

阿里巴巴实时计算平台JStorm Turbo

封仲淹@alibaba

Background



- Apache Storm PMC（管理委员会成员）
- 中国最早的Storm团队
 - Storm 0.5.1/0.5.4/0.6.0/0.6.2/0.7.0/0.7.1
 - Jstorm
0.7.1/0.9.0/0.9.1/0.9.2/0.9.3/0.9.3.1/0.9.4/0.9.4.1/0.9.5/0.9.5.1/0.9.6/0.9.6.1/0.9.6.2/0.9.6.3/0.9.7/0.9.7.1/0.9.7.2/0.9.8/0.9.8.1/2.0.4/2.1.0/2.1.1
- 公司职责：
 - 应用开发
 - Jstorm 平台开发
 - jstorm/storm技术支持
 - 运维所有集群

1

2

3

4

5

业务场景

架构

创新点

下一步

未来

jstorm 最大的5个业务方 (每日流入1.81PB)

1

2

3

4

5

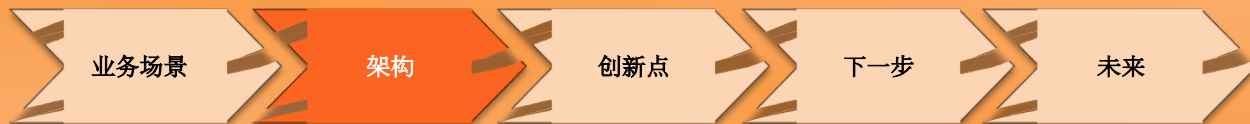
| 业务 | 业务介绍 |
|----------------------|---|
| Alimama | 钻展结算 / 直通车/防恶意平台 / 搜索营销 / basecamp / 精准定向/ |
| Eagleye & Tlog | 实时跟踪线上十几w机器日志，eagleye链路分析，监控调用链Tlog，实时分析日志，产生系统监控报表 |
| Nut & velocity | 集团和小微 风控实时监控系統，防止资损 |
| RDS/DRC/Oceanbase 监控 | 反数据库入侵/RDS/数据订正 / 库存对账 / drc 实时监控 / oceanbase 指标监控 |
| 支付宝 海纳百川 & mmonitor | 无线监控平台， 商户业务监控平台, 新春红包后台服务监控 |

业务场景分类

1 2 3 4 5

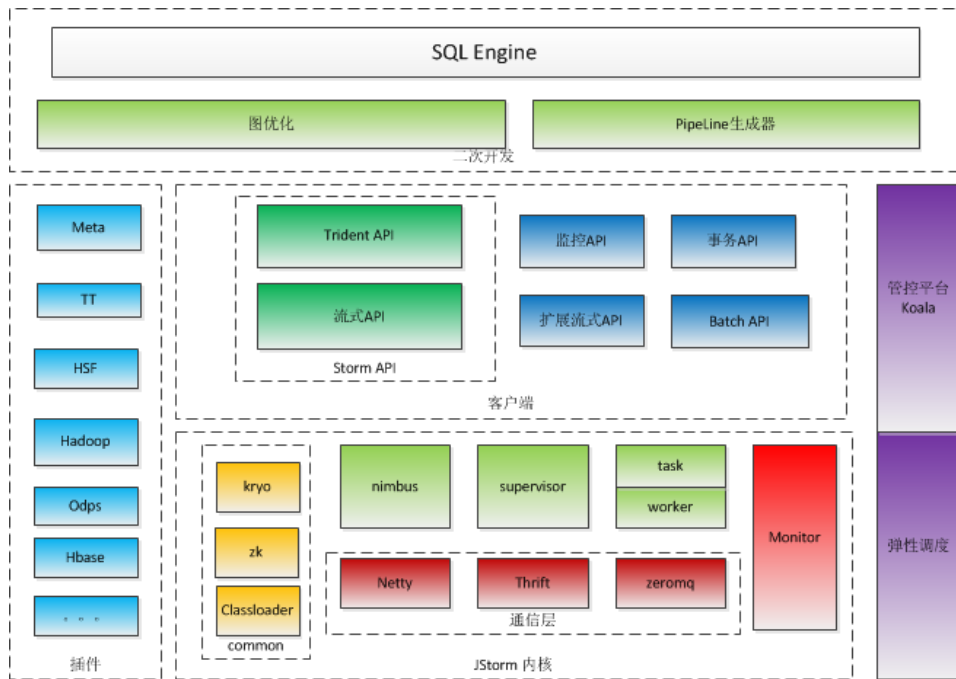
- 1600个任务, 3000+机器)
 - **风控系统**：集团nut／支付宝velocity
 - **审计类**：AMG 库存对账/DAM数据库安全审计系统 / 阿里妈妈p4p结算 / 阿里妈妈实时反作弊 / 。。。
 - **统计分析类**：eagleeye/ae pv－uv计算 / BI 统计/...
 - **监控系统**：tlog／rds-monitor/oceanbase-monitor/菜鸟雷达／去啊监控 / 云os监控 / 。。。
 - **管道类**：一号项目内数据库实时解析，RDS, 统一日志
 - **实时推荐**：tpp
 - **调度系统**：ae 邮件实时分析，1688网销宝

- 1
- 2
- 3
- 4
- 5



Jstorm Turbo 架构

1 2 3 4 5



弹性调度



2个方向:

大规模化:

轻量化（单元化）:

大规模化:

jstorm-on-yarn, 一个物理集群上部署多个逻辑集群

轻量化

Jstorm-on-docker: 集团测试机器上划小资源给jstorm,

弹性扩容缩容

一台虚拟机即可部署， 一键安装

- 1
- 2
- 3
- 4
- 5



创新点



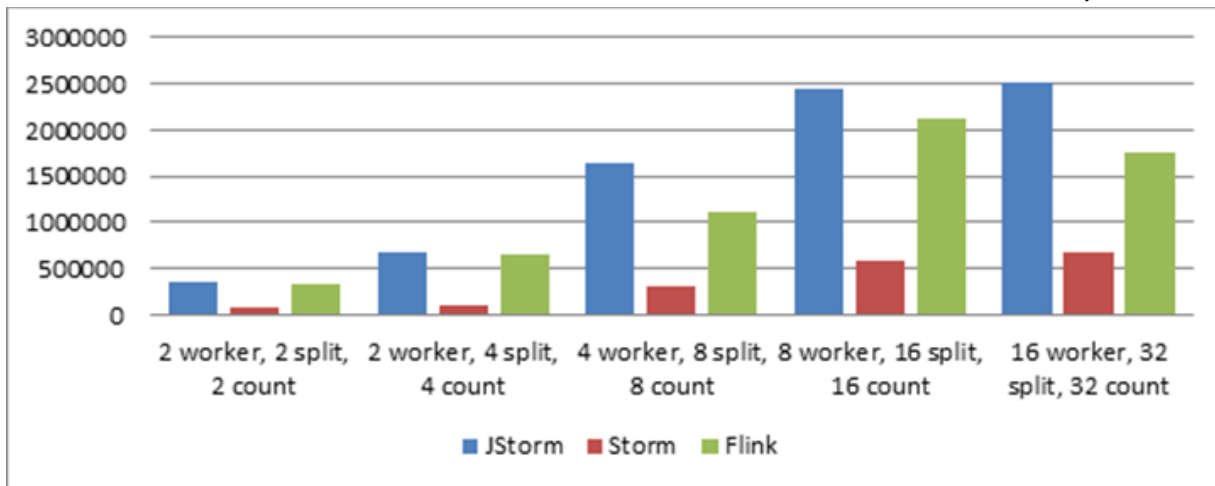
在storm的基础上，在稳定性 / 性能 / 功能特性上前进一大步

- 提高稳定性
 - 全新监控系统
 - 限流控制
 - 资源隔离
 - Nimbus HA（中央控制器HA）
 - 超大规模部署（突破Storm 不能超过400台上限）
 - TopologyMaster

性能提升

1 2 3 4 5

- WordCount: 是storm的5~6 倍， 比flink 快 10% ~ 20%



性能提升

1. 重新设计数据结构
2. 多级batch机制
3. 增加线程池单独处理反序列化
4. 去掉总发送线程和总接受现场，降低瓶颈
5. 在spout内部读写分离
6. 解决disruptor和spout 空数据 cpu 空跑问题.
7. 降低采样对性能影响.
8. 优化所有复杂函数， 比如rotatingmap / acker
9. GC调优

高级特性

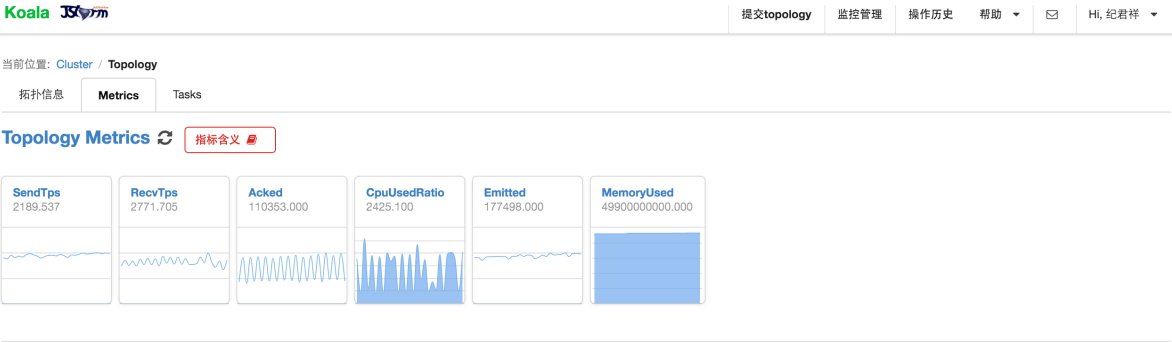


- 在api上，jstorm是storm的超集
 - 自定义监控与报警
 - 自定义调度
 - Exactly-once/window 消息支持
 - 自定义日志
 - 优雅关闭
 - 管控平台
 - 动态加载配置 / 扩容 / 重启
 - 自定义内存使用
 - 多netty策略
 - Classloader — 解决类冲突

管控平台koala

- 1
- 2
- 3
- 4
- 5

- 用户管理
- 应用管理
- 集群管理
- 监控



Component Metrics (最近1分钟)

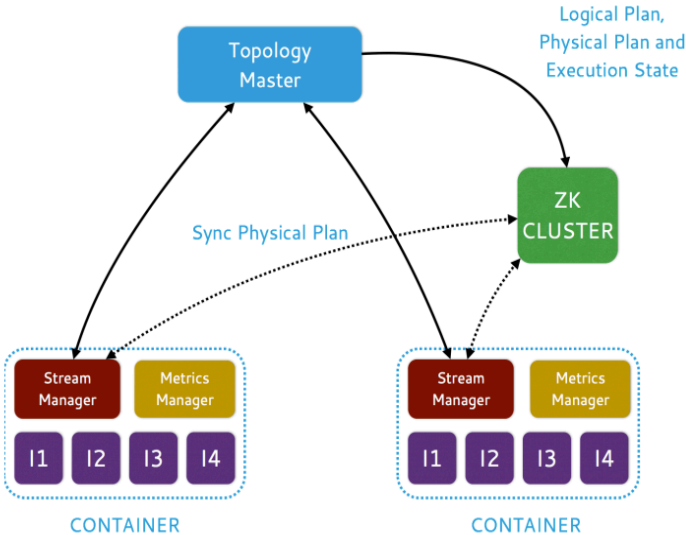
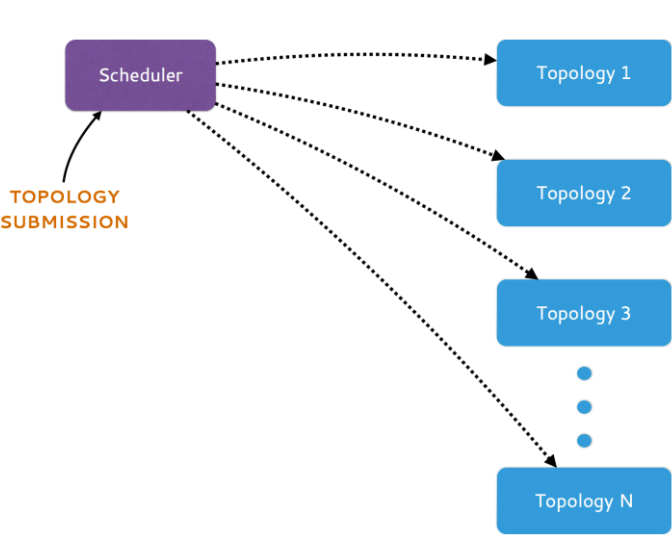
| Component | Emitted | Acked | SendTps | RecvTps | ProcessLatency | DeserializeTime | SerializeTime | ExecutorTime |
|---------------------------|---------|--------|-----------|-----------|----------------|-----------------|---------------|---------------|
| cdnLogSpout (2) | 9,088 | 0 | 127.538 | 0 | 0 | 0 | 0 | 1,470,000 |
| aplusResourceLogSpout (4) | 24 | 0 | 0.385 | 0 | 0 | 0 | 0 | 606,476.19 |
| aplusPcLogSpout (64) | 78,487 | 0 | 1,145.632 | 0 | 0 | 0 | 37.037 | 1,289,436.893 |
| aplusMobileLogSpout (4) | 24 | 0 | 0.39 | 0 | 0 | 0 | 0 | 1,654,375 |
| resMetricsCalcBolt (32) | 192 | 0 | 3.076 | 0 | 0 | 0 | 0 | 0 |
| resIpParserBolt (32) | 192 | 0 | 3.076 | 0 | 0 | 0 | 0 | 0 |
| ipParserBolt (32) | 79,281 | 77,597 | 806.632 | 1,318.425 | 120.69 | 21.898 | 0 | 0 |

与Twitter Heron 区别

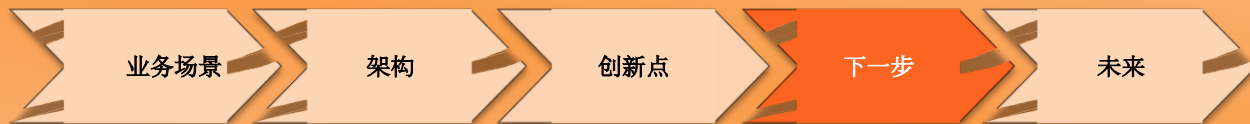
1 2 3 4 5

| 核心设计思想 | Twitter Heron | Alibaba JStorm |
|------------|--------------------------------------|--|
| 大规模部署 | 类似storm-on-mesos | jstorm-on-yarn |
| App-master | TopologyMaster | TopologyMaster |
| 反压 | BackPressure | Backpressure |
| 监控系统 | VIZ/Heron Tracker | Koala |
| 性能上 | 是Storm 0.8.2 的10倍，但只是fieldgrouping场景 | Storm 0.8.2 的10倍多，不仅是filed group，如果是localshuffle，会超过20倍的提升 |

Twitter Heron 架构



- 1
- 2
- 3
- 4
- 5



下一步



- 帮社区重写storm 2.0
- Jstorm-on-yarn / jstorm-on-docker 调优
- 提升开发效率
 - Exactly-once 框架性能优化
 - 业务window 机制

1

2

3

4

5

业务场
景

架构

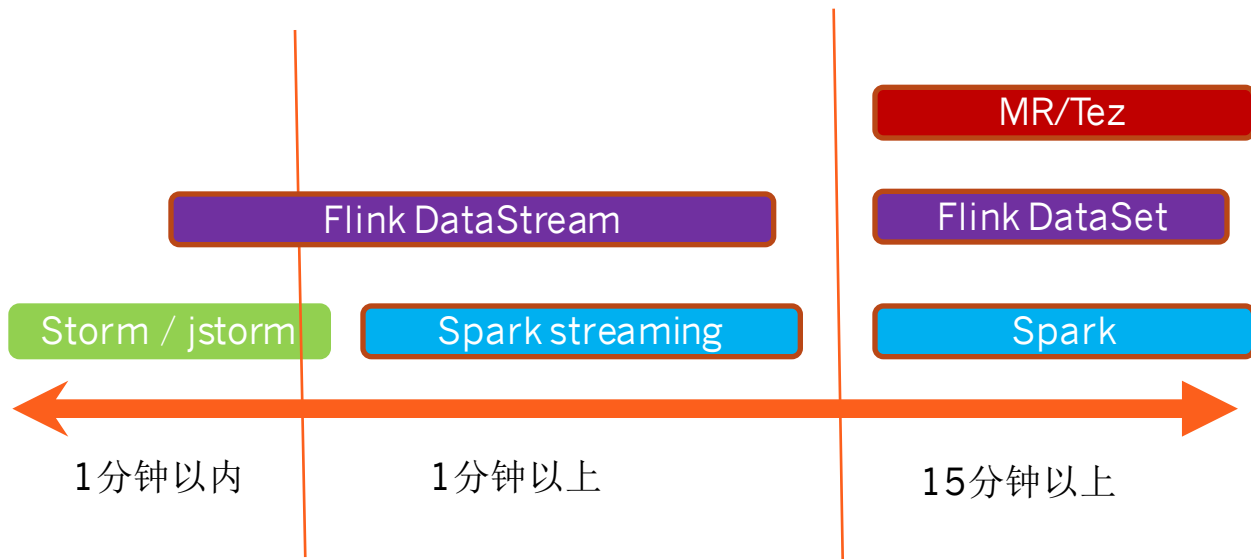
创新点

下一步

未来

未来

1 2 3 4 5





longda

