

大型监控系统设计与应用 实践

郑永宽 京东云 产品研发部总监



TOP100Summit

全球软件案例研究峰会

时间：11月15~17日

地点：北京国际会议中心

100个年度最值得学习案例

MPD工作坊（深圳站）

时间：9月21~22日

地点：深圳博林圣海伦酒店

20个3小时大时段沙盘课程

MPD

100

MPD

MPD

MPD工作坊（北京站）

时间：7月06~07日

地点：北京国家会议中心

20个3小时大时段沙盘课程

MPD工作坊（上海站）

时间：10月26~27日

地点：上海

20个3小时大时段沙盘课程

CONTENT

01 | 需求背景

02 | 京东云监控实践

03 | 监控系统设计

04 | 未来展望

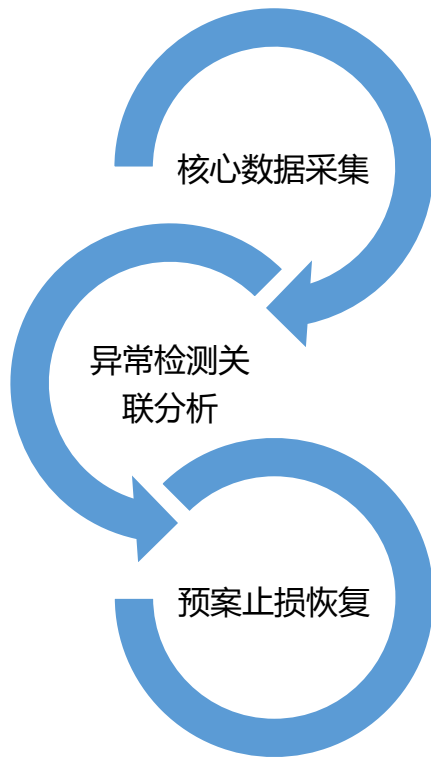
01

需求背景

需求背景

- **监控是运维的生命线**
- **缩短异常生命周期MTTR**
 - See->know->act
- **期望监控系统:**
 - 丰富的数据采集手段
 - 多维度数据实时聚合计算
 - 异常检测, 告警准确及时
 - 可定制的dashbord, 定位问题
 - 根因推荐定位, 辅助决策
 - 预案平台, 快速止损
 - 易用性、高可用、可扩展

99.99%



02 京东云监控实践

京东云监控体系---监控标准

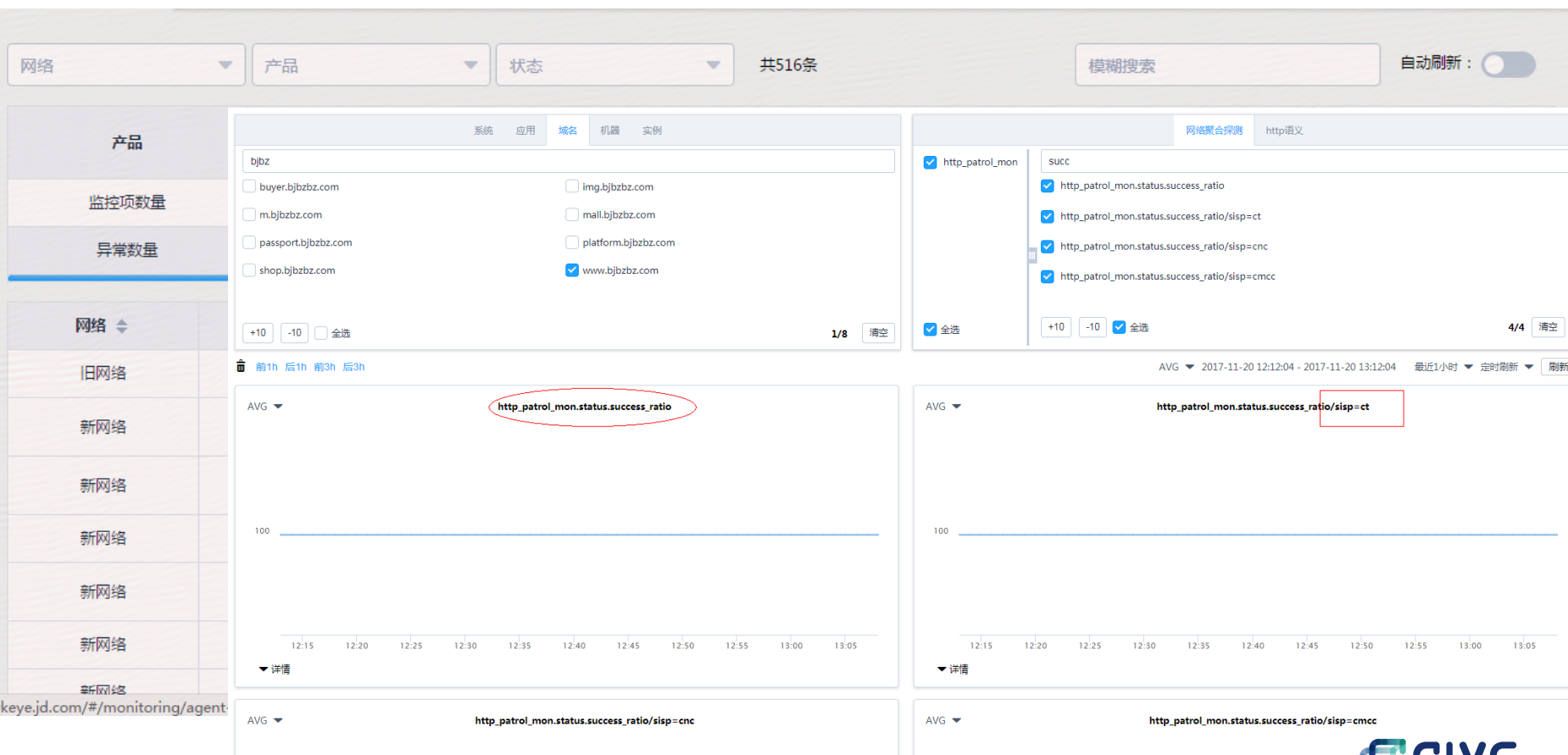


京东云监控体系---业务监控

用户侧使用情况 => (监控手段:自定义、外网域名)

业务监控

- 京东云官网的页面的访问状态；流程监控(模拟创建子网流程)
- 30+省市节点模拟用户访问；产生分运营商成功率



京东云监控体系---应用监控

应用监控

函数方法监控, JVM性能

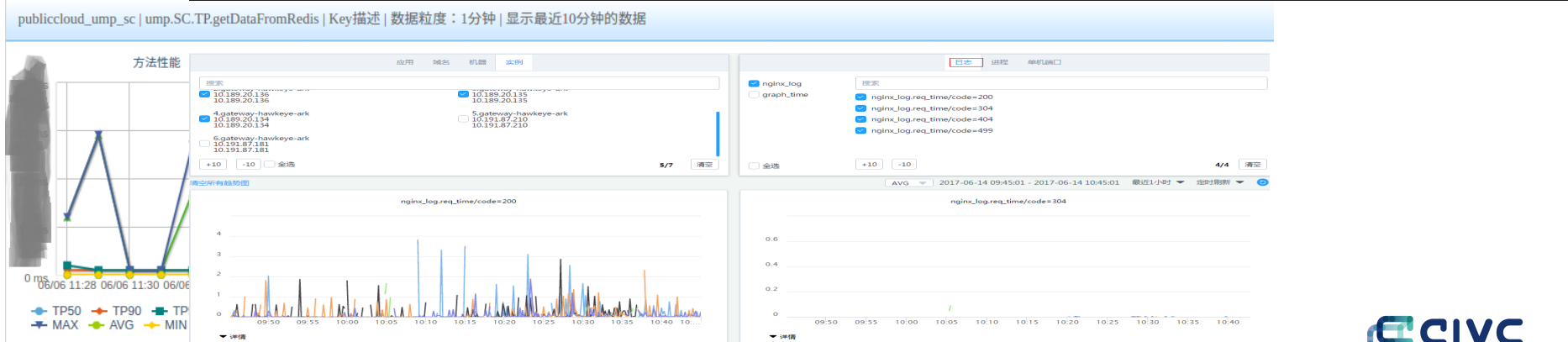
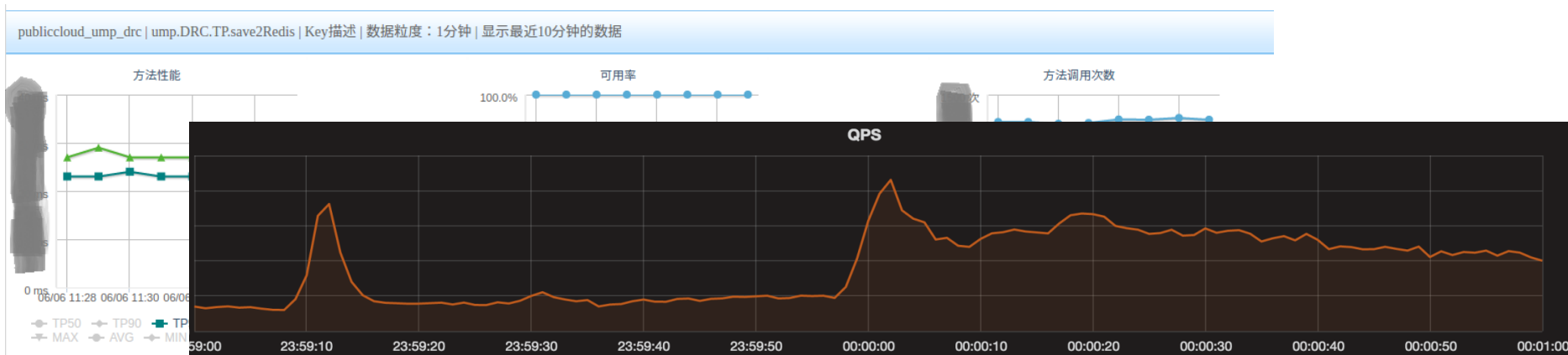
流量, QPS, 延迟, 错误率, 命中率...

常见开源软件监控

Lib库接入, 埋点

日志/自定义接入

组件监控

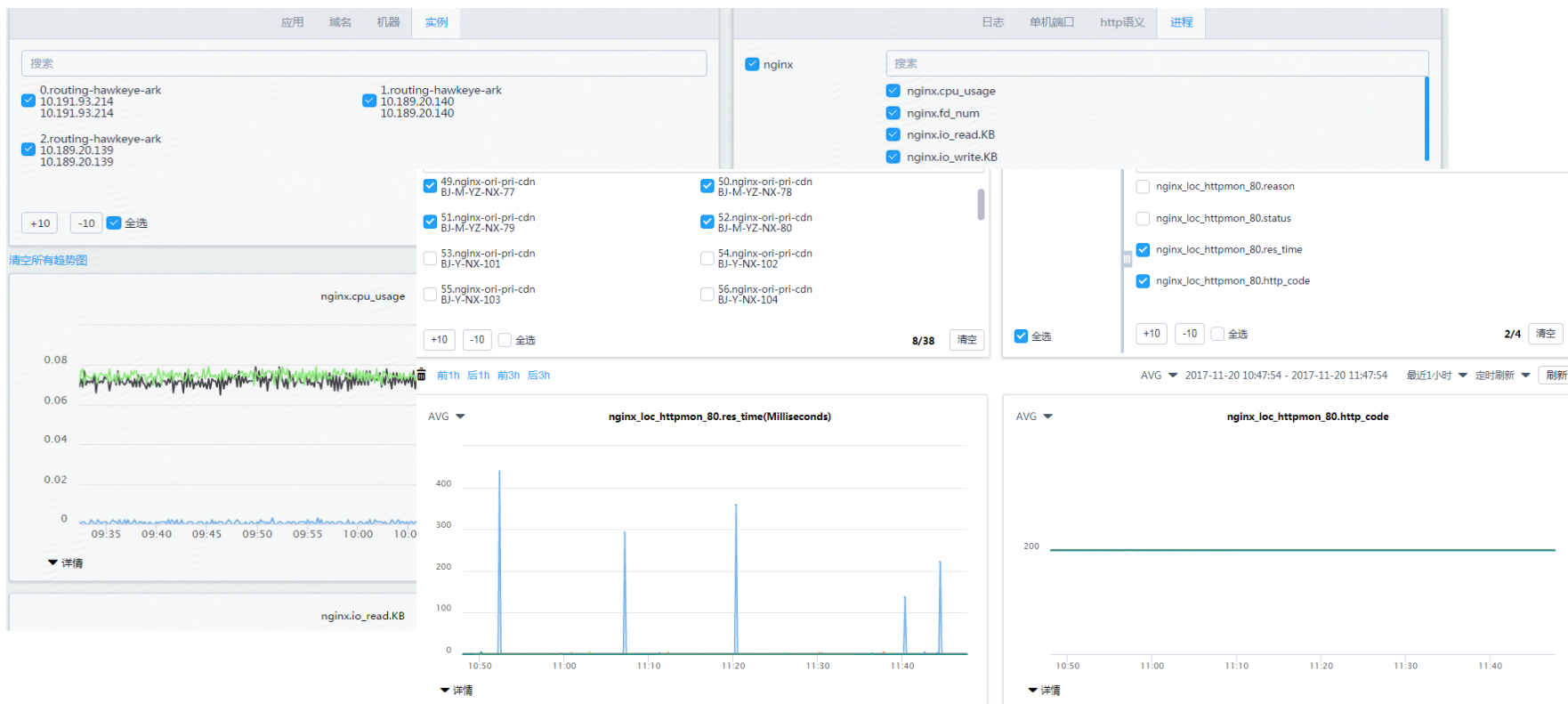


京东云监控体系---存活性监控

程序在机器上是否存活=>(监控手段:进程监控、端口监控)

存活性监控

- 进程存活状态、数目、占用资源
- 端口探活，模拟http/https/tcp/udp等协议通程序交互



京东云监控体系---基础设施

机器资源监控 => (监控手段: 机器监控、死机监控)

基础监控

- 200+ 监控项自动采集、支持物理机/容器/虚拟机...
- 死机支持ping、ssh探测，支持对假死判断

The screenshot displays the JD Cloud monitoring console interface for machine resource monitoring. It includes a sidebar with a tree view of monitoring objects, a central configuration panel for selecting monitoring items, and a dashboard with two line charts and a summary table.

Configuration Panel:

- Search: 选择具体监控对象的名字
- Selected items:
 - 172.19.6.151 ORI-CLOUD-HUABEI-JCS-151
 - 172.19.6.152 ORI-CLOUD-HUABEI-JCS-152
 - 172.19.6.153 ORI-CLOUD-HUABEI-JCS-153
 - 172.19.6.154 ORI-CLOUD-HUABEI-JCS-154
 - 172.19.6.155 ORI-CLOUD-HUABEI-JCS-155
 - 172.19.6.156 ORI-CLOUD-HUABEI-JCS-156
- Monitoring items selected: mem (checked), load, disk, system.
- Additional items: mem.cached.KB, mem.buffers.KB, mem.total.KB, mem.free.percent, mem.free.KB, cpu.sys/core=22.

Time Selection: 进行时间选择: AVG 2017-05-04 09:57:12 - 2017-05-04 10:57:12 最近1小时 不刷新

Charts:

- cpu.idle:** Line chart showing CPU idle percentage over time for various IP addresses.
- mem.free.KB:** Line chart showing free memory in KB over time for various IP addresses.

Summary Table:

监控对象	MAX	MIN	AVG	LATEST	SUM
172.19.6.156	1806428	1352752	1671636	1381676	200596320
172.19.6.153	1109880	857468	1043815.8	1062176	125257896

京东云监控标准实施 —— 监控打分与配置推荐

打分&推荐

- 践行监控标准
- 运维人员可以查漏补缺
- 管理者对整体稳定性有直观认识
- 推荐配置形成最佳实践



存活监控包括端口监控、进程监控和死机监控，以此监控您的机器和服务是否存活。

<input type="checkbox"/>	范围	节点名称	监控类型	采集配置名称	推荐信息	操作	使用场景
<input type="checkbox"/>	APP	log-test	端口监控	portmon	80	启用 忽略	提供端口存活和...
<input type="checkbox"/>	APP	test-deploy-pa...	进程监控	procmon	/export/Instan...	启用 忽略	提供进程状态监控

更新配置

- * 监控类型: 端口监控
- * 方法: 端口探测
- * 名称: portmon
- * 范围: 应用
- * 节点: log-test
- * 采集周期: 60s
- * 地址: 本机IP 80

生成推荐配置

云翼

优(88)

监控覆盖度: 64

报警: 100

报警处理能力: 100

京东云监控标准实施—告警处理

报警分级:不同级别对应不同处理方式

- P0, 立即处理, 监控系统要保证及时发送, P0报警对应应有预案, 预案需要定期演练
- P1, 可以延迟处理, 如果是固定的机械动作, 通过自动化平台进行自动处理; 每天进行定期例行dashbord检查处理
- P2, 一般用于根因定位里面的辅助决策

处理流程:

- 接受报警后, 通过报警历史页面:
 - 通过看图定位出现什么问题
 - 通过事件流图查看是否有上线影响
 - 通过查看采集/报警配置, 是否快速修改阈值
 - 通过ACK/恢复, 进行人工确认
- 每天定期进行巡检, 关注未恢复的报警
 - 处理类似报警优先级比较低的, 比如磁盘<20%,避免升级

监控平台能力:

- 告警方式多样:电话、短信、邮件、微信、咚咚
- 预案平台: 固话机械性动作
- 报警统计: 协助管理人员推进, 消除隐患
- Dashbord: 定期巡检
- 干预手段丰富: ACK、暂停等
- 报警合并: 减少对人的打扰

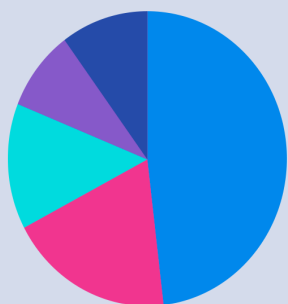


报警历史记录表格显示了报警的详细信息，包括主机名、IP地址、报警名称、当前状态、报警级别、监控对象、报警时间、恢复时间、持续时间以及操作选项。表格中有一行报警状态为“未恢复”，其报警名称为“remote_pingmon”，报警级别为“P2”，监控对象为“atus”，报警时间为“2017-11-20 14:25:00”，持续时间为“5分42秒”。

主机名	IP地址	报警名称	当前状态	报警级别	监控对象	报警时间	恢复时间	持续时间	操作
HOST	172.19.54.138	remote_pingmon	已恢复	P2	atus	2017-11-20 14:25:00	--	5分42秒	恢复 ACK
HOST	172.19.53.53(172.19.53.53)	remote_pingmon	未恢复	P2	atus	2017-11-20 14:22:00	--	8分42秒	恢复 ACK
HOST	172.19.57.42(A04-R08-07-42-6001566.XCLOUD.COM)	remote_pingmon	未恢复	P2	atus	2017-11-20 14:17:00	--	19分42秒	恢复 ACK
HOST	172.19.53.50(172.19.53.50)	remote_pingmon	未恢复	P2	atus	2017-11-20 14:16:00	--	14分42秒	恢复 ACK
HOST	172.19.54.99(172.19.54.99)	remote_pingmon	未恢复	P2	atus	2017-11-20 14:14:00	2017-11-20 14:28:00	14分	恢复 ACK
HOST	172.19.54.104(172.19.54.104)	remote_pingmon	未恢复	P2	atus	2017-11-20 14:12:00	--	18分42秒	恢复 ACK
INSTANCE	104jcloud-docker-lbaas-jcloud-aaa117.2.18.15.163	cc_controller_gw_data_error	未恢复	P1	log_cc_controller/c_c_controller_gw_data_error	2017-11-20 14:08:00	2017-11-20 14:09:00	1分	恢复 ACK

京东云监控标准实施——故障定位

故障原因



- 变更
- 程序异常
- 网络&资源
- 中间件
- 其他

- 定位边界VS定位根因 → **止损优先**
- 京东云快速迭代升级 → **变更可视化**

上线操作
配置变更
初始化任务
平台全局事件
...

请输入关键字进行过滤

- 工具产品研发部
 - Ark-内部版
 - Ark-测试节点
 - 交接机器给大数据的deepl...
 - 云翼自己的测试节点
 - 用户使用的华东测试二期...
 - 井亮亮交接的测试机器
 - 运维组不在我们这的机器
 - 用户使用的华东预发二期

报警统计 报警列表 变更事件 2019-04-10 15:06:54 - 2019-04-10 16:06:54 最近1小时 刷新

搜索操作对象, 操作内容, 操作人

事件类型	操作范围	操作对象	操作内容	开始时间	结束时间	操作人	
部署事件	应用	git-test	上线: 部署成功	2019-04-11 12:00:09	2019-04-11 12:00:09	wuxuelian7	>
当前版本: ci-test-e68d91cd-0411150259 环境: 预发布				分组: lf-pub,mjq-pub,bjyz-pub 部署方式: 包部署			
服务树变更	分组	hb	修改分组 tower-api. hb: env: ...	2019-04-11 12:00:09	2019-04-11 12:00:09	liuhaoran	>
第三方事件	应用	app1	重启实例0.app1	2019-04-11 12:00:09	2019-04-11 12:00:09	liuhaoran	>

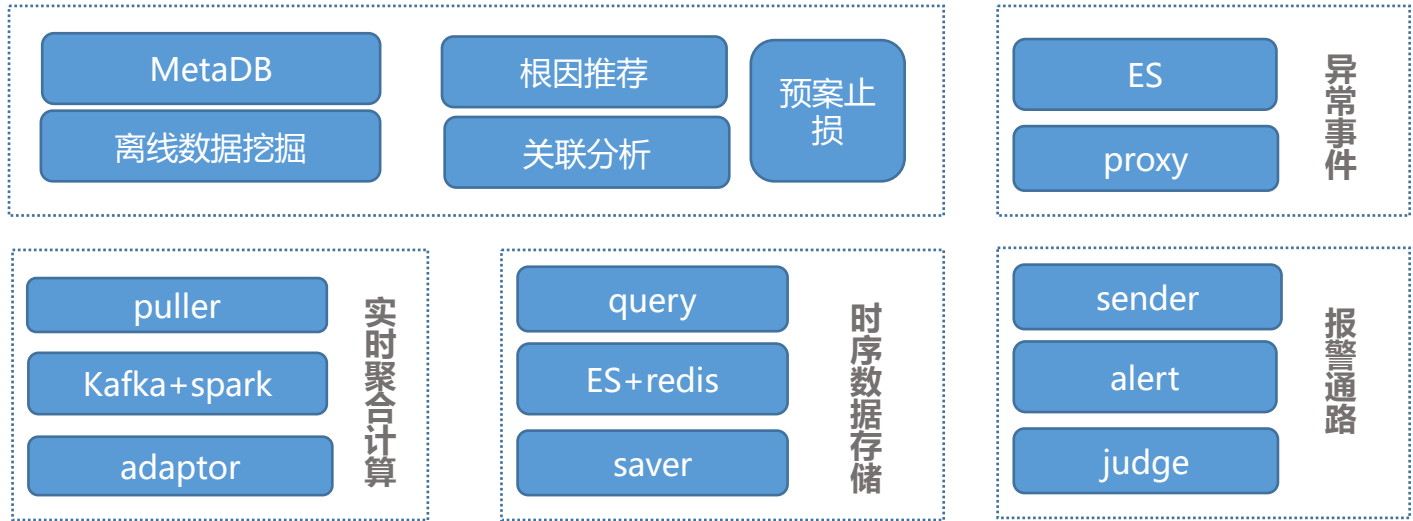
03 监控系统设计

典型监控系统架构图

数据展示



数据处理



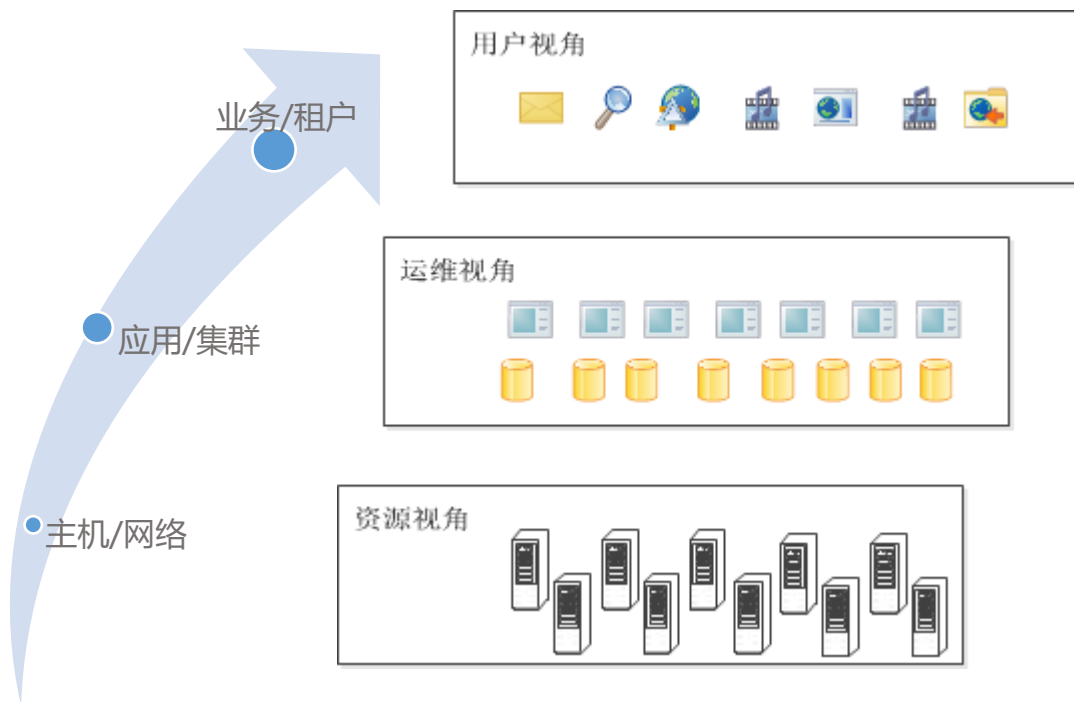
数据采集



数据抽象



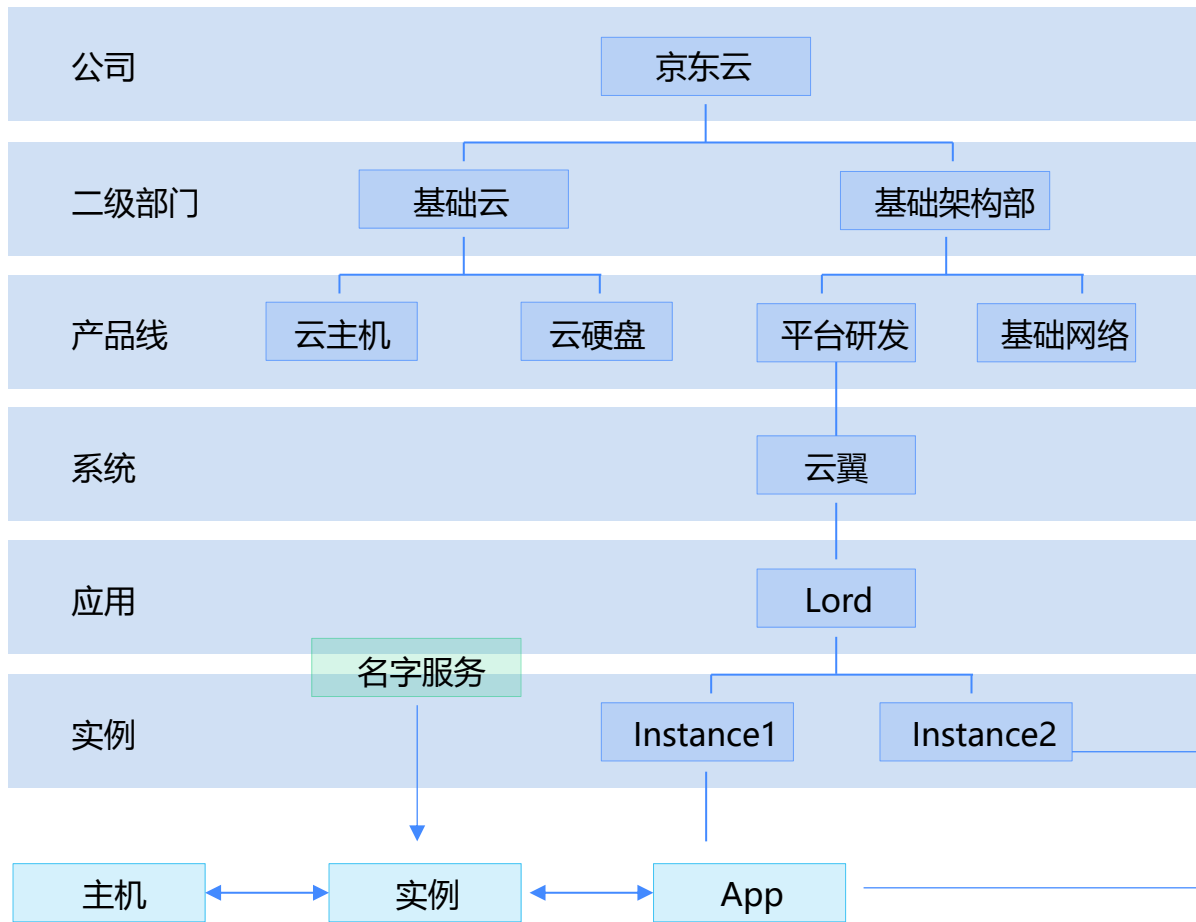
统一运维世界认知-CMDB



- **面向业务的资源关联管理**
 - 业务->应用->集群->主机(网络)
 - 提供统一入口管理
- **基于CMDB的名字服务**
 - 提供资源快速正查反查服务
- **基于CMDB监控配置服务**
 - 提供业务/应用/集群的配置

CMDB——服务与资源管理

服务树与名字服务示意图



服务树

- 业务组织架构信息
- 应用从属关系
- 角色管理与基于角色的权限控制
- 运维基础meta数据

资源管理

- 全流程机器管理
- 机器资源池的管理
- 机器所属信息展示搜索



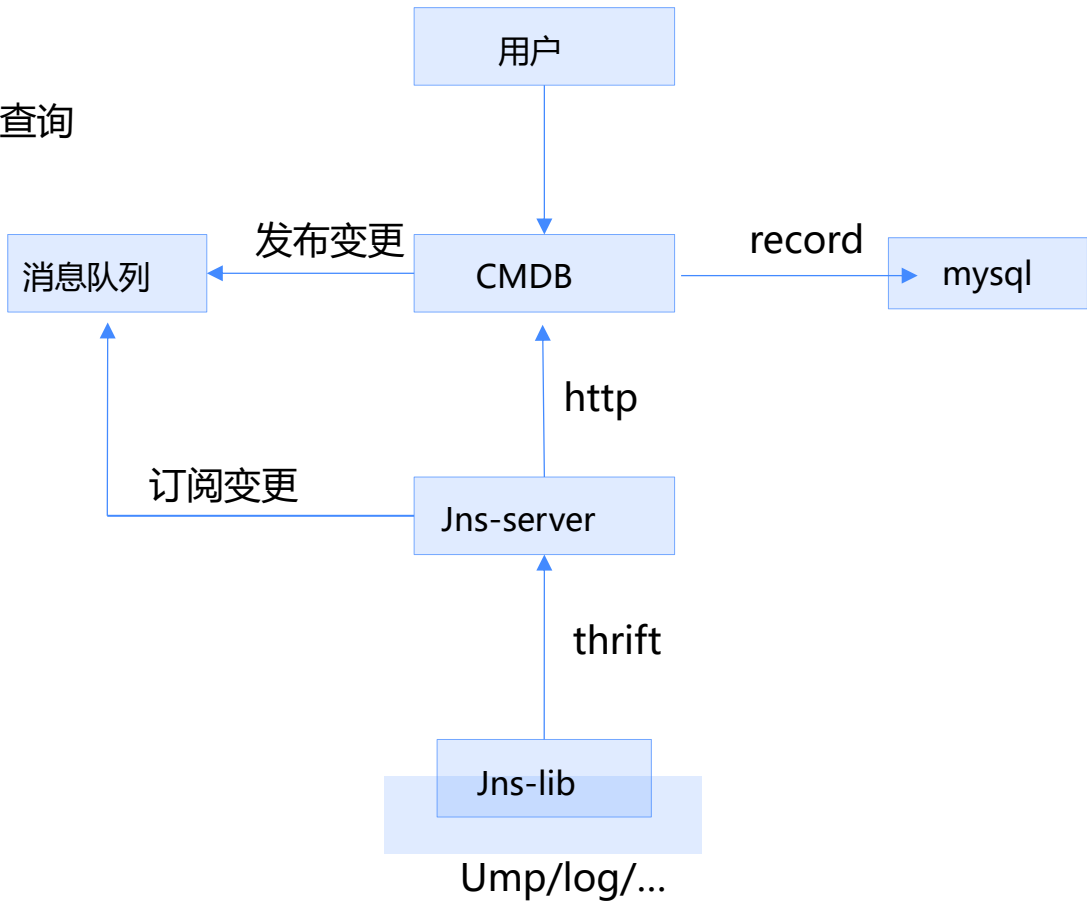
CMDB——名字服务 JD Naming Service

JNS功能

- 全量名字信息同步到lib,提供正向反向的查询
- 查询的数据缓存内存,提升查询效率
- 能够快速增量更新变更信息
- 服务解耦合

维护实例-App-主机之间的对应关系

- | | |
|----|-----|
| 部门 | 产品线 |
| 系统 | 应用 |
| 分组 | 实例 |
| 机器 | |



JNS整体架构图

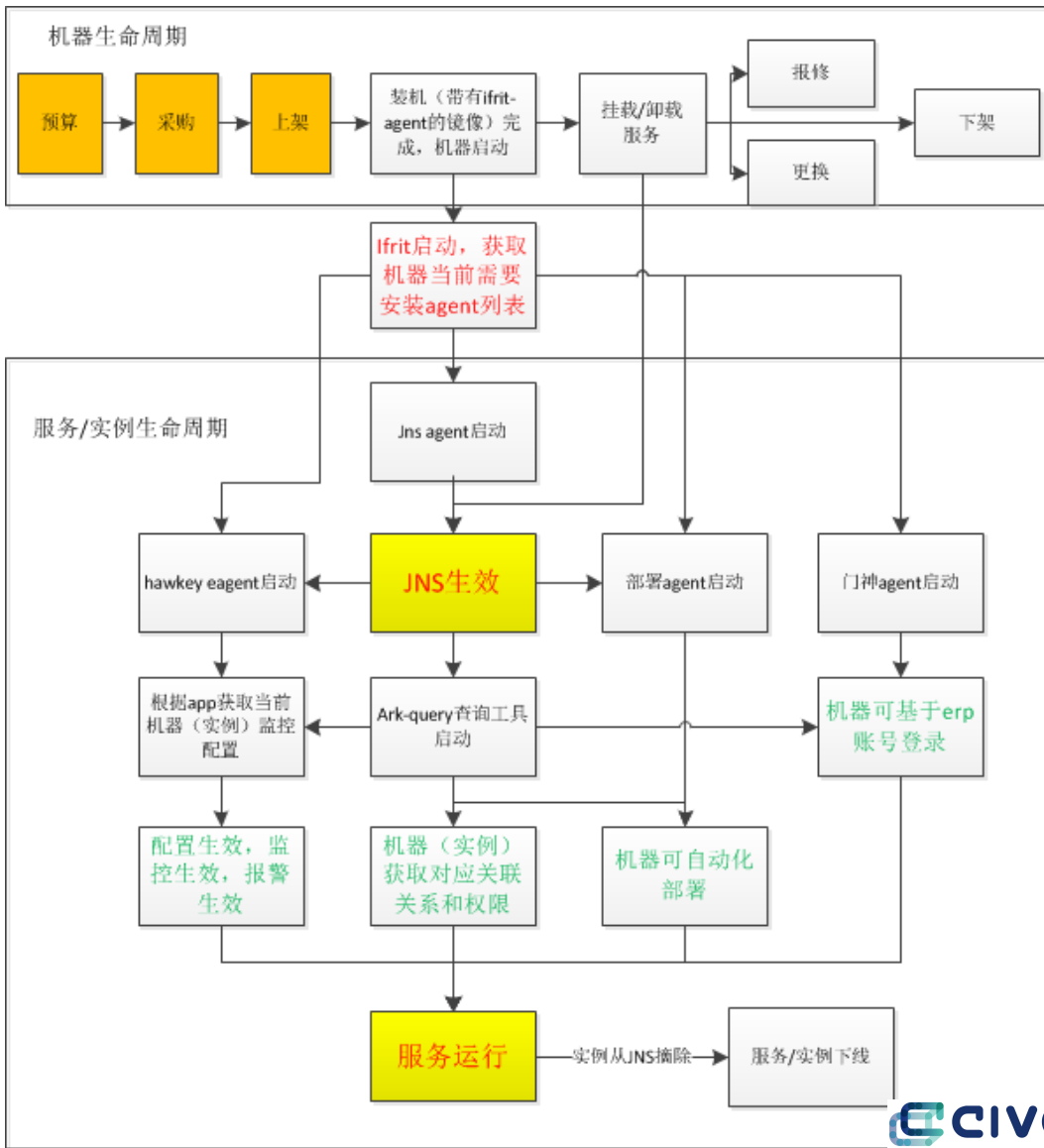
CMDB——机器与服务生命周期

生命周期管理

机器生命周期与服务生命周期解耦

自动化

高效与稳定



数据采集:标准化过程

采集是**数据标准化过程**，形式A->形式B

- 标准化采集的重要性

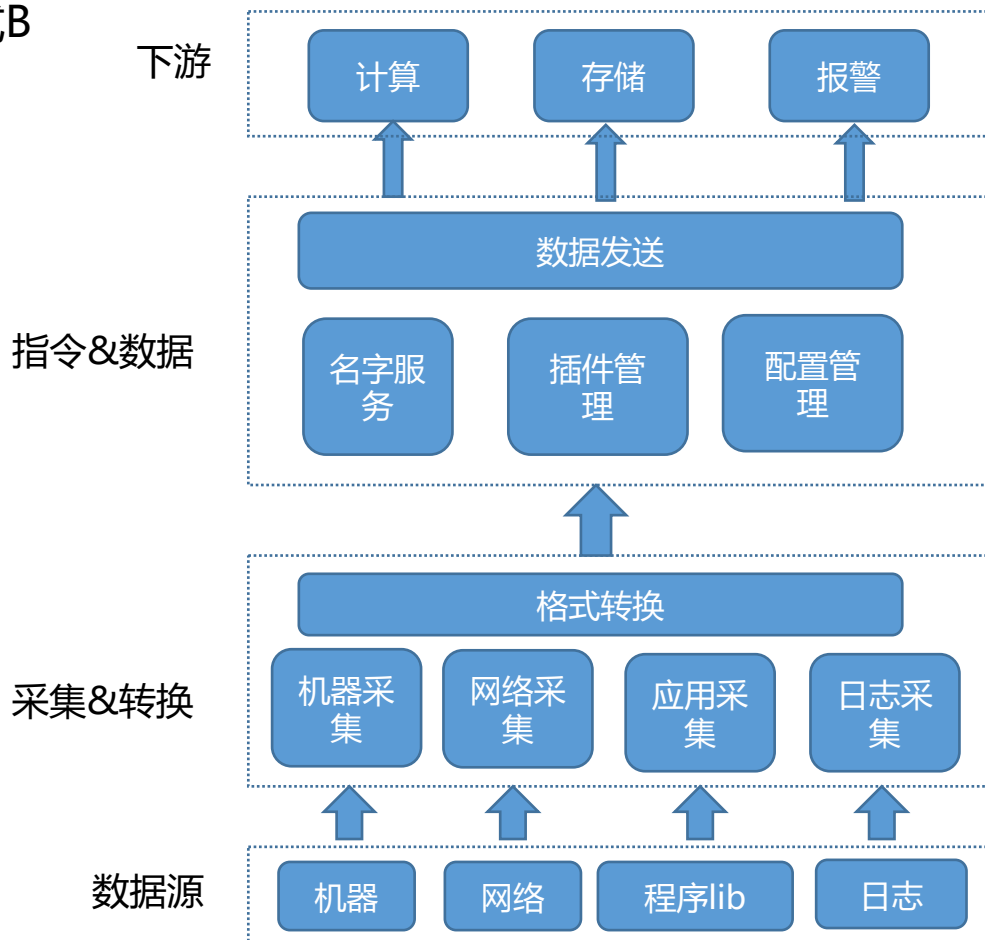
设计要点:

- 基于名字服务的配置管理
- 插件式管理,便于扩展
- 下游三路发送, 互不影响

- ✓ mem.free.KB
- ✓ mem.free.percent
- ✓ mem.usable.percent
- ✓ mem.buffers.KB
- ✓ mem.cached.KB
- ✓ mem.total.KB



```
[root@JD ~]# cat /proc/meminfo
MemTotal:      65920768 kB
MemFree:       54066168 kB
Buffers:       332348 kB
Cached:        10078472 kB
```



时序数据存储

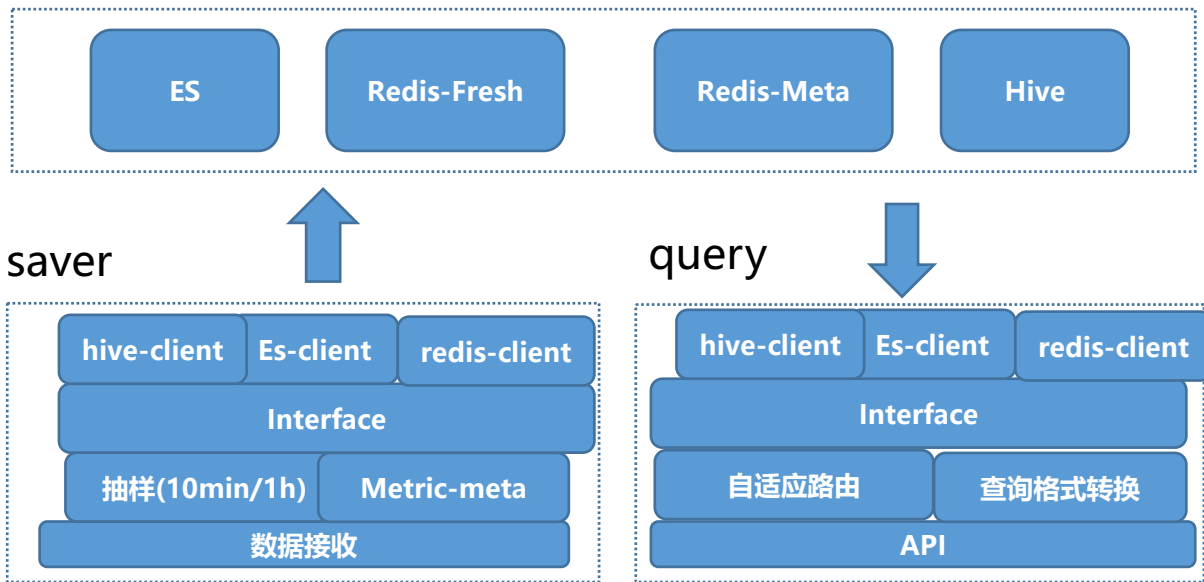
需求

- 写多读少，需要根据各种维度进行检索
- 最近一小时数据读取频繁，有数据热点
- 各种时间段的读取需求，一年数据秒出
- 数据用于离线分析
- ES/redis故障能快速恢复

设计

- 结合查询+写入，选择ES为主存
- 选择redis作为最新值/热点存储
- 写入抽样，查询自适应路由
- 抽象接口，便于添加各种下游
- Saver/query可以进行机房调度

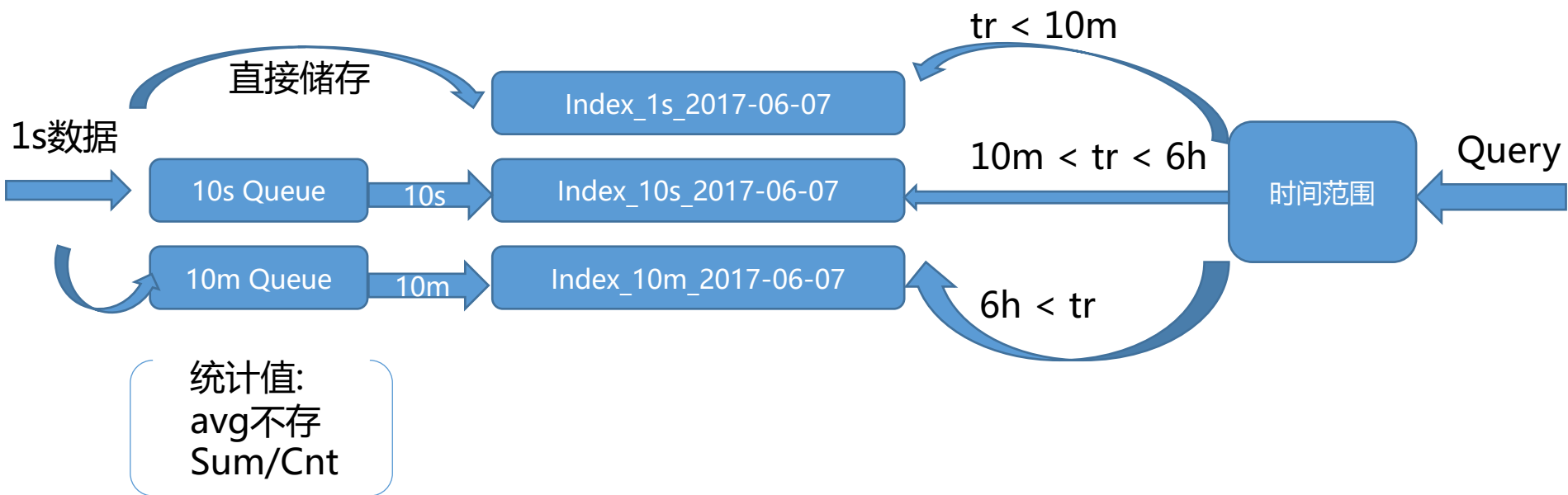
分机房部署



时序数据存储—抽样&自适应路由

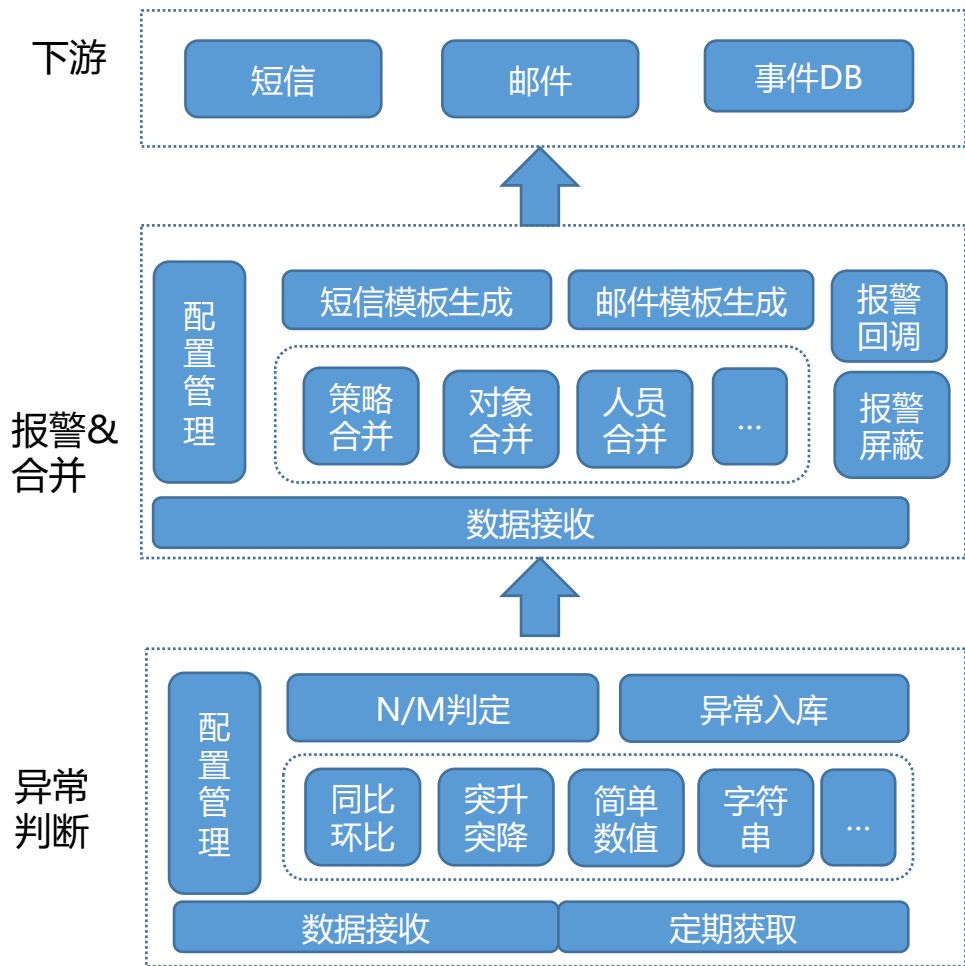
一年数据秒出：

- 存储实时计算抽样，写入不同ES索引index
- 根据查询的时间跨度，自动选择存储的时间粒度



报警通路

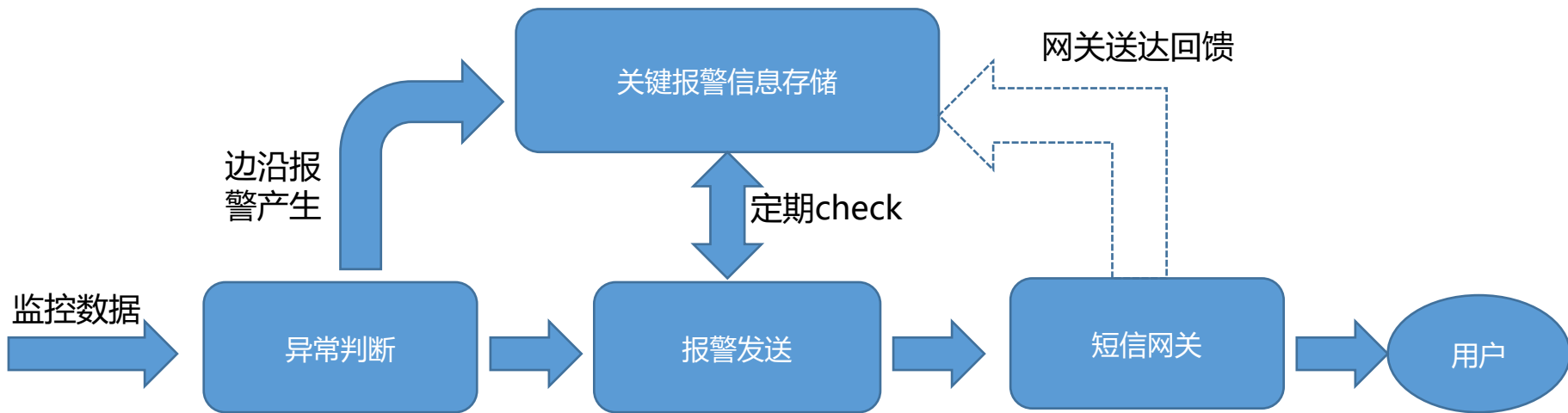
- 准确判定数据异常
- 报警及时发送，能屏蔽干预，避免报警风暴



- **丰富异常检测算法:**
 - 支持同环比/突升突降
 - 支持数值/字符串报警
- **丰富报警合并策略:**
 - 按人员/策略/对象合并
 - 报警分级&报警方式

报警通路—边沿报警不丢失

- **边沿报警**：正常→异常，异常→正常
- 模块重启导致消息丢失，收不到恢复报警



| 总结

- **京东云体系监控涵盖**
 - 基础设施/应用/服务
- **典型监控体系设计**
 - 数据抽象(CMDB先行)
 - 数据采集(标准化过程+资源控制)
 - 聚合计算(圈定范围/算子+重复数据)
 - 时序存储(读写正交/抽样+自适应)
 - 报警通路(异常判断/报警发送+不丢报警+场景算法)

04

未来展望

未来展望

- **问题发现**

- 采集标准化、无配置化
- 报警阈值离线分析,自动设置
- 报警事件自动升级

- **问题定位**

- 异常同原始日志关联
- 事件关联推荐算法

- **问题解决**

- 预案平台(预案的应用商店)
- 分场景进行自动处理

-



Q&A



关注京东云开发者社区，获取分享PPT