

Tumbling Windows vs Sliding Windows区别与联系

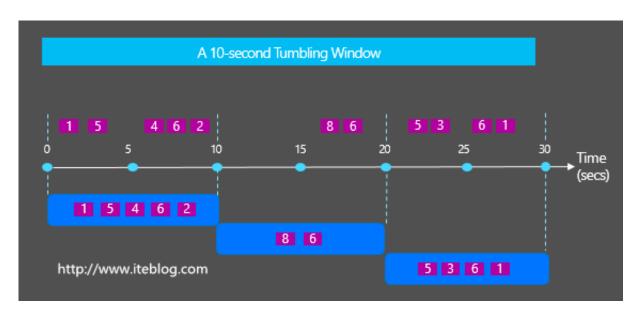
在流系统中通常会经常使用到Windows来统计一定范围的数据,比如按照固定时间、按个数等统计。一般会存在两种类型的Windows: Tumbling Windows vs Sliding Windows,它们很容易被初学者混淆,那么Tumbling Windows vs Sliding Windows之间到底有啥区别与联系呢?这就是本文将要展开的。

Tumbling的中文意思是摔跤,翻跟头,翻筋斗;Sliding中文意思是滑动,滑行、滑落;下跌;打滑等意思。中文意思之间不太好区分。

Tumbling Windows vs Sliding Windows之间最大的区别就是:Tumbling windows是不可能重叠的;而Sliding Windows是存在重叠的。

Tumbling window

Tumbling window在流数据中进行滚动,这种窗口不存在重叠,也就是说一个events/data只可能出现在一个窗口中,如下图所示:

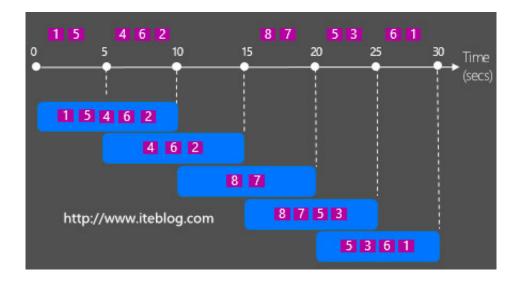


从上图可以看出, Tumbling window之间是不存在重叠的。Tumbling window的创建可以基于数量(比如每5个元素构成一个窗口)或者基于时间(如上图每隔10s创建一个窗口)。

Sliding Window

Sliding Window是在流数据中进行滑动,窗口之间可以重叠,它可以在传入的数据流中进行平滑聚合。Sliding Window如下图所示:





和Tumbling window一样,我们可以根据需求创建基于数量或者基于时间窗口。

Tumbling Windows vs Sliding Windows区别



微信扫一扫,加关注 即可及时了解Spark、Hadoop或者Hbase 等相关的文章 欢迎关注微信公共帐号:iteblog hadoop

过往记忆博客(http://www.iteblog.com) 专注于Hadoop、Spark、Flume、Hbase等 技术的博客,欢迎关注。

Hadoop、Hive、Hbase、Flume等交流群: 138615359和149892483

如果想及时了

解Spark、Hadoop或者Hbase相关的文章,欢迎关注微信公共帐号:iteblog_hadoop

我们可以将Tumbling Windows看作是Sliding Windows的特殊情况,当Sliding Windows的滑动时间和窗口时间是一样的时候,这时候Sliding Windows窗口之间就不会重叠,这就是Tumbling Windows。

在当前的Spark版本中(甚至是Spark 2.0)是没有提供直接创建Tumbling Windows的API,但是我们可以通过将Sliding Windows的window duration和slide duration设置成一样来构成Tumbling Windows(可以参见<u>《Spark 2.0介绍:Spark SQL中的Time Window使用》</u>)。

而在Flink中存在非常丰富的窗口函数,包括Global Windows、Tumbling Windows、Sliding



Windows以及Session Windows等等,这也是Flink功能强大的一面。本博客将会在后面的文章中单独介绍Flink的窗口函数,欢迎关注/

。过往记忆大数据技术博客微信公共账号:iteblog_hadoop。

本博客文章除特别声明,全部都是原创! 原创文章版权归过往记忆大数据(过往记忆)所有,未经许可不得转载。 本文链接:【】()