

# 大数据架构下的质量工具建设

孔祥云

2018/4/20





# 目录

CONTENTS

1

质量工具面临的数据挑战

2

稳定性监控系统的数据挑战

3

实时应对方案

4

智能接口平台的数据挑战

5

智能应对方案

# 我是谁？



《京东系统质量保障技术实战》作者之一，多年测试领域从业经验，擅长功能测试、自动化测试、测试工具开发等。

在京东任职期间参与开发EAT测试框架、Mock接口测试平台、自动部署系统、商城质量门户等；主导应用大数据技术的接口稳定性监控平台、接口变更探测工具等。

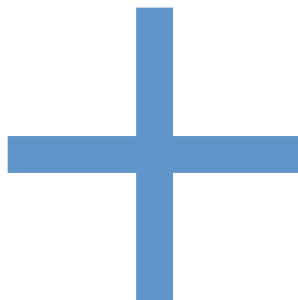


# 质量工具面临的数据挑战

规模

千万级别数据

数据存储瓶颈



计算

操作百万级别数据

亚秒级响应

# 稳定性监控系统的挑战——背景

接口

监控

稳定性

# 稳定性监控系统的数据挑战——背景



- 接口的稳定性数据采集后，虽然直观，但是不能对整个应用有好的表达
- 一般查询只提供了单个接口的方式，不能查询较长时间、多接口的稳定性数据
- 不是基于业务场景的，非开发人员无法映射到具体的业务场景上

# 稳定性监控系统的数据挑战——目标

- 直观展示多维度稳定性趋势

多时间跨度



多业务场景

服务名称
获取单个商品的广告页
设置单个商品的广告页
通过类目和属性类型, 获取类目的属性列表

多层次



# 稳定性监控系统的数据挑战——难点分析

## ●实时性

- 每秒稳定性数据采集频次：

$$100 (\text{应用}) * 100 (\text{接口}) * 5 (\text{分钟}) / 60 (\text{秒}) = 833 \approx 1000$$

- 及时反映稳定性变化



采集频率5分钟



变化频率10分钟



# 稳定性监控系统的数据挑战——难点分析

## ● 海量数据

### ➤ 数据量大:

单个应用:

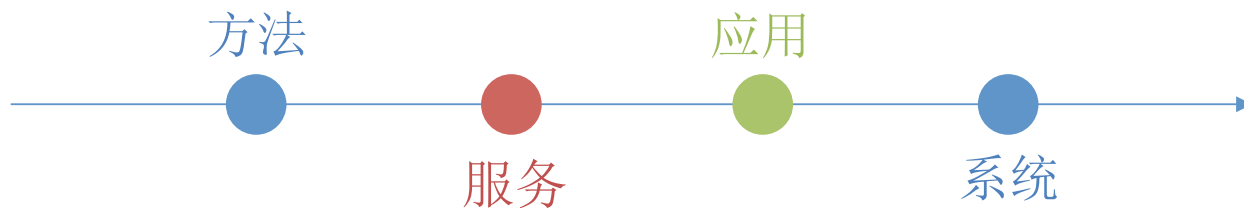
$$100 \text{ (接口)} * 24 \text{ (小时)} * 60 \text{ (分钟)} / 5 \text{ (分钟)} =$$

28800

$$28800 \text{ (条)} * 365 \text{ (天)} =$$

10,512,000

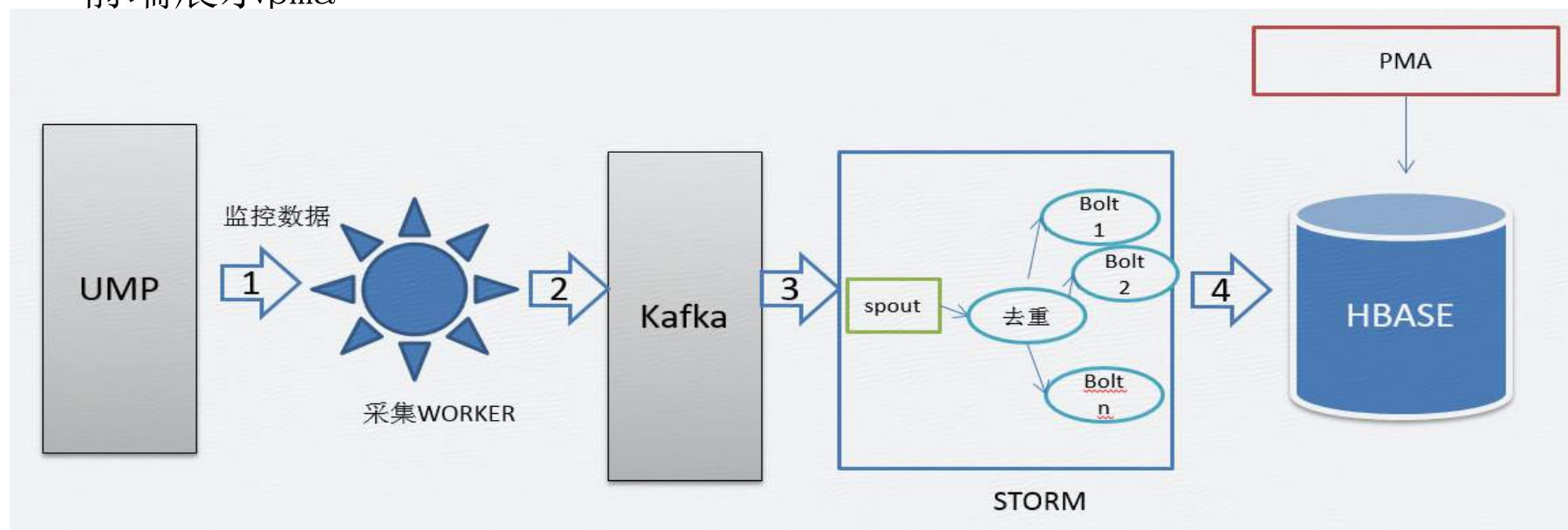
### ➤ 维度多:



# 实时应对方案

## ●涉及到的技术模块

- ✓ 采集worker
- ✓ 实时计算storm
- ✓ 前端展示pma

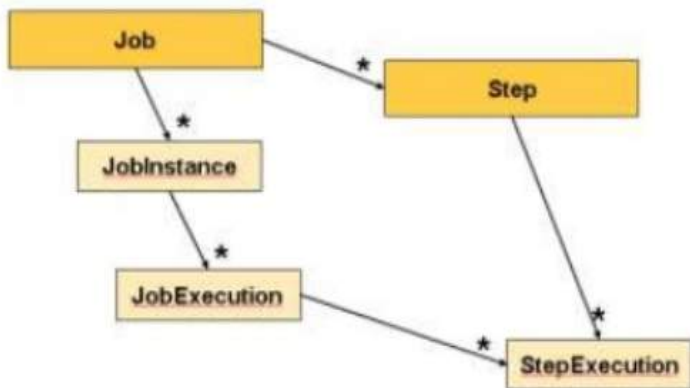


# 实时应对方案

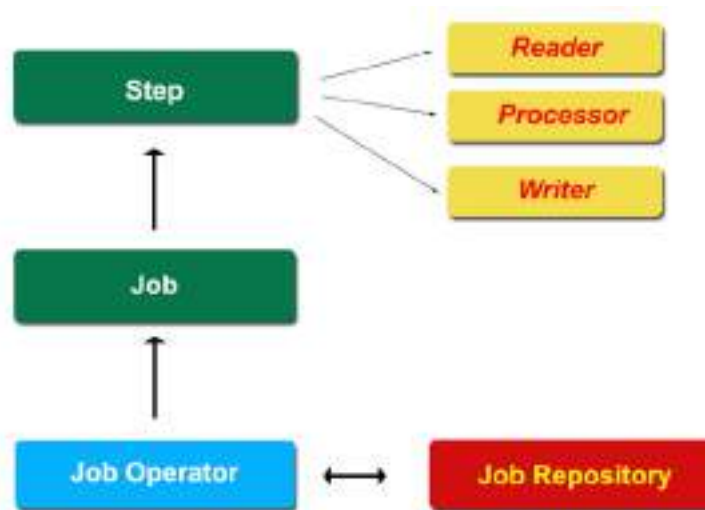
- 采集worker的及时性



并行

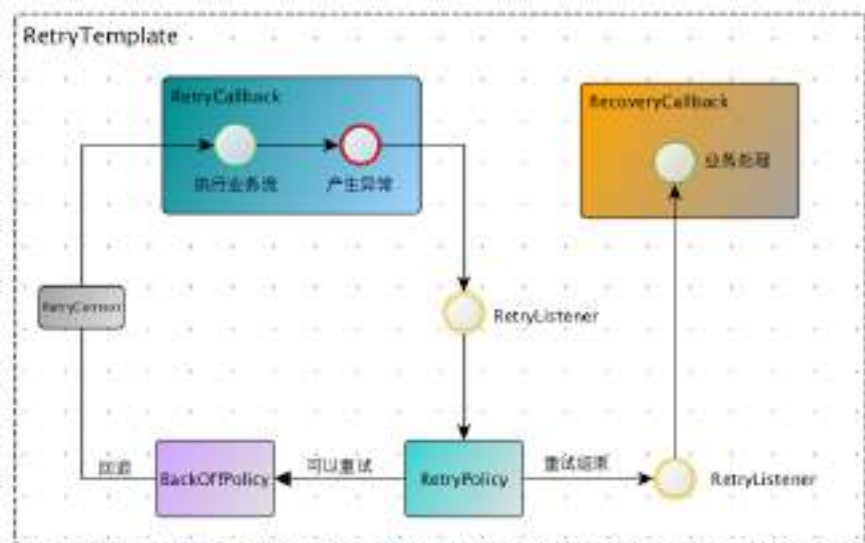


解耦



# 实时应对方案

- 采集worker的失败重采机制



# 实时应对方案

- 及时展示变化数据的大数据方案

离线计算



准实时计算



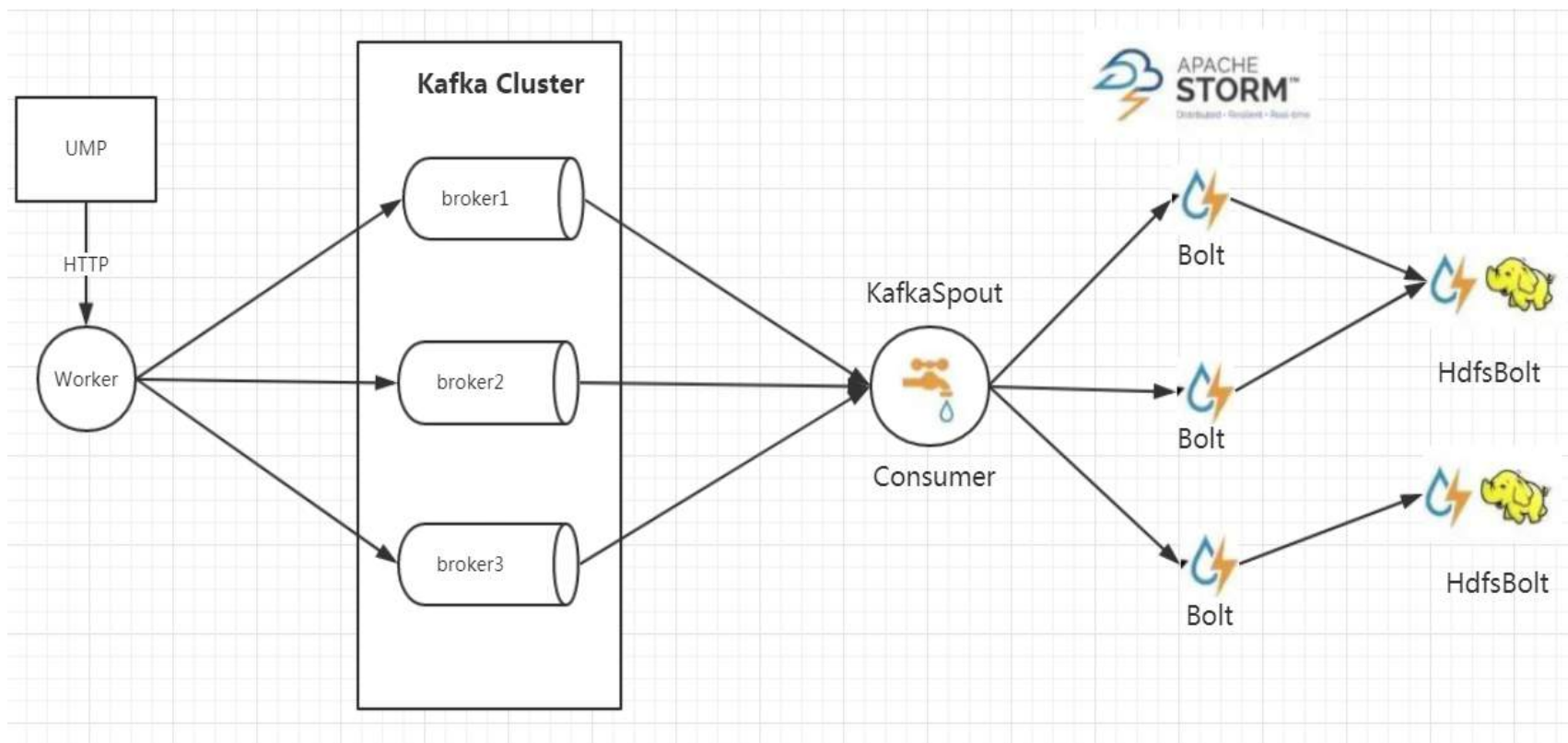
实时计算



# 实时应对方案

## ● Storm 特性

### ➢ 实时性



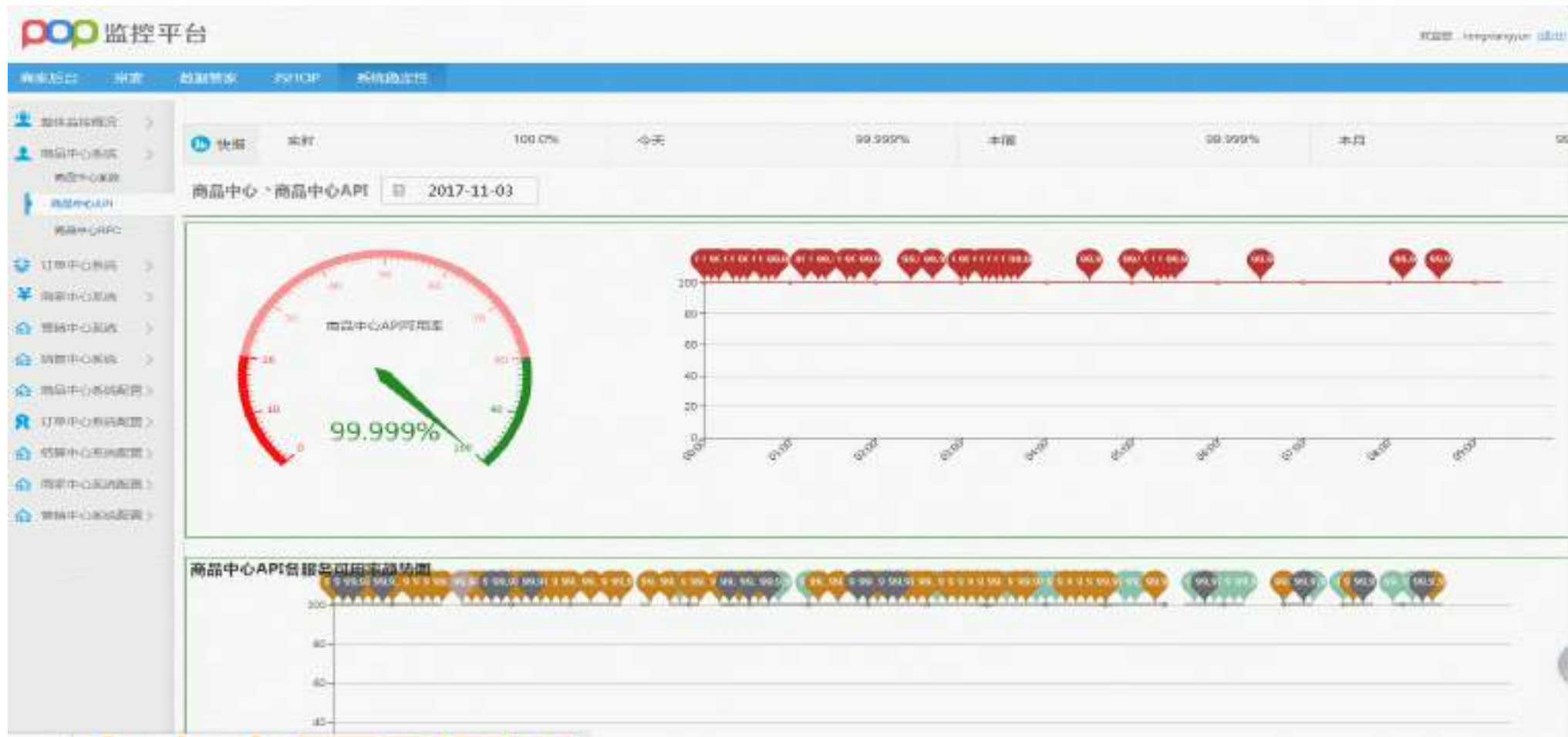
# 实时应对方案

## ●Storm使用中的问题及解决

- 数据落盘方案：HBase
  - ✓ 数据条数2000万+, 数据量PB级别
  - ✓ 横向领域扩展：新维度tag
  - ✓ 纵向字段扩展：新增字段
- 数据计算方案：Redis
  - ✓ Set集合



# 实时应对方案——效果展示



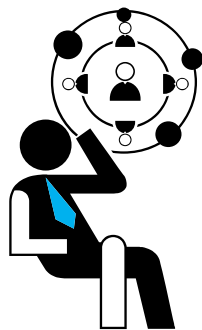


# 智能接口平台的数据挑战

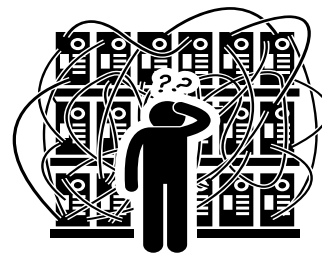
- 背景



数据准备



回归用例



用例选择

# 智能接口平台的数据挑战

- 目标

真实数据



分类

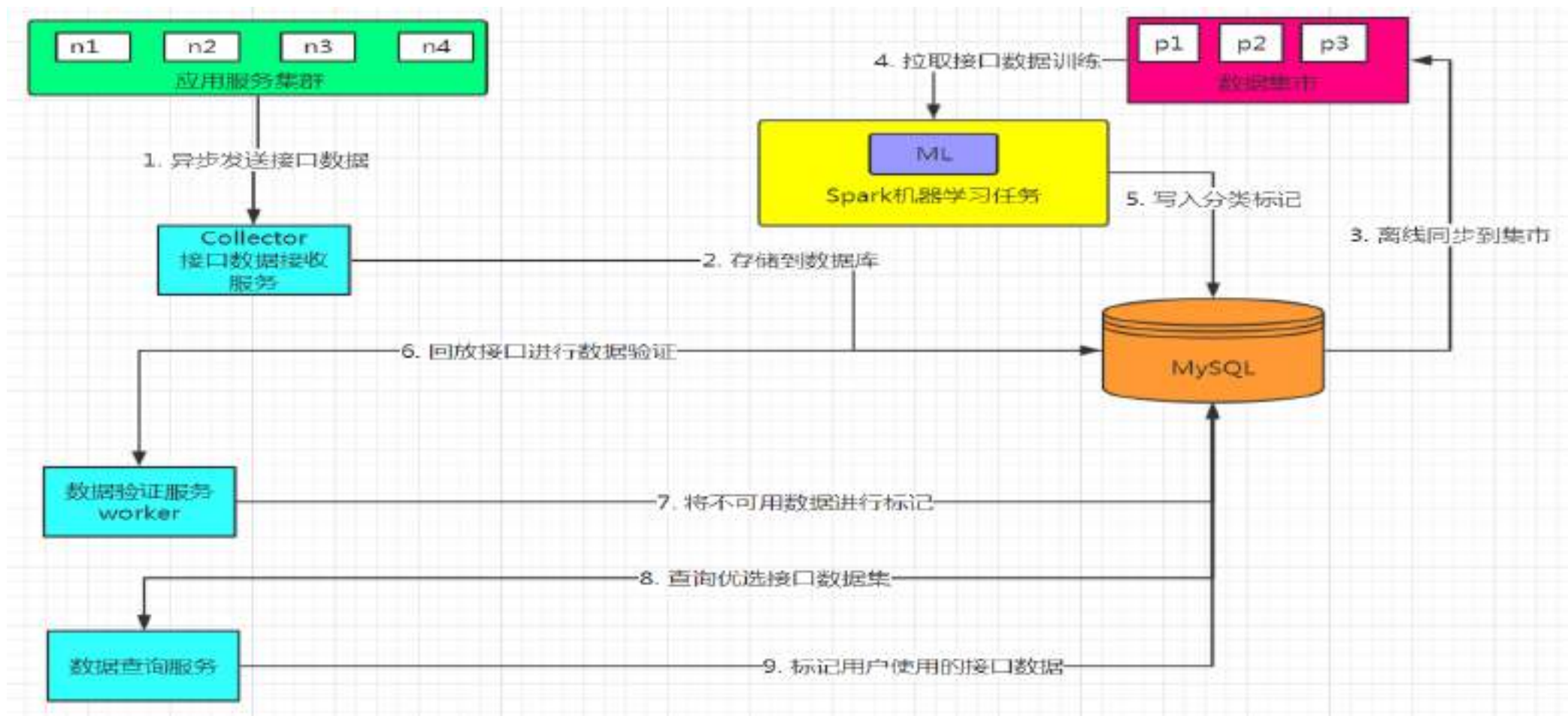
# 智能接口平台的数据挑战

## ● 数据挑战

- 接口：1000（目前接入100个接口）
- 每天积累数据：50万+
- 每天训练量：100万+



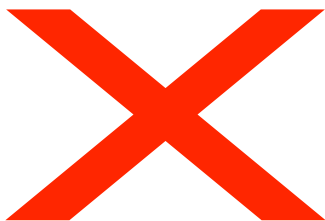
# 智能应对方案



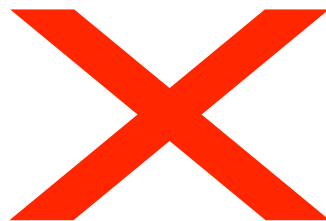
# 智能应对方案

## ●采集方案选择

日志



中间件



应用Filter



# 整体方案之机器学习

## ●机器学习工具选择



ML

	R	Spark
机器学习类库	√	√
语言支持	R语言	Scala, Java, Python, R
计算性能	几十万数据	百万以上级别数据



# 整体方案之机器学习

## ● 数据建模过程中的点滴

◆ Kmeans K中心点选择——轮廓系数

MLlib内置的WCSS类方法评估聚类模型的性能

WCSS for k=2 : 0.685

WCSS for k=3 : 0.71

WCSS for k=4 : 0.653

WCSS for k=5 : 0.676

WCSS for k=10 : 0.51465

WCSS for k=20 : 0.4684





# 整体方案之机器学习

- 数据建模过程中的点滴
  - ◆ 聚类后业务评估

A	C	G	I	L	O	P	Q	S
	category	jdPrice	orderId	skuId	stockNum	venderId	wareId	cluster
1	12212	108	6.59E+10	17864734287	4350	706581	10843606511	2
2	12212	108	6.59E+10	17864734287	4350	706581	10843606511	2
3	11155	98	NA	18814990660	429	662392	10915138394	3
4	12247	88	NA	19884419064	200	672184	10342535983	3
5	11151	58	NA	11896479497	504	662392	10299307162	3
6	1440	70	6.61E+10	1985486232	17	104149	1960383930	1
7	1440	70	6.61E+10	1985486232	17	104149	1960383930	1

# 整体方案之未来规划

1  
数据分类  
优选数据集

2  
精准测试  
接口覆盖率

3  
无人值守测试  
Code Diff

谢谢

