

Kafka (/archive/?tag=Kafka)

Kafka Streams 与其他流处理框架的对比

"各种流处理框架的特点总结与对比"

Posted by Clinat on June 23, 2019

"Kafka Stream, Storm, Spark Streaming, Flink"

流处理

数据的业务价值随着时间的流失而迅速降低，因此在数据发生后必须尽快对其进行计算和处理。而传统的大数据处理模式无法满足数据实时计算的需求。

在诸如实时大数据分析、风控预警、实时预测、金融交易等诸多业务场景领域，批量处理对于上述对于数据处理时延要求苛刻的应用领域而言是完全无法胜任其业务需求的。

流处理框架

Kafka Streams

Kafka Streams是一个构建应用程序和微服务的客户端库，并且输入数据和输出数据均是保存在Kafka集群上的。Kafka Streams主要有如下特点：

- 非常简单的客户端库，可以非常容易的嵌入到任何java应用程序与任何应用程序进行封装集成。
- 使用Kafka集群作为消息层，没有外部依赖。
- 支持本地状态存储。
- 提供了快速故障切换分布式处理和容错能力。
- 提供了非常方便的API。

- 采用one-record-at-a-time的消息处理方式，实现消息处理的低延迟。

但是Kafka Streams的设计目标是足够轻量，所以很难满足对大体量的复杂计算需求，并且数据的输入和输出均是依靠Kafka集群，对于其他的数据源需要借助Kafka connect将数据输入到Kafka主题中，然后在通过Kafka Streams程序进行处理，并通过Kafka connect将主题中的数据转存到其他数据源。

所以Kafka Streams更适合计算复杂度较小，数据流动过程是Kafka->Kafka的场景。

Storm

Storm是一个较早的实时计算框架，并且在早起应用在了各大互联网企业。具有高性能，低延迟，分布式，容错，消息不丢失等特点。

但是，从其他方面看，Storm只能算是个半成品，ack功能非常不优雅，开启该功能之后会严重影响吞吐量，并且没有提供exactly once功能，所以给大家带来一种印象：流式系统只是用于吞吐量较小，低延迟不精确的计算。而对于精确的计算常常交给批处理系统进行处理（Lambda架构），所以需要维护两套系统。

Spark Streaming

采用微批的方式，提高了吞吐性能。Spark streaming批量读取数据源中的数据，然后把每个batch转化成内部的RDD。Spark streaming以batch为单位进行计算，而不是以record为单位，大大减少了ack所需的开销，显著满足了高吞吐、低延迟的要求，同时也提供exactly once功能。但也因为处理数据的粒度变大，导致Spark streaming的数据延时不如Storm，Spark streaming是秒级返回结果（与设置的batch间隔有关），Storm则是毫秒级。

但是Spark Streaming的优点是可以与Spark大环境进行有效的结合。

Flink

Flink 是一种可以处理批处理任务的流处理框架。Flink 流处理为先的方法可提供低延迟，高吞吐率，近乎逐项处理的能力，并且提供了复杂计算的能力。

框架对比

	Storm	Trident	Spark Streaming	Flink	Samza	Kafka streams
数据流模型	原生	微批	微批	原生	原生	原生
状态存储	不支持状态管理	本地存储, 外部数据库	多种状态存储方式	多种状态存储方式	本地存储, Kafka主题	本地存储, 日志变更主题
时延	低	高	高	低	低	低
吞吐量	低	高	高	高	高	高
保障机制	at-least-once	exactly-once	exactly-once	exactly-once	at-least-once	exactly-once
容错机制	record ack	record ack	RDD based, checkpoint	checkpoint	Kafka log-base	Kafka log
成熟度	较多不足, 但实际应用比较广泛	Storm基础上改进	流行的框架之一, Spark大环境	较新的流处理框架, 性能非常优秀	基于Kafka作为数据源	完全基于Kafka集群实现
定位	框架	框架	框架	框架	框架	类库

PREVIOUS

GITHUB PACKAGE_LOCK.JSON 安全警报
 (/2019/06/22/PACKAGEJSONLOCK/)

NEXT

KAFKA STREAMS 设计原理之线程模型
 (/2019/06/23/KAFKASTREAMS_THREADMO)

在 HUX BLOG 上还有

3 个月前 · 6条评论

中国高等教育的系统性失败 - 黄玄的博客

2 年前 · 3条评论

Vim 与中文输入法 - 黄玄的博客

1 年前 · 1

My Progl Lang

Disqus 无法加载。如果您是管理员，请参阅[故障排除指南](#)。

FEATURED TAGS (/archive/)

Kafka (/archive/?tag=Kafka)

GC (/archive/?tag=GC)

(/archive/?tag=GC)

FRIENDS

(/archive/?tag=GC)

(/archive/?tag=GC) (/archive/?tag=GC)CLQ (http://www.baidu.com)