数据驱动IT运维数字化转型

------主机监控产品介绍

常州慧鹰科技

SMARTEAGLE
慧鹰

目录

Contents

Part.

1

SMARTEAGLE
慧鹰

产品定位

慧鹰SEHM（SmartEagle Host Manger）是慧鹰科技最新研发的一代以数据运维为核心的运维平台。借助于大数据分析技术、人工智能算法，可视化技术实现以数据驱动的主机、业务的高保障运维管理。

关于云鹰

目标客户

- 服务器众多、业务系统需要高可用
- 业务系统部署结构错综复杂，业务保障需要加强
- 客户运维规划发展支撑需要更多数据支撑

产品主要功能

- 全息大数据技术、全局态势感知，全面感知当前、将来的运维状态
- 大数据技术实现趋势预测，用于异常检测、可用性监测、辅助决策
- 场景化分析模型，辅助根原因定位（场景如：主机硬件异常、主机性能瓶颈、应用假死、主机容量风险、业务高可用性）

产品价值

- 通过大数据挖掘，实现主机、业务运维态势以及健康度全面掌控，助力IT部门数字化转型。
- 专注业务高可用运维，落实SLA实践，保障服务双方之间所提供服务品质

关于云鹰

产品定位不同

本产品是大数据分析在IT运维领域的场景挖掘，属于大数据分析产品范畴，核心是“数据挖掘、数据分析”，是主机管理在“深度”上的拓展，从而实现全面掌控。

数据采集不同

不再局限于IT协议采集的单维度数据，横向扩展数据采集方式。对管理对象数据采集的频率、指标数量也远大于运维监控产品。

核心技术不同

本产品采用人工智能技术，对数据进行机器学习、深度学习，进而进行模式识别。

用户感知不同

传递给用户的是人工智能算法计算后的结果，提供的是动态阀值预警，非固定阀值；呈现的是主机当前和将来健康度值，非主机当前状态的颜色标识。

Part.

2

SMARTEAGLE
慧鹰

产品架构

本节主要阐述产品的技术原理以及系统架构

2.1.1

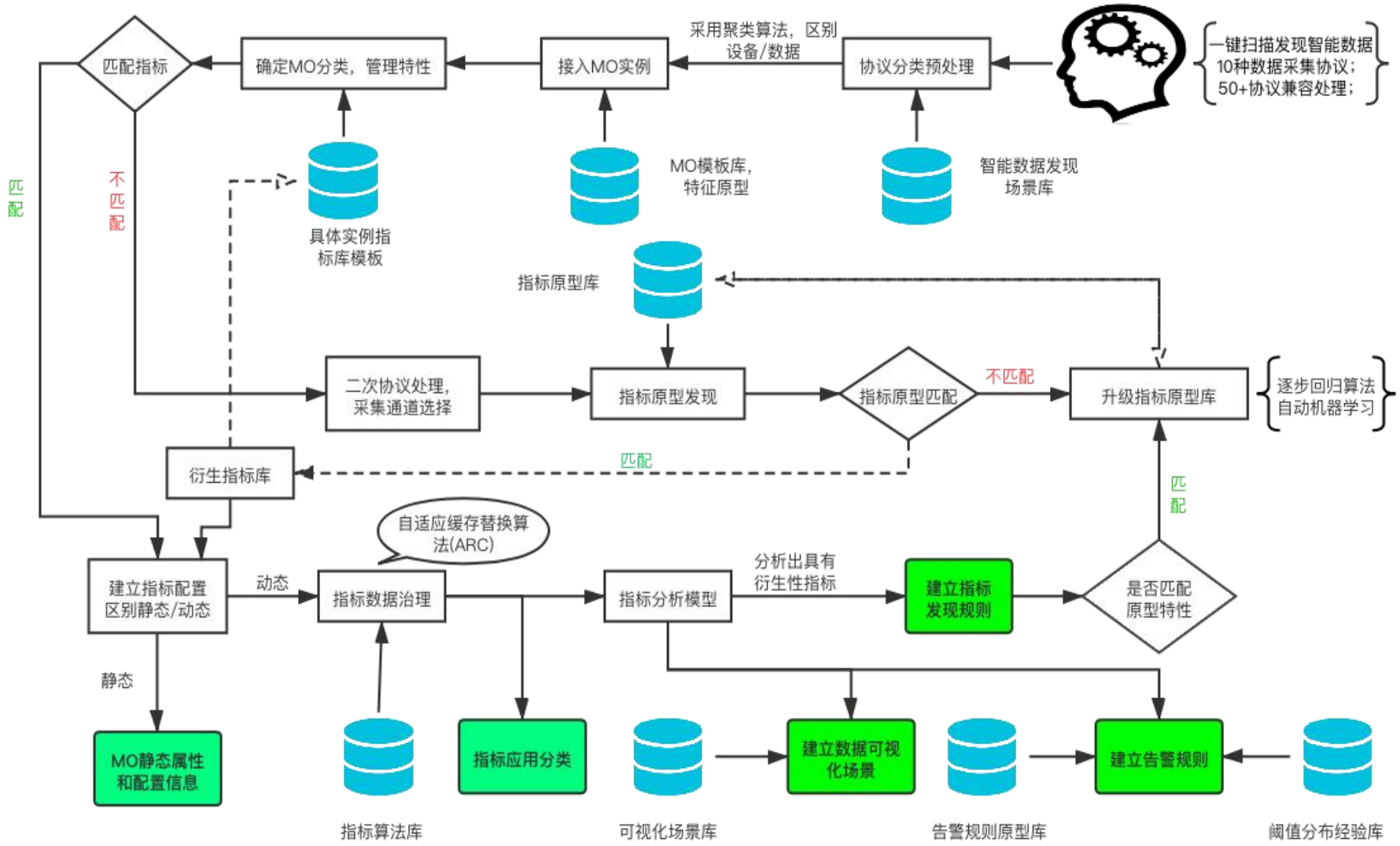
全息大数据技术

多数据源采集，通过植入探针，数据包分析，日志分析，爬虫抓取， ETL以及协议采集。数据库集群+时序数据库+NOSQL数据库存储，提升系统访问性能。数据可视化技术，通过交互式实时数据可视化大屏来帮助运维人员发现并诊断业务问题辅助决策。

2.1.2

基于原型感知技术

扫描自动发现智能数据，包括硬件和软件，通过数据发现感知匹配场景库，对于纳入管理的智能数据，通过原型感知技术，自动识别设备特征以及设备内部结构和配置信息。从而连接物联、互联、业务系统，构建全息数据。



基于神经网络等人工智能技术-----技术原理

1 模式识别：从中长周期中寻找数据特征

机器学习、深度学习的周期可以是几周、几个月、甚至几年，寻找特征（如上下限、指标相关性、周期特征、趋势特征、季节特征等）的规律。

2 趋势预测：异常检测、辅助决策

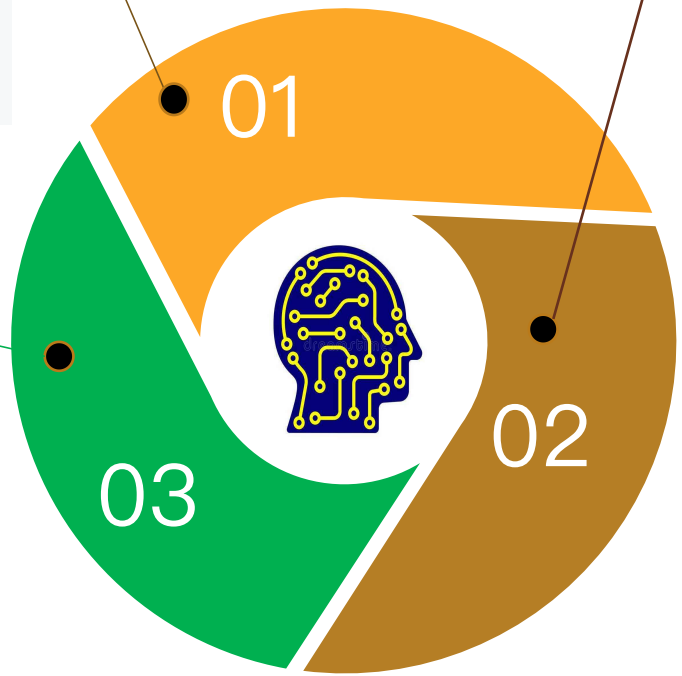
异常检测：
通过历史数据特征，验证当前的数据状态，以发现特征异常（如趋势异常）；

容量预测：
通过历史数据预测未来（如各种资源），合理规划IT资源；

辅助决策：
通过不断的数据挖掘（如一段时间后的各种对象运行趋势），实现全面掌控，从而辅助决策。

3 辅助分析异常根源

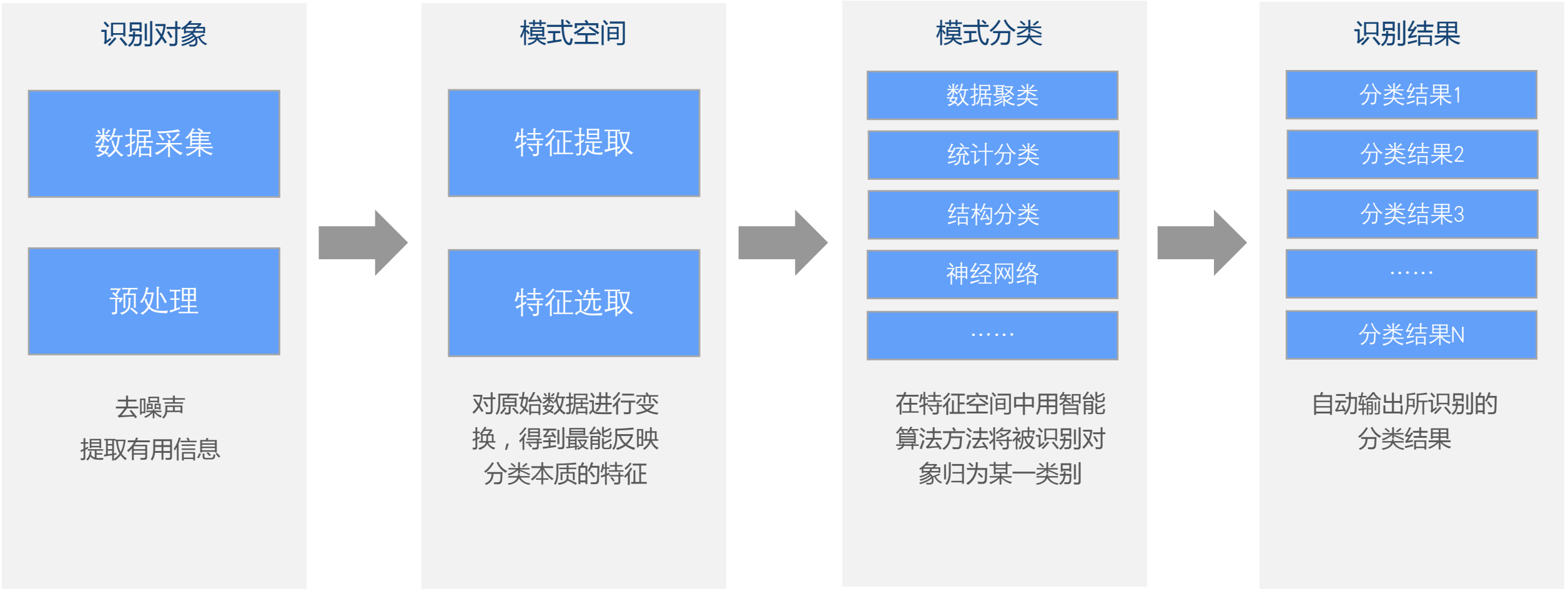
在发现特征异常后，启动异常检测系列算法，结合相关性分析，系统将自动标记出异常点，更进一步，通过**场景化分析模型**辅助分析异常根源。



基于神经网络等人工智能技术——技术原理

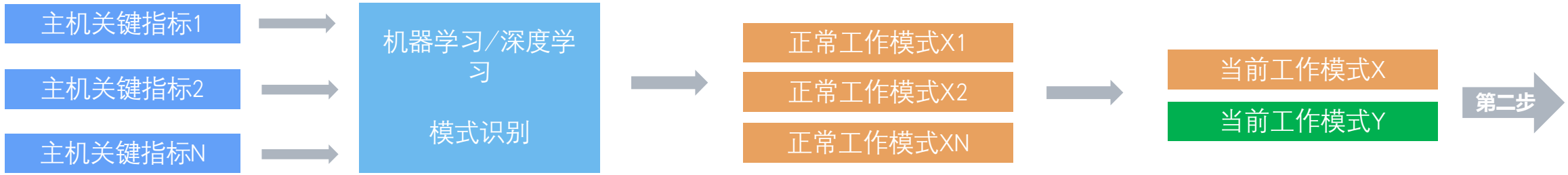
这一次，重磅突破，运维界首次引入模式识别技术！

模式识别技术当前成功应用于图像识别（人脸识别）、语音识别、医学诊断、自动驾驶.....

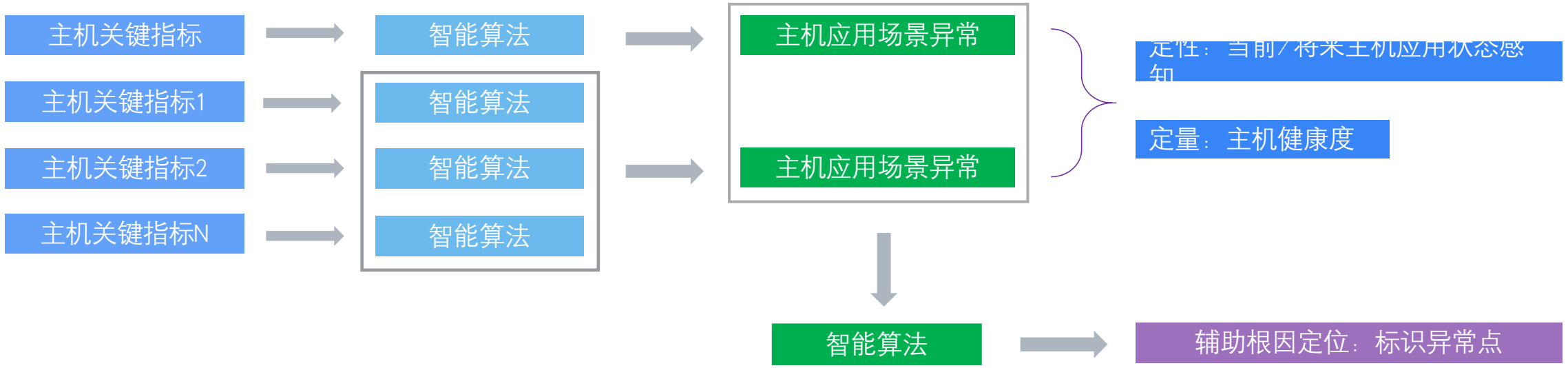


基于神经网络等人工智能技术------技术原理

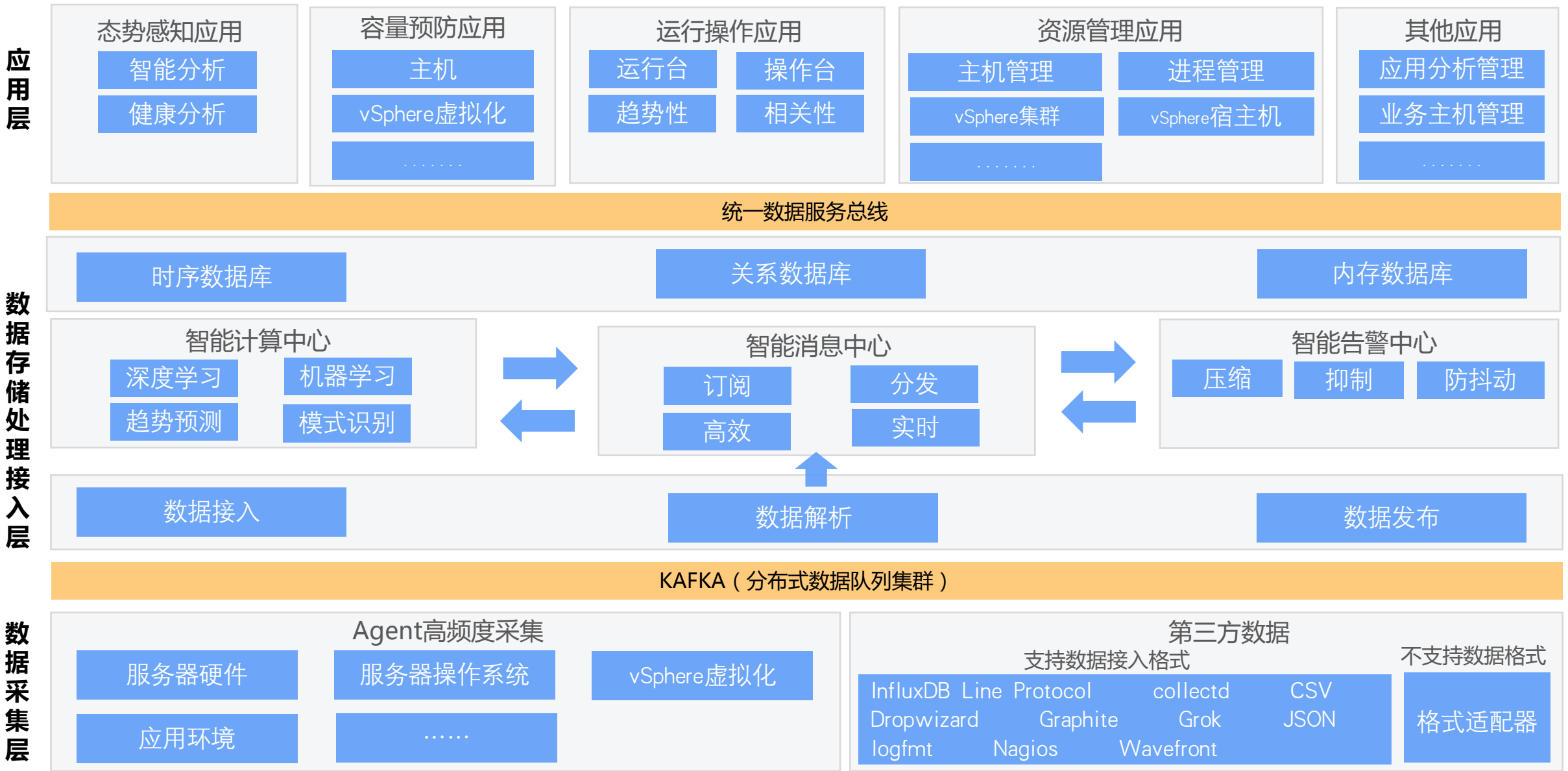
第一步：深度机器学习--模式识别及判定



第二步：异常检测、根因辅助定位



2.3.1 系统架构


应用层
态势感知应用
智能分析
健康分析
容量预防应用
主机
vSphere虚拟化
运行操作应用
运行台
操作台
趋势性
相关性
资源管理应用
主机管理
进程管理
vSphere集群
vSphere宿主机
其他应用
应用分析管理
业务主机管理
统一数据服务总线
数据存储处理接入层
时序数据库
关系数据库
内存数据库
智能计算中心
深度学习
机器学习
趋势预测
模式识别
智能消息中心
订阅
分发
高效
实时
智能告警中心
压缩
抑制
防抖动
数据接入
数据解析
数据发布
KAFKA（分布式数据队列集群）
数据采集层
Agent高频度采集
服务器硬件
服务器操作系统
vSphere虚拟化
应用环境
……
第三方数据
支持数据接入格式
InfluxDB Line Protocol collectd CSV
Dropwizard Graphite Grok JSON
logfmt Nagios Wavefront
不支持数据格式
格式适配器

2.3.2 数据架构图


数据呈现
平台
运行台
趋势性
相关性
应用
智能分析
健康分析
容量预防
数据存储
时序数据库
关系数据库
内存数据库
数据建模
可视化建模
算法管理
算法模型管理
计算规则管理
数据采集
服务器硬件数据采集
服务器系统数据采集
vSphere虚拟化数据采集
应用性能数据采集
其他对象采集
原始数据
ETL
数据接收
KAFKA
分布式数据队列集群
消费
流式/批量处理
数据处理
数据存储
实时算法
智能告警
使用场景

目录

Part.

3

功能介绍

通过CEHM实现的数据运维会给我们带来哪些运维场景实践？

SMARTEAGLE
慧鹰

典型功能一 ——

全局态势感知：当前、未来健康感知

全局态势感知：当前、未来健康感知----主机运维的现状及困惑

- 众多主机、海量数据，究竟哪些需要重点关注、关注什么指标？

全局态势感知：当前、未来健康感知----主机稳态/异常识别

- 通过模式识别算法识别主机稳态/异常运行

主机硬件

CPU性能

内存性能

磁盘性能

网络性能

应用进程状况

进程运行环境

容量资源

模式识别算法

全局态势感知：当前、未来健康感知----异常检测辅助根因定位

非稳态运行即启动异常检测算法

- 主机硬件
- CPU性能
- 内存性能
- 磁盘性能
- 网络性能
- 应用进程状况
- 进程运行环境
- 容量资源
- ……

异常检测算法

健康度评价

标识异常点

场景化根因定位

3.1.4

全局态势感知：当前、未来健康感知----智能分析过程动态感知

预测与实测发生背离，发出预警，便于提前介入，避免重大故障发生。

场景分析模型结果：呈现健康度和风险状况

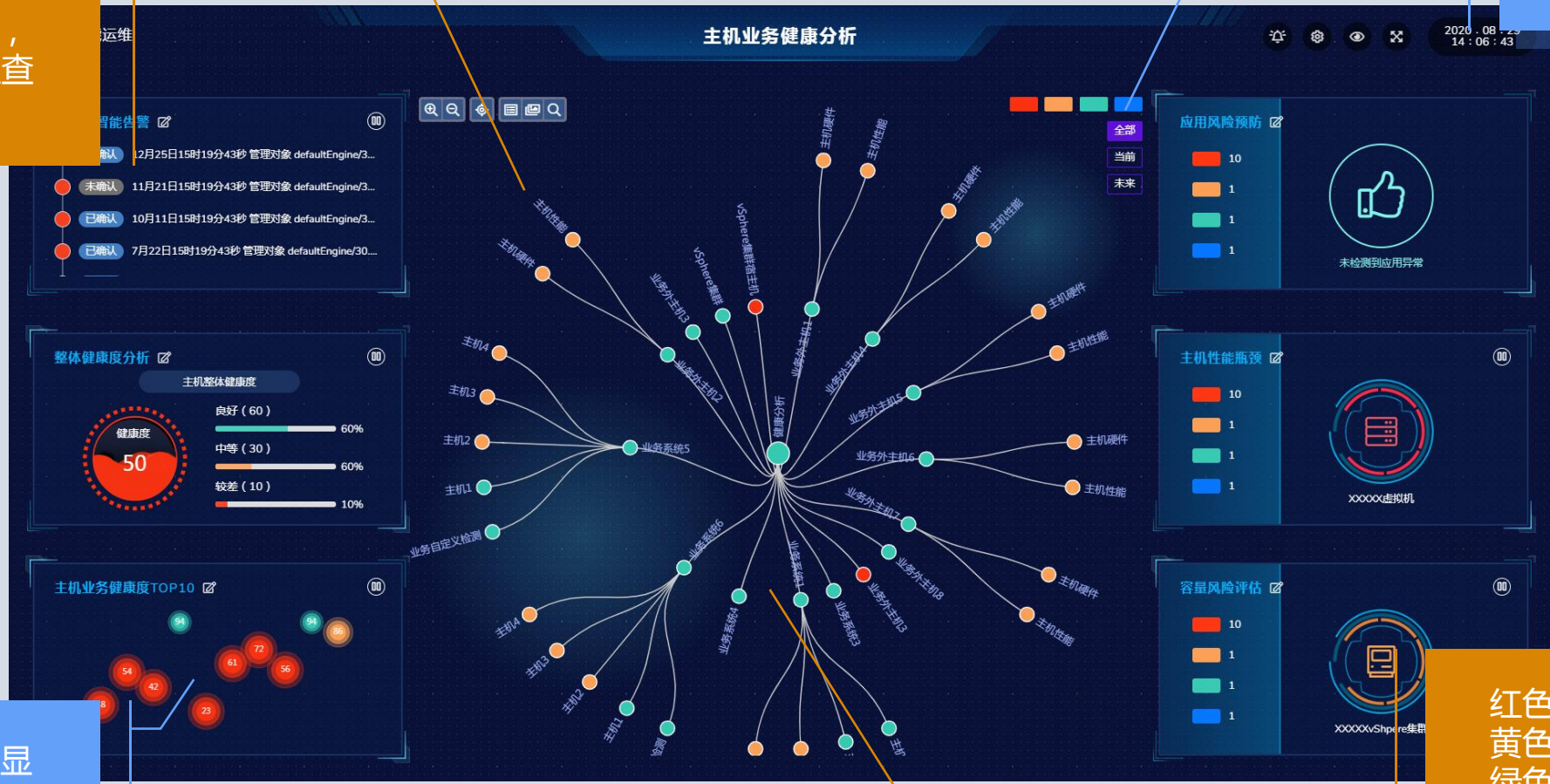
基于人工智能技术，对历史数据进行机器学习、深度学习。

3.1.5

全局态势感知：当前、未来健康感知----全局：当前/未来健康感知

全面呈现：主机健康度、业务健康度、vSphere容量风险，并可展开分析模型查看风险点。

切换了解到当前/未来状况



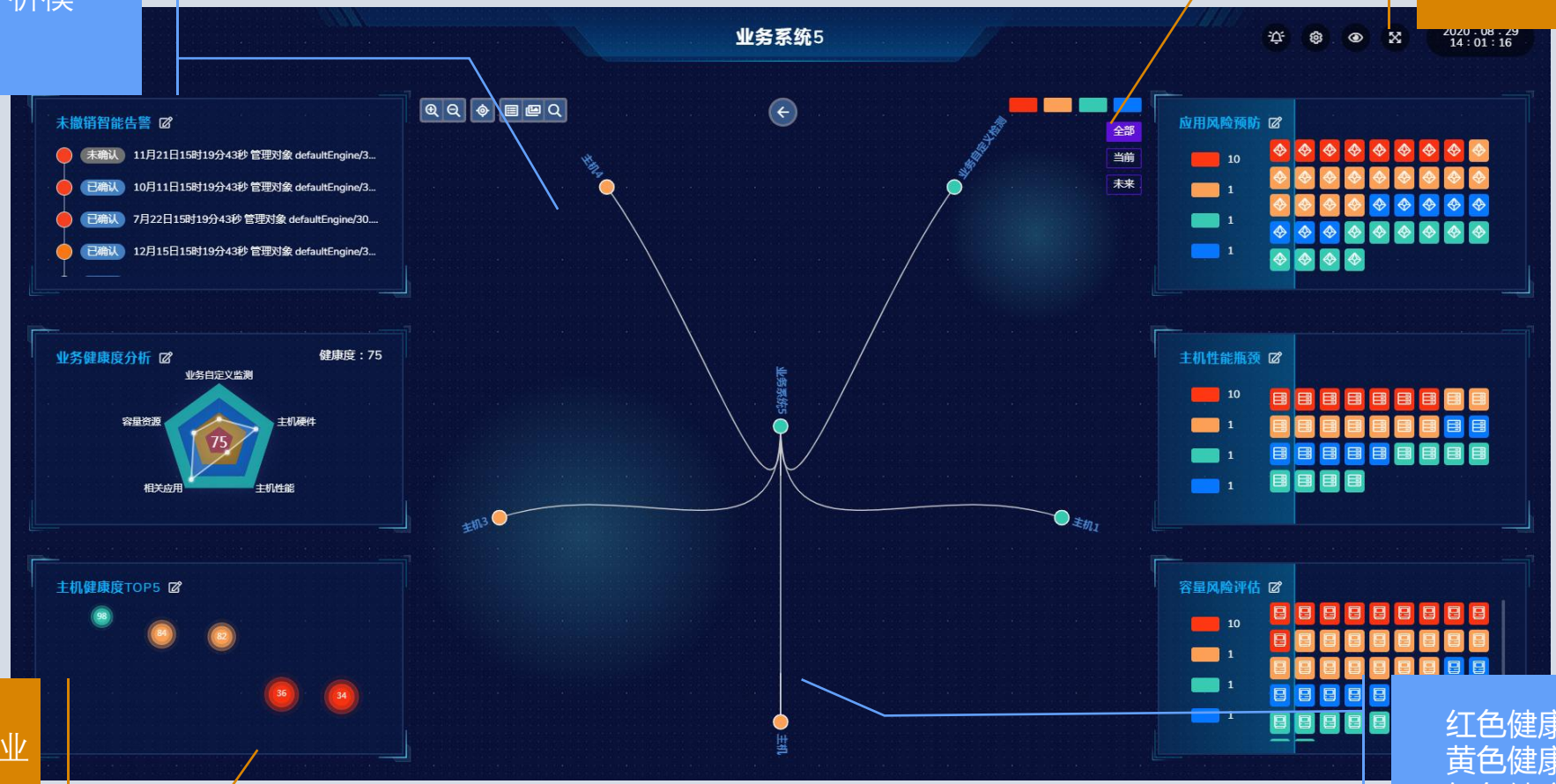
面板轮播方式显示关注点

红色健康度较差
黄色健康度中等
绿色健康度良好

3.1.6 全局态势感知：当前、未来健康感知----业务系统健康感知

全面呈现：业务健康度，并可展开分析模型查看风险点。

切换显示当前/未来业务系统状况



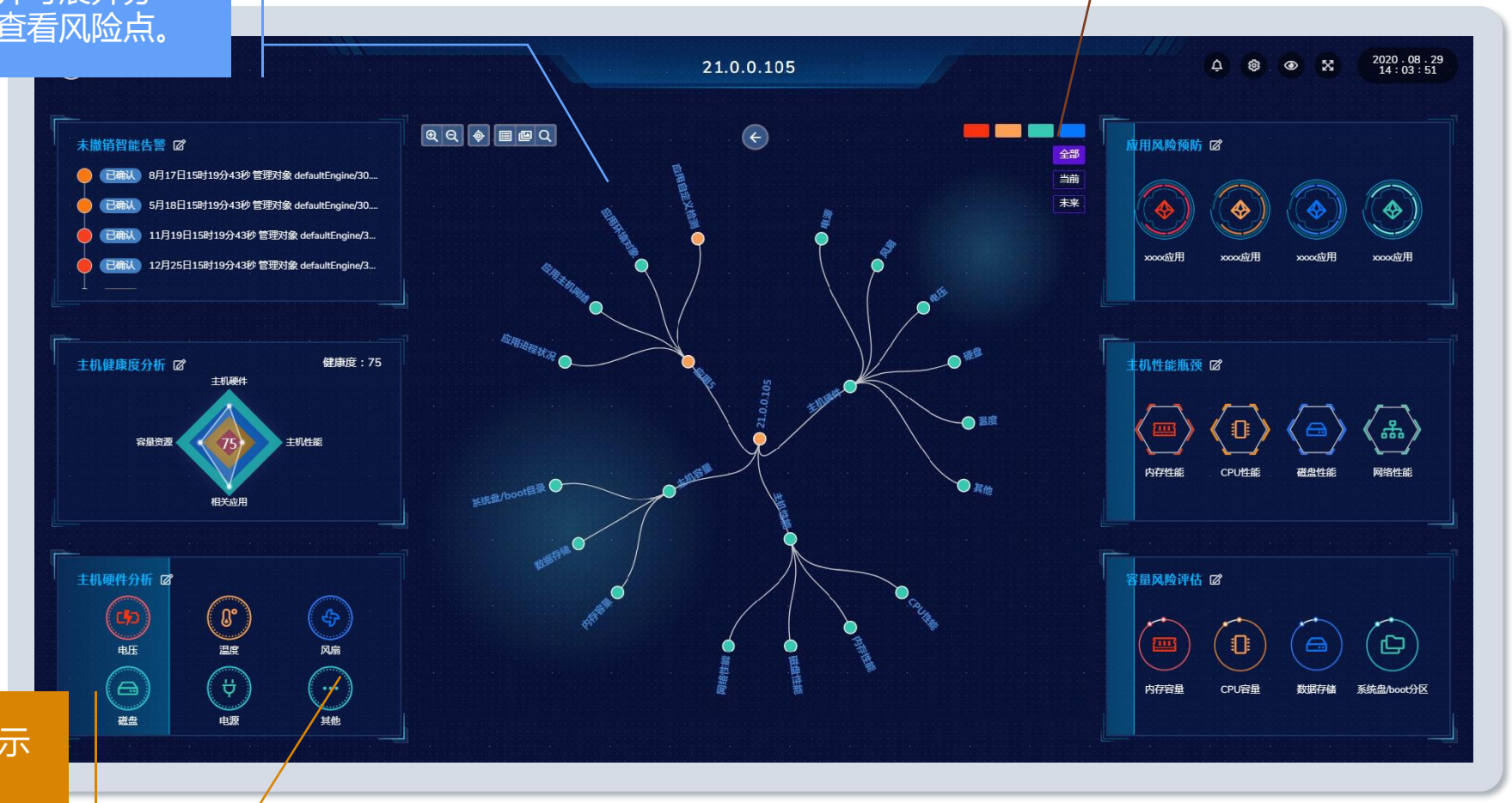
面板轮播方式显示业务系统关注点。

红色健康度较差
黄色健康度中等
绿色健康度良好

全局态势感知：当前、未来健康感知----单台主机健康感知

全面呈现：主机健康度，并可展开分析模型查看风险点。

切换显示当前/未来主机状况



面板轮播方式显示主机关注点。

典型功能二 ——

趋势预测：异常检测、辅助决策

趋势预测：异常检测、辅助决策

单台主机健康感知

模式识别及判定

趋势预测

异常检测

辅助决策

辅助根因定位

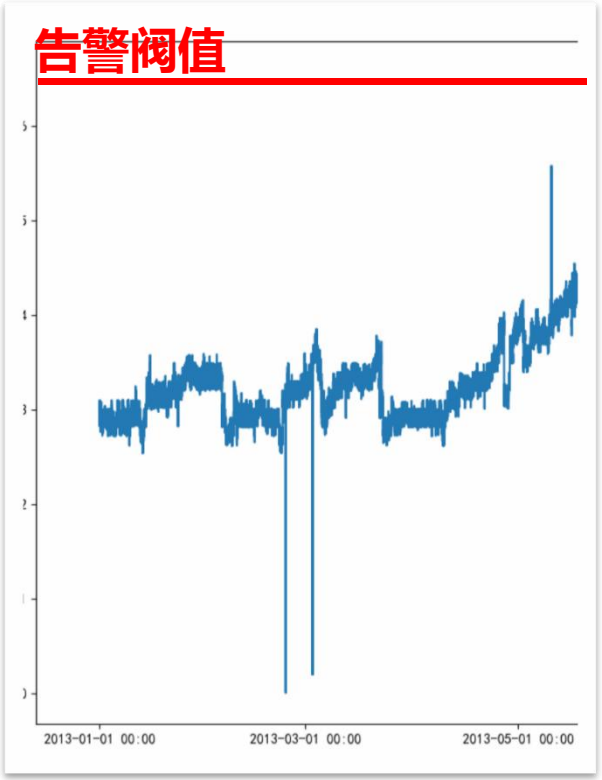
典型功能二 —— 场景1

趋势预测：用于异常检测

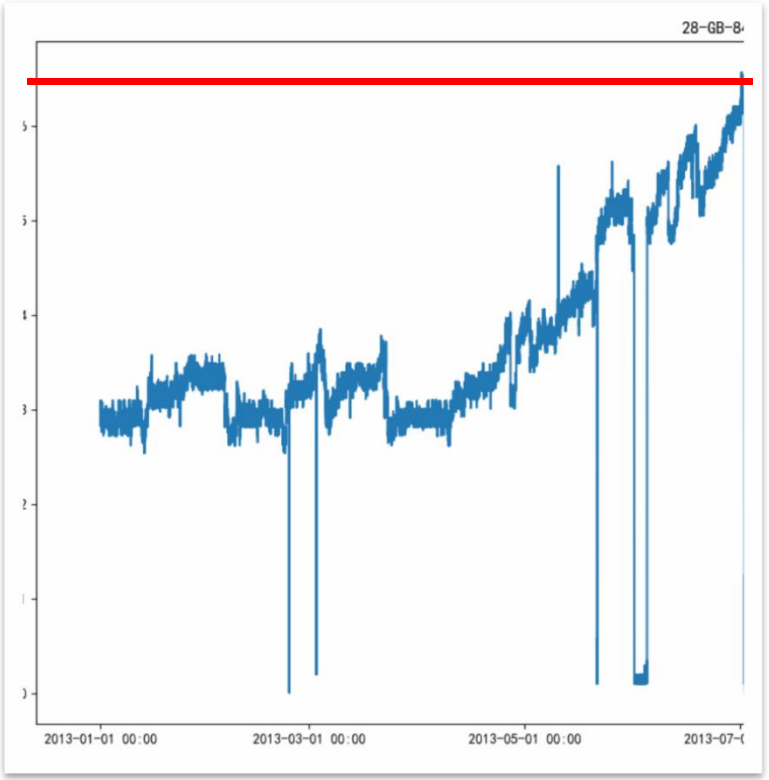
场景1：异常检测

我有监控系统，设置了很多阈值告警， 没有告警就真的太平了吗？

- 有问题吗？应该正常的吧



- 哦，有问题了，但一个月过去了



- 紧急排障

- 已经检修了，应该正常的吧？

- 心惊胆颤，如履薄冰

场景1：异常检测

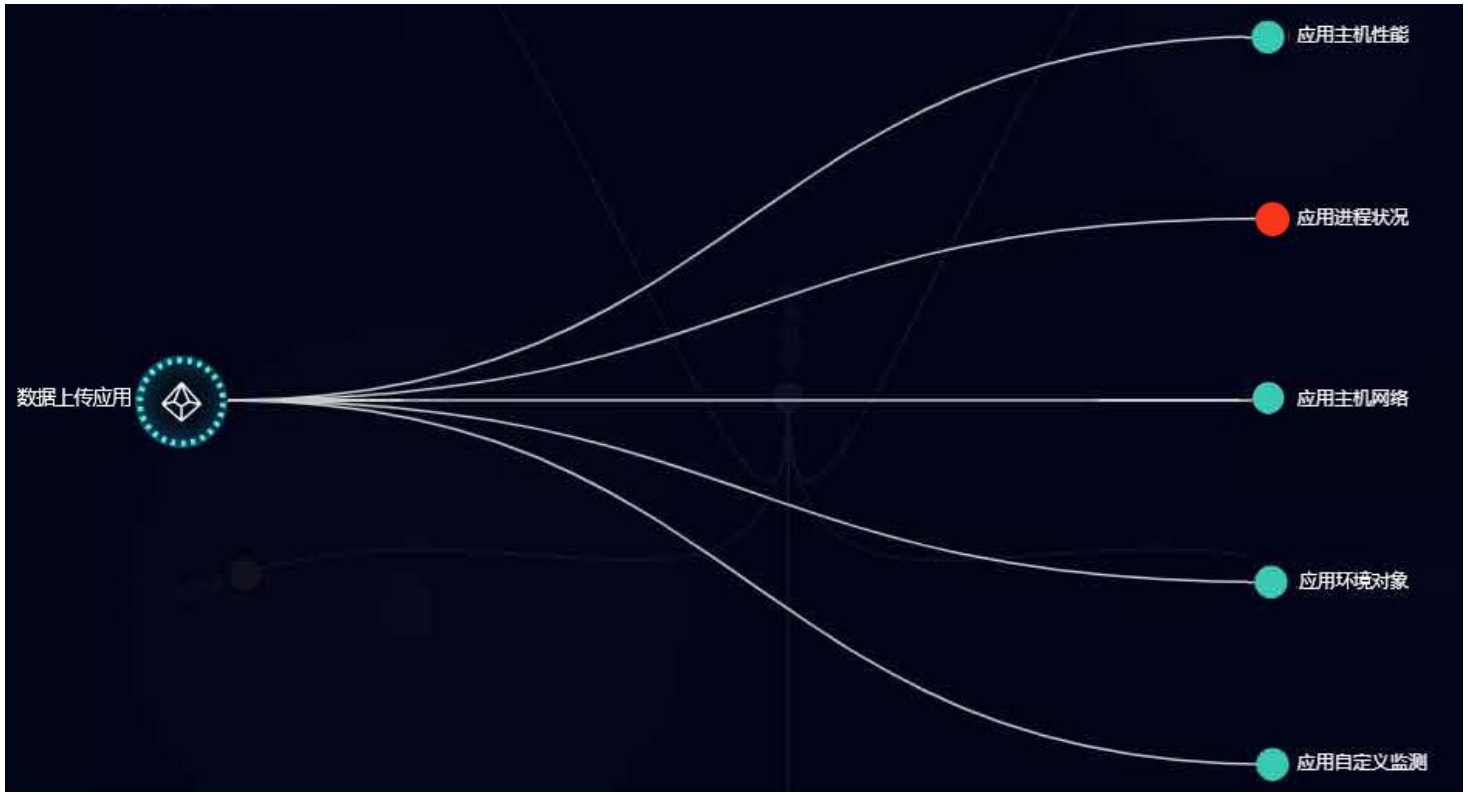
一个真实的“宕机”事件

发现问题

客户某业务与上级单位无法正常通信，硬件部分未见明显异常，操作系统还在运行，但操作响应非常慢，很卡顿。为尽快恢复数据上传，通常重启服务器，重启服务器后，不规则时间内故障重现。

分析问题

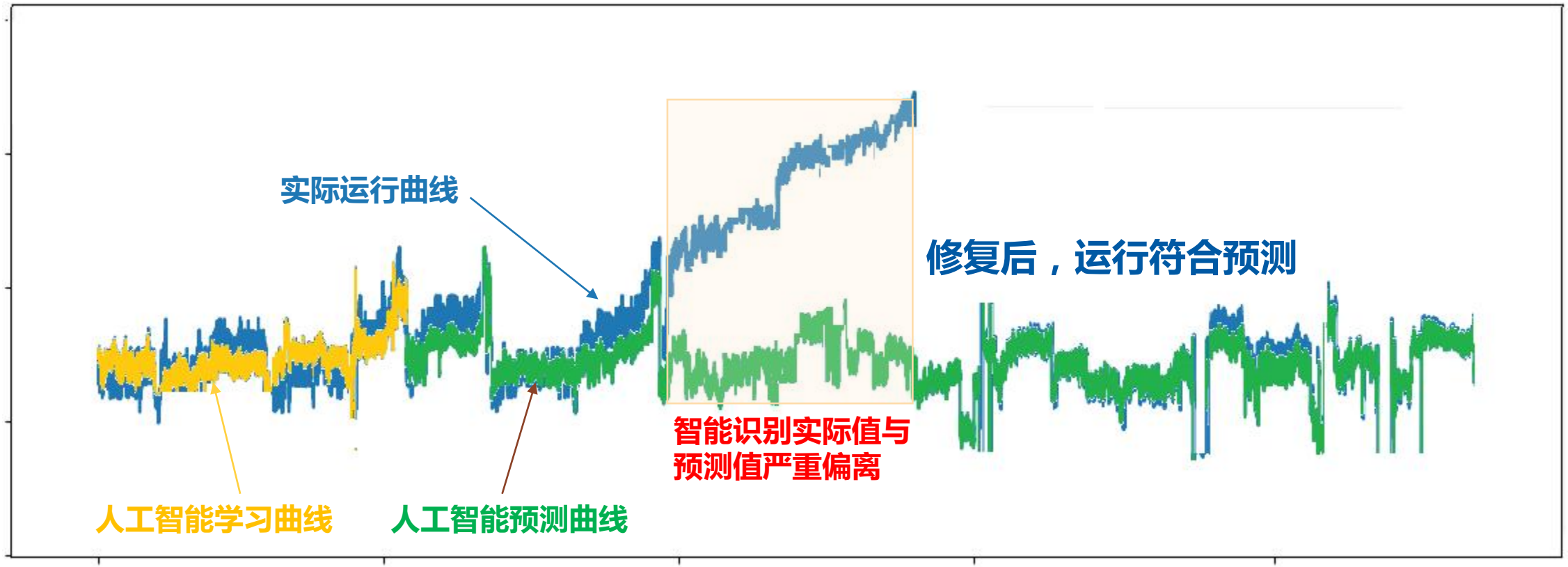
将“数据上传应用”加入应用分析模型，对应用主机性能、应用进程状况、应用主机网络、应用环境对象进行全面的智能算法分析，并加入需要自定义的监测项，发现应用进程状况呈现“红色预警”。



场景1：异常检测

发现句柄数趋势异常

定位问题　　该应用进程句柄数实际值与预测值发生了严重趋势背离，确定故障是句柄数过多引起。



场景1：异常检测

句柄数异常导致应用宕掉

解决问题

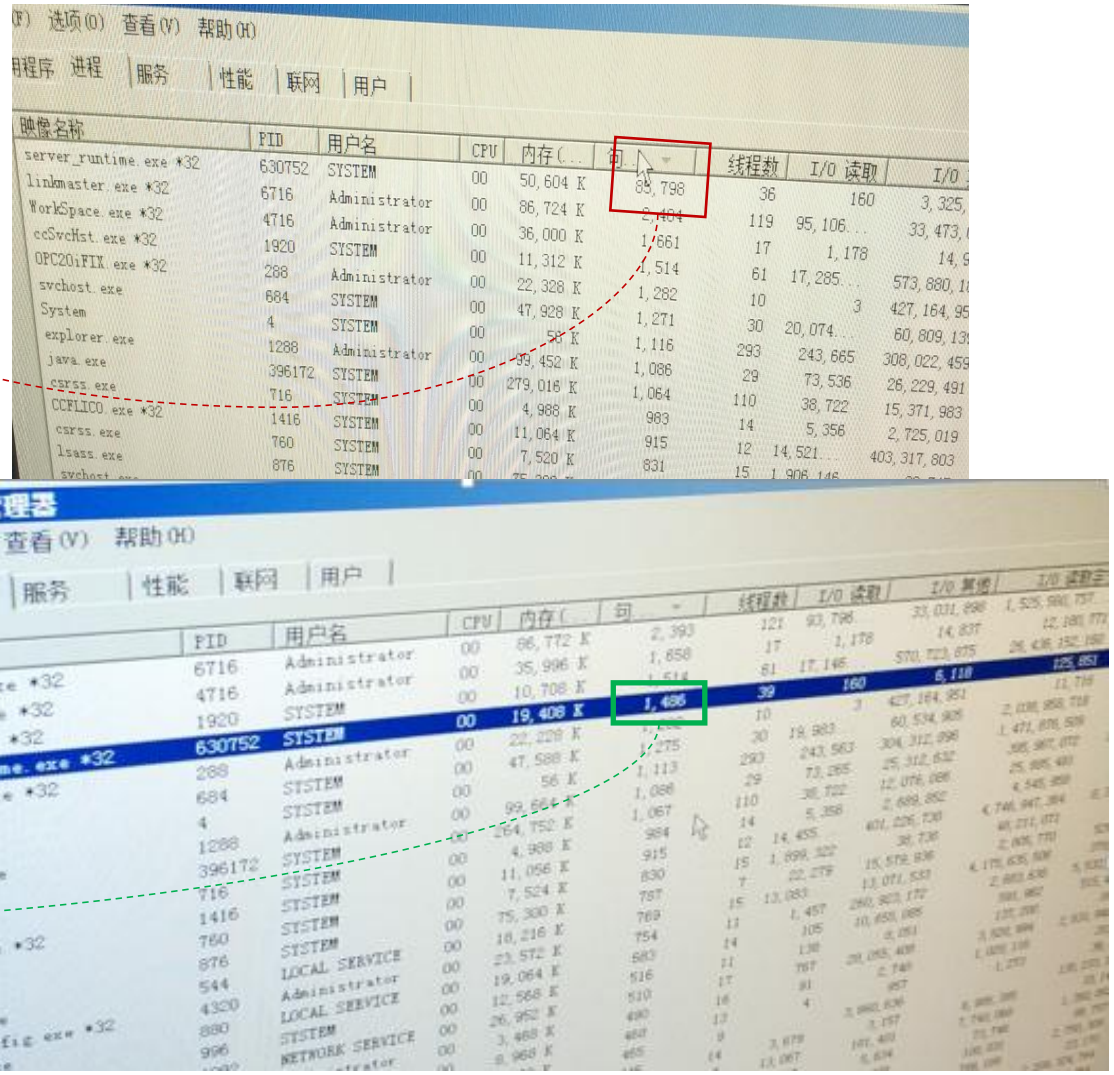
该业务系统有守护机制，主动自检，当发现服务进程异常，会自动启动服务进程，但在这过程中出现了问题，旧的资源没有释放，新的又起来，导致句柄数持续上升，最终业务宕掉。

故障发生时句柄数

85,798

修复运行后句柄数

1,486

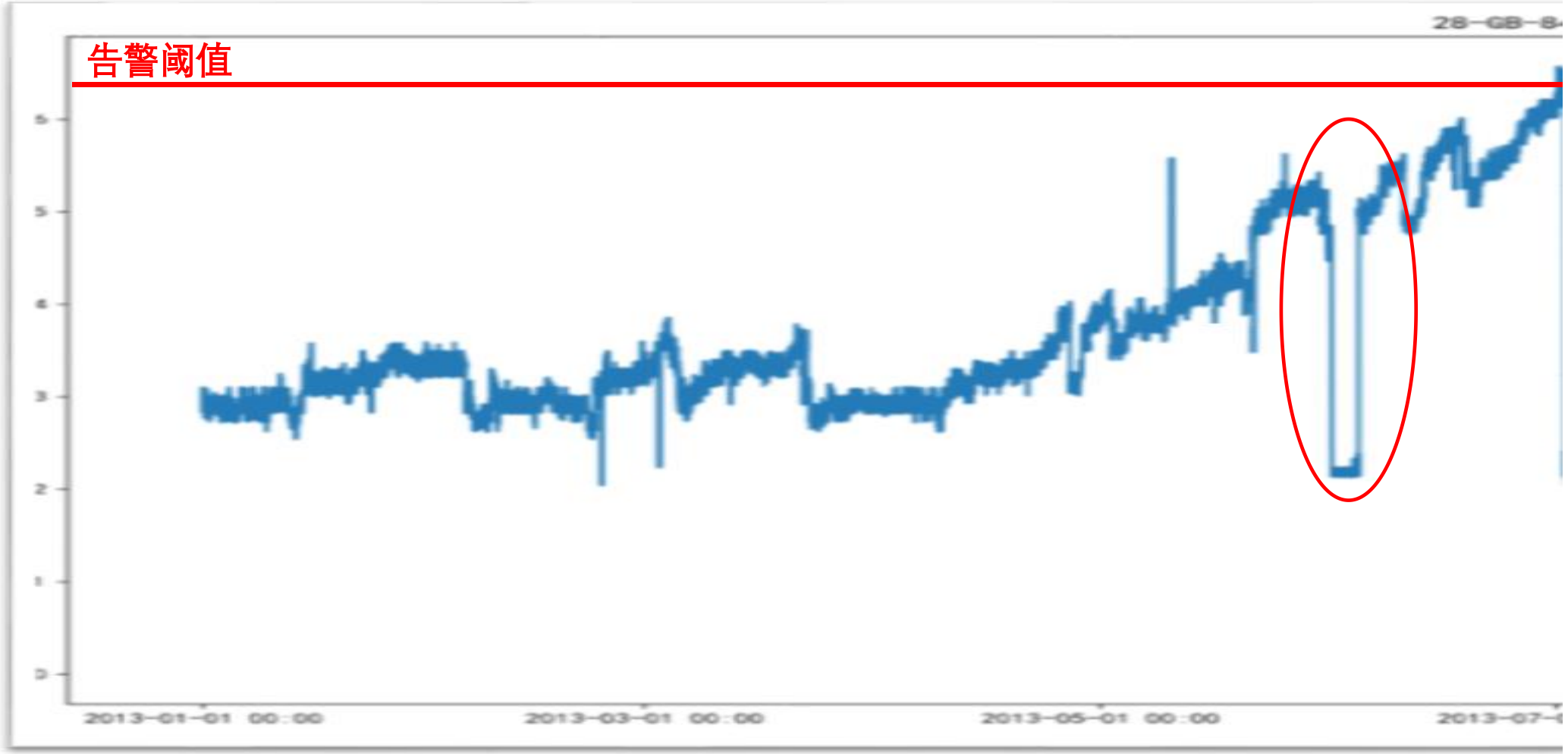


典型功能二 ——场景2

趋势预测：用于异常波动

3.2.2 场景2：异常波动

- 正常情况下短期的波动幅度较小，突然性的下跌或上涨容易被忽略；很有可能应用此时停止了运行。



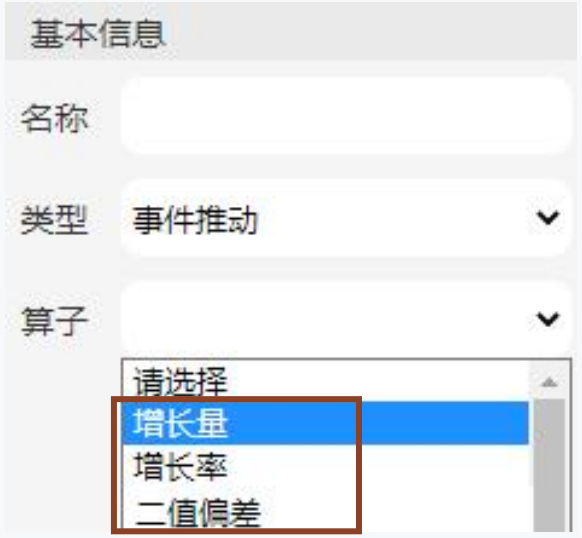
场景2：异常波动

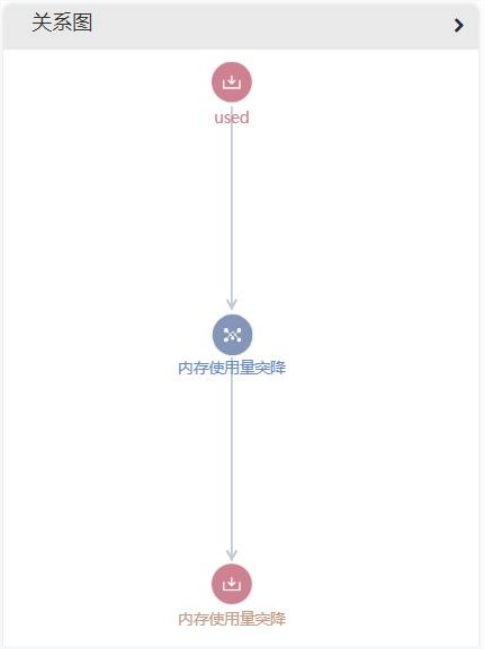
异常波动判定业务运行异常

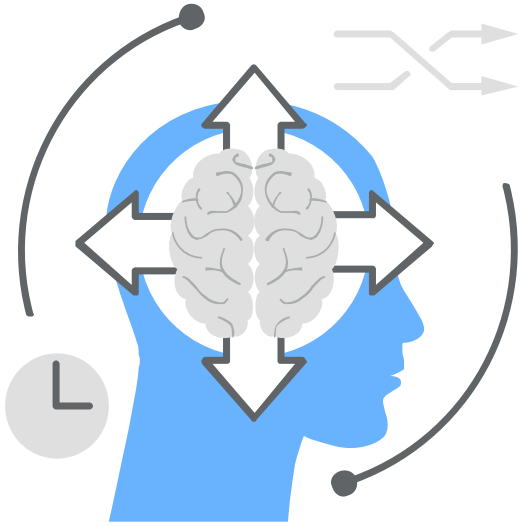
- 应用系统运行稳定后，通常一些系统指标变化也趋于稳定，如：内存的使用量。

  通过增长量、增长率、二值偏差算法智能识别突增、突降异常变化现象，预防以后更大故障发生；如：内存使用量突然下降很多，后来又恢复正常，经查是业务系统应用几个进程自动重启，需排查重启原因，避免以后更大故障发生。

三种智能算法 → 创建计算规则 → 实时监测预防







典型功能二 ——场景3

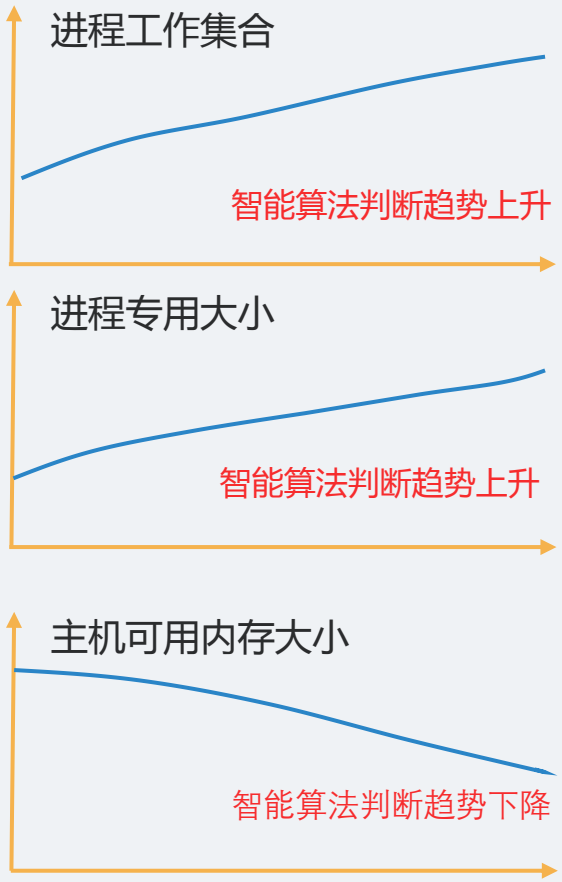
趋势预测：用于早期风险防范

场景3：风险防范

早期风险防范：内存泄露预防

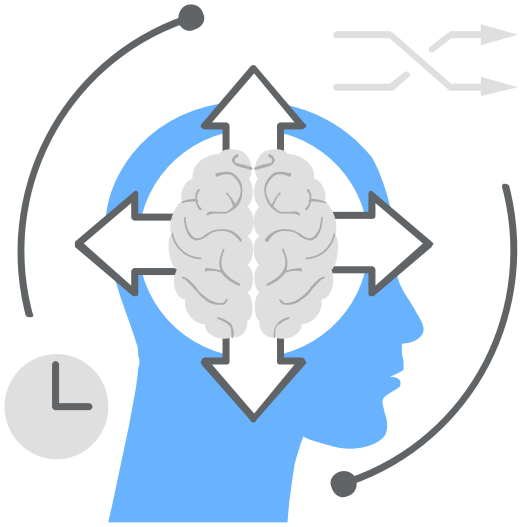
对内存相关的多指标进行趋势预测，可早期判断内存泄漏。

同一组数据，有不同的结果



内存泄漏预警

早期风险防范：内存泄漏预防



典型功能二 ——场景4

趋势预测：用于辅助决策

场景4：辅助决策

- 目前业务状况下，主机性能有没有瓶颈
- CPU、内存、数据存储这些容量资源够用吗
- 将来1个月、3个月够用吗？将来1年需要增加多少

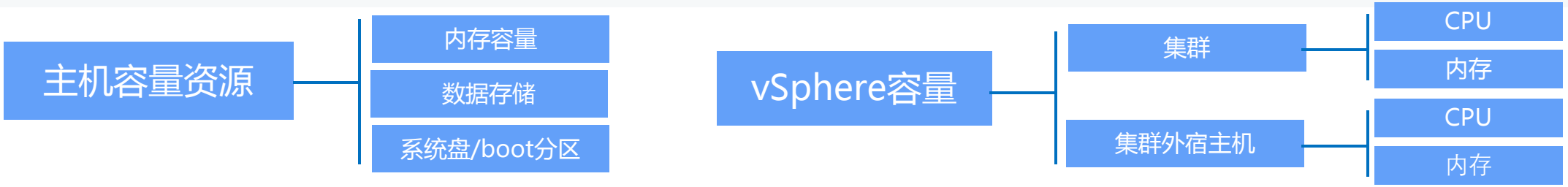
这些一无所知，无法做好IT资源规划

3.2.4

场景4：辅助决策

预置智能算法，通过对容量使用历史数据的学习，从当前和将来两个时间角度，直接显示容量分析结果：

- 当前时间角度告知容量目前使用情况；
- 将来时间角度告知容量还能使用多长时间，同时告知后续各段时间所消耗的容量大小，极大方便用户有效做好容量的预防工作以及规划工作。

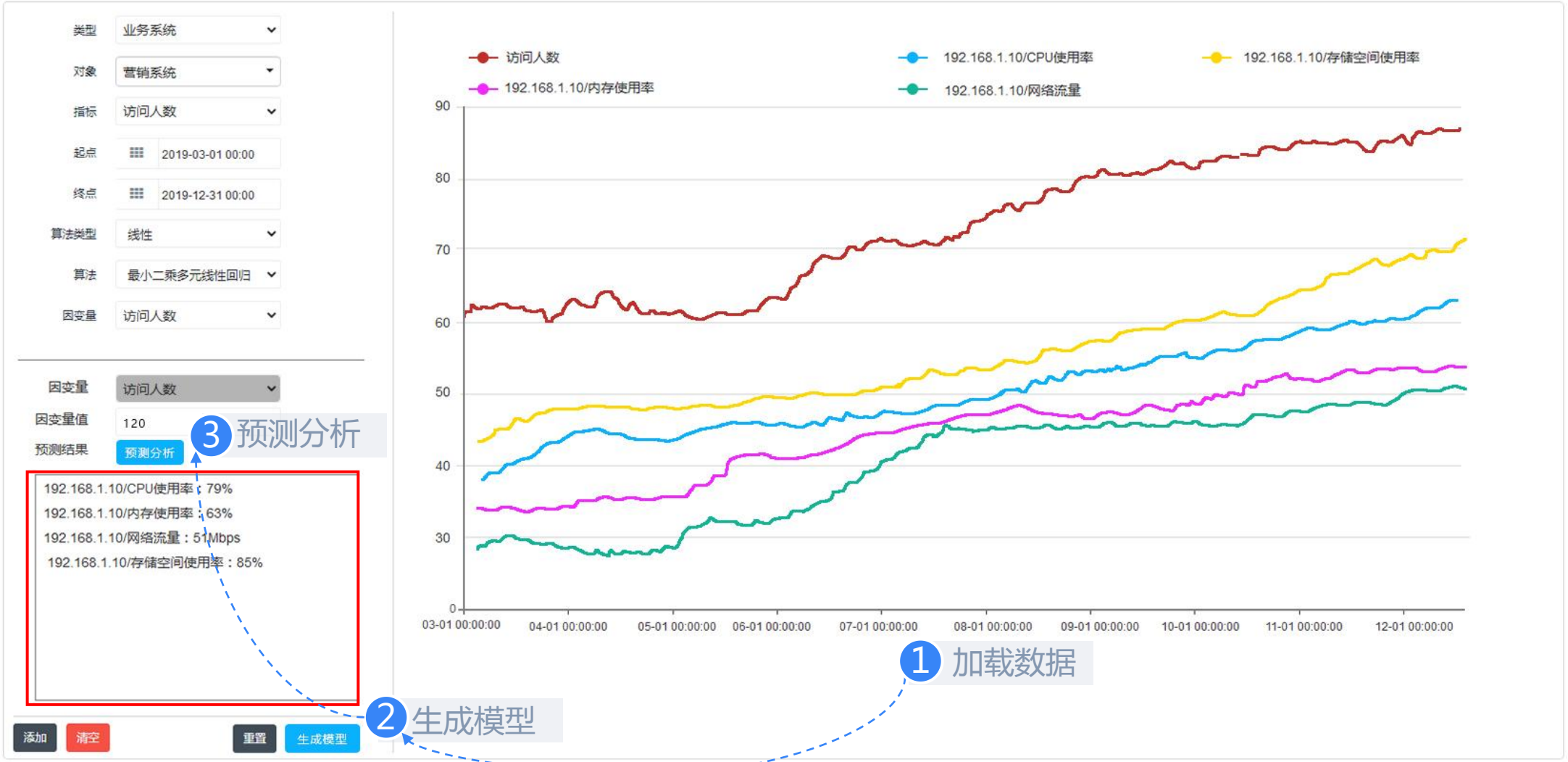


名称	所属虚拟中心	类型	总CPU(GHz)	平均可使用CPU(GHz)	平均CPU使用率(%)	预计可使用时间	将来1个月CPU使用(GHz)	将来2个月CPU使用(GHz)	将来3个月CPU使用(GHz)	将来1年CPU使用(GHz)
30.0.1.80	虚拟中心二	宿主机	19.95	1.93	90.33	<1个月	20.12	22.22	24.32	43.57
30.0.1.81	虚拟中心二	宿主机	19.95	9.6	51.88	>=1年	10.95	11.55	12.15	17.65
30.0.1.82	虚拟中心二	宿主机	19.95	10.37	48.03	>=1年	9.94	10.3	10.66	13.96
30.0.1.83	虚拟中心二	宿主机	21.33	2.19	89.74	>=1个月	20.24	21.34	22.44	32.5
30.0.1.84	虚拟中心二	宿主机	21.33	3.22	84.91	>=2个月	19.19	20.26	21.34	31.18
30.0.1.85	虚拟中心二	宿主机	21.33	6.57	69.2	>=3个月	15.36	15.96	16.56	22.06
集群一	虚拟中心一	集群	85.31	7.56	91.14	>=1个月	81.56	85.37	89.18	124.11
集群二	虚拟中心一	集群	85.31	56.11	34.23	>=1年	30.1	31	31.9	40.15

3.2.4 场景4：辅助决策

根据智能算法，通过对主机或业务系统关键指标的学习，预测主机或业务将来使用情况：

案例：客户某业务系统上线9个月后，访问人数越来越多，预计将达到120人左右，主要资源使用会怎样？需要做提前预防？

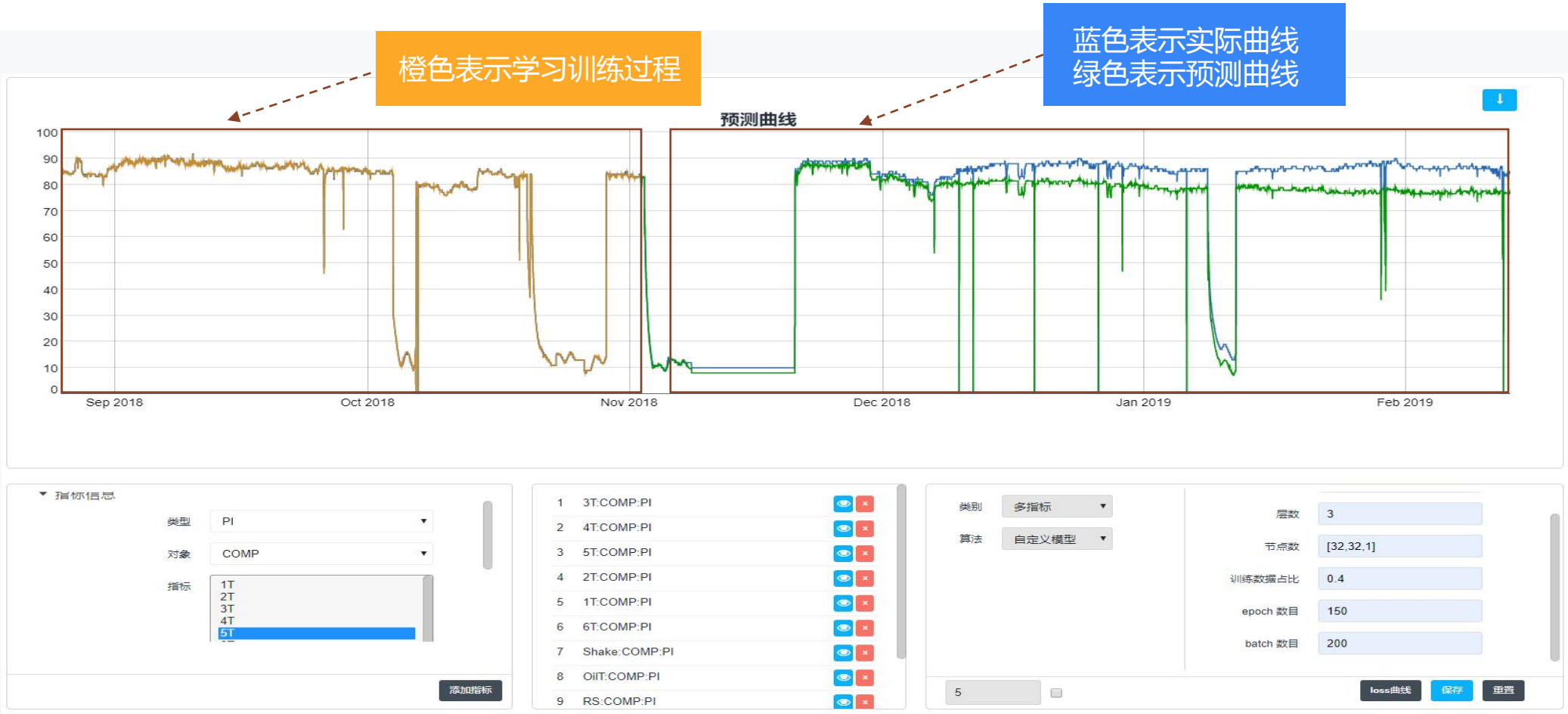


典型功能三——

场景化分析模型、辅助根因定位

人工智能算法——单指标、多指标建模预判，提前发现隐患

通过对历史运行情况的深度学习，图形化方式数字建模，对未来趋势进行预测，能够精准的预判风险点，较传统告警方式大大提前，同时也能够抑制众多无效告警。



算法模型选优——损失值、预测损失值、偏离值

关键参数越小越好

算法名称	损失值	预测损失值	偏离值
单指标神经网络	0.094795	0.0792191003673168	30.49309870822382
单指标神经网络	0.450422	0.33440034979311845	0.2964656776128223
单指标神经网络	0.065881	0.07809035537134895	30.844518820823836
单指标神经网络	0.074952	0.0844675170050727	31.755335328471524
多指标神经网络	0.000012	0.000006801367918...	0.022016431331634...
多指标神经网络	0.342355	0.3653975857628716	-0.14990858409424...

智能算法——内置了精选的45种常用智能算法

平台算法库储备了**超过2000种**算法，根据场景需要，可通过**导入**算法方式实现**动态添加**。

	名称	类型	分类	描述	操作
1	单指标神经网络	流式处理算子	nn	深度学习算法	详情
2	多指标神经网络	流式处理算子	nn	深度学习算法	详情
3	趋势系数	批处理算子	trend	机器学习算法，常用于趋势分析。	详情
4	肯德尔相关性系数	批处理算子	correlation	相关性算法，常用于指标间相关度分析。	详情
5	斯皮尔曼相关性系数	批处理算子	correlation	相关性算法，常用于指标间相关度分析。	详情
6	皮尔逊相关性系数	批处理算子	correlation	相关性算法，常用于指标间相关度分析。	详情
7	最优相关性系数	批处理算子	correlation	分别调用相关性算法分析，选择相关度最高值。	详情
8	区间聚合	批处理算子	stats	机器学习算法。	详情
9	神经网络预测	批处理算子	forecast	深度学习算法。	详情
10	一阶指数平滑预测	批处理算子	forecast	机器学习算法。	详情
11	Holt-Winter季节性模型预测	批处理算子	forecast	机器学习算法，常用于趋势分析。	详情
12	二阶指数平滑预测	批处理算子	forecast	机器学习算法。	详情
13	加权移动平均模型预测	批处理算子	forecast	机器学习算法。	详情
14	增长率模型预测	批处理算子	forecast	机器学习算法，常用于趋势分析。	详情
15	天真模型预测	批处理算子	forecast	机器学习算法。	详情
16	移动平均模型预测	批处理算子	forecast	机器学习算法。	详情
17	回归分析	流式处理算子	regression	机器学习算法。	详情
18	延迟算子	流式处理算子	utils	机器学习算法。	详情

« 1 2 3 » 第 1/3 页 第 1 ~ 18 项 共 45 项

场景1：应用假死

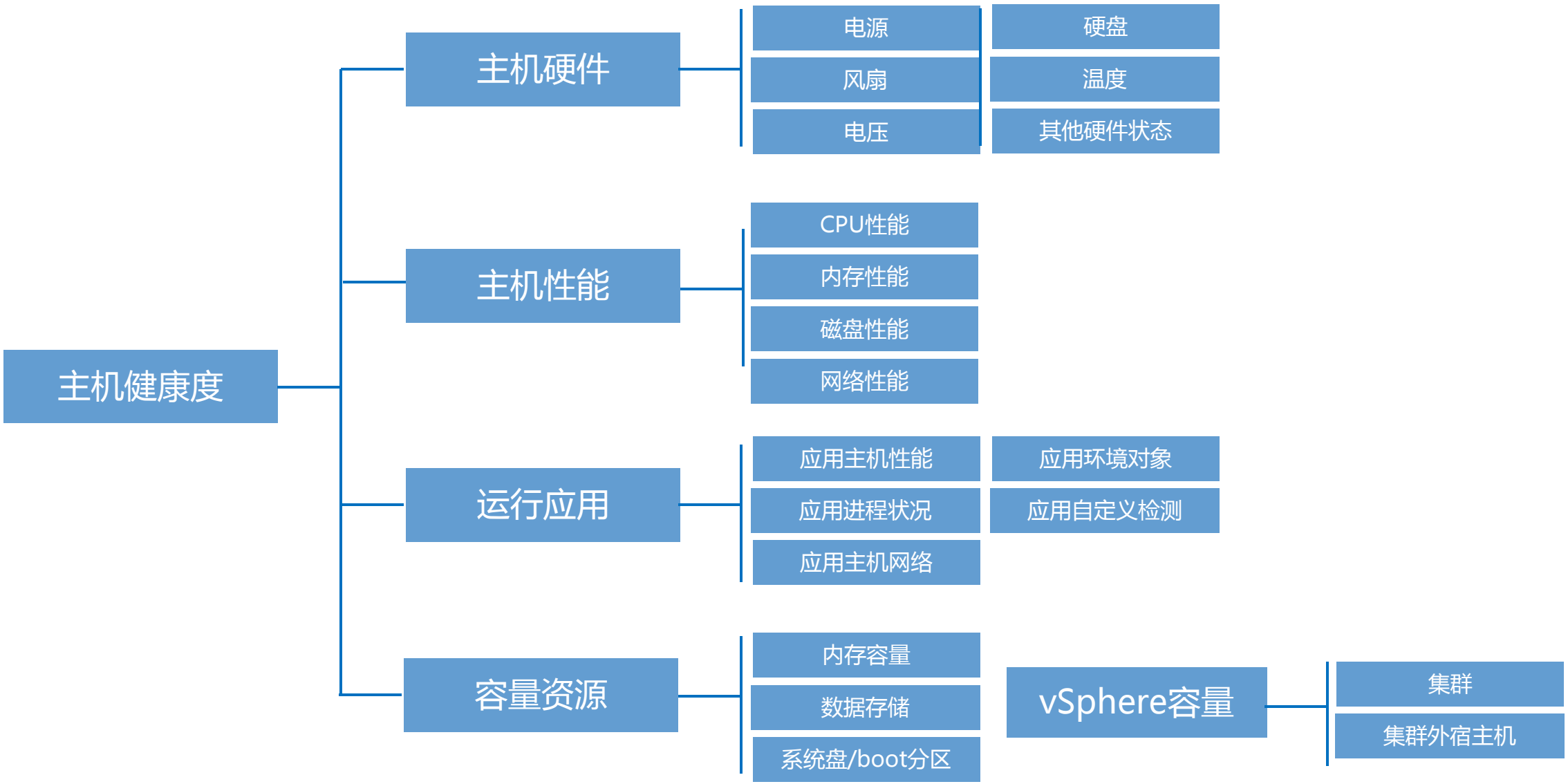
某应用无法访问，究竟是中间件服务问题、操作系统问题、还是网络层面问题？

客户端没有响应	应用WEB无法访问
中间件进程存活	Tomcat服务问题？
操作系统运行正常	操作系统问题？
网络连接正常	网络层面问题？

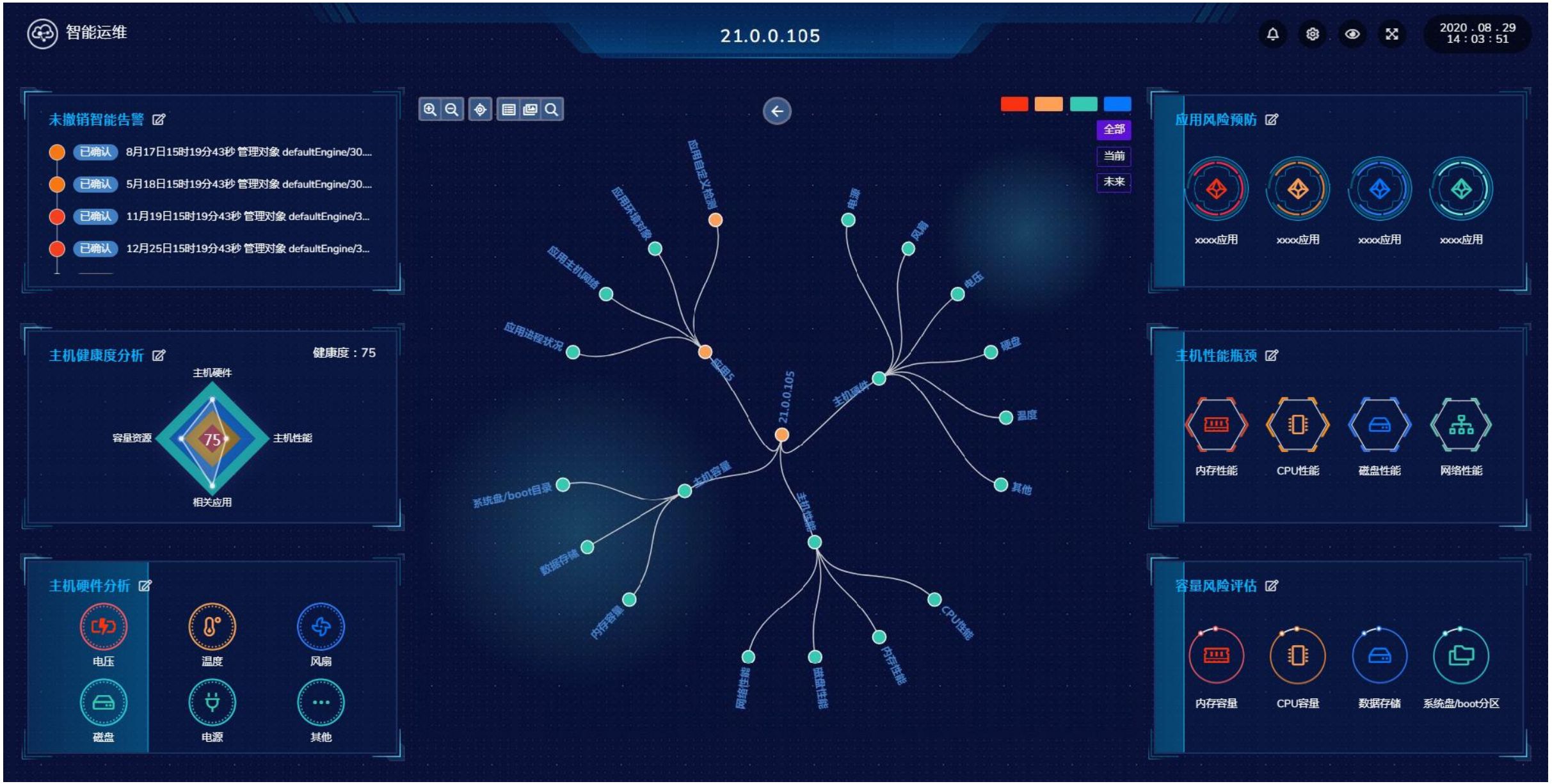
主机模型构建

主机在应用系统中处于核心位置：下能感知网络情况，上能感知应用情况！

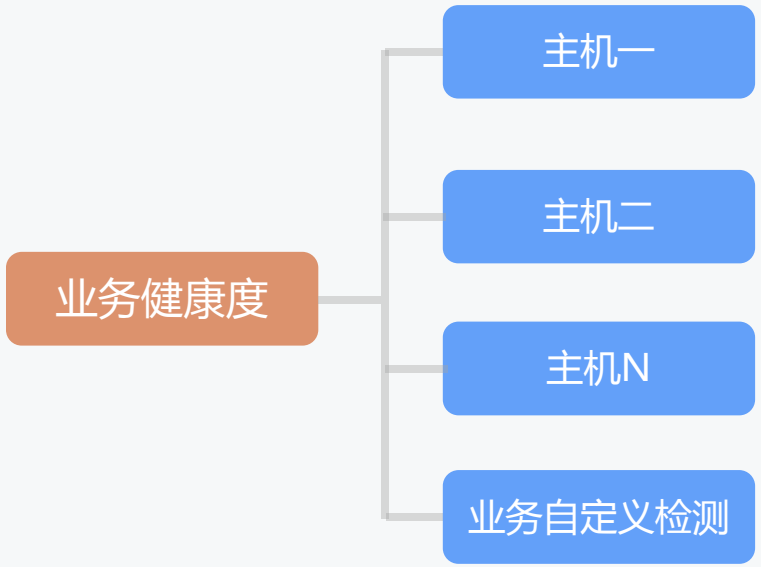
场景化分析模型

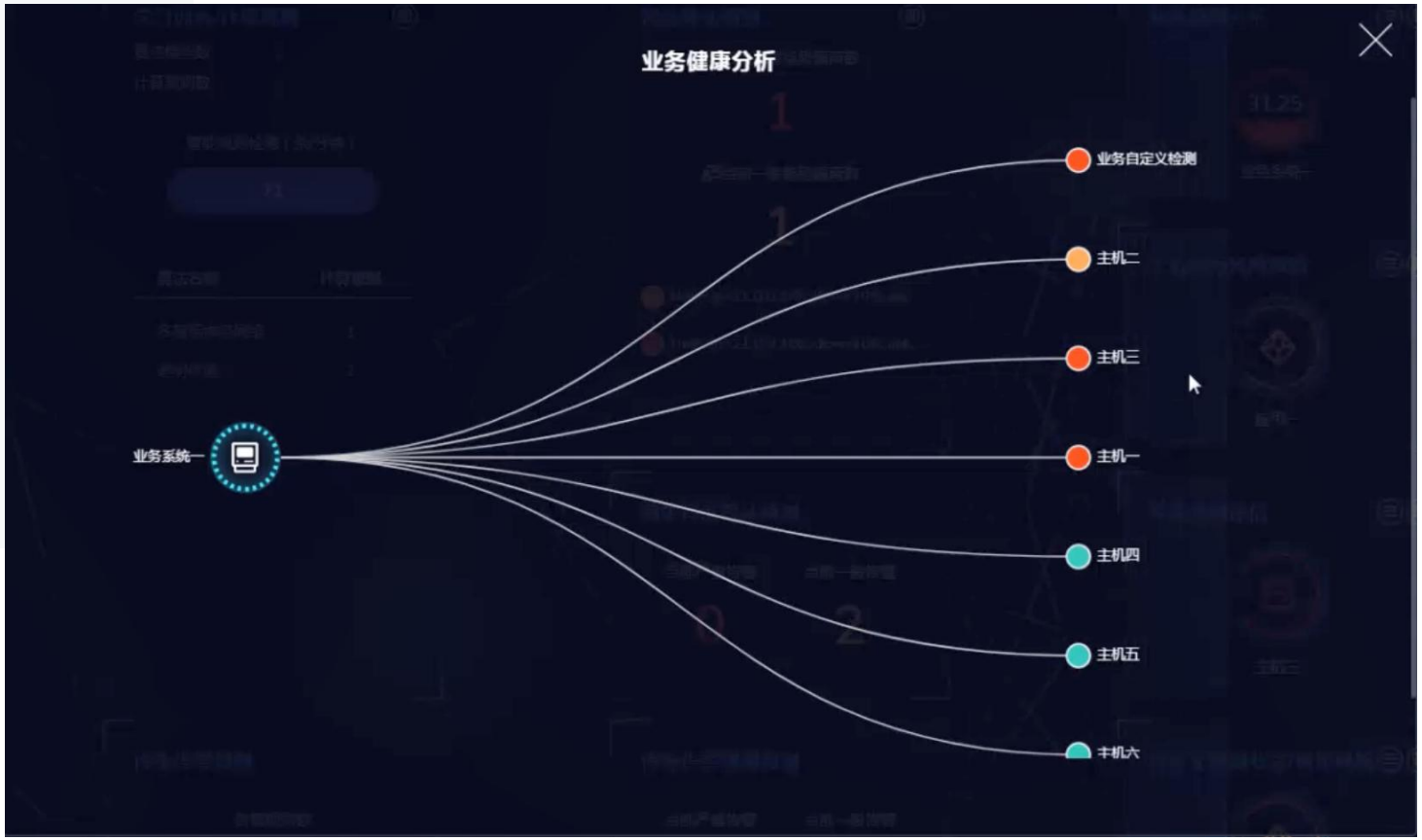

主机健康度
主机硬件
电源
风扇
电压
硬盘
温度
其他硬件状态
主机性能
CPU性能
内存性能
磁盘性能
网络性能
运行应用
应用主机性能
应用进程状况
应用主机网络
应用环境对象
应用自定义检测
容量资源
内存容量
数据存储
系统盘/boot分区
vSphere容量
集群
集群外宿主机

3.3.1 场景化分析模型界面

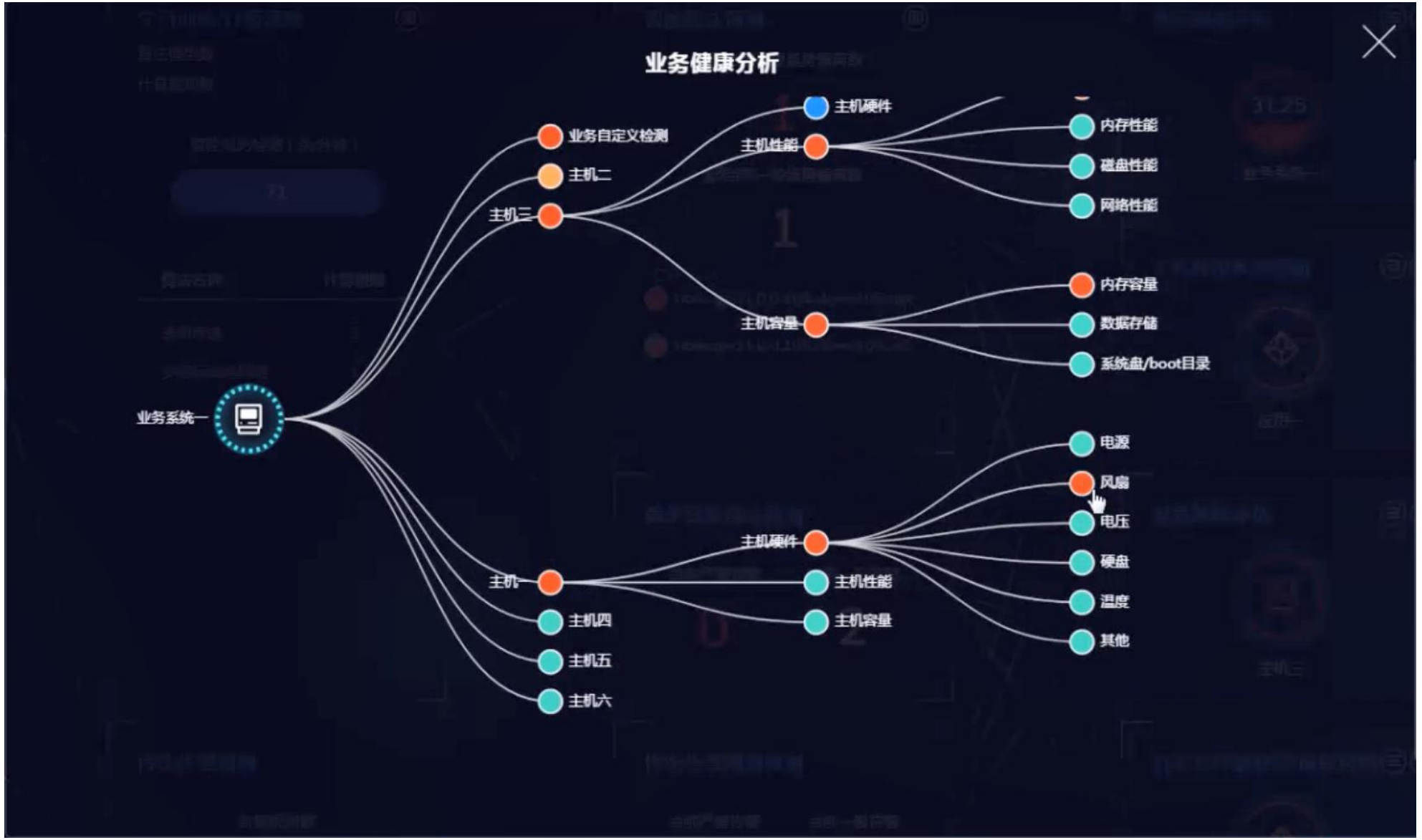


3.3.1 场景化分析模型


业务健康度
主机一
主机二
主机N
业务自定义检测


业务健康分析
业务系统一
业务自定义检测
主机二
主机三
主机一
主机四
主机五
主机六

3.3.1 场景化分析模型界面

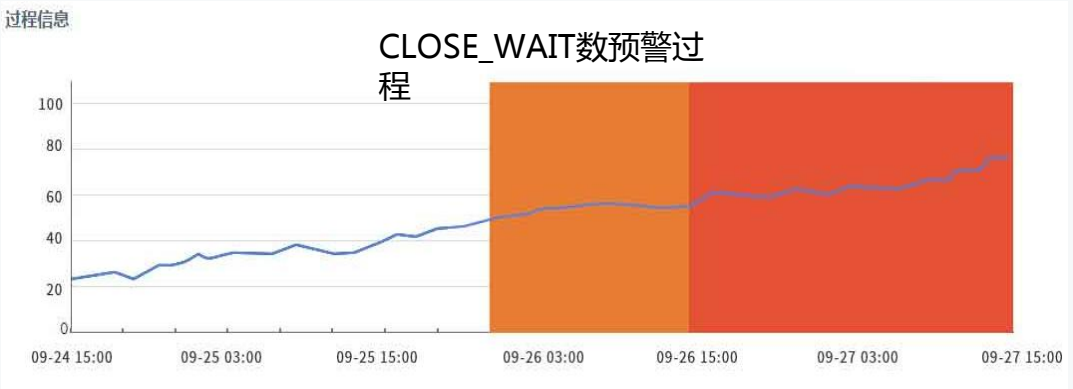


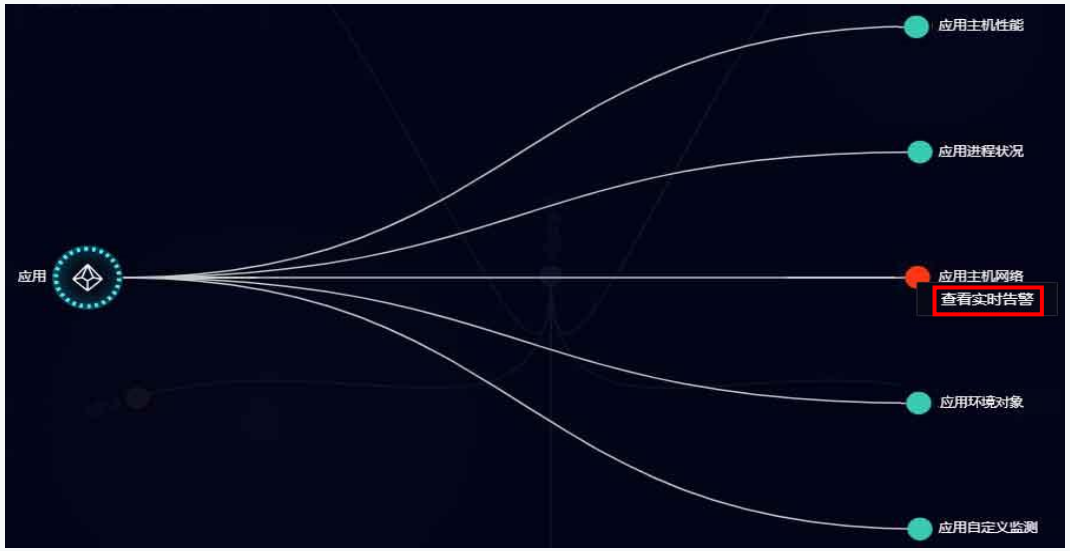
应用假死分析

场景化分析模型—>辅助根因定位

- 根据应用分析模型快速定位根源

客户端没有响应	应用WEB无法访问
进程存活，telnet端口没有回应	Tomcat服务假死
JVM环境	JVM性能正常
主机CPU、内存、磁盘、流量/丢包	主机性能正常
大量TCP 连接 **CLOSE_WAIT** 应用程序关闭socket不及时	应用网络表现异常





场景2：关联指标判定

- 同时触发多个指标告警，指标间是否有关联？采集指标很多，是否与未设置告警规则的指标也存在关联？

告警指标一

告警指标二

告警指标N

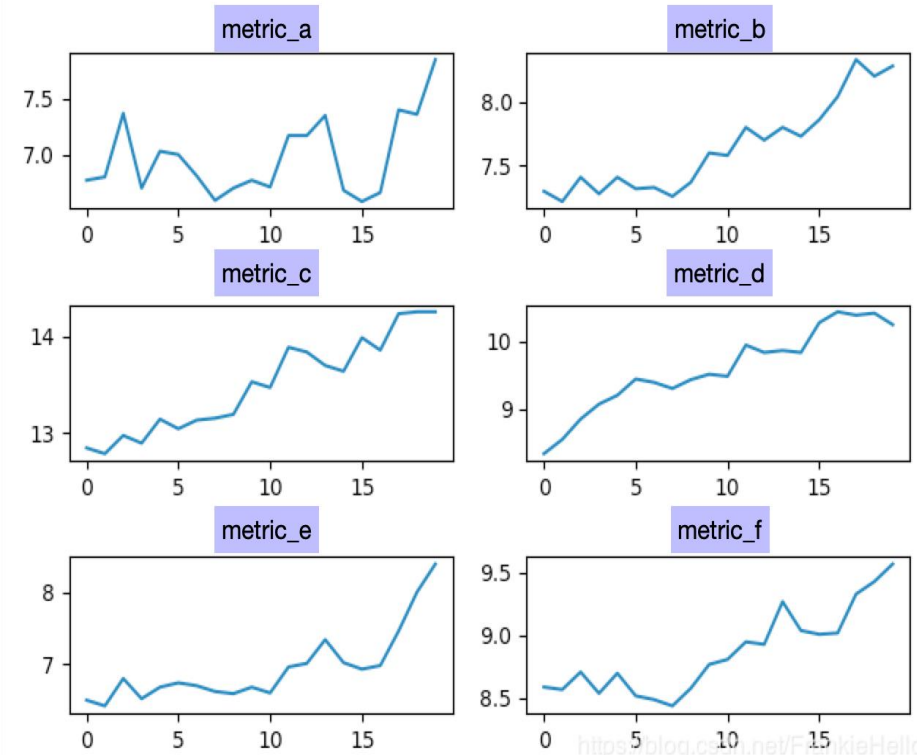
告警指标间存在关联特性？

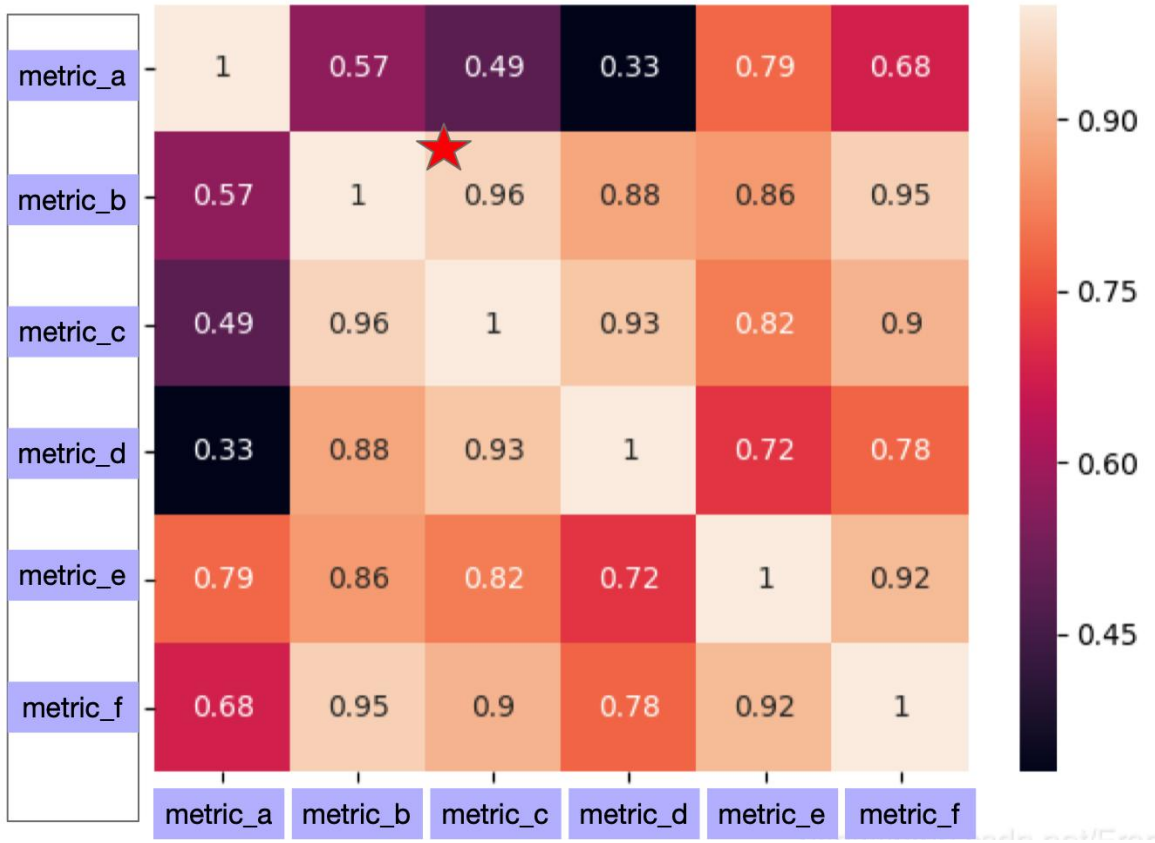
与未设置告警规则的指标有关联？

3.3.2 指标间相关性实现原理

- 指标间的关联关系，凭借自己的经验，同时对比查看关键指标的变化，这种方式既耗时又费力，也容易出错。

  借助人工智能技术来进行分析，这样可以发现或者验证经验认知中没有或已有的关联关系，极大提高了工作效率，更便于辅助分析一些突出表现在指标类的疑难故障。





3.3.2 指标间相关性应用案例

辅助排查故障根因

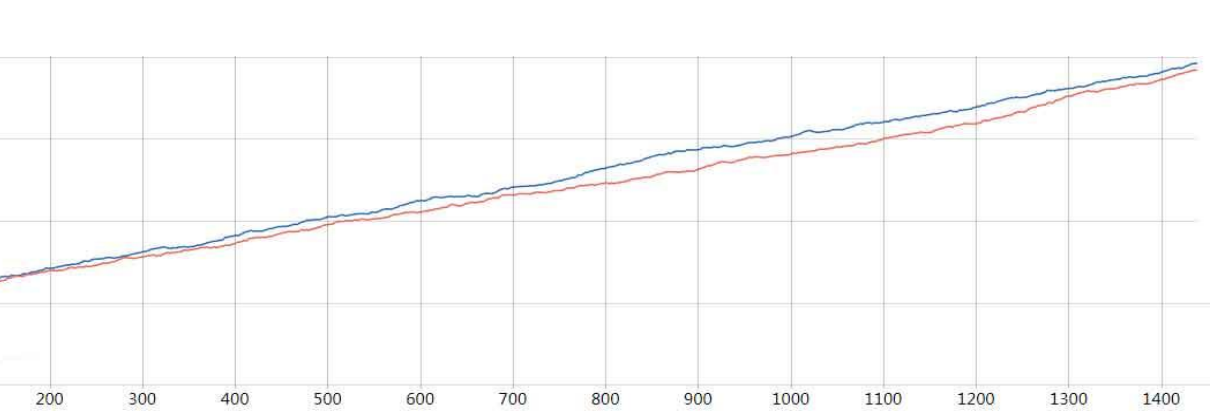
故障案例

某客户陆续收到主机CPU使用率偏高、HTTP响应时间慢、CPU温度升高、主机线程数增加等多条告警信息，如何排查？

相关性分析

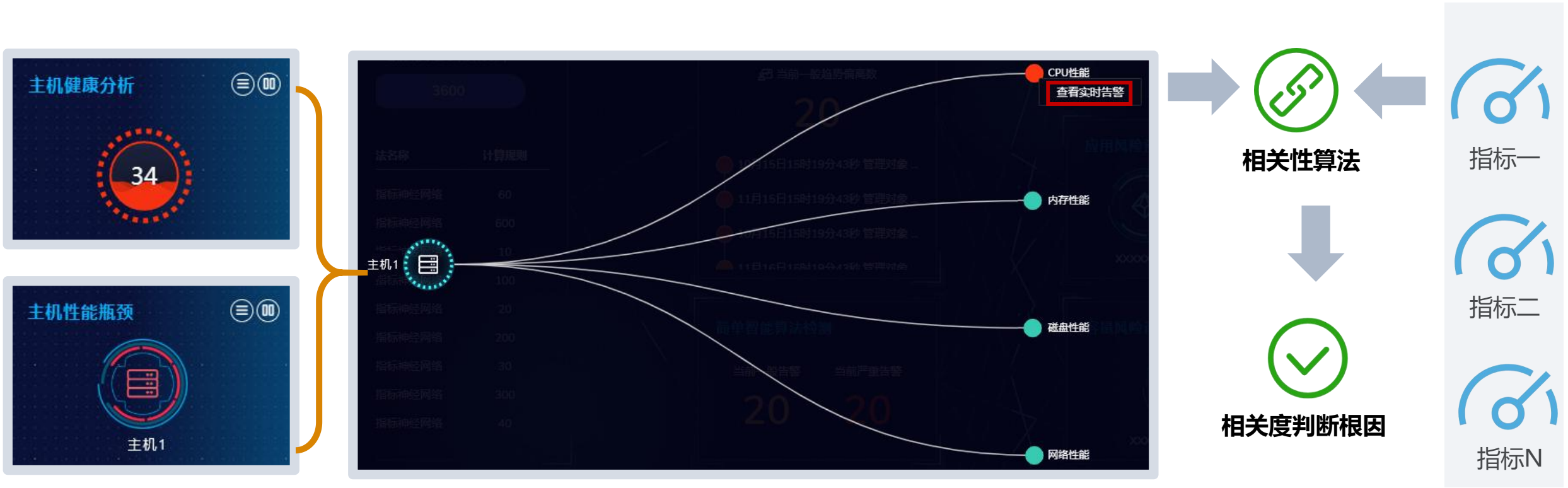
通过使用相关性智能算法分析，最终锁定是主机线程数增多导致CPU利用率偏高。

	CPU使用率	HTTP响应时间	主机线程数	内存使用率	CPU温度	磁盘IO
CPU使用率	1.000	0.778	0.992	0.055	0.780	0.637
HTTP响应时间	0.778	1.000	0.778	0.040	0.677	0.509
主机线程数	0.992	0.778	1.000	0.055	0.779	0.637
内存使用率	0.055	0.040	0.055	1.000	0.055	-0.114
CPU温度	0.780	0.677	0.779	0.055	1.000	0.543
磁盘IO	0.637	0.509	0.637	-0.114	0.543	1.000



- 智能算法判断指标间变化相似度系数。
- 从变化趋势图看，主机线程数与CPU使用率基本一致。

小结：场景化分析模型—>辅助根因定位



健康度差或触发风险

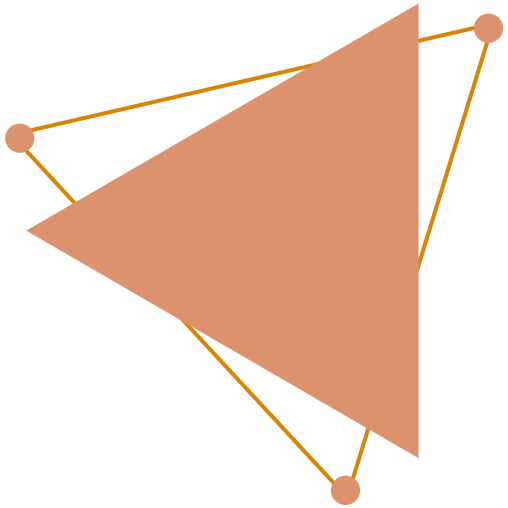
分析模型定位到红色问题点，并查看原因

如遇疑难问题
相关性算法辅助深层定位

01 我们的价值

趋势预测

通过深度机器学习，进行了模式识别，具备了预测的能力，趋势预测可用于：早期风险防范、日常异常检测、IT资源容量预测、掌握运行趋势，真正辅助决策； 比如预测主机性能、应用在未来几天是否会发生风险？发生什么样的风险？容量还能用多久？



全局态势感知

几百上千台服务器，日常是凭经验在检查，或者定期巡查，或者通过监控软件的告警再进行检查等等，但都是单机、单指标、单点的概念，缺乏总体、立体思维、缺乏相互作用、影响分析、缺乏对未来的感知等等，而我们是实现全局态势感知，几百上千台服务器，只让你看需要看的，就是与平时模式不同的，有异动的，才是需要关心的.....

辅助根因分析

在发现特征异常后，系统将自动标记出异常点，并形成直观认识的健康度，通过分析模型方式辅助分析异常根源。

总结

02 监控软件由于定位不同，均不具备上述能力，其注重“广度”，我们注重“深度”，与我们显著不同。

03 大数据分析：经验非常重要，经验涉及到：业务经验、算法经验、项目经验

业务经验

我们公司已积累20年运维经验，超过8000家用户服务经验，积累了丰富的运维经验

算法经验

多年前，即成立了大数据工作组，进行大数据技术研究

项目经验

我们公司大数据分析平台产品已在工业生产中投运，有丰富的成功大数据项目经验

这三者经验我们都结合起来了，这就是优势

Part.

4

SMARTEAGLE
慧鹰

常见问题

面临5G网络的未来发展，以及万物互联的时代即将到来，慧鹰SEHM在实现IT运维的基础上，未来还能够做点什么？

常见问题1

问

你提到了数字化转型，怎么体现出来的？

答

上线一段时间后，通过不断的数据分析挖掘，能够找出各种IT资源的运行趋势，实现全面掌控，从而为数字化转型的决策提供数据支持。

常见问题2

问

你们需要学习多久的数据？怎么提高预警的准确性和实用性？

答

学习周期

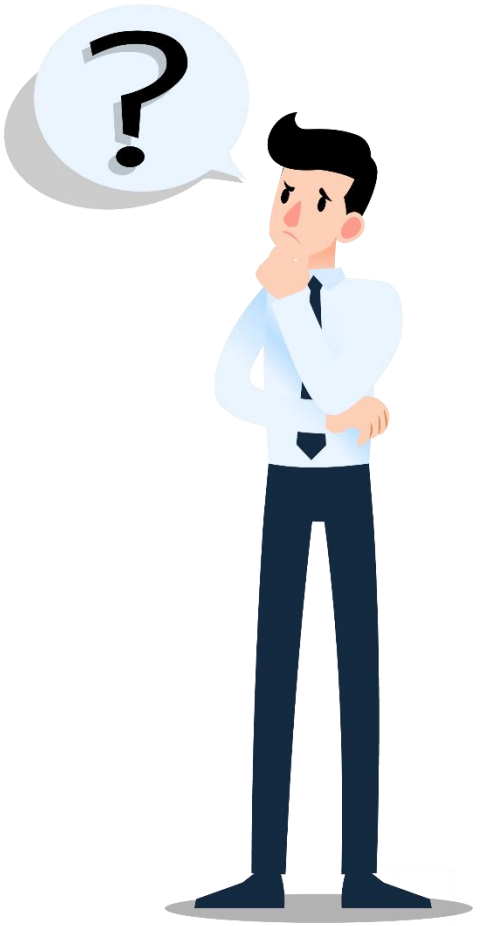
主要取决于用户该场景下的数据规律分布周期，极端的说，如果习1年的数据，以我们的工程经验来看，通常是需要学习1-6个月这个规律是每1天都在重复，那么我们学习1天的就可以了， 如果是规律是1年，那么需要学的历史数据；

准确性和实用性

首先，完整学习了规律性周期的数据，准确性本身就会非常高。

其次，为了保证预警的实用性，高价值，我们特别引入了深度学习中的高阶应用：模式识别技术，通过模式识别技术，再结合趋势预测，就大大提高了预警的价值度。

最后，模式识别、趋势预测，数据越丰富越精确，价值越高，所以，随着时间推移、数据积累，将会越用越有价值。



常见问题3

问

为什么只有主机大数据分析？网络设备、存储等支持吗？

答

大数据分析在IT运维领域的应用目前都属于在探索；
主机大数据分析是我们大数据分析平台面向IT运维领域的第一个大数据分析场景；
网络、存储等会在后续支持。当然，非常期待与专家合作，共同探讨大数据分析场景，快速推出，合作共赢。

我们产品紧接着的下一步重磅功能：

知识图谱：

当我们进行搜索时，搜索结果右侧的联想，来自于知识图谱技术的应用；

我们几乎每天都会接收到各种各样的推荐信息，从新闻、购物到吃饭、娱乐。个性化推荐也来自于知识图谱技术的应用；

搜索、地图、个性化推荐、互联网、风控、银行……越来越多的应用场景，都越来越依赖知识图谱。

静待知识图谱在IT运维领域的应用！！！

SMARTEAGLE
慧鹰

感谢您的观看

THANK YOU FOR YOUR LISTENING