Flink在小米的应用与实践

夏军 · 小米 / 高级工程师

Apache Flink China Meetup 北京
- 2019年09月21日

Contents

目录 >>

PART 01

小米流计算演进

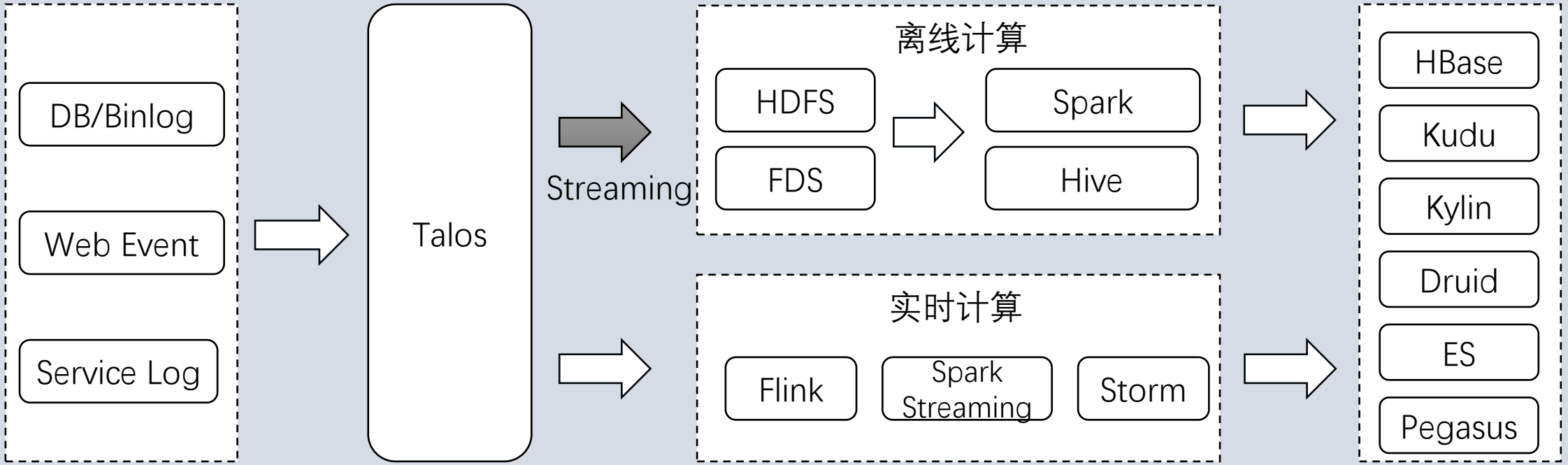
小米流计算演进

Apache Flink

2014.02	2015.07	2019.01

基本架构

Apache Flink



集群概况

Apache Flink

7000亿	150+	7w+	450TB +
日处理数据	作业规模	CPU	内存

PART 02

小米相关改造

Talos适配

Apache Flink

Talos简介

Talos是小米自研的消息队列，提供标准的Topic语义封装，用于实现业务上下游的数据解耦。

目前已经应用于小米业务的各个场景，也为Streaming计算提供数据源支持。

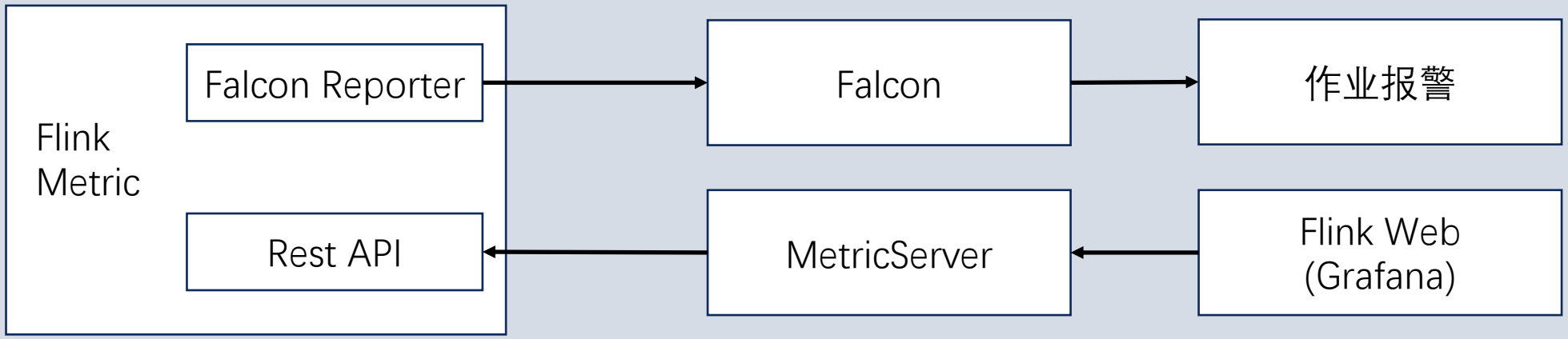
Flink Talos Source

- FlinkTalosConsumer，TalosTableSource
- 自动follow上游Partition变化
- 基于Thrift进行反序列化

Flink Talos Sink

- FlinkTalosProducer，TalosTableSink
- 默认Partitioner支持

监控与报警

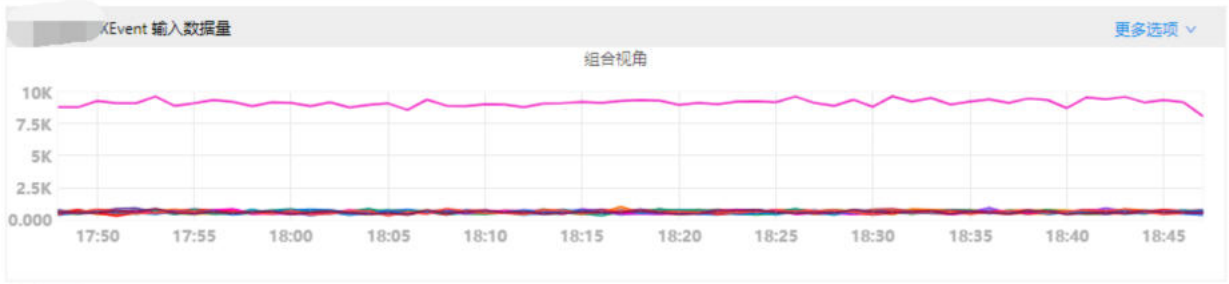


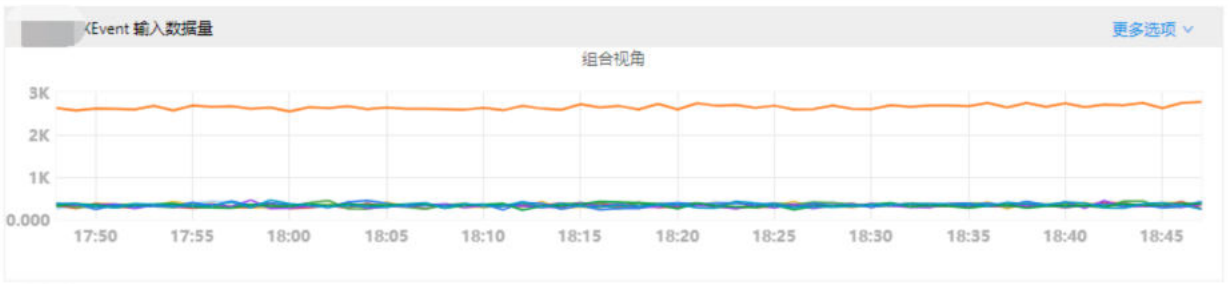
Falcon数据：通过定制化FalconRepoter，将数据推动到Falcon实现数据监控与报警，主要包括消费积压、作业异常等需要报警的case。

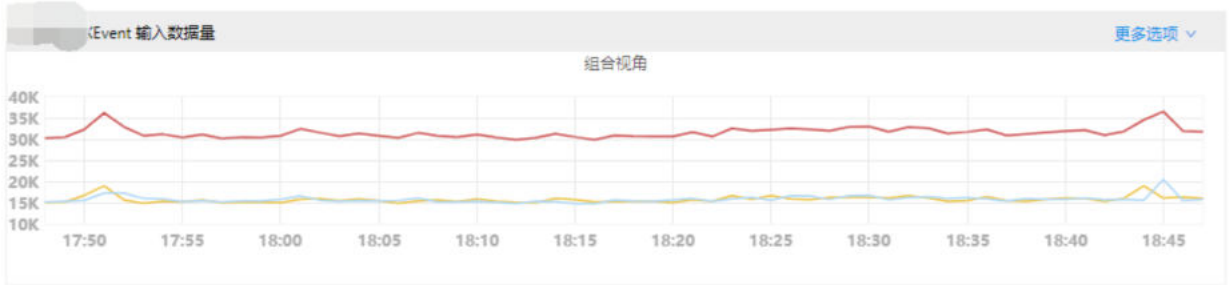
Flink Web Metric数据展示：通过定制化MetricServer将Flink Metric数据转换为Json，并实现数据基于timestamp的缓存。FlinkWeb通过嵌入Grafana页面，实现Metric数据的定制化展示。主要包括 流量信息，checkpoint信息，jvm信息，gc信息等。

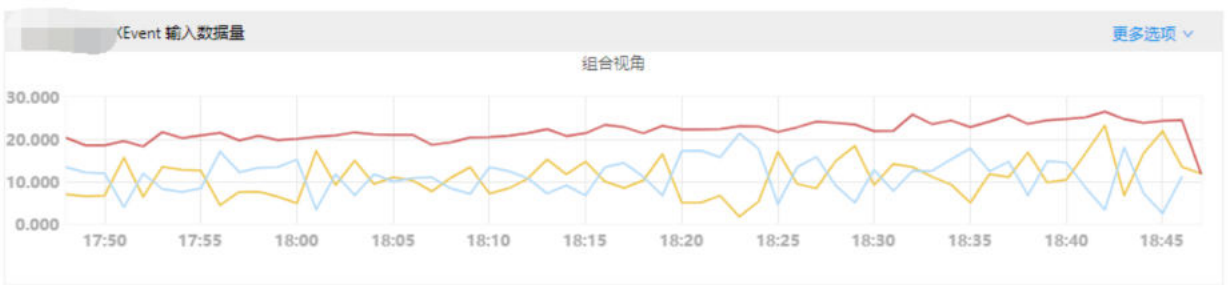
监控与报警

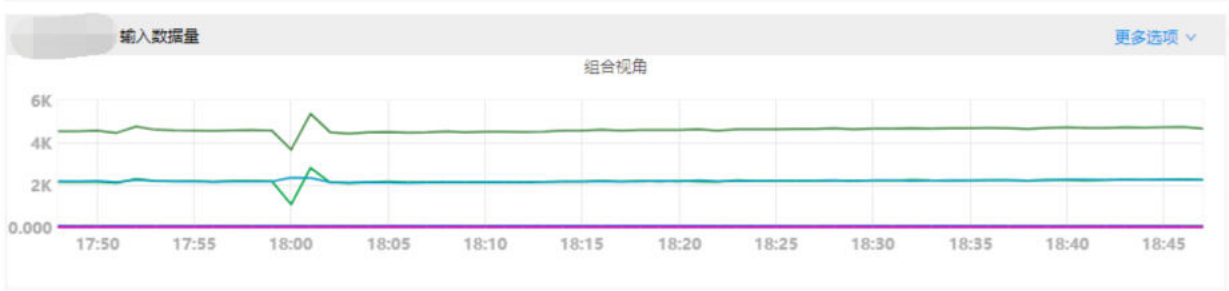
Apache Flink

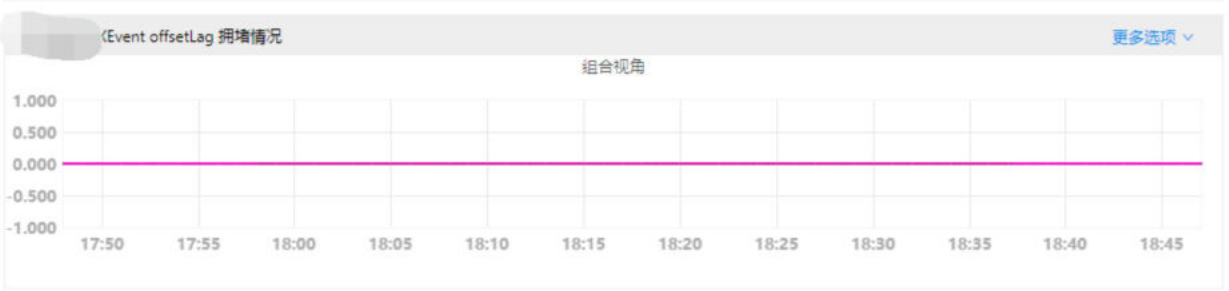






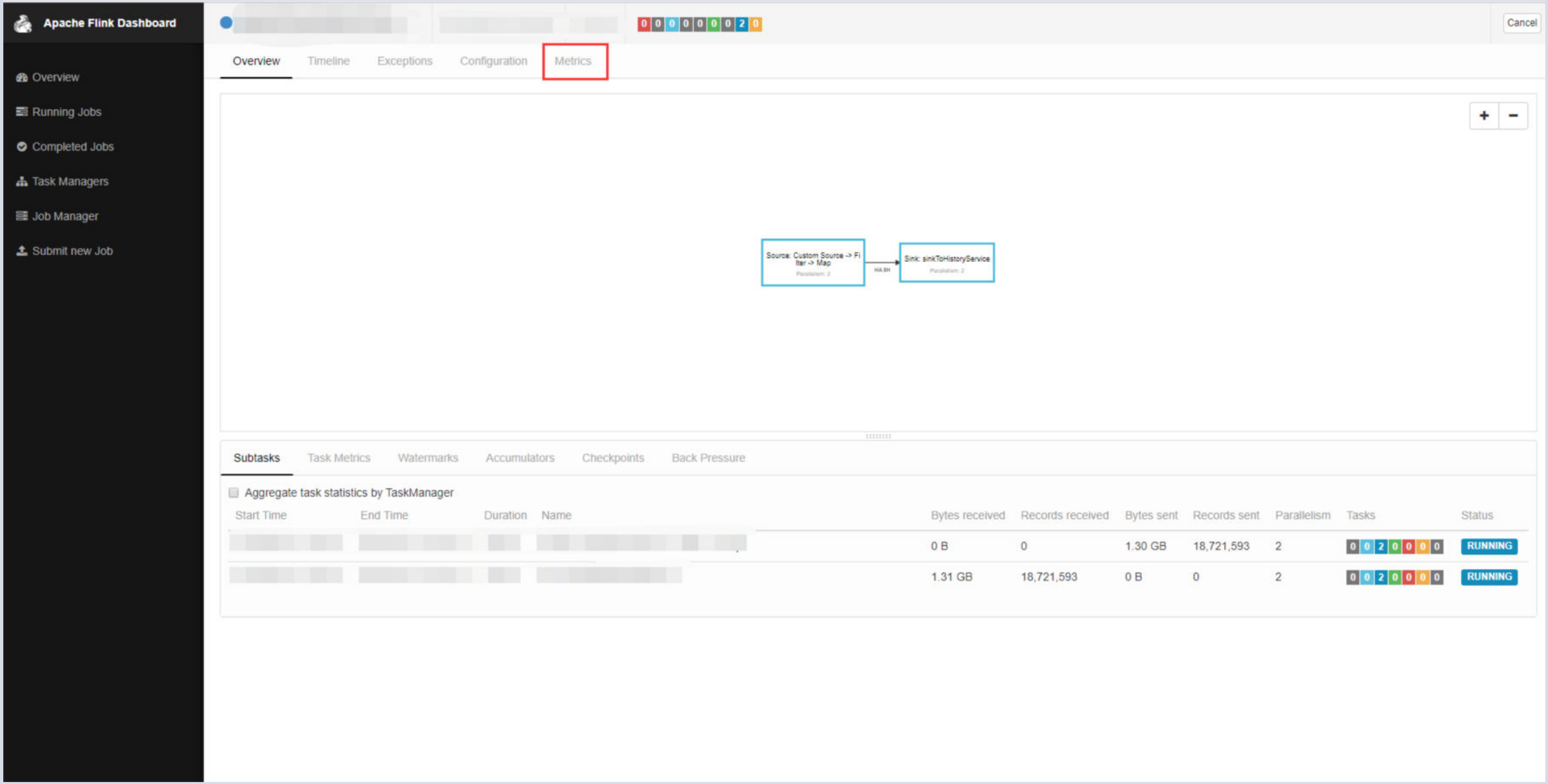






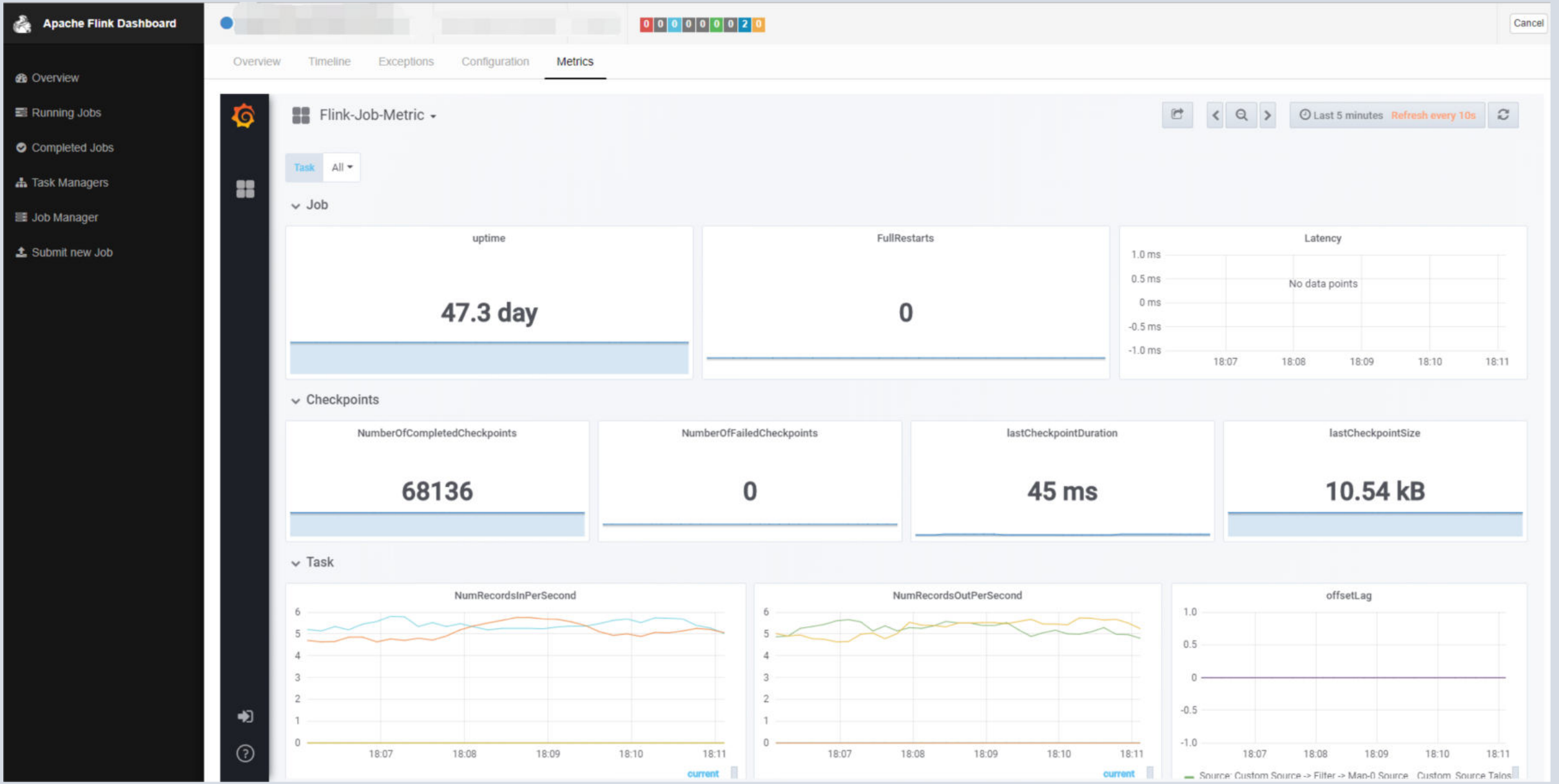
监控与报警

Apache Flink



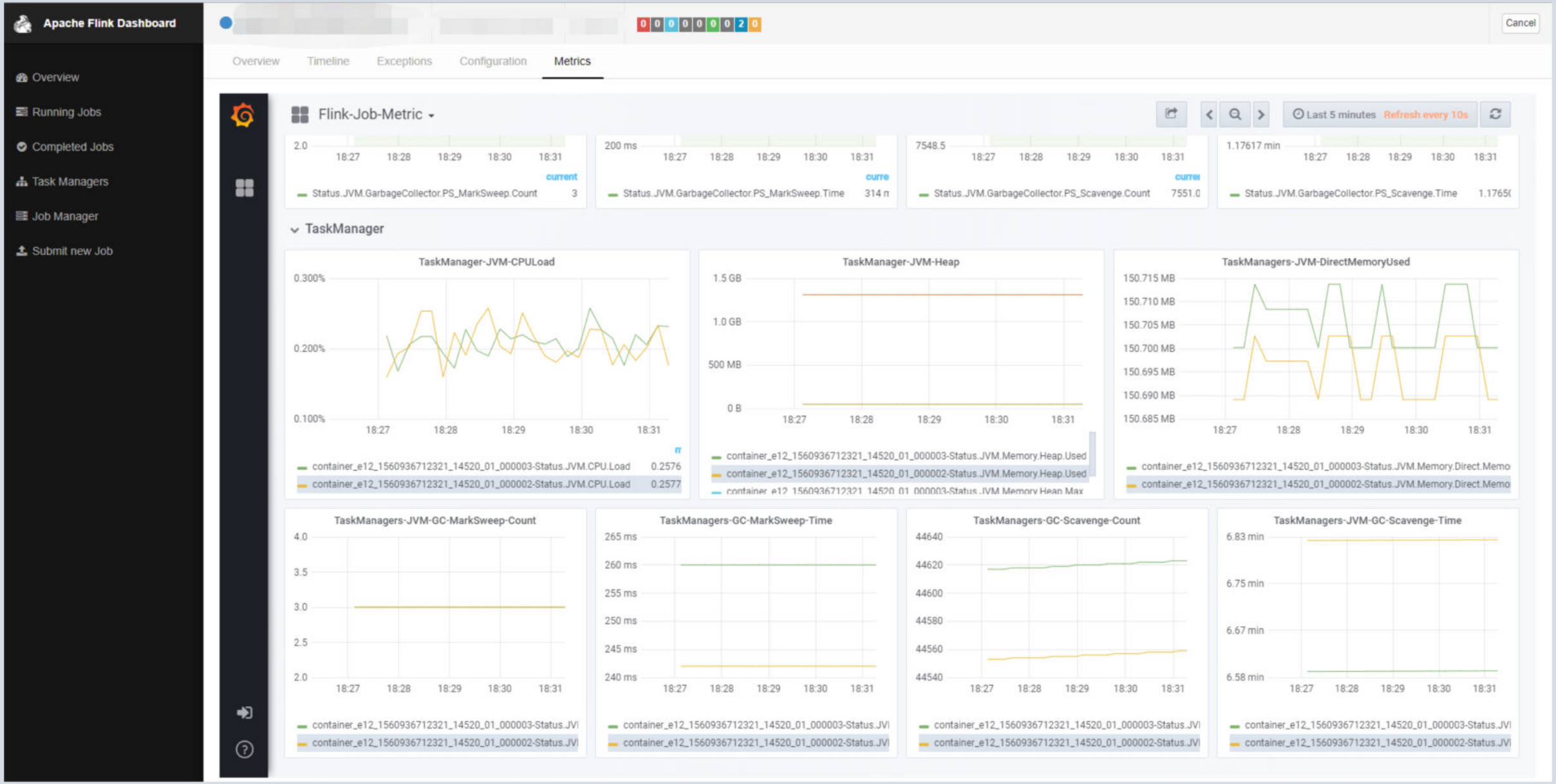
监控与报警

Apache Flink

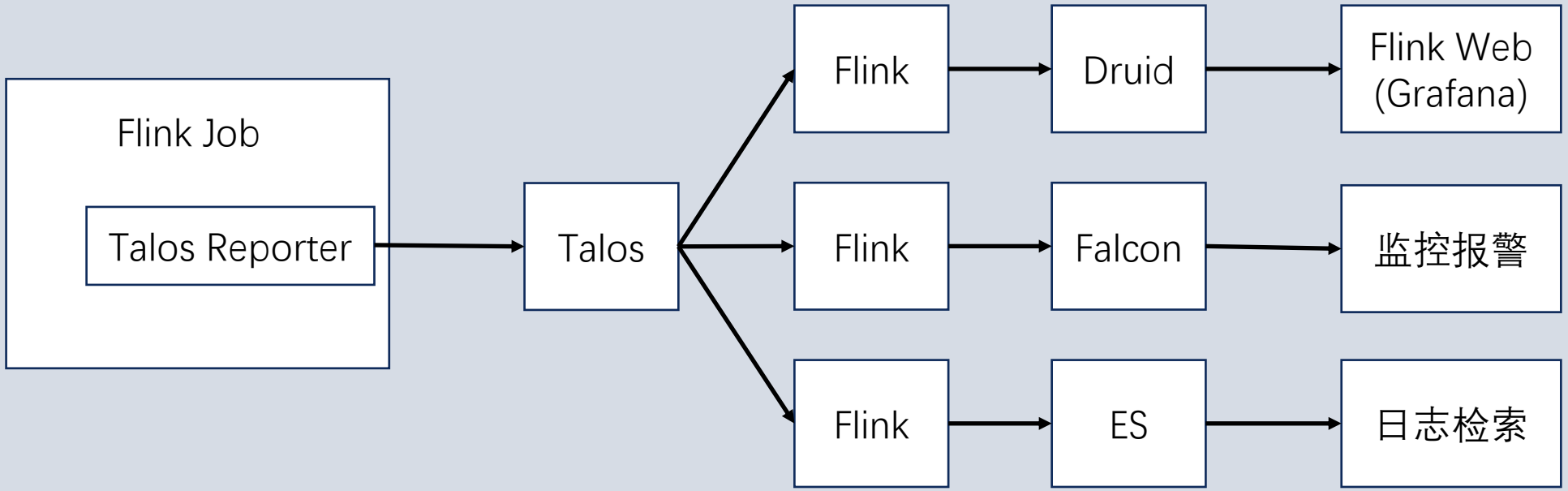


监控与报警

Apache Flink



Flink数据体系（重构中）



数据收集：所有数据通过Reporter收集到消息队列Talos

Metric展示：Talos + Flink + Druid + Grafana的方式实现Metric数据展示；

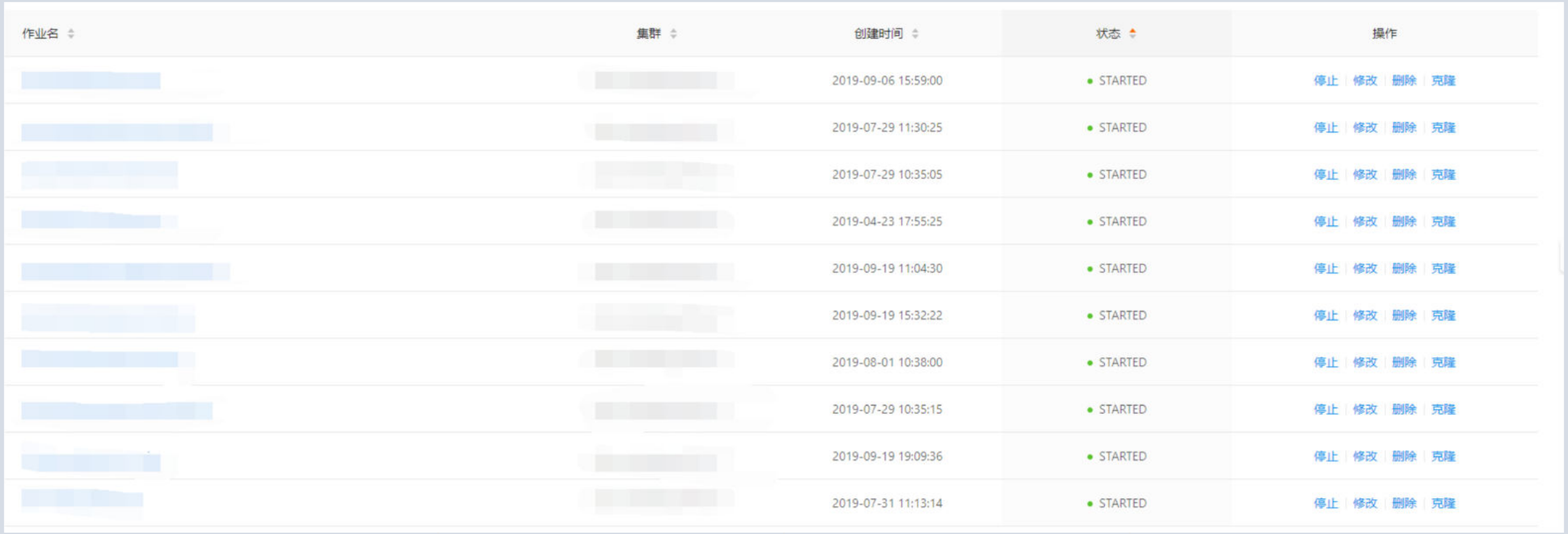
监控报警：Talos + Flink + Falcon的方式实现监控报警；

日志检索：Talos + Flink + ES的方式实现日志检索；

上述三种方案均会和Flink Web进行整合；

作业管理

Apache Flink



作业名	集群	创建时间	状态	操作
		2019-09-06 15:59:00	STARTED	停止 | 修改 | 删除 | 克隆
		2019-07-29 11:30:25	STARTED	停止 | 修改 | 删除 | 克隆
		2019-07-29 10:35:05	STARTED	停止 | 修改 | 删除 | 克隆
		2019-04-23 17:55:25	STARTED	停止 | 修改 | 删除 | 克隆
		2019-09-19 11:04:30	STARTED	停止 | 修改 | 删除 | 克隆
		2019-09-19 15:32:22	STARTED	停止 | 修改 | 删除 | 克隆
		2019-08-01 10:38:00	STARTED	停止 | 修改 | 删除 | 克隆
		2019-07-29 10:35:15	STARTED	停止 | 修改 | 删除 | 克隆
		2019-09-19 19:09:36	STARTED	停止 | 修改 | 删除 | 克隆
		2019-07-31 11:13:14	STARTED	停止 | 修改 | 删除 | 克隆

通过流式作业管理平台实现了Flink Streaming作业的提交/停止/删除/克隆等基本操作，同时支持作业运行历史，Flink Web连接等集成。

作业管理

Apache Flink

创建作业

Please select

* 作业名：

1~80个字符，只能包含数字、字母、"_"和"-"，不能以"_"或"-"开头

* Kerberos账户：

* Yarn队列： root.default

* Driver Memory： 1G

示例：512M,1G

* Driver Cores： 1

* Number Executors： 1

* Per Executor Memory： 1G

示例：512M,1G

* Per Executor Cores： 1

其他框架参数：

spark.yarn.driver.memoryOverhead=512
spark.yarn.executor.directMemoryOverhead=512

PART 03

小米典型应用

Hive数据转储

Apache Flink

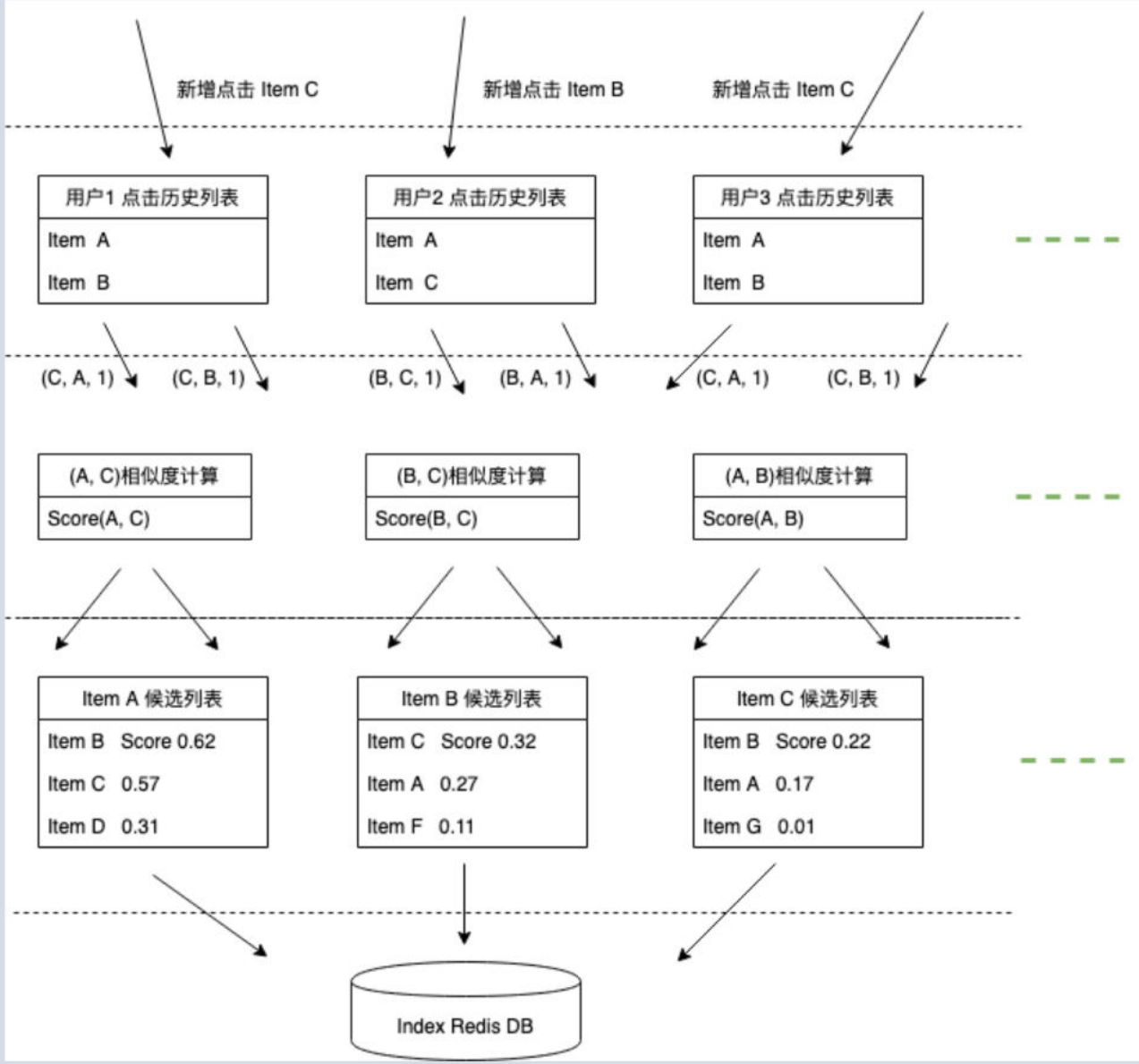
基于Message Event进行Hive partition选择，彻底解决数据延迟带来的Partition错乱问题

基于BucketingSink 重写Buckter实现

```
val fileSink = new BucketingSink[Tuple2[BytesWritable, BytesWritable]](defineOutputHDFSFile)
fileSink.setBucketer(new MiddleDataDateTimeBucketer[Tuple2[BytesWritable, BytesWritable]])
fileSink.setWriter(new SequenceFileWriter[BytesWritable, BytesWritable]())
fileSink.setBatchSize(1024 * 1024 * 400) // 每个part超过 400MB 就截断
fileSink.setBatchRolloverInterval(20 * 60 * 1000); // 超过 20min 也截断

outputStream.startNewChain().addSink(fileSink).name("rollingHdfsSink")
```

实时 Item CF索引



更新频率：Spark Streaming 天级更新 + 小时级更新 => Flink 秒级更新

收益：图文浏览时长提升15s，实时度提升50%

峰值QPS：7w/s

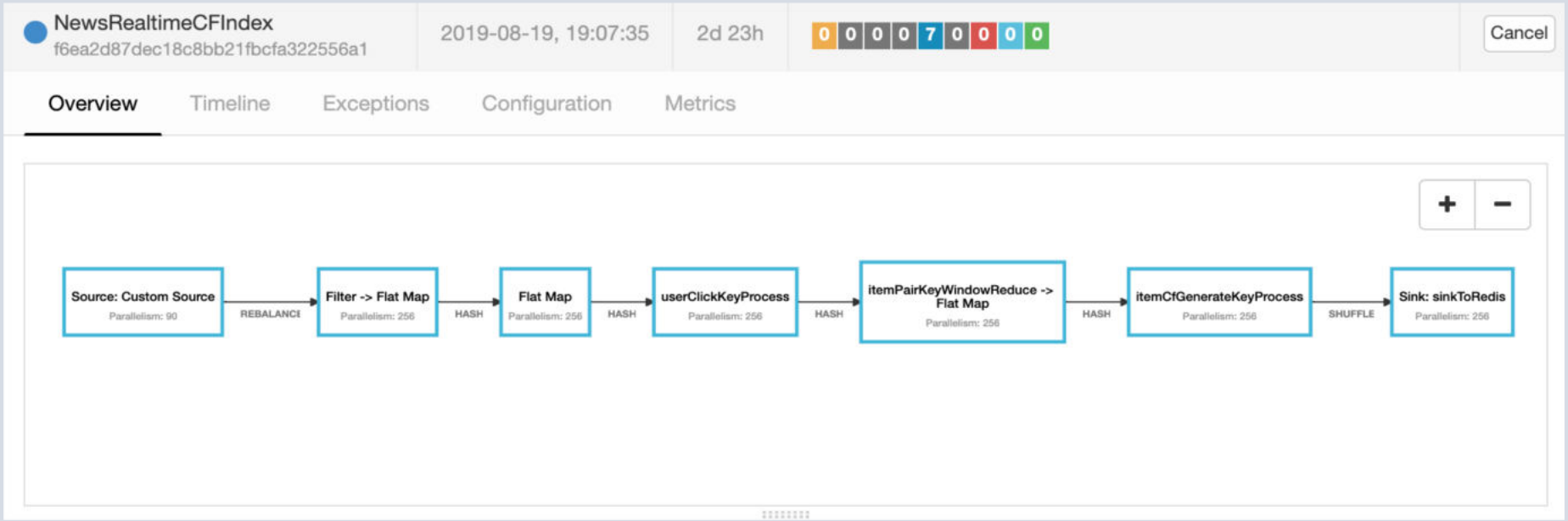
State：27GB

实时 Item CF索引

Apache Flink

```
val itemCfStream = datasource.filter(defineEntryProcessor.filterLogLine(_)).setParallelism(env.getParallelism)
  .flatMap(defineEntryProcessor.extractUserClickEntry(_))
  .keyBy(_.itemId)
  .flatMap(new HbaseFlatMapFunction)
  .keyBy(_.userId)
  .process(new UserRecentClickFunction()).name("userClickKeyProcess")
  .keyBy(defineEntryProcessor.generateSortedPairHashKey(_))
  .timeWindow(Time.seconds(5))
  .reduce((a, b) => defineEntryProcessor.reduceToSortedItemPairFunction(a, b)).uid("itemPairKeyWindowReduce")
  .flatMap(defineEntryProcessor().transformItemPairToItemSequence(_))
  .keyBy(_.itemA)
  .process(new ItemCFGenerateFunction(defineEntryProcessor())).uid("itemCfGenerateKeyProcess")

//sink to index Redis
itemCfStream.shuffle.addSink(new IndexSinkDatabaseFunction(defineBusinessType())).name("sinkToRedis")
```



实时Impression拼接

State大小：800GB+

结果数据：17TB/天

基于connect，CoProcessFunction和Timer实现

PART 04

未来规划

未来规划

Apache Flink

- Flink 1.9 + Flink SQL
- 平台化建设
- 资源优化 + 动态调度
- 更多的社区参与

FLINK FORWARD

Organized by 阿里云

Created by ververica

Flink Forward Asia

全球最大的 Apache Flink 官方会议

预计 **2000+** 参会人员， 2019年11月28-30日 @北京国家会议中心

国内外一线厂商悉数参与

阿里巴巴、腾讯、字节跳动、intel、 DellEMC 、Uber、美团点评、Ververica ...

大会官网，查看更多

Apache Flink

Apache Flink 社区微信公众号「 Ververica」

Meetup动态 / Release 发布信息 / Flink 应用实践

THANKS

Apache Flink China Meetup

· BEIJING