

领英对实时流计算的应用和探索

Stream Computing at LinkedIn: the application and the trail blazing

吕海 (Hai Lu)
数据基础架构资深工程师

背景介绍

LinkedIn

- 数据规模：在全球超过5亿用户，超过9百万公司...



Stream Infrastructure Team @LinkedIn

- 开发并维护Apache Kafka, Apache Samza, Brooklin, Databus等多个与实时流计算相关的开源项目

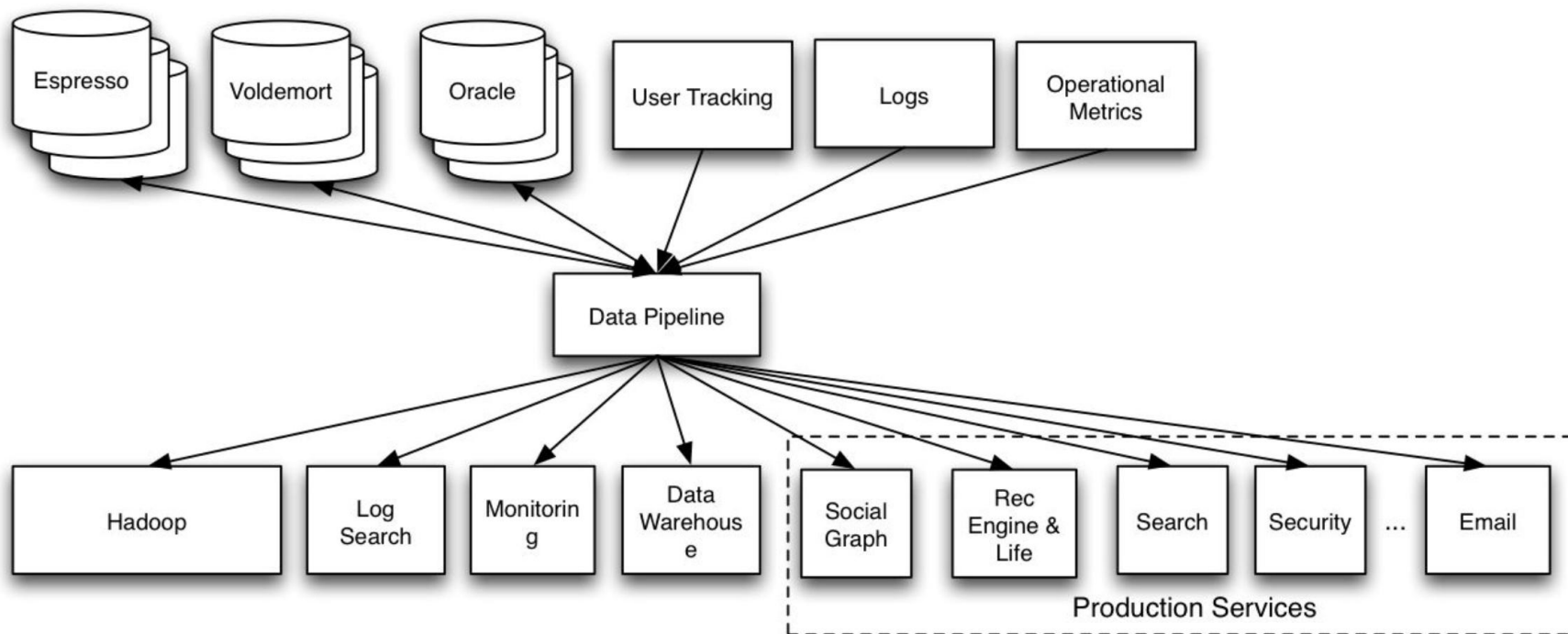


大纲

- Apache Kafka简介与应用场景
- Apache Samza简介
- Apache Samza应用场景和功能介绍
- Apache Samza最新进展

Apache Kafka

数据的搬运工



Apache Kafka

数据规模

- 每天约3万亿条消息; 0.5PB数据写, 2PB数据读
- 超过5万个话题(Topics), 1百万个分区(Partitions)

应用

- 用户追踪数据
- 日志 / 系统监控数据
- 应用间通信
- 在线与离线系统之间数据传输(数据库 / 前端 \leftrightarrow Hadoop)

Apache Kafka: 高效且稳定可靠的系统

应用于Espresso数据库的主从备份和跨数据中心同步(复制)

Durability

RF = 3
min.insync.replicas = 2
Disable unclean leader election
Rack awareness

acks = "all"
block.on.buffer.full = true
retries = Integer.MAX_VALUE

Low latency

Bump up num.replica.fetchers

max.block.ms = 0
Reduce linger.ms
Use LZ4

Ordering

max.inflight.requests = 1
close(0) in callback on send
failure

Nice-haves

Large message support
JBOD (RF3 is costly with RAID-10)

大纲

- Apache Kafka简介与应用场景
- Apache Samza简介
- Apache Samza应用场景和功能介绍
- Apache Samza最新进展

Apache Samza

开源的分布式实时流处理平台

Apache顶级项目

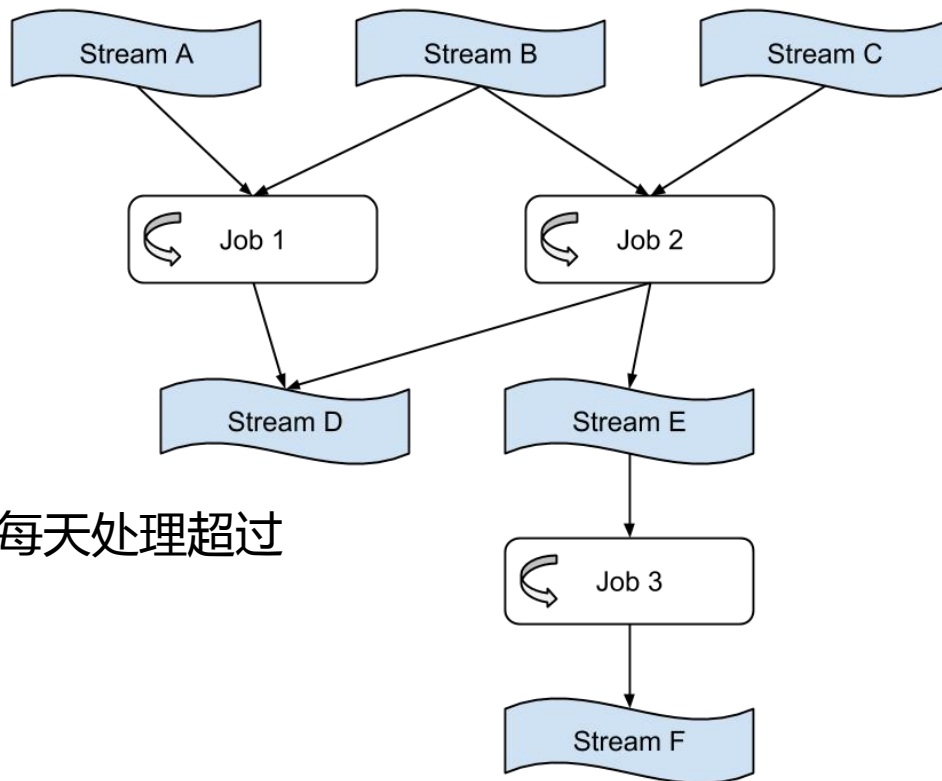
- 14 Committers, 62 Contributors

在领英大规模的部署

- 超过300个应用; 10000+YARN容器; 每天处理超过2500亿条消息

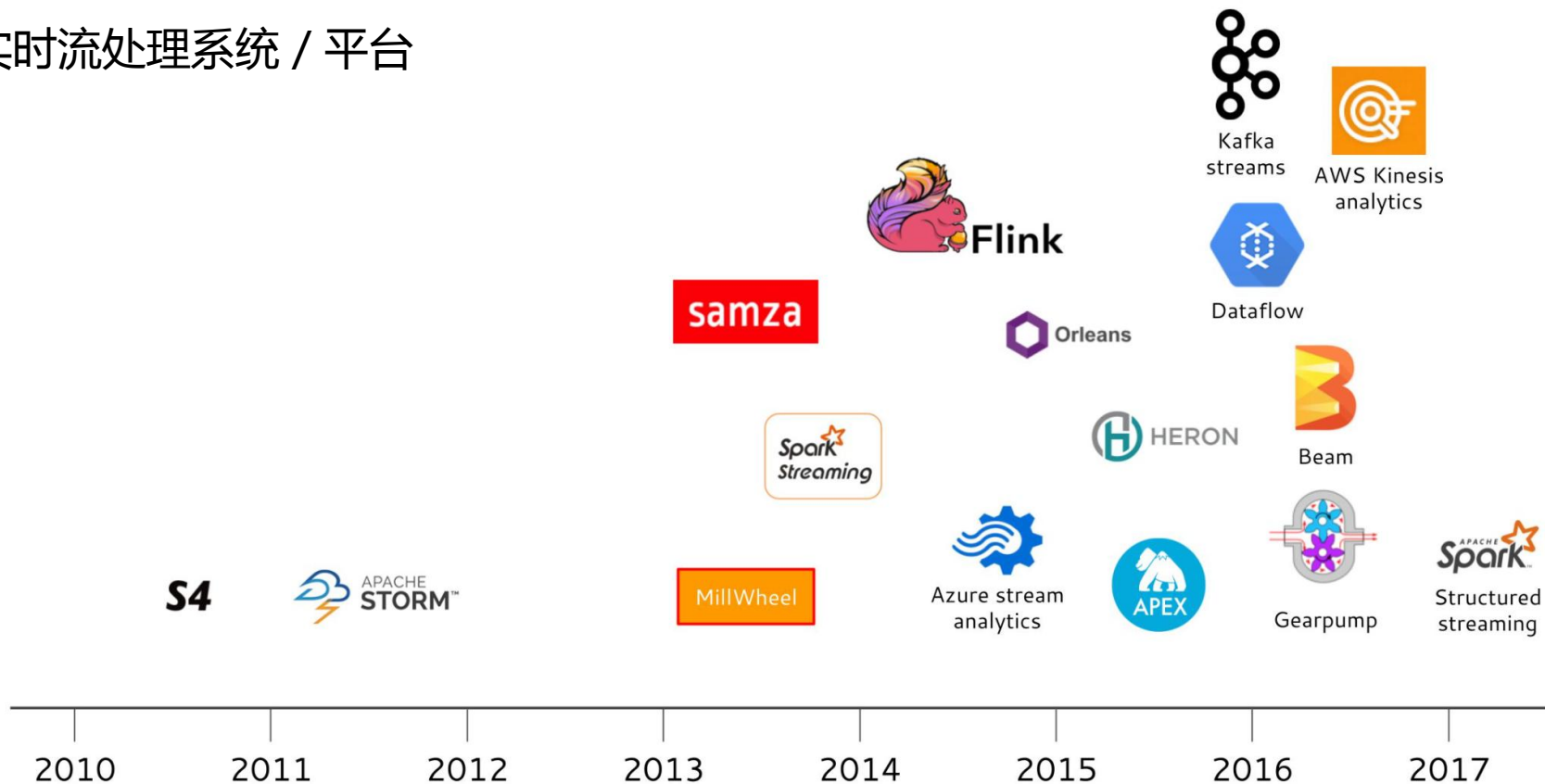
广泛的用户

- LinkedIn, Uber, VMWare, Intuit, TripAdvisor, Slack...



Apache Samza

实时流处理系统 / 平台



Apache Samza

主要特点

- 带状态的实时流处理(Stateful Processing)
- 高性能，高可靠(在LinkedIn等公司经过大规模的应用)
- 统一的流处理与批处理API
- 灵活的部署方式
- ...

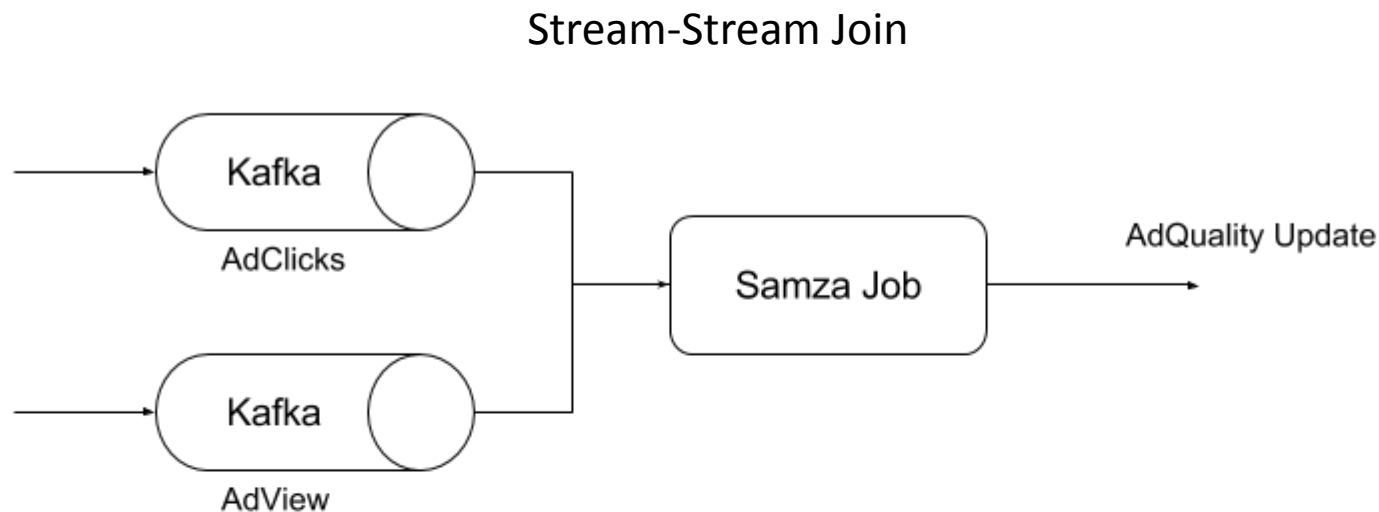
大纲

- Apache Kafka简介与应用场景
- Apache Samza简介
- Apache Samza应用场景和功能介绍
- Apache Samza最新进展

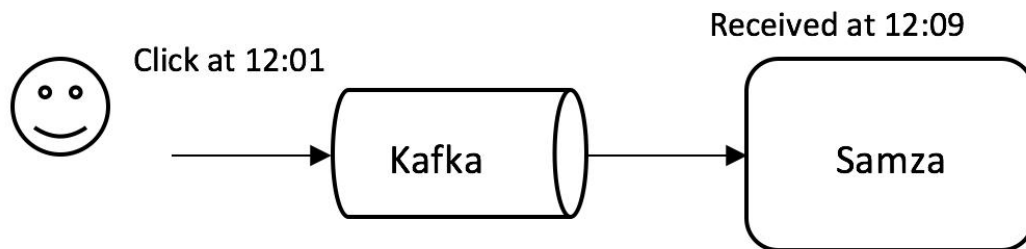
Apache Samza: 应用场景1

实时的广告质量评估

- 如果广告在被浏览后2分钟内被点击则是“好”广告
- 使用时间窗口(Windowing)



Apache Samza: 应用场景1



如何处理事件延迟(Late arrival events)

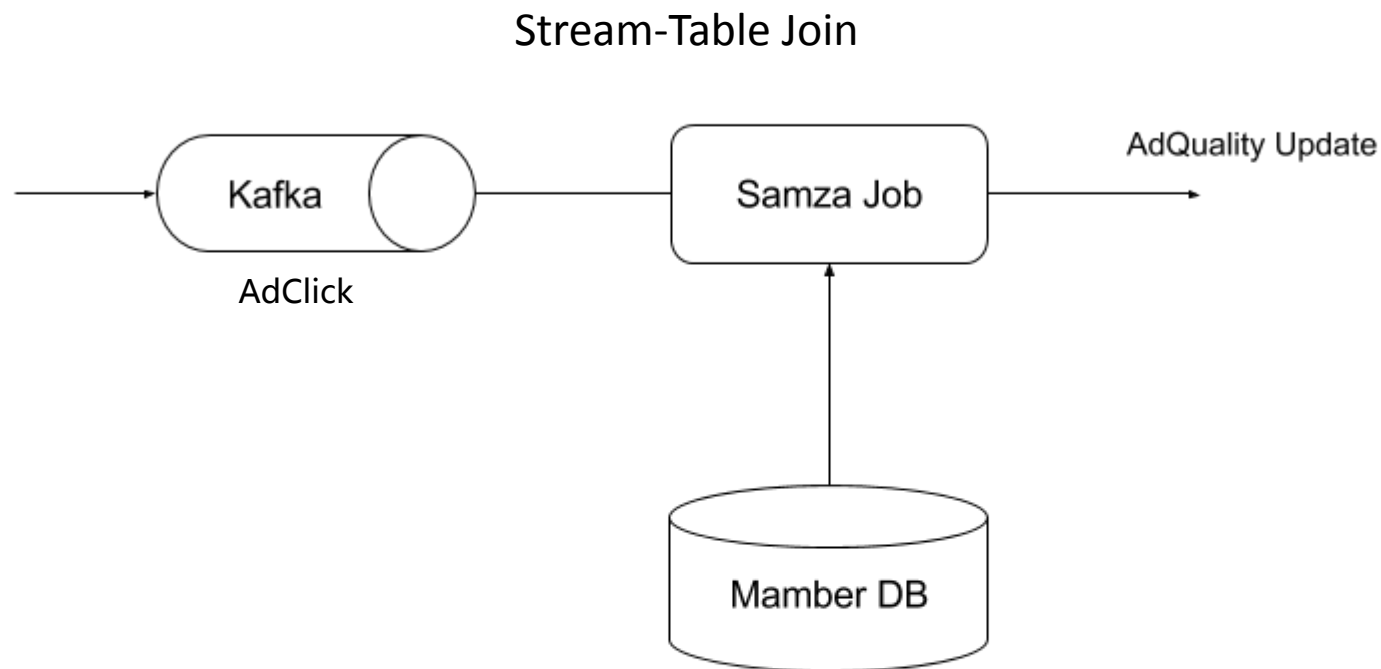
- 所有事件在本地存储(RocksDb)
- 找到受影响的窗口进行再处理
- 受Google MillWheel启发

本地状态会在Kafka中备份，用于恢复

- Host Affinity避免冷启动

Apache Samza: 应用场景1

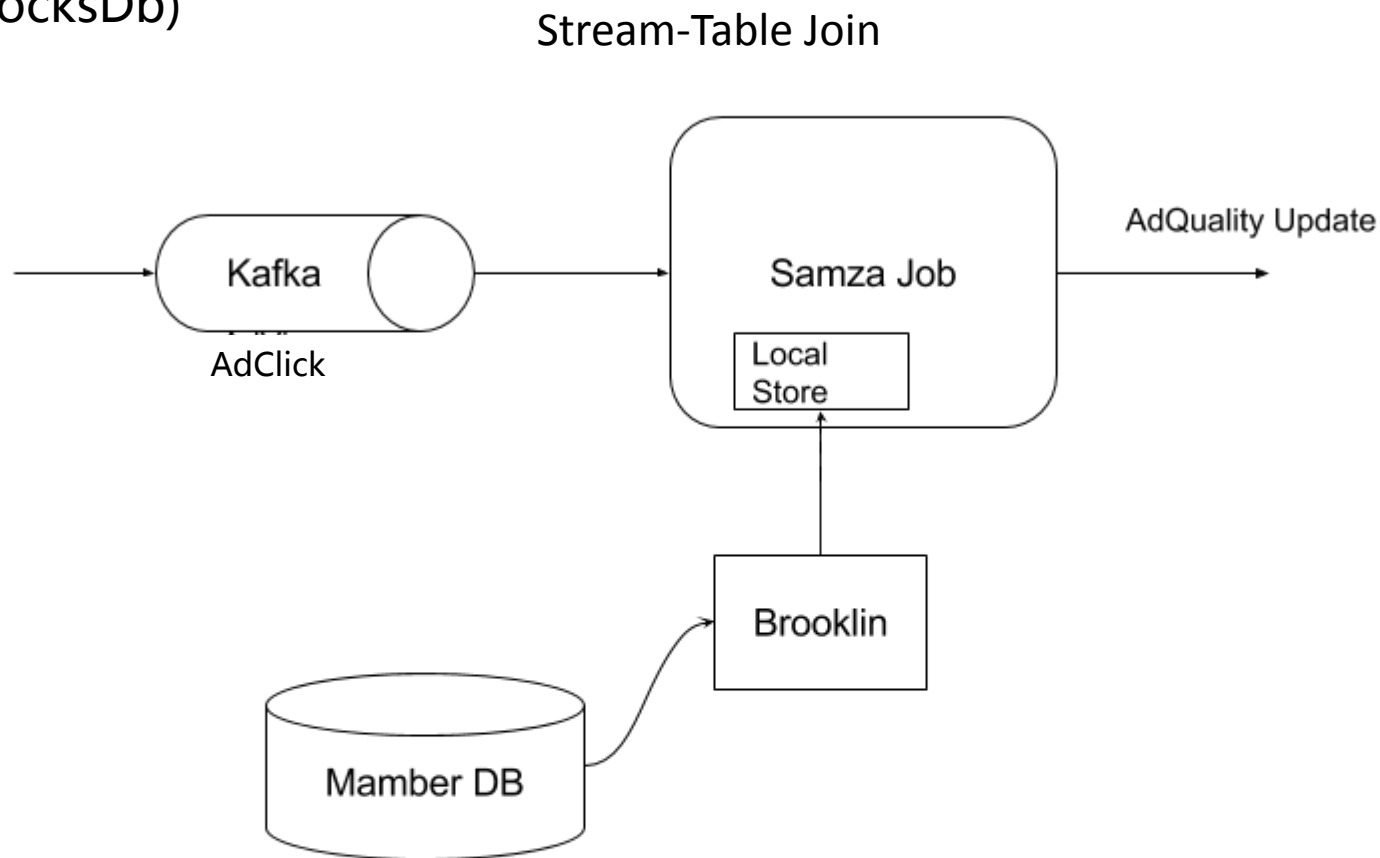
当需要连接用户数据时



Apache Samza: 应用场景1

使用本地数据库(RocksDb)

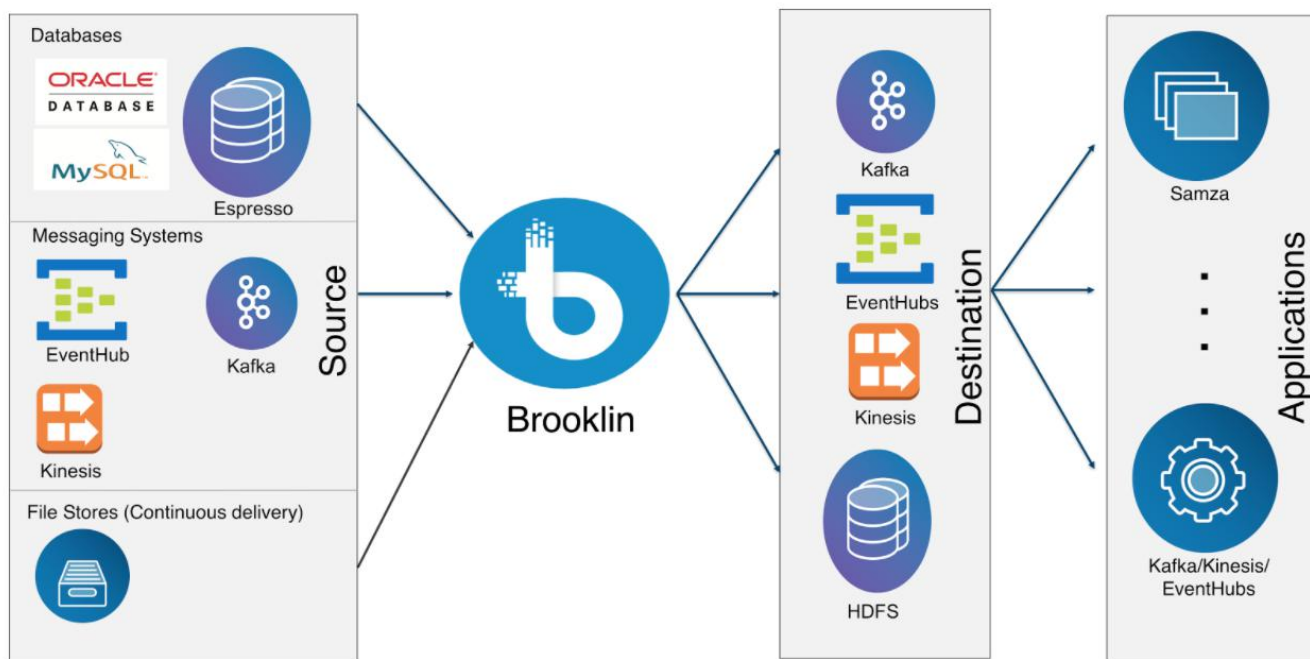
- 100x性能提升
- 保护主数据库



Brooklin

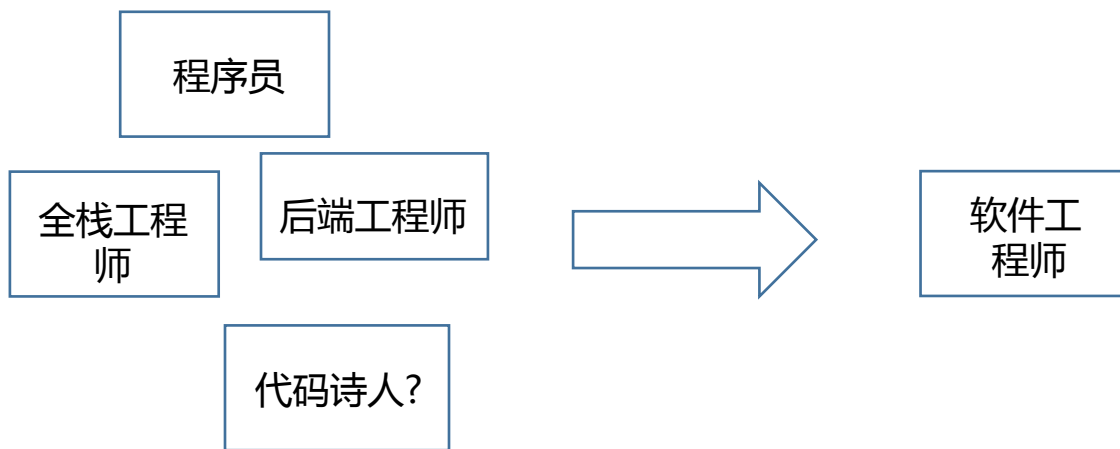
什么是Brooklin

- 通用的数据导入系统
- 支持多种数据源与目标(Kafka, Espresso...)
- 可作为Samza IO



Apache Samza: 应用场景2

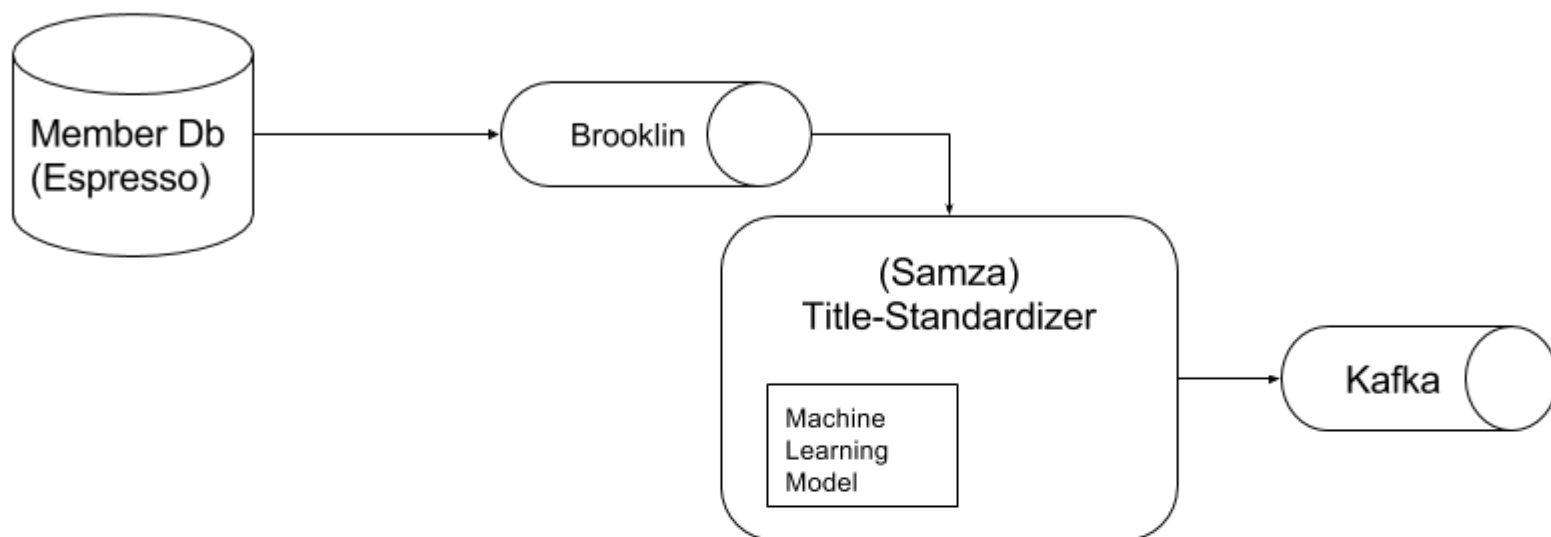
对职位名称进行标准化，以便作为机器学习模型的输入(特征标准化)



Apache Samza: 应用场景2

将机器学习模型载入Samza任务中

实时读取数据库中的更新并输出职位标准化的结果



Apache Samza: 应用场景2

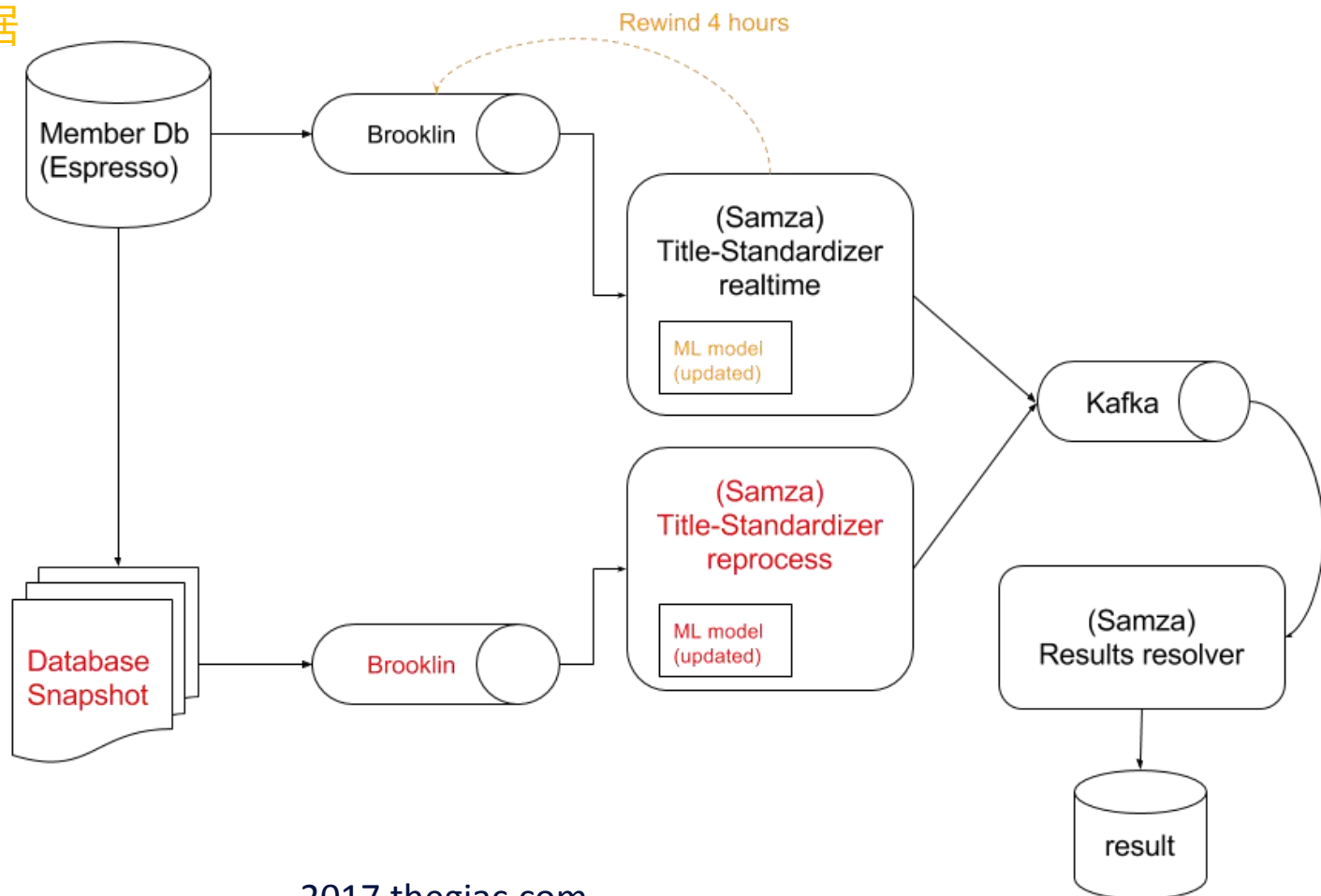
重处理(Reprocessing)

- 人为造成的bug，导致需要重放几个小时以前甚至几天以前的事件
- 机器学习模型改变或业务逻辑改变，需要重新处理整个数据库

重处理(Reprocessing)

重放4小时以前数据

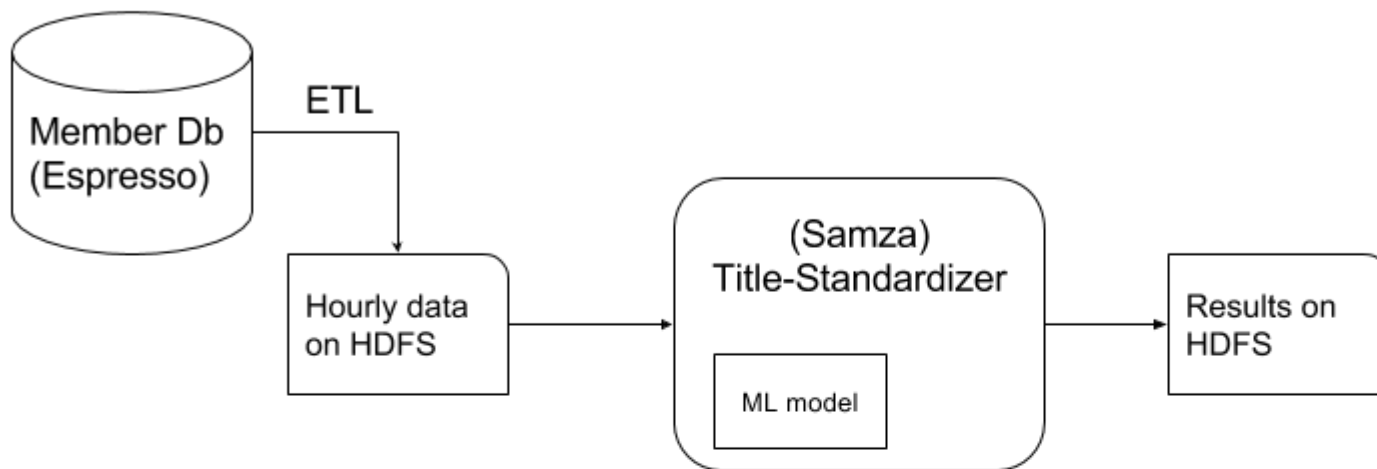
重处理所有数据



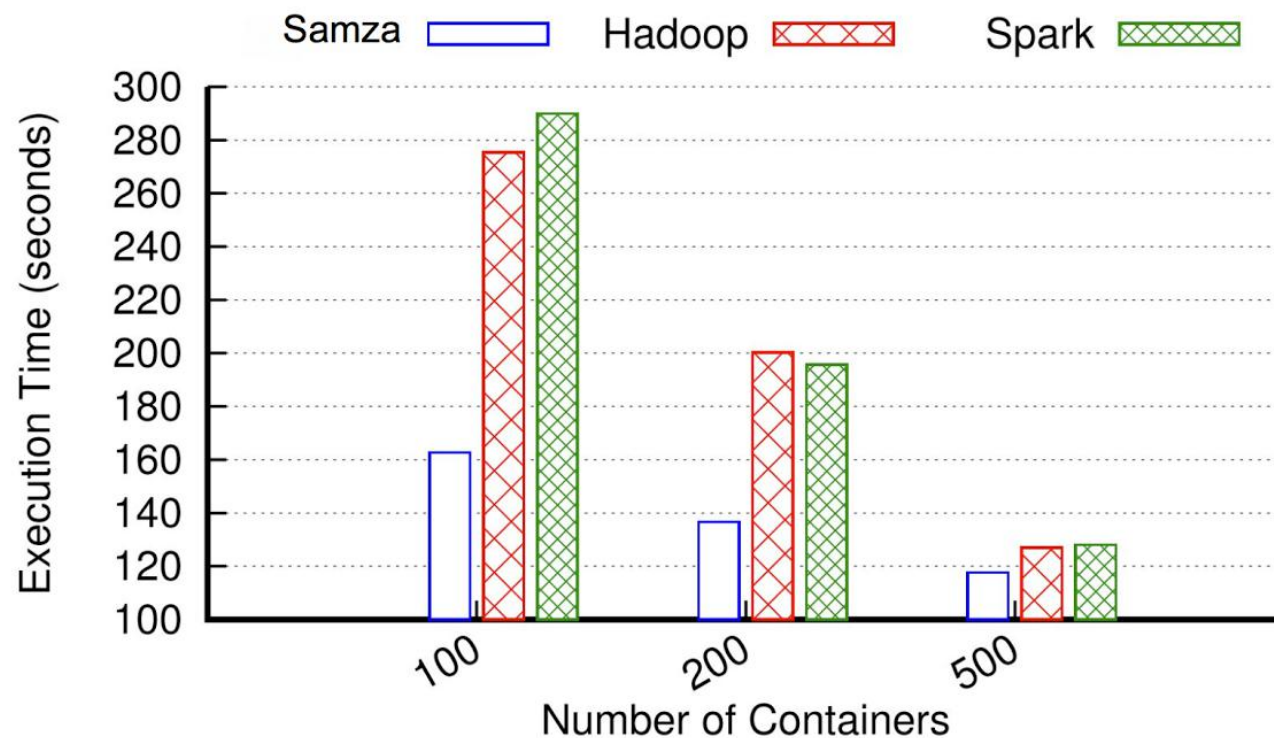
Apache Samza: 应用场景2

Samza on Hadoop

- 将Samza部署在Hadoop集群,读写HDFS
- 利用离线计算更强大的计算资源进行重处理
- 以HDFS作为输入对业务逻辑进行试错和A/B测试



Samza on Hadoop



Profile count,
group-by country

500 files

250GB input

其它应用场景

通知控制系统

反作弊系统

离线任务迁移到实时任务

...



大纲

- Apache Kafka简介与应用场景
- Apache Samza简介
- Apache Samza应用场景和功能介绍
- Apache Samza最新进展

Apache Samza: 与Apache Beam的整合

为什么加入Apache Beam

- 多语言支持(Python, Java)
- 在线与离线的整合
- 更多IO接口



Apache Samza: Samza SQL

基于Apache Calcite

为什么做Samza SQL

- 让更多人能使用实时流处理技术
- 更简便的实时数据查询
- 更便捷地创建实时流处理应用

Apache Samza: Samza SQL

```
1 public class PageViewCountApplication implements StreamApplication {
2     @Override
3     public void init(StreamGraph graph, Config config) {
4         MessageStream<PageViewEvent> pageViewEvents = graph.getInputStream("pageView");
5         MessageStream pageViewCount = graph.getOutputStream("pageViewCount");
6
7         pageView
8             .partitionBy(m -> m.memberId)
9             .window(Windows.keyedTumblingWindow(m -> m.memberId, Duration.ofMinutes(5)),
10                 initialValue, (m, c) -> c + 1)
11             .map(MyStreamOutput::new)
12             .sendTo(pageviewPerMember);
13     }
14 }
```



```
1 INSERT INTO kafka.pageViewCount
2     SELECT memberId, count(*) FROM kafka.pageViewStream
3     GROUP BY memberId, TUMBLE(current_timestamp, INTERVAL '5' MINUTES)
```

Q&A

GIAC | 全球互联网架构大会
GLOBAL INTERNET ARCHITECTURE CONFERENCE

GIAC

全球互联网架构大会

GLOBAL INTERNET ARCHITECTURE CONFERENCE



扫码关注GIAC公众号

2017.thegiac.com