Spark Streaming在唯品会的实践

毛玮

wei.mao@intel.com

About Us

- Dedicate to Big Data Ecosystem
- Mainly focus on *Spark!
- Corporate with YOU

FREE !

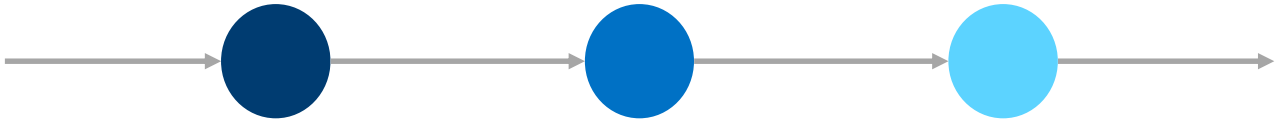
E-mail: wei.mao@intel.com

Overview

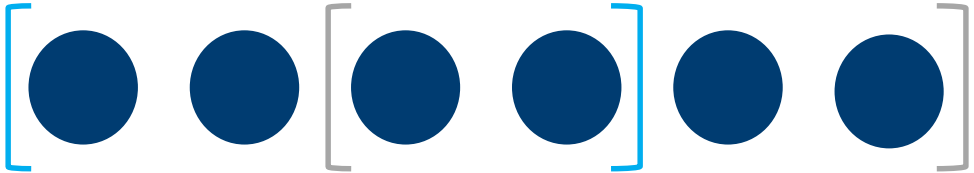
- *Storm VS Spark Streaming
  - Ease of use
  - SQL support
  - Throughput
  - Latency

Streaming Applications

- ETL Operations

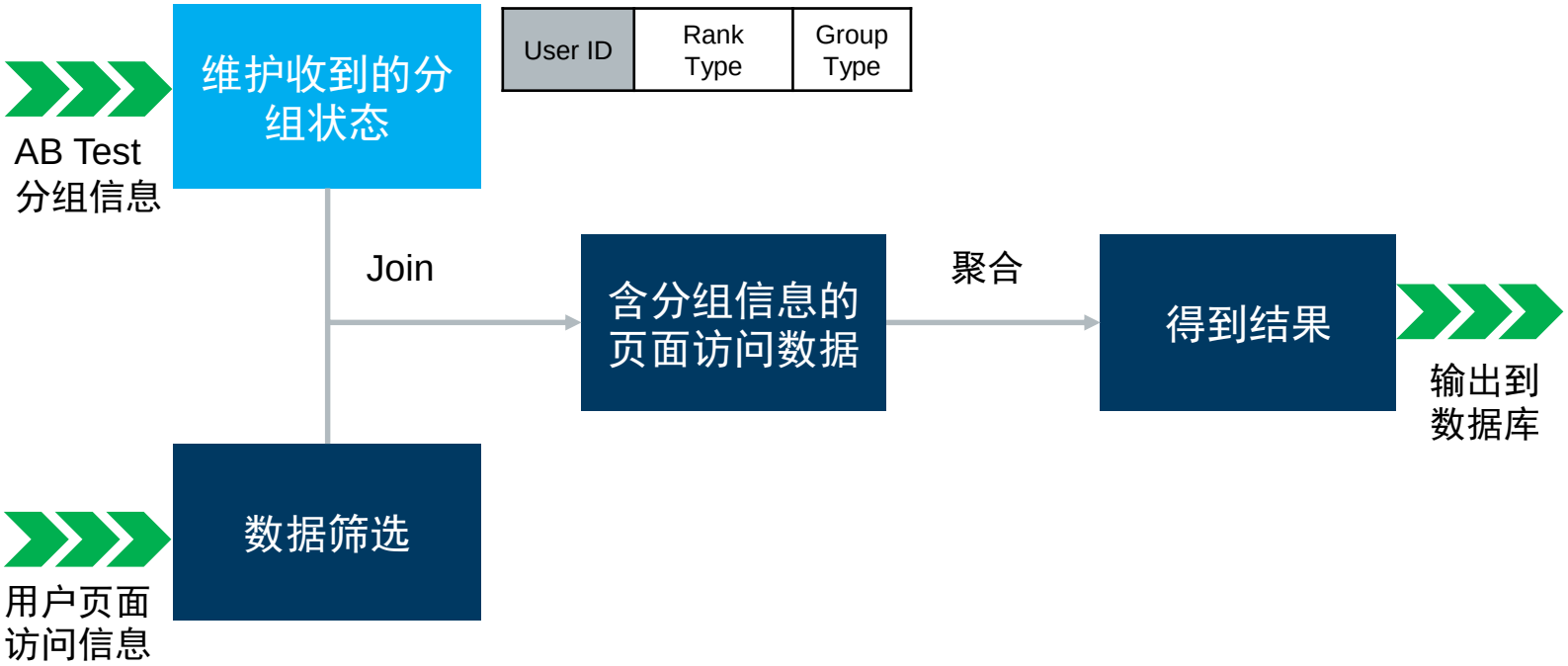


- Analyze in bounded interval (Windowing & State)



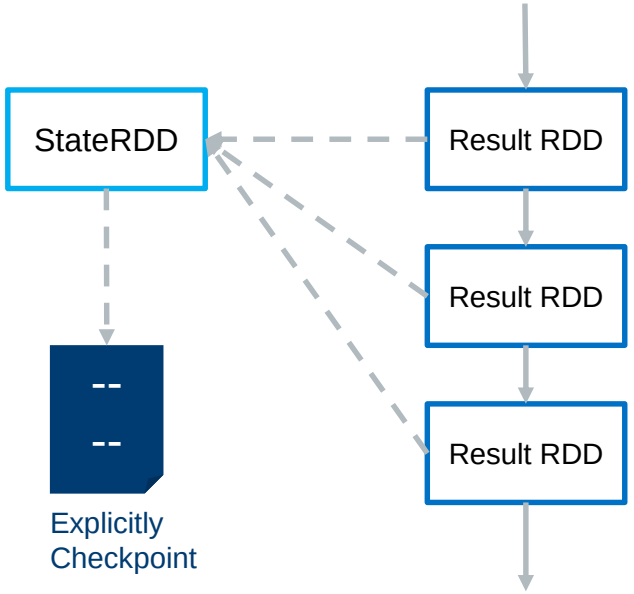
- Machine Learning, Pattern Recognition, etc.

Dynamic Recommendation System



State Management

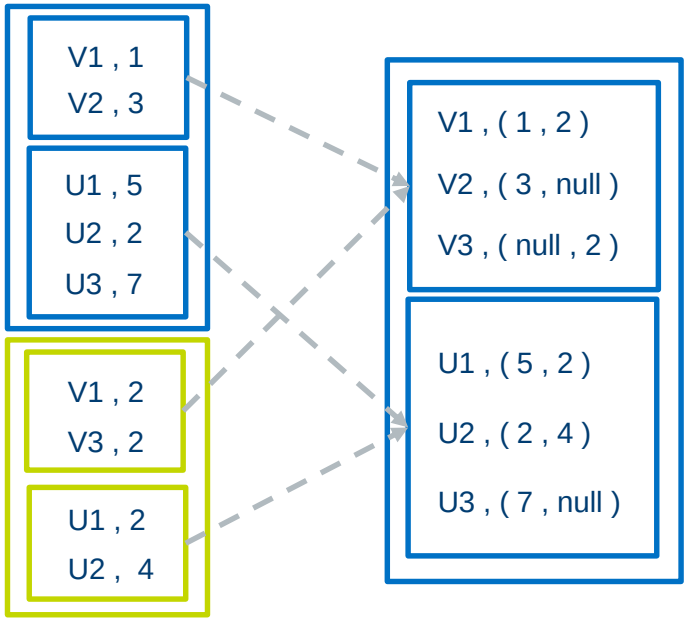
- External StateRDD
  - Not support High Availability
  - Data Skew
  - Complex Process
  - Growing Files

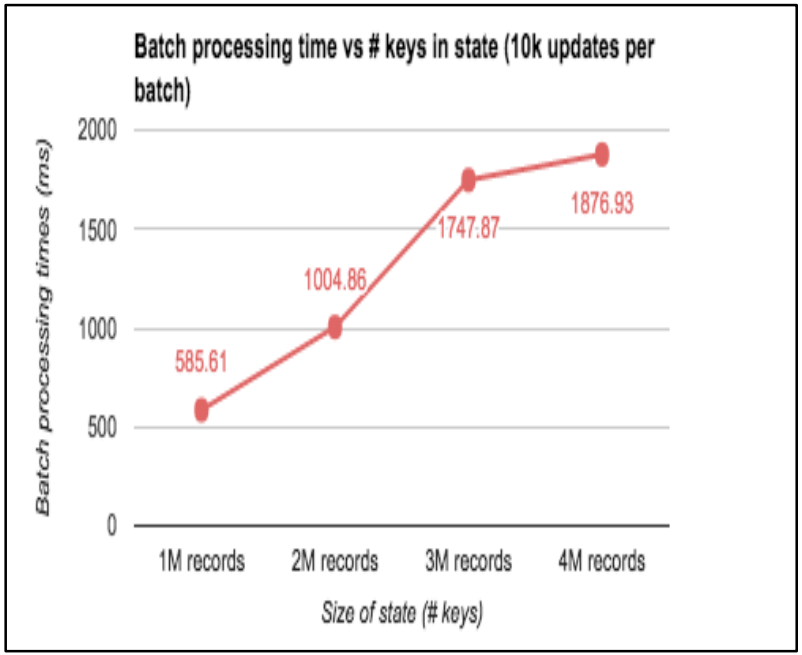


- Spark Streaming Native State Operator

updateStatesByKey

- Internal: RDD.cogroup(rdd: RDD)



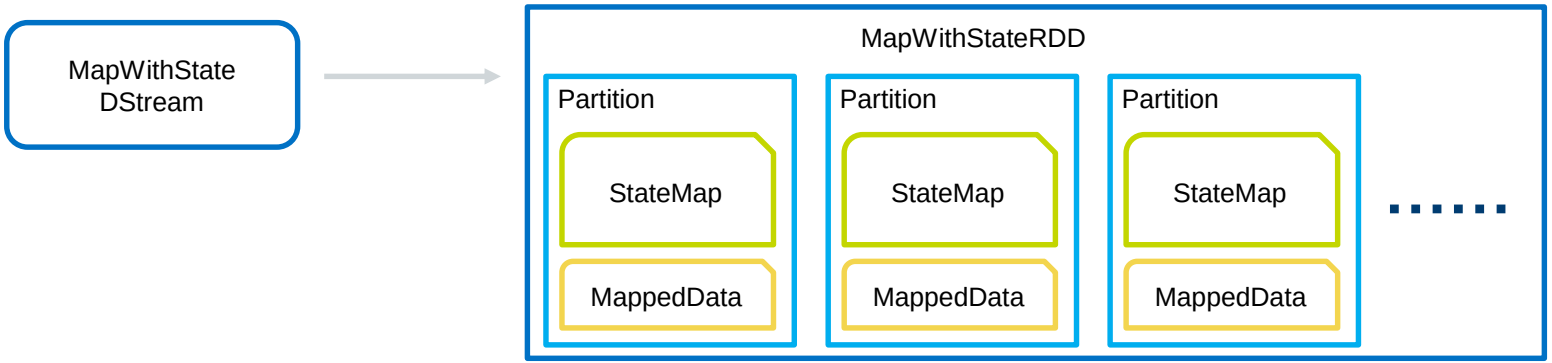


*图片引用自https://docs.google.com/document/d/1NoALLyd83zGs1hNGMm0Pc5YOVgiPpMHugGMk6COqxxE/edit#heading=h.ph3w0clkd4em
*英特尔不控制或审计本文提及的第三方基准测试数据或网址。请访问提及的网站，以确认提及的数据是否准确

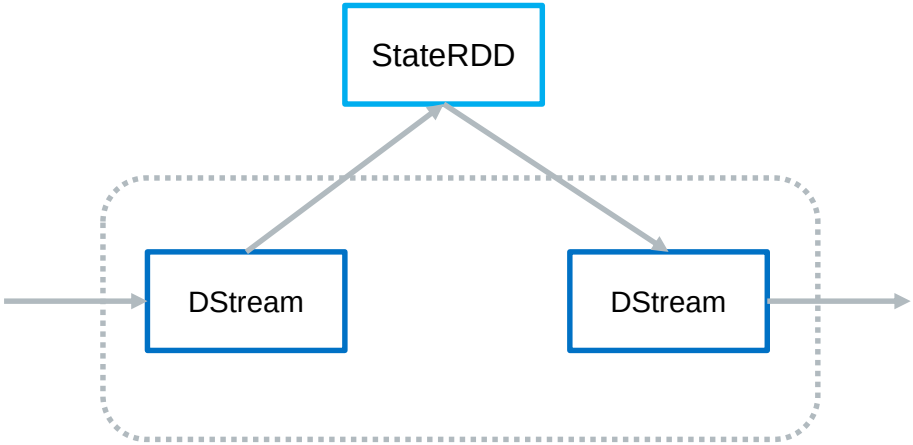
mapWithState （Spark 1.6）

JIRA: SPARK-2629

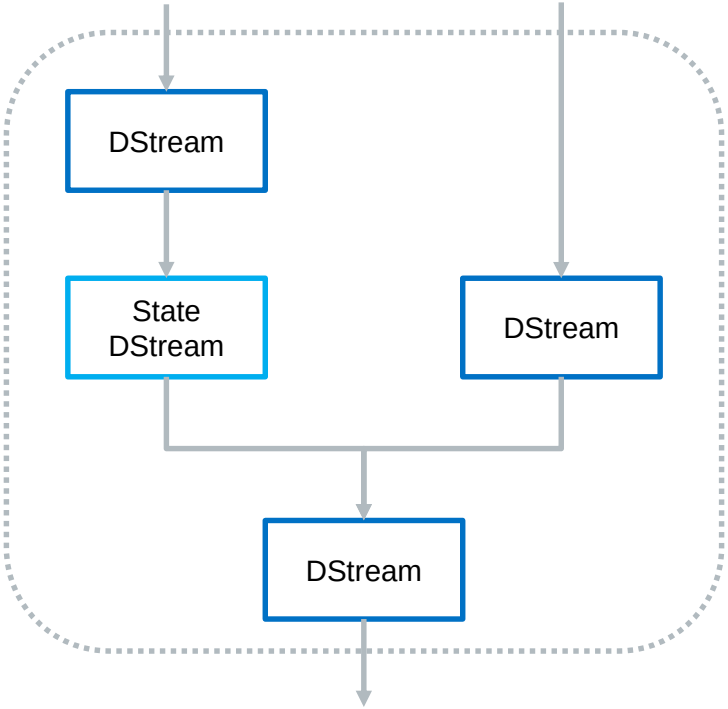
- Need for optimized state management to avoid scanning every key
- Native support timeout mechanism
- Return items other than state



Refactor



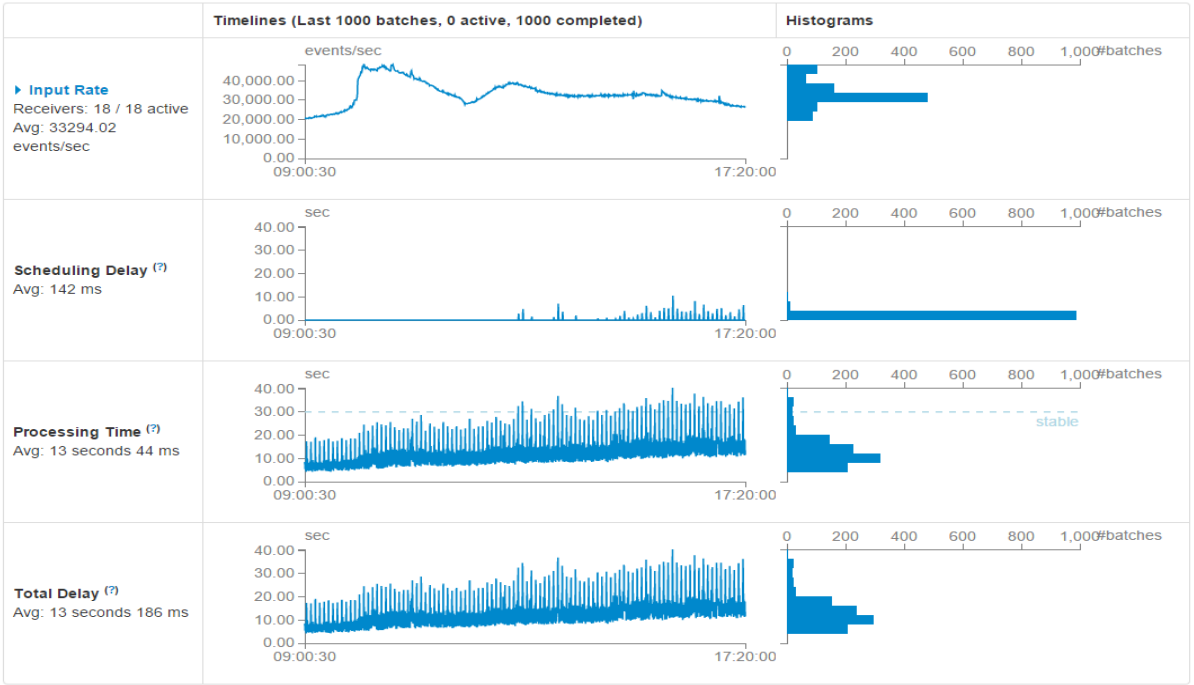
- ~~Not support High Availability~~
- ~~Data Skew~~
- ~~Complex Process~~
- ~~Growing Files~~



Test Result

Streaming Statistics

Running batches of **30 seconds** for **13 hours 28 minutes 6 seconds** since **2016/05/17 03:52:20** (**1615** completed batches, **1168145503** records)



Test Result

16 Partitions

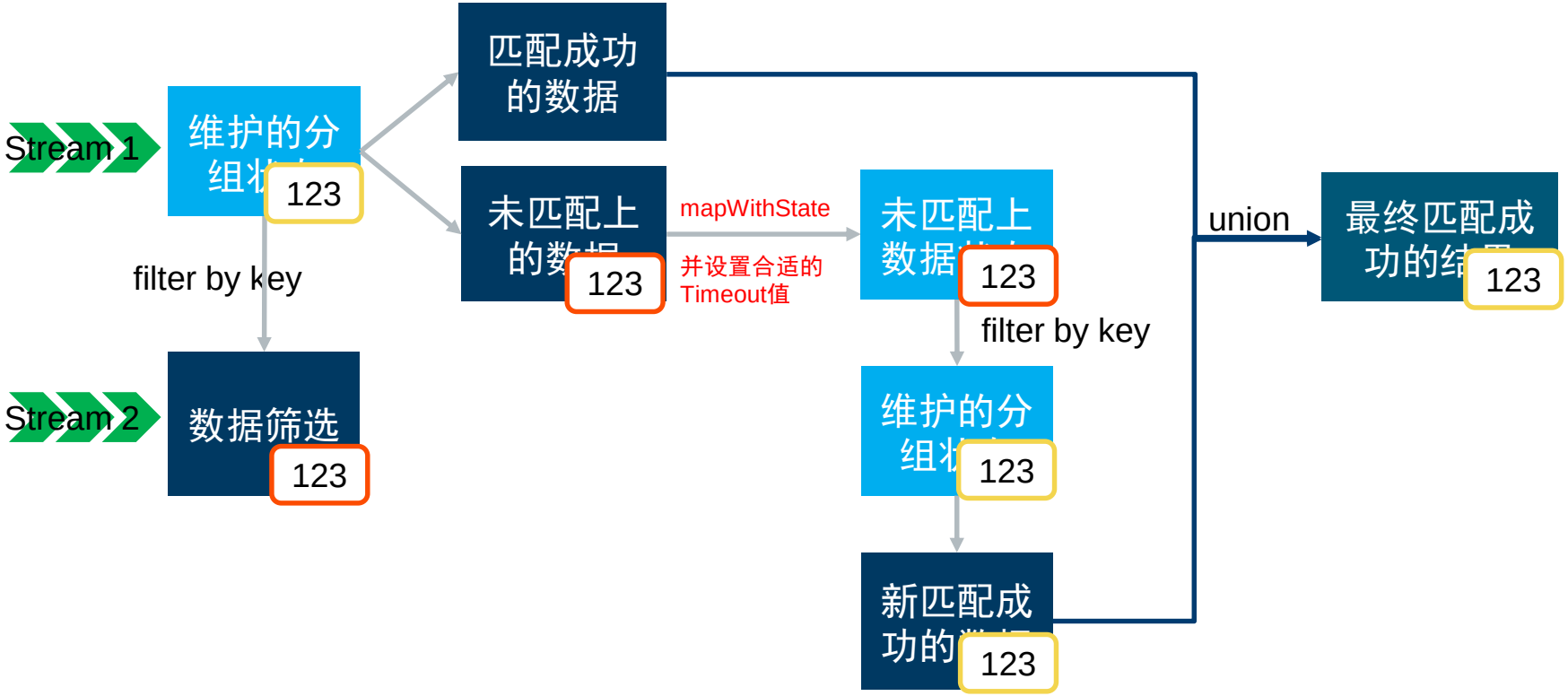
Block Name ▲	Storage Level	Size in Memory	Size on Disk	Executors
rdd_214872_0	Memory Serialized 1x Replicated	33.8 MB	0.0 B	gdata-spark6-202-218.idc:64343
rdd_214872_1	Memory Serialized 1x Replicated	33.9 MB	0.0 B	gdata-spark6-202-224.idc:41781
rdd_214872_10	Memory Serialized 1x Replicated	33.9 MB	0.0 B	gdata-spark6-202-225.idc:60006
rdd_214872_11	Memory Serialized 1x Replicated	33.9 MB	0.0 B	gdata-spark6-202-223.idc:13420
rdd_214872_12	Memory Serialized 1x Replicated	33.8 MB	0.0 B	gdata-spark6-202-220.idc:30859
rdd_214872_13	Memory Serialized 1x Replicated	33.9 MB	0.0 B	gdata-spark6-202-224.idc:27236
rdd_214872_14	Memory Serialized 1x Replicated	33.8 MB	0.0 B	gdata-spark6-202-222.idc:32962
rdd_214872_15	Memory Serialized 1x Replicated	33.9 MB	0.0 B	gdata-spark6-202-218.idc:64343
rdd_214872_2	Memory Serialized 1x Replicated	33.8 MB	0.0 B	gdata-spark6-202-226.idc:34334
rdd_214872_3	Memory Serialized 1x Replicated	33.9 MB	0.0 B	gdata-spark6-202-219.idc:51460
rdd_214872_4	Memory Serialized 1x Replicated	33.9 MB	0.0 B	gdata-spark6-202-221.idc:24808
rdd_214872_5	Memory Serialized 1x Replicated	33.8 MB	0.0 B	gdata-spark6-202-225.idc:33949
rdd_214872_6	Memory Serialized 1x Replicated	33.8 MB	0.0 B	gdata-spark6-202-223.idc:54510
rdd_214872_7	Memory Serialized 1x Replicated	33.8 MB	0.0 B	gdata-spark6-202-222.idc:16580
rdd_214872_8	Memory Serialized 1x Replicated	33.8 MB	0.0 B	gdata-spark6-202-219.idc:36846
rdd_214872_9	Memory Serialized 1x Replicated	33.8 MB	0.0 B	gdata-spark6-202-220.idc:13587

Event Delay

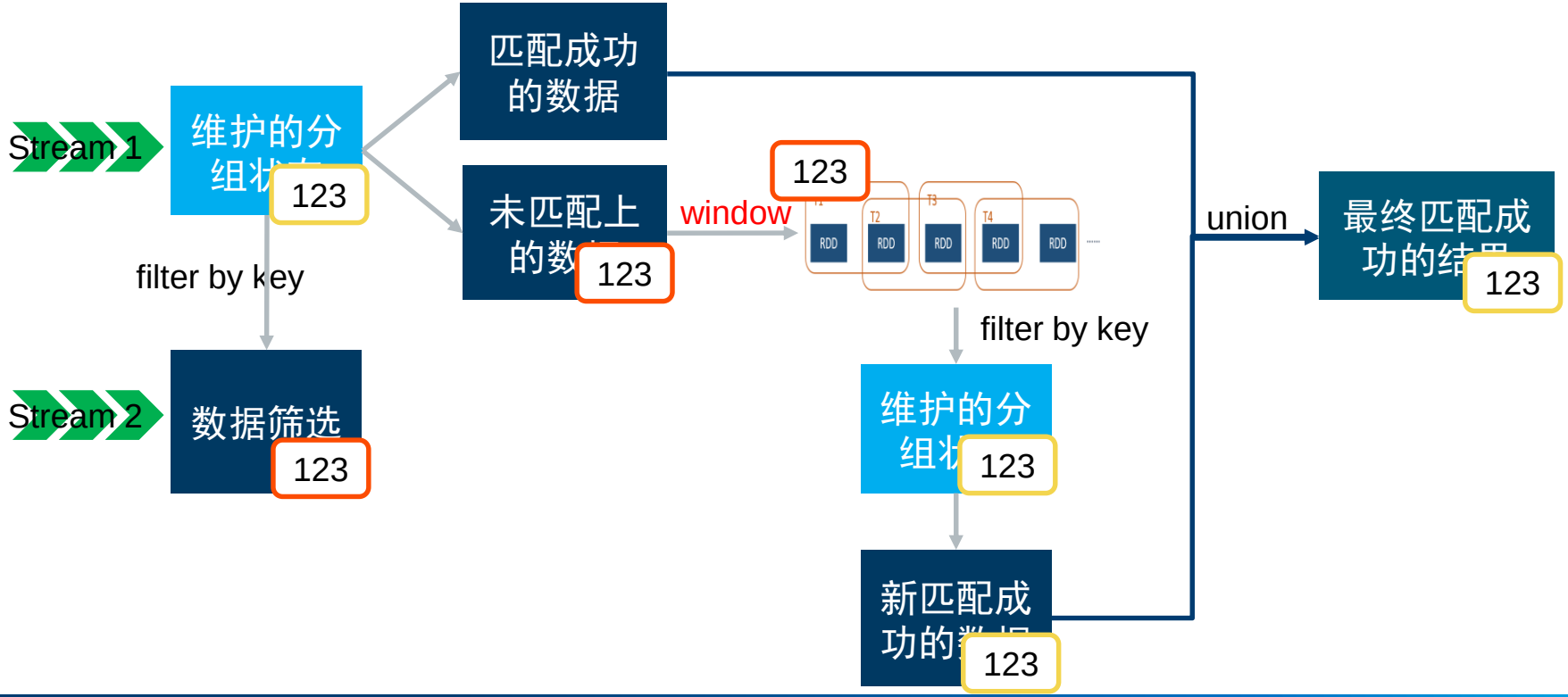
- Spark Streaming doesn’t support event time.

- Get approximate result with exist operator
  - mapWithState
  - window

Work Around: MapWithState

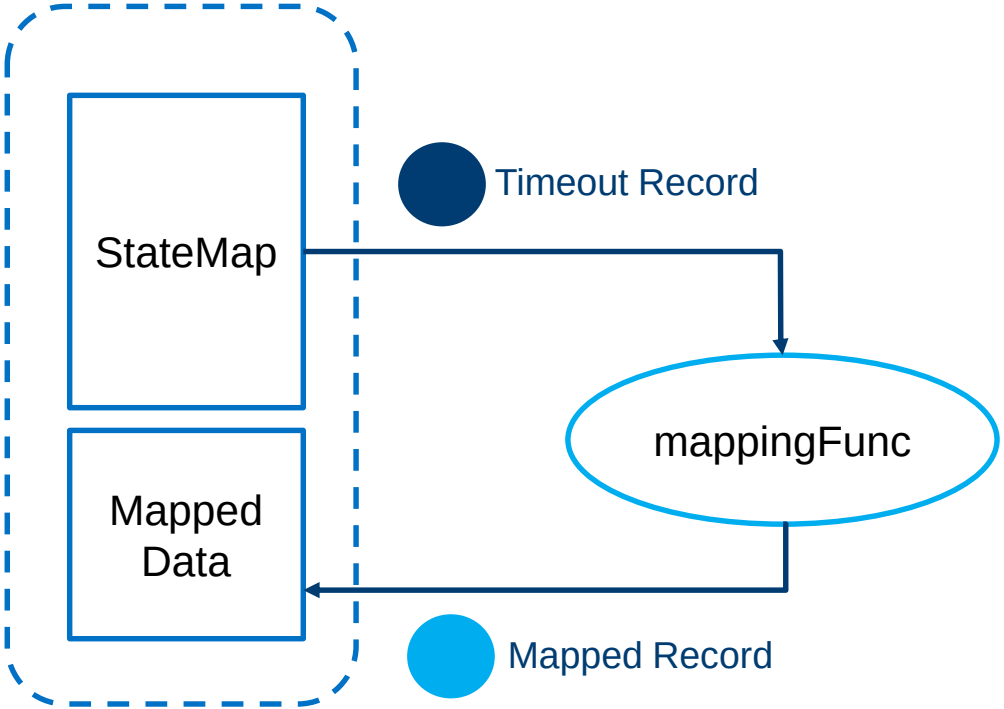


Work Around: Window



N.B. Timeout in MapWithState

- Timeout Records as part of output
- “isTimeout = true”
- Check before calling State.update()



N.B. Memory Concern

- Size of single cached RDD
  - org.apache.spark.util.collection.OpenHashMap
  - spark.streaming.sessionByKey.deltaChainThreshold

- Amount of all cached RDD
  - RememberDuration = 2 * CheckpointDuration
  - CheckpointDuration = 10 * BatchInterval
  - Worse with “window” operator.

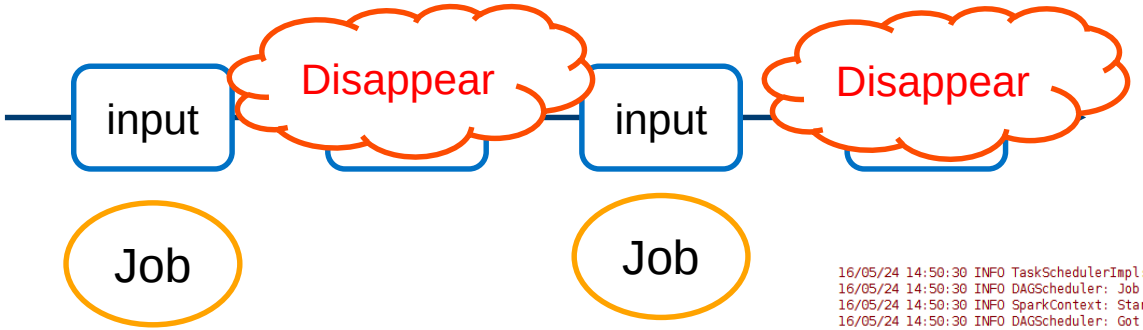
N.B. SqlContext recover from Checkpoint

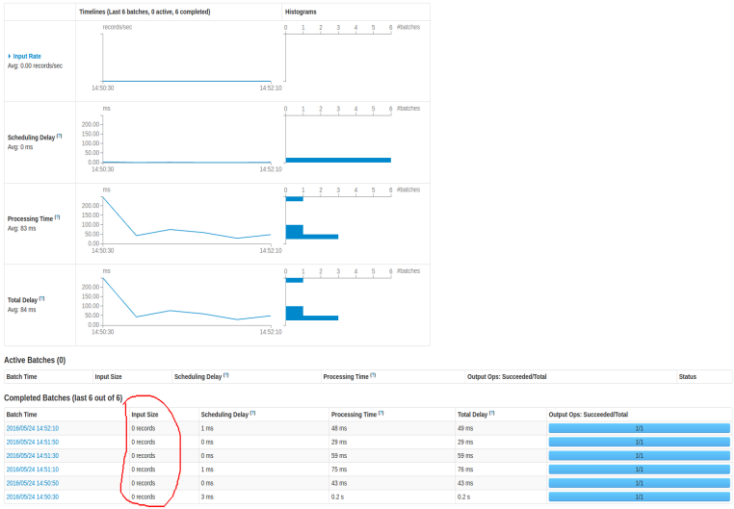
- SQLContext is NOT stored in Checkpoint.

- Soluation: Create SQLContextSingleton:
  - if (instance == null) => create a new SQLContext
  - if (instance != null) => return instance

N.B. Streaming UI Bug

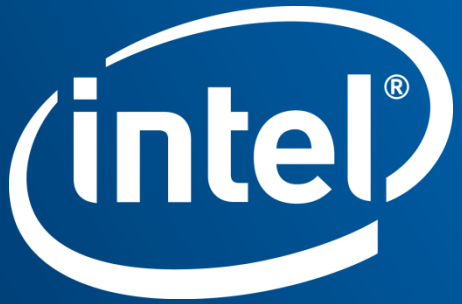
- Part of input “disappear” in UI
- SPARK-15480






```
16/05/24 14:50:30 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
16/05/24 14:50:30 INFO DAGScheduler: Job 0 finished: count at BugRepo.scala:33, took 0.209925 s
16/05/24 14:50:30 INFO SparkContext: Starting job: count at BugRepo.scala:36
16/05/24 14:50:30 INFO DAGScheduler: Got job 1 (count at BugRepo.scala:36) with 2 output partitions
16/05/24 14:50:30 INFO DAGScheduler: Final stage: ResultStage 1 (count at BugRepo.scala:36)
16/05/24 14:50:30 INFO DAGScheduler: Parents of final stage: List()
count number: 6
16/05/24 14:50:30 INFO DAGScheduler: Missing parents: List()
16/05/24 14:50:30 INFO DAGScheduler: Submitting ResultStage 1 (UnionRDD[2] at window at BugRepo.scala:30), which has no missing parents
16/05/24 14:50:30 INFO MemoryStore: Block broadcast_1 stored as values in memory (estimated size 3.2 KB, free 2.4 GB)
16/05/24 14:50:30 INFO MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 2039.0 B, free 2.4 GB)
16/05/24 14:50:30 INFO BlockManagerInfo: Added broadcast_1_piece0 in memory on 10.239.10.37:44450 (size: 2039.0 B, free: 2.4 GB)
16/05/24 14:50:30 INFO SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1012
```

We are making it better!


intel®

本文并未（明示或默示、或通过禁止反言或以其他方式）授予任何知识产权许可。

英特尔未做出任何明示和默示的保证，包括但不限于关于适销性、适合特定目的及不侵权的默示保证，及履约过程、交易过程或贸易惯例引起的任何保证。

本文件包含研发中的产品、服务和/或程序信息。这里提供的所有信息可在不通知的情况下随时发生变更。请联系您的英特尔代表，获得最新的预测、计划、规格和路线图。

描述的产品可能包含可能导致产品与公布的技术规格有所偏差的、被称为非重要错误的设计缺陷或错误。一经要求，我们将提供当前描述的非重要错误。

英特尔技术特性和优势取决于系统配置，并可能需要支持的硬件、软件或服务才能激活。更多信息，请见Intel.com，或从原始设备制造商或零售商处获得更多信息。

英特尔不控制或审计本文提及的第三方基准测试数据或网址。请访问提及的网站，以确认提及的数据是否准确。

在特定系统的特殊测试中测试组件性能。硬件、软件或配置的差异将影响实际性能。当您考虑采购时，请查阅其他信息来源评估性能。关于性能和基准测试程序结果的更多信息，请访问http://www.intel.com/performance

英特尔、英特尔标识是英特尔公司在美国和或其他国家的商标。

*其他的名称和品牌可能是其他所有者的资产。