



Jstorm介绍

longda@alibaba

The image features a stack of books against a dark teal background. The top book is open, showing its pages. A white circular graphic with a teal segment is centered over the open book. Inside the circle, the text '已有什么' is written in white, and the number '1' is written in teal below it.

已有什么

1

- 国内最早使用storm的团队之一
 - Storm 0.5.1/0.5.4/0.6.0/0.6.2/0.7.0/0.7.1
 - Jstorm
0.7.1/0.9.0/0.9.1/0.9.2/0.9.3/0.9.3.1/0.9.4/0.9.4.1/0.9.5/0.9.5.1/0.9.6/0.9.6.1/0.9.6.2/0.9.6.3/0.9.7/0.9.7.1/2.0.4
- 公司内职责：
 - 运维
 - 技术支持
 - 应用开发

公司内部近1200台机器，遍及公司所有BU, 超过400个具体应用， 1.5 PB

阿里云 飞天二部 云产品监控

阿里妈妈事业部 共享平台、钻展、交易结算

支付宝数据仓库、无线实时分析、商户实时监控

共享事业部 调用链监控EagleEye、Tlog实时报表

统一推荐平台

Alimonitor

来往事业部 – 安全防控

OS事业群 – 云os实时报表

航旅事业群 – 机票实时监控、营销监控

安全事业群 --交易秩序技术

。 。 。

Tlog/eagleeye

Rds

Oceanbase

阿里妈妈实时反作弊

Ctu 风控

BI 实时监控

- 使用者角度的高级特性

- 自定义调度
- 自定义日志
- 自定义Metrics
- 优雅shutdown

- 动态扩容/重启
- 内存使用
- 网络传输高级特性
- classloader

- 自定义每个worker的cpu权重/内存大小
- 自定义任务分配
 - 任务的分配完全自己控制
 - 可以将多个component绑定到某个worker（如spout/bolt）
- 强制某个component的task运行不同的机器上
- 强制某个topology运行在某些机器上
- 强制使用历史调度

- 可以使用自己的log4j配置
- 可以切换到logback
- 可以重定向终端输出到指定文件（/dev/null）
- 日志可以增加常用tag
（clustername/hostname/topologyname/workerid/taskid
）

- 使用java metrics
 - 自定义采样
 - Web ui 可以展示
- 使用Alimonitor
 - 最新Alimonitor提供报警接口， topology自定义报警
 - 所有metrics都送往Alimonitor， 自己配置报警和曲线图

- Topology 进入killed 状态
- Nimbus 等待Timeout时间
 - Spout会触发进入deactive状态
- 从zk remove topology
- Supervisor 对每个worker执行一次kill命令
- Supervisor 等待task.cleanup.timeout.sec
 - 保证每个task的cleanup能成功完成
- Supervisor对每个worker执行5次kill -9 命令

- Rebalance
 - 扩容时，可以不杀死任何worker，不影响现有数据流
- Restart
 - 可以重新调整所有的设置，重新load configuration
 - 可以对任何一个component并发进行重新修改

- 修改worker 内存大小 -- worker.memory.size
- 修改gc 方式
 - worker.gc.childopts
 - 使用自定义调度接口WorkerAssignment
- 队列模型切换
 - 固定长度的disruptor和无限长度的blockingdequeue
 - 修改disruptor长度

- 同步模式和异步模式切换
- 异步模式下的缓冲机制
- Netty thread pool 设置

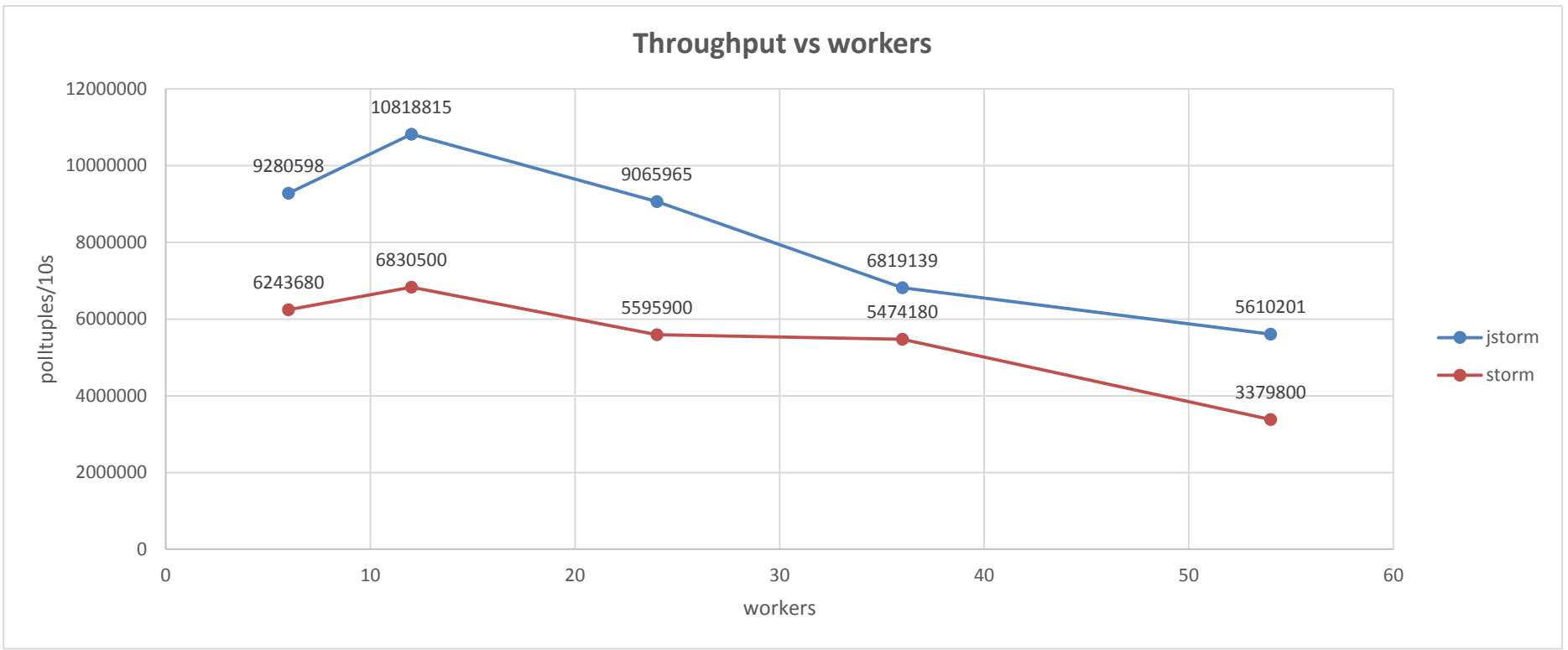
- 当topology jar和jstorm jar发生冲突时， 需要打开classloader
 - topology.enable.classloader

- 系统角度的高级特性
 - 资源隔离
 - Nimbus HA
 - 性能优化
 - RPM 安装做了大量系统设置

- 多集群物理隔离， 利用管理控制台管理多个集群
- 单集群内部
 - Cgroup， share + limit 方式
 - 也可以用自定义调度， 部署topology在某些机器上

- Nimbus HA,
 - 稳定可靠，随便切换
 - 不对第三方依赖

- 6 Servers (24core/98G)
- 18 Spout/18 Bolt/18 Acker



设计上做了大量优化

- 额外增加反序列线程

- Redesign采样代码，减少采样性能影响

- 让Spout发送和接收隔离，加强发送性能

- 发送数据前自动batch(新功能)

- 去掉几处CPU空转操作

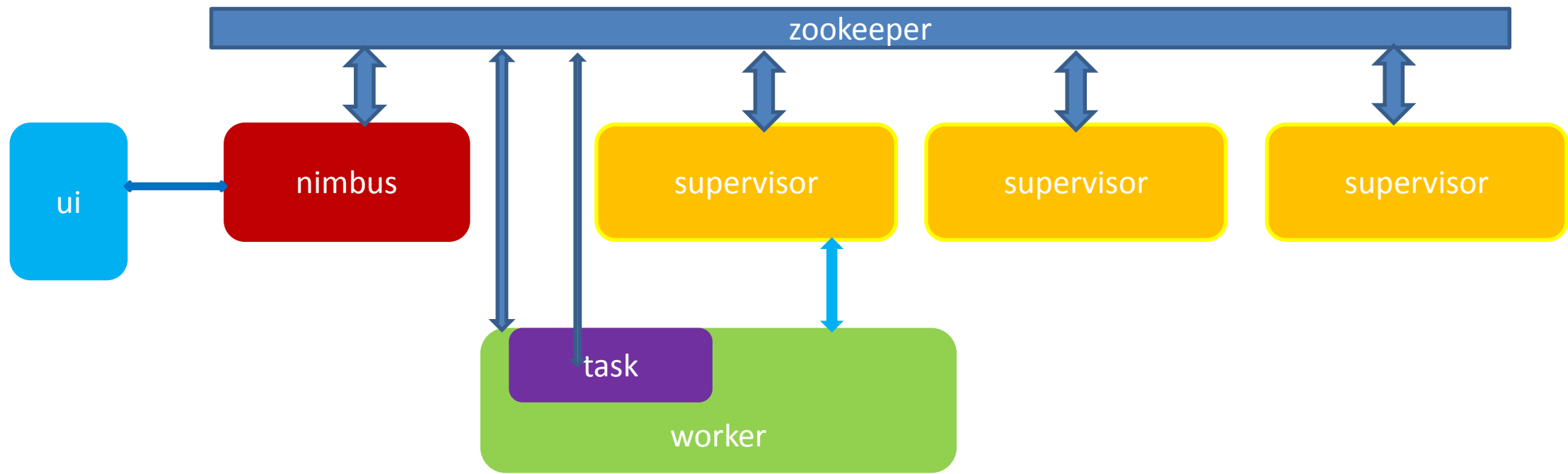
- 优化ack代码

- 优化缓冲map性能

- 对GC做了特别优化

- Java比clojure更底层

JStorm层技术架构



- 内核演化
 - Topology Manager —— 下一个版本
 - BackPressure —— 下一个版本
 - Zk数据结构重构 —— 下一个版本
 - 超大型Topology优化 —— 下一个版本
 - 集群规模2000+ node —— 下一个版本
- Topology自动扩容和缩容 —— 下下一个版本
- Koala（集成用户管理，集群管理，应用管理，监控系统）
 - 全新监控系统 —— 下一个版本

已经明确的：

1. Merge Apache
2. SQL Engine 重构
3. 新的编程框架
4. 弹性计算



欢迎有志之士 加入

QQ: 32147704

谢
谢！