

# 流计算——让业务实时，让数据发电

阿里云流计算产品经理——付空

1  
为什么要用流计算

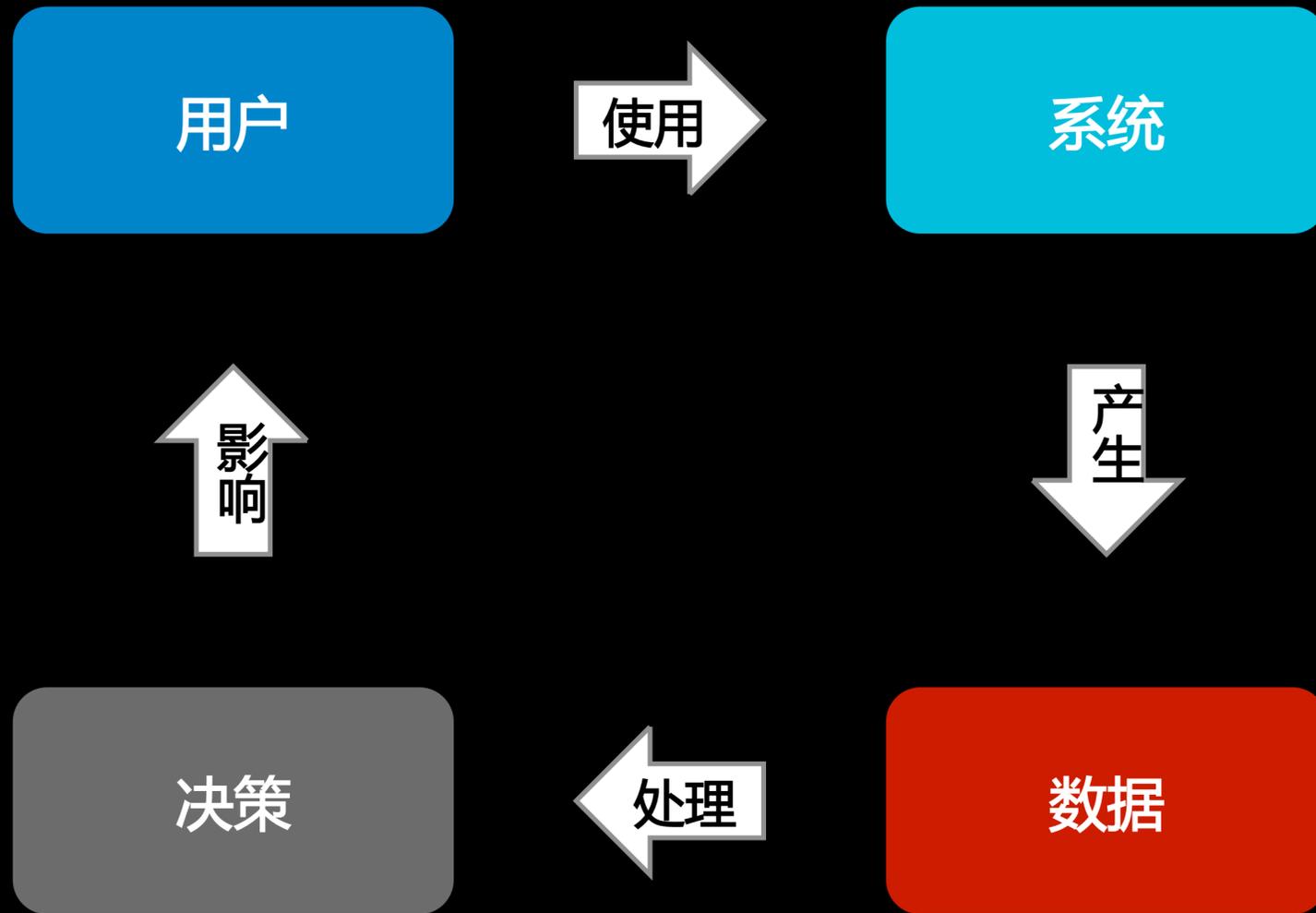
2  
为什么要用阿里云流计算

3  
如何用流计算

**1**  
为什么要用流计算

**2**  
为什么要用阿里云流计算

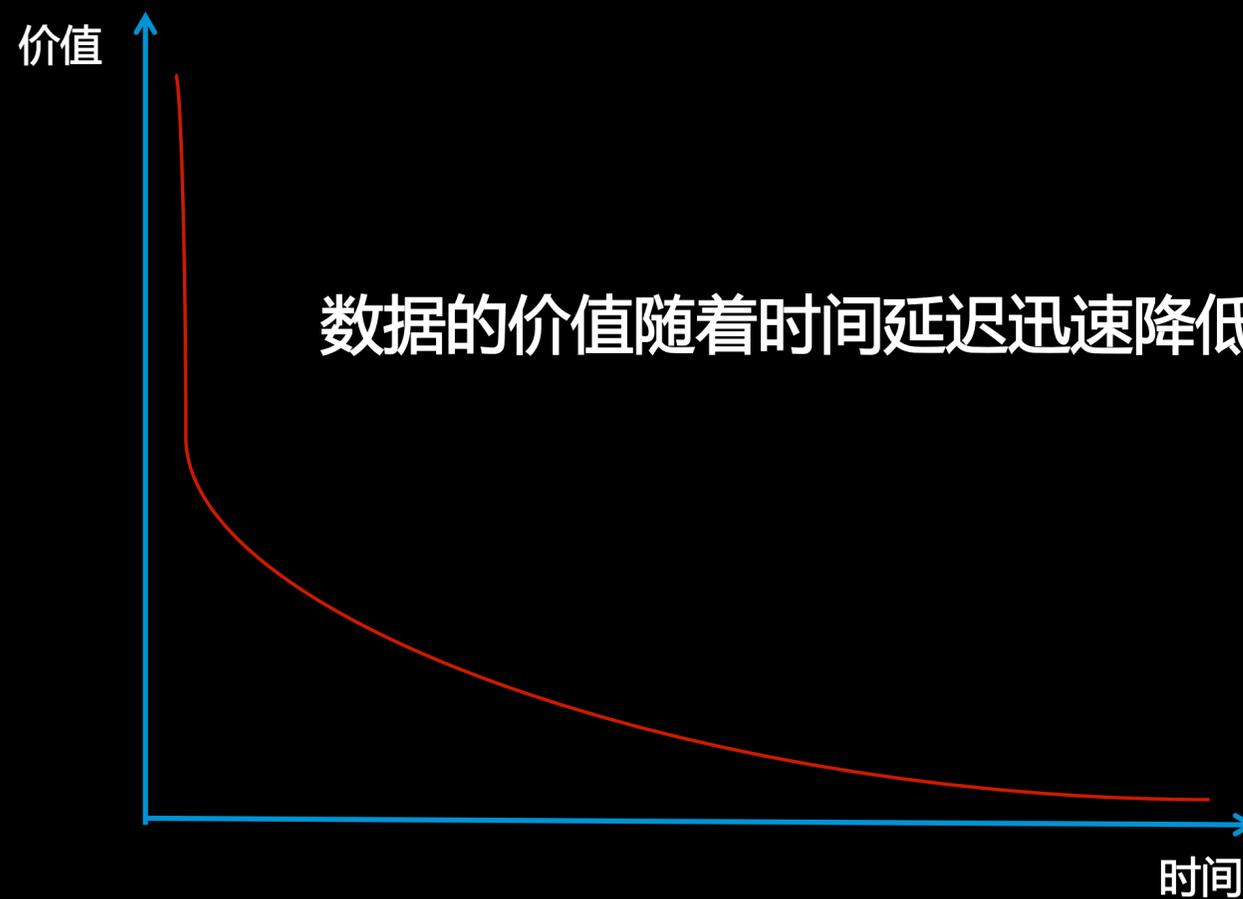
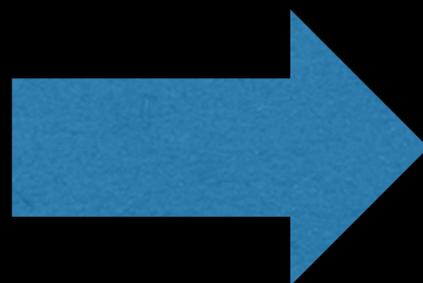
**3**  
如何用流计算



商业和数据形成闭环



业务数据



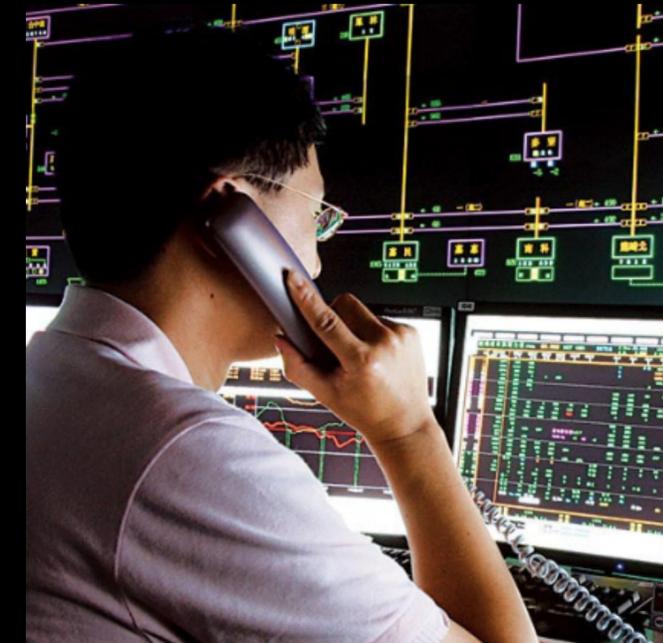
价值时效



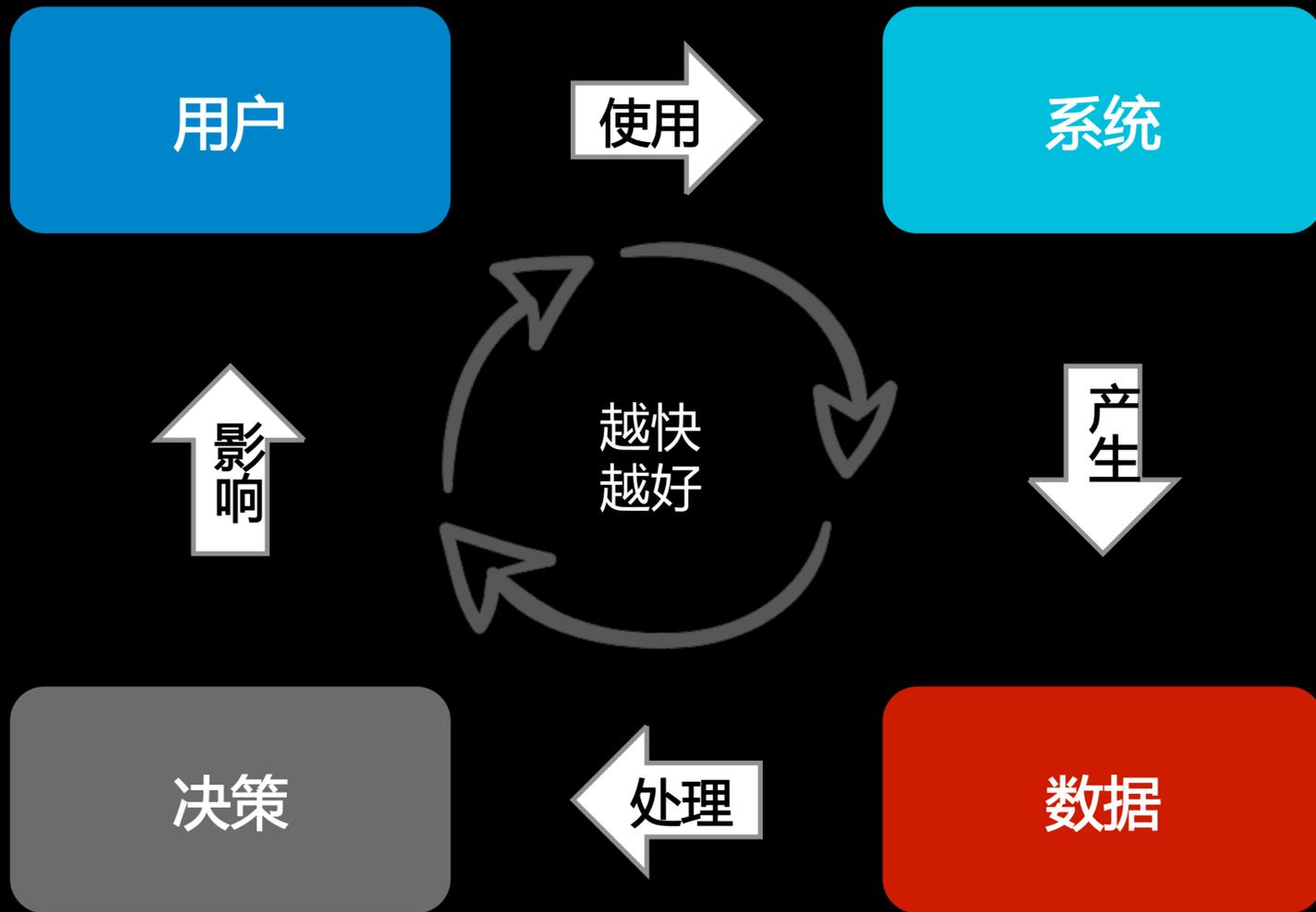
实时推荐



异常&欺诈检测



实时调度



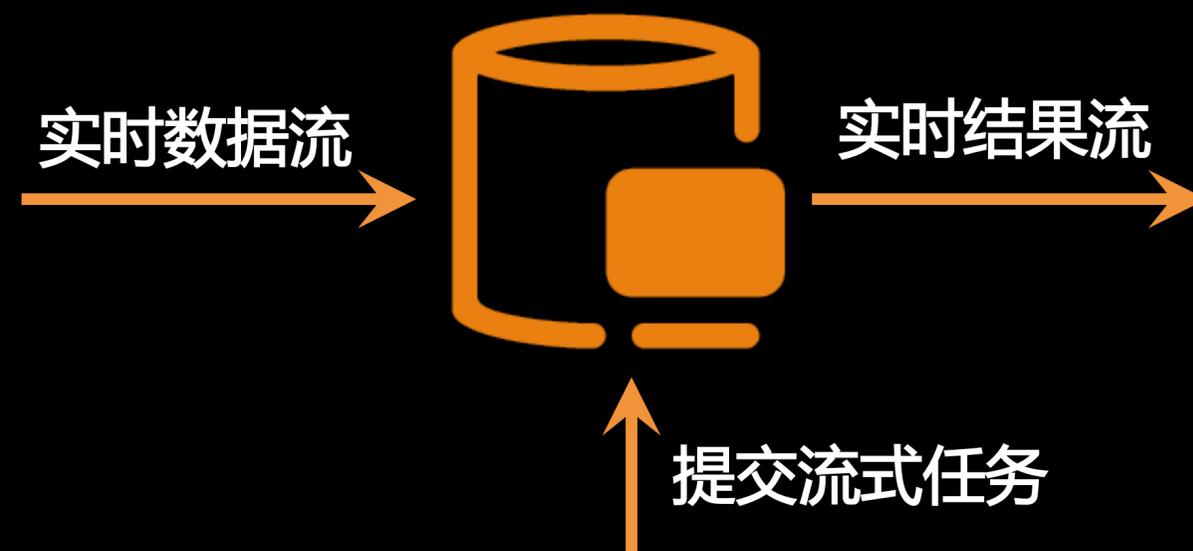
越快越有竞争优势  
大数据实时化——流式处理

### 离线（批量）计算



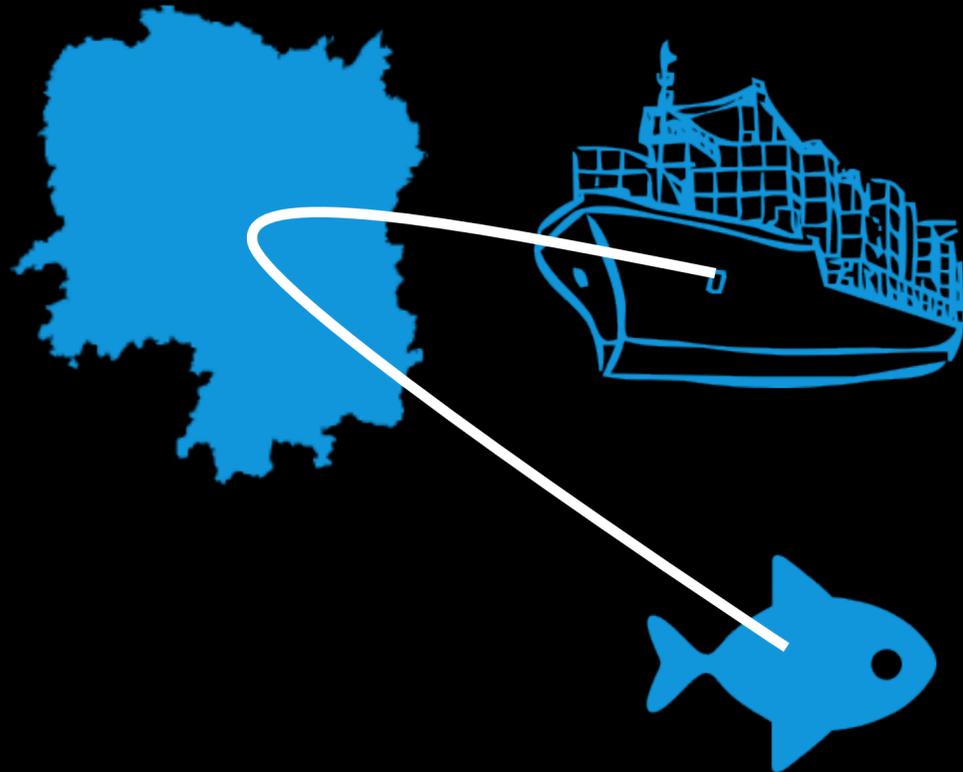
批量计算是一种批量、高时延、主动发起的计算任务

### 流计算



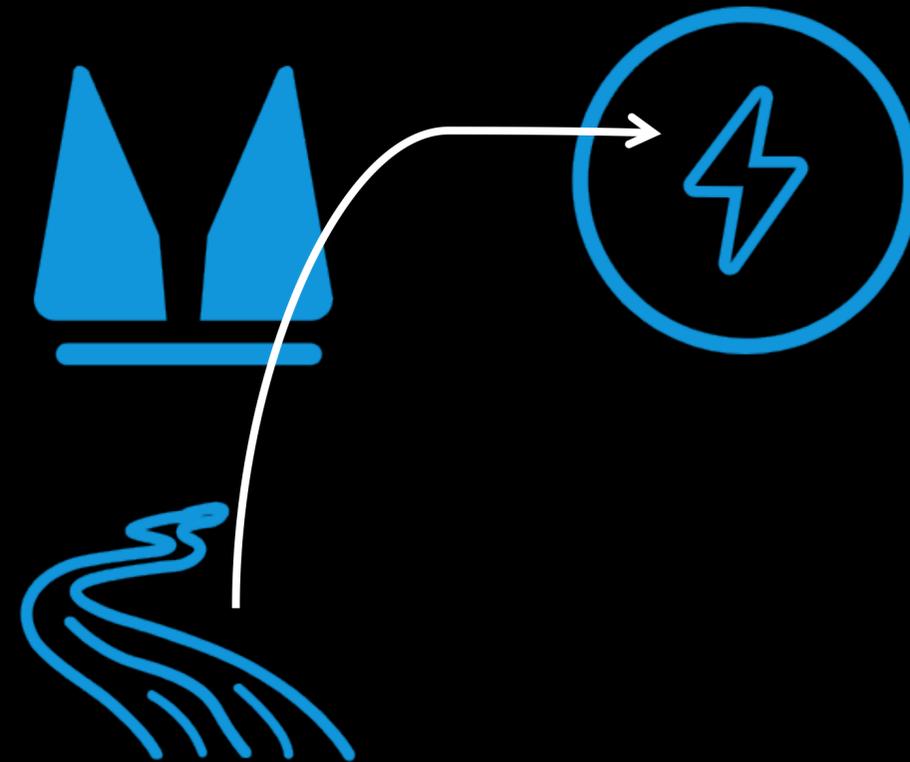
流计算是一种持续、低时延、事件触发的计算任务

离线（批量）计算



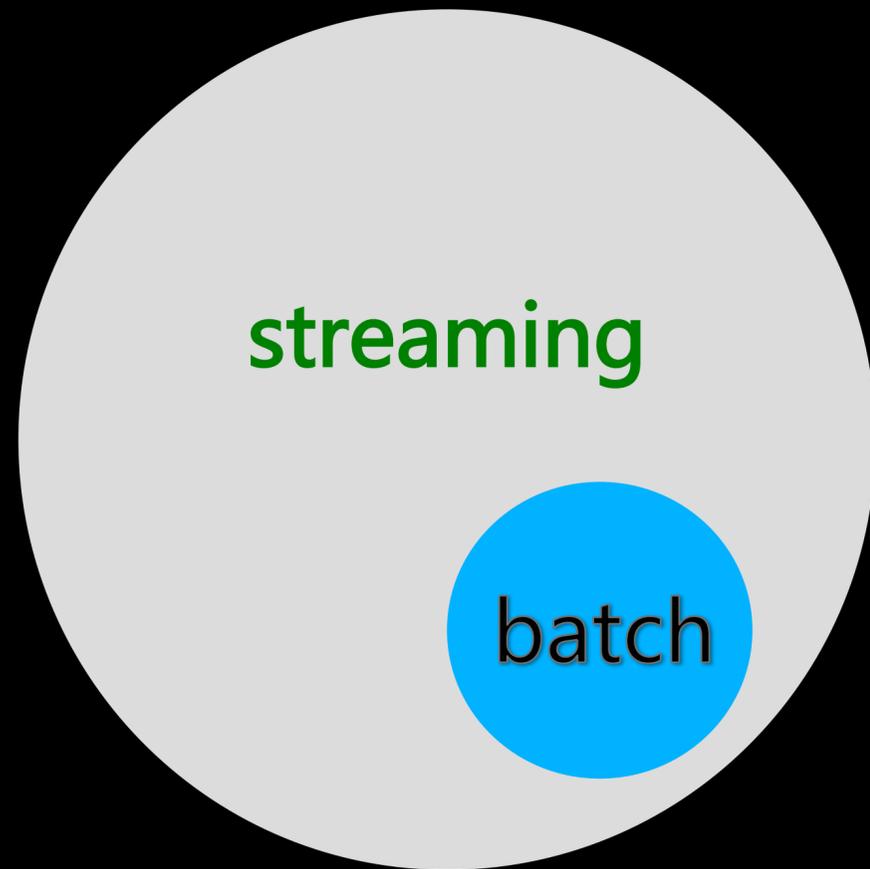
开船去湖里打鱼

流计算

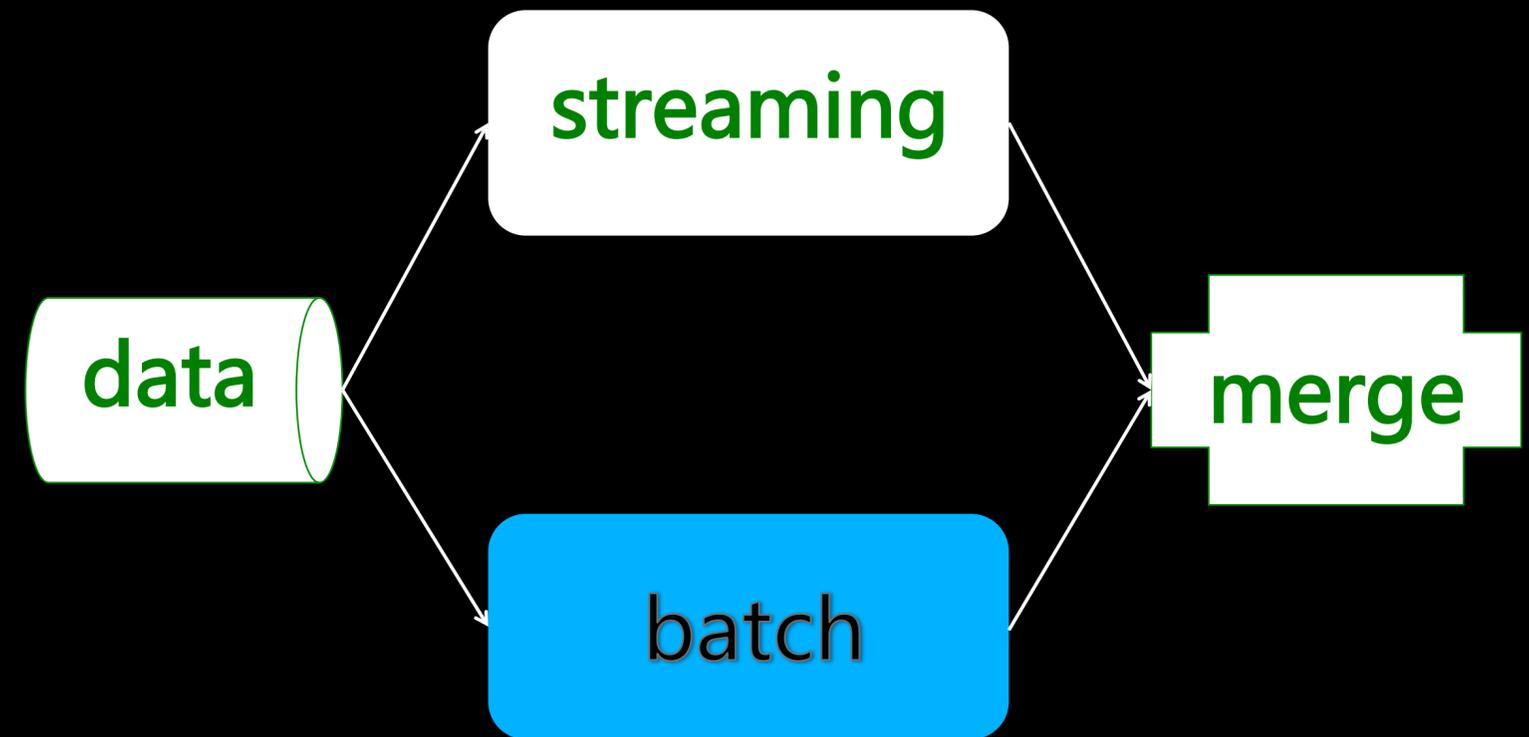


拦河建坝发电

## 理论



## 实践



1  
为什么要用流计算

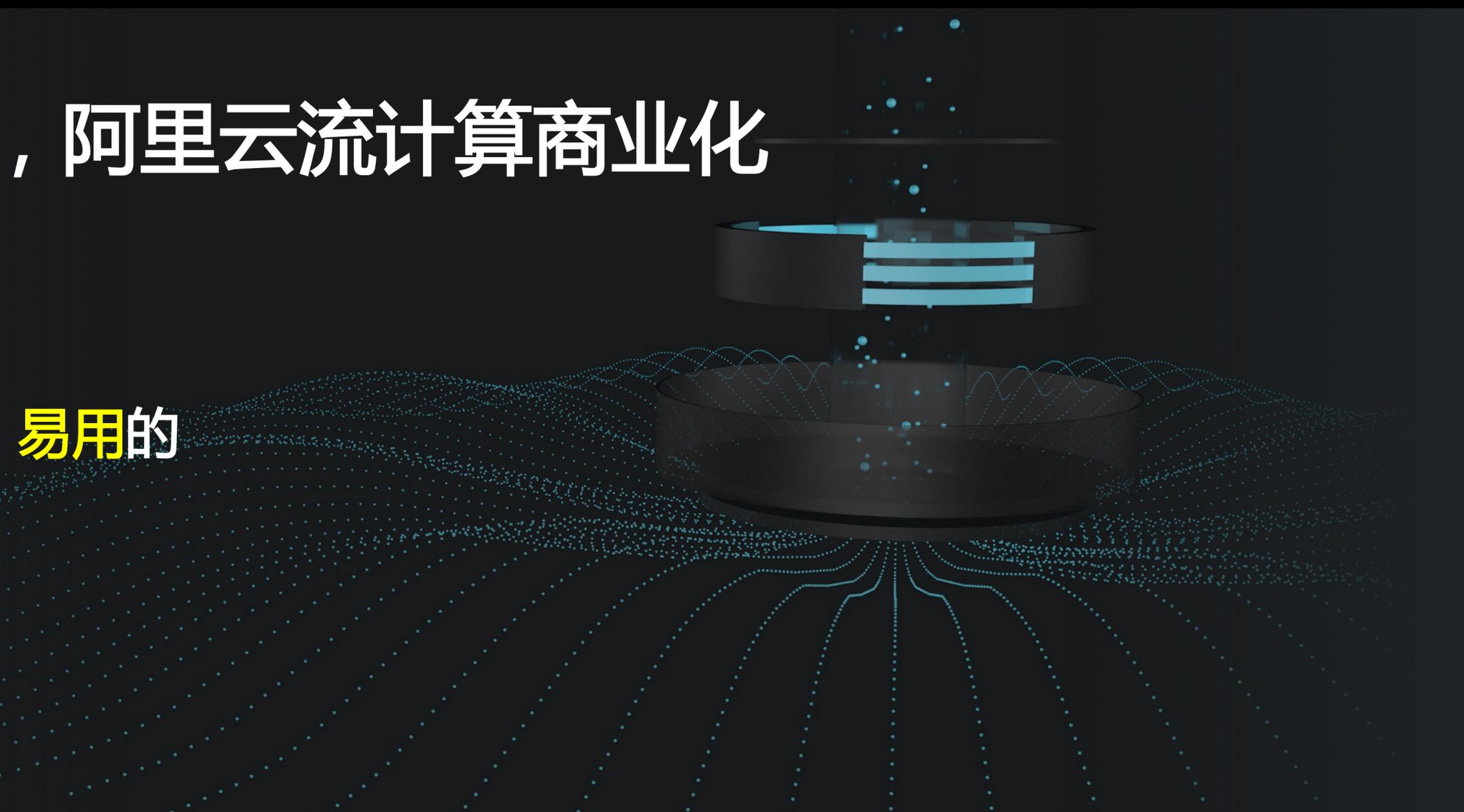
2  
为什么要用阿里云流计算

3  
如何用流计算

# 2018年3月21日，阿里云流计算商业化

2016年9月开始公测

**一站式、高性能、稳定、易用的**  
流式大数据处理平台



# 阿里云流计算



SQL

快速上手



一站式平台

全链路、全周期、全托管



数据生态

对接阿里云各种数据存储



Blink

高性能、稳定



阿里流计算开发平台 - STREAM x

安全 | <https://stream.console.aliyun.com/zh/light/#/workspace/k63dw3ofhtpd9l8rhs9cqgcg/operate/ecf9mns3ysyucrs3vpire16a/dashboard>

阿里流计算开发平台 fukong\_test **一站式平台** 总览 开发 运维 用户: fukong\_g

作业列表 搜索作业名称... [作业运维 / test\\_11\\_stream\\_join](#) | 运行 | 启动位点: 2018-03-08 09:43 | 业务延时: 0 s [暂停](#) | [停止](#) | [监控](#)

作业名称	运行状态	运行信息	数据曲线	FailOver	CheckPoints	JobManager	TaskExecutor	血缘关系	属性参数
test_11_stream_join	运行	Task状态	创建: 0	运行: 7	失败: 0	完成: 0	调度: 0	取消中: 0	已取消: 0
test_04	停止	计算耗时	输入TPS	输入RPS	输出RPS	输入BPS	消耗CU	启动时间	运行时长
test_06	停止	7314 ms	0 Block/s	0 条/s	0 条/s	0 B/s	1.34 CU	2018-04-17 07:32:38	42天 6小时 21分钟 2秒
test_07udf	停止	Vertex拓扑 <a href="#">切换视图</a>   <a href="#">列表模式</a>   <a href="#">放大</a>   <a href="#">缩小</a>							

ID	Name	Status	InQ max	OutQ max	RecCnt sum	SendCnt sum	TPS sum	Delay	Task
0	Source: DataHubTableSource...	RUNNING	0 (0.00%)	0 (0.00%)	0	0	0	0	2
1	Source: DataHubTableSource...	RUNNING	0 (0.00%)	0 (0.00%)	0	0	0	0	2
2	InnerJoin(where: =(a, a0)), joi...	RUNNING	0 (0.00%)	0 (0.00%)	0	0	0	0	2

• 开发、调试、运维、报警

• 数据source、sink、维表管理

• 完善的支持团队：业务问题排查、引擎特性修改

# 数据生态

无缝对接云上**10+**种数据存储

- IoT Hub
- Data Hub
- SLS
- MQ
- ADS
- RDS
- HybridDB
- HiTSDB
- ElasticSearch
- TableStore



# Blink引擎

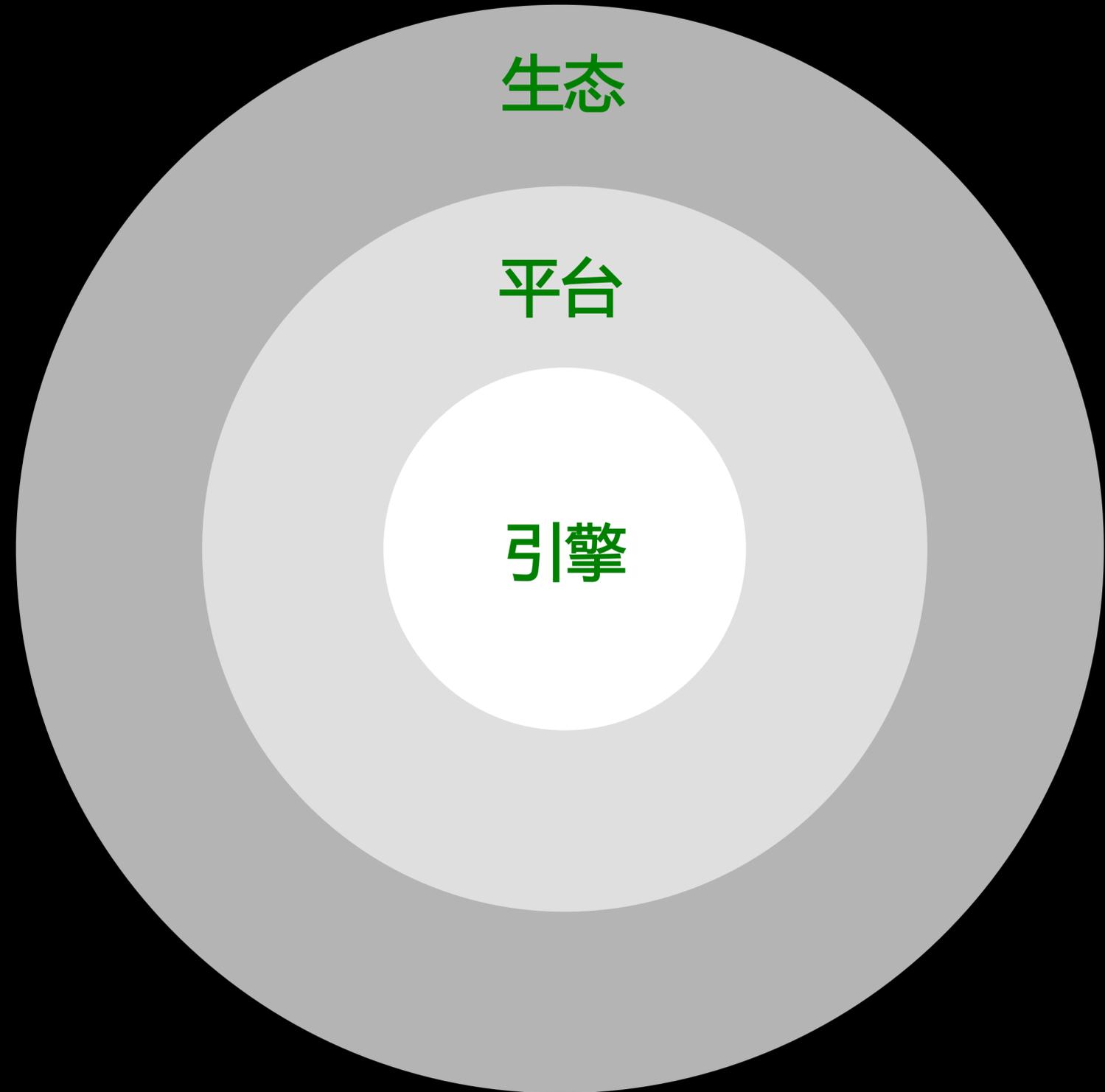
## Flink的企业版

二级调度

增量checkpoint

异步IO

关键部分性能提高**10倍**



# Why Blink ?

阿里巴巴对大数据处理的思考



趋势



大数据继续发展

批流共存

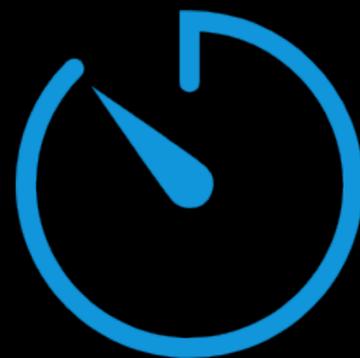
流的比例会逐渐增加



# 阿里巴巴对Blink的思考



引擎



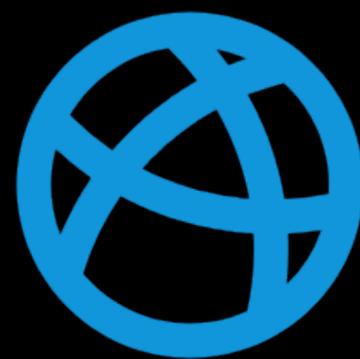
低延迟



exactly-once



大规模



复杂计算



批流统一



	低延迟	exactly-once	大规模	复杂计算	批流统一
Spark	NO				
Kafka Stream			No	No	No
Storm		No			No
Flink					

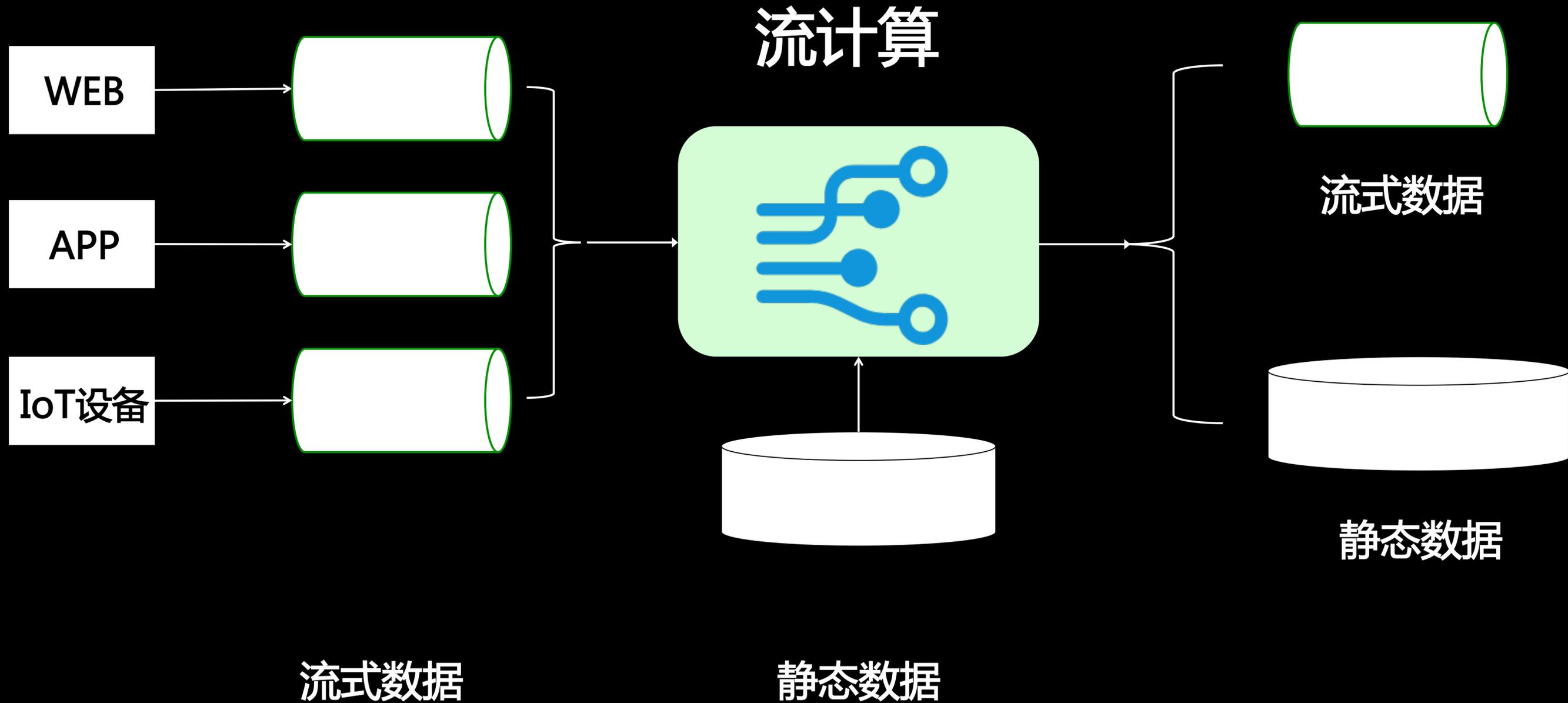
	Apache Storm	Spark Streaming	Flink
Model	Native	Micro-Batch	Native
API	Low Level	High Level	High Level
状态	No	Yes	Yes
准确性	At-Least-Once	Exactly-once	Exactly-once
窗口	No	Limited	Yes
时延	Low	High	Low
吞吐	Low	High	High
SQL	No	Yes	Yes

1  
为什么要用流计算

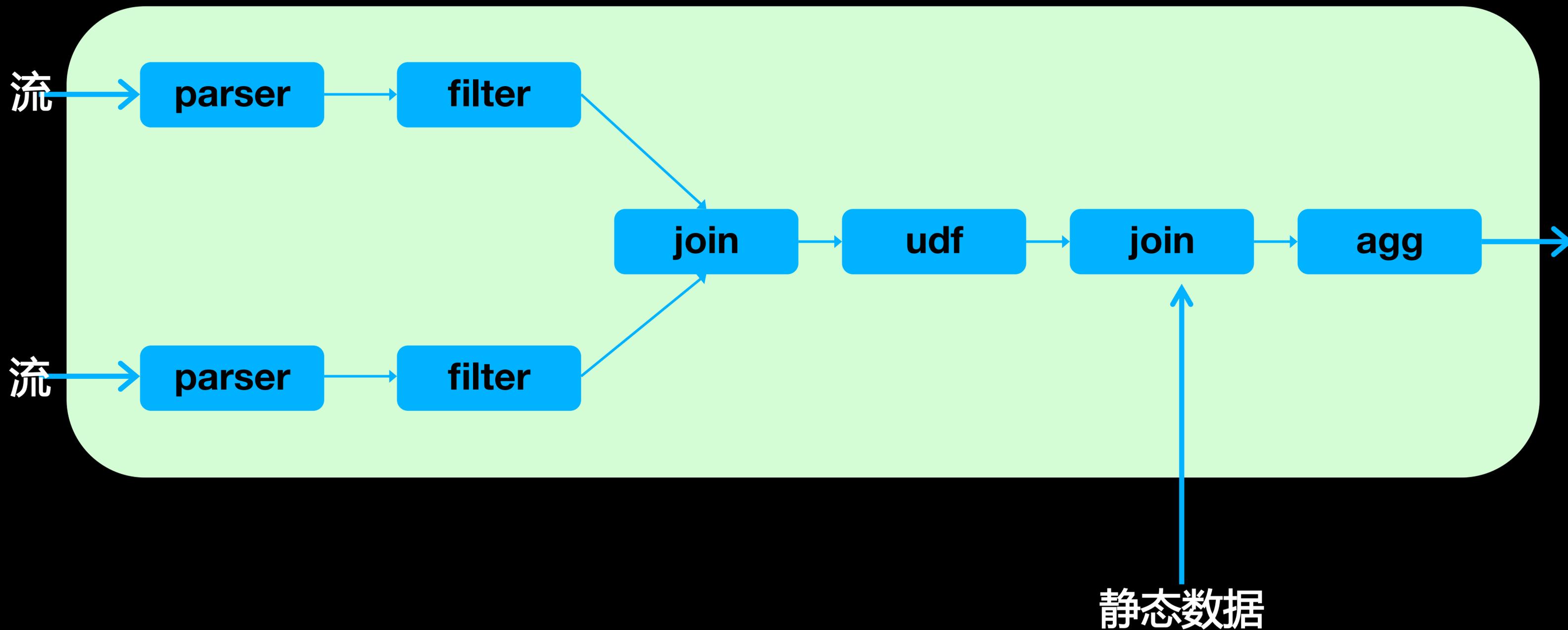
2  
为什么要用阿里云流计算

3  
如何用流计算

# 典型架构



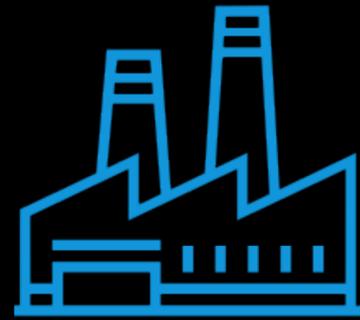
# 流计算



# 典型场景



实时推荐



工业IoT



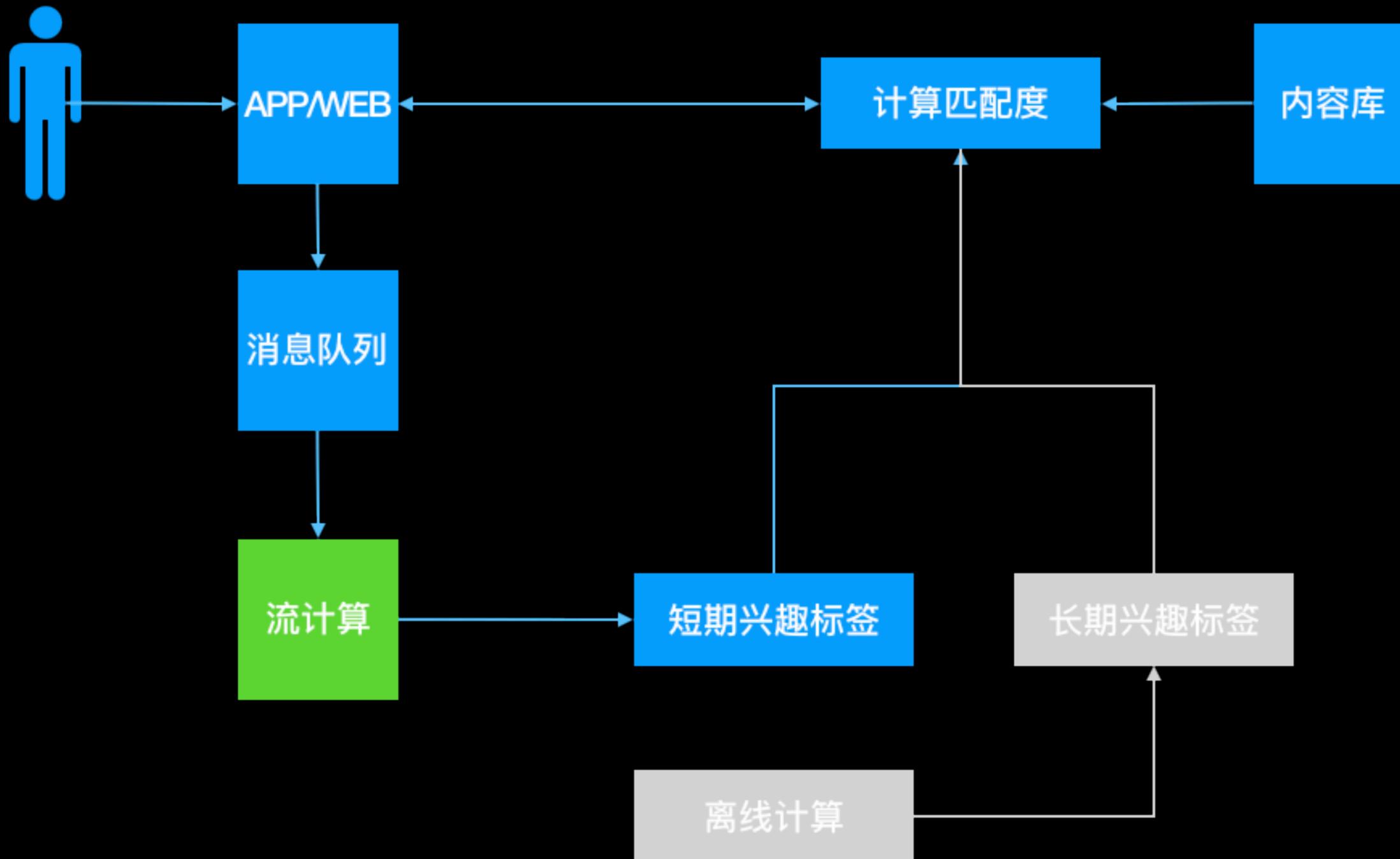
欺诈检测

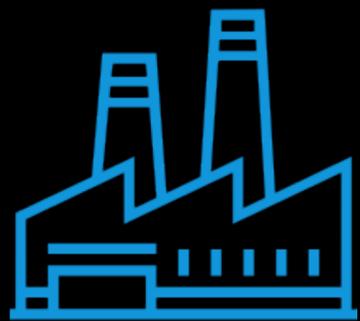


实时报表

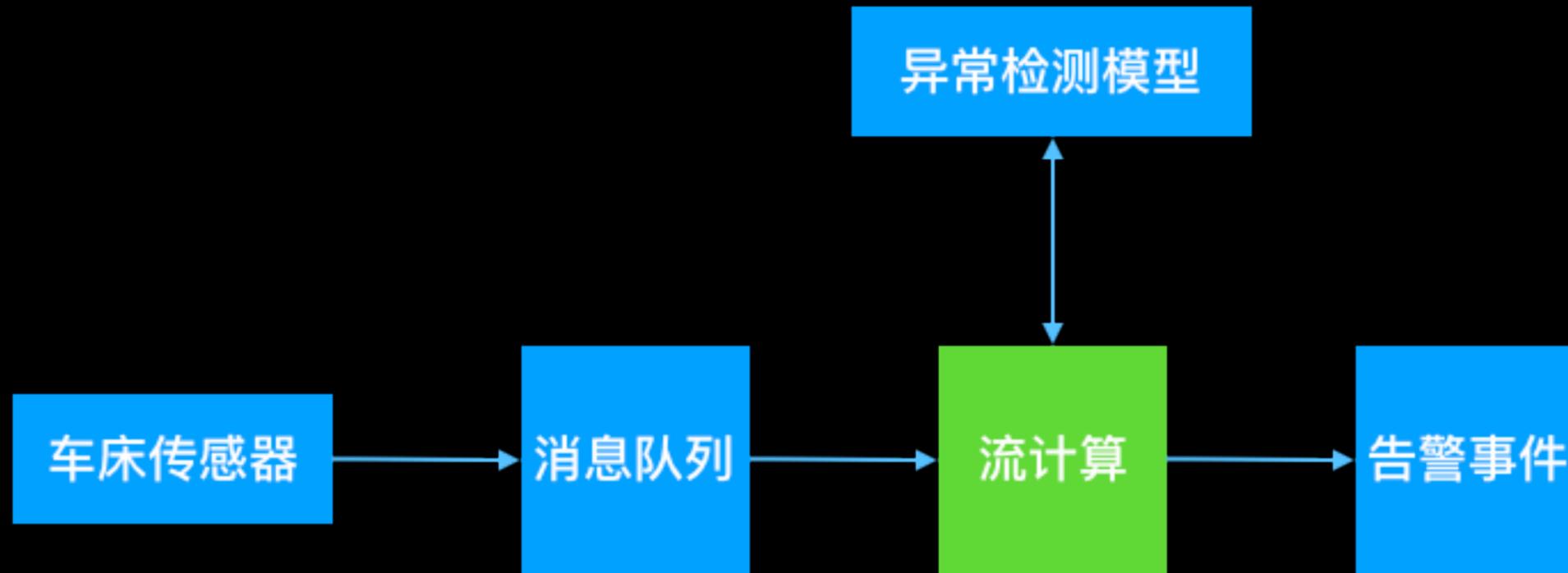


实时推荐



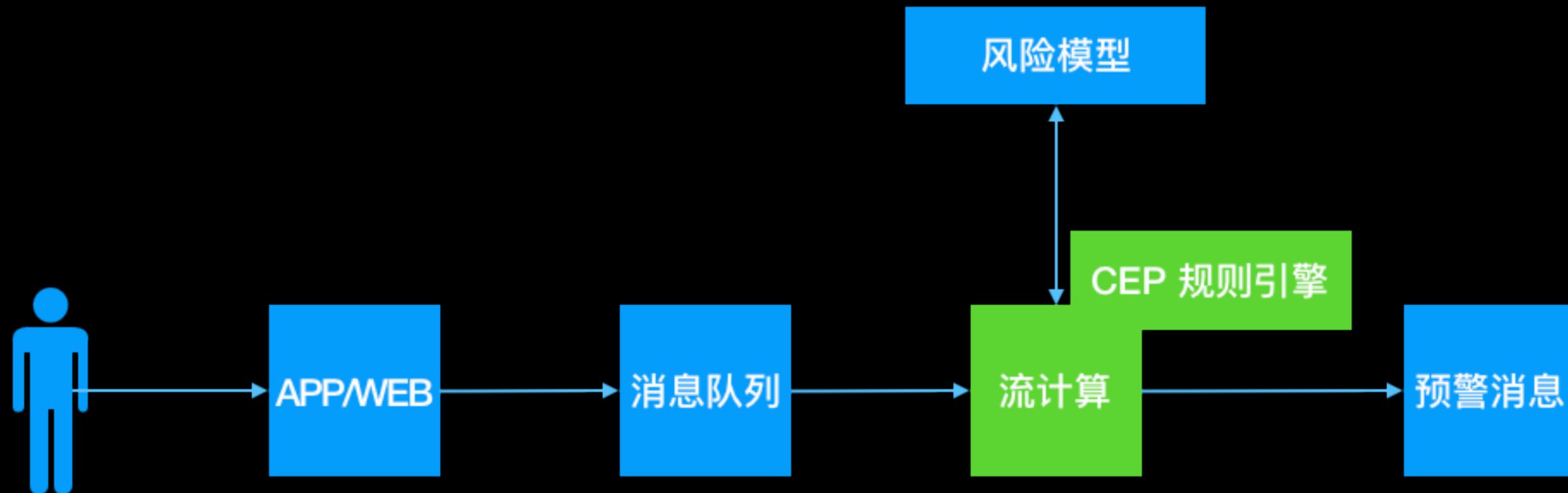


工业IoT



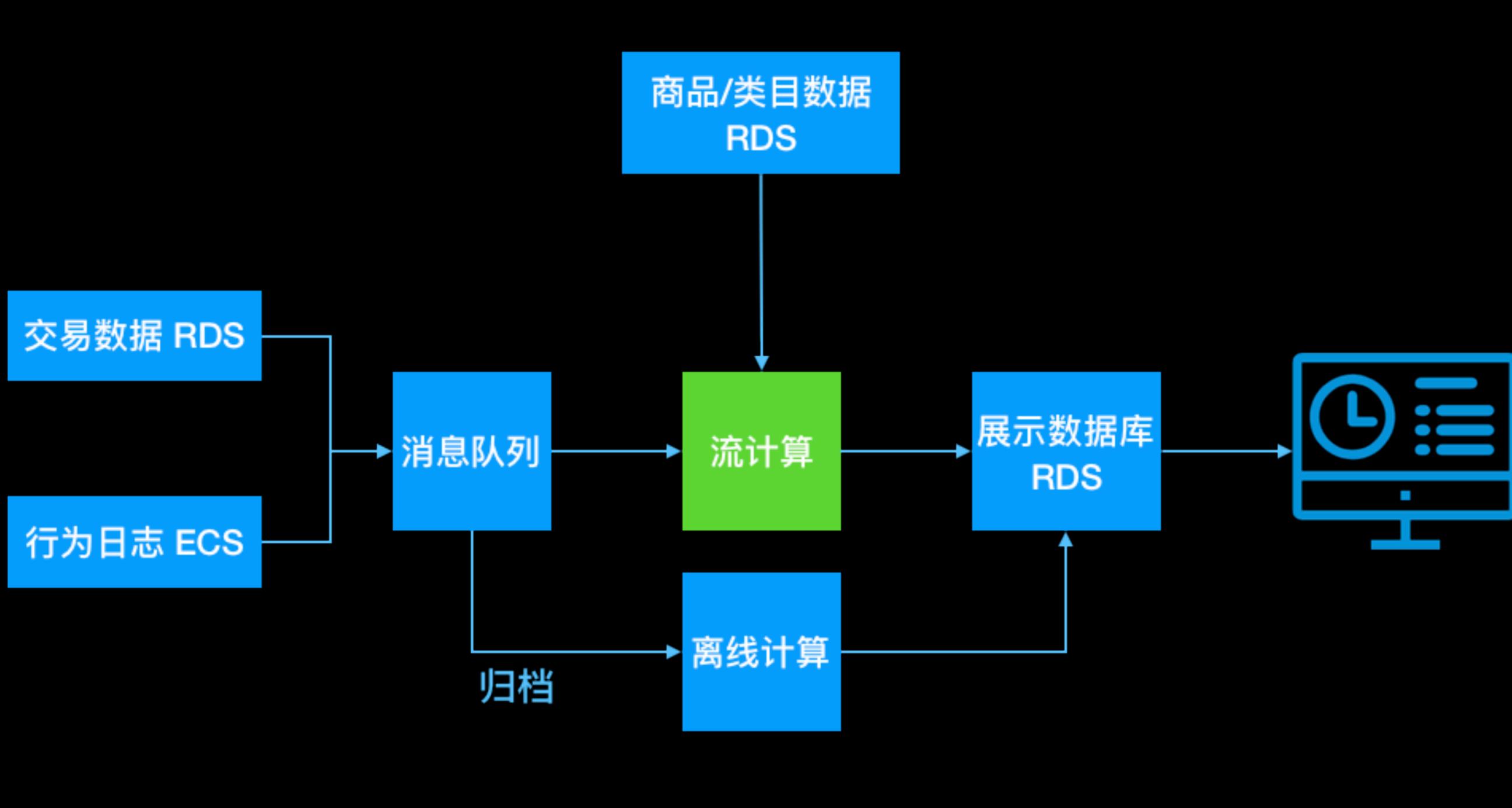


欺诈检测





实时报表



# 典型案例



天猫双十一大屏



城市大脑



工业大脑

# 我们的用户

互联网



金融



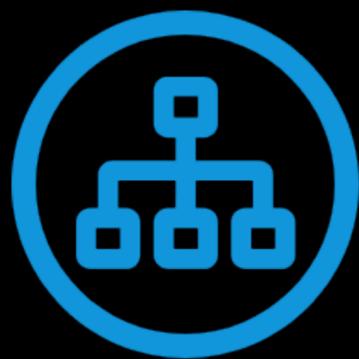
IoT



国企/政府



# 新功能速览



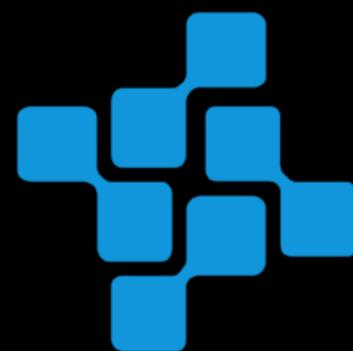
独享集群

UDX、VPC、GPU、FPGA



Datalake ETL

数据清洗、同步、分析



CEP

流式复杂事件处理

6月7日邀测

# 新功能速览

复杂事件处理 事件检测、异常检测

- 流式数据
- 跨多条记录
- 规则&模式匹配



CEP

数据流



模式



事件

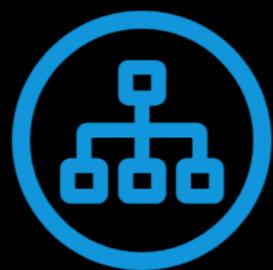
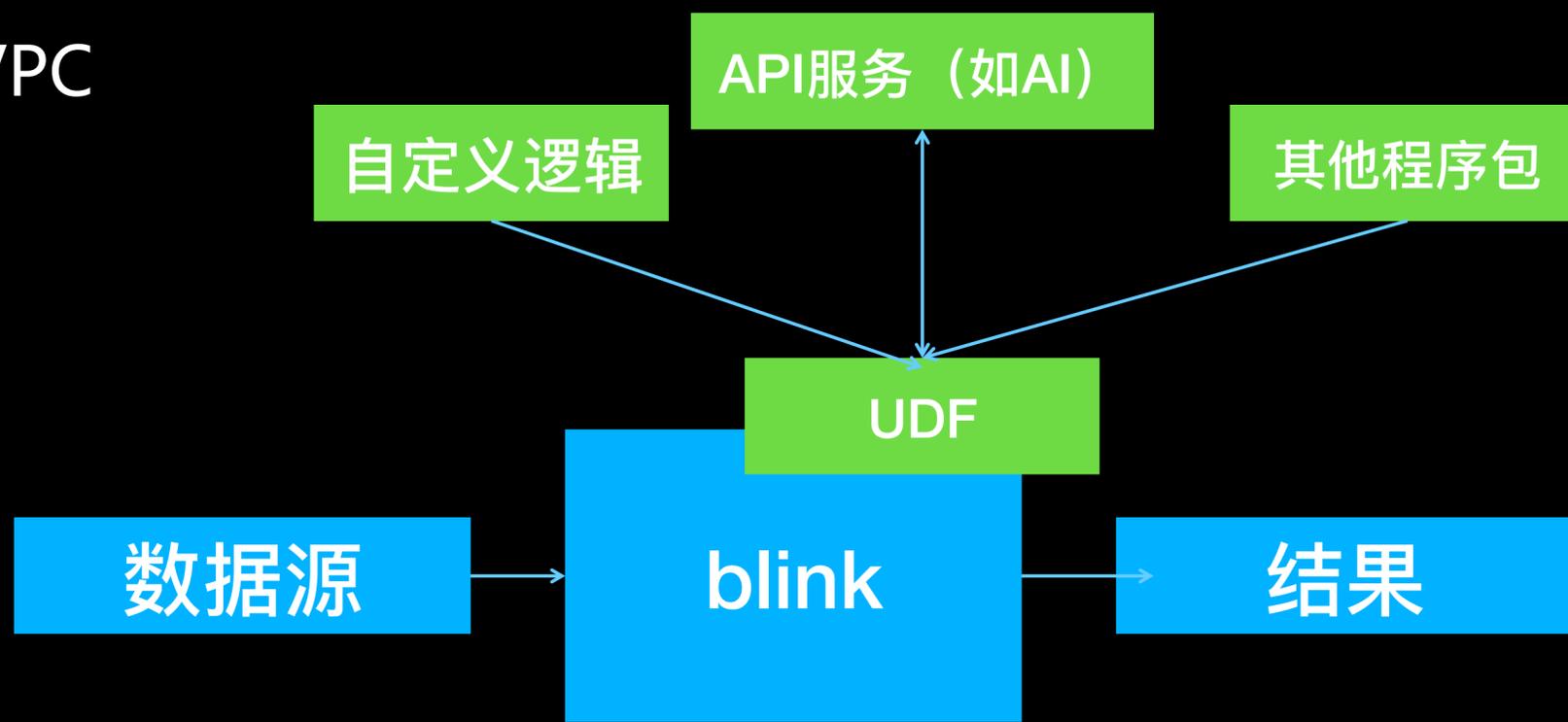


# 新功能速览

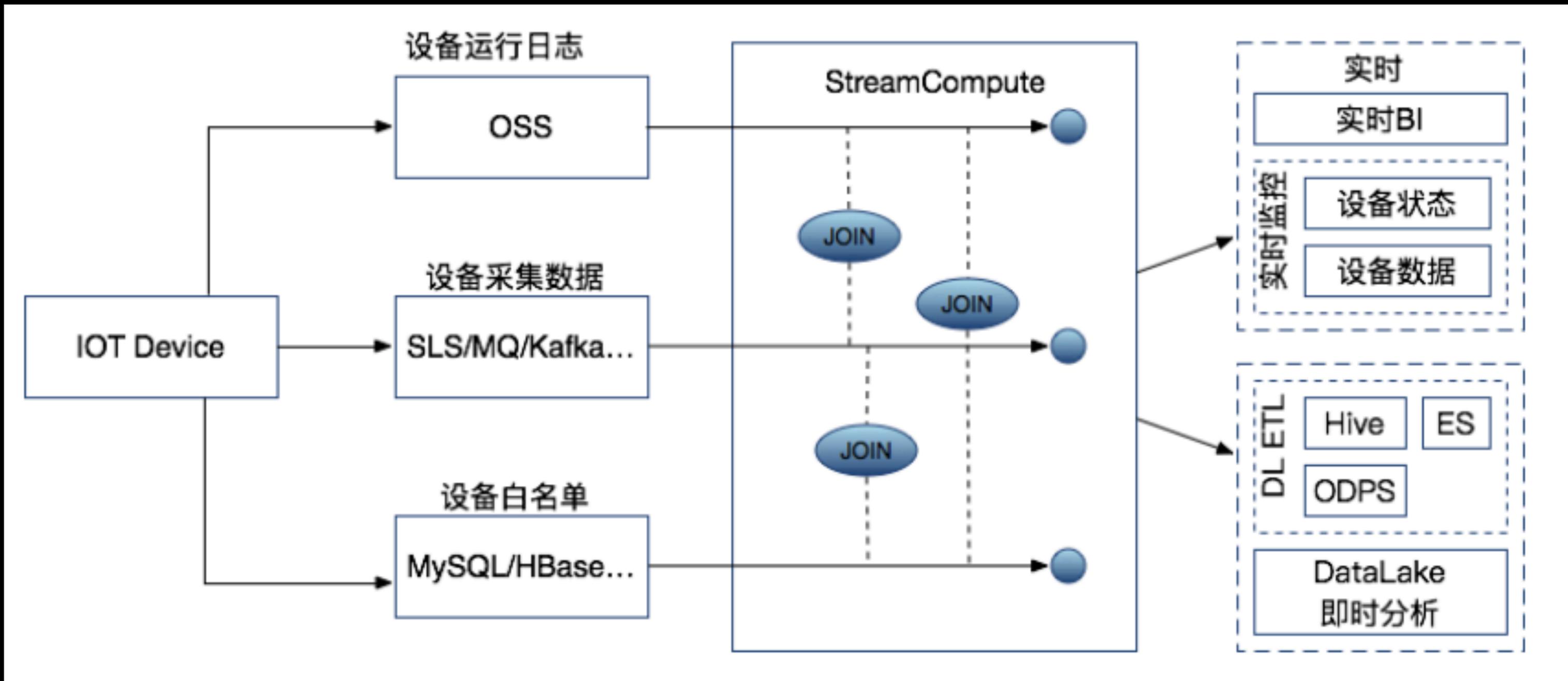
## 小集群 blink on ecs

更开放，适用场景更多

- UDF
- FPGA&GPU
- VPC



小集群



# DT时代，数据是新能源 流计算

## 让业务实时，让数据发电

