ZABBIX 2020 Conference CHINA

演讲主题

云原生体系下的监控能力演进

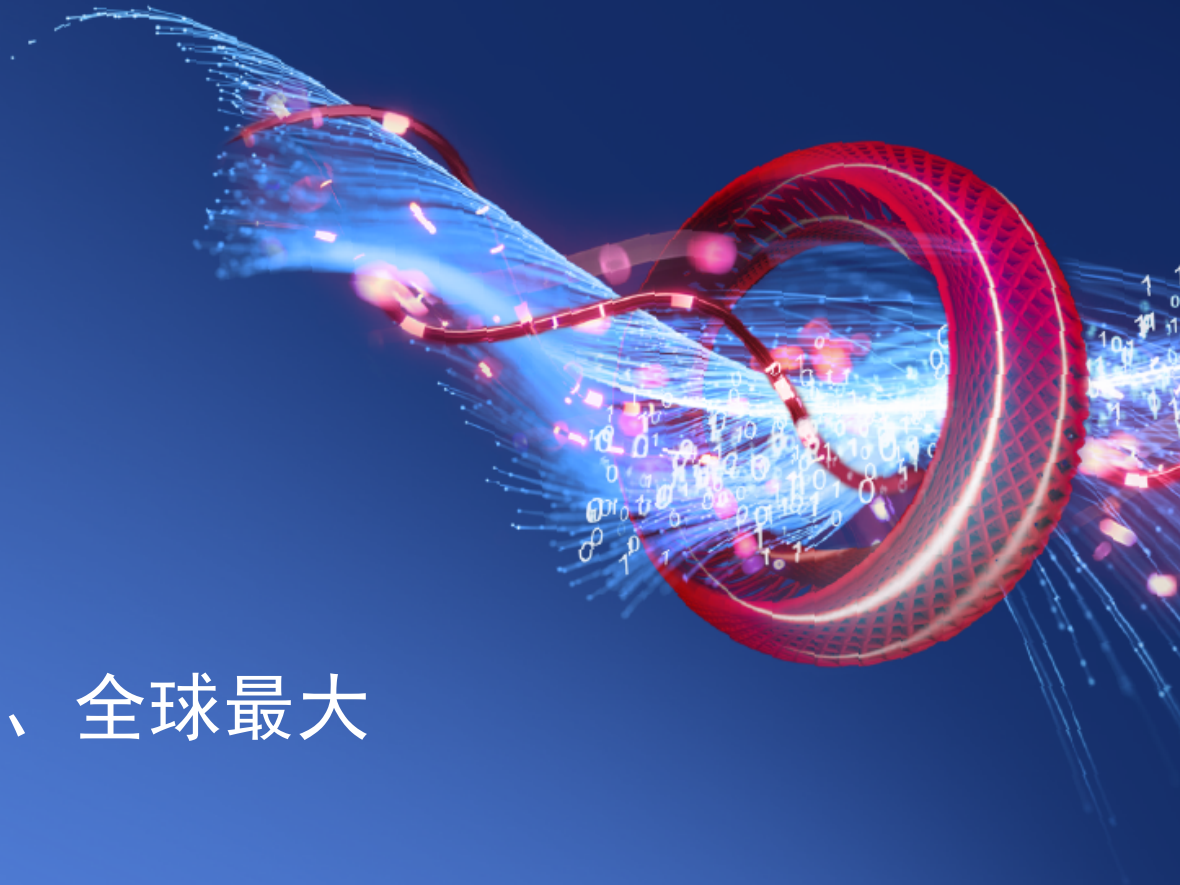
演讲嘉宾

王漫雪 技术经理 中移在线服务有限公司

目 录

C O N T E N N T S

ZABBIX 2020 Conference CHINA



01

背景

——全国集中维护、全球最大

- 中移在线公司简介
- 业务需求升级
- 积极应对挑战

ZABBIX 2020 Conference CHINA

中移在线公司简介

服务宗旨

移动全网渠道运营
集中支撑者

移动全网集中服务
提供者

移动全网业务
后台集中处理者

发展历程

2014

- 10月注册成立
- 全集团集中化、专业化运营试验田

2016

- 31省呼叫业务完成划转
- 奠定全网集中化运营基础

2017

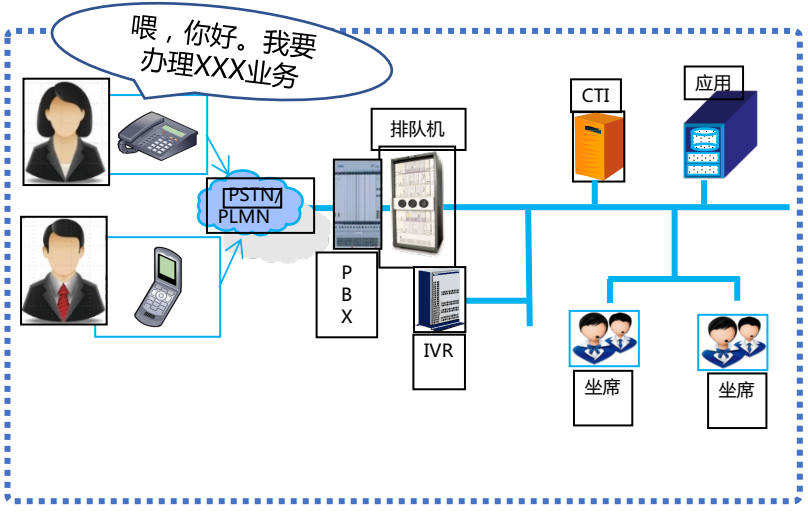
- 成为全国客服行业规模最大、实力最强的领军式企业。

2018

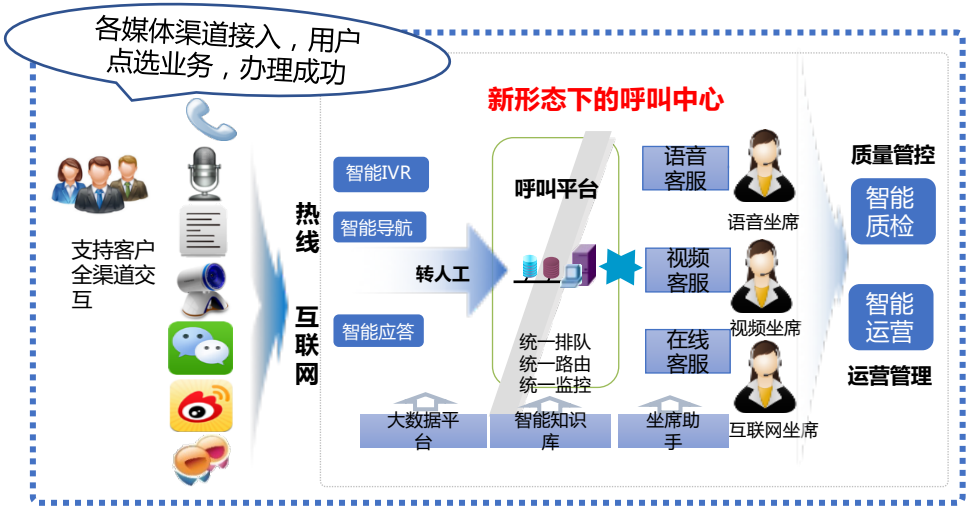
- 全集团首批入选国资委国企改革“双百行动”三家公司之一

业务需求升级

你觉得服务场景只有这些



实际服务场景



为满足多媒体渠道用户接入，公司进行了**服务、业务、运维**升级。

- **服务升级**：支持传统语音、文本、图片、视频、短语音、微信、微博等**多渠道、多数据**内容接入。
- **业务升级**：由软硬一体设备升级为**纯软件系统**，实现全媒体CTI、IVR、互联网接入网关、软交换、中继网关、媒体加速、用户终端**一体化运营**；将**人工智能（AI）、大数据技术**应用于IVR、机器人应答、质检、外呼等核心业务。
- **运维升级**：本部及31省分公司设备及业务系统，**集中化**配置、部署、上线、发布、迁移、维护、监控、告警、资源管控、故障处理。

积极应对挑战

面临的运维挑战

多

用户多,IT规模接近一线互联网企业

- **9亿**用户, **超1亿**微信粉丝，月服务超亿次，微博矩阵粉丝3038万（居行业首位），10086APP超五千万用户量
- 20000+服务器，**30000+Pod**

难

业务变化快，运维环境复杂

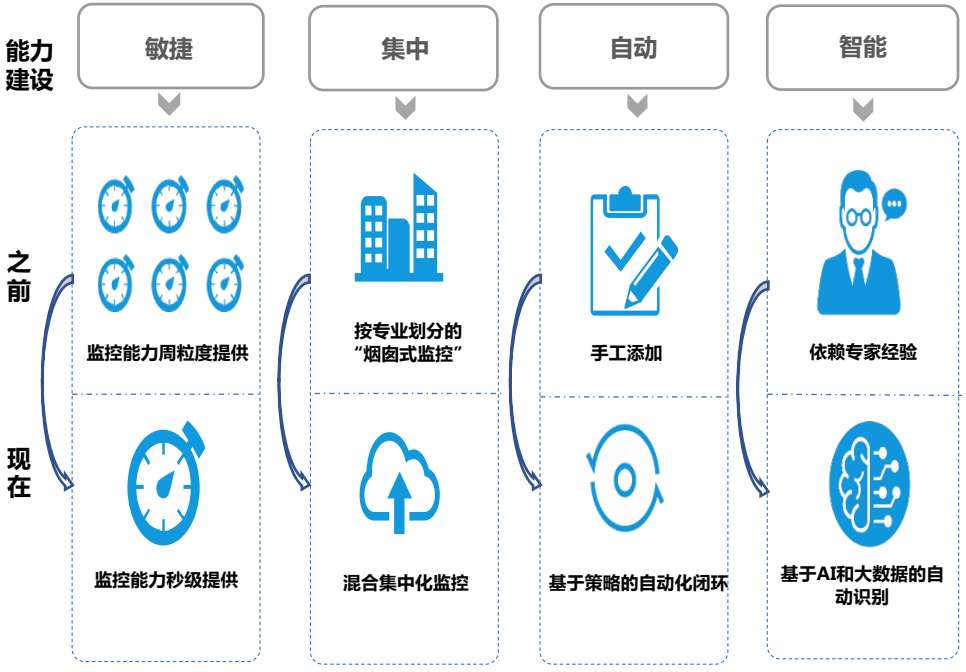
- 支撑全国营销活动，总部/分公司/省公司多级协同
- 日均上线 **120** 次，日处理 500 例工单
- 技术新：微服务/云计算/容器 ...

高

要求高，提供电信级服务

- 99.999% 的可靠性
- 15秒 接通要求
- 7*24 小时保障

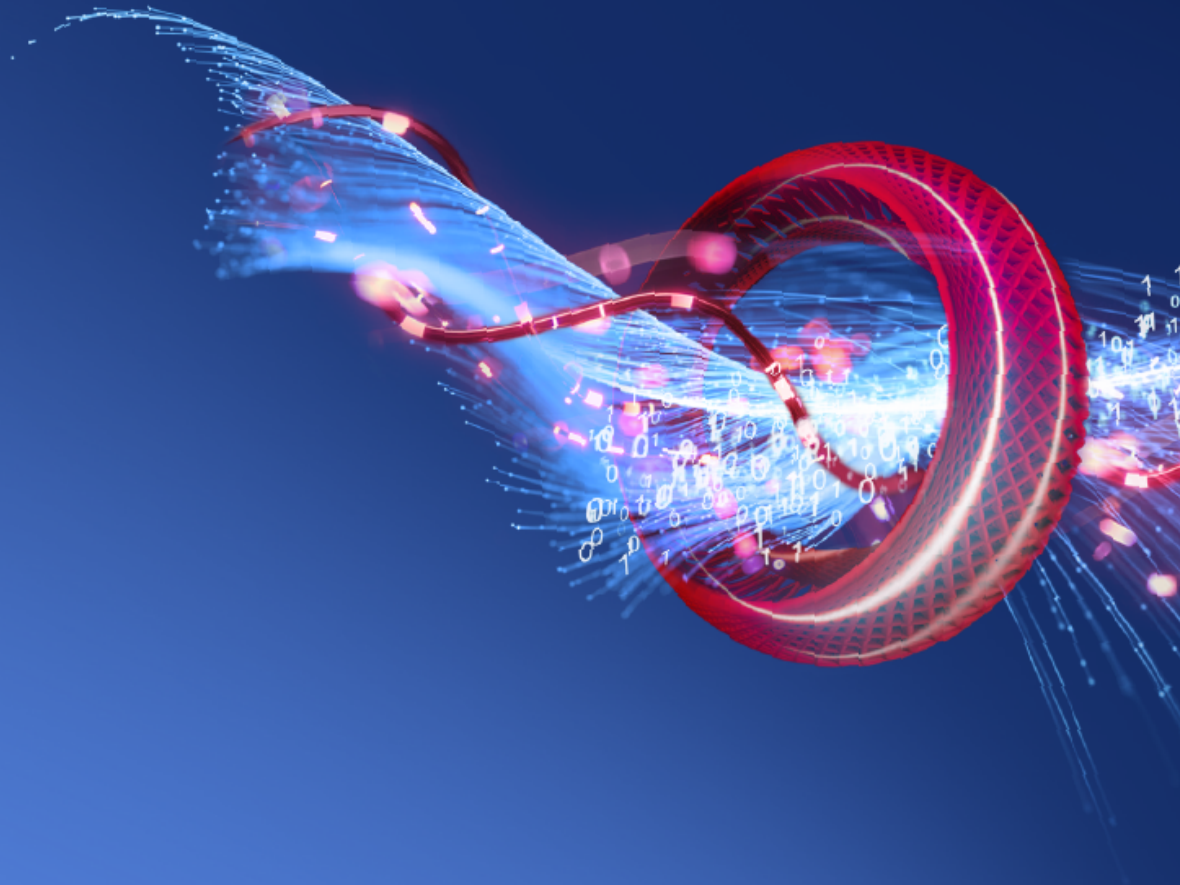
积极应对



02

出路
——选择开源

- 拥抱开源
- 统一监控平台
- 小结



ZABBIX 2020 Conference CHINA

拥抱开源

拥抱开源，站在巨人肩膀上

快速设施: 利用Zabbix的成熟能力，**1个月**快速完成监控系统的能力建设

覆盖范围全: Zabbix自带的官方模板以及社区的各种模块，可以快速实现多操作系统、各主流中间件的监控覆盖

实时稳定: Zabbix非常成熟，可实现秒级数据采集，线上3年基本无故障发生

可视化看板: 可与Grafana很好的结合，快速实现丰富可视化看板的制作

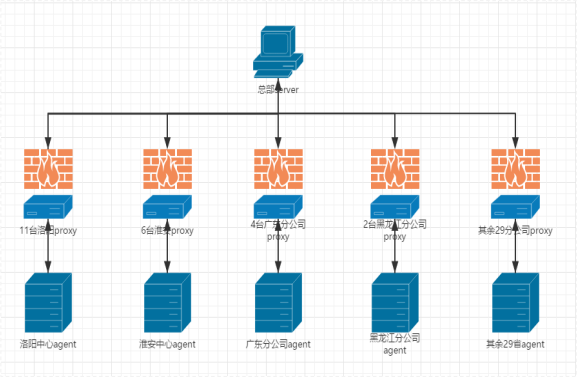
高效低成本：Zabbix本身资源消耗极低，主要是数据库需要物理硬件支撑，比Prometheus占用资源少

统一监控平台：集中建设、统一管控、边缘节点标准化

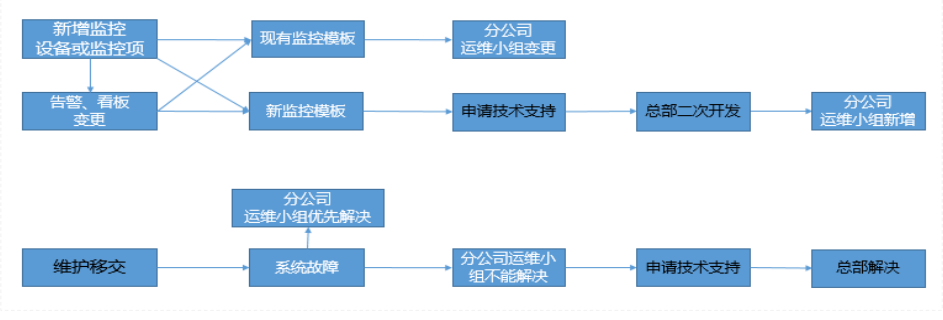
为了更快速的建立监控能力、更全面的管控系统质量，在线服务公司统一监控平台采用了总部集中建设、统一管控，分公司标准化接入的建设模式。

全网集中：

- *总部负责监控能力建设、边缘节点的标准化，所有监控数据的上收、分析、展现与通知。*
- *分公司提供资源，遵照标准化、封装后的监控模板进行监控资源的维护与管理。*

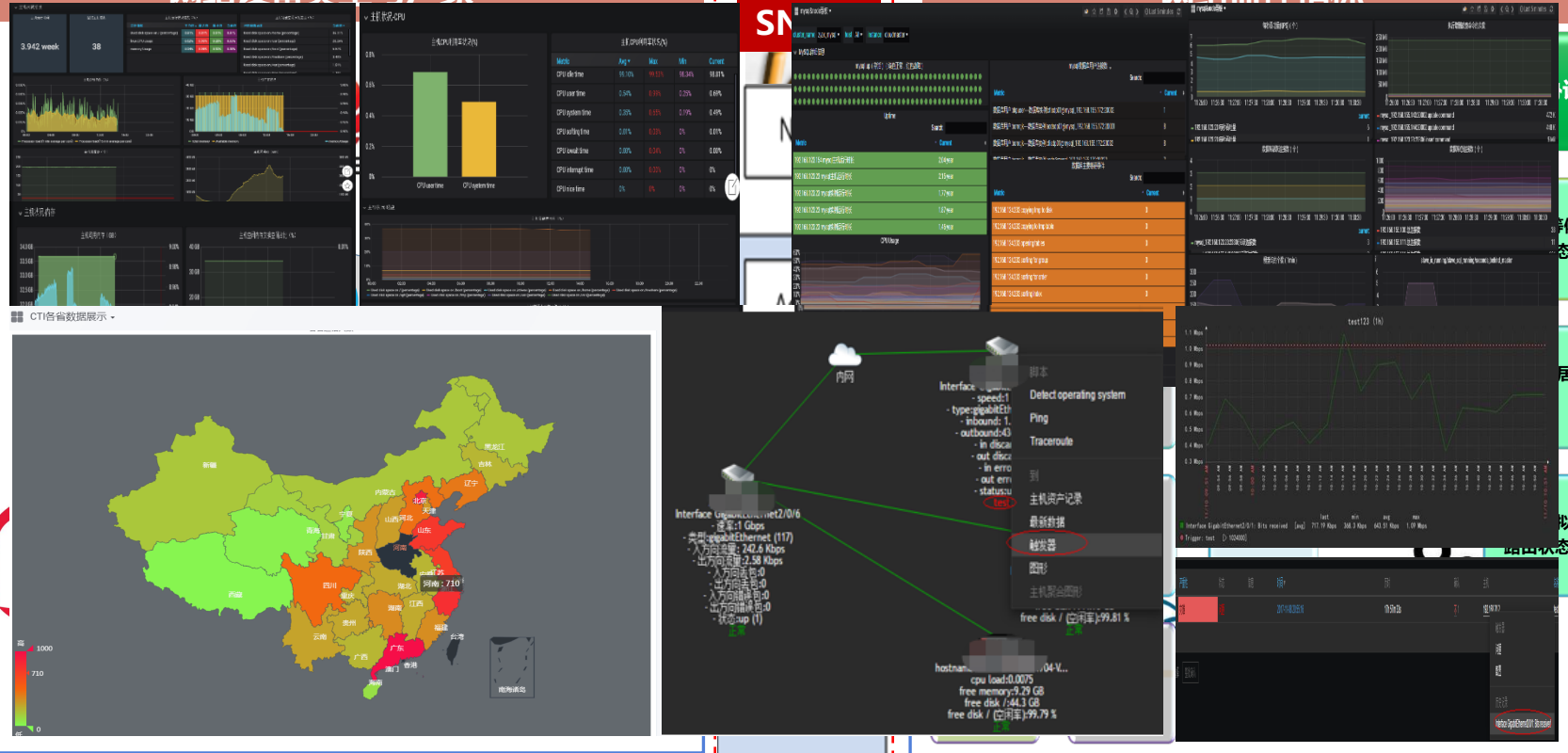


A	B	C
监控对象	监控种类	模板个数
网络设备	华为、思科、锐捷、华三	12
主机	suse、centos、redhat、windows、ubuntu、arm	5
数据库	mysql、oracle	3
中间件	tomcat、redis、elasticsearch、rocketmq、kafka、nginx	6
进程	cpu、内存、存活、端口	3
日志	关键字、日志不刷新	2
拨测	拨测状态	2
硬件设备	华为	1
自定义指标	命令类、sql类、接口类、脚本类	3
总计		37



一些小总结：广泛、丰富、多样、灵活

看板可灵活制定，分钟级完成配置。图表多样化展现：折线图、柱状图、饼图、区域图、拓扑图等。

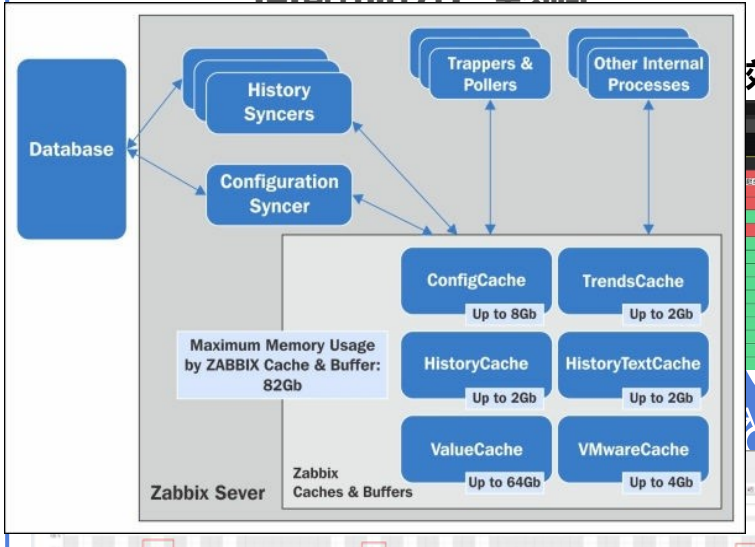


一些小总结：半年时间

2 万 主机	200 万 监控项	90 万 触发器	30 万 报警
84 Proxy	400+ DashBoard	545 用户数	1.3 K 动作

一些小总结：zabbix系统优化

- **Zabbix配置的同步机制**



- **Zabbix的配置表比较多，大容量局点关联查询sql耗时很长**

如数据库控制sql执行时间的max_execution_time配置不合理，会导致无法将相应配置表数据同步到zabbix server以及proxy的cache，从而导致出现大量监控项无法正常采集及消息队列积压现象。以下为zabbix_server.log相应日志：

```
30550:20181119:185248.301 slow query: 8.970014 sec, "select i.itemid,i.hostid,i.status,i.type,i.value_type,i.key_,i.snmp_community,i.snmp_oid,i.port,i.
snmpv3_securityname,i.snmpv3_securitylevel,i.snmpv3_authpassphrase,i.snmpv3_privpassphrase,i.ipmi_sensor,i.delay,i.trapper_hosts,i.logtimefmt,i.params,i
.state,i.authtype,i.username,i.password,i.publickey,i.privatekey,i.flags,i.interfaceid,i.snmpv3_authprotocol,i.snmpv3_privprotocol,i.snmpv3_contextname,
i.lastlogsize,i.mtime,i.history,i.trends,i.inventory_link,i.valuemapid,i.units,i.error,i.jmx_endpoint,i.master_itemid from items i,hosts h where i.hosti
d=h.hostid and h.status in (0,1) and i.flags<>2"
 30550:20181119:185302.460 slow query: 8.981808 sec, "select i.itemid,f.functionid,f.function,f.parameter,t.triggerid from hosts h,items i,functions f,t
riggers t where h.hostid=i.hostid and i.itemid=f.itemid and f.triggerid=t.triggerid and h.status in (0,1) and t.flags<>2"
 30550:20181119:185322.258 slow query: 18.383914 sec, "select distinct t.triggerid,t.description,t.expression,t.error,t.priority,t.type,t.value,t.state,
t.lastchange,t.status,t.recovery_mode,t.recovery_expression,t.correlation_mode,t.correlation_tag from hosts h,items i,functions f,triggers t where h.hos
tid=i.hostid and i.itemid=f.itemid and f.triggerid=t.triggerid and h.status in (0,1) and t.flags<>2"
 30550:20181119:185438.702 slow query: 9.202654 sec, "select i.itemid,i.hostid,i.status,i.type,i.value_type,i.key_,i.snmp_community,i.snmp_oid,i.port,i.
snmpv3_securityname,i.snmpv3_securitylevel,i.snmpv3_authpassphrase,i.snmpv3_privpassphrase,i.ipmi_sensor,i.delay,i.trapper_hosts,i.logtimefmt,i.params,i
.state,i.authtype,i.username,i.password,i.publickey,i.privatekey,i.flags,i.interfaceid,i.snmpv3_authprotocol,i.snmpv3_privprotocol,i.snmpv3_contextname,
i.lastlogsize,i.mtime,i.history,i.trends,i.inventory_link,i.valuemapid,i.units,i.error,i.jmx_endpoint,i.master_itemid from items i,hosts h where i.hosti
d=h.hostid and h.status in (0,1) and i.flags<>2"
 30550:20181119:185453.166 slow query: 9.095964 sec, "select i.itemid,f.functionid,f.function,f.parameter,t.triggerid from hosts h,items i,functions f,t
riggers t where h.hostid=i.hostid and i.itemid=f.itemid and f.triggerid=t.triggerid and h.status in (0,1) and t.flags<>2"
```

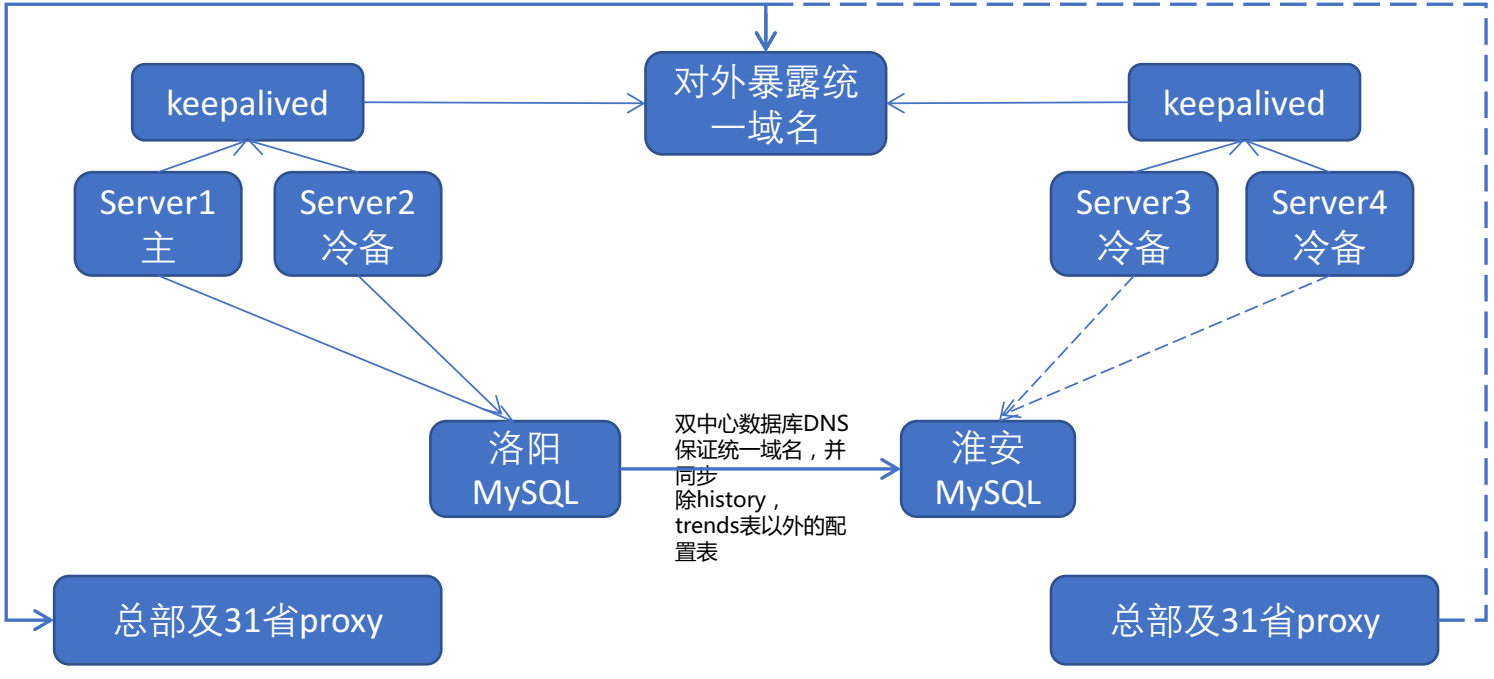
- **数据库sql执行超时配置建议**

3、降低server的pollers、java pollers、pingers、trappers等进程数配置

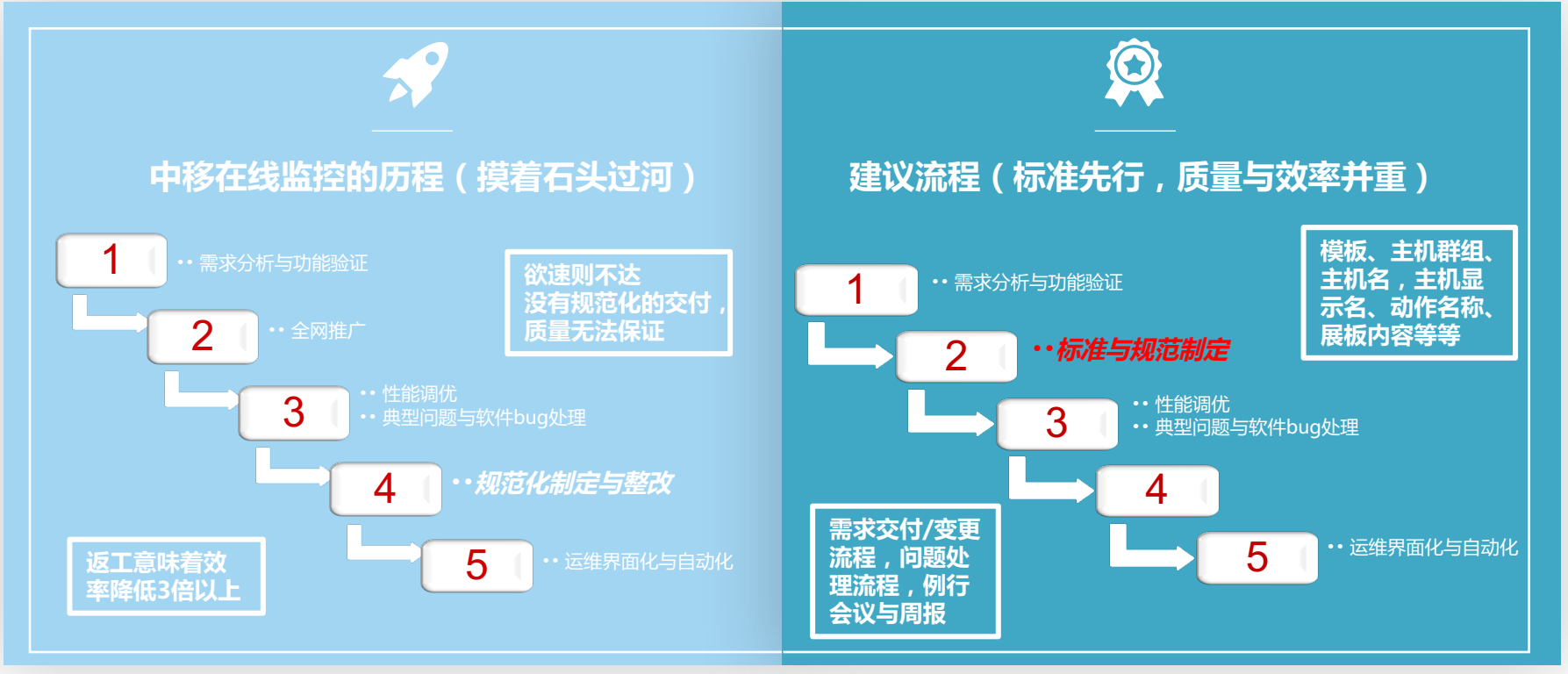
根据现网的数据库IO处理性能以及局点规模合理配置数据库超时相关参数，将max_execution_time设置为超过目前zabbix server同步配置sql执行时长的2倍以上，并定期检查zabbix_server.log日志的相应执行时长，或者增加自监控告警。

一些小总结：Zabbix双中心高可用方案探索

通过南北双中心部署Zabbix系统，对外暴露统一域名提供服务，并开发对应的proxy一键切换server能力，实现北中心机房异常时，第一通过快速切换Zabbix server域名解析，完成server服务切换。第二通过自动化脚本一键切换所有proxy对应server。保证单中心异常时，实现10min内一键监控服务的快速恢复。

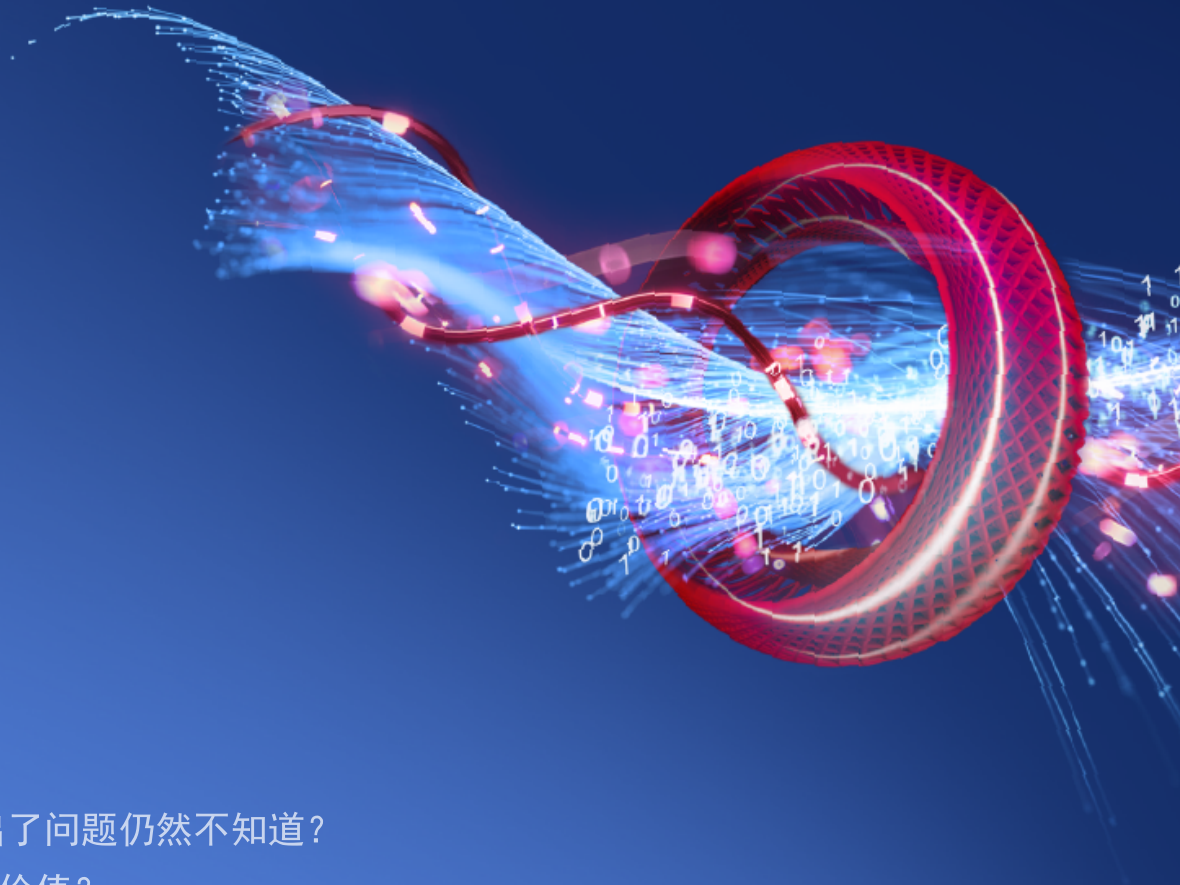


经验点



03

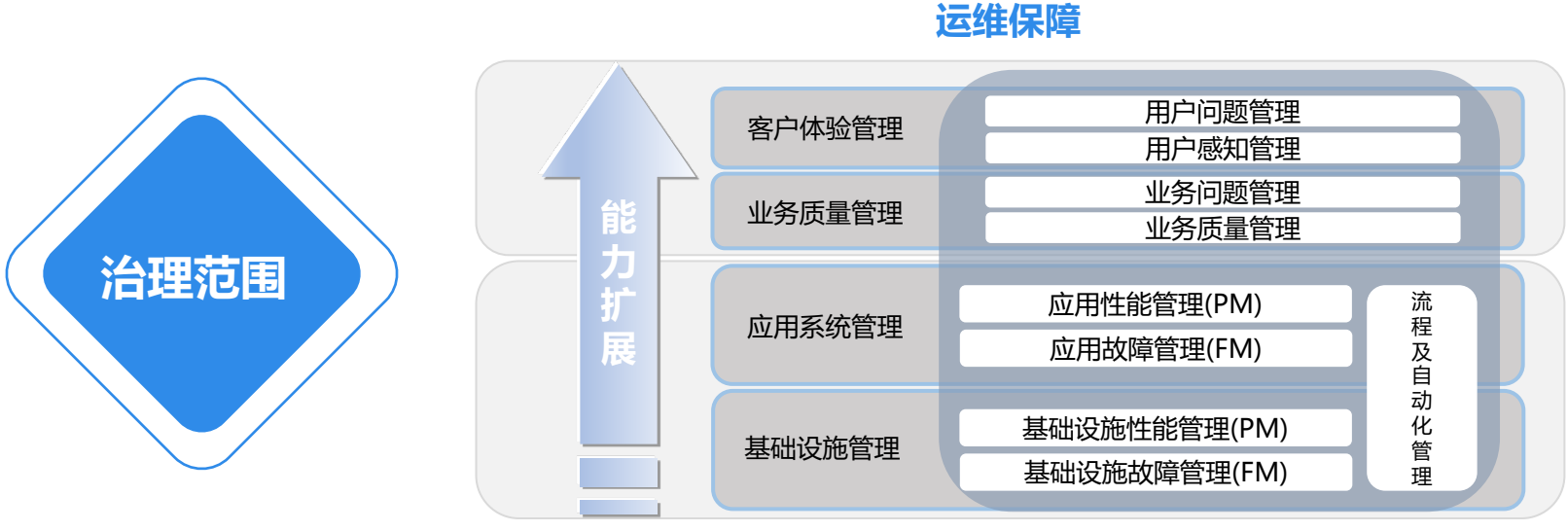
转型
——几个问题



- 问题一：200w监控指标，业务出了问题仍然不知道？
- 问题二：海量的日志是否有利用价值？
- 问题三：容器上的监控怎么做？
- 小结

问题一：200w监控指标，业务出了问题仍然不知道

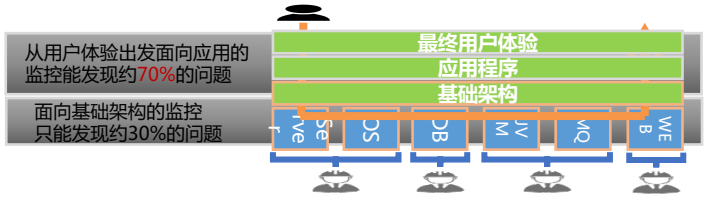
- 以业务质量和客户体验为核心，以可管控、可视化、可度量为目标。
- 全网集中建设、集中管控、边缘节点标准化接入。
- 软件监控+硬件监控一网打尽，运维数据统一、融合、流动，建立多层次度量体系。
- 以用户体验出发，建立端到端全链路监控，告警+投诉预警+客服联动形成完整闭环管理。



业务及应用质量可感知，是监控的核心

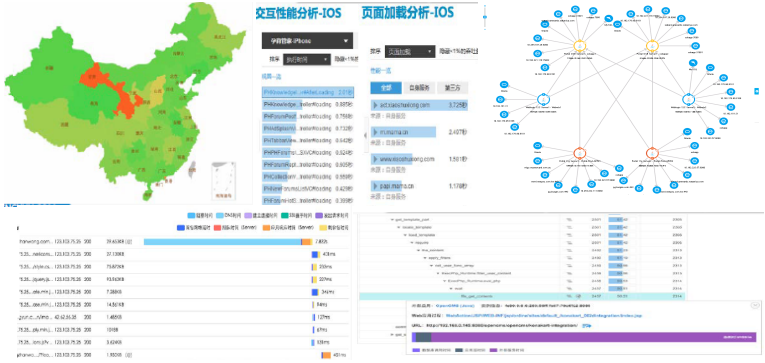
在强化基础设置监控的基础上，补充应用性能监控和业务质量监控能力，保障业务的稳定性和客户感知。

应用性能监控



应用性能监控将前台页面与后端服务以及用户网络环境真正串联，做到端到端全链路、代码级监控。

用户体验评分 → 前端交互体验 → 网络切片 → 应用调用拓扑 → 代码定位追踪

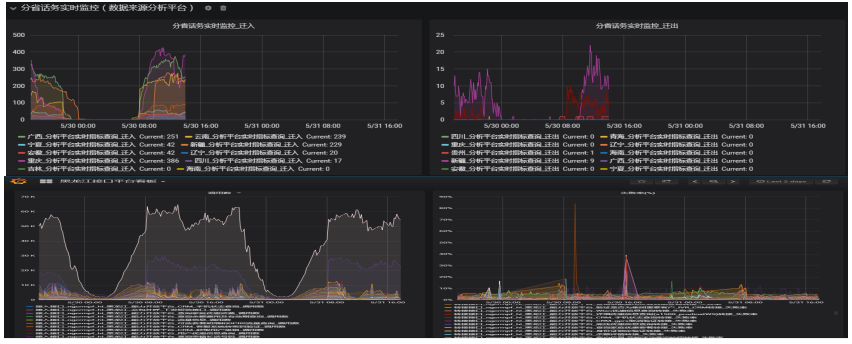


业务质量监控

梳理业务系统核心功能模块 → 梳理功能模块的核心监控指标 → 评审监控指标的提取方式及有效性 → 监控看板制作

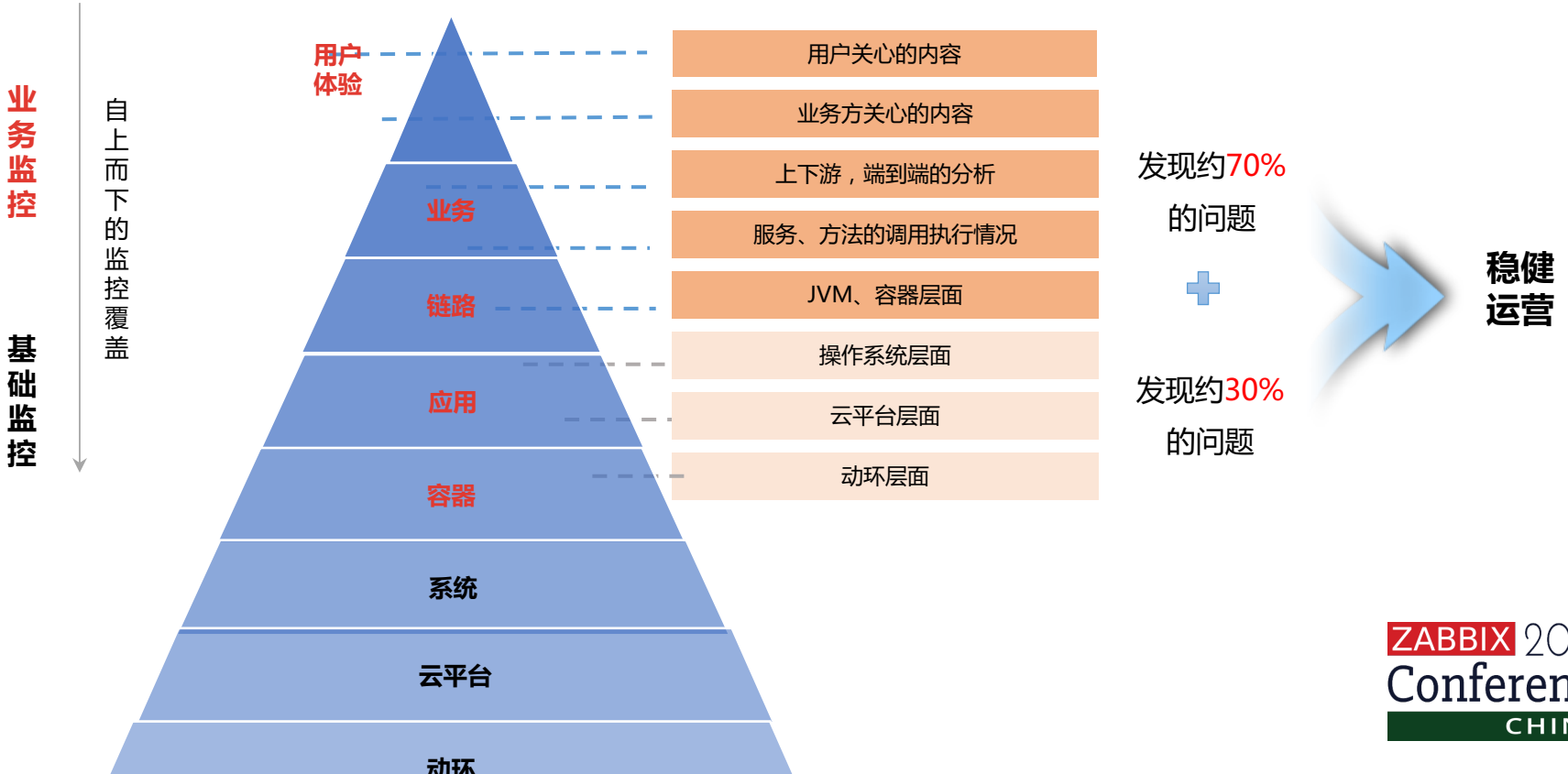
参考Google SRE五项黄金指标

1：速率：请求速率，请每秒请求数量。
2：错误：错误率，即每秒错误数量。
3：延迟：响应时间，包括队列 / 等待时间，以毫秒为单位。
4：饱和度：即过载程度，指标与资源利用率相关，也可通过队列深度进行直接衡量。
5：利用率：资源或系统的繁忙程度，通常表示为 0% 至 100%。



ZABBIX 2020 Conference CHINA

云化架构下的监控分层

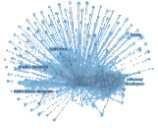


问题二：海量的日志是否有利用价值？

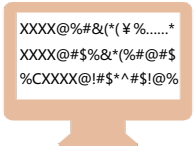
对于亚健康状态，异常日志比系统故障更早出现。由于海量日志存储在海量网元中，不同厂商日志标准不统一且可读性差，往往很难鉴别真正触发异常的日志。

挑战

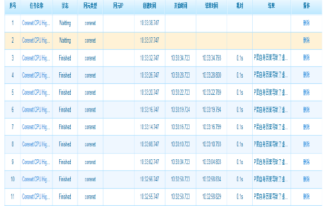
海量日志保存在海量网元中，缺乏统一视图



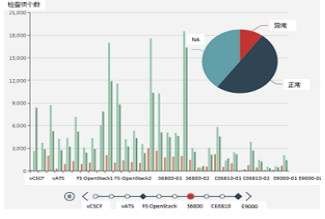
不同厂商设备的日志缺乏统一标准，可读性差



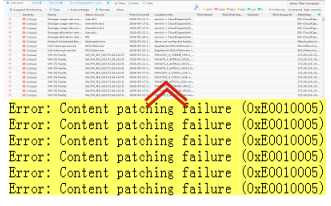
①跨厂商设备日志统一查询



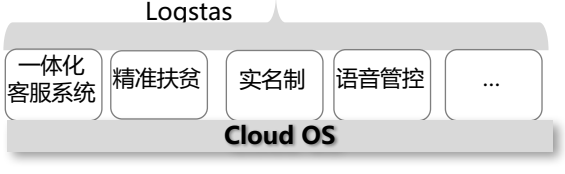
②异常日志统计

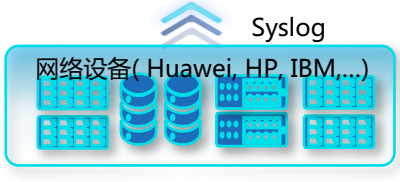


③异常日志分析与告警推送



统一日志分析





价值

日志统一采集，统一呈现，异厂商设备日志统一查询

针对异常日志进行统计，实时推送异常日志告警，提升亚健康网络问题定位效率

统一日志平台

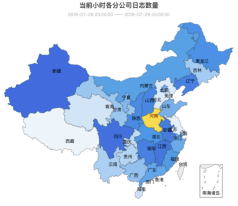
采集灵活

- **日志接收方式：**HTTP、TCP、UDP、文件监听、SYSLOG
- **插件化：**在数据采集层针对不同的数据类型以插肩袖形式进行部署采集，便捷轻量
- **支持对多类数据采集模式，**采集数据种类包括文件、网络设备、数据库、容器、用户行为、前端异常等，

下一步将支持对采集插件实行可视化部署配置

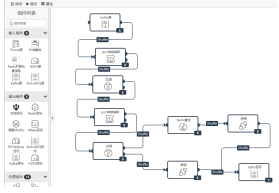
跨网传输

- **支持实时的跨机房传输：**在一个监控看板掌控全国各地区的设备运行状态，任何一台机器出现异常日志都能及时告警定位，真正做到**全面监控纳管**



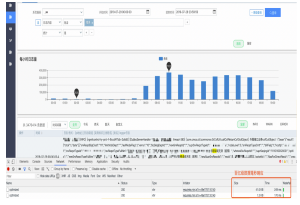
实时计算

- 使用流处理的实时计算能力将明细日志在传输过程中完成聚合分析，计算的结果直接用于监控告警展示，极大提升了运维人员在故障发生时的**决策能力**



快速检索

- 支持对海量日志全文检索和按字段精准查询，支持根据日志相似性进行聚类统计。



精准下钻

- 支持全链路的日志追踪功能，**清晰定位故障根源。**
- 支持上下文查询便于在业务故障排查中**快速查找相关故障信息。**

应用情况

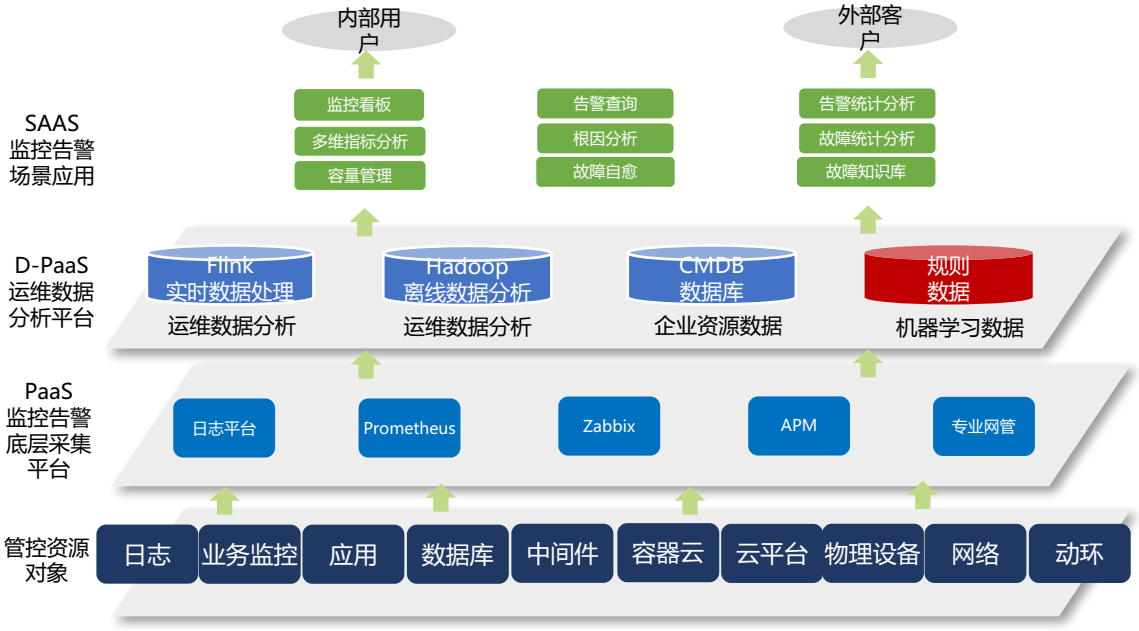
50T+ 600亿行+ 日写入	100万行+ 秒索引	10万+ 毫秒返回 日查询	20000+ 客户端
覆盖8000+ 服务器 范围	500人+ 用户	31省 + 洛阳淮安 机房	400节点+ 集群

问题三：容器上的监控怎么做？

2019年，公司进行了底层技术的升级，云平台从以Openstack虚拟化为主体升级为容器云，目前生产环境有30000+Pod实例在运行。目前我司的以我们选取与其天然支持的Prometheus+Granfana进行监接，直接进行日志关键字监控。

监控对			
容器云监	云化服务监控	业务入口访问量监控	node_exporter,对入口nginx的请求量进行监控，业务高峰期为0即产生告警
		业务核心nginx状态监控	node_exporter,对核心nginx请求的响应码及最大响应时间进行统计，非200或大于3s比率过高即触发告警
		业务接口运服务治理	node_exporter,对服务调用的核心csf接口的成功率进行统计，成功率过低及响应时间过高即触发告警
		业务接口运行状态	node_exporter,对服务调用接口平台的成功率及响应时长进行统计，成功率过滤或响应时间过高即触发告警
		业务核心日志异常关键字监控	node_exporter,通过大数据的手段对服务的日志进行清洗过滤，统计错误关键字的次数，并对产生错误关键字的pod进行监控告警
		dubbo线程池状态监控	node_exporter,对dubbo线程池的active状态进行监控，超过75%即触发告警
		核心业务量监控	node_exporter,对业务的核心业务量指标进行统计，如接通率，订单量，当前通话人数等核心指标过低即触发告警

构建能力开放、自主可控的监控告警体系建设



“一平台、四体系、三能力”的自动化运维体系

一个平台：监控告警平台

四大体系

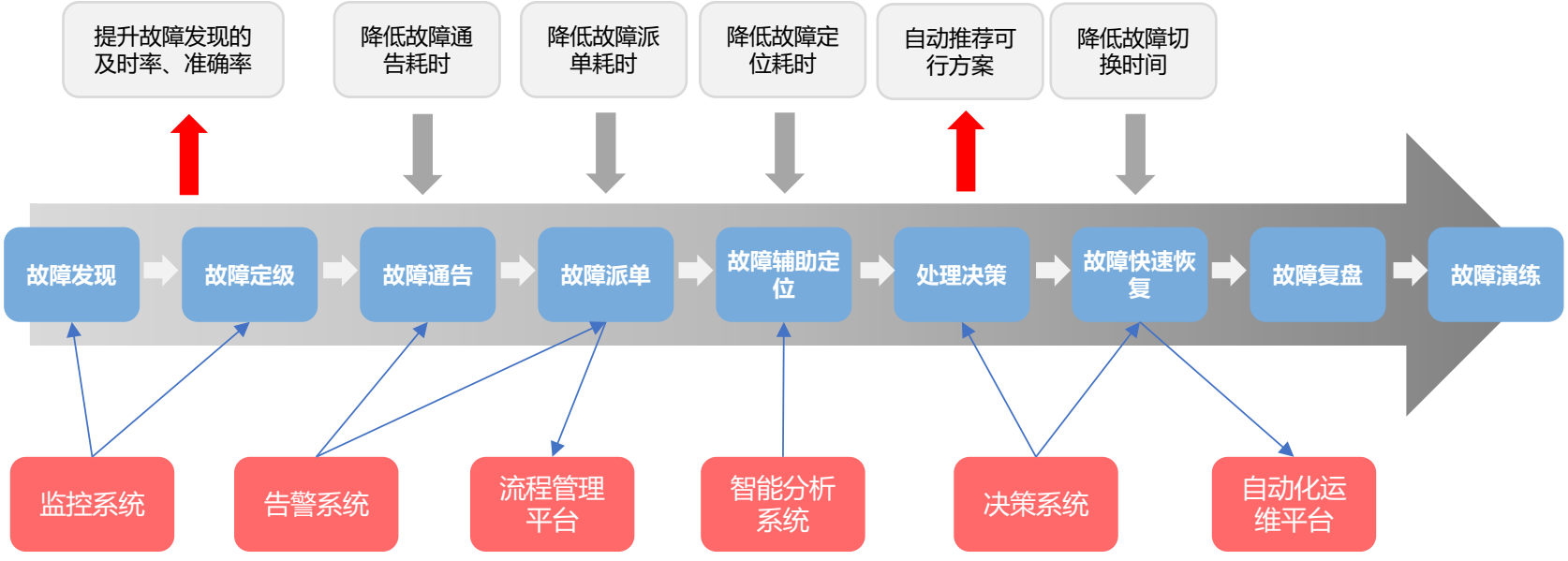
- 四位一体联动体系：控、监、管、营四位一体
- 决策分析支撑体系：基于大数据分析、人工智能算法支撑日常故障处理
- 开发运维协同体系：建立高效组织协同机制，开放运维能力
- 运维服务评估体系：运维服务评估360度可度量

三项能力

- 自动化能力：高效执行，提高人员效能
- 数据化能力：运行状态透明化，数字化驱动系统优化
- 智能化能力：利用人工智能算法辅助运维进行决策分析

实现端到端的自动化、智能化运维场景

通过平台基础能力建设，结合规则、算法、流程引擎、故障树等技术方式，实现故障运维的端到端自动化处理，提升运维工作效率和质量。

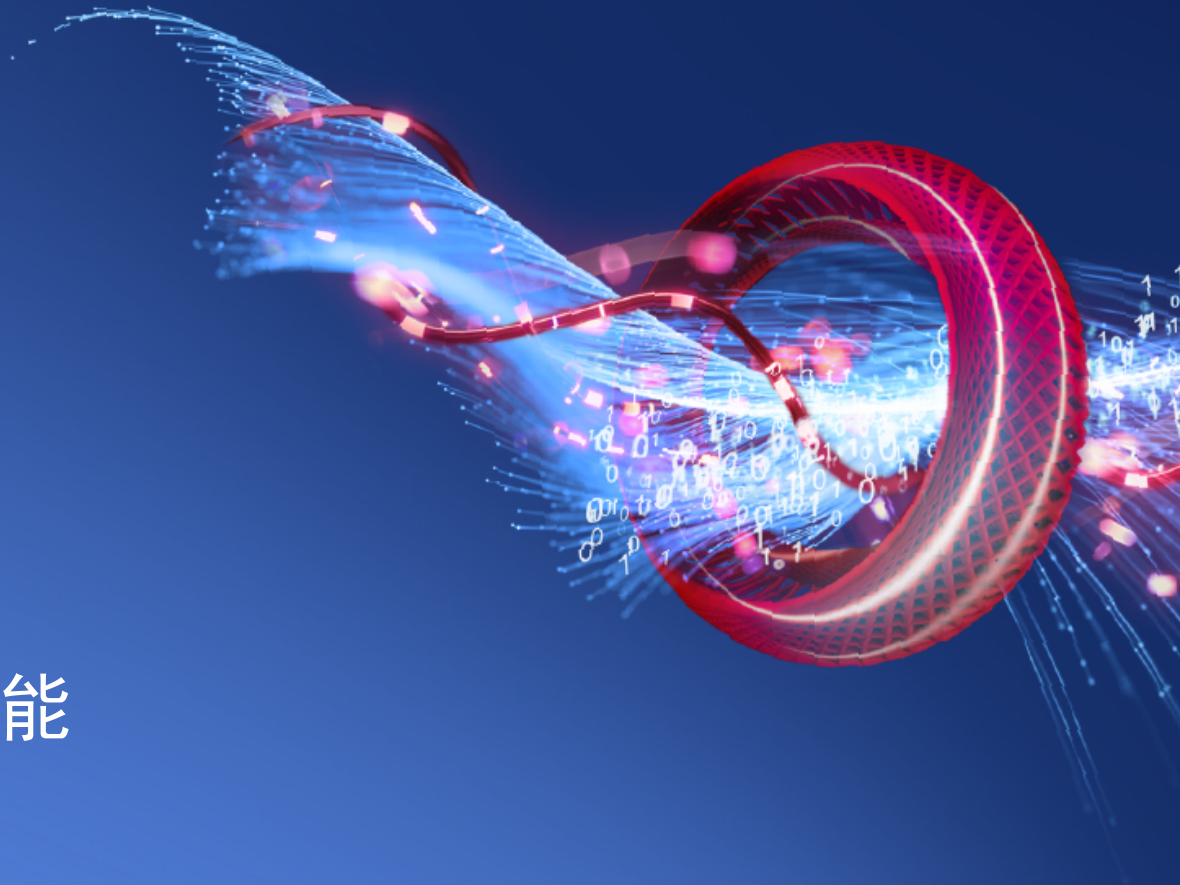


04

沉淀

——让监控多些可能

- 自动化提速
- 数据化赋能



ZABBIX 2020 Conference CHINA

1、自动化提速

痛点

- 大型公司基础资源多，业务广，线上变更频繁，监控配置任务量大
- 监控添加不是一蹴而就，需要反复调整，重复工作量大
- 开源工具使用门槛高，大多没有好用的web界面，需要培训才能灵活使用
- 中移在线公司业务/工作人员遍布全国各省，基础资源达到上万级别，业务变更频繁，统一管理难度系数高

应对方案

1. 监控能力标准化、流程化、模块化
2. 二次开发、自动化
3. 配置界面化 数据展示界面化

当前监控自助可实现各类监控自助式增-删-改-查，已覆盖公司***85%***监控需求，目前只需2个自有人员即可维护整个监控系统。实现主机资源申请流程一键添加基础监控、告警联系人增删改查、监控覆盖度一键查询、zabbix定期巡检等

监控分类	监控对象
网络设备	华为、思科、锐捷、华三
主机	Suse、CentOS 、Redhat 、Windows、Ubuntu
数据库	Mysql、Oracle
中间件	Tomcat、Redis、Es、Mq
进程	CPU、内存、存活、端口
日志	关键字
拨测	拨测状态
自定义指标	命令类、SQL类、脚本类
带外硬件	华为、联想、戴尔、惠普、AMAX、浪潮等

1、自动化提速-服务自助化的底线

CMDB 中 CI 类型	监控对象	一级告警标签	二级告警标签	默认必选	严重告警阈值
主机监控接收人	PC服务器	通断告警	宕机	是	
		性能告警	mem	是	日志关键字，不能分配内存；
			cpu	是	计算型：95%，持续15分钟；非计算型85%；load avg 2.5
			文件	是	95%&剩余量小于256G&持续增长
	虚拟机	通断告警	宕机	是	
		性能告警	mem	是	日志关键字，不能分配内存；
			cpu	是	计算型：90；非计算：80
			文件	是	95&小于256&增长量
	网络设备	通断告警		是	
	网络设备	性能告警		是	
服务监控接收人	业务进程	通断告警		是	
		性能告警		是	
	日志	日志关键字告警		是	
	mysql	通断告警		是	
		性能告警		是	
	oracle	通断告警			
		性能告警			
	elasticsearch	通断告警		是	
		性能告警		是	
	redis	通断告警		是	
		性能告警		是	
	rocketmq	通断告警		是	
		性能告警		是	
	容器	通断告警		是	
	容器	性能告警		是	cpu: limit,70%；mem: 95%
自定义监控接收人	拨测	拨测告警			
	tomcat-jmx	业务告警			
	nginx	业务告警			
	接口	业务告警			
	dubbo status	业务告警			
	自定义监控	业务告警			

当监控自助化服务开放不久，就发现了一些问题：监控增删改的权限放给了用户。用户出于各种原因会将一些关键监控去除或者阈值设置过高，导致故障无法及时发现。

用户便利 **VS** 监控管理

资深专家划分17种网元监控类型，设定监控类型的**不可更改的生命线。**确保在服务外放的同时，及时保证“**挂了我知道**”。

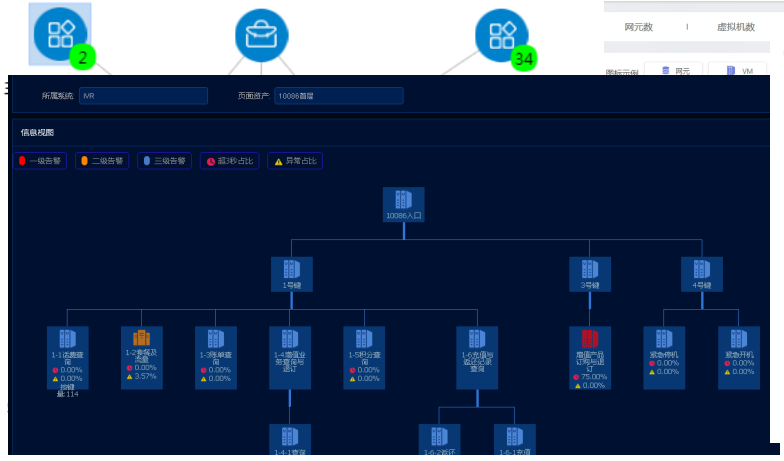
2、数据化赋能

抽取Zabbix、Prometheus、告警平台、日志平台、CMDB等数据，最终装载到大数据分析平台中，进行多维度的数据分析。

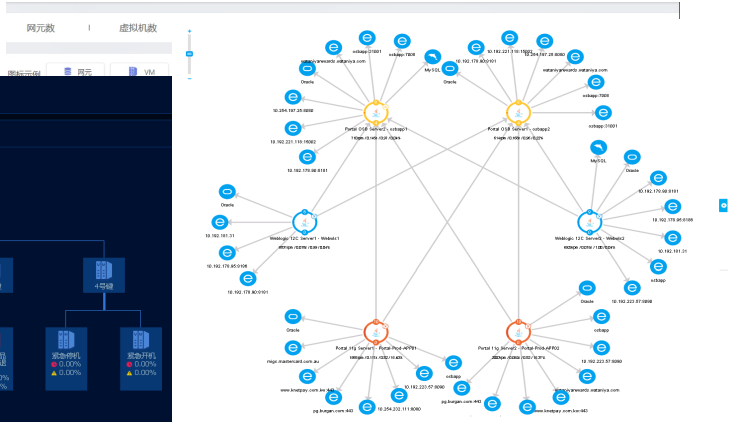
应用扩展—根因分析

监控告警数据与资源数据深度结合打造一个法典，**两套拓扑**，实现**四看一报**，即看全貌，看局部，看纵向，看横向，报故障。达到故障及时发现，根因清晰定位，影响范围一眼可见。

纵向拓扑



横向拓扑



与CMDB深度结合打造**纵向拓扑**，实现业务系统与基础设施的使用关系关联

与企业消息总线数据关联，与日志数据结合，打造业务调用关系**横向拓扑。**

应用扩展—故障自愈

通过监控告警实时数据，通过自动化能力，实现简单故障自愈。

自动扩容

CPU/Mem 百分比配置

request_cpu

Dubbo线程池占比

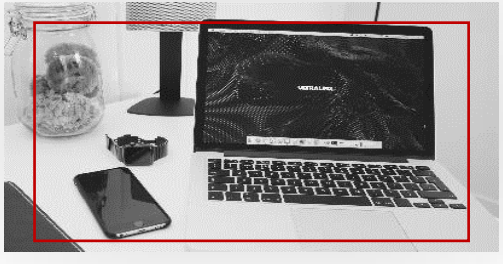
HTTP连接数

自动重启

宕机监控

拨测结果

NGINX 502个数



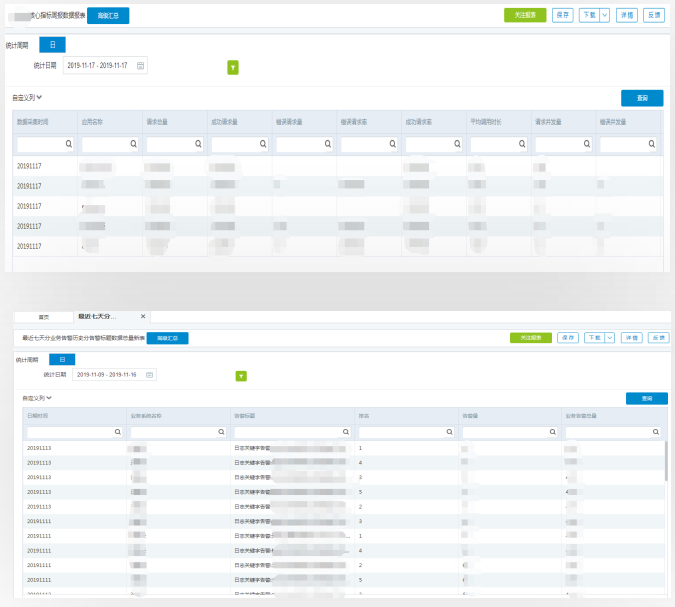
自动接口关闭

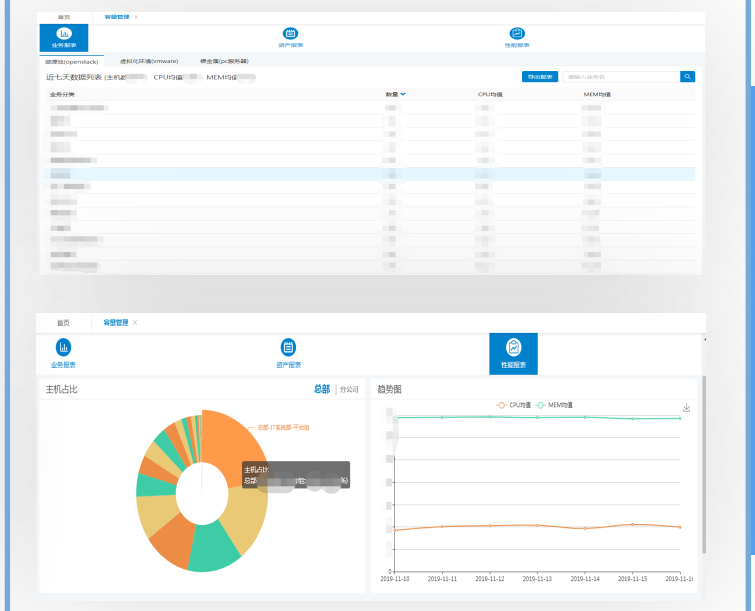
接口失败率

接口超时率

应用扩展—容量管理

大数据分析报表





容量管理平台

ZABBIX 2020 Conference CHINA

现在的数字

3.4 万 主机	2606 万 监控项	78 万 触发器	198 万 报警
600亿 日志	1700+ DashBoard	975 用户数	84 个 Proxy

05

蜕变

——AIOPS在监控告警方面的尝试

- 阈值正确设定
- AI预测拟合曲线
- 智能化运维

ZABBIX 2020 Conference CHINA

当前的主要矛盾是：海量的告警和有限的专家

运维主管

工程师小明

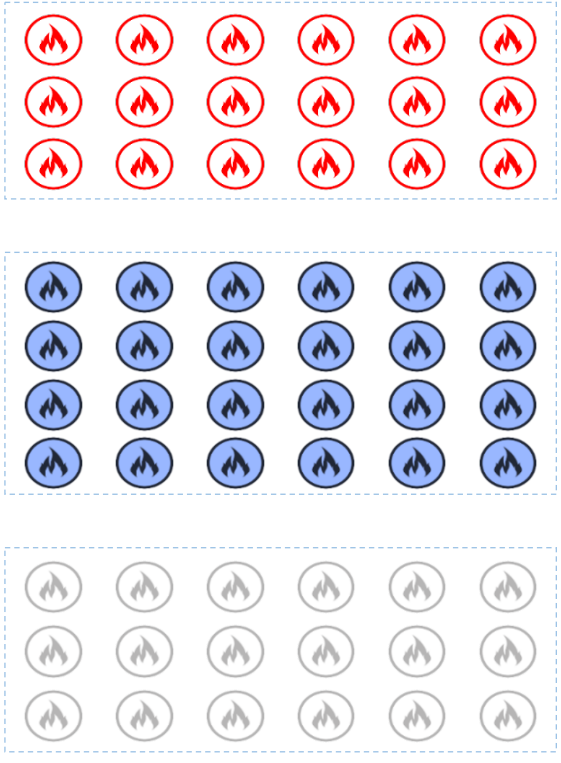
监控要“多而全”，
一个问题都不能放过！

VS

告警要“少而精”，不要重复和误报

2606万+ 监控指标，**198万+** 告警/天，**2000+** 短信/每人每天

阈值正确设定是平衡“多而全”和“少而精”的关键手段之一



告警

正常告警

- 缺少压缩&关联

误报

- **阈值不合理：80%**
- 监控能力不足：10%
- 人员配置失误：10%

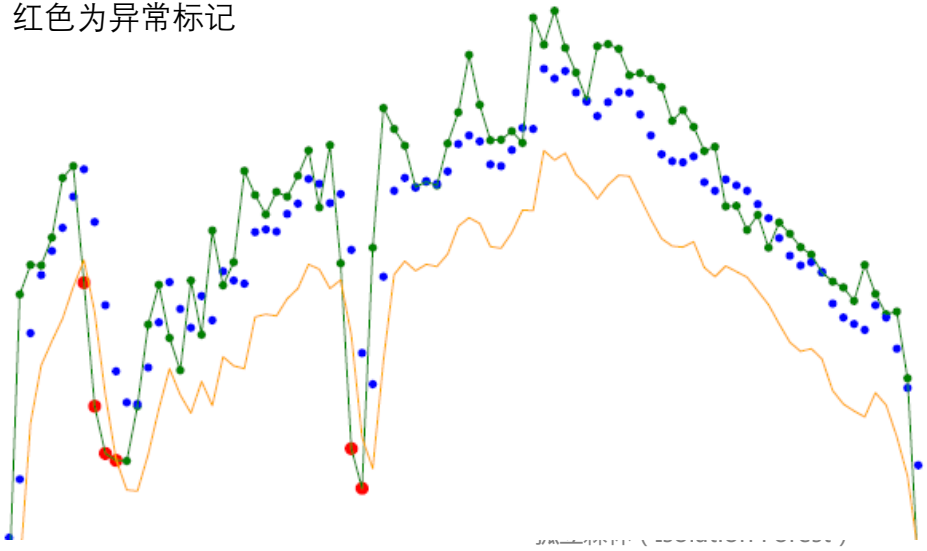
漏报

- **无法设定阈值：70%**
- 无监控：30%

阈值设定从依靠专家经验向智能动态设定演进

基于结构化的时序数据，通过AI预测拟合曲线，进行异常检测

绿色为实际指标；蓝色为预测指标；黄色为预测方差区间；
红色为异常标记



历史数据分析

历史数据读取和清洗

- 数据抽取ETL
- 断点修复
- 数据间隔调整
- 自相关性分析

- Moving Average移动平均滤波 (ARIMA)
- Exponential Smoothing指数平滑滤波 (Holt-Winters)
- N*sigma统计检测

日同比（Day over Day method）
箱线图（Box-whisker plot）

异常判定

途径一：N-sigma方差

途径二：专家标记

智能化运维并不是我们想象的那样遥不可及

告警准确率
提升到80%

数据

- 海量数据源（性能指标、日志、告警）
- 可以迭代预测、迭代标注……

告警覆盖率
提升到95%

算法

- TensorFlow等成熟算法库
- 针对不同场景，可选择不同算法，如LSTM用于趋势预测、ARIMA用于回归过滤异常

告警配置人力下降60%

计算

- 轻量化
- 虚拟机部署，4C32G即可起步

AIOps常见应用场景

未来


深度
日志异常检测、
告警压缩&关联、
告警规则生成、
容量管理、性能管理等
智能故障发现
让智能化在更多运维领域落地开花
广度

一则招聘广告

- 享受互联网般技术挑战
- 国企稳定待遇
- 与客户交互产生的海量数据，包括语音、文本、图像等数据
- 郑州、北京、上海、深圳研发中心，31省会城市
- 公司年轻、人员年轻、扁平化管理

【招聘岗位】

系统架构师、运维开发、应用运维、数据库运维、大数据运维、数据分析、容器云开发、云计算开发、JAVA开发

中移在线

官方网址：http://online.10086.cn/

加入我们！

联系我们

Contact us

Zabbix 中国致力于为国内用户提供培训、咨询、以及其他的专业技术支持。也为国内的用户搭建交流学习的平台。

138-1772-0274

china@zabbix.com

www.grandage.cn
www.zabbix.com/cn

上海市徐汇区虹梅路1905号

Zabbix开源社区

Zabbix中国

Zabbix_China

Zabbix_team

Zabbix 开源社区

加入技术交流群

ZABBIX 2020
Conference
CHINA

THANK YOU