腾讯社交广告 Tencent Social Ads | 广点通

CSDN CCTC 中国云计算技术大会 Cloud Computing Technology Conference

中国 Spark 技术峰会 2016

Spark Streaming 在腾讯广点通的应用

关于腾讯广点通

腾讯社交广告（*Tencent Social Ads*）

- 基于腾讯社交网络体系的广告平台

流量覆盖

- QQ 客户端、手机 QQ
- 微信
- QQ 音乐客户端、腾讯新闻客户端

主动型效果广告

- 有效实现更加智能的广告匹配和高效的广告资源利用

广告形式

- Banner 广告、插屏广告、开屏广告、信息流广告等

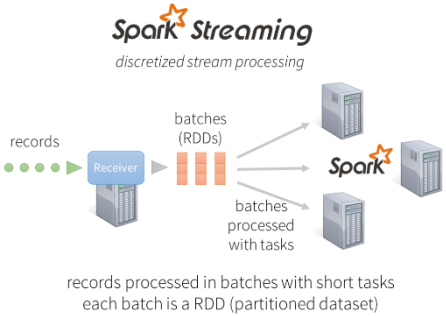
广点通与 Spark Streaming

Spark 0.x ~ 1.1

- 调研、试验

Spark 1.2 ~ 1.6

- 使用
- 1.2 加入了 driver 故障恢复



社区贡献

Spark Streaming 源码解析系列

「腾讯·广点通」技术团队荣誉出品

本系列内容适用范围：

```
* 2016.02.25 update, Spark 2.0 全系列 √ (2.0.0-SNAPSHOT 尚未正式发布)
* 2016.03.10 update, Spark 1.6 全系列 √ (1.6.0, 1.6.1)
* 2015.11.09 update, Spark 1.5 全系列 √ (1.5.0, 1.5.1, 1.5.2)
* 2015.07.15 update, Spark 1.4 全系列 √ (1.4.0, 1.4.1)
```

- *概述*
  - 0.1 Spark Streaming 实现思路与模块概述
- *模块 1：DAG 静态定义*
  - 1.1 DStream, DStreamGraph 详解
  - 1.2 DStream 生成 RDD 实例详解
- *模块 2：Job 动态生成*
  - 2.1 JobScheduler, Job, JobSet 详解
  - 2.2 JobGenerator 详解
- *模块 3：数据产生与导入*
  - 3.1 Receiver 分发详解
  - 3.2 Receiver, ReceiverSupervisor, BlockGenerator, ReceivedBlockHandler 详解
  - 3.3 ReceiverTraker, ReceivedBlockTracker 详解
- *模块 4：长时容错*
  - 4.1 Executor 端长时容错详解
  - 4.2 Driver 端长时容错详解
- *StreamingContext*
  - 5.1 StreamingContext 详解

Google 广点通 spark streaming

左侧图片来源：https://databricks.com/blog/2015/07/30/diving-into-spark-streamings-execution-model.html

Agenda

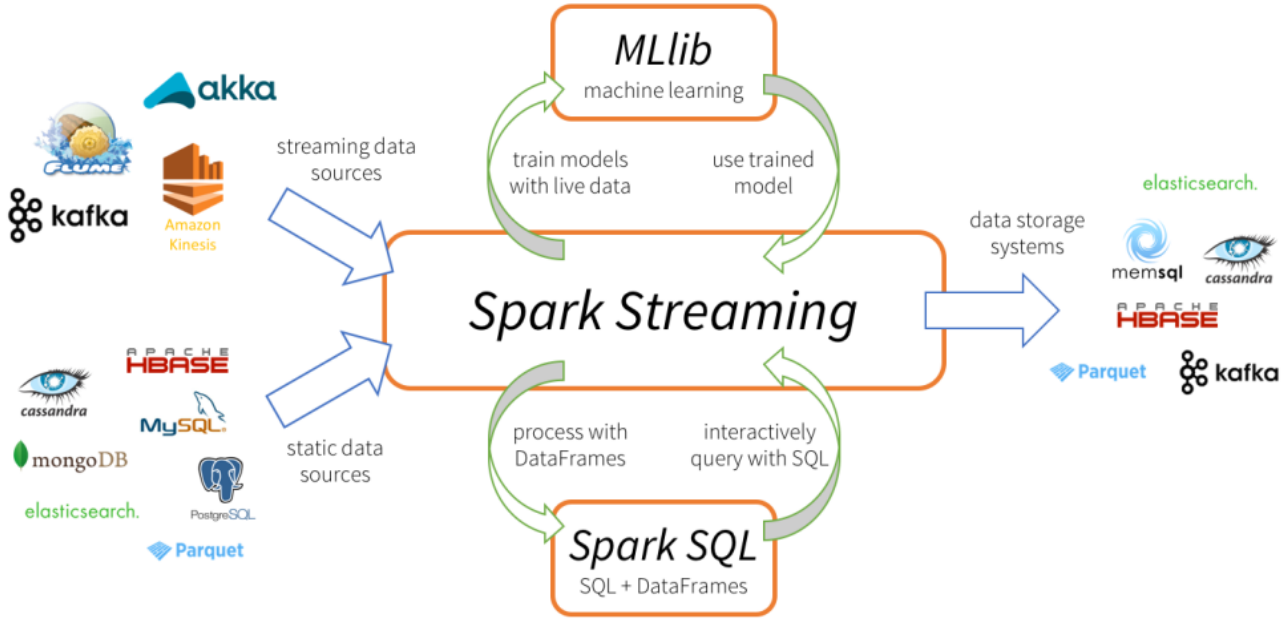
概述

- Spark Streaming 基本架构
- Spark Streaming at 广点通

特性与应用

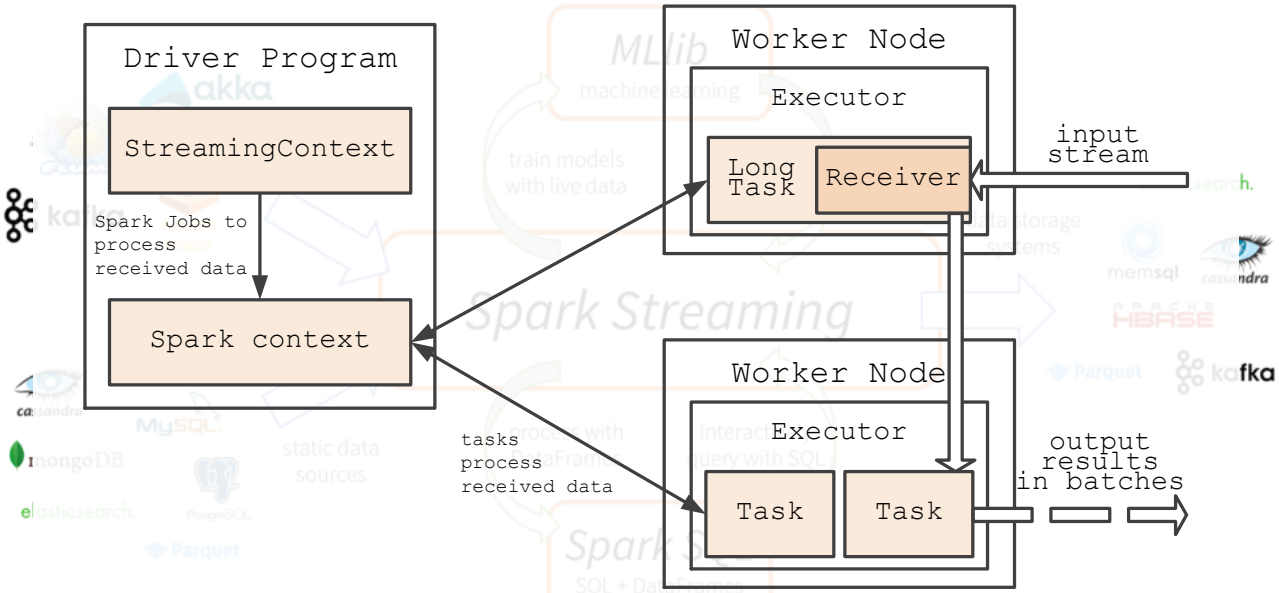
优化经验

Spark Streaming 基本架构



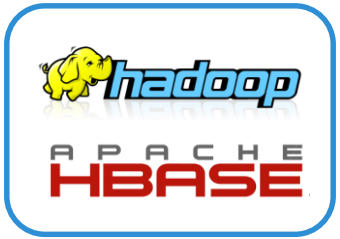
图片来源：http://www.datanami.com/2015/11/30/spark-streaming-what-is-it-and-whos-using-it/

Spark Streaming 基本架构



Spark Streaming at 广点通

腾讯 Spark 技术栈（powered by 数据平台部）



TDBank

message queue

Spark Streaming
监控与拉起

Spark Streaming at 广点通

#apps	#cores	data volume	peak rate
~30	~2000	~70 TB/day	~600 k/sec

Agenda

概述

特性与应用

- (1) exactly-once
- (2) 可靠状态
- (3) batch 调度

优化经验

特性与应用-(1) exactly-once

Spark 执行单元

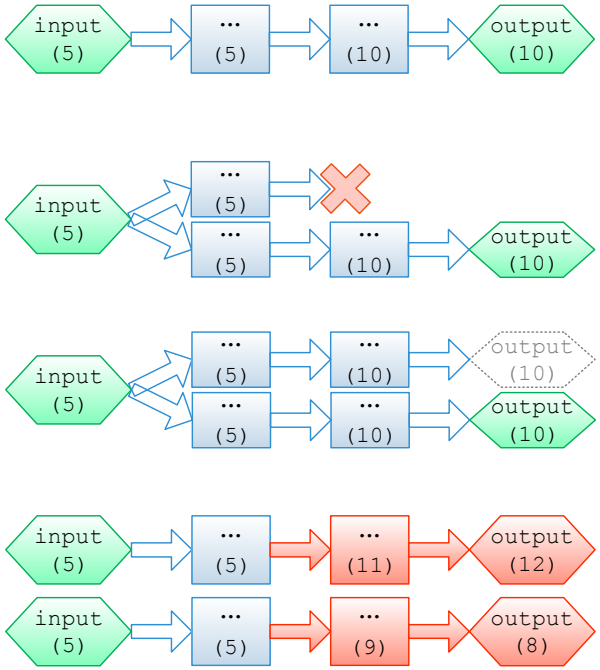
- 任务（即一批数据）
- 一批数据全部成功/全部失败

Task 重做

- 失败重做：task 重做、stage 重做
- 推测执行：另一个节点同时做
- Committer: 任务唯一成功

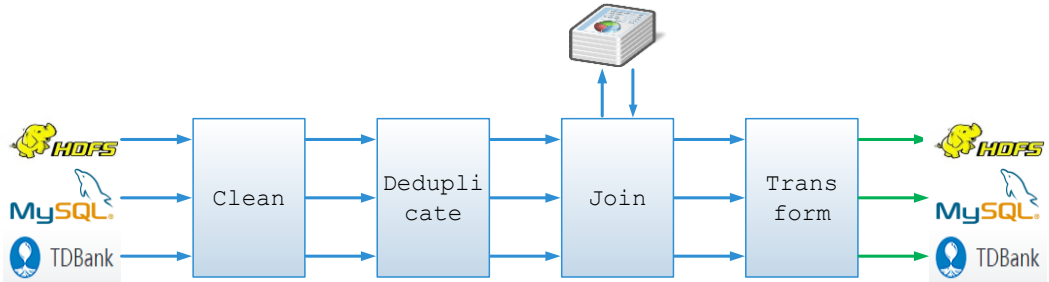
其它系统

- Storm: at-most-once
  at-least-once
- MapReduce: exactly-once

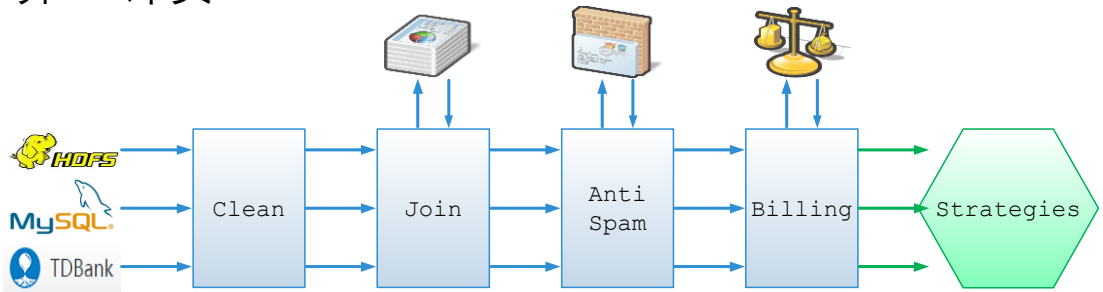


特性与应用-(1) exactly-once

应用：实时准确数据转移



应用：反作弊 + 计费！！！



特性与应用-(2) 可靠状态

Spark Streaming 天然面向状态

- RDD: Resilient Distributed Datasets
- RDD lineage
  - 容错：重做
- rdd.checkpoint()
  - HDFS, S3… 等

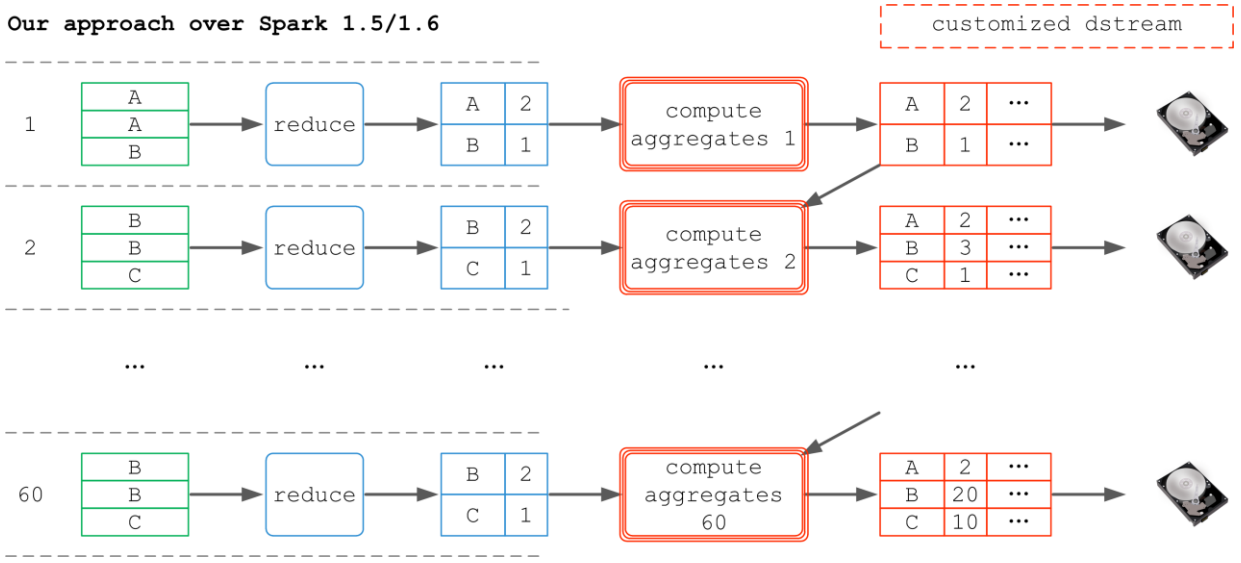
可靠状态管理

- Spark 1.5: updateByKey()
- Spark 1.6: [kv store] mapWithState()
- Spark 2.0: [kv store] StateStore

图片来源：https://databricks.com/blog/2015/06/22/understanding-your-spark-application-through-visualization.html

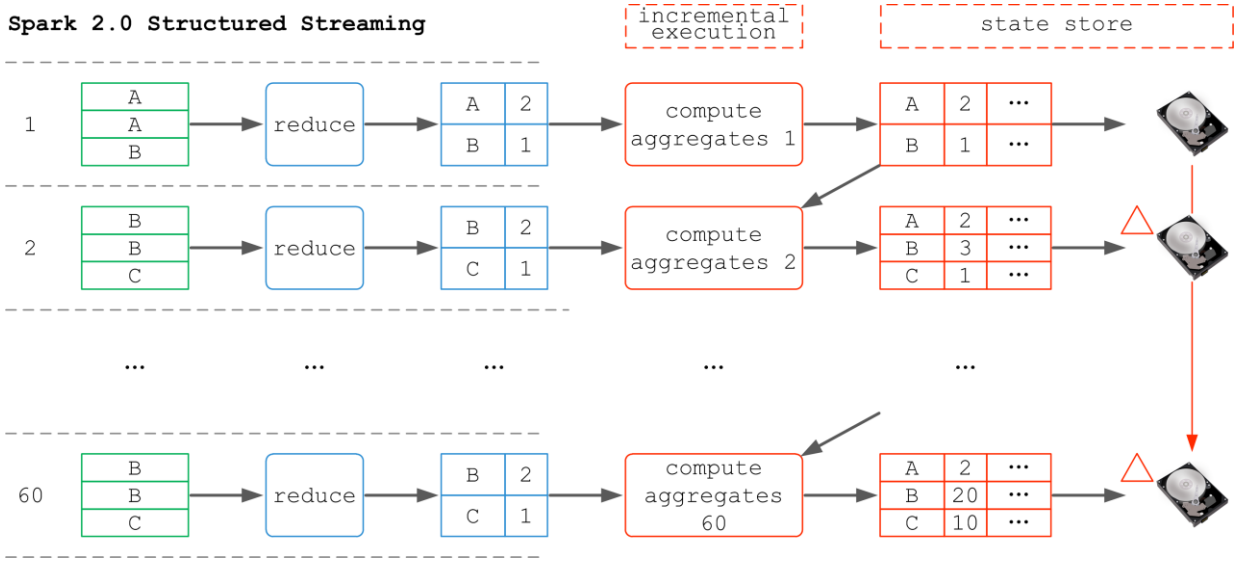
特性与应用-(2) 可靠状态

应用：跨 batch 聚合（pv/uv 计算，记录去重，微额记账等）



特性与应用-(2) 可靠状态

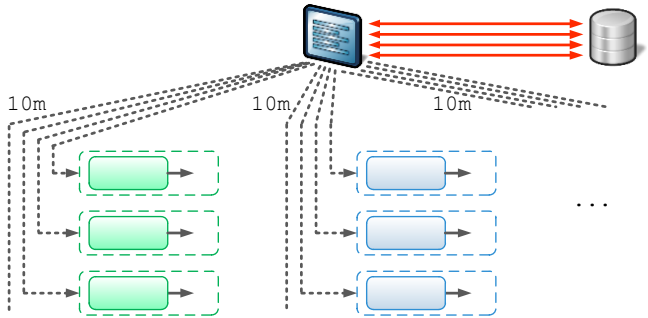
应用：跨 batch 聚合（pv/uv 计算，记录去重，微额记账等）



特性与应用-(3) 快速 batch 调度

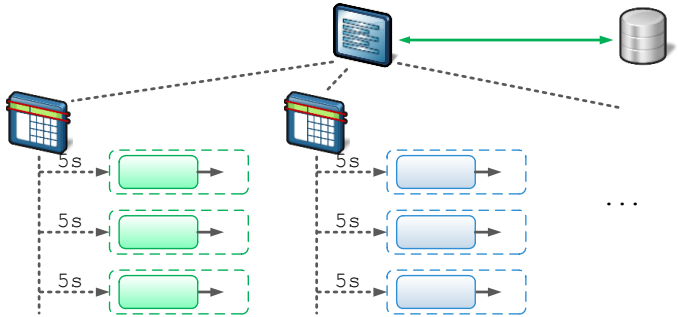
MapReduce 的实例调度

- 一级调度系统（oozie 等）
  - 最小间隔 10 min
  - 进程调度
  - 需要一定启动时间



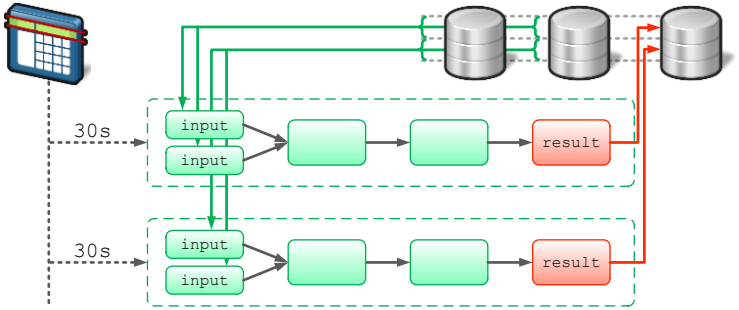
Spark Streaming 的 batch 调度

- 一级调度系统（oozie 等）
- 二级调度系统
  - Driver / JobScheduler 调度
  - 间隔 1s ~ 60s
  - 进程常驻+线程调度、无启动时间

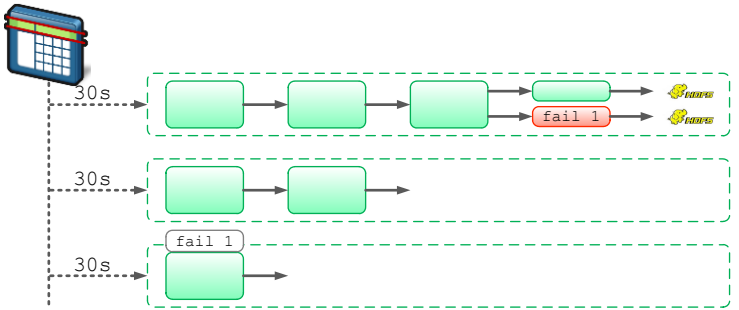


特性与应用-(3) 快速 batch 调度

应用：数据指标监控



应用：复杂 pipeline 的
未成功数据唯一快速重试



Agenda

概述

特性与应用

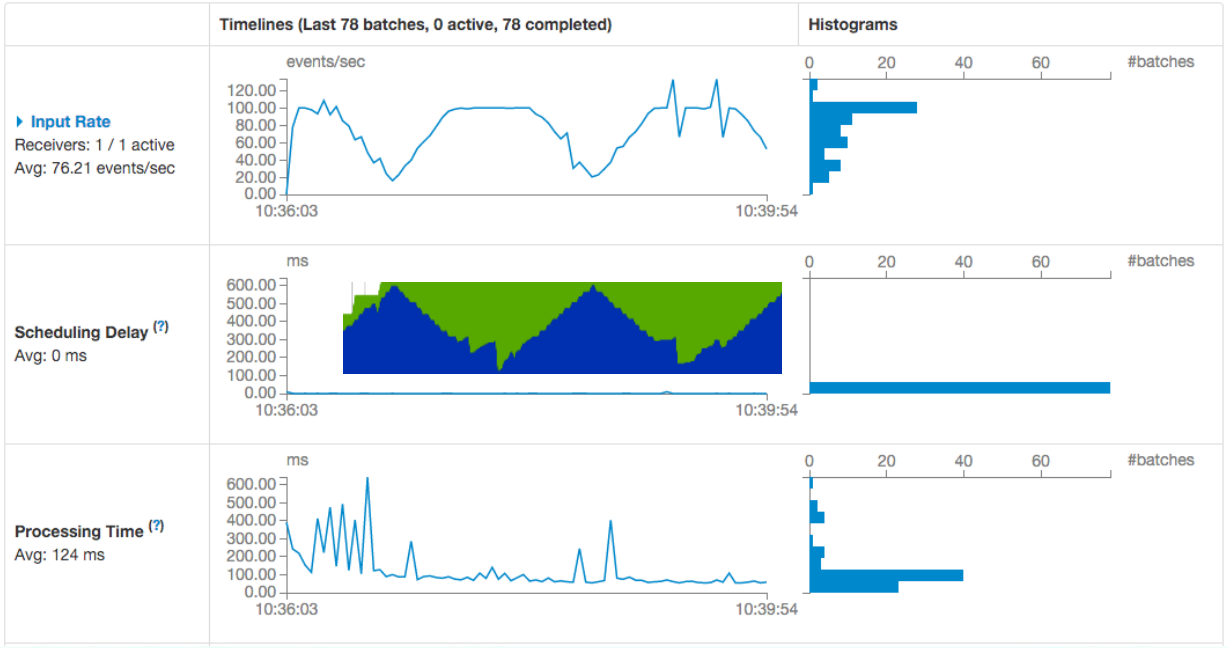
优化经验

- (1) 增加 Memory Back Pressure
- (2) 为 Spark 增加新特性（无需编译 Spark 优化）
- (3) SparkSQL API > RDD API
- (4) async execution within a task
- (5) try-cath
- (6) concurrentJobs 开启
- (7) Spark 远程调试

优化经验

(1) 增加 Memory Back Pressure

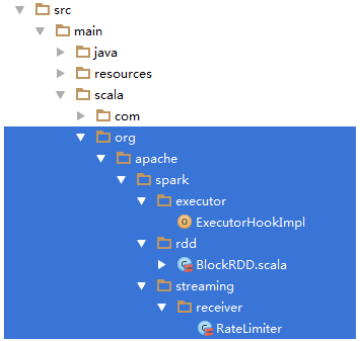
- Receiver 的接收速率随 Executor 的内存使用动态放缩，避免 OOM



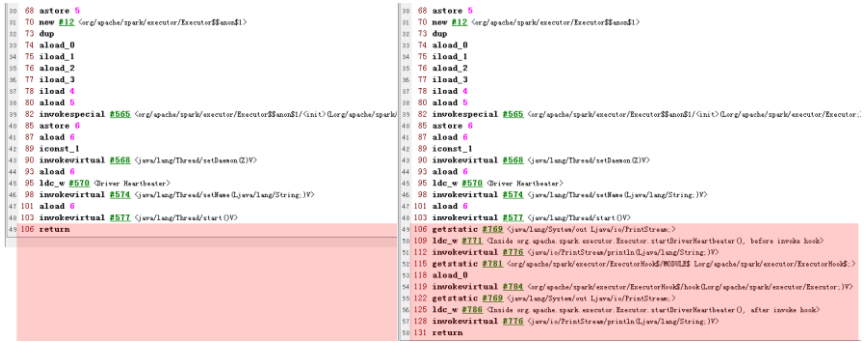
优化经验

(2) 为 Spark 增加新特性（无需编译 Spark 工程）

- A. 直接改源文件 ***.scala
  - src 原包名下，如 src/o/a/s/streaming/receiver/RateLimiter.scala
  - 运行参数：spark.driver/executor.userClassPathFirst=false; spark.driver/executor.extraClassPath=app.jar
- B. 直接改字节码 ***.class
  - resources 原包名下，如 resources/o/a/s/executor/Executor.class
  - 运行参数：spark.driver/executor.userClassPathFirst=false; spark.driver/executor.extraClassPath=app.jar



A

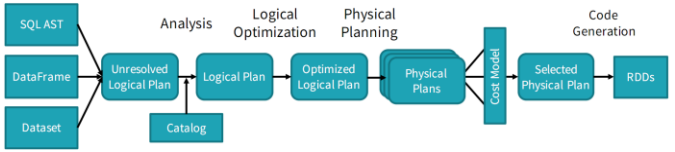


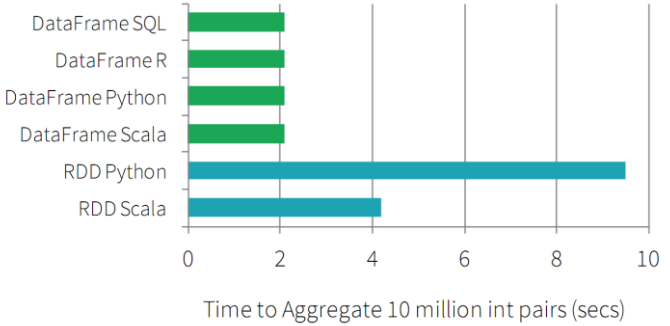
B

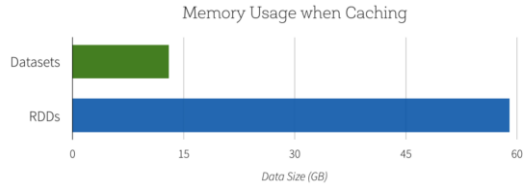
优化经验

(3) SparkSQL API > RDD API

- Run faster
  - Catalyst Optimizer
  - Tungsten Engine
    - Memory Management
    - Cache-aware Algorighms
    - Whole Stage Codegen
- Spark 1.x
  - ```
    dstream.foreachRDD { rdd =>
      rdd.toDF().select…
    }
    ```
- Spark 2.x: Structured Streaming
  - ```
    spark.….stream.….startStream()
    ```





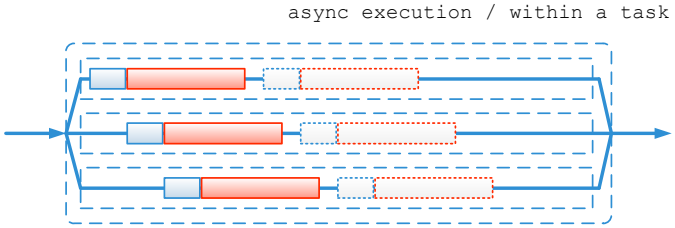


图片来源：Real-time Spark: From interactive queries to streaming

优化经验

(4) async execution within a task

- 应对外部操作：线程池+异步



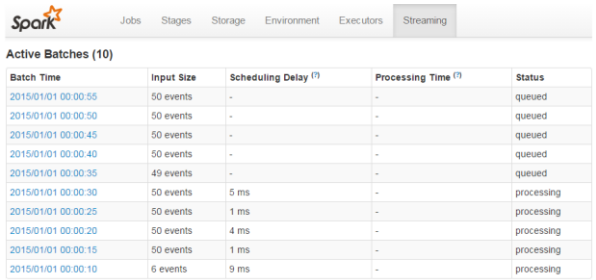
(5) try-catch

- task 的错误，会在 driver 端抛出
- 屏蔽 "Could not compute split" 等问题

```
val inputDStream = ssc.fileStream("...")
inputDStream.foreachRDD(rdd => {
  try {
    // do something
  } catch {
    case NonFatal(e) => errorHandler(e)
  }
})
```

(6) concurrentJobs 开启

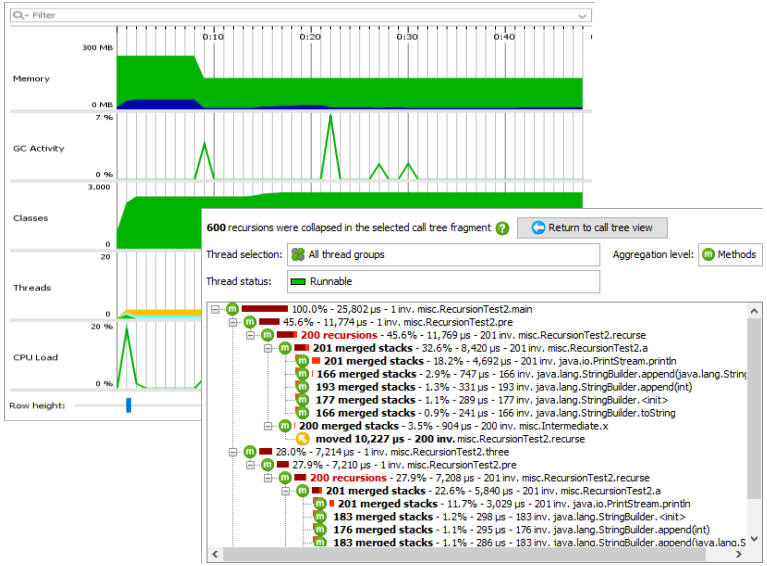
- 内部参数，同时执行 n 个 output
  - 一般 1 个 batch 对应 1 个 output
- spark.streaming.concurrentJobs = n



优化经验

(7) Spark 远程调试

- your profiler 按需分发



```
73  private def cmds() = {
74    val LS = "ls -trhl;"
75    Array(
76      s"mkdir ${dirPath};",
77      LS,
78      s"cd ${dirPath};",
79      s"cp -L ../${srcJarPath} ${desJarPath};",
80      LS,
81      s"jar -xf ${desJarPath};",
82      LS,
83      s"tar zxf ${resourcesPath};",
84      LS,
85      s"cmd to enable your profiler..."
86    )
```

图片来源：https://www.ej-technologies.com/products/jprofiler/whatsnew91.html

总结与展望

总结

特性	应用
(1) exactly-once	实时准确数据转移 反作弊 + 计费！！！
(2) 可靠状态	跨 batch 聚合 (pv/uv 计算，记录去重，微额记账等)
(3) 快速 batch 调度	数据指标监控 复杂 pipeline 的未成功数据唯一重试

展望

- From Lambda Architecture(MR + Storm) to Spark Streaming
- Spark 2.0: Structured Streaming
  - High-level streaming API built on Spark SQL engine
  - Event time, windowing, sessions, sources & sinks

Thanks!