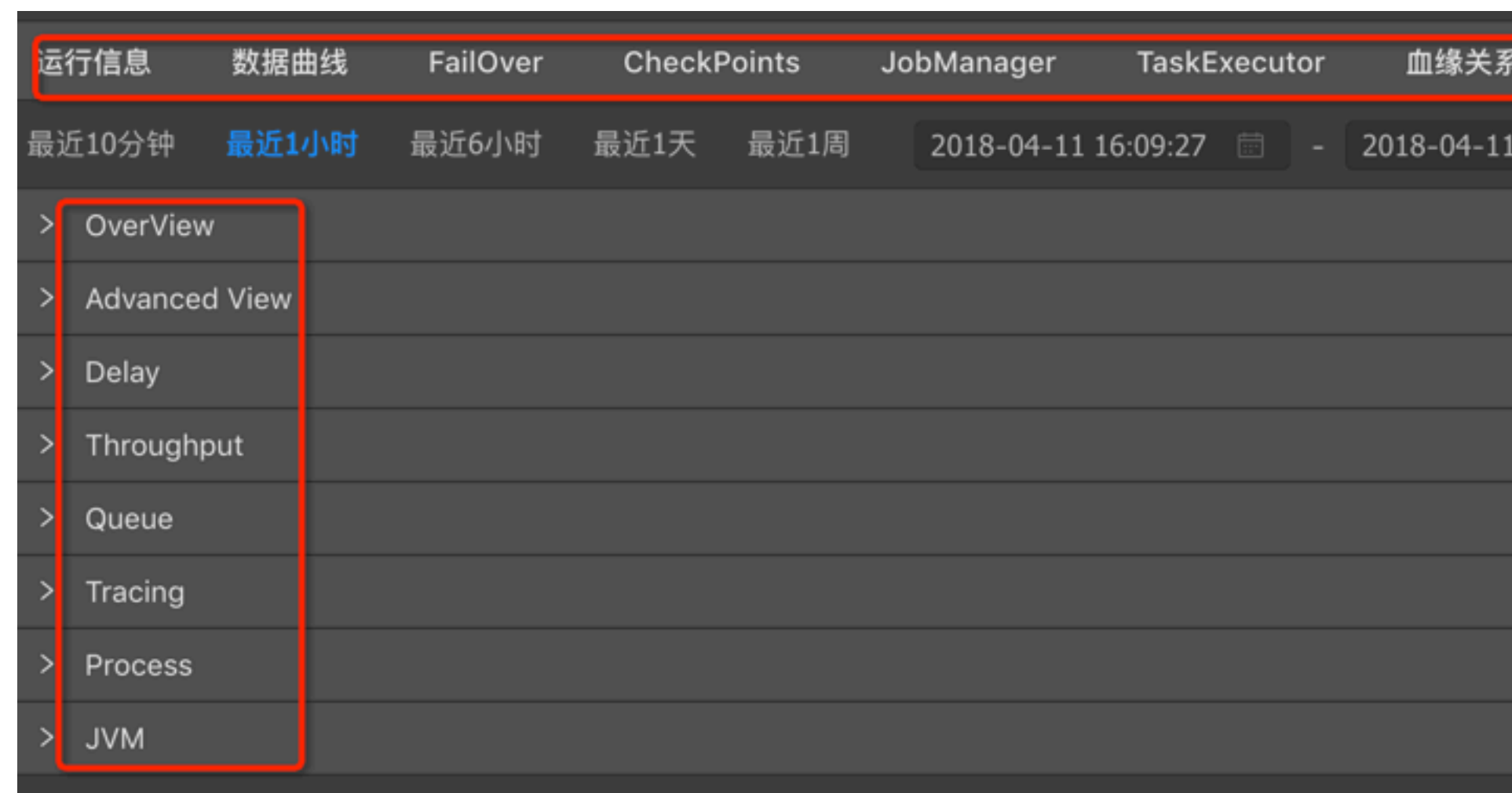
Shard ID	System Time	a (STRING)	b (STRIN
0	2018-04-11 16:03:19	静夜思	
0	2018-04-11 16:03:19	静夜思	
0	2018-04-11 16:03:19	静夜思	
0	2018-04-11 16:03:19	静夜思	
0	2018-04-11 16:03:19	静夜思	
0	2018-04-11 16:03:19	静夜思	
0	2018-04-11 16:03:19	静夜思	
0	2018-04-11 16:03:19	静夜思	

< 1 2 >





丰富的指标曲线让任务健康状况一目了然

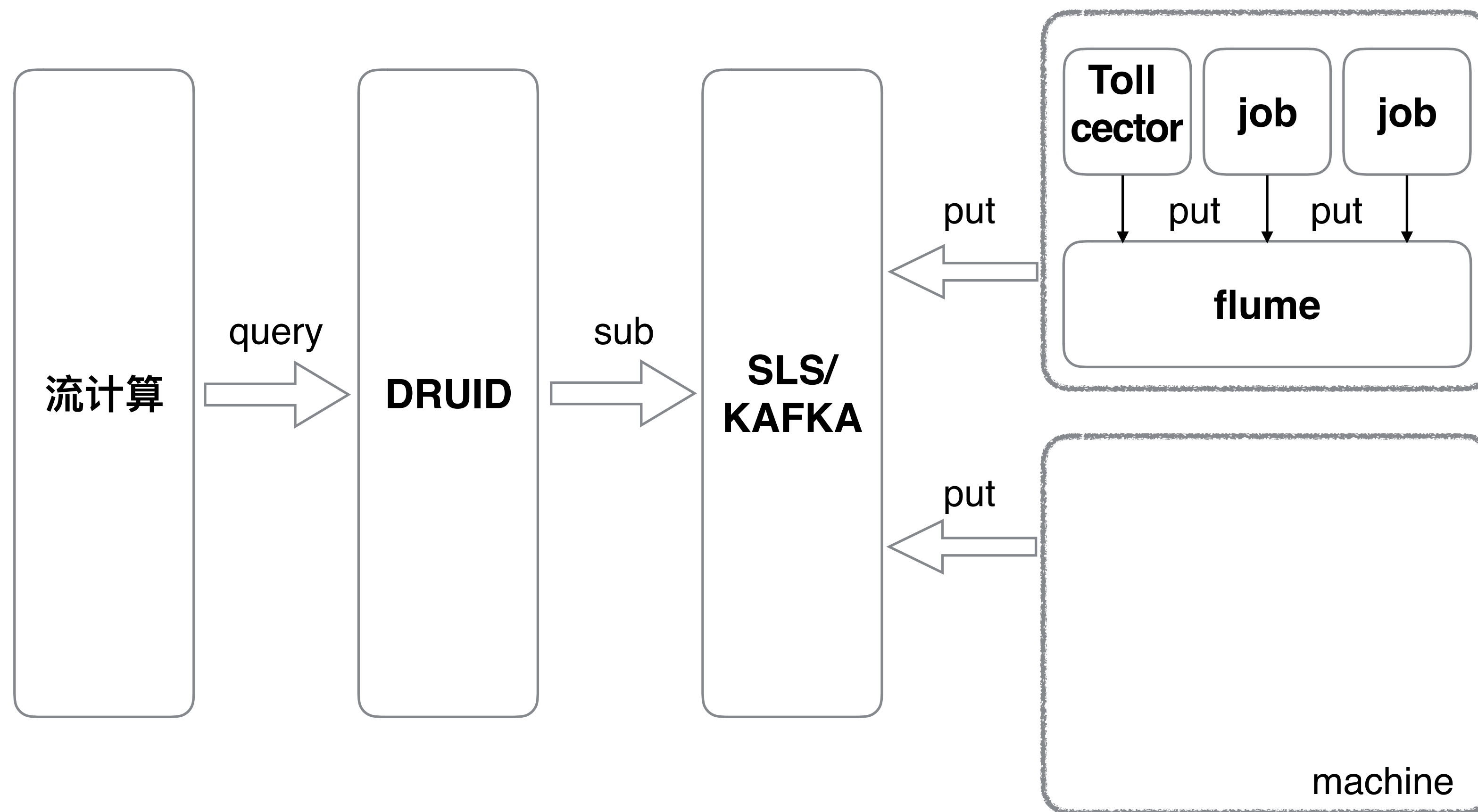


丰富的指标曲线让任务健康状况一目了然

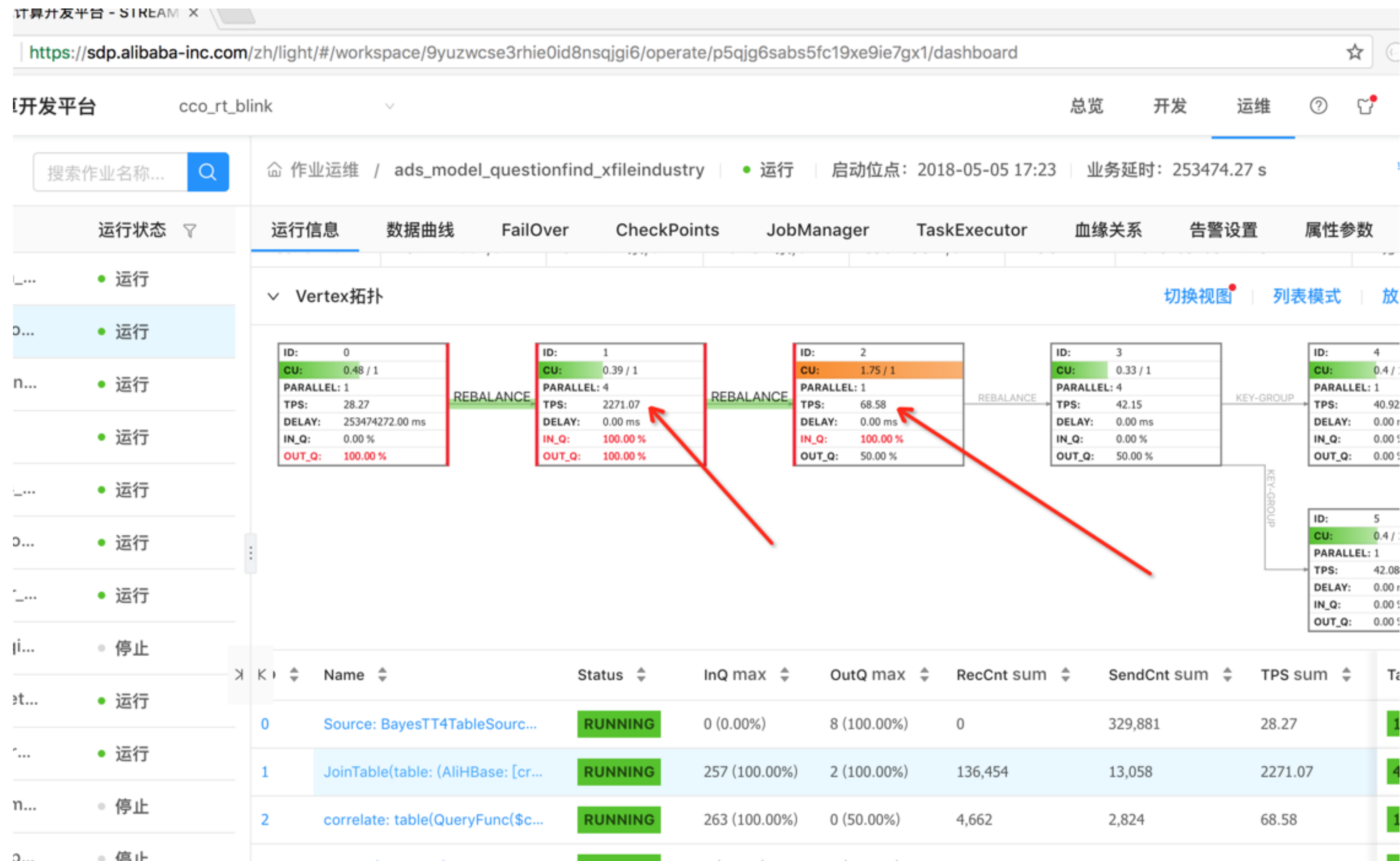


# 基于Flink的流计算平台

## 指标采集流程



调优





← → ↺ ⌂ ⓘ zprofiler.alibaba-inc.com/dynamic/threads.htm?ip\_port=hadoop0216.et2.tbsite.net:1036 ☆

**ZProfiler** Heap Dump Thread Dump GC Log HotMethod MethodTracing **Insight** ZDebugger 乔燃 (王刚) 退出

动态视图  
系统环境  
系统动态  
GC阶段耗时  
内存动态  
Load  
Hot Method  
Method Tracing  
膨胀锁  
线程动态  
单线程动态

控制区  
手动刷新线程列表 暂停采样

线程列表 【根据间隔时间内cpu损耗百分比排列】

🔍	[cpu=100.0%]	[sample_time=4.99s]	[name=where: (= (order_id, ttp_order_id)), join: (order_id, buyer_id, total_price, pass_seg_count, is_wireless, ttid, order_attributes, -> select: (buyer_id, CASE(=(is_wireless, 'Y'), 1, 0) AS is_wireless
🔍	[cpu=100.0%]	[sample_time=4.99s]	[name=where: (= (order_id, ttp_order_id)), join: (order_id, buyer_id, total_price, pass_seg_count, is_wireless, ttid, order_attributes, -> select: (buyer_id, CASE(=(is_wireless, 'Y'), 1, 0) AS is_wireless
🔍	[cpu=0.2%]	[sample_time=4.99s]	[name=pool-3-thread-1] [tid=150] [status=WAITING] [blocked_count=0]
🔍	[cpu=0.1%]	[sample_time=4.96s]	[name=Timer for 'kmonitor' metrics system] [tid=34] [status=TIMED_WAITING] [blocked_count=0]
🔍	[cpu=0.1%]	[sample_time=4.96s]	[name=Flink-MetricRegistry-1] [tid=35] [status=TIMED_WAITING] [blocked_count=0]
🔍	[cpu=0.1%]	[sample_time=4.96s]	[name=flink-scheduler-1] [tid=43] [status=TIMED_WAITING] [blocked_count=0]
🔍	[cpu=0.0%]	[sample_time=4.96s]	[name=main] [tid=1] [status=WAITING] [blocked_count=30]
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=ShortCircuitCache_Cleaner] [tid=258] [status=TIMED_WAITING] [blocked_count=0]
🔍	[cpu=0.0%]	[sample_time=4.96s]	[name=Reference Handler] [tid=2] [status=WAITING] [blocked_count=186]
🔍	[cpu=0.0%]	[sample_time=4.96s]	[name=Finalizer] [tid=3] [status=WAITING] [blocked_count=915]
🔍	[cpu=0.0%]	[sample_time=4.96s]	[name=Signal Dispatcher] [tid=5] [status=RUNNABLE] [blocked_count=0]
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=IPC Parameter Sending Thread #8] [tid=519] [status=TIMED_WAITING] [blocked_count=1]
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=LeaseRenewer:admin@et2prod2] [tid=522] [status=TIMED_WAITING] [blocked_count=993]
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=DataStreamer for file /blink_rest_server/root.TRAV.fliggy_trip/online_ads_biz_trvl_train_thomas_pay/rocksdb_cp/fc311faf-bd5a-4c72-a767-d945471a9ac1 block BP-1538344543-11.180.34.2
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=ResponseProcessor for block BP-1538344543-11.180.34.220-1504493465110:blk_2019487142_946919585] [tid=1299] [status=RUNNABLE] [blocked_count=9]
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=DataStreamer for file /blink_rest_server/root.TRAV.fliggy_trip/online_ads_biz_trvl_train_thomas_pay/rocksdb_cp/fc94f789-5f9a-405d-87f9-e320886f5507 block BP-1538344543-11.180.34.22
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=ResponseProcessor for block BP-1538344543-11.180.34.220-1504493465110:blk_2019487146_946919589] [tid=1301] [status=RUNNABLE] [blocked_count=5]
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=DataStreamer for file /blink_rest_server/root.TRAV.fliggy_trip/online_ads_biz_trvl_train_thomas_pay/rocksdb_cp/cf3e4a47-098b-4427-9774-085b13118297 block BP-1538344543-11.180.34
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=ResponseProcessor for block BP-1538344543-11.180.34.220-1504493465110:blk_2019487148_946919591] [tid=1303] [status=RUNNABLE] [blocked_count=8]
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=DataStreamer for file /blink_rest_server/root.TRAV.fliggy_trip/online_ads_biz_trvl_train_thomas_pay/rocksdb_cp/613434da-c014-4da4-9e2b-6ef23387d623 block BP-1538344543-11.180.34
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=ResponseProcessor for block BP-1538344543-11.180.34.220-1504493465110:blk_2019487209_946919652] [tid=1305] [status=RUNNABLE] [blocked_count=8]
🔍	[cpu=0.0%]	[sample_time=4.96s]	[name=Curator-ConnectionStateManager-0] [tid=30] [status=WAITING] [blocked_count=0]
🔍	[cpu=0.0%]	[sample_time=4.96s]	[name=main-SendThread(11.141.158.10:12181)] [tid=31] [status=RUNNABLE] [blocked_count=1]
🔍	[cpu=0.0%]	[sample_time=4.96s]	[name=main-EventThread] [tid=32] [status=WAITING] [blocked_count=4]
🔍	[cpu=0.0%]	[sample_time=4.96s]	[name=Curator-Framework-0] [tid=33] [status=WAITING] [blocked_count=0]
🔍	[cpu=0.0%]	[sample_time=4.96s]	[name=RMI Reaper] [tid=37] [status=WAITING] [blocked_count=0]
🔍	[cpu=0.0%]	[sample_time=4.96s]	[name=GC Daemon] [tid=38] [status=TIMED_WAITING] [blocked_count=0]
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=flink-akka.actor.default-dispatcher-16] [tid=296] [status=WAITING] [blocked_count=1]
🔍	[cpu=0.0%]	[sample_time=4.96s]	[name=RMI Scheduler(0)] [tid=40] [status=TIMED_WAITING] [blocked_count=0]
🔍	[cpu=0.0%]	[sample_time=4.99s]	[name=flink-akka.actor.default-dispatcher-15] [tid=297] [status=WAITING] [blocked_count=0]







## 02

## 基于Flink的流计算平台



智能调优 + 手动  
灵活的资源配置



02

## 基于Flink的流计算平台

< 返回实例列表

创建报警规则

前往此实例控制台

刷新

监控图表

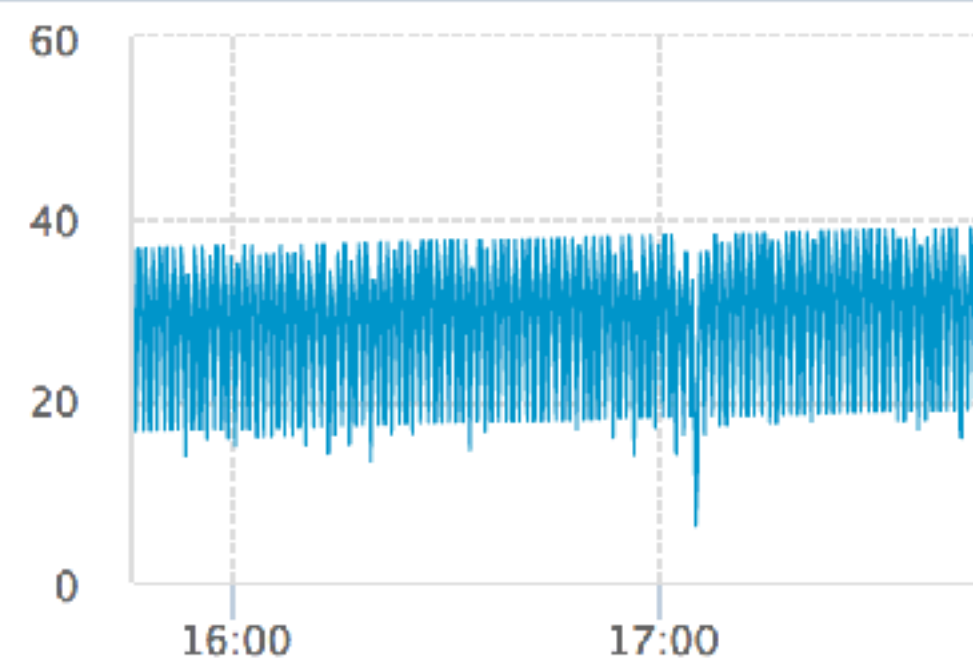
报警规则



业务延迟(秒)



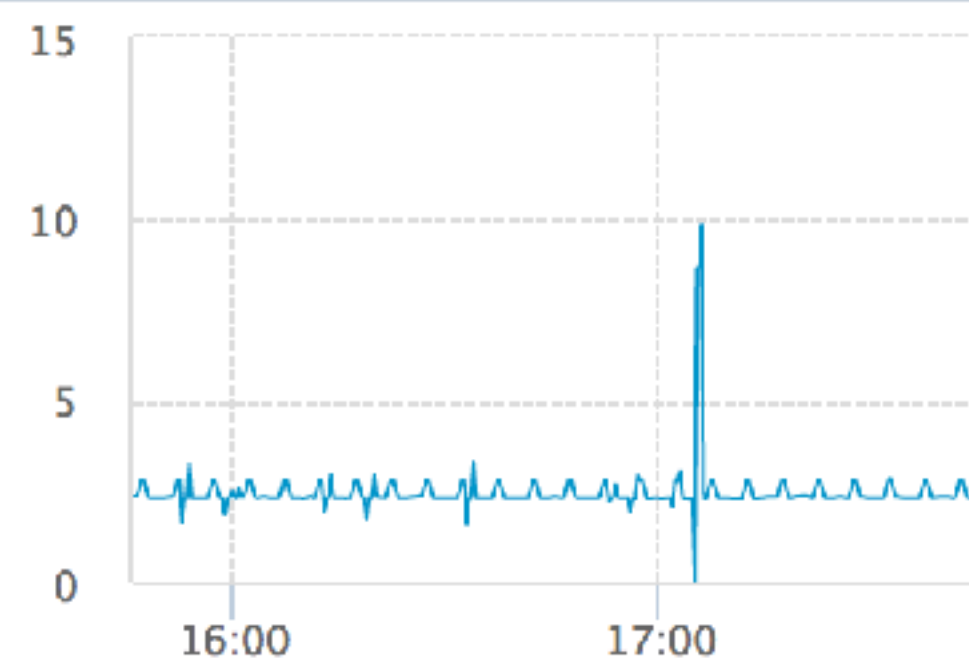
周期: 20s 聚合方式: Average



读入RPS(RPS)



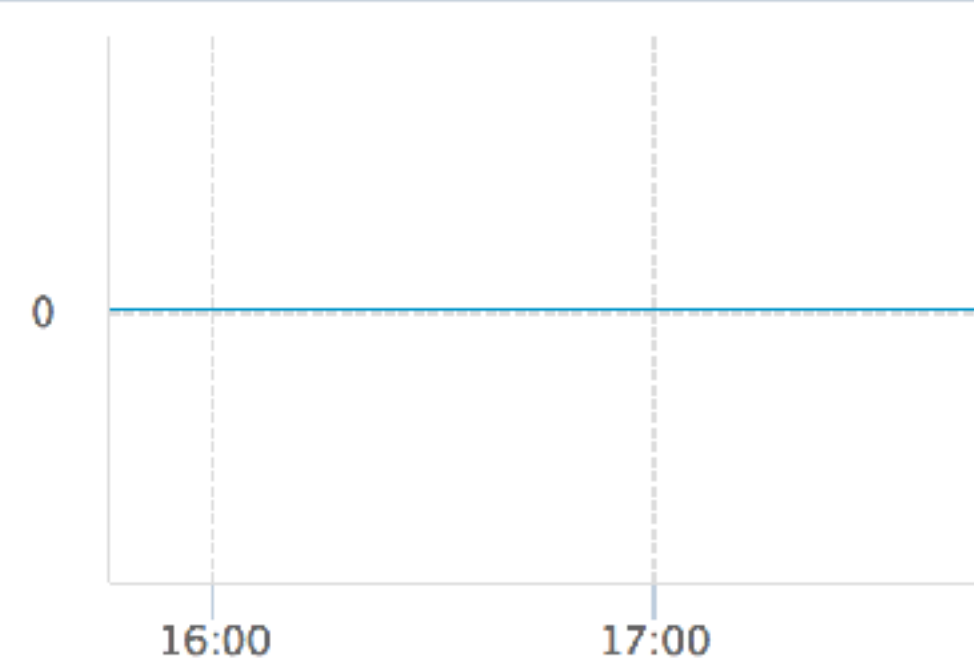
周期: 20s 聚合方式: Average



写出RPS(RPS)



周期: 20s 聚合方式: Average



FailoverRate(%)

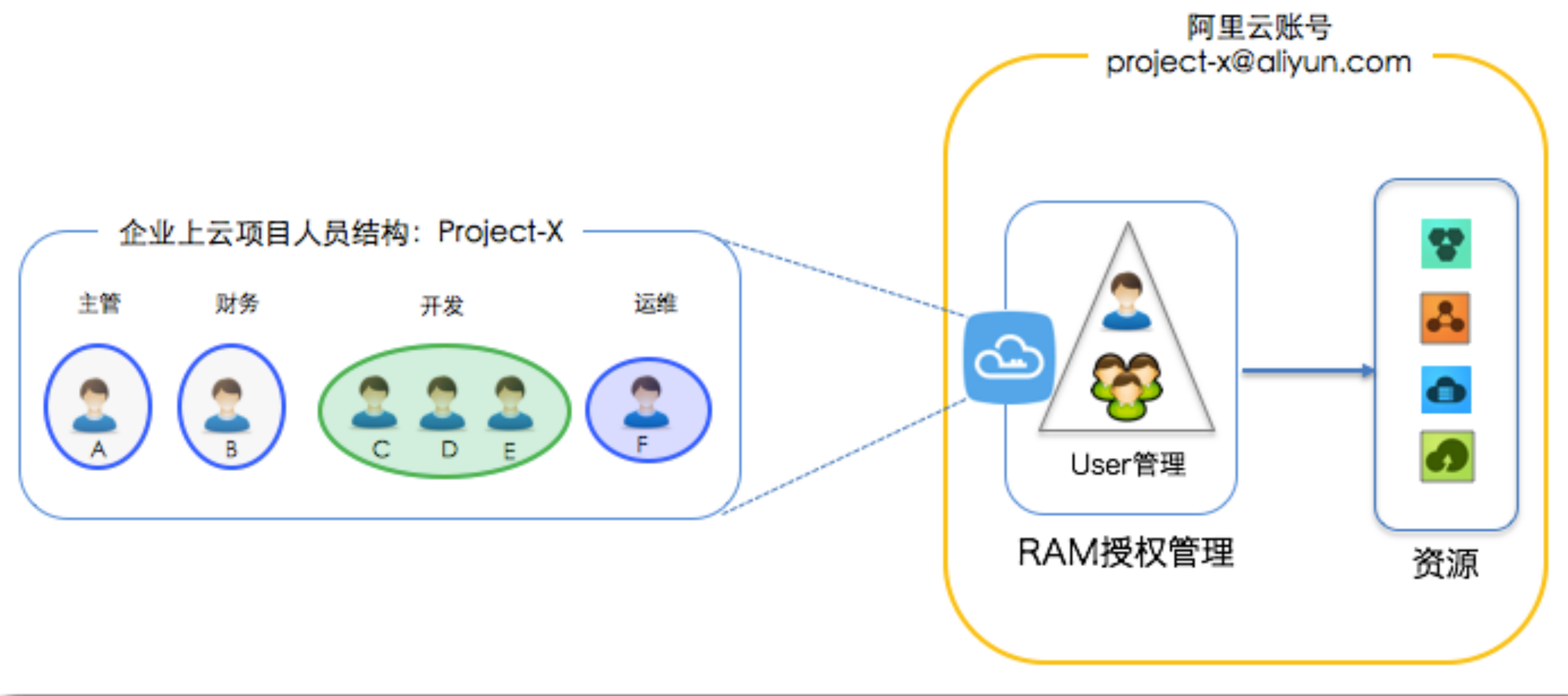


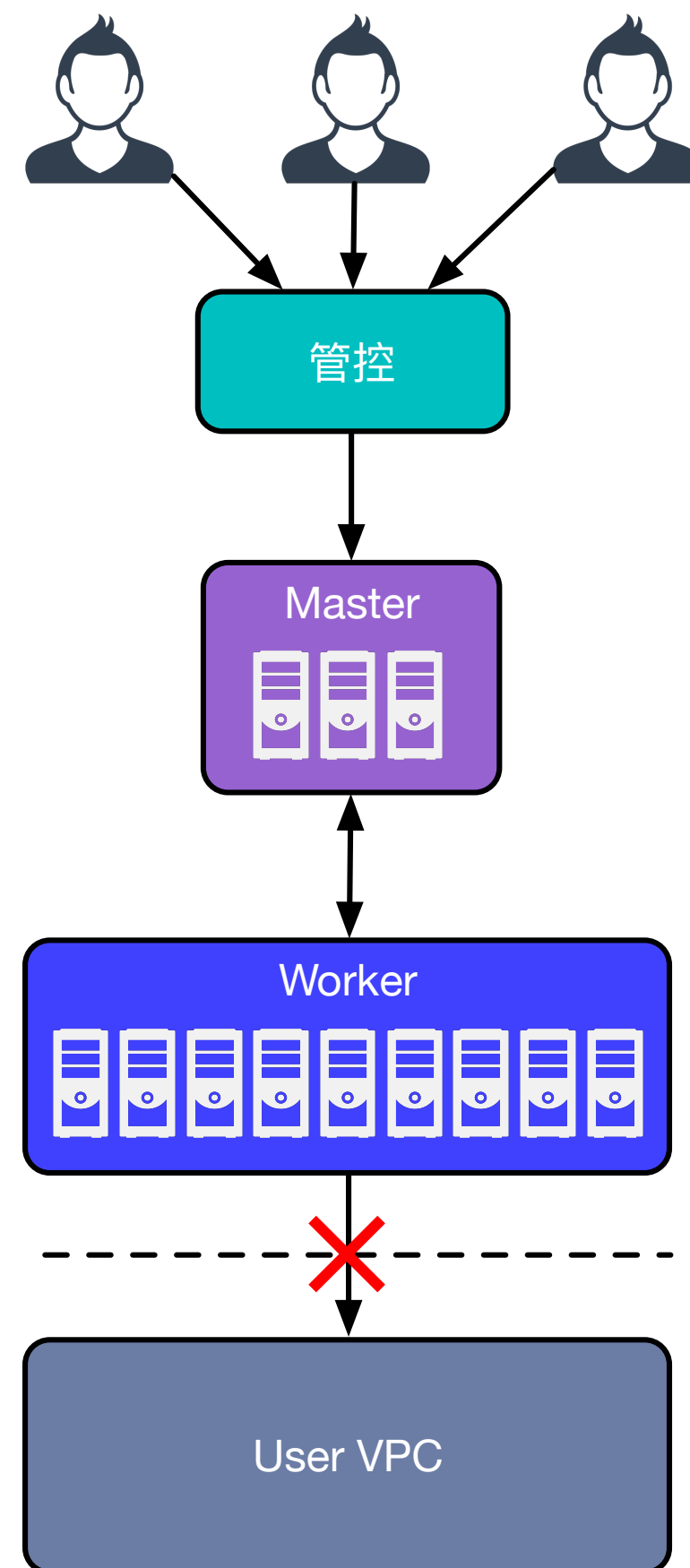
周期: 20s 聚合方式: Average



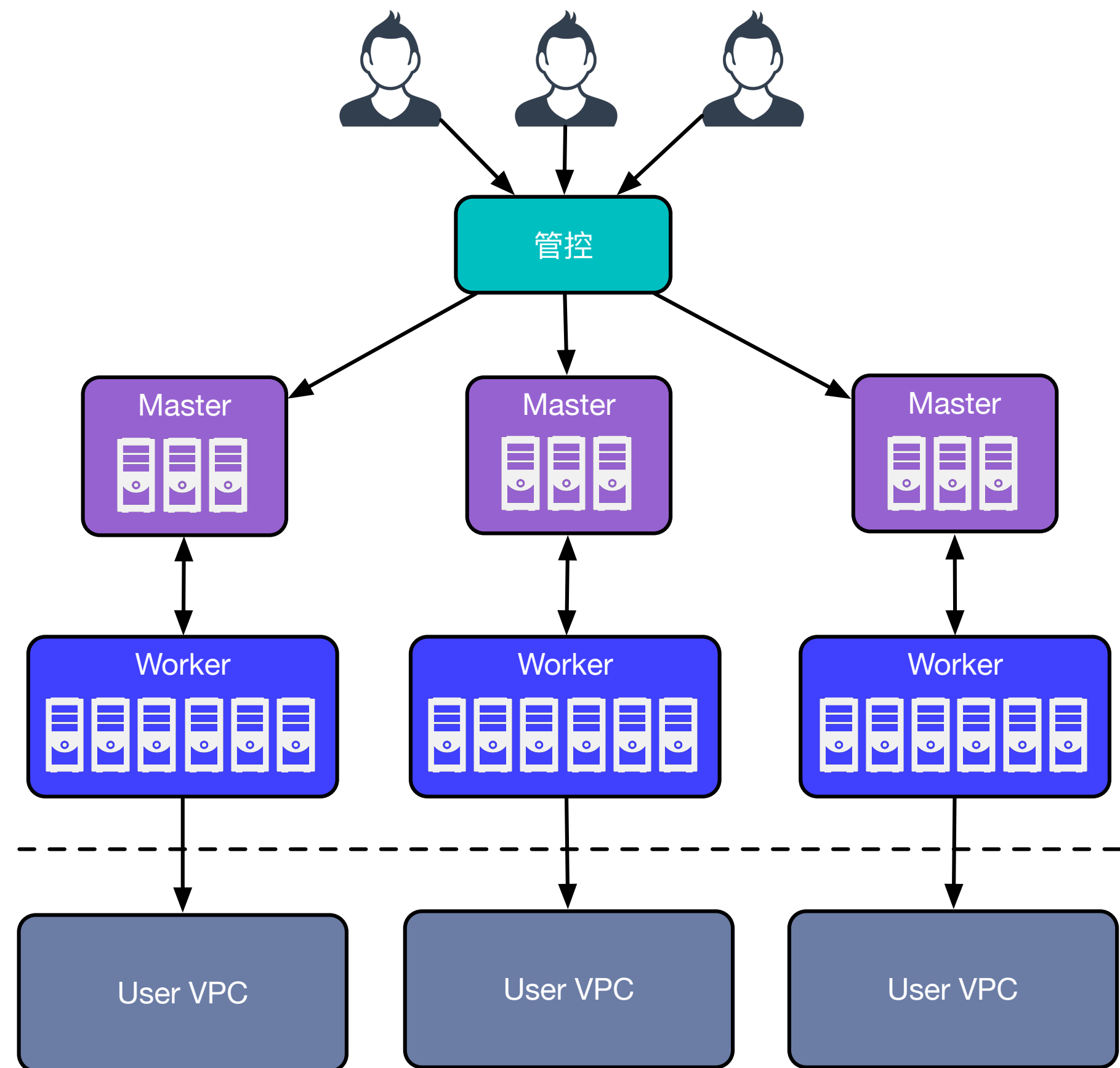
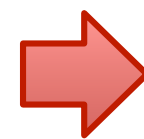
统一云监控配置告警

- 1.支持子用户授权
- 2.支持带IP/SSL限制的授权
- 3.支持带MFA限制的授权
- 4.带时间限制的授权
- ...

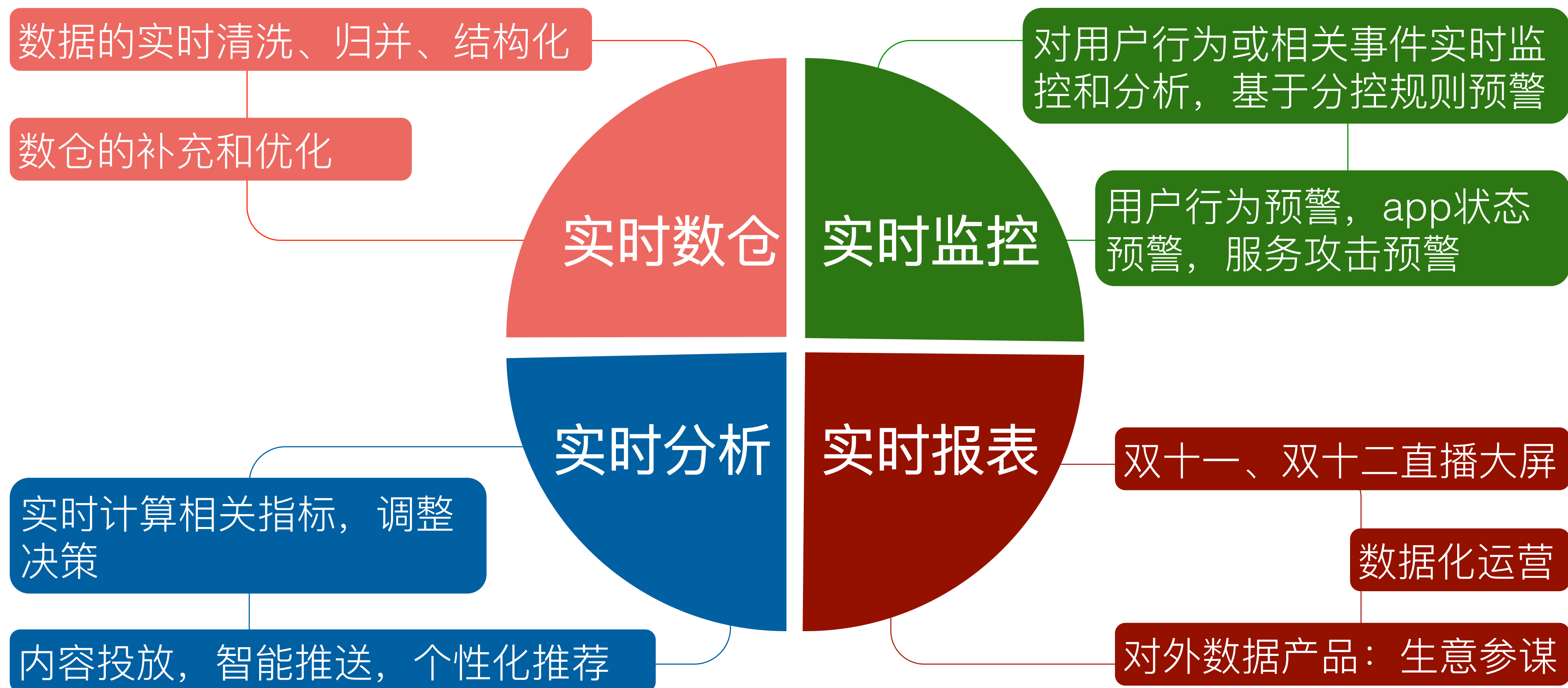




逻辑隔离



物理隔离





Thanks

Q&A