

美团Flink实践与应用

孙梦瑶

基础数据部





孙梦瑶 美团点评 研发工程师

毕业于北京科技大学，2017年加入美团点评数据平台。
从事实时计算平台建设与维护相关工作，致力于打造高效、可靠、易用的基础设施和解决方案。

CONTENTS

目录

01

• 现状与背景

02

• Flink在美团的实践

03

• Flink在美团的應用

04

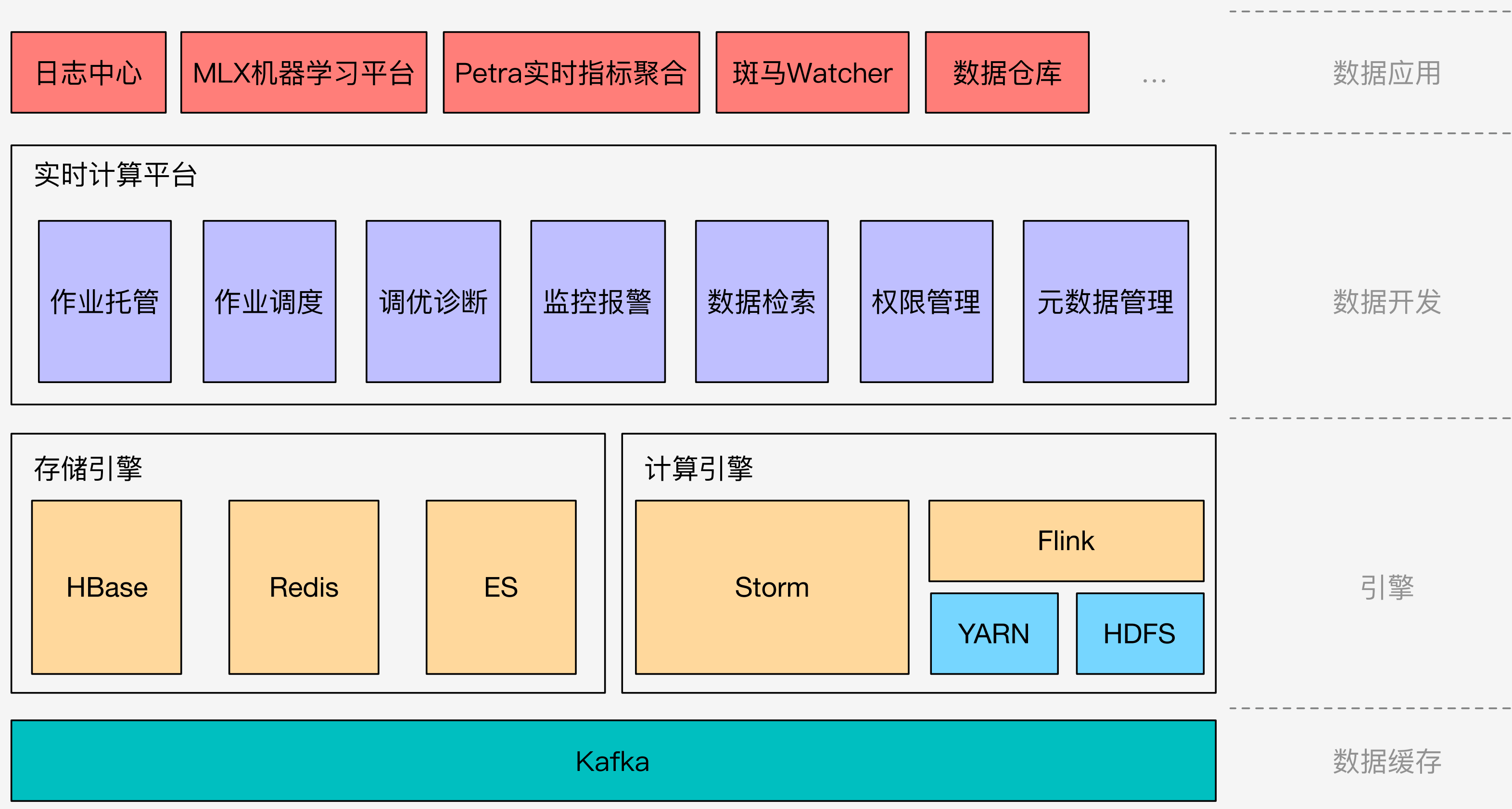
• 未来展望

01. 现状与背景

- 实时平台架构
- 实时平台现状
- 痛点和问题
- Flink探索的关注点

Why  Flink?

01. 现状与背景 实时平台架构



01. 现状与背景

实时平台现状

近万

平台运行作业总数



数千台

集群节点规模



万亿级

天级消息量



千万条

高峰期秒级消息量



01. 现状与背景

痛点和问题

- ☹️ 实时计算**精确性**问题
- ☹️ 流处理中的**状态管理**问题
- ☹️ 实时计算**表义能力**的局限性
- ☹️ 开发调试**成本**高



01. 现状与背景

Flink探索关注点

- ☺ Exactly once计算能力
- ☺ 状态管理能力
- ☺ 窗口/Join/时间处理等等
- ☺ SQL/Table API



02. Flink在美团的实践

- 稳定性实践
- Flink平台化
- 生态建设

How to use Flink?

02. Flink在美团的实践

稳定性实践-资源隔离

- 资源隔离的**考虑**：分场景、按业务
 - 高峰期不同，运维时间不同；
 - 可靠性、延迟需求不同；
 - 应用场景，重要性不同。
- 资源隔离的**策略**：
 - YARN打标签，节点物理隔离；
 - 离线DataNode与实时计算节点的隔离。



02. Flink在美团的实践

稳定性实践-故障容错

- 节点/网络故障
 - JobManager HA
 - 自动拉起
- 上下游容错
 - Flink Kafka 08 读写异常重试
- 容灾
 - 多机房
 - 流热备

02. Flink在美团的实践

Flink平台化-作业管理

名称	TaskSet	配置	状态	操作	配置
			DOWN	操作	调试配置
		1.0.3	DOWN	操作	调试配置
		0.0.2	ACTIVE	操作	调试配置
		0.0.1	DOWN	操作	调试配置
		1.0.1	ACTIVE	操作	调试配置
		1.0.7	DOWN	操作	调试配置

启动生产
均衡生产
结束生产
命令日志
查看日志

作业运行状态

作业基本操作

作业基本配置 [app_data_rcwh_flink_test]

基本配置 生产配置 调试配置 生产版本 Metrics 历史 依赖 延迟

TaskSet: [模糊] 引擎: Flink
用户组: data 名称: rcwh_flink_test
Git仓库: [模糊] 相对目录: [模糊]
报警方式: 大象 邮件 短信
组织架构: [模糊] 拓扑类型: App
服务事业部: [模糊]
负责人邮箱: [模糊]
备注信息: [模糊]

基本信息配置

FlinkFrame配置

基本配置 生产配置 调试配置 生产版本 Metrics 历史 依赖 延迟

主函数: [模糊]
并发度: 1
TaskManager数量: 1
Flink版本: 1.4.0
从上一次结束 Savepoint启动:
Flink arguments: 选填

JobManager内存/MB: 1024
TaskManager内存/MB: 1024
JDK版本: jdk1.8.0_112

资源并发配置

02. Flink在美团的实践

Flink平台化-监控报警

作业状态

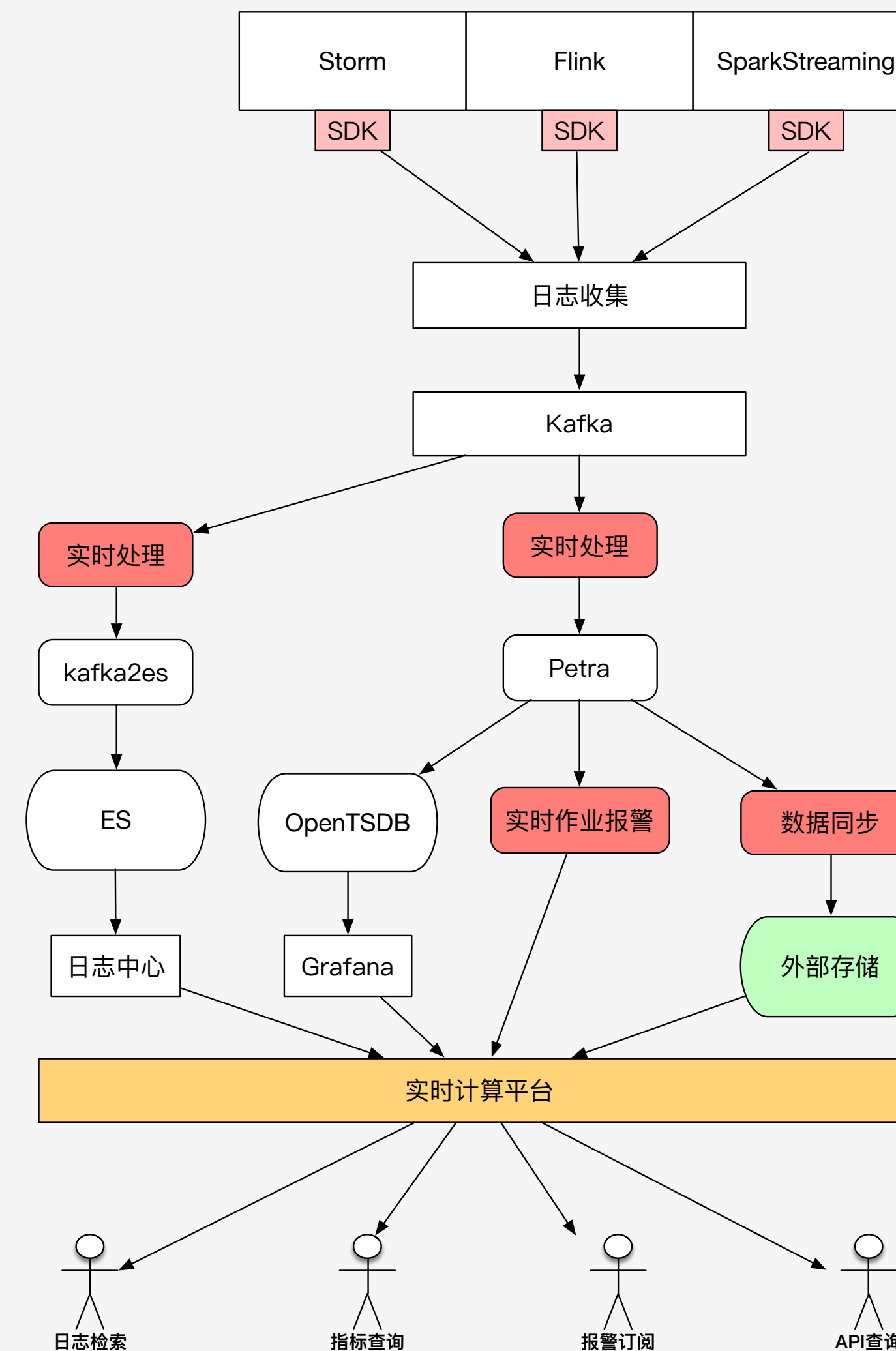
处理延迟

自定义
Metrics

02. Flink在美团的实践

Flink平台化-调优诊断

- 实时计算引擎提供**统一的日志和指标方案**
- 为业务提供按条件过滤的**日志检索**
- 为业务提供自定义时间跨度的**指标查询**
- 基于日志和指标，为业务提供**可配置的报警**



02. Flink在美团的实践

Flink平台化-调优诊断

日志查询条件

日志名
索引名
起始时间 2018/08/27 16:48:27 now
结束时间 2018/08/27 17:48:27 now
15m 30m 1h 3h 12h 1d 7d 14d
查询 GO

日志查询结果

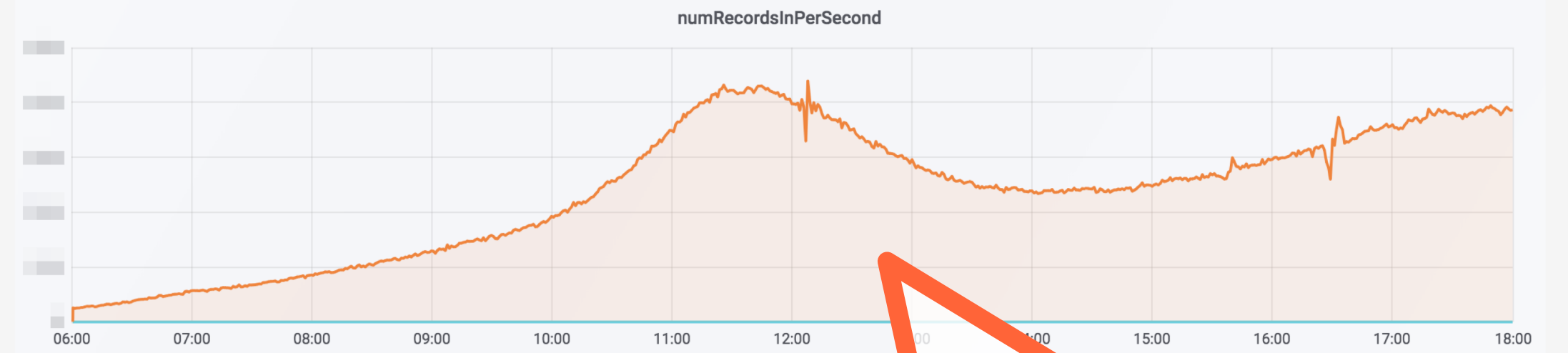
raw json kv

- 全选
- 反选
- es_timestamp
- job_name
- mt_appkey
- mt_clientip
- mt_logger_name
- mt_servername

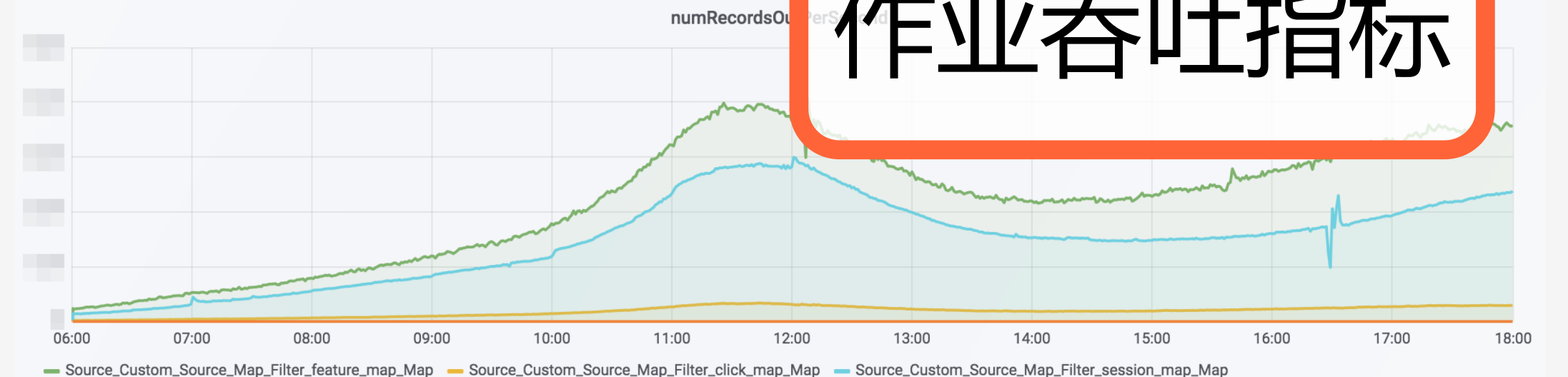
applicationID	containerID	es_timestamp	message	mt_datetime	mt_level	mt_thread
application_1535179038827_1658522	container_e59_1535179038827_1658522_01_000015	2018/08/27 17:48:27 +0800	bu is null,itemType:ARTICLE,itemId:	2018-08-27 17:48:27+0800	ERROR	SimpleConsumer-Source:Custom_Sourc->_Map->_Filter->_Map->_Filter->_exposure->_Map->_broker-2958_()
application_1535179038827_1658522	container_e59_1535179038827_1658522_01_000015	2018/08/27 17:48:27 +0800	bu is null,itemType:ARTICLE,itemId:	2018-08-27 17:48:27+0800	ERROR	SimpleConsumer-Source:Custom_Sourc->_Map->_Filter->_Map->_Filter->_exposure->_Map->_broker-2963_()

起始时间 2018-08-27 06:00:00 终止时间 2018-08-27 18:00:00 快速选择 过去12小时 查询

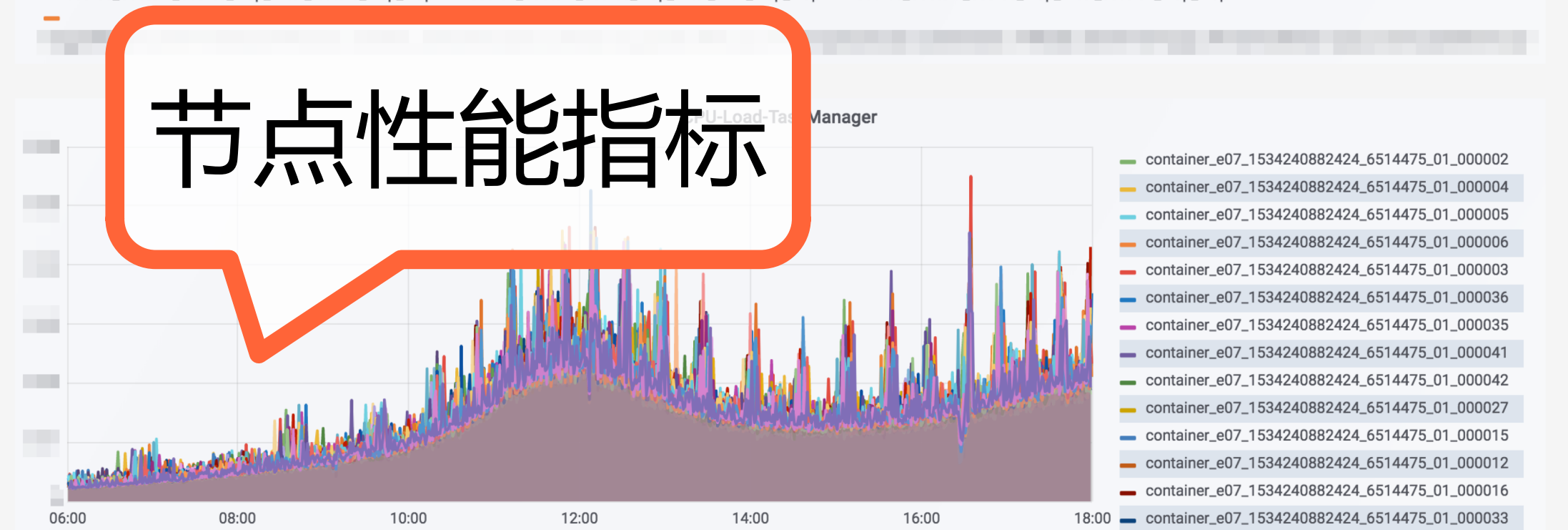
Metrics帮助wiki



作业吞吐指标

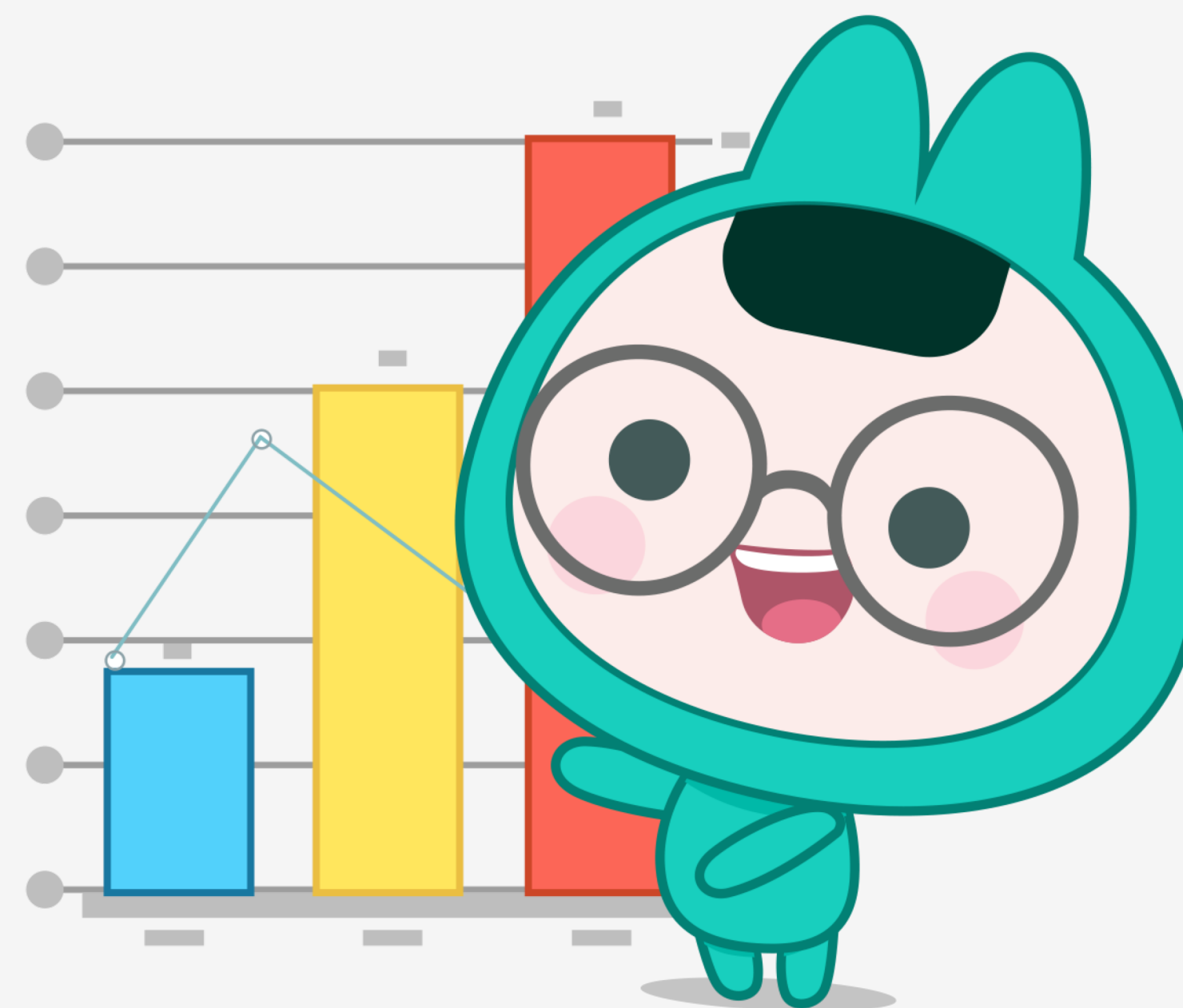


节点性能指标



02. Flink在美团的实践 生态建设

- 适配业务MQ
 - 对接线上数据
 - 一次同步，多次消费
- 封装Kafka读写组件
 - 屏蔽地址、基础配置
 - 权限管理
 - 指标监控



03. Flink在美团的应用

- Petra实时指标聚合
- MLX机器学习平台

What to do with Flink?

03. Flink在美团的应用

Petra实时指标聚合

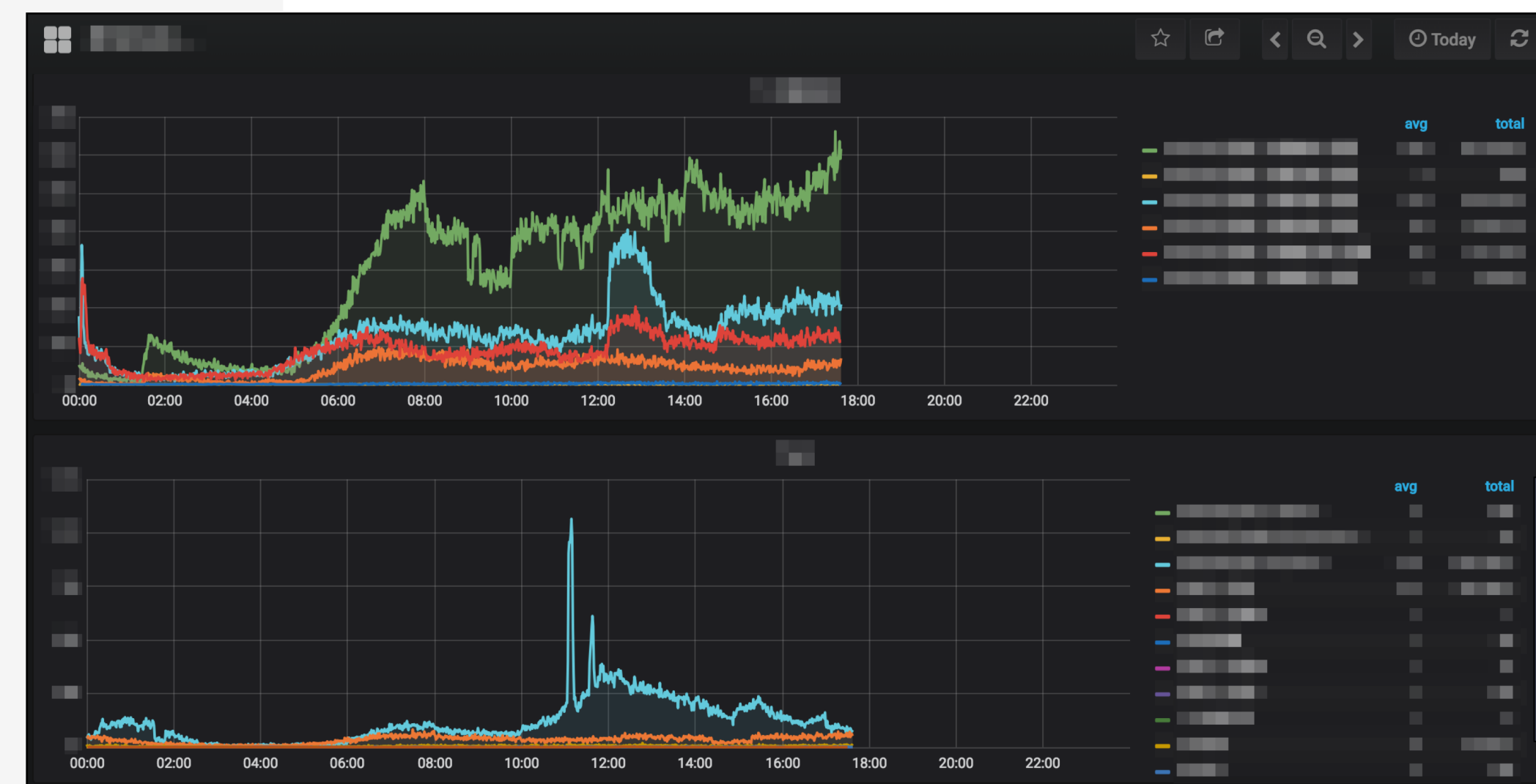
- 业务指标聚合服务
- 对业务系统的指标进行统一的聚合、展示
 - XX服务的95%的响应时间
 - XX数据每个类型的计数
 - 每分钟XXXX的成功率
 - 如：交易成功率 = 支付成功数 / 下单数

聚合规则配置

× 删除配置 + 新增配置 批量导入 Search

<input type="checkbox"/>	metric名称	聚合类型	聚合维度	聚合粒度	数据延迟时间	同步falcon	操作
<input type="checkbox"/>	__fail_count	counter	job_name	60, 3600, 86400	300	False	📄
<input type="checkbox"/>	__ack_count	counter	host, job_name	60	300	False	🗑️

Showing 1 to 2 of 2 rows



03. Flink在美团的应用

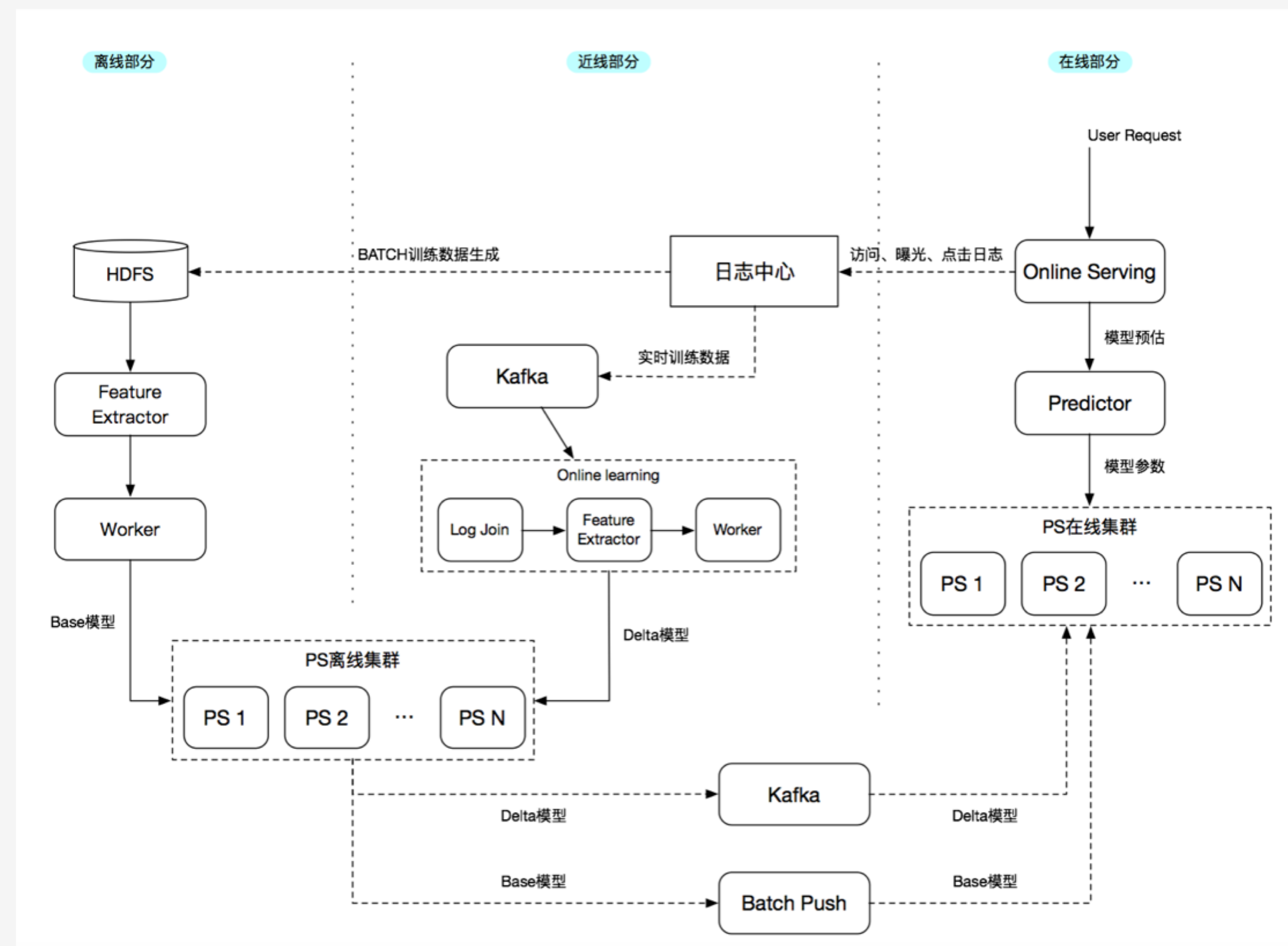
Petra实时指标聚合

- 业务场景特征
 - 基于业务时间（事件时间）
 - 多业务维度聚合，如应用、通道、机房等
 - 复合指标计算
 - 如：交易成功率 = 支付成功数 / 下单数
 - 低延迟，秒级结果输出
- Exactly once的精确性保障
 - Flink checkpoint机制
 - 维度计算中数据倾斜
 - 热点key散列
 - 对晚到数据的容忍能力
 - 默认分钟级，最高支持小时级

03. Flink在美团的应用

MLX机器学习平台

- MLX：大规模机器学习平台
- 模型训练、特征提取、在线预估等
- 近线训练部分
- 流Join
- 大窗口



04. 未来展望



状态管理

- 统一管理
- 大状态性能优化



SQL

- 开发效率提升



新场景探索

So what's next?

Q&A

CODE A BETTER LIFE

一行代码 亿万生活



更多技术资料
欢迎关注“美团点评技术团队”



招聘：后台开发岗位
邮箱：sunmengyao02@meituan.com