



ClickHouse 与流计算

2024年7月20日



Jove Zhong
Co-Founder, Timeplus

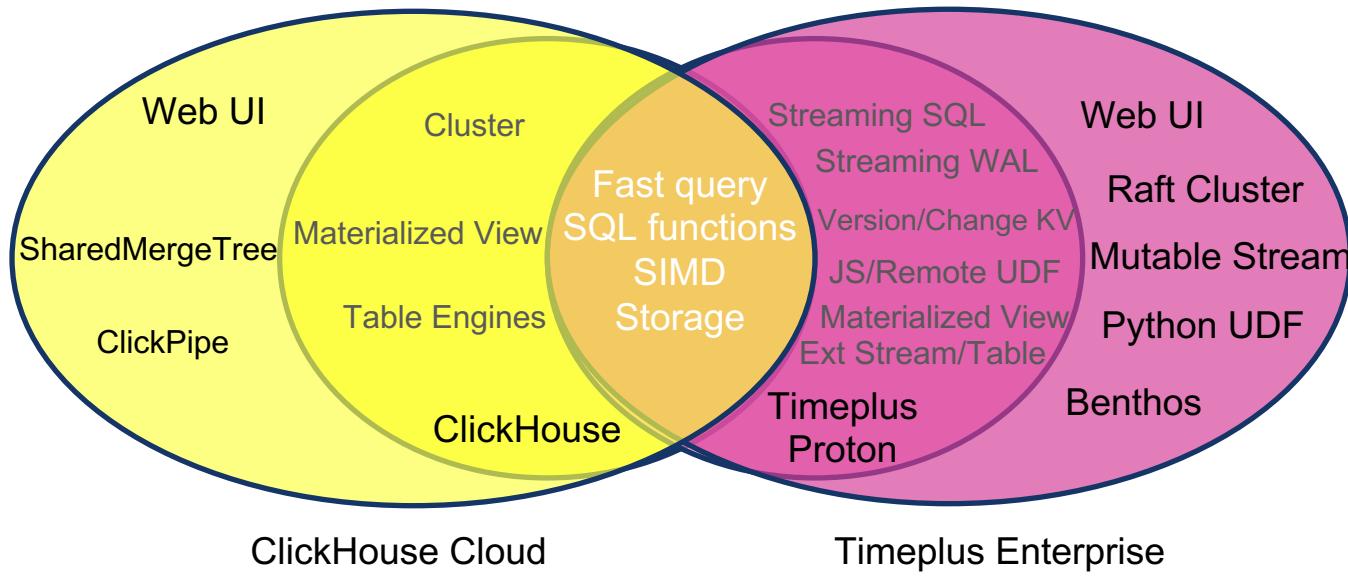


Timeplus 公众号

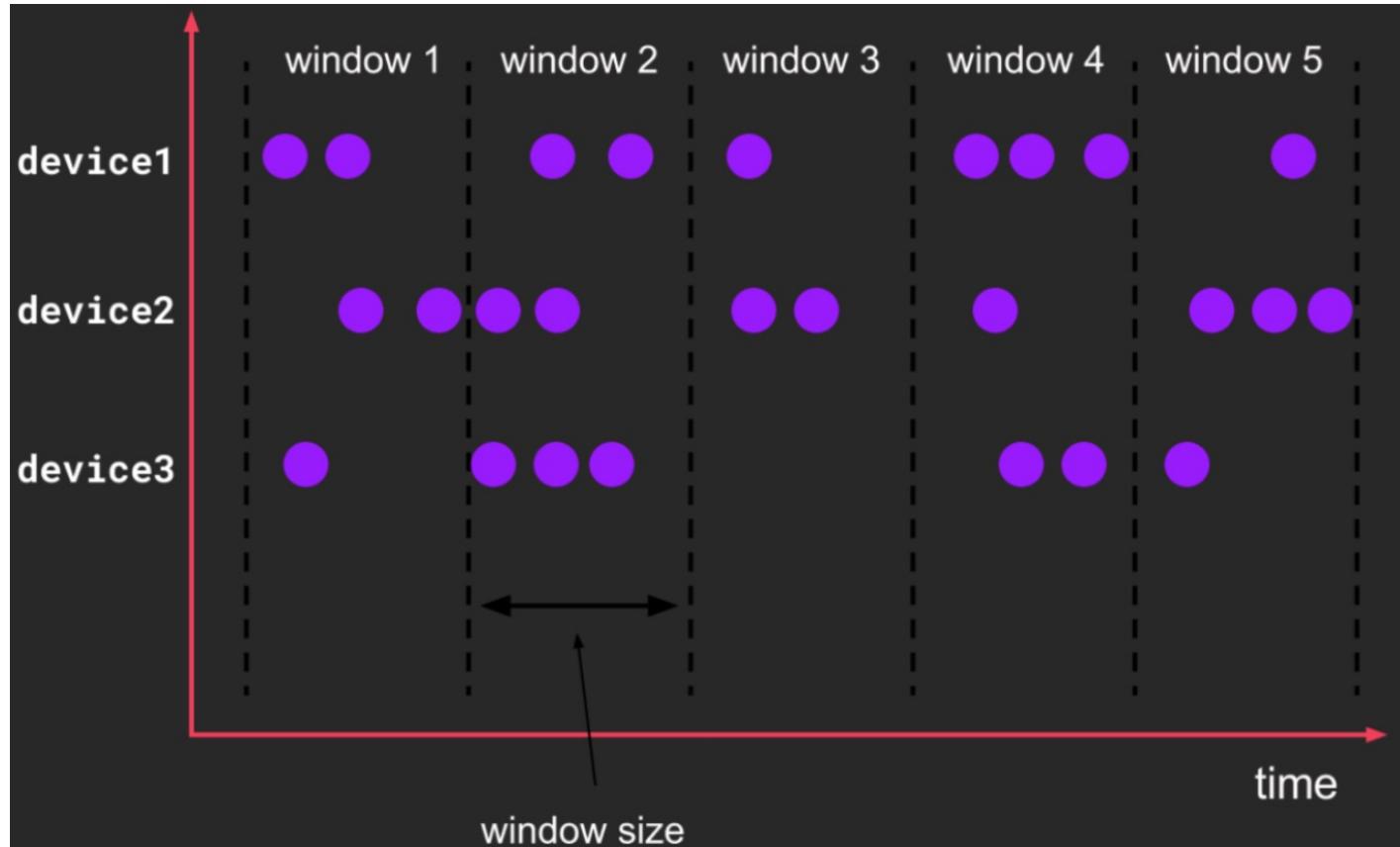


GitHub 开源

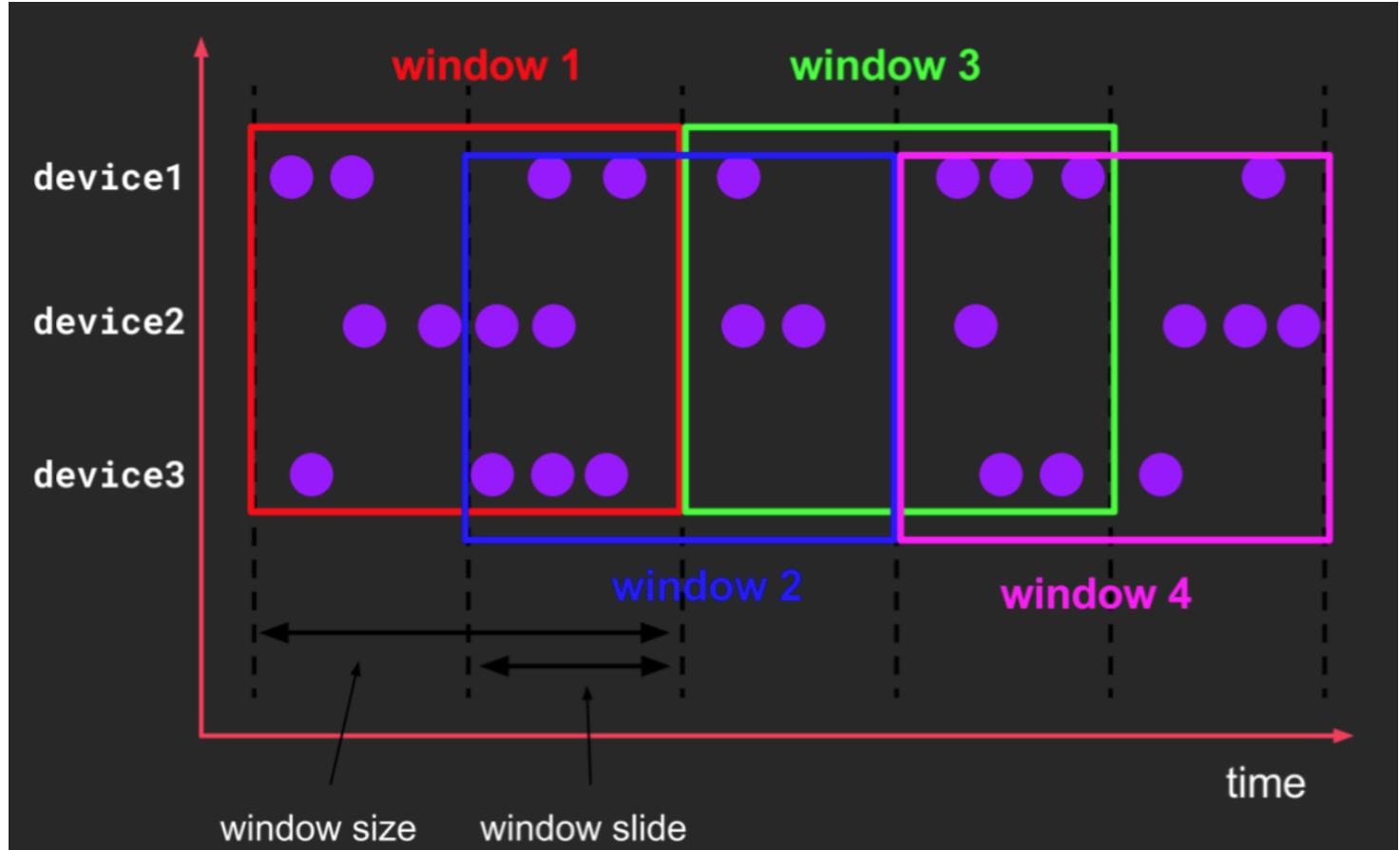
Timeplus ❤️ ClickHouse



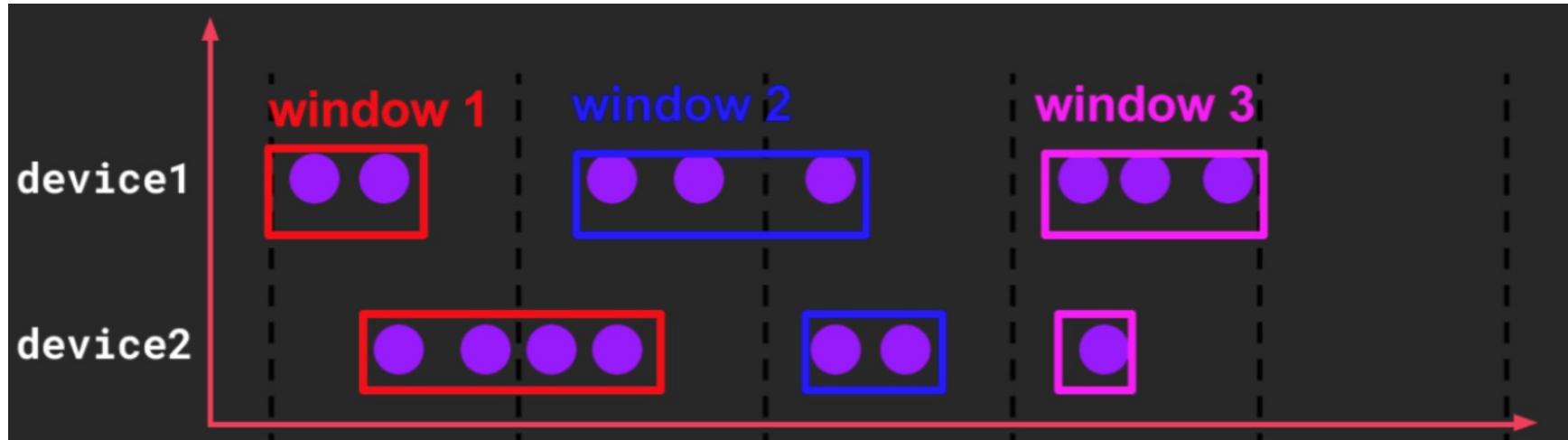
Tumble Window 固定间隔窗口



Hopping Window 滑动窗口

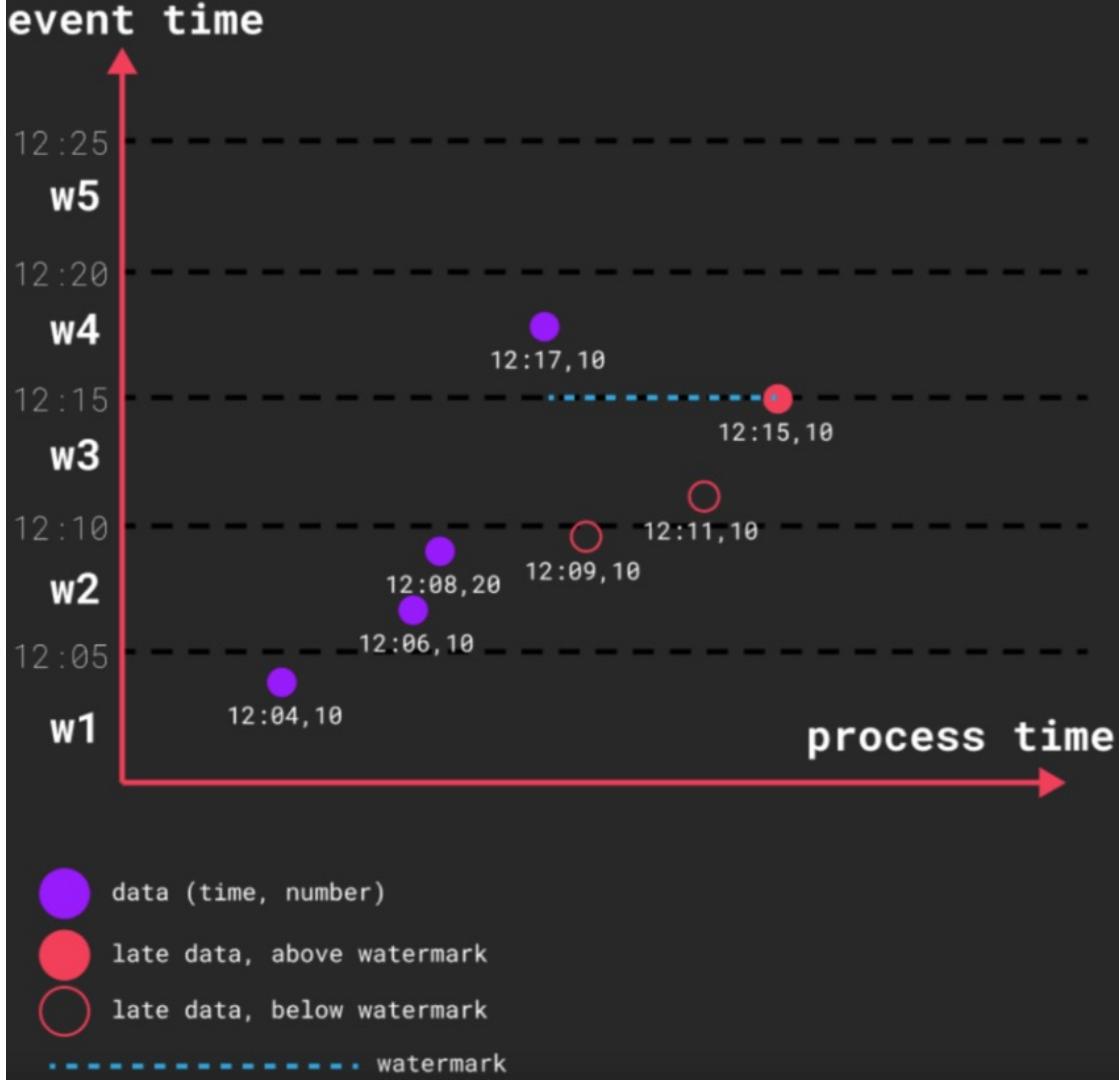


Session Window 变长会话窗口



Watermark

利用水位线来判断
迟到时间
选择性放弃



来点刺激的



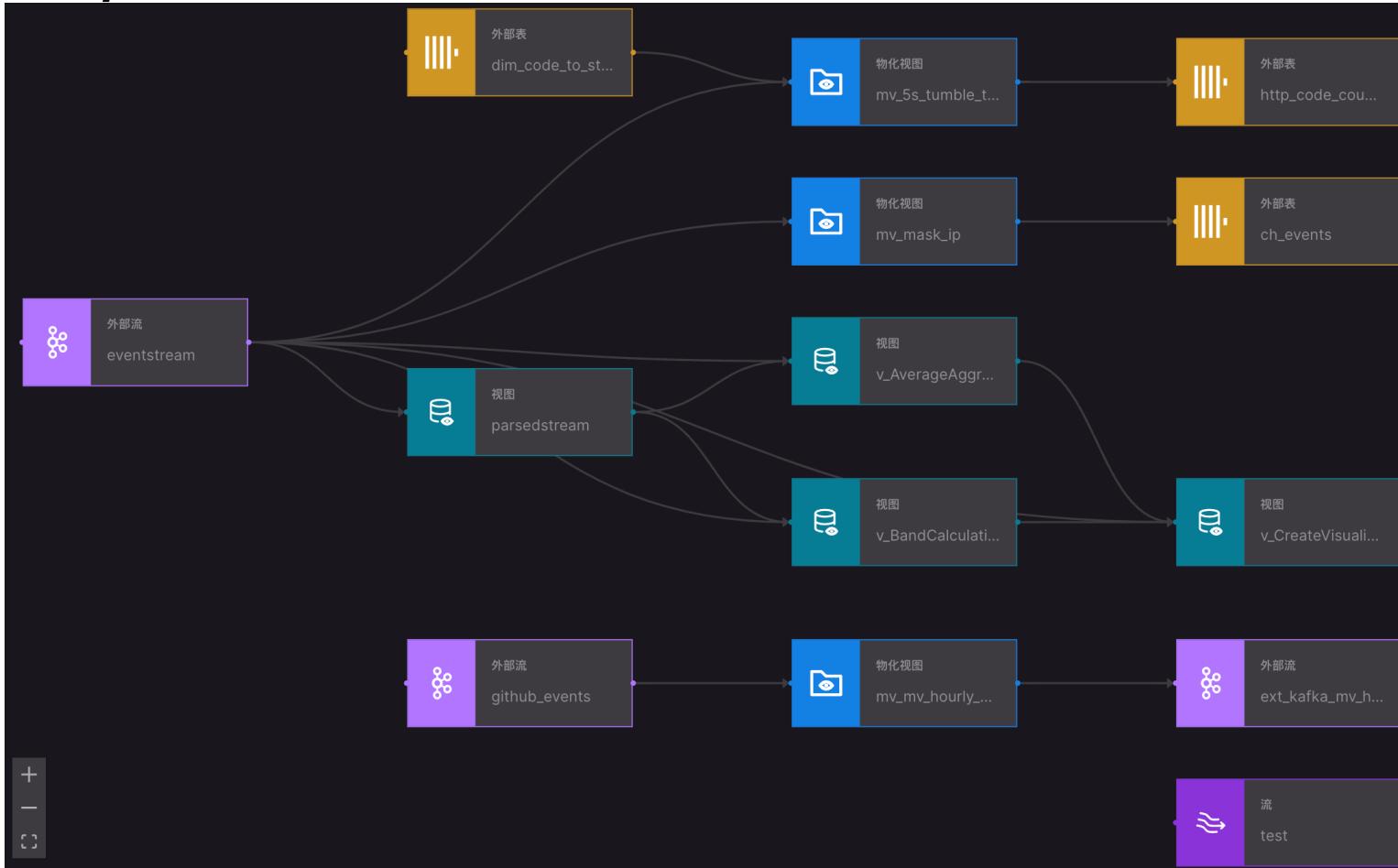
```
CREATE RANDOM STREAM test(
    a float32 default rand(),
    b float32 default rand()%1000/10,
    c string default 'region_'||to_string(rand()%5),
    d int32 default floor(rand_uniform(0,100))
)
SETTINGS eps=10_000_000
;

select count() from test;

select count() from tumble(test,1s) group by window_start;

select c, count() from test group by c;
```

ETL for Kafka/ClickHouse



External Stream -> Materialized View -> External Table

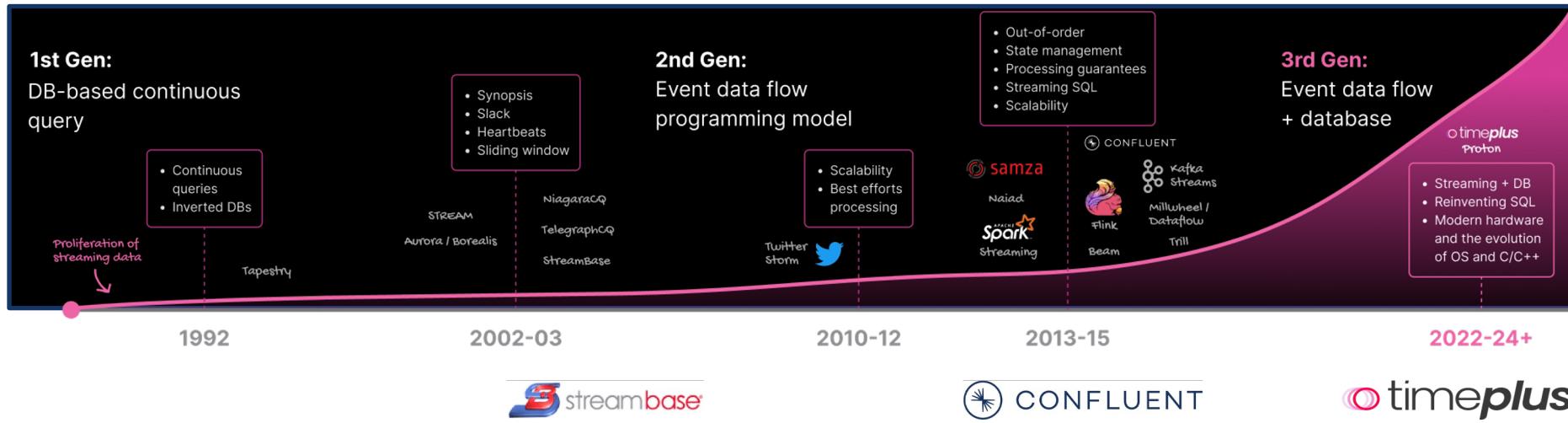
```
-- read the topic via an external stream
CREATE EXTERNAL STREAM frontend_events(raw string)
    SETTINGS type='kafka',
              brokers='redpanda:9092',
              topic='owlshop-frontend-events';
```

```
-- clickhouse table
CREATE TABLE events
(
    _tp_time DateTime64(3),
    url String,
    method String,
    ip String
)
ENGINE=MergeTree()
PRIMARY KEY (_tp_time, url);
```

```
CREATE EXTERNAL TABLE ch_local
SETTINGS type='clickhouse',
        address='clickhouse:9000',
        table='events';
```

```
CREATE MATERIALIZED VIEW mv INTO ch_local AS
    SELECT now64() AS _tp_time,
           raw:requestedUrl AS url,
           raw:method AS method,
           lower(hex(md5(raw:ipAddress))) AS ip
    FROM frontend_events;
```

实时分析技术演进 - 复杂场景推动流式数据融合处理



Gartner: 统一(融合)平台是一种相对较新的基础设施软件，支持需要处理动态流数据和静态历史数据的运营或分析业务应用程序,

实时数据融合处理引擎&平台

驱动于日益增长的复杂实时流式数据处理场景需求



Web UI Console

分布式集群

Timeplus Proton 引擎(单一二进制文件)

统一SQL

API & 客户端

内置UDF引擎

Python

JavaScript

内置高速热数据存储/缓存 + 热数据极速OLAP查询分析

亚秒级流式数据处理

数据驱动实时运营支持

实时数据管道

实时流式计算

实时事件处理

实时统计分析



1. 基于C++的**亚秒级分布式流式数据处理**，行处理和Mini-Batch处理双模自适应，向量化流式数据处理
2. 统一SQL，流式数据关联热数据分析建模
3. 基于UDF的SQL自由拓展，Remote UDF, JavaScript UDF, Python UDF
4. 灵活部署模式，同一核心引擎多角色部署

Timeplus与Flink对比：独创的核心技术



Target User Cases	Analysts/Data Engineer/Developer	Data Application Developer
Runtime	Native (C++)	Java/JVM
Stream SQL	Yes	Yes
Kernel Level IO Scheduling Optimization	IO Uring	No
Exact-Once Processing	Yes(checkpoint)	Yes(checkpoint)
Built-in stream storage	Support	No
Vectorized Stream Computation	Yes	No
Built-in persistent Storage (历史存储)	Included	No
Built-in historic data query optimization (JIT, Vectorized Query Execution, built-in compression)	Yes	N/A
API	REST/Python/Java	Java/Scala/Python
Extensibility	UDF/UDAF	Java Code
Data development interaction	Console UI, WYSIWYG	CLI

Timeplus的创新技术优势



融合 海量数据查询与高吞吐量流式计算于一体

- 独特的向量化流式数据计算技术
- 基于列存及向量化计算的高性能历史存储及计算
- 基于内存计算以及TFF的新一代高性能内建流式数据融合历史数据关联计算技术
- 数仓级SQL，强大的数仓挖掘分析能力



数据实时可修改以及动态实时更新语义

- 不仅提供Append-only的数据流的支持
 - 同时提供基于主键的数据更改的语义
 - 能够支持复杂的流式数据分析功能
- 提供基于JSON和正则表达式的读时建模，无需在写入数据时候的固定模型，极大的提供了分析的灵活性，能够更容易的应对变化的业务数据和规则



端到端的性能卓越的计算引擎 - 极速响应&低延迟

- 基于增量计算模型
- Zero-拷贝/反序列化
- Single C++ binary设计，充分利用硬件/内核底层的计算能力
- 百万级EPS下亚秒级延迟，十万级EPS下毫秒级延迟

开放 统一的体系

- Streaming SQL为核心：Timeplus和Snowflake, Confluent, Oracle等一起参与SQL标准化定义
- SDK/API为开放系统，允许客户自己选择自己的开发语言，无需创造新的脚本
- Cloud Native：对K8S支持以及Computing/storage的分离和无限扩展性

Timeplus典型应用场景 - 亚秒级实时计算



金融风控

去年的监管要求：
过去3分钟的滑动窗口



新的监管要求：
过去1秒至过去5分钟
共 300 个滑动窗口



Timeplus实时计算 – 金融：券商自营量化交易实时指标服务

Timeplus价值

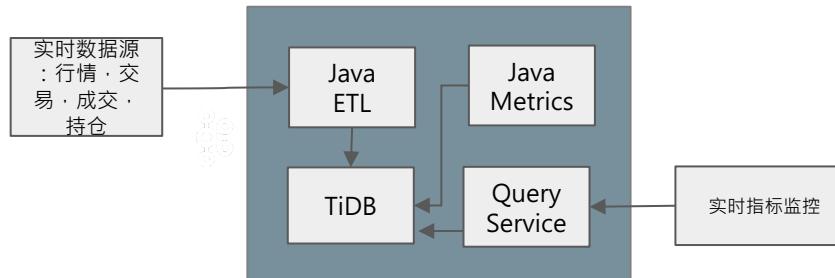
- 十毫秒级实时指标计算性能：亚秒级复杂维度数据推动下的流式数据窗口计算能力显著提升实时指标计算性能 5-10倍
- 集群管理+横向扩展及高可用：Multi-raft集群及能力实现实时指标计算、数据存储及查询的横向扩展及高可用
- 一站式实时数据中心显著提升开发效能：一站式实时数据中心显著提升开发效能：基于Timeplus平台使用统一SQL以及UDF一站式完成实时数据集成，实时数据计算，历史数据回放/回测，交互式查询分析，开发效能提高~5倍

量化交易监控指标平台

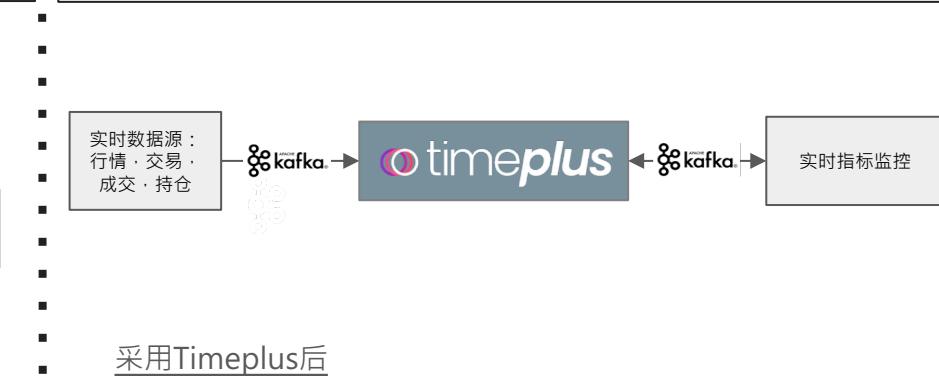
- 300+ 实时量化交易监控指标
- 每日数据吞吐量：100,000 EPS · 5小时每天
- 实时指标类型：交易监控，Alpha视图，链路/系统监控
- 实时指标延迟：十毫秒 ~ 分钟级

实时指标样例：多头收益率

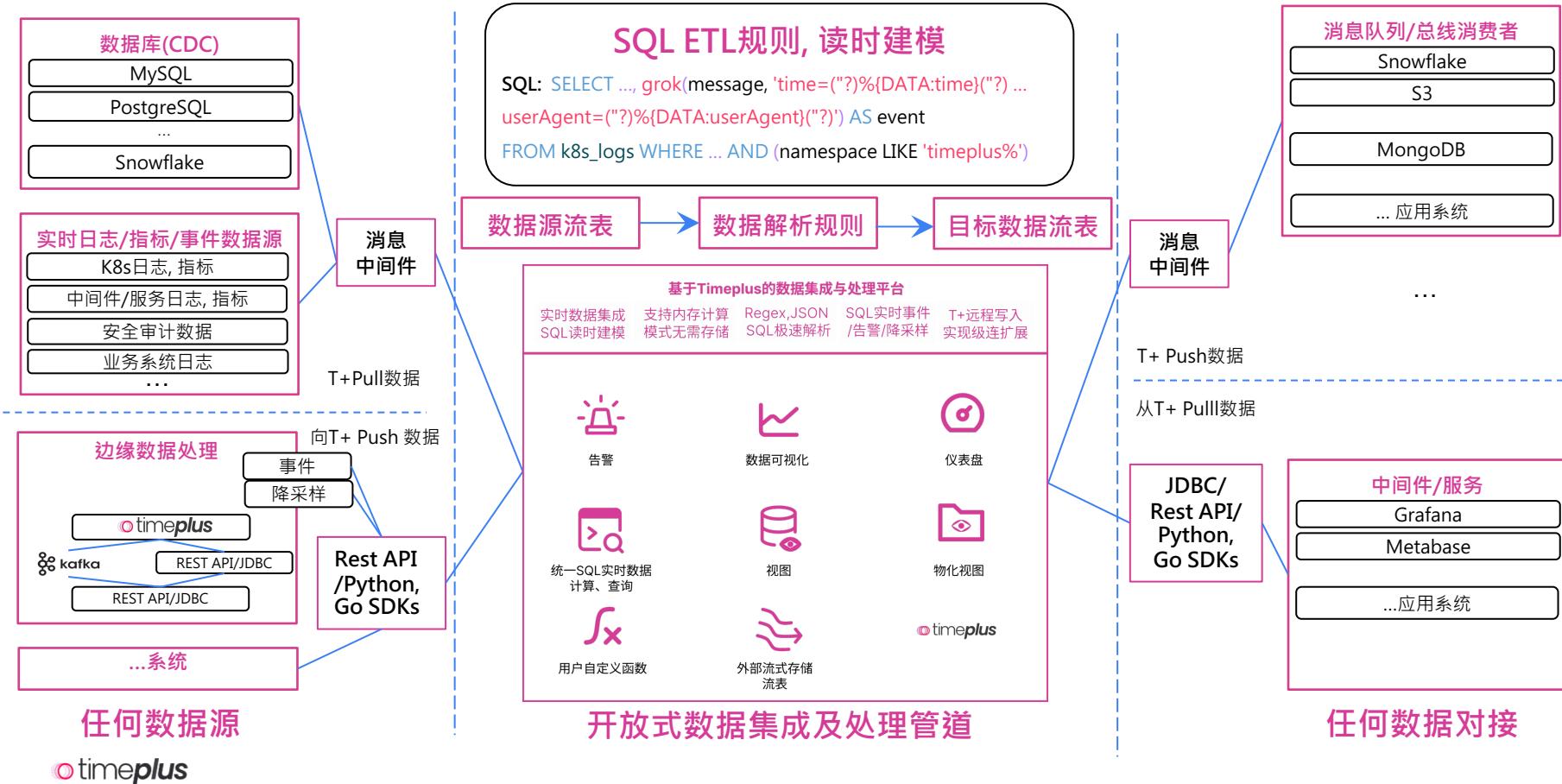
- 输入：成交数据，持仓数据，行情数据
- 时间窗口：1分钟滚动窗口，实时持续输出结果，延迟(P90)10ms
- 计算分组：对3000只股票*20个账户的每一个组合进行计算



采用Timeplus前



Timeplus典型应用场景 - 实时数据集成及处理



流式数据的Schema On Read

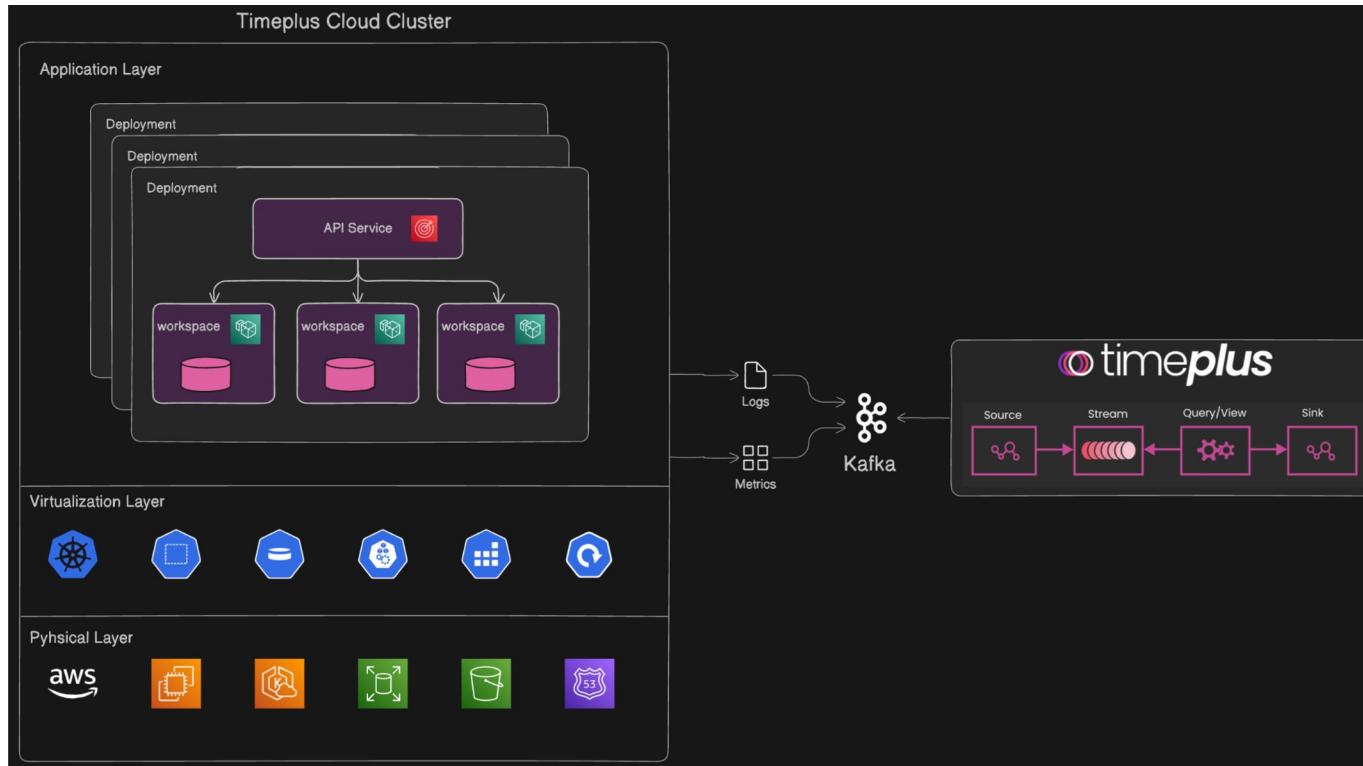
```
SELECT  
    cast(raw:counter:value, 'float') AS  
cpuupload, to_time(raw:timestamp) AS  
time, raw:tags.node AS node  
FROM  
eks_k8s_metrics  
WHERE  
(raw:namespace = 'host') AND (raw:name  
= 'load15')
```

直接分析非结构化或半结构化数据流，无需创建格式化流表

```
{  
    "counter": {  
        "value": 172946.23  
    },  
    "kind": "absolute",  
    "name": "cpu_seconds_total",  
    "namespace": "host",  
    "tags": {  
        "collector": "cpu",  
        "cpu": "0",  
        "host": "vector-herzl",  
        "mode": "idle",  
        "node": "ip-10-0-1-156.us-west-  
2.compute.internal"  
    },  
    "timestamp": "2022-09-  
29T23:47:04.544883775Z"  
}
```

Timeplus 时序数据及日志集成 – ITOPS：实时IT运维中心

覆盖基础设施运行指标、应用程序运行指标及业务指标的全域IT运营指标计算及监控场景

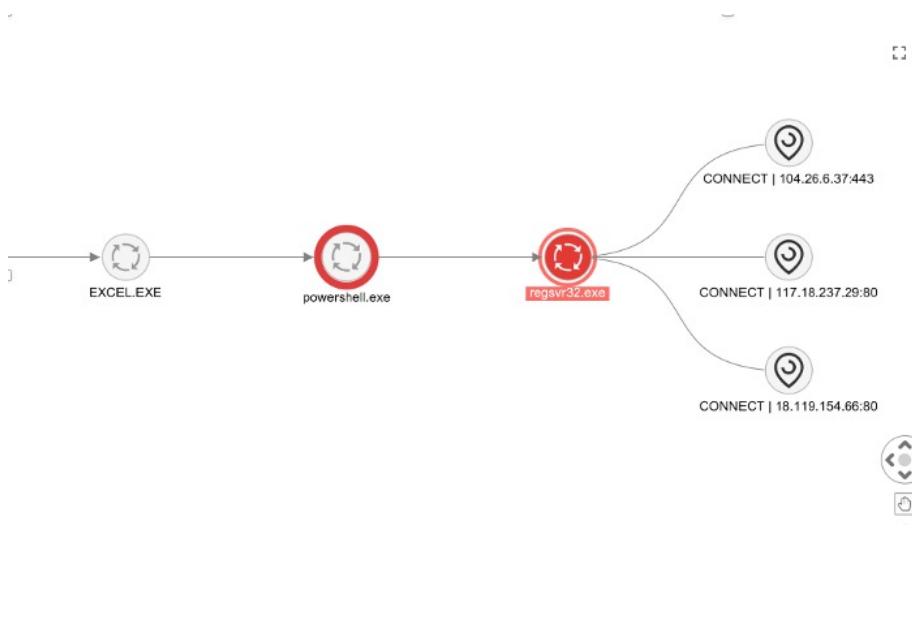


- **业务层指标**
计费和使用情况（计算/存储）、每个Workspace成本、收入/利润
- **应用层指标**
用户活动、应用程序错误、源、接收器和查询的吞吐量和状态
- **基础设施层指标**
CPU、内存、磁盘、节点和容器

Timeplus 事件处理 – 安全运营：复杂事件处理(CEP)

T+轻松实现开放式的基于复杂事件处理(CEP)模式匹配的安全检测规则

MITRE T1218.010: Squiblydoo attack 实时复杂事件检测



```
SELECT raw.`process`:`pid` as pid,
mt_1218_attack(_tp_time, raw) FROM threat GROUP BY pid
```

UDF: mt_1218_attack.js

```
var PROTON_FUNC = {
  has_customized_emit: true,
  initialize: function() {
    this.infected = false;
    ....
    this.result = [];
  },
  process: function (event_ts, events) {
    for (let i = 0; i < event_ts.length; i++) {
      let event = JSON.parse(events[i]);
      if (this.infected == false && event['event']['category'] == 'library' && event['dll'][name] == 'scrobj.dll') {
        this.infected = true;
        this.infect_raw_event = JSON.stringify(event);
        this.infect_ts = event_ts[i];
      } else if (this.infected == true && event['event']['category'] == 'network') {
        this.attack_ts = event_ts[i];
        this.attack_raw_event = JSON.stringify(event);
        this.result.push({
          'infect_ts': this.infect_ts,
          'infect_evt': this.infect_raw_event,
          'attack_ts': this.attack_ts,
          'attack_evt': this.attack_raw_event
        });
      }
    }
  }
}
```

Timeplus事件处理 - 安全运营：开放式安全事件分析平台

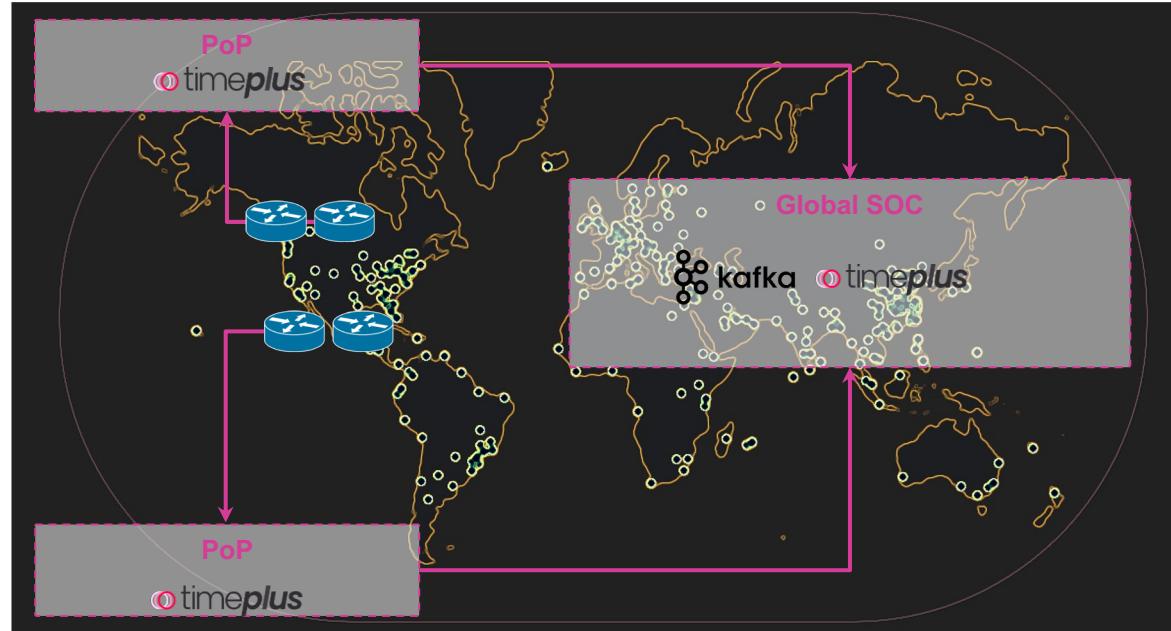
某全球头部线上游戏厂商分布式DDOS解决方案，基于开放规则的高性能安全事件处理/分析

边缘部署的挑战：

- 每秒超过1百万IPFIX事件
- 高基数键值
- 30s hopping window, 1s 滑动
- 10至30条DDoS规则
- 低延迟（亚秒级无堆积）
- 单机架配置：6核8线程核心处理器 & 12GB内存

Timeplus价值：

- 为边缘计算提供单一组件流处理引擎
- 实时警报和分析
- 避免无用数据输入，确保数据质量
- 减少下游数据，实现成本节约



云端挑战：

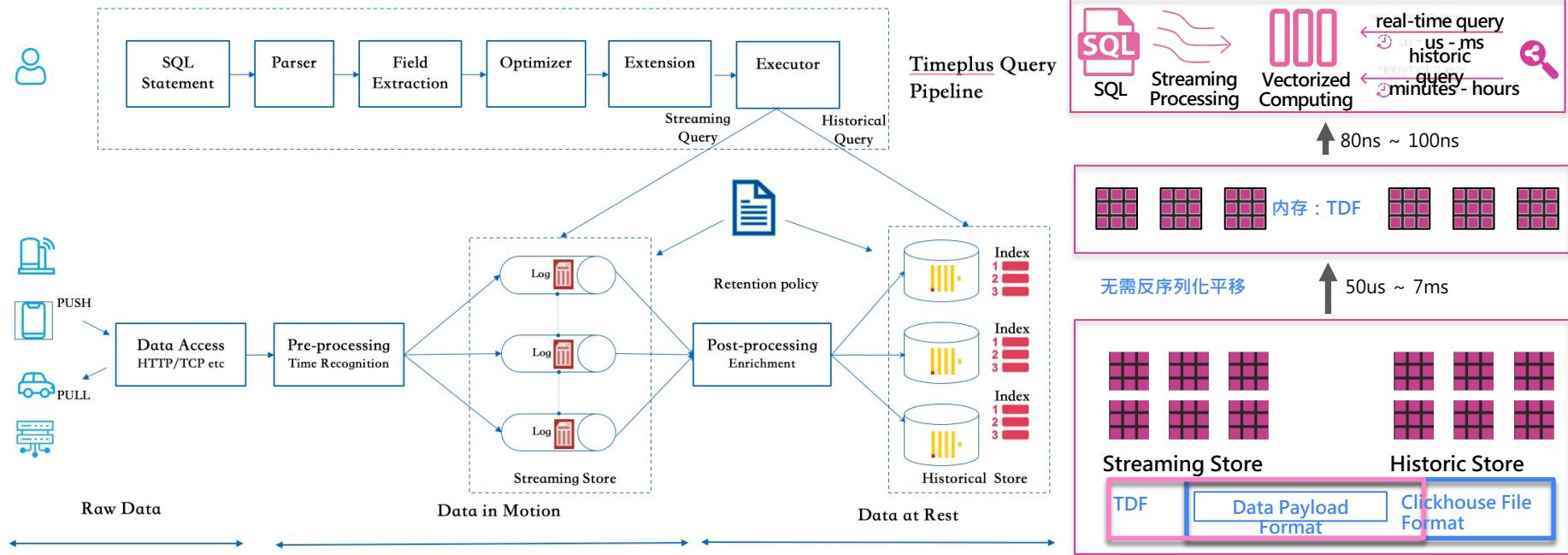
- 汇聚PoP事件处理，每秒超过7百万事件
- 更复杂的规则
- 大规模可扩展性
- 对延迟敏感（秒级）

Timeplus价值：

- 与 Confluent 无缝集成，实现层次化缓解
- 在全球中心支持复杂的流数据增强、关联和警报
- 高度可扩展且可靠的平台

Timeplus技术架构 – 流批存储的融合

一个统一的流式数据和历史数据存储



欢迎提问，有礼品



github.com/timeplus-io/proton



Timeplus 公众号



GitHub 开源