



2017第八届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2017

# 数据库智能优化系统 探索与实践

乔红麟(洪林)

# Who am I?

乔红麟，花名洪林。2015年3月加入阿里巴巴数据库技术团队，目前负责阿里巴巴数据库智能优化产品CloudDBA开发。2007年毕业于北京邮电大学，先后从事过服务器虚拟化(Xen/KVM/VMware vSphere)，存储备份(EMC/DataDomain)，分布式存储等相关技术开发工作。

Mail: [kenny.qhl@alibaba-inc.com](mailto:kenny.qhl@alibaba-inc.com)  
[qiaohonglin@gmail.com](mailto:qiaohonglin@gmail.com)



# Agenda

Why

阿里巴巴数据库服务的诉求

What

数据库智能优化系统的思考

How

CloudDBA关键技术

Now

数据库智能优化系统的探索

Q & A

Q & A

# Agenda

Why

阿里巴巴数据库服务的诉求

What

数据库智能优化系统的思考

How

CloudDBA关键技术

Now

数据库智能优化系统的探索

Q & A

Q & A

# 阿里数据库场景及服务诉求

## 场景分析

- 交易/支付型应用(OLTP)
- 业务高速发展，数据库规模急剧增长
- 数据规模大，业务场景复杂，业务负载变化常态化
- 海量性能和运维数据
- 业务热点(双11/双12/618)明显

## 服务诉求

- 持续的数据库稳定性/可用性要求
- 规模化运维能力，DBA专家服务具备良好的扩展性
- 持续主动优化，全局规模化优化，降低计算/存储成本
- 如何基于数据实现自诊断自优化
- 极致的弹性伸缩和资源调度能力

# Agenda

Why

阿里巴巴数据库服务的诉求

What

数据库智能优化系统的思考

How

CloudDBA关键技术

Now

数据库智能优化系统的探索

Q & A

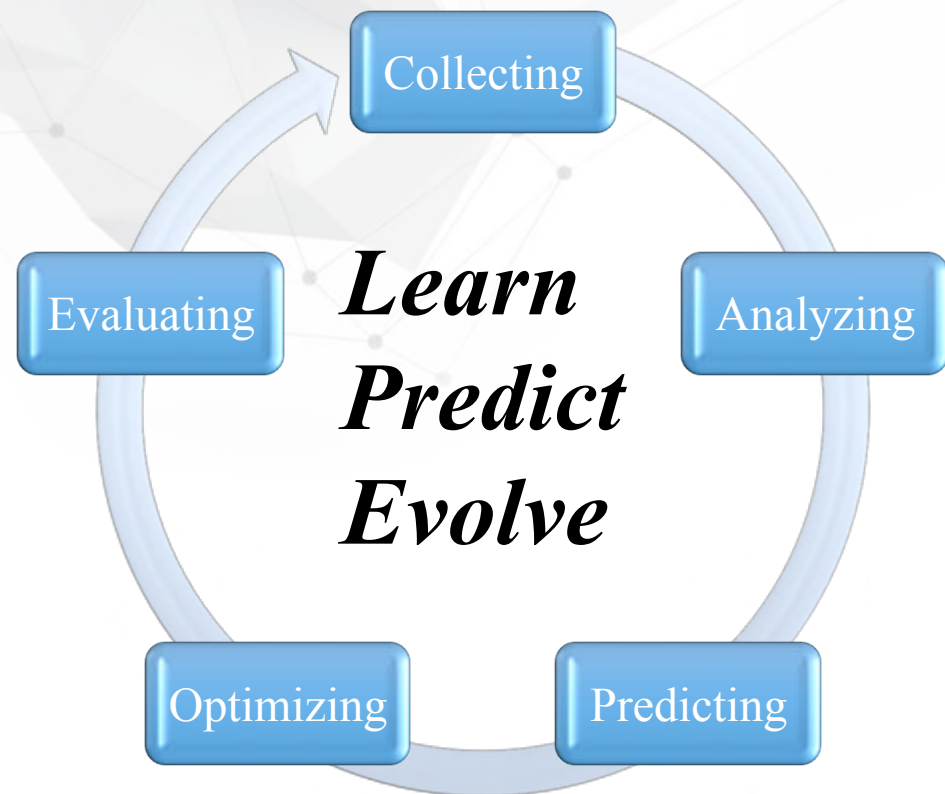
Q & A

# 传统数据库优化方式的局限

- 被动优化(监控/报警/慢SQL/应用报错等)
- 耗时耗力且低效，较难形成闭环
- 受限于人的局限性,复杂场景缺乏扩展性
- 决策过程中数据价值缺失
- DBA专家紧缺与数据库服务要求之间的矛盾越来越突出
- .....



# 数据库智能优化系统的特点



数据驱动

流程闭环

持续优化

自诊断

自优化

.....



# 数据库智能优化系统的价值



## 专家经验产品化

DBA专家优化经验转服务产品化，降低人工成本，提高服务质量和效率



## 信息透明

向开发人员提供全面专业的数据库信息分析和展示



## 自助化服务

全面专业的诊断优化建议，诊断流程闭环，提供自助优化服务



## 主动优化

持续主动优化数据库性能，提高数据库稳定性



## 降低成本

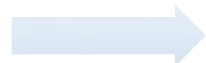
发现规模化优化点，优化资源管理，降低计算和存储成本



## 自诊断、自优化

基于海量数据和机器学习的自诊断、自优化能力

# CloudDBA@Alibaba



CloudDBA

用户身边的数据库专家

# Agenda

Why

阿里巴巴数据库服务的诉求

What

数据库智能优化系统的思考

How

CloudDBA关键技术

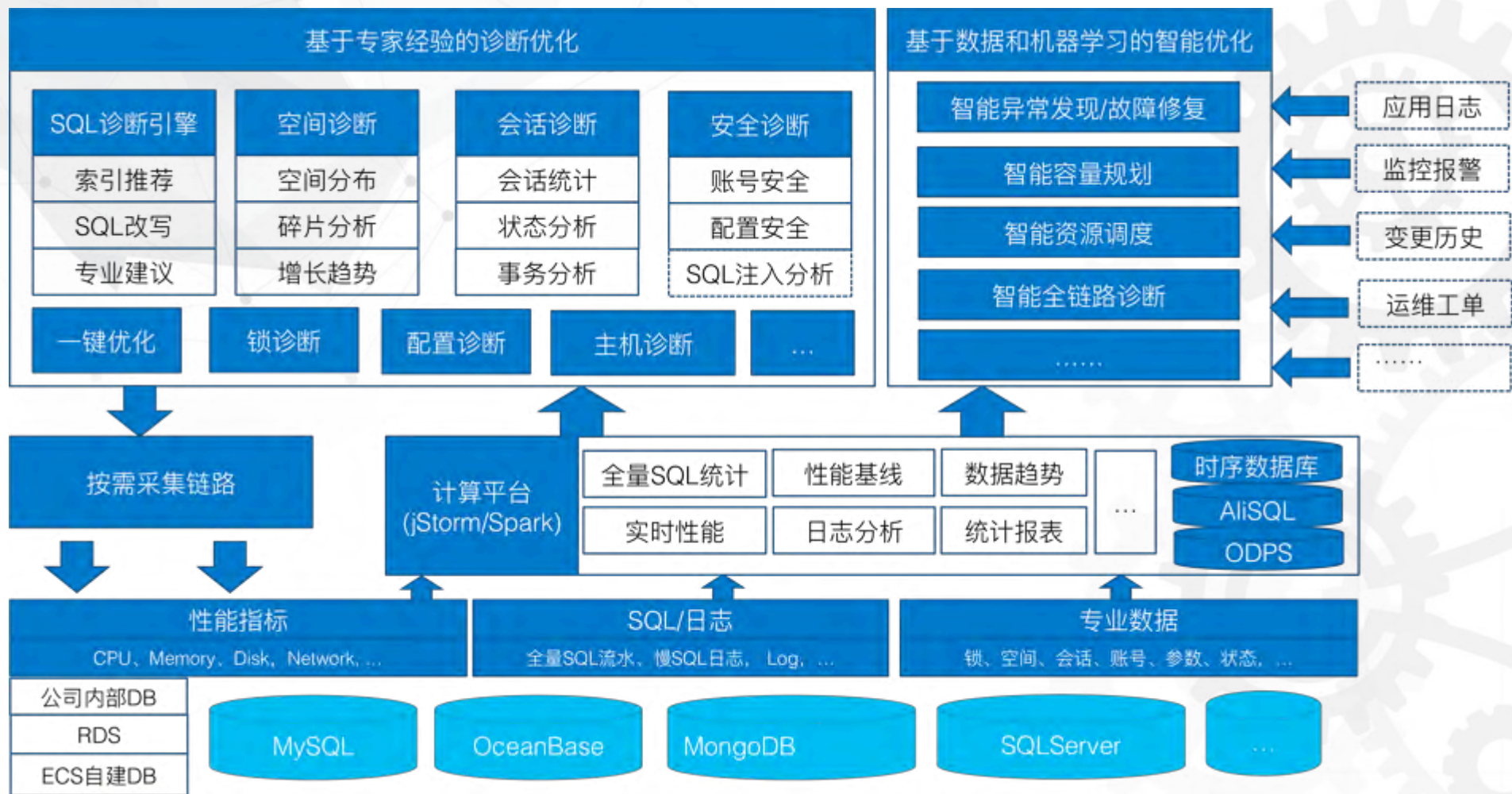
Now

数据库智能优化系统的探索

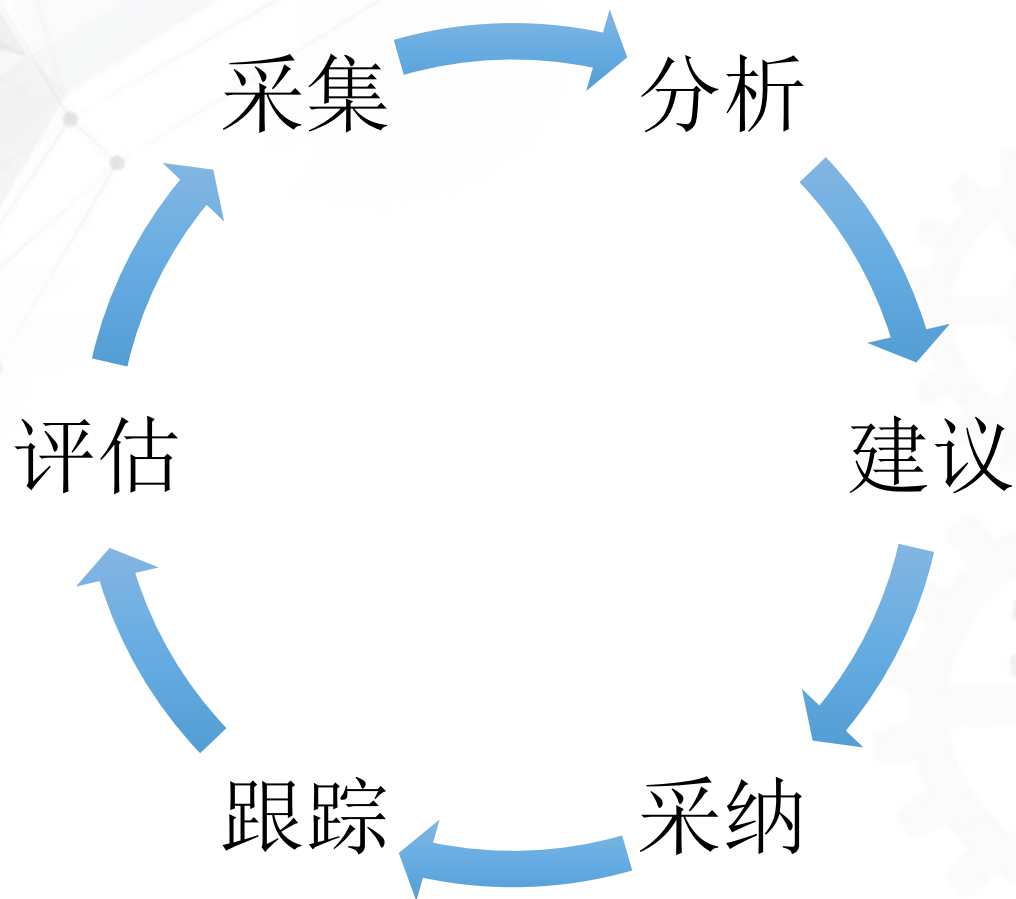
Q & A

Q & A

# CloudDBA – 用户身边的数据库专家



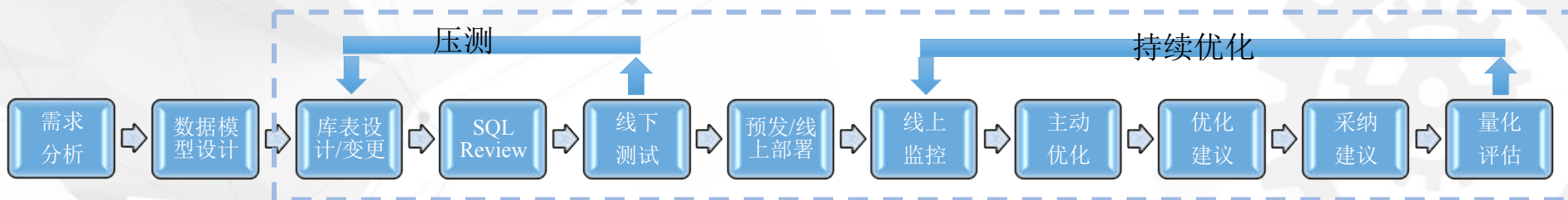
# 优化闭环



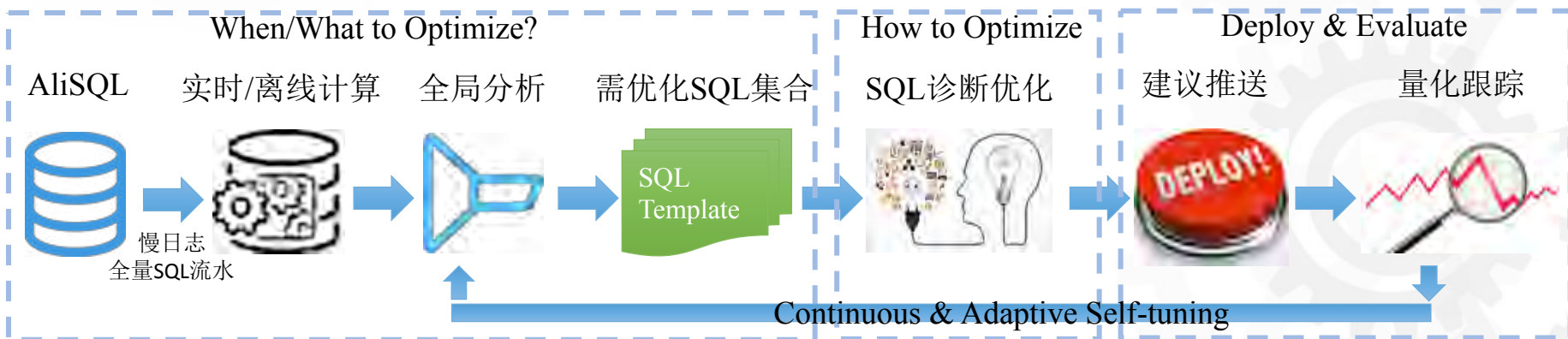


# SQL优化 - 闭环

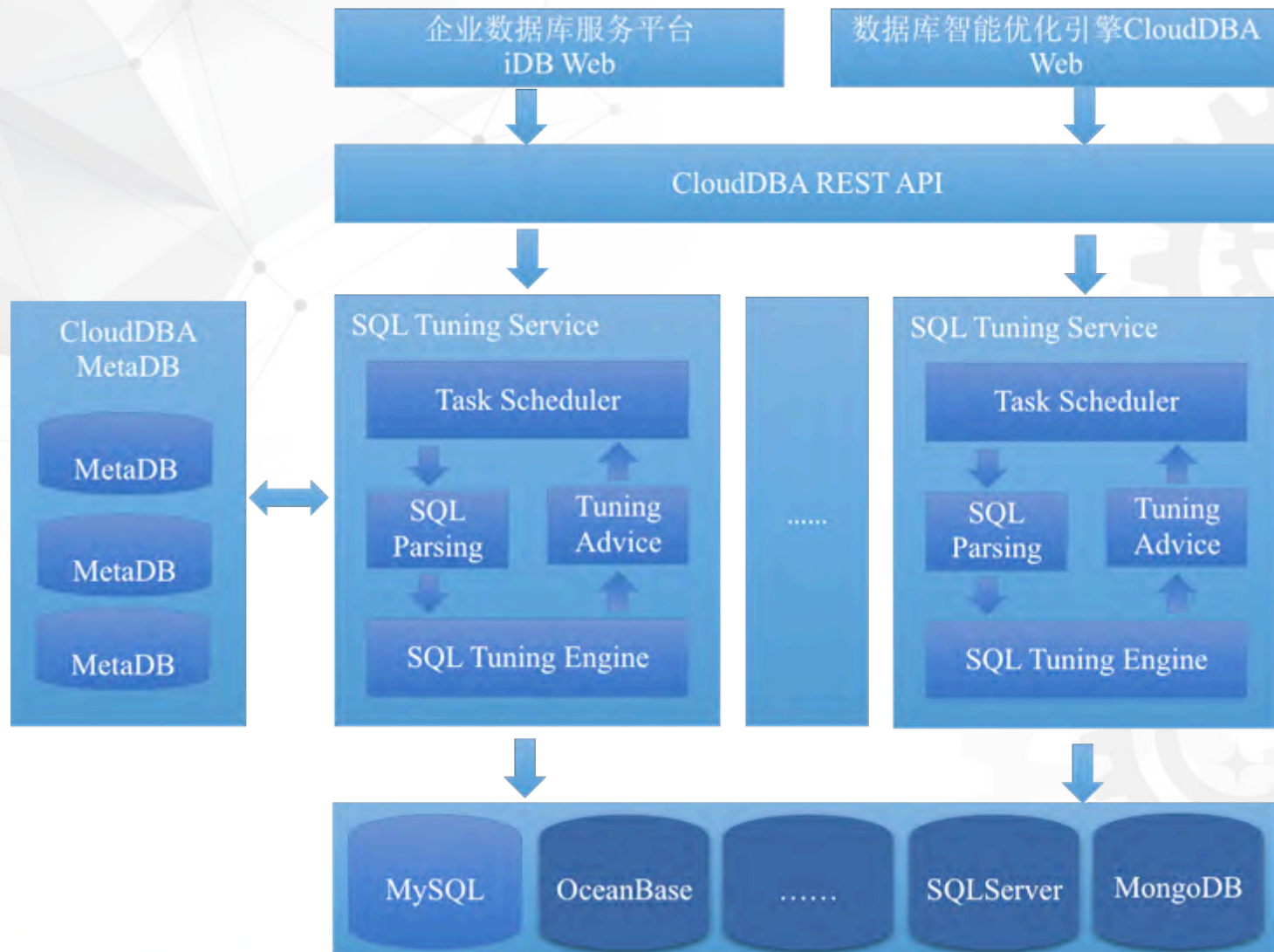
## 开发流程闭环



## 优化流程闭环

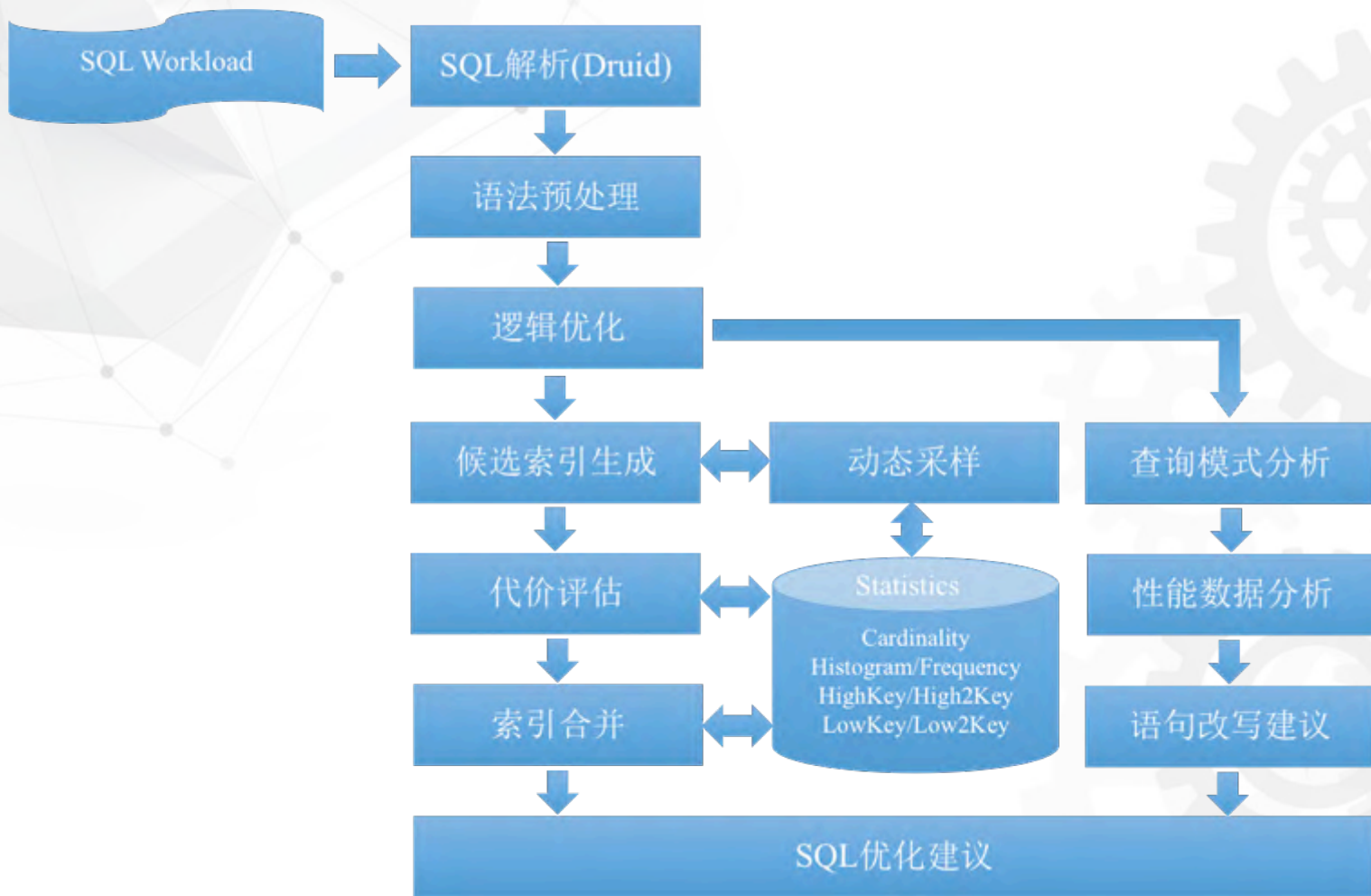


# SQL优化 - 引擎

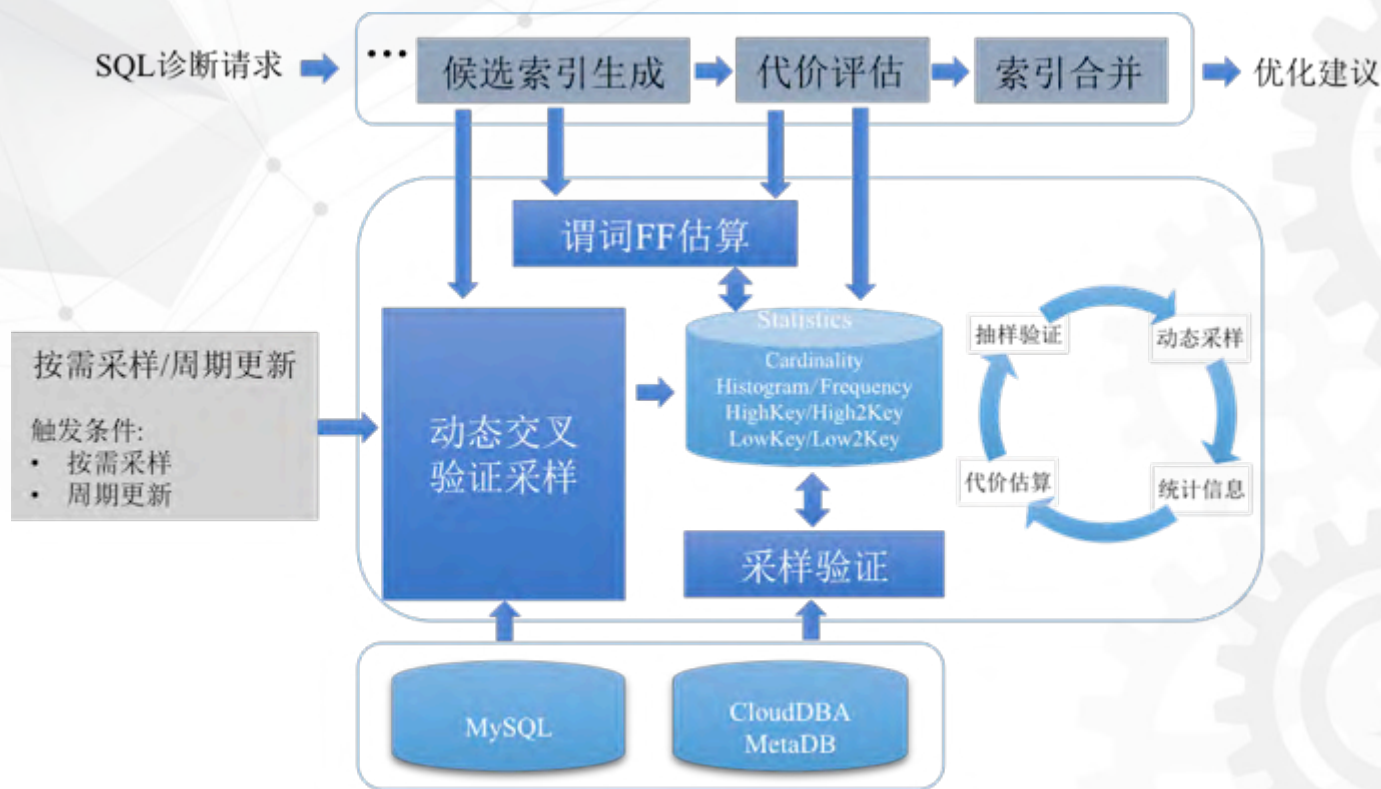




# SQL优化 - 流程



# SQL优化 - 动态采样



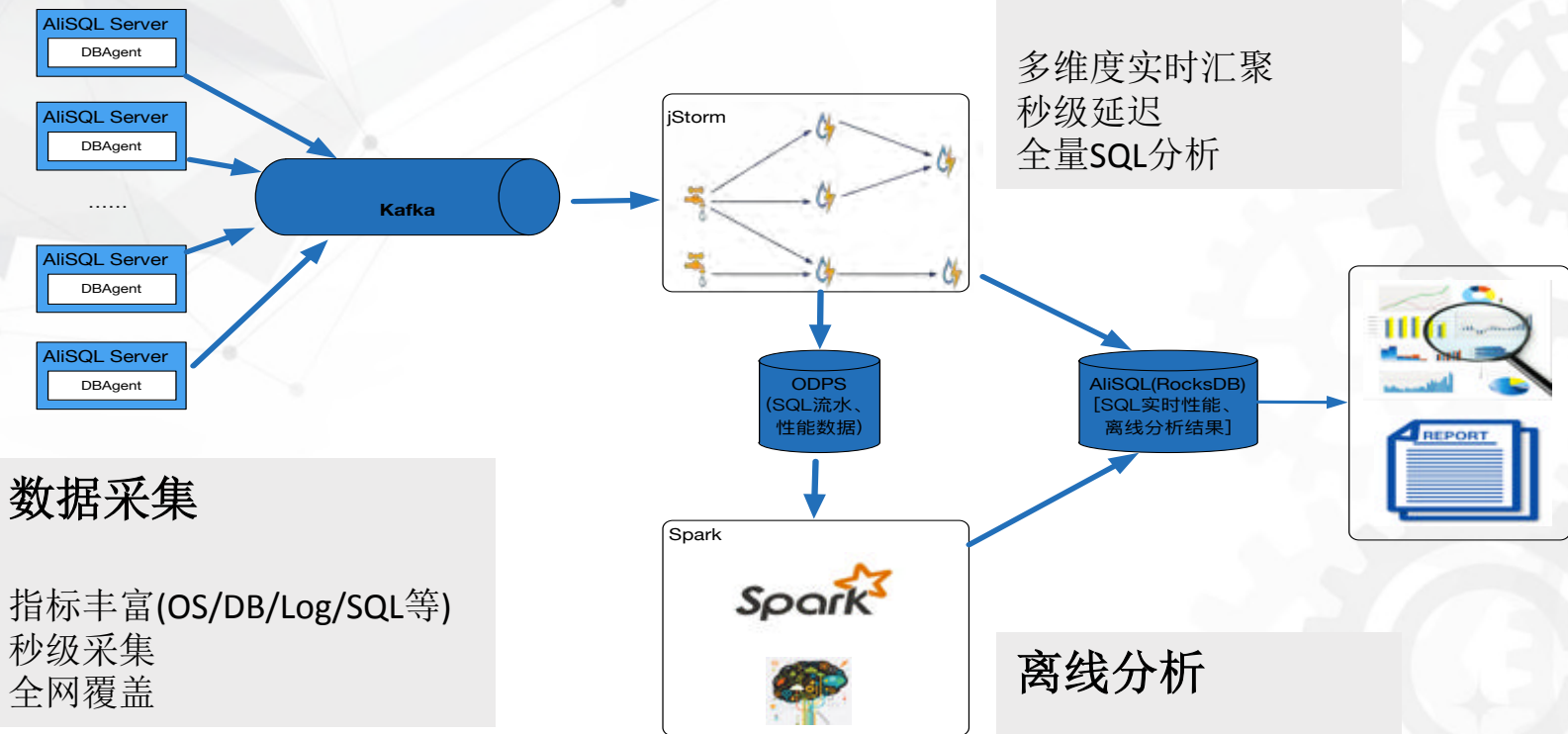
# 空间优化

- 详细空间信息(空间占比, 增长趋势,读写流量等)
- 表空间碎片回收
- 无用/重复索引清理
- 无流量表清理
- 容量预测/扩容评估
- 数据生命周期管理

# 数据化 - 数据通道

## 实时分析

多维度实时汇聚  
秒级延迟  
全量SQL分析



## 数据采集

指标丰富(OS/DB/Log/SQL等)  
秒级采集  
全网覆盖

## 离线分析

性能成本模型  
容量预测  
超大数据量

# 数据化 – 全量SQL采集计算

- 全网SQL流水实时采集
  - AliSQL内核实现轻量级全部SQL流水采集
  - 指标丰富
    - SQL文本/请求来源/用户信息/时延/逻辑读/物理读/锁等待时间等
- 秒级采集，秒级计算，秒级展示
- 日常SQL实时采集计算量达到千万级SQL/s
- CloudDBA诊断优化最重要的基础数据

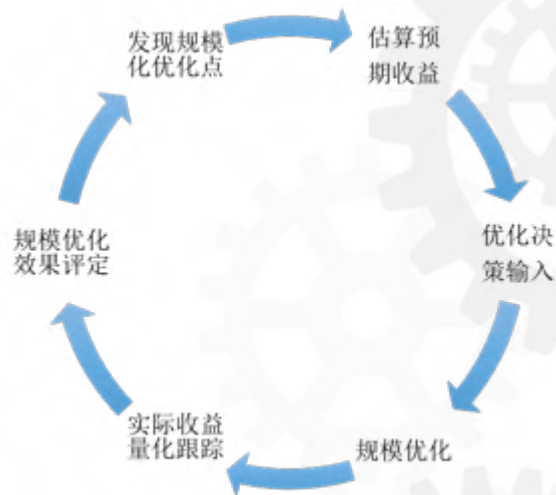
# 数据化 – 全量SQL价值

- 开发人员自助查看全部SQL性能数据(信息透明)
- 全局优化
  - 基于SQL Workload优化
  - SQL健康度
  - 关联分析
  - 量化跟踪
- 性能热点分析
- 业务模型分析
- SQL审计
- .....



# 数据化 – 性能成本度量系统

- 量化全网优化收益
- 建立全网性能成本模型
- 全网性能成本基线
- 发现规模优化点
- .....





# Agenda

Why

阿里巴巴数据库服务的诉求

What

数据库智能优化系统的思考

How

CloudDBA关键技术

Now

数据库智能优化系统的探索

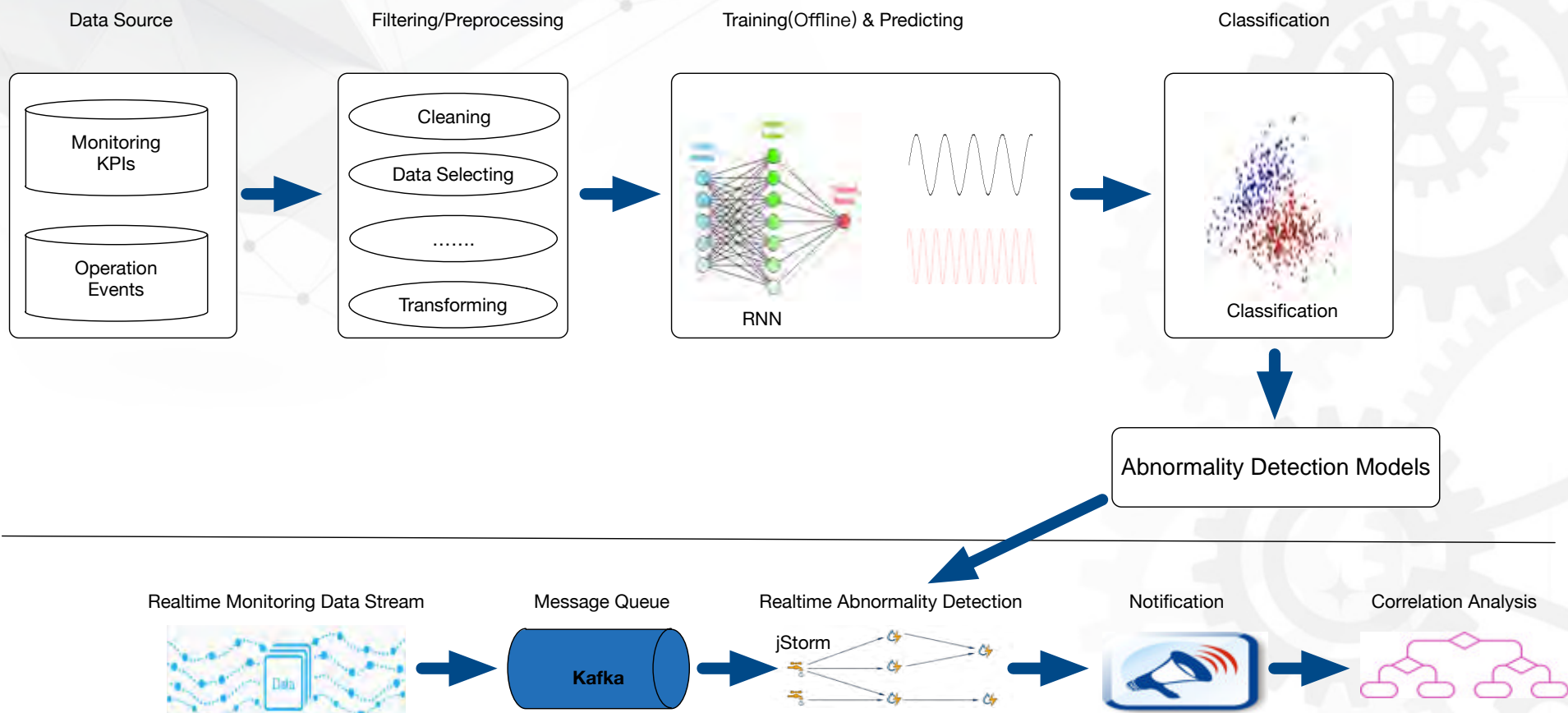
Q & A

Q & A

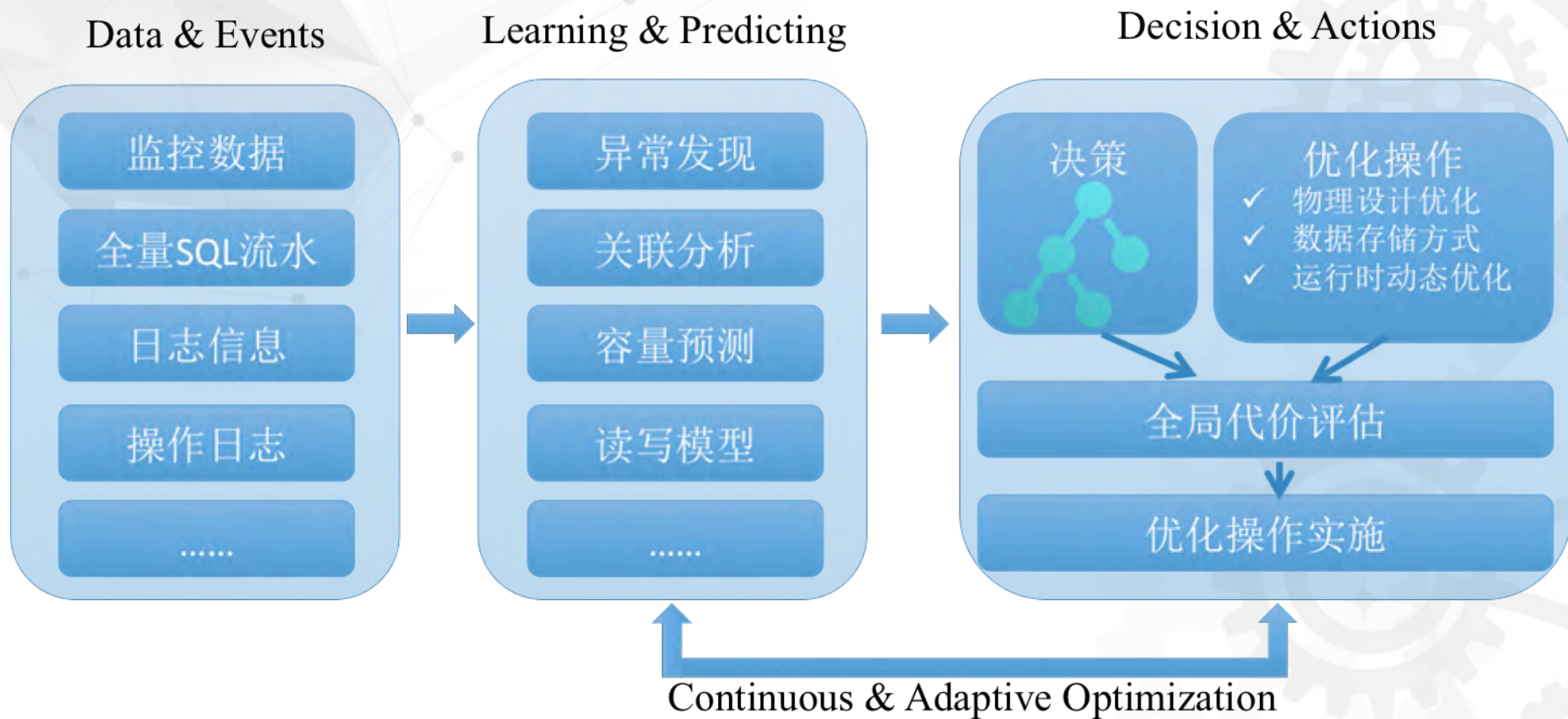
# 自诊断 – 被动到主动

- 实时准确发现异常
- 自动关联分析定位异常原因
- 持续自适应算法优化
- 事后发现到事前预警
- 大规模场景具有扩展性

# 自诊断 – 智能异常发现



# 自优化 – Learn, Predict, Evolve



# Summary

Why

阿里巴巴数据库服务的诉求

What

数据库智能优化系统的思考

How

CloudDBA关键技术

Now

数据库智能优化系统的探索

# Q & A

***We are hiring!!!***

**Java技术专家/高级专家**

- ✓ 数据库诊断优化
- ✓ Java后端开发
- ✓ 数据处理/数据分析(Storm/Spark等)

**算法专家/高级专家**

- ✓ 数据挖掘
- ✓ 机器学习

Mail: [kenny.qhl@alibaba-inc.com](mailto:kenny.qhl@alibaba-inc.com)  
[qiaohonglin@gmail.com](mailto:qiaohonglin@gmail.com)







# THANKS

SequeMedia  
威拓传媒

IT168.com

ITPUB

ChinaUnix