



**QCon** 全球软件开发大会  
INTERNATIONAL SOFTWARE  
DEVELOPMENT CONFERENCE

BEIJING 2017

# 海尔实时计算平台技术选型与实践

海尔电器 - 肖云



促进软件开发领域知识与创新的传播



关注InfoQ官方信息  
及时获取QCon软件开发者  
大会演讲视频信息



扫码，获取限时优惠

**ArchSummit**  
全球架构师峰会 2017 [深圳站]

2017年7月7-8日 深圳·华侨城洲际酒店

咨询热线：010-89880682

**QCon**

全球软件开发大会 [上海站]

2017年10月19-21日

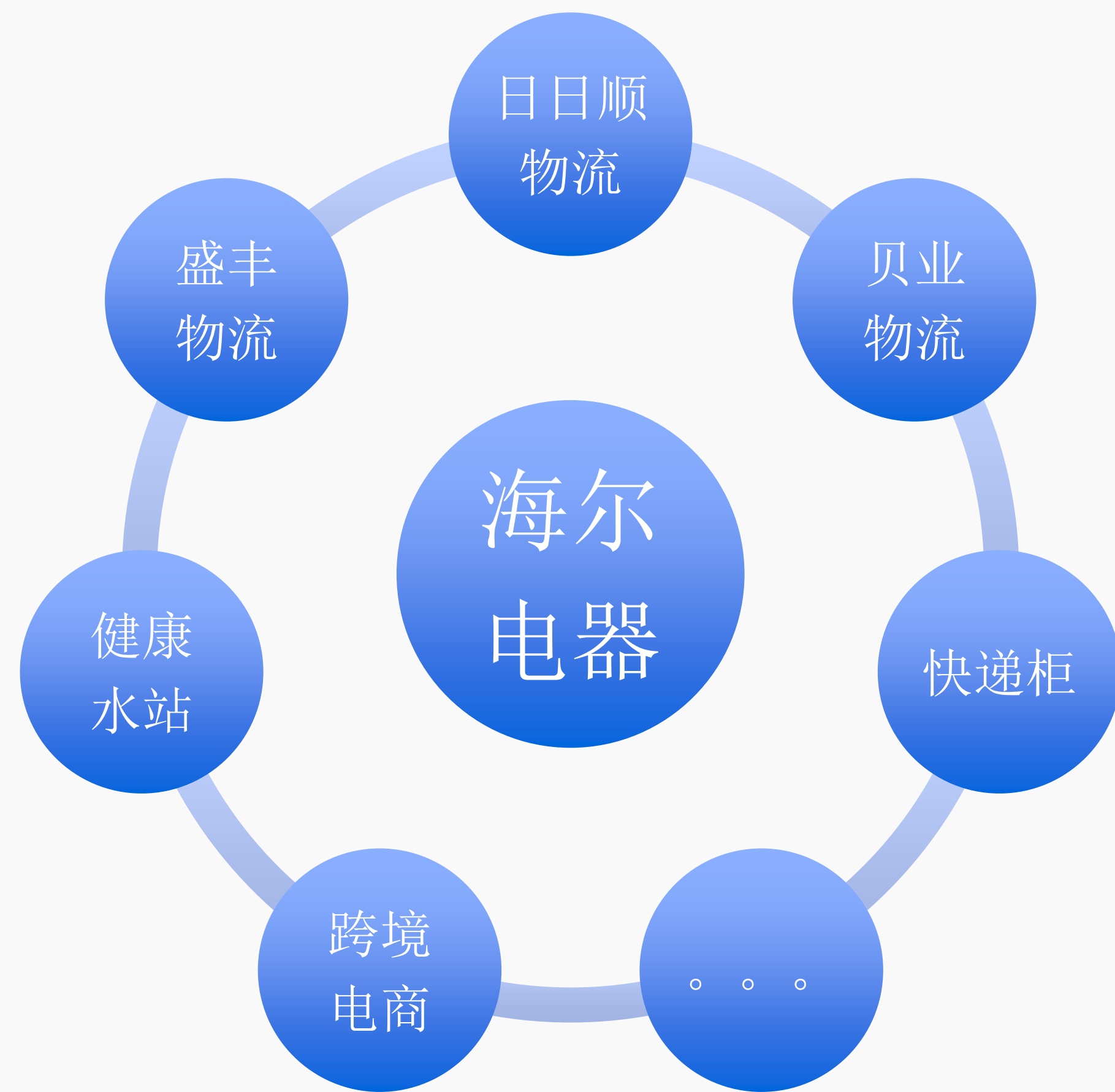
咨询热线：010-64738142

# 个人介绍

- 方正电子新媒体开发总监
- 中投视讯研发总监
- 海尔电器资深架构师

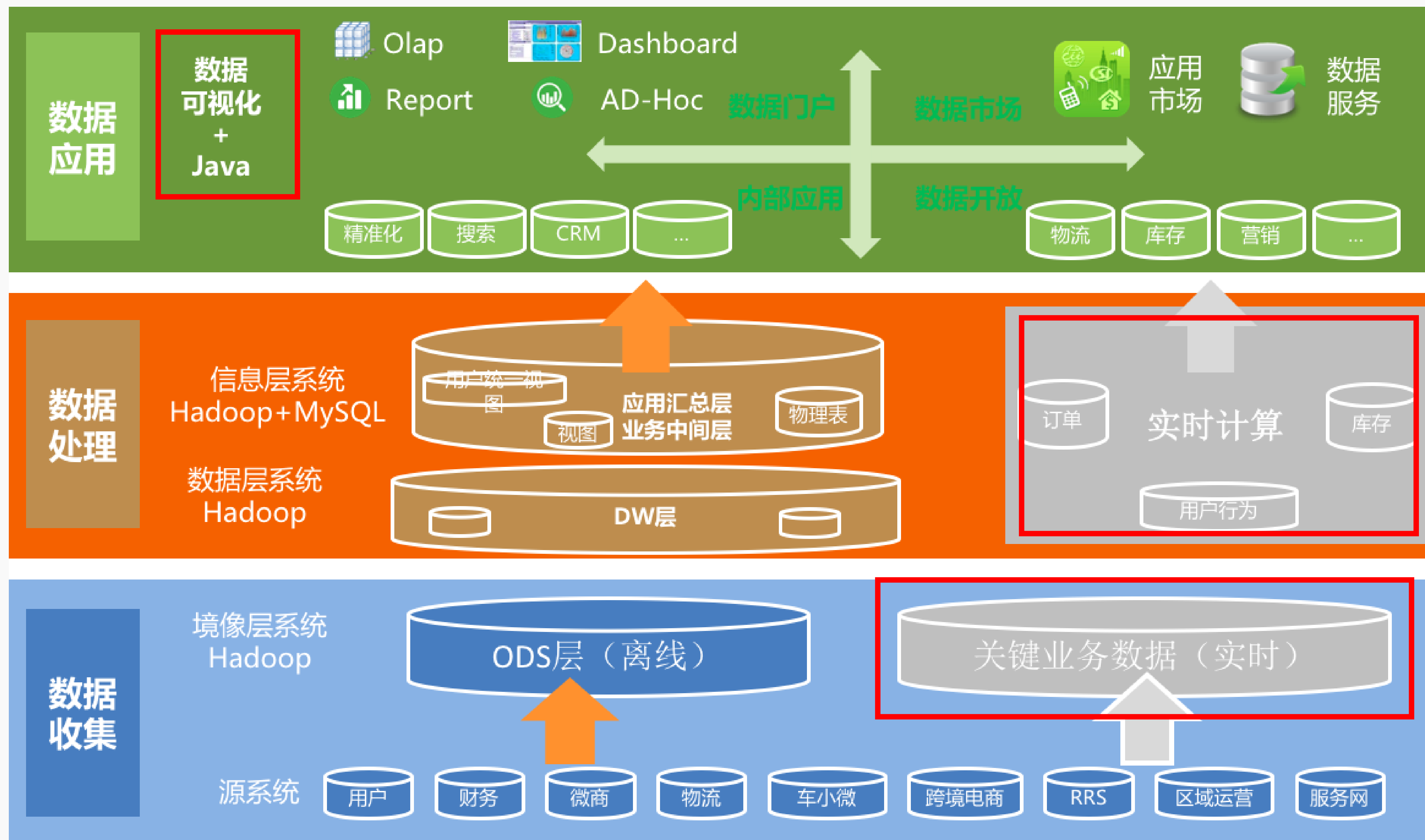


# 公司介绍



- 实时计算平台背景
- 开源技术选型与实践
- 开源技术改造经验

# 背景 - 海尔大数据总体规划



元数据管理

数

数据质量

据

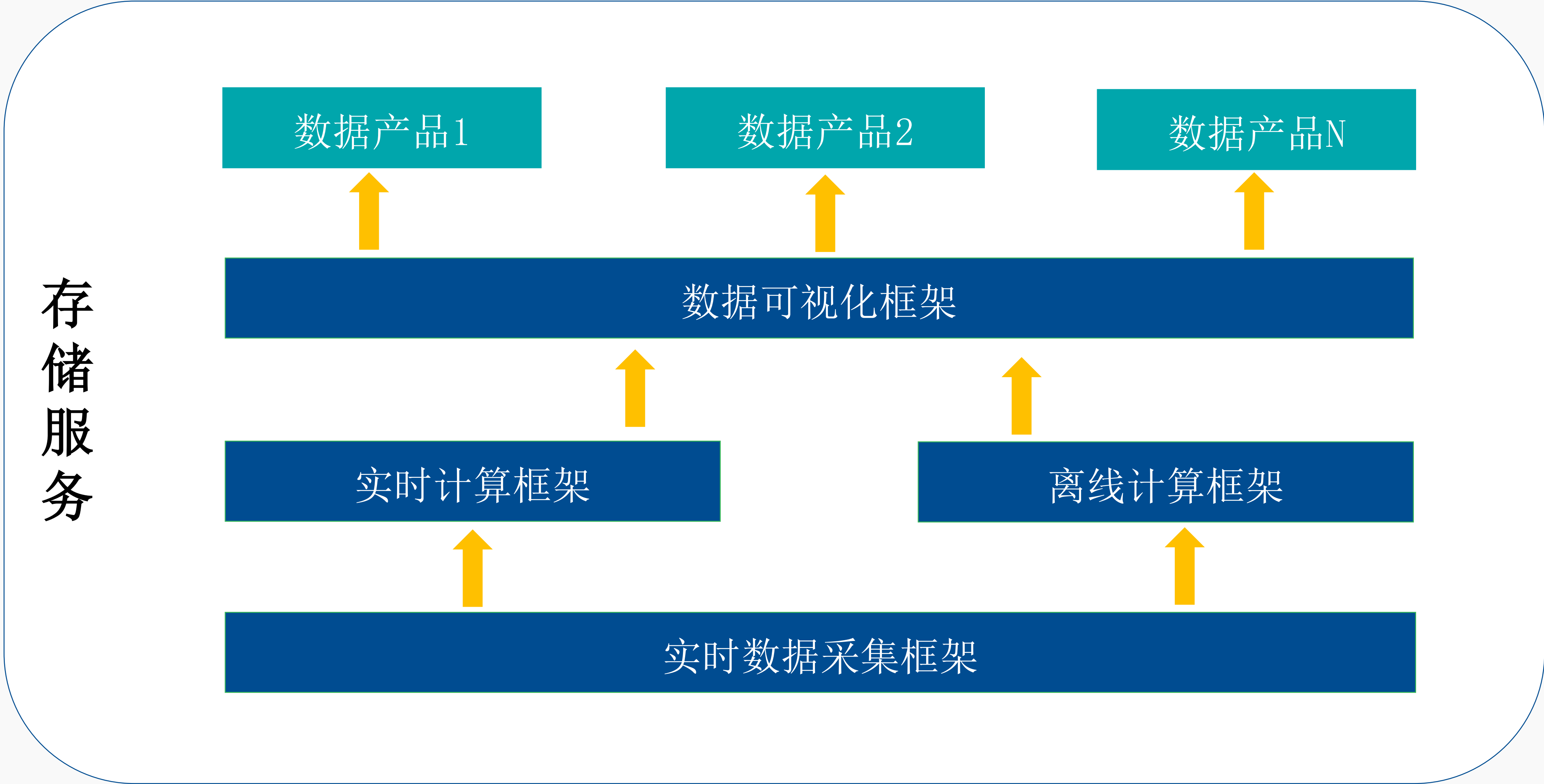
管

数据安全

控

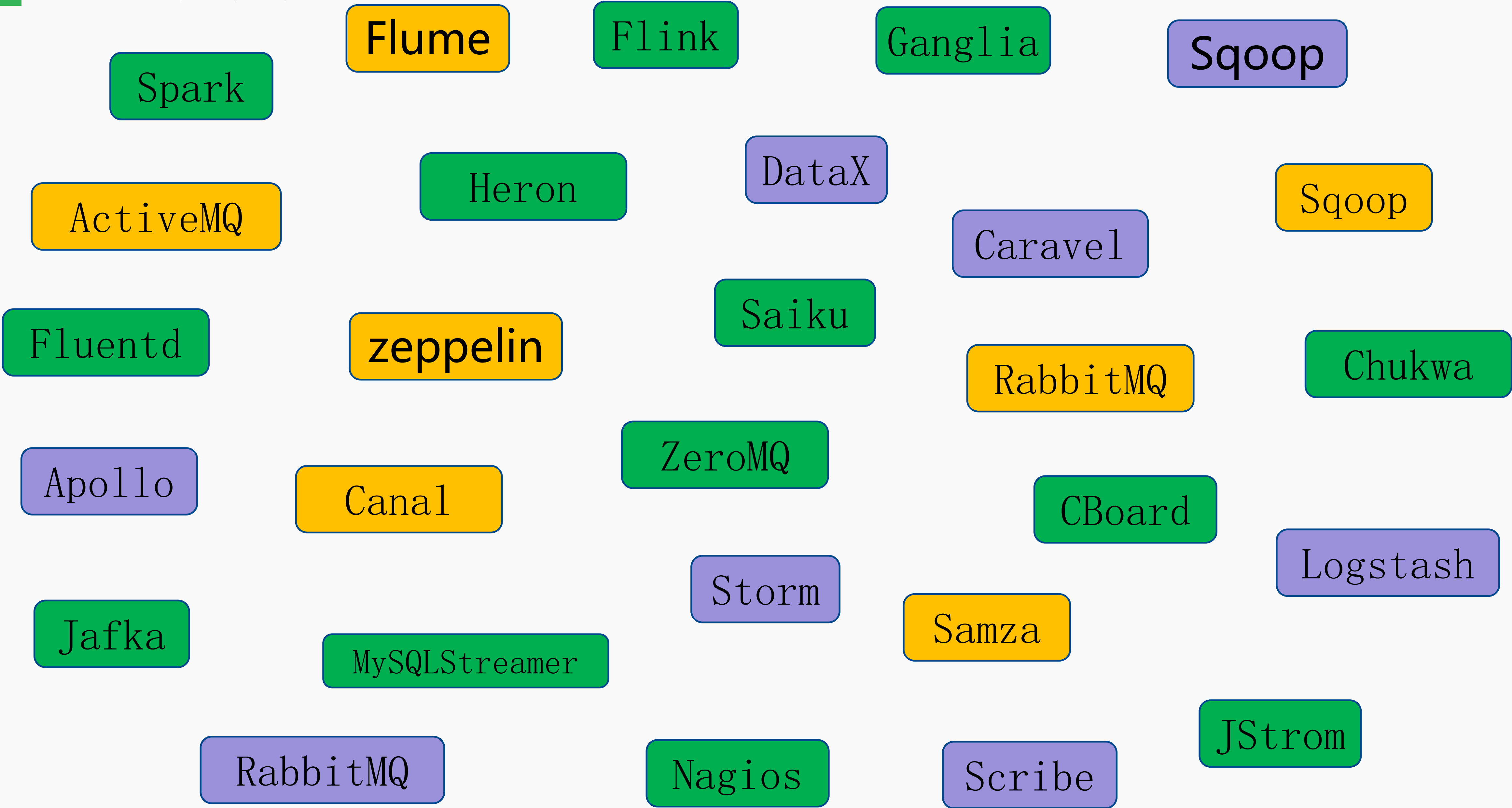


# 实时计算平台框架





# 可选的开源技术





# 实时数据采集技术选型要求

- 完整
- 低延时
- 不影响业务系统性能

# 实时数据采集 - 数据如何获取？

## 代码埋点：

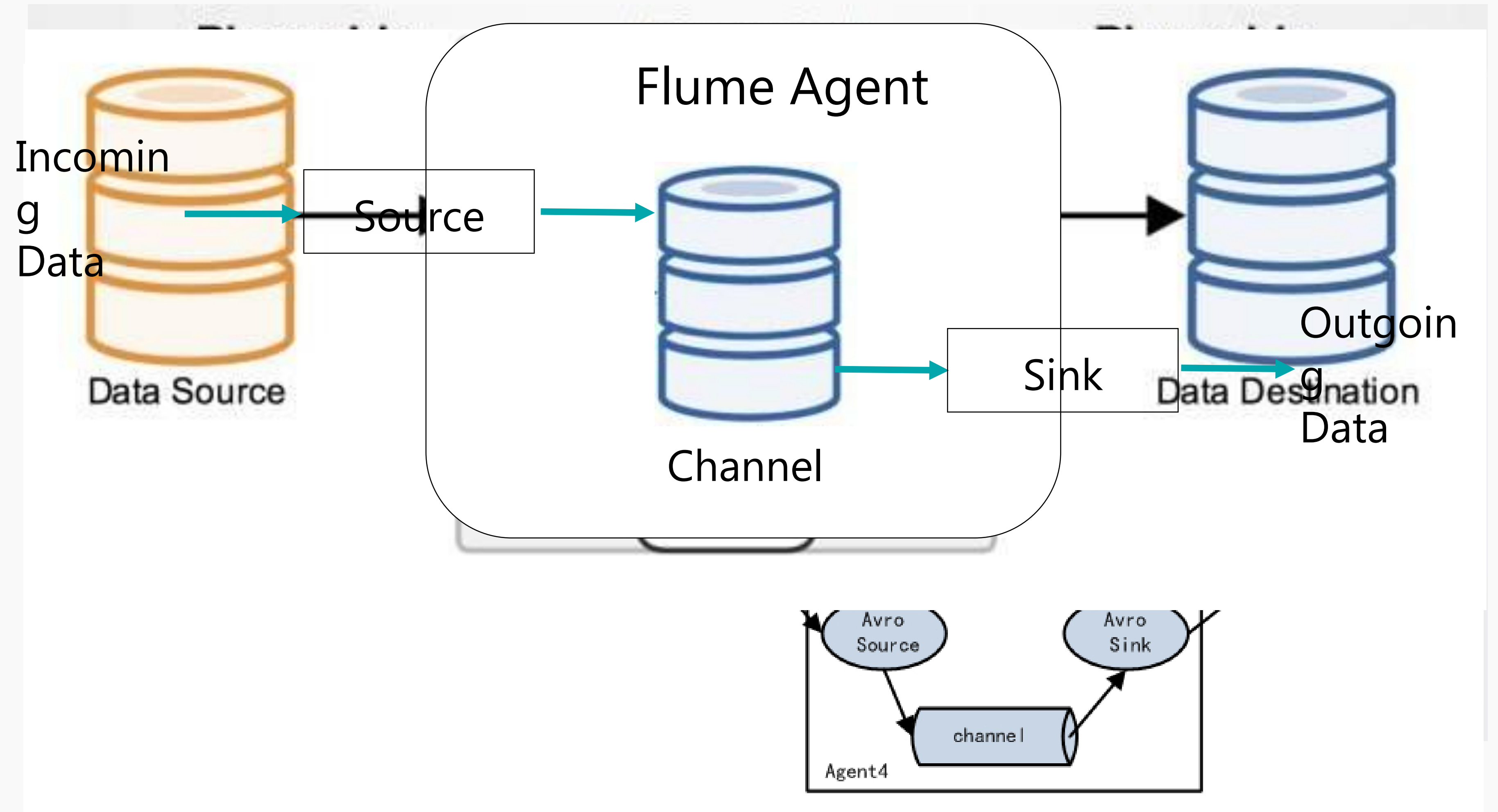
- 优点：采集能力强
- 缺点：时间、人力成本大

## 可视化埋点：

- 优点：成本低，速度快
- 缺点：行为记录信息少，支持的分析方式少

# 日志收集可选技术

- Flume
- Fluentd
- Logstash
- Scribe

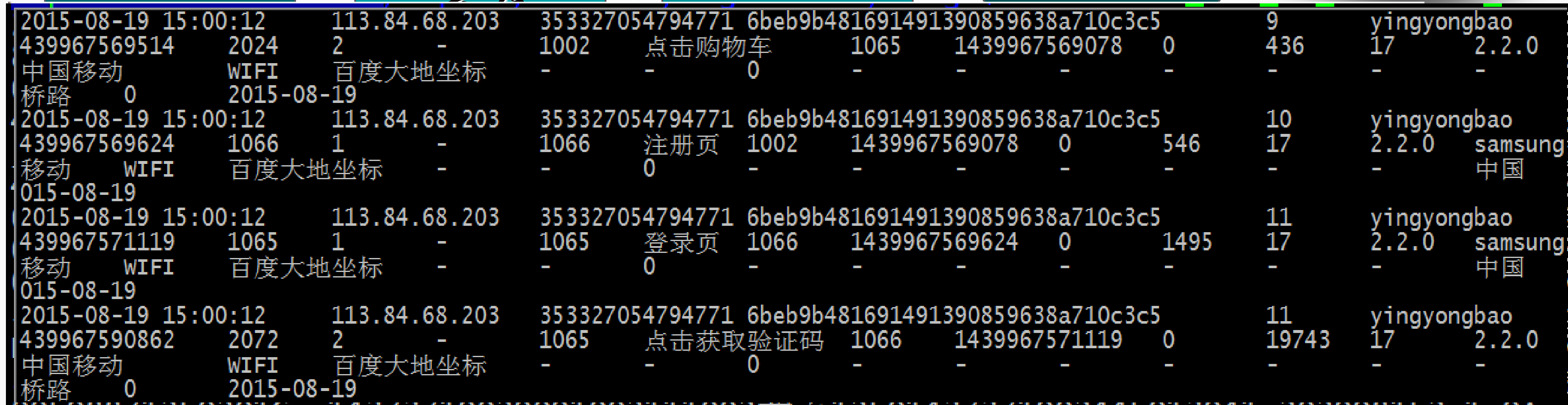


## Ganglia :

- Source : 接收的event数
- Channel : Channel中拥堵的event数
- Sink : 已经处理的event数



100



# 实时数据采集新要求



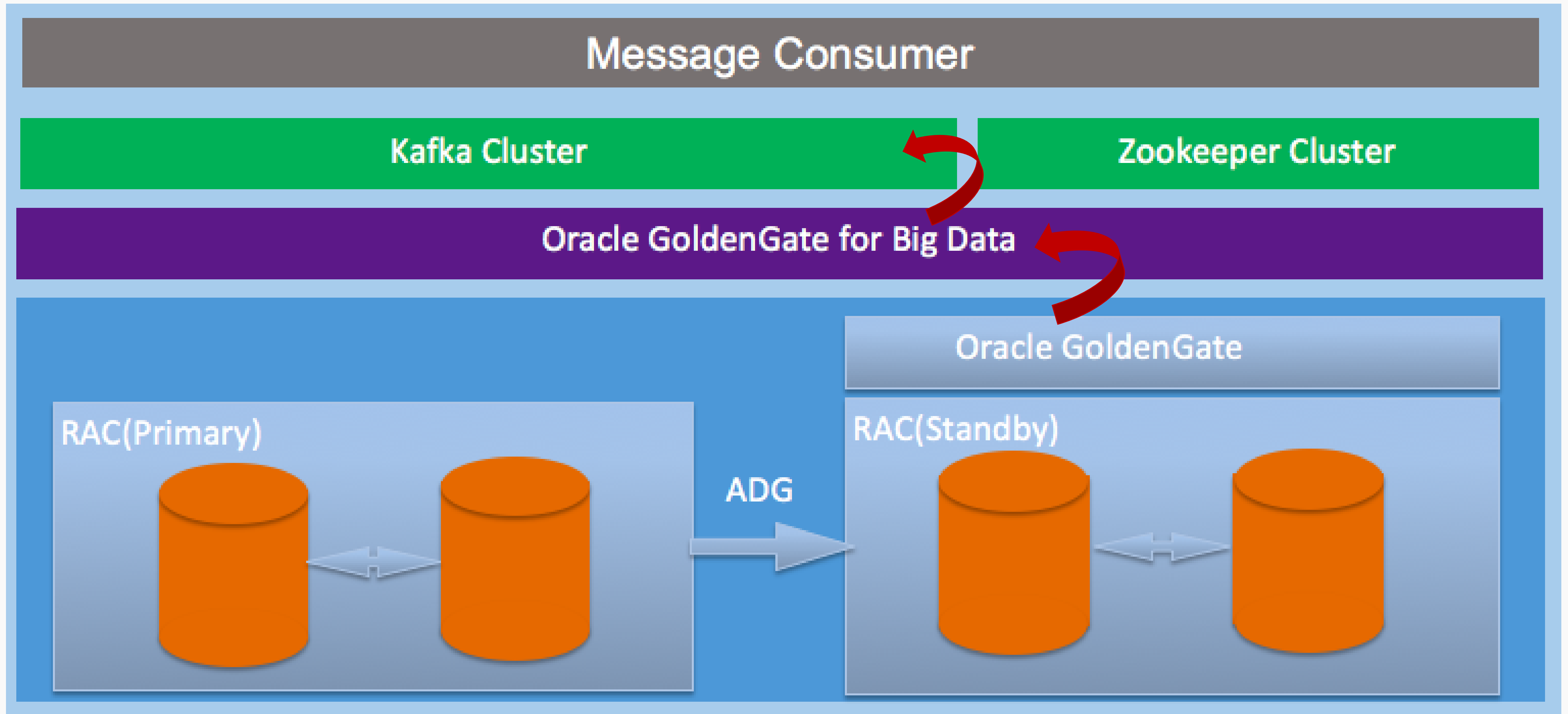
**业务系统无法配合修改，怎么办？**

# Change Data Capture ( CDC )

CDC的几种方法：

- 触发器
- 时间戳
- 全表比对
- 日志对比

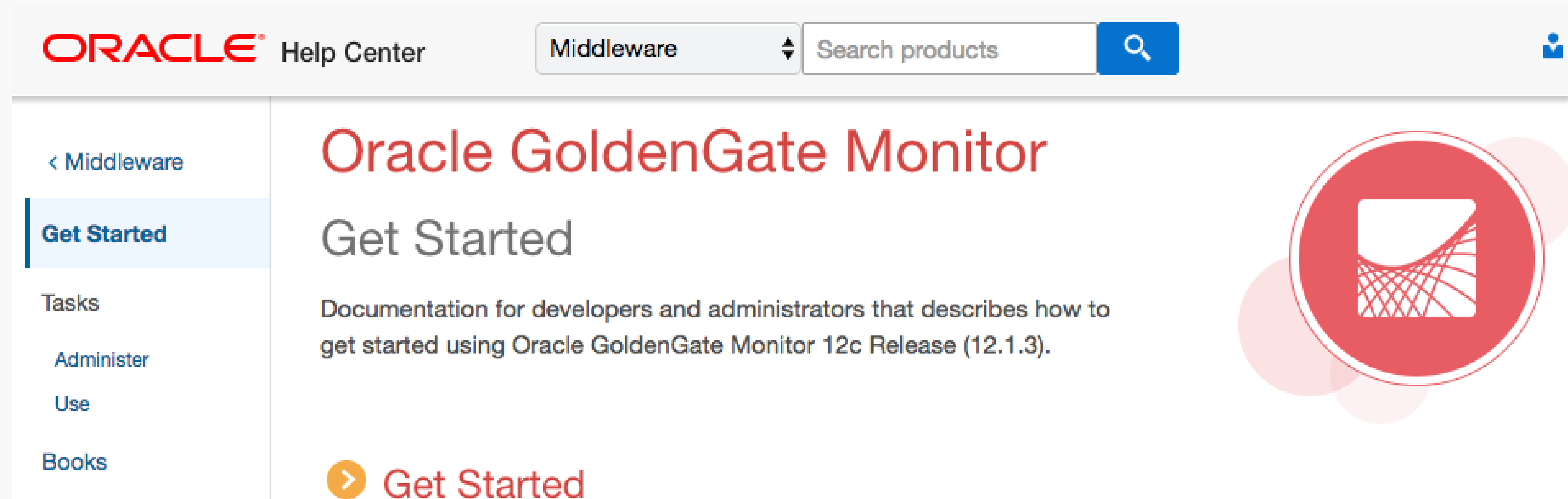
# Oracle CDC建议方案 - OGG





## OGG监控可选方案：

- GoldenGate Director
- GoldenGate Monitor




# Mysql CDC建议方案

Mysql CDC可选方案：

- Canal
- Yelp MySQLStreamer

 **Yelp** / **mysql\_streamer**

 Watch 30

 Star 119

 Fork 38

 Code

 Issues 9

 Pull requests 4

 Projects 0

 Pulse

 Graphs

MySQLStreamer is a database change data capture and publish system.

 1,070 commits

 1 branch

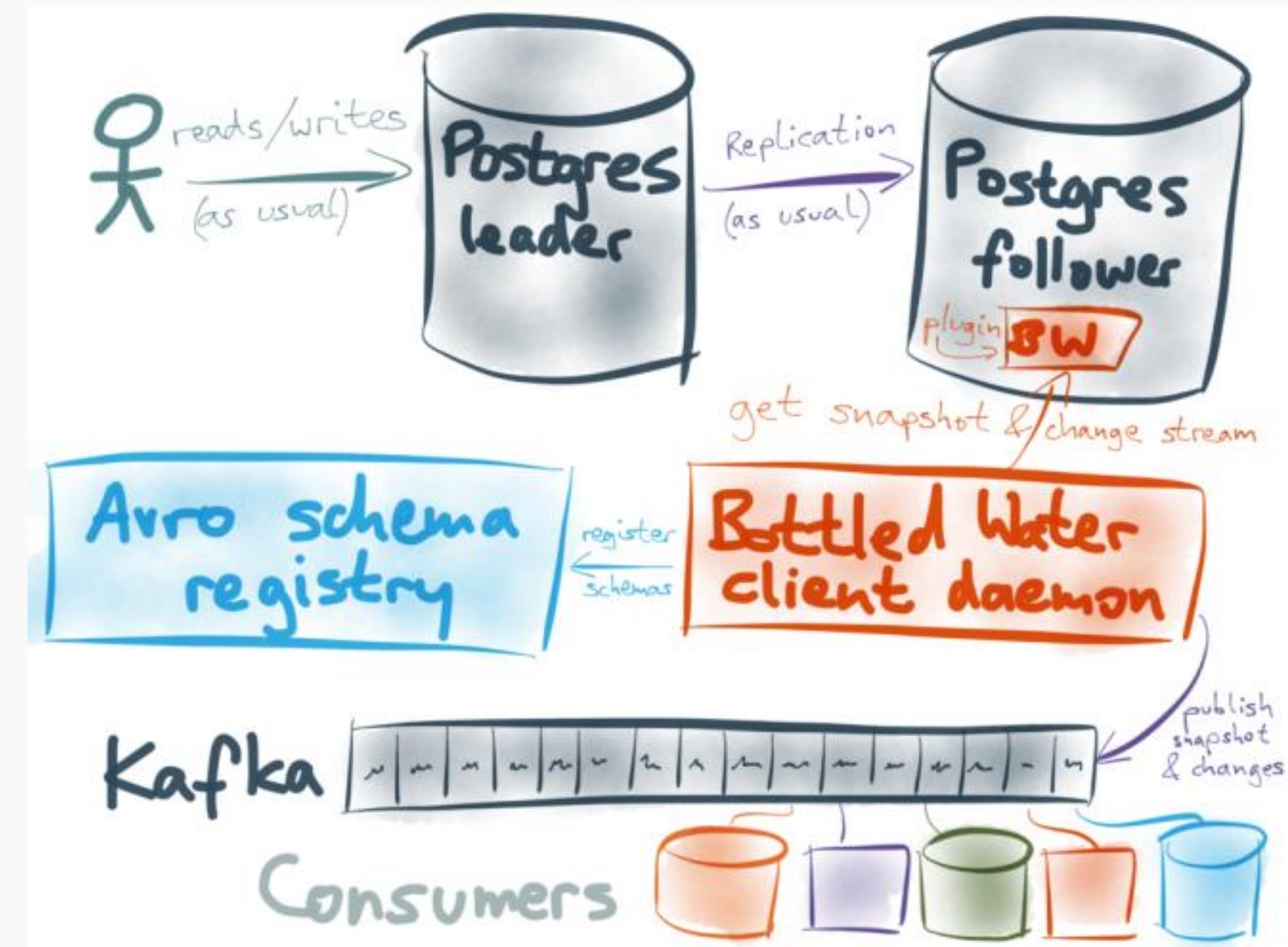
 0 releases

 15 contributors


# Postgresql CDC 建议方案

## BottledWater:

- 支持postgresql 9.4及以后版本
- 几乎不影响数据库本身性能
- 事务一致性的输出
- 容错




上图摘自confluent官网


 [Features](#) [Business](#) [Explore](#) [Pricing](#)

This repository


Search

[Sign in](#) or [Sign up](#)


 [confluentinc / bottledwater-pg](#)

 Watch


132


 Star


1,054


 Fork


99


 Code


 Issues 40

 Pull requests 1

 Projects 0

 Wiki

 Pulse

 Graphs

Change data capture from PostgreSQL into Kafka

<http://blog.confluent.io/2015/04/23/b...>

# 为何要引入消息队列

Flume适合：

- 日志收集
- 日志传输
- 拦截数据

消息队列适合：

- 消息持久化
- 解耦

推荐：Flume + 消息队列



# 消息队列可选方案

- Kafka
- Jafka
- RocketMQ
- RabbitMQ
- ActiveMQ
- Apollo

# 消息队列使用场景及选型

**事务可靠性场景（ ActiveMQ/RabbitMQ/RocketMQ ）：**

- 重要业务数据的异步处理

**吞吐量优先场景（ Kafka/RocketMQ ）：**

- 大规模数据的传输和收集

# Kafka web**监控管理界面**

## Kafka web**管理界面**：

- Kafka Web Console
- Kafka Manager
- KafkaOffsetMonitor
- Uber开源Kafka监控工具--Chaperone

# 流式计算可选方案

## 仅流处理框架：

- Storm
- JStorm
- Samza
- Heron

## 混合框架：

- Spark
- Flink



# Core Storm or Storm Trident ?

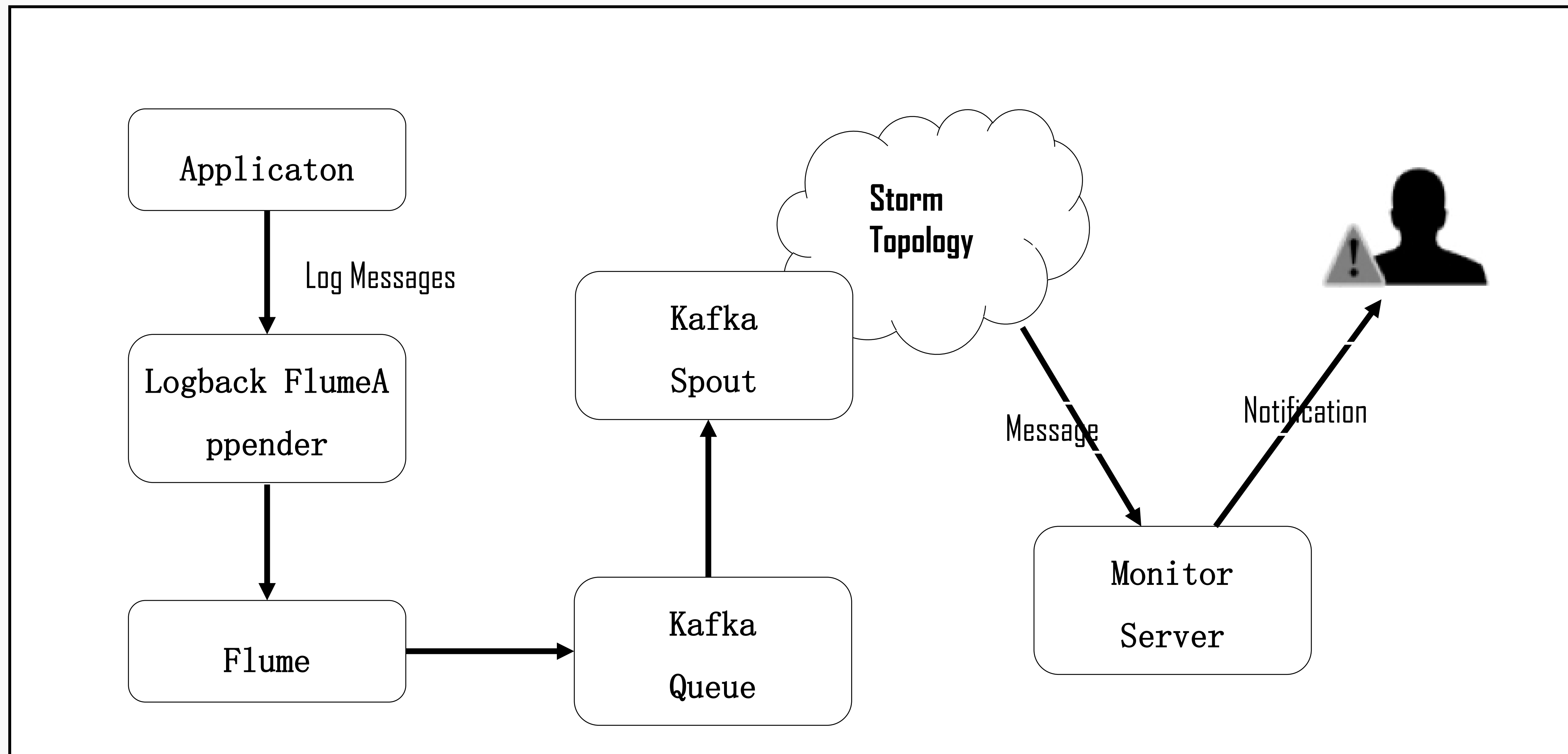
## Storm Trident:

- Batch
- Exactly-once

## 选型考虑 :

- 性能
- 状态

# Storm流式日志处理常见架构

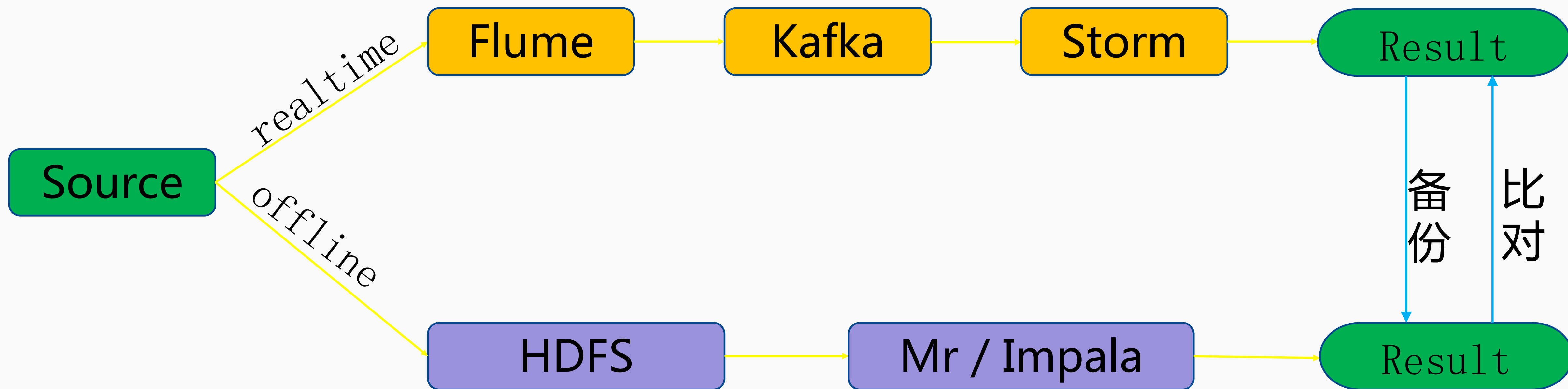


# 实时计算质量保证



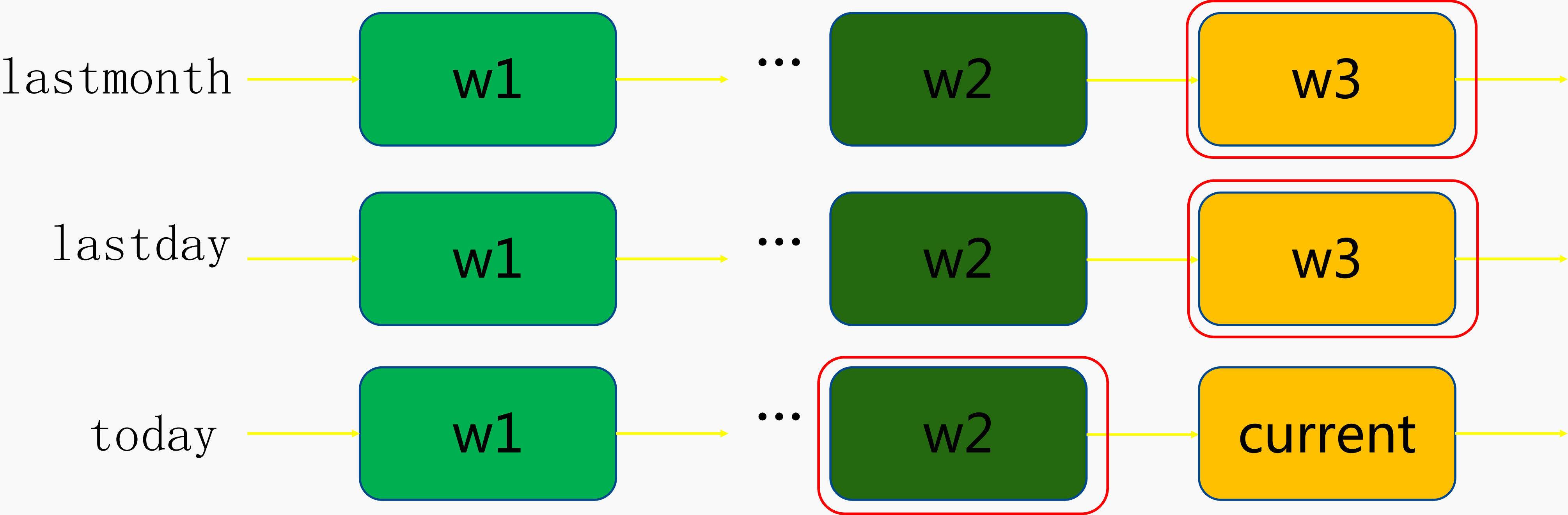
实时计算结果正确性如何验证？

## 方案一：结果导向





# 方案二：经验值



# ■ 海尔实时计算平台技术选型与实践

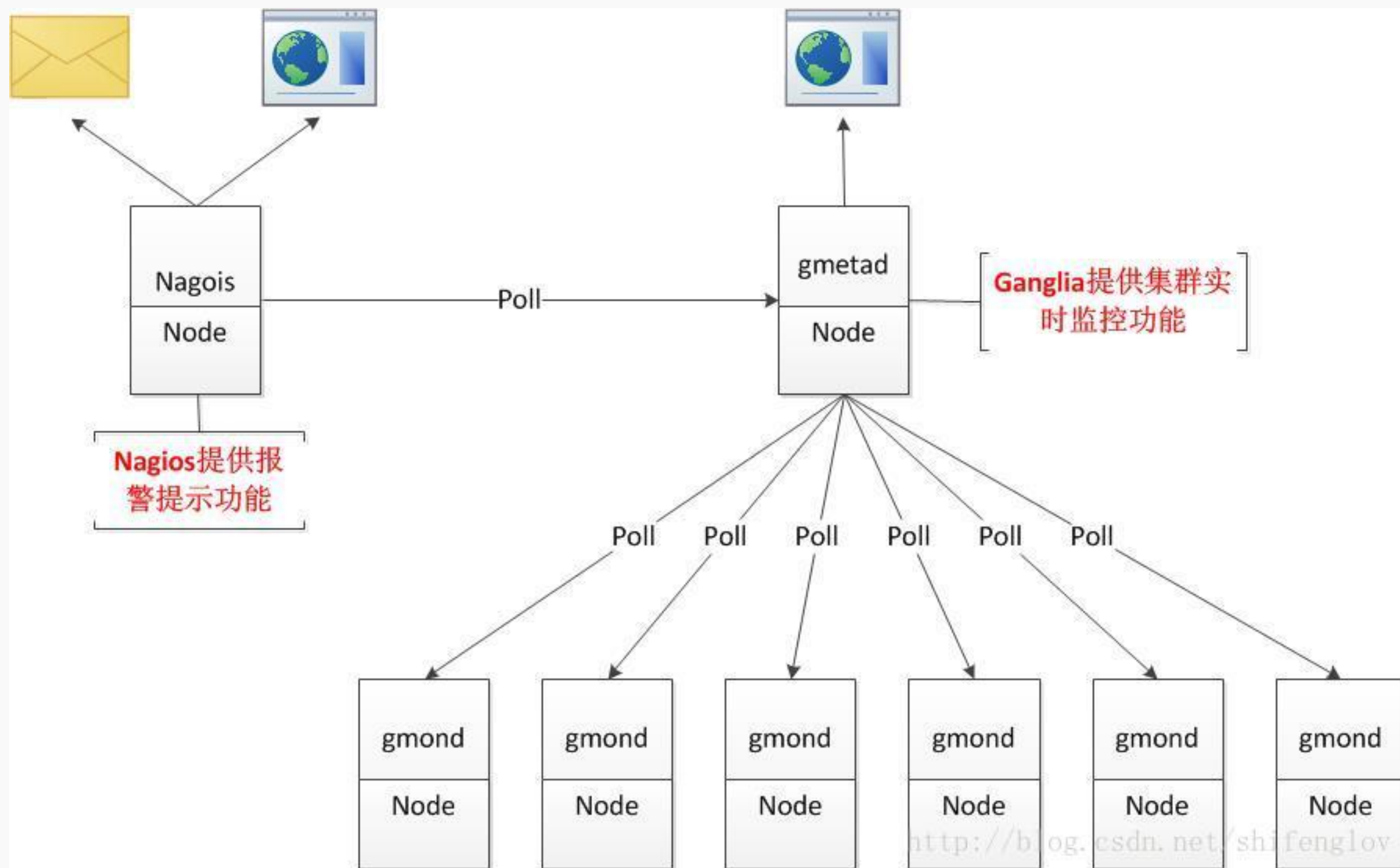


实时计算平台如何监控和报警？



# 平台监控

- Ganglia
- Nagios



# 开源技术改造经验 - 数据可视化

# 数据可视化目标



目标：为公司所有数据分析类项目，提供数据洞察和展示工具。

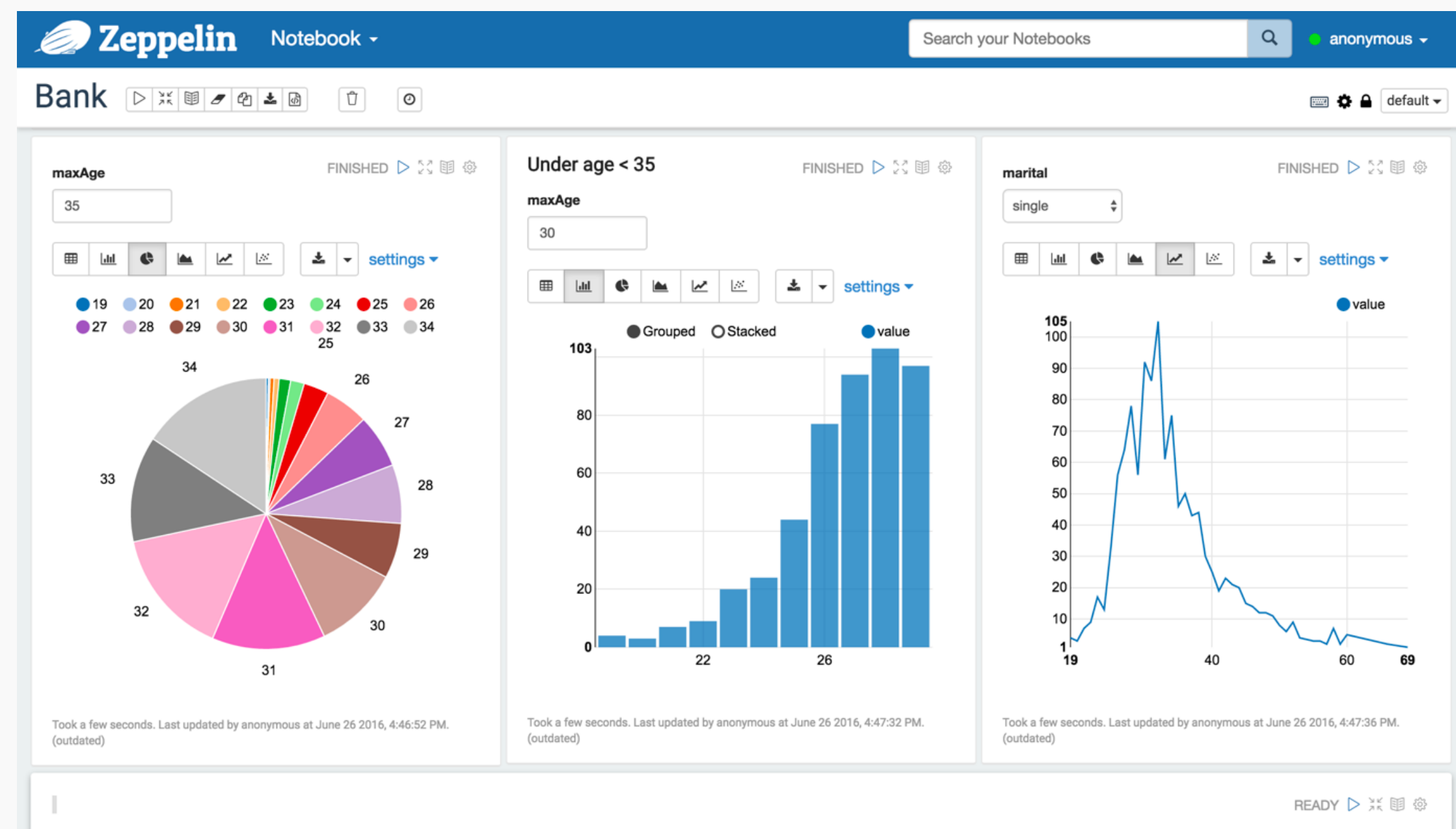
# 数据可视化可选方案

- Caravel
- Saiku
- Zeppelin


# Zeppelin主要特性

主要特性：

- 数据提取
- 数据发现
- 数据分析
- 数据可视化和协作



# Zeppelin优势 - 多语言支持

 Notebook ▾

Search your Notebooks 🔍

anonymous ▾

Elasticsearch Tests

▶ ⌵ 📖 📝 ⬇ ⬆ ⬇

🗑 ⌚

🔍 ⚙ 🔒 default ▾

ⓧelasticsearch

FINISHED ▶ ⌵ 📖 ⚙

```
search /logs { "fields": [ "date", "request.headers" ], "query": { "match": { "status": 500 } } }
```

📊

📈

📉

📉

📈

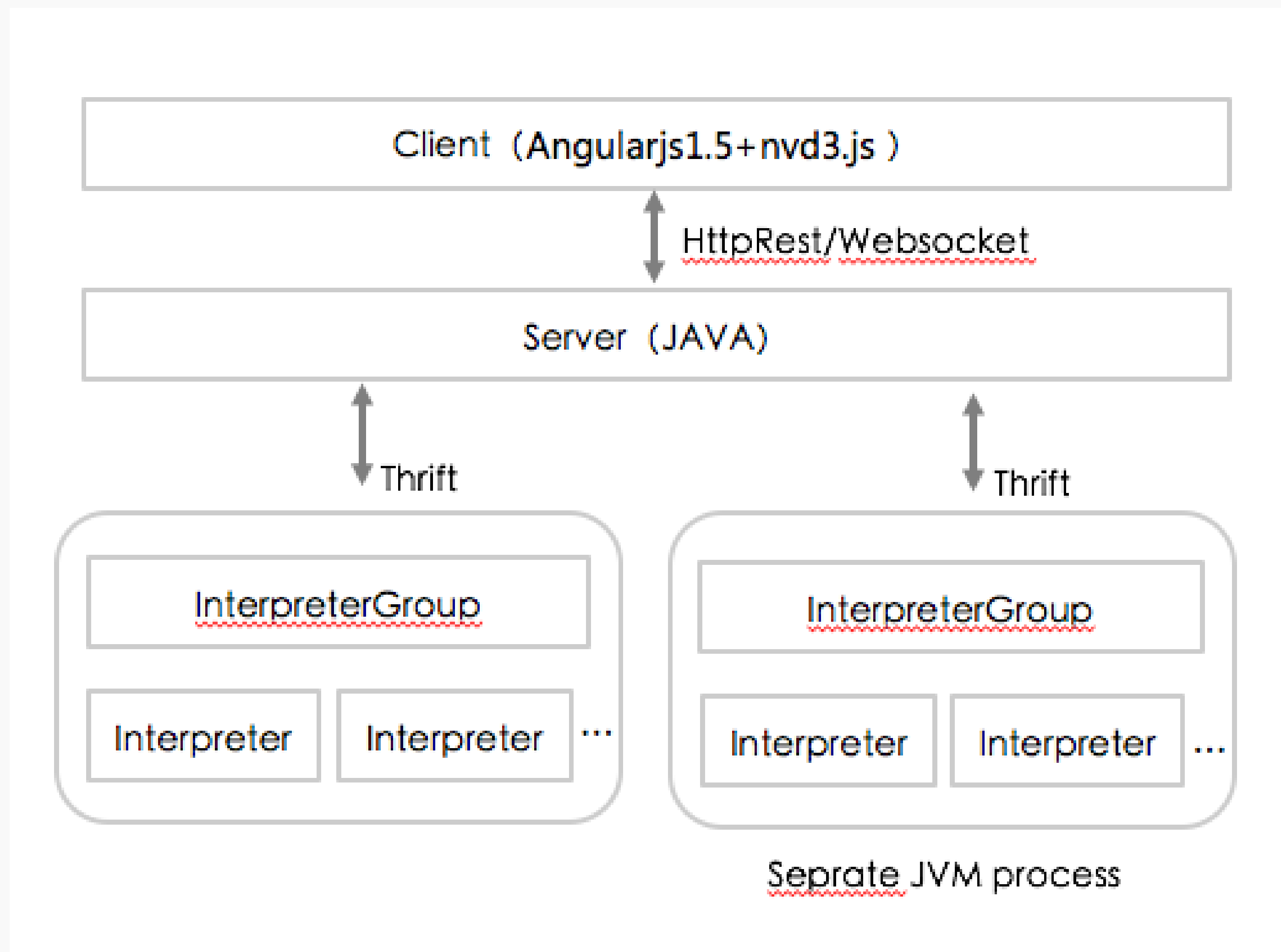
📉

Took 6 seconds (outdated)



## Zeppelin技术：

- 基于JAVA开发
- 前端Angularjs1.5 + nvd3.js



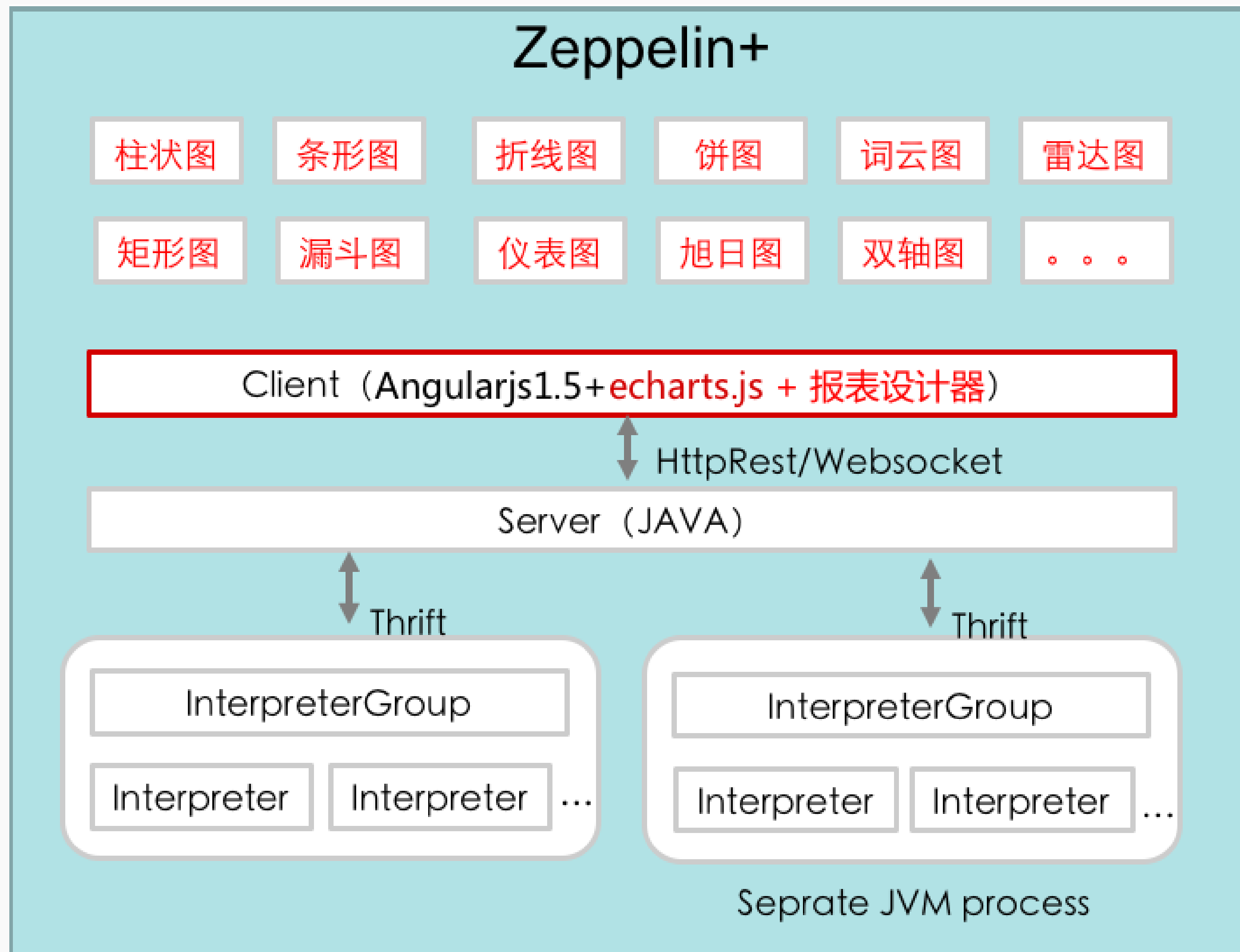
# Zeppelin改造

## Server :

- 支持echarts json数据
- 前端的计算逻辑都改到后端

## 前端 :

- 前端nvd3.js替换成echarts
- 更强大易用的报表设计器



# 改造后的zeppelin报表设计器

select  
ORDER\_NO as 订单号,  
SOURCE as 订单来源

FINISHED ▶ ⚙

字段:  
T 订单号  
T 订单来源  
T 原始订单  
📅 订单时间  
📅 年  
📅 季度  
📅 月  
📅 日  
📅 年聚合  
📅 季度聚合  
📅 月聚合  
📅 日聚合  
T 订单类型  
T 所属中心  
T 所属仓编码  
T 批次  
T 配送方式  
T 省  
T 市  
T 区  
T 取消标记  
T 目的国代码  
T 订单标记  
T 市场等级  
# 数量

维度  
数值 📊 商家 (去重计数) ×  
数值 📈 品牌 (去重计数) ×

全部  
商家  
品牌

颜色  
大小  
标签

工具提示  
详细信息

双轴图

70  
60  
50  
40  
30  
20  
10  
0

TMALL  
JINGD  
YIJIAN  
XIAOMI  
EGYMCOM  
QCJ

13  
13  
8  
1  
2  
1  
2  
1  
1  
1

13  
13  
8  
1  
2  
1  
2  
1  
1  
1

商家  
品牌

图表标题  
双轴图

图表参数设置

图表备注

图表颜色主题

图表类型

图内筛选器

坐标轴  
标题  
单位  
最大值  
最小值  
间隔  
☐ 是否显示

次轴坐标轴  
标题  
单位  
最大值  
最小值  
间隔  
☐ 是否显示

数据执行区、多层钻取

行列控制区

2、双轴异图支持

3、高级计算：同比、环比

4、度量维度排序

1、图表参数设置

2、图内筛选器

3、图表辅助线

4、图表预警

4、图表坐标轴控制

5、图表图例显示控制

1、字段设置

2、添加计算字段

3、字段分组

1、图表颜色分组

2、图形大小控制

3、图表标签控制

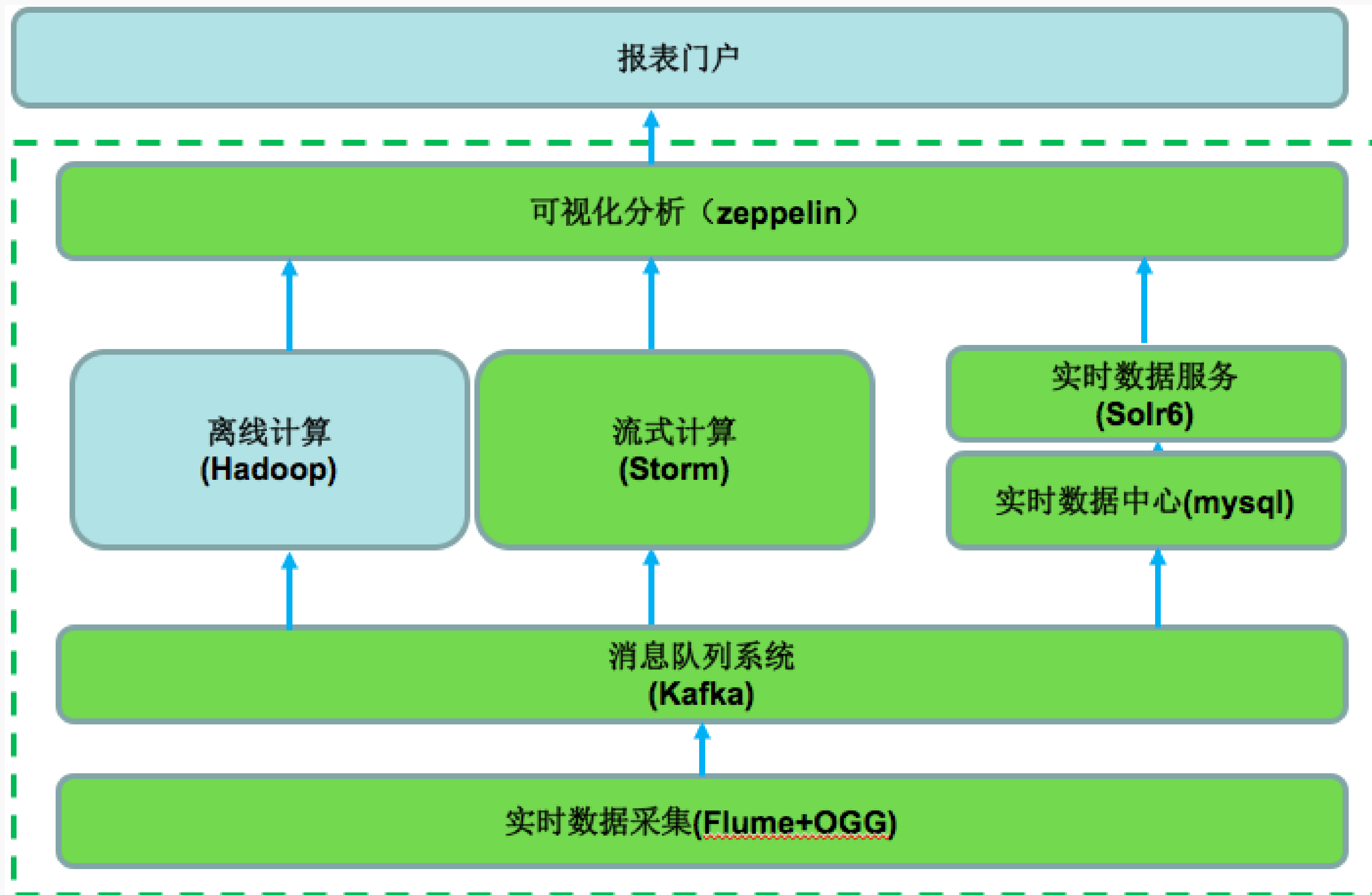
4、图形提示控制

图表辅助设置区

各种消息队列对比.pdf  
开源MQ技术介绍.pptx  
Kafka+FlumeNG+Stor....pptx  
Kafka+FlumeNG+Stor....pptx  
培训课程Kafka.ppt  
苏宁实时计算公开版.pptx

显示全部 ×

## 选型后的实时平台技术架构





关注QCon微信公众号，  
获得更多干货！

# Thanks!



主办方 **Geekbang** > **InfoQ**  
极客邦科技