数字转型 架构演进

SACC

2019 中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2019

IT168.com ChinaUnix.net ITPUB

2019年10月31-11月2日 | 北京海淀永泰福朋喜来登酒店

DevOps到AIOps-智能化故障处理系统

- 一.背景
- 二.问题
- 三.解决
- 四.规划
- 五.Q&A

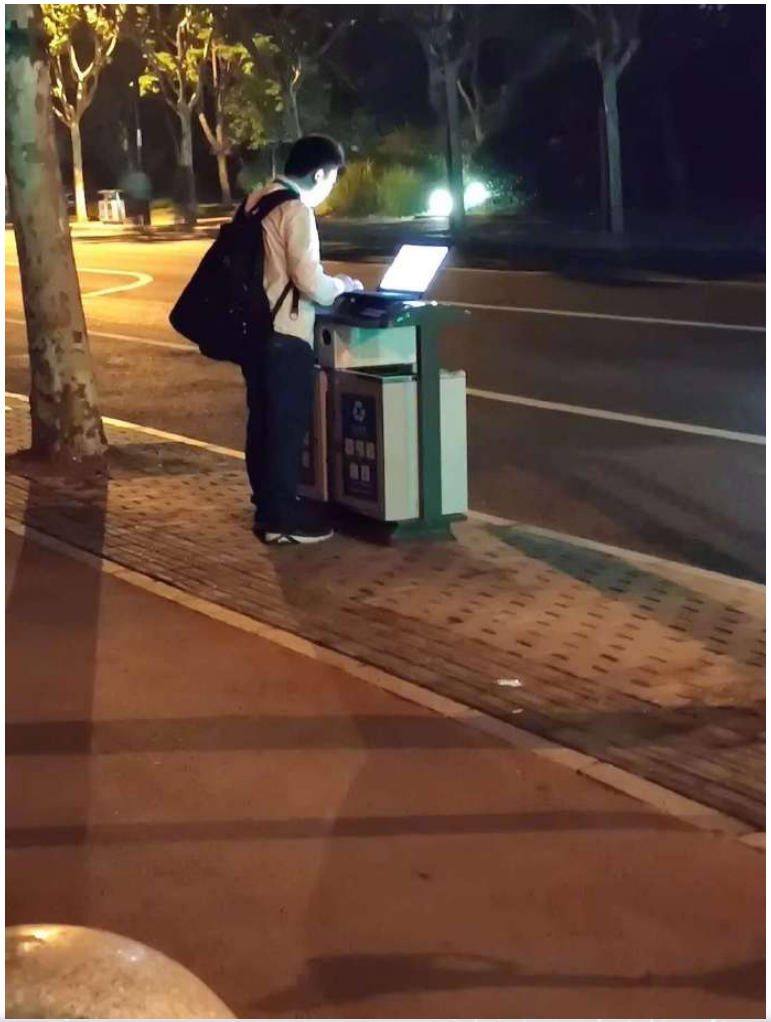
陈永清@翼课网

DevOps到AIOps-智能化故障处理系统

- **一.背景** ⟵
- 二.问题
- 三.解决
- 四.规划
- 五.Q&A

陈永清@翼课网

一.背景

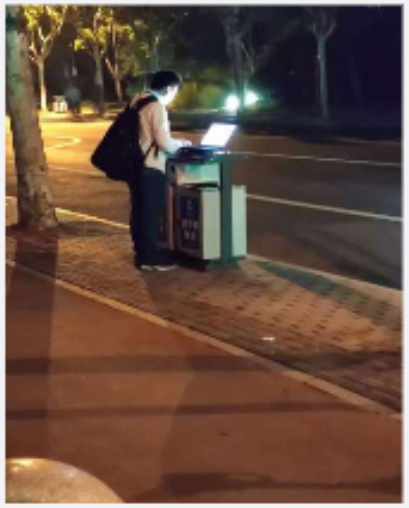


6月22日 09:34 来自 华为P8

//@RogerZhuo:运维，后端突然崩了//@马少平THU: 程序员，突然想到了一个bug//@昵称已被占用2018_://@石锋强19880101://@RevengeRangers:也可能是个运维//@今天WB倒闭了吗:我赌程序员，设计师和编辑才不会这样，拖稿才是他们干的事~~~

@互联网的那点事

他可能是个程序员，也可能是个设计，也可能是个编辑…但他也可能是个丈夫，是个父亲… 他就是你！成年人的世界哪有容易二字！



6月21日 23:37 来自 iPhone客户端 已编辑　583 | 374

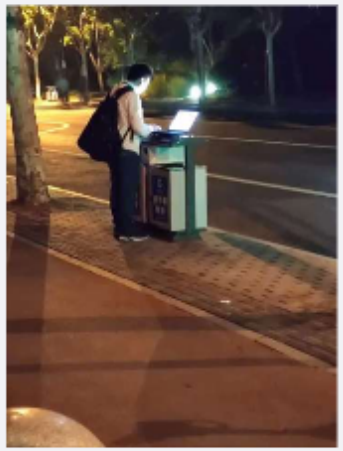
一.背景

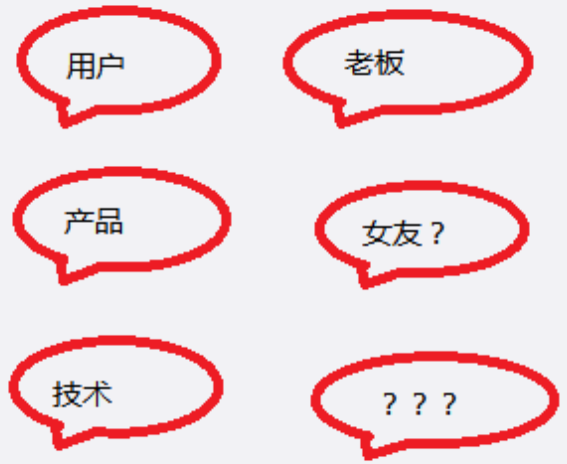
6月22日 09:34 来自 华为P8

//@RogerZhuo:运维，后端突然崩了//@马少平THU: 程序员，突然想到了一个bug//@昵称已被占用2018_:////@石锋强19880101://@RevengeRangers:也可能是个运维//@今天WB倒闭了吗:我赌程序员，设计师和编辑才不会这样，拖稿才是他们干的事~~~

@互联网的那点事

他可能是个程序员，也可能是个设计，也可能是个编辑…但他也可能是个丈夫，是个父亲… 他就是你！成年人的世界哪有容易二字！





6月21日 23:37 来自 iPhone客户端 已编辑　583 | 374 | 1340

一.背景

- 1.一个系统，不可能没告警(故障)。
- 2.处理告警很痛苦。
- 3.不处理影响用户满意度。
- 4.不处理影响公司营收。
- 5. 处理了，处理好了，产品好用了，用户满意了，公司也有利了。

- 以上，当告警发生时，
- 1.技术人员需要在极短时间，接受各种压力，心情是焦虑的，茫然的，错愕的，担忧的，惆怅的。
- 2.用户需要承受使用产品过程中的不爽，不痛快。
- 3.公司需要承受指责。
- 4.客服需要承受漫骂。

DevOps到AIOps-智能化故障处理系统

- 一.背景
- **二.问题** ⟵
- 三.解决
- 四.规划
- 五.Q&A

陈永清@翼课网

二.问题-界定

- 1.如何高效、精准、快速的处理告警(故障)?
- 2.什么时间处理。
- 3.谁处理。
- 4.处理到什么程度。

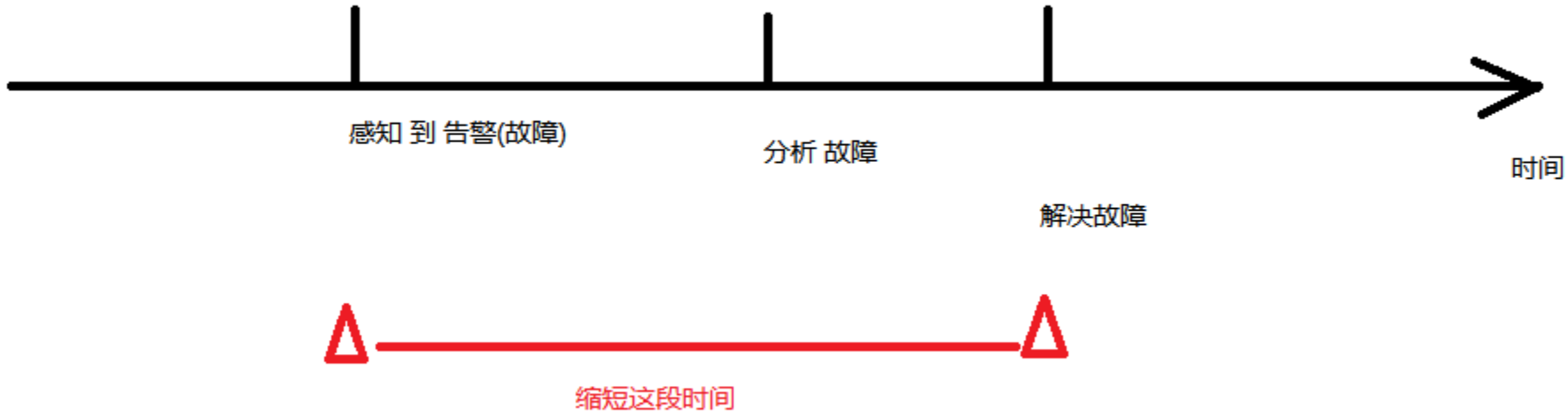
- **需要 定量、定性 分析**

二.问题-“4个三”定方向

- 通过"4个三"，做定量定性分析。
- 三个步骤看流程，感知、分析、解决 是处理故障的三个步骤。我们从分析环节入手。
- 三个维度找方向，影响最大的、频率最高的、最难处理的 告警 找到‘痛点’。
- 三个集合做决策，告警+决策点+原因 三个集合，找到关联性。
- 三个10做定量，针对过去1年的10大类告警，以DBA人力需要10分钟以上分析出告警原因，现在要系统10秒内分析出结果。

二.问题-故障生命周期

- 通过"4个三"，做定量定性分析。
- 三个步骤看流程，感知、分析、解决 是处理故障的三个步骤。我们从分析环节入手。



二.问题-找痛点

- 通过"4个三"，做定量定性分析。

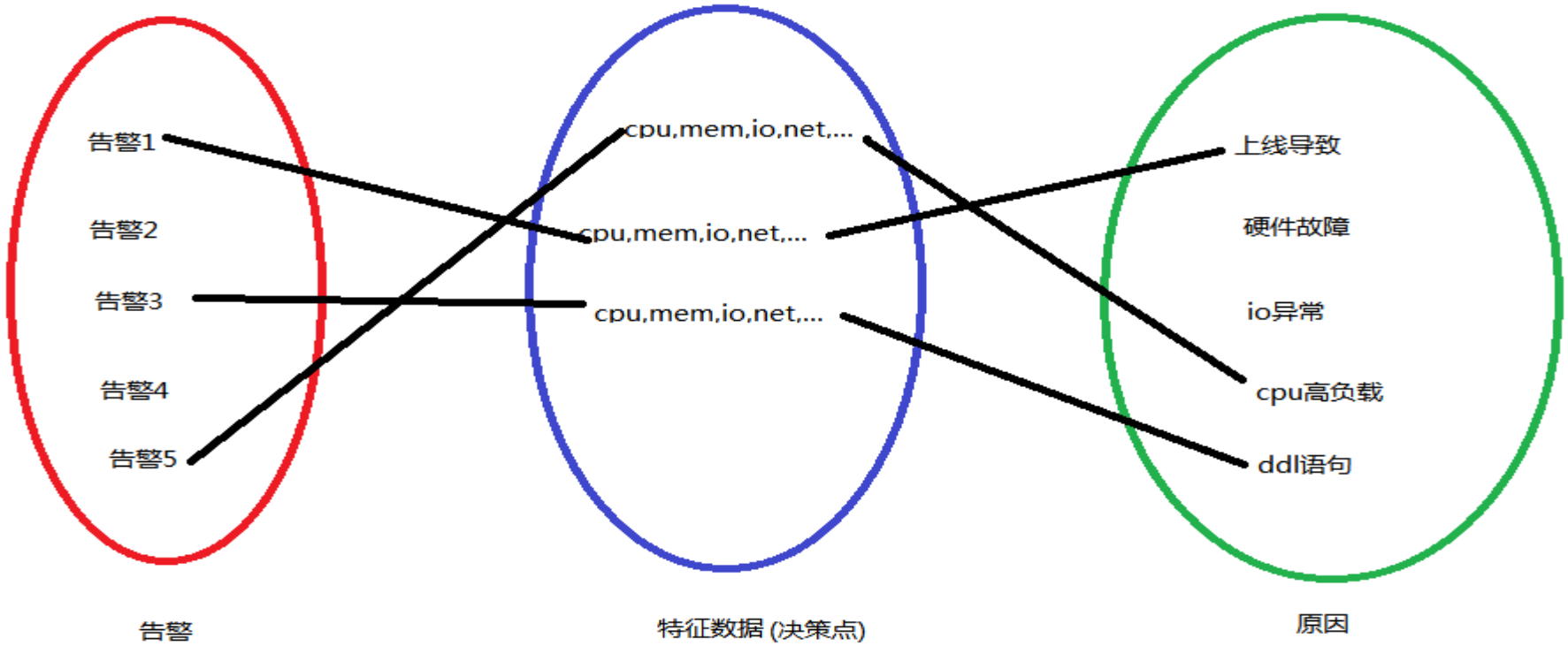
- 三个维度找方向，频率最高的、影响最大的、最难处理的 告警 找到‘痛点’。

- 从过去1年，统计告警类型和对应出现次数，按照 出现次数最多，影响最大，最难处理，三个维度，来决定，哪些告警(故障) 是痛点，最应该首先被智能化处理的。
- AAAAA类告警　10000次，影响xxx，处理难度sss
- BBBBB类告警　500次，影响xxx，处理难度sss
- CCCCC类告警　400次，影响xxx，处理难度sss

二.问题-三个集合找关联

- 通过"4个三"，做定量定性分析。

- 三个集合做决策，告警+决策点+原因 三个集合，找到关联性。



二.问题-10秒内分析出核心告警原因

- 通过"4个三"，做定量定性分析。

- 三个10做定量，针对过去1年的10大类告警，以DBA人力需要10分钟以上分析出告警原因，现在要系统10秒内分析出结果。

- 10大类告警
- 过去人需要10分钟分析出原因
- 现在需要做到10秒内分析出原因

二.问题-“4个三”定方向

- 通过"4个三"，做定量定性分析。
- 三个步骤看流程，感知、分析、解决 是处理故障的三个步骤。我们从分析环节入手。
- 三个维度找方向，影响最大的、频率最高的、最难处理的 告警 找到‘痛点’。
- 三个集合做决策，告警+决策点+原因 三个集合，找到关联性。
- 三个10做定量，针对过去1年的10大类告警，以DBA人力需要10分钟以上分析出告警原因，现在要系统10秒内分析出结果。

- **总结**，先 解决 **最痛**的点，从 故障**分析** 环节入手，达到提升 **准确率**和**效率** 的目的。从而 减轻大家的痛苦，让用户满意。

DevOps到AIOps-智能化故障处理系统

- 一.背景
- 二.问题
- **三.解决** ⟵
- 四.规划
- 五.Q&A

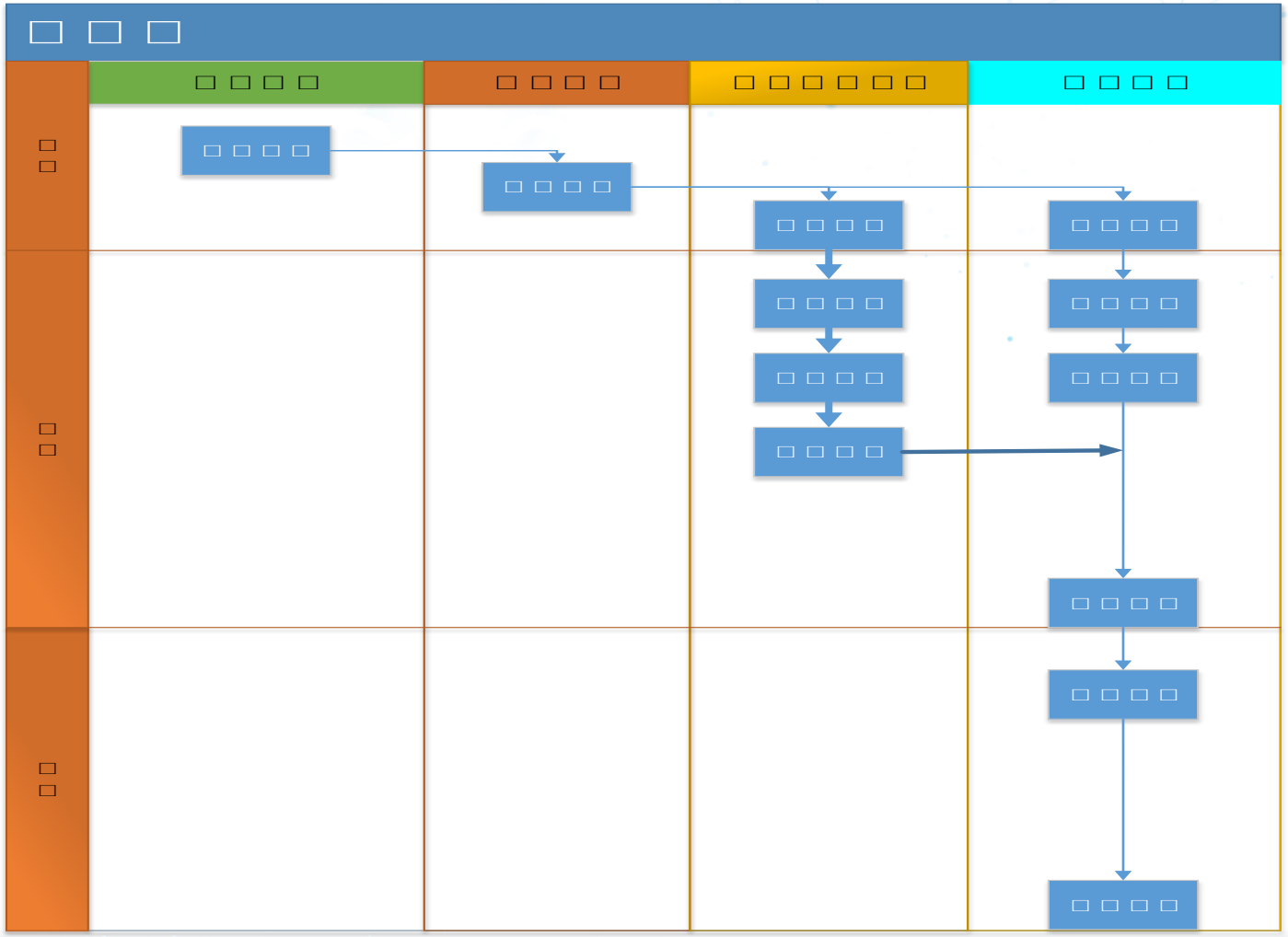
陈永清@翼课网

三.解决

- 1.业务流
- 2.数据流
- 3.架构设计
- 4.算法

三.解决-业务流

- 1.业务流



三.解决-数据流

- 2.数据流
- 采集
- 存储
- 计算

三.解决-数据流-采集

- 出异常时，技术人员需要排查的监控项数据，需要采集

- 采集系统特征数据

  包括4大维度数据：硬件、网络、操作系统、应用程序

  硬件：硬件如cpu、memory、硬盘、网卡、温度等状态数据。

  网络：吞吐量流入/流出、吞吐率流入/流程、丢包率等。

  操作系统：socket状态、cpu、mem、io 等使用量、使用率等。

  应用程序：并发量、错误率、mysql global status 、processlist、innodb status 、slow query 、lock 性能指标数据等。

- 一个时刻，一共108项。每项，再取3-8个时刻，组成几百维的多维时序数据。

三.解决-数据流-采集

- **难点：**

  1.不能对线上系统侵入性太大

  2.不能因采集数据影响太大线上系统性能

  3.需要扩展的、灵活的增加、变更采集项目

- 对时序性的、几百维度的、不影响或少影响线上、线上系统的 状态数据做 采集 是个难题。

三.解决-数据流-存储

- **难点：**

1.数据维度多、扩展添加、实时性要求高

2.对业务侵入性小

3.存储后方便的读

三.解决-数据流-采集和存储

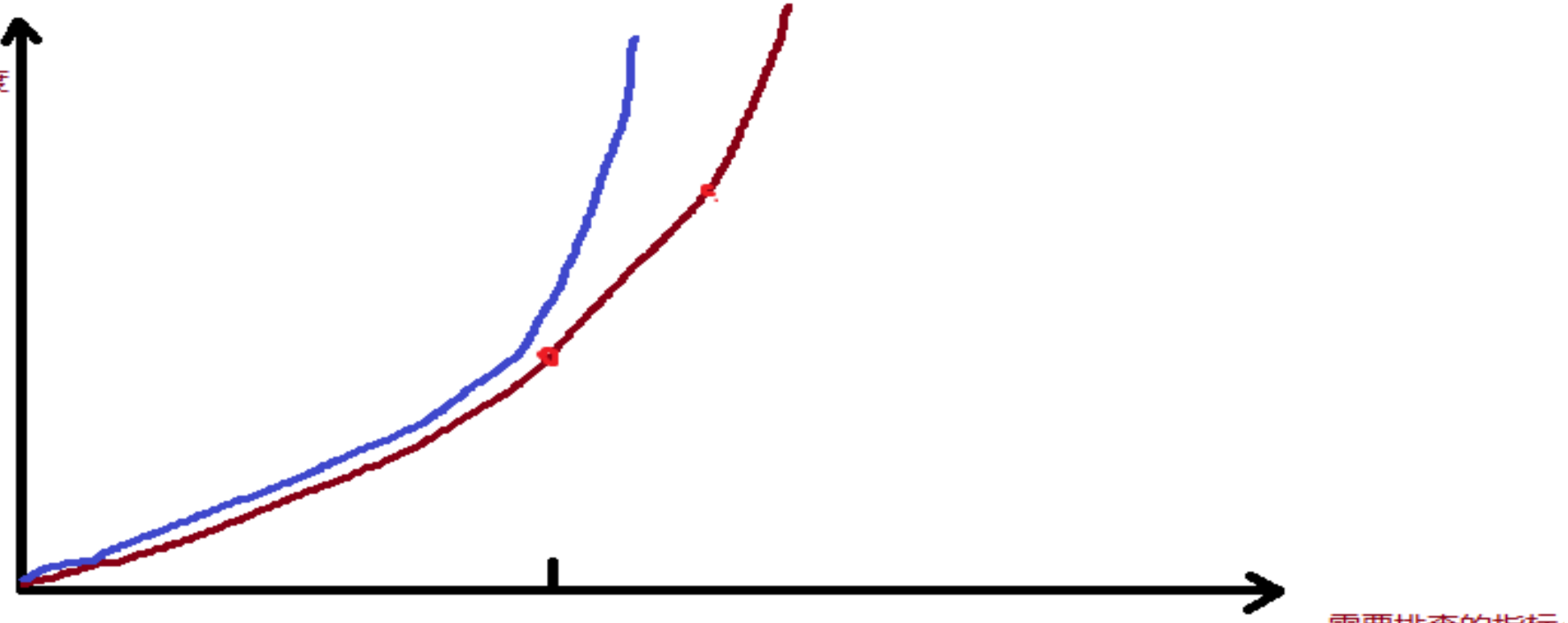
- **解决：**

  1.主要的：通过**监控系统** 完成 **采集**和**存储**

  2.辅助的：通过主机/进程打本地日志做采集和存储

三.解决-数据流-计算

- **难点**



结论1：系统越复杂，故障时需要排查的**指标个数越多**，那么分析**问题的复杂度越大**

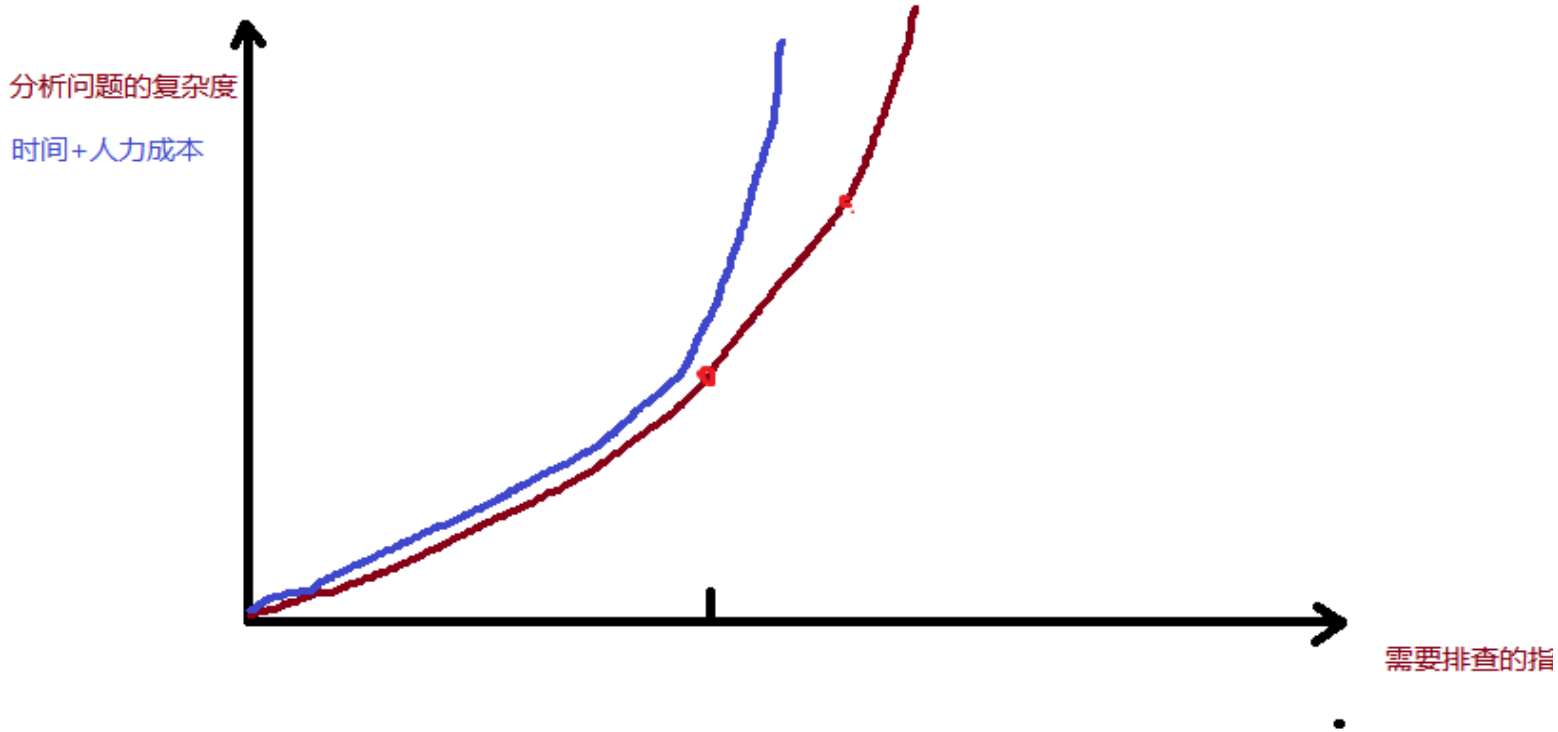
结论2：系统越复杂，故障时需要排查的**指标个数越多**，那么时间+人力**成本越大**

三.解决-数据流-计算

- **解决**

1.AIOPS：用AI的技术，解决OP的问题

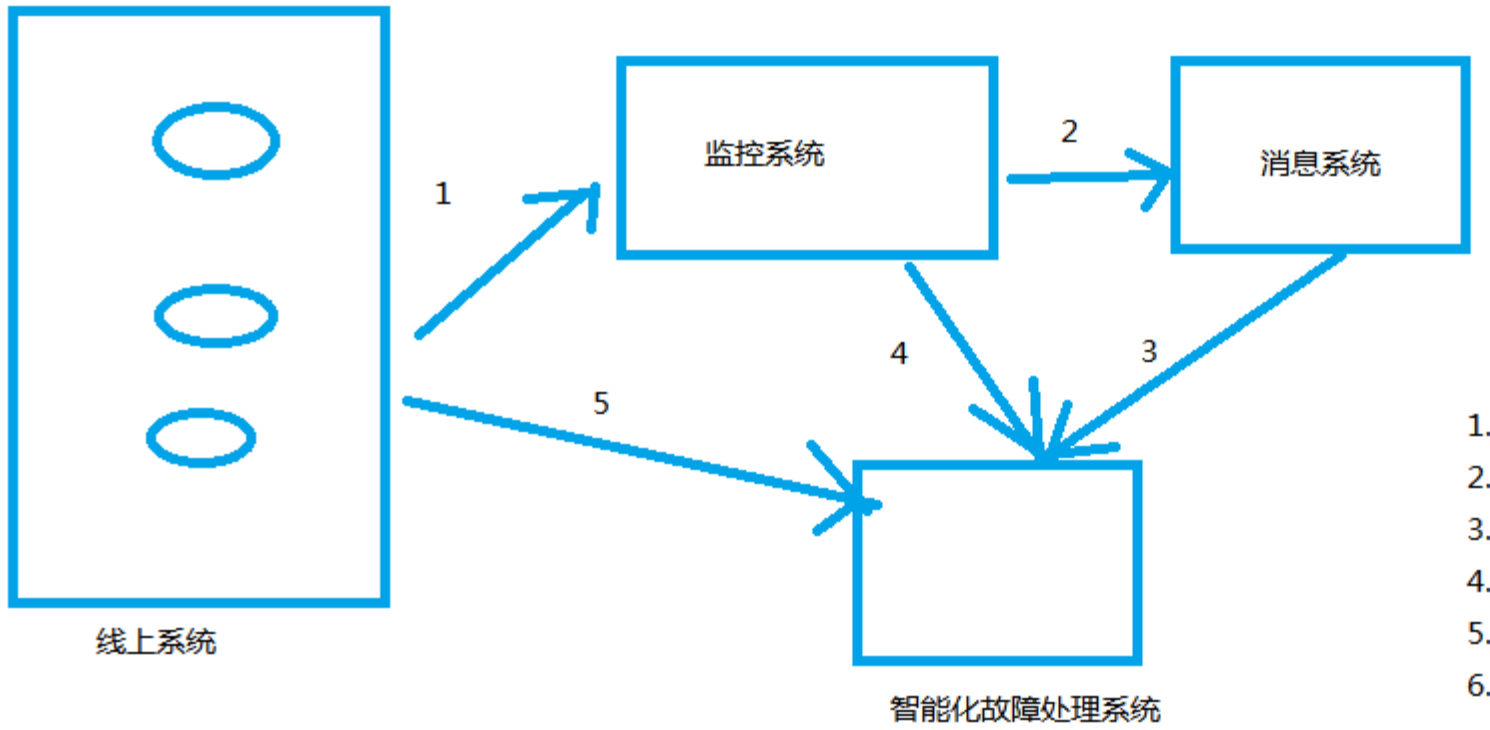
2.有监督机器学习



结论1：系统越复杂，故障时需要排查的**指标个数越多**，那么分析**问题的复杂度越大**

结论2：系统越复杂，故障时需要排查的**指标个数越多**，那么时间+人力**成本越大**

三.解决-数据流-计算



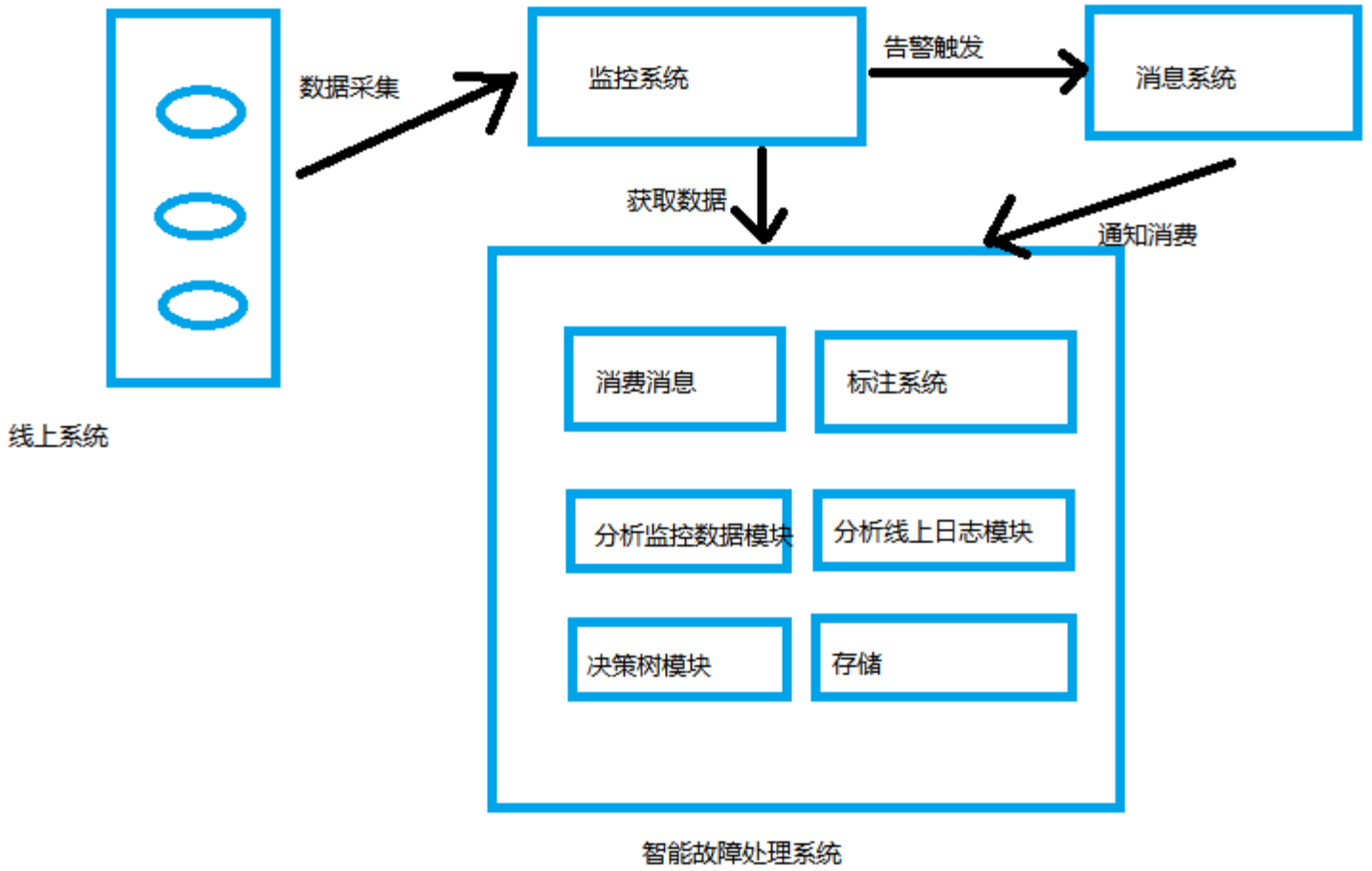
1.监控系统 监控线上系统。
2.出异常触发告警，消息系统受到msg。
3.智能故障处理系统，消费msg，感知告警。
4.智能故障处理系统，从监控系统获得数据。
5.智能故障处理系统，从线上服务器获得数据。
6.进行决策分析。

三.解决-架构设计

- 数据采集
- 数据存储
- 数据计算
- 数据标注

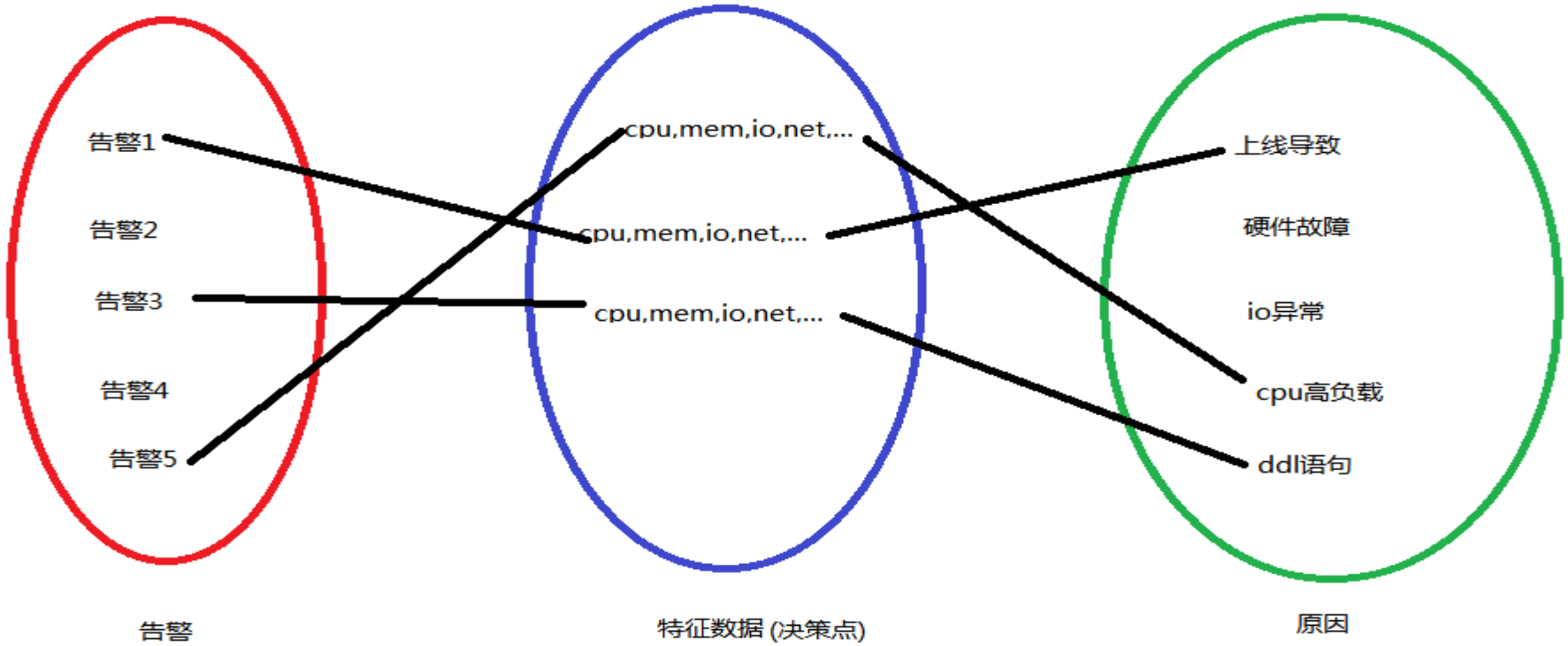
三.解决-架构设计-采集、存储、计算

- 数据采集
- 数据存储
- 数据计算



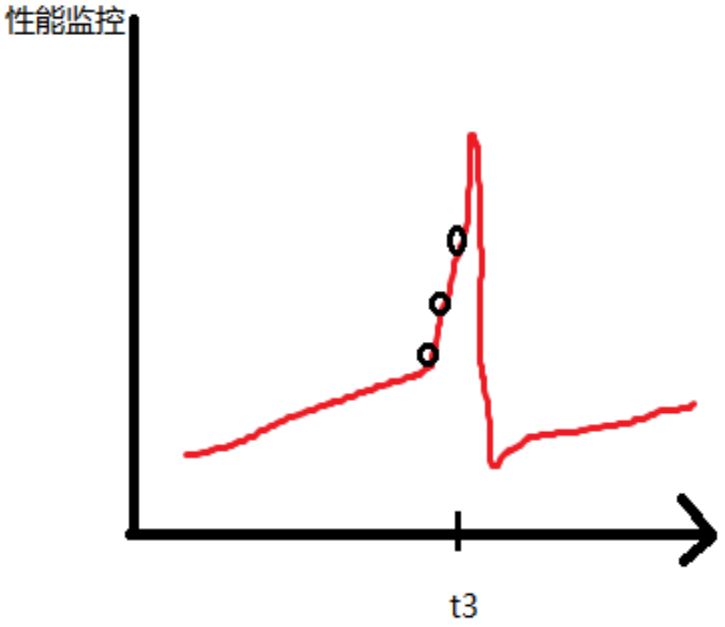
三.解决-架构设计-标注

- 数据标注



三.解决-架构设计-标注

- 数据标注



标注：t3时刻，发生故障。建立 告警 系统状态 告警原因 关联性

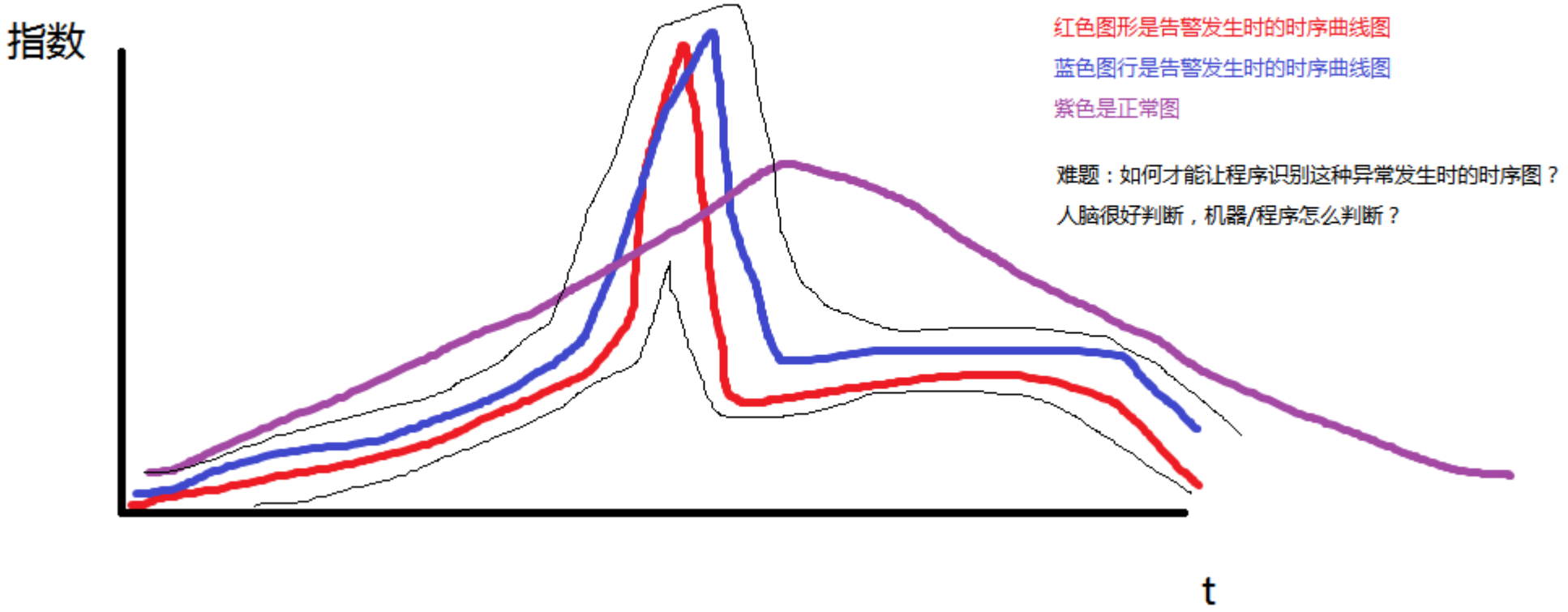
比如：

数据库慢查询告警	硬件 网络 操作系统 数据库 4大维度状态时序数据	磁盘坏快导致
监控产生	监控产生	人工标注

三.解决-算法

- **难点：**

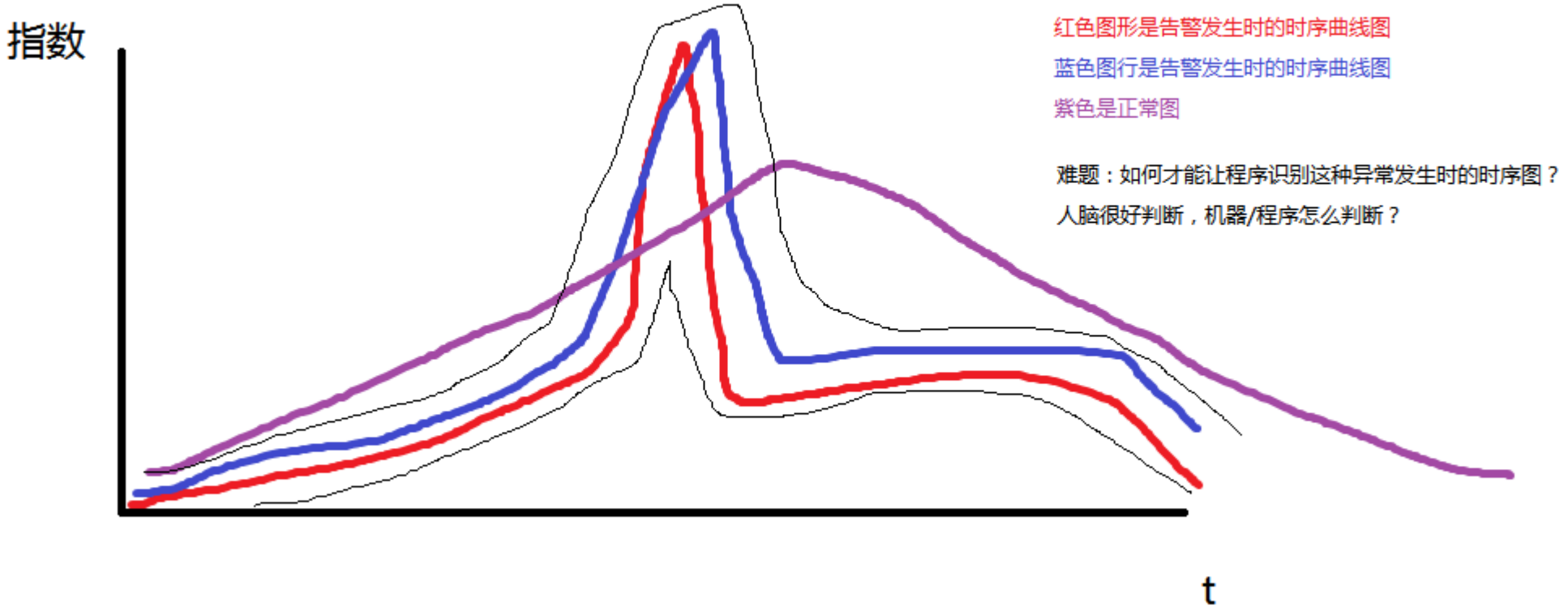
  肉眼很好识别这个异常，机器/程序怎么识别



三.解决-算法

- **解决：**

  时序图形编码算法

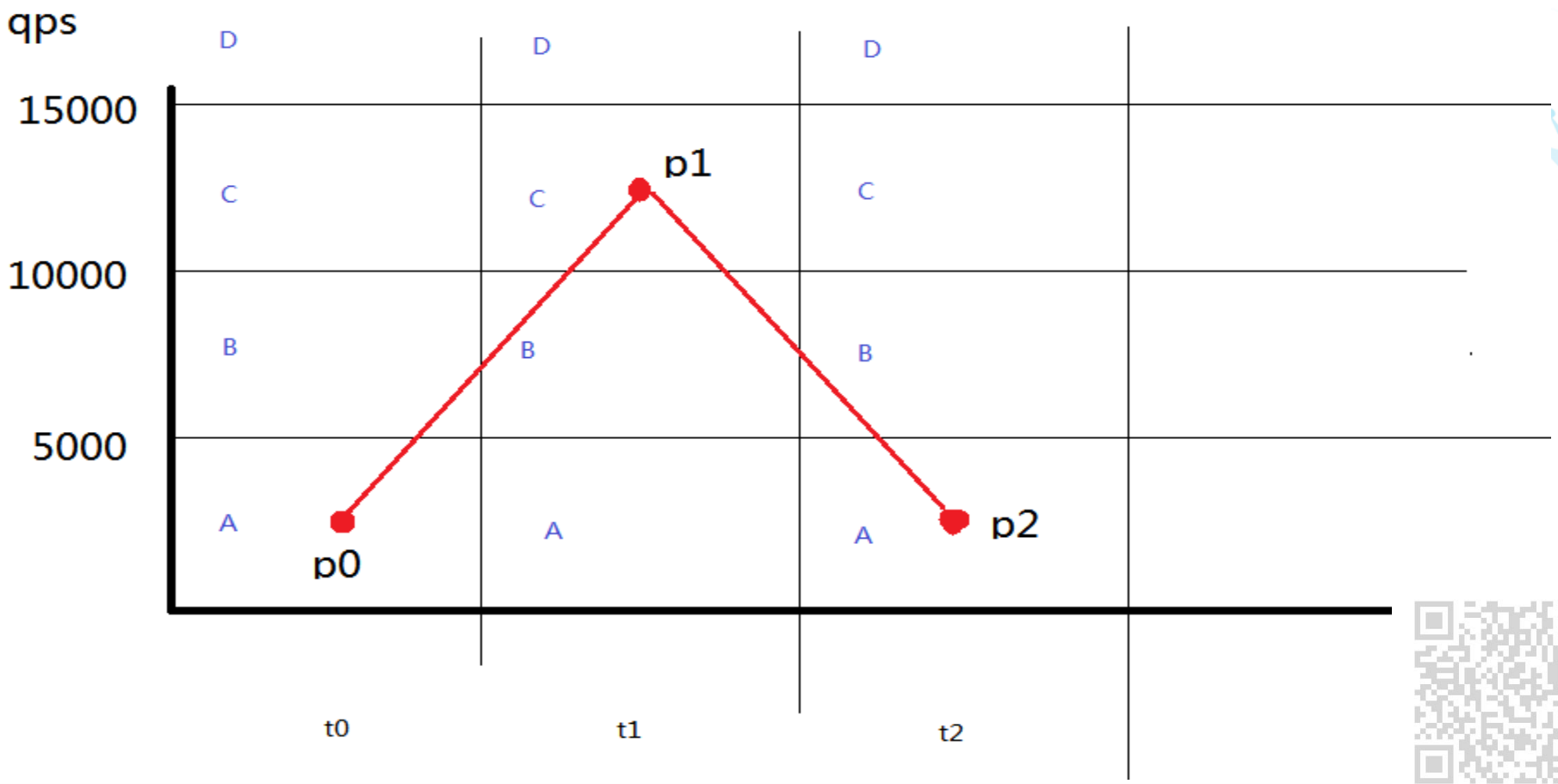


三.解决-算法

- 时序图形编码算法

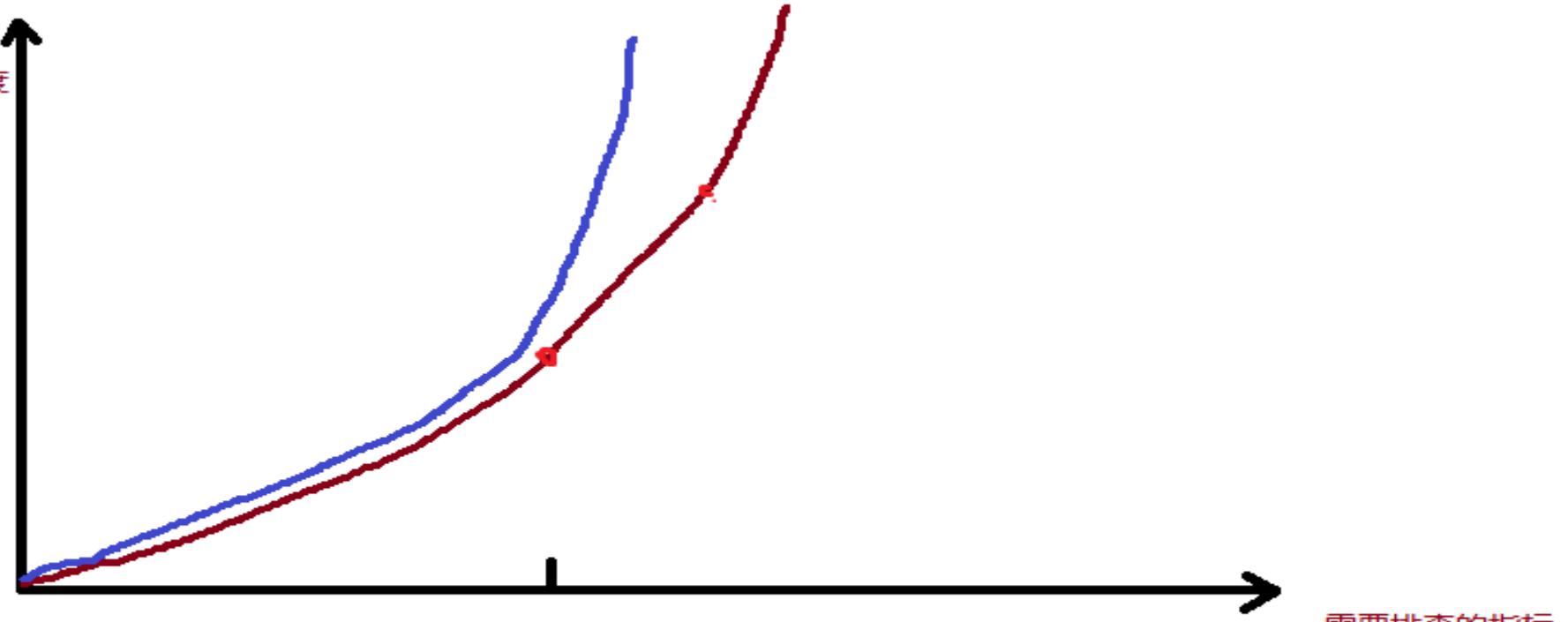
1.对区域进行划分
2.对子区域进行编码
3.对时序点进行编码
4.对时序曲线进行编码

p0,p1,p2
转换成
ACA



三.解决-算法

- **难点：几十，几百，甚至千个、万个 监控数据怎么排查故障**



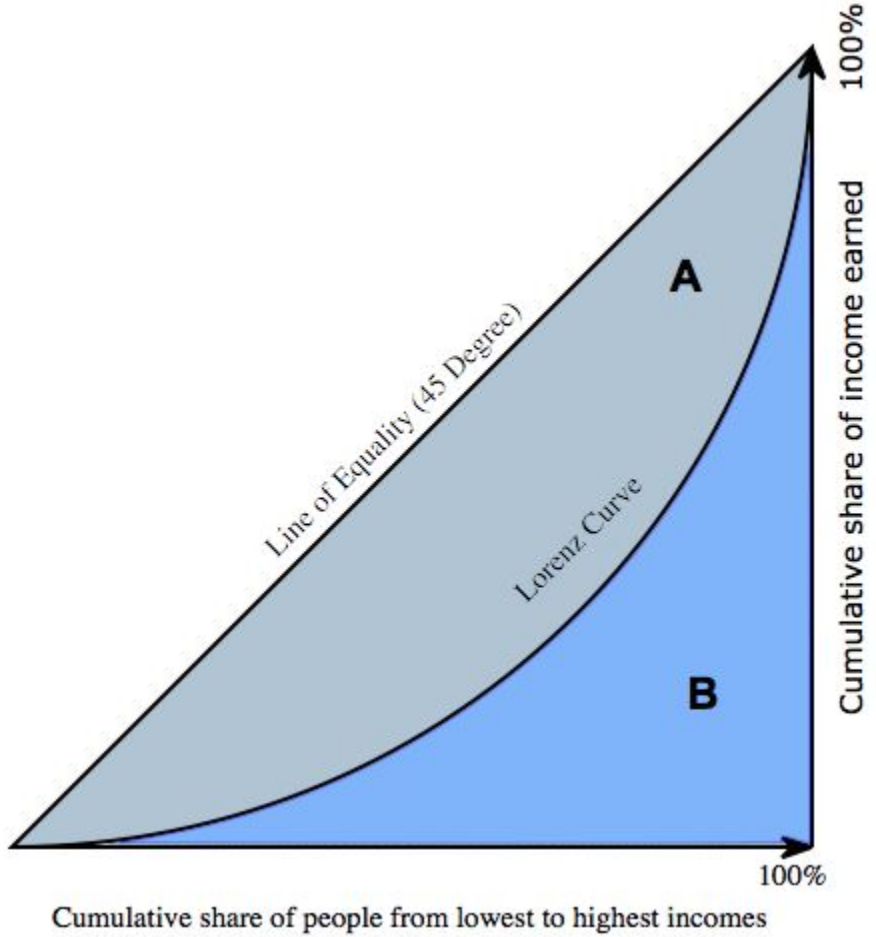
结论1：系统越复杂，故障时需要排查的**指标个数越多**，那么分析**问题的复杂度越大**

结论2：系统越复杂，故障时需要排查的**指标个数越多**，那么时间+人力**成本越大**

三.解决-算法

- 基尼指数

- gini=A/(A+B)



三.解决-算法

- 决策算法-CART

使用gini 系数做分类依据

特征数据：

t0,硬件状态,cpu,io,mem,status

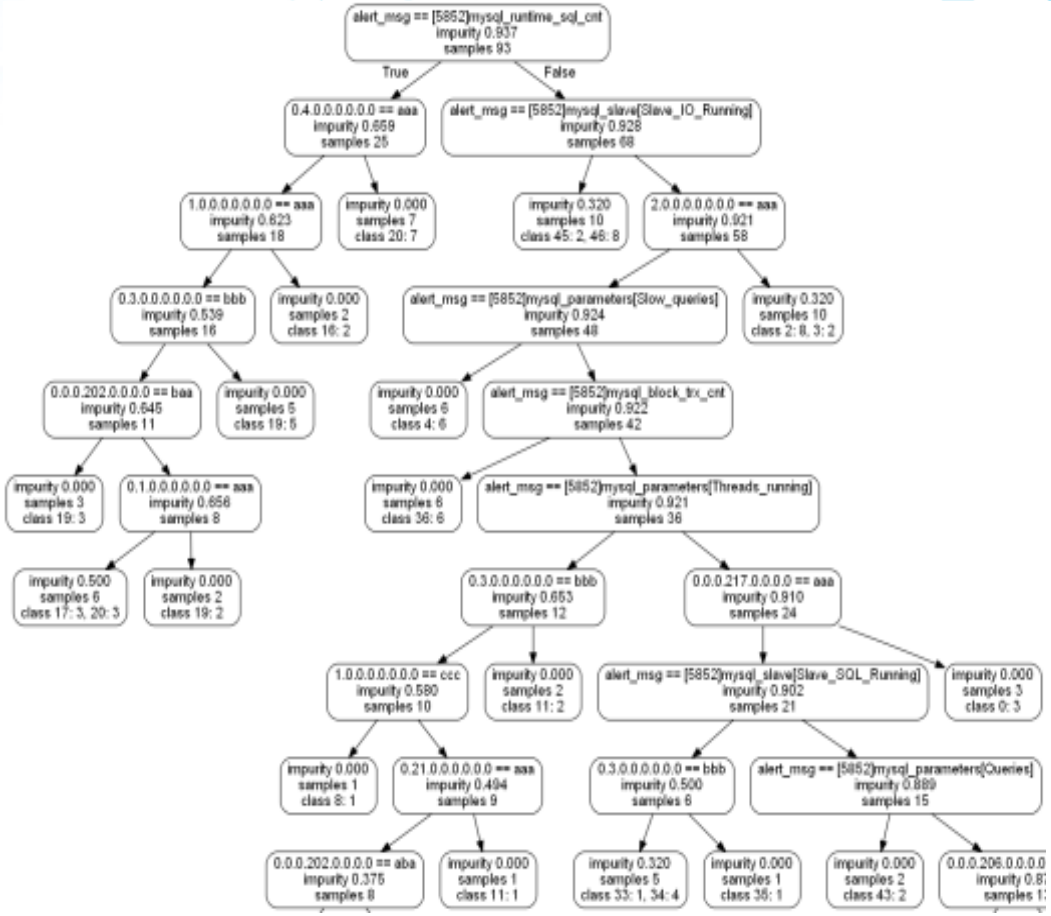
t1,硬件状态,cpu,io,mem,status

t2,硬件状态,cpu,io,mem,status

t3,硬件状态,cpu,io,mem,status

人工标注结果：

原因1



1.通过人工标注，获得标注数据。

2.有监督学习，使用标注数据，跑CART算法，得到决策树模型，生成决策树。

3.输入故障时的特征数据，通过决策树，做故障决策，分析故障原因。

4.不断迭代，提升泛化性，如随机森林，如采集更多数据。

DevOps到AIOps-智能化故障处理系统

- 一.背景
- 二.问题
- 三.解决
- **四.规划** ⟵
- 五.Q&A

陈永清@翼课网

四.规划

- AIOPS
- 智能化故障处理

四.规划-AIOPS

- 目标：以合适的契机、用恰当的资源、解决有挑战的技术难题，从而创造价值

- 策略：

  1. AIOPS ，用AI的技术，解决OP的问题

  2.先试点，再推广

  3.找准痛点，三个维度，频次高、影响大、解决难 的问题

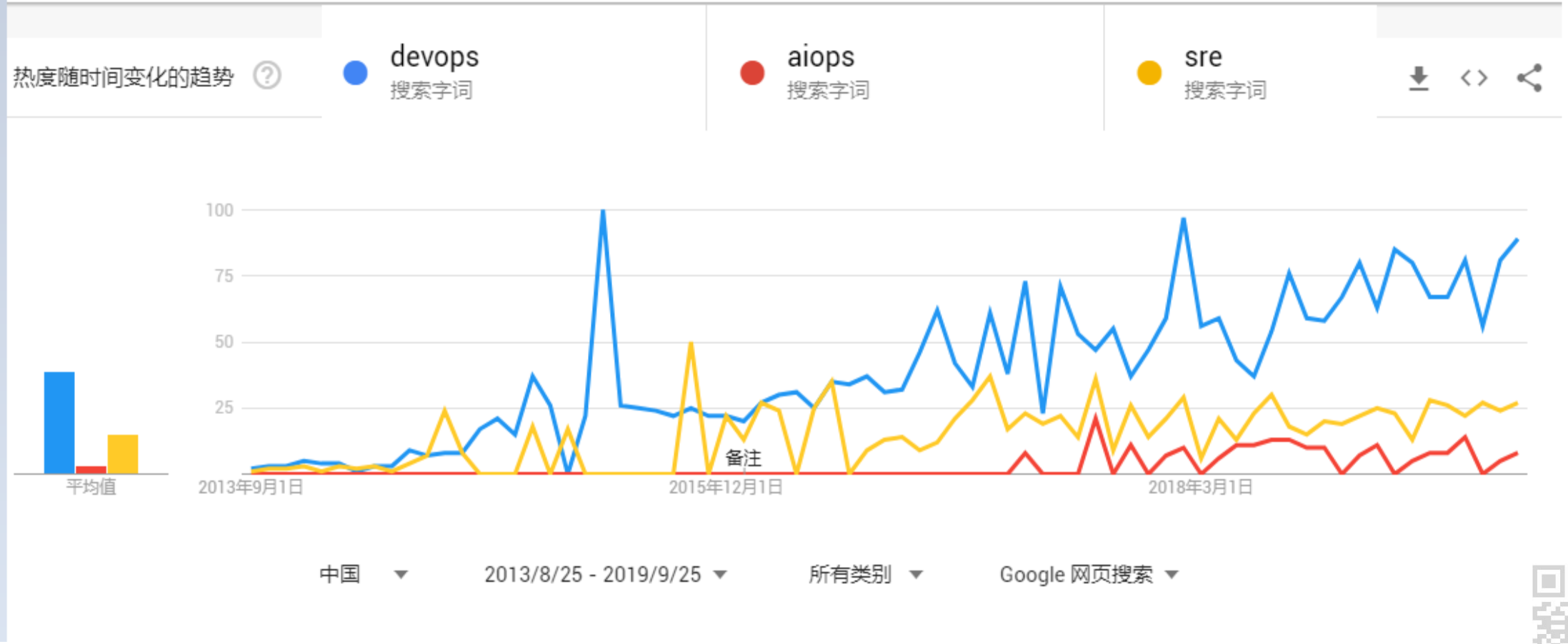
- 技术路线：

  1.抽象出需求，学习AI技术，储备知识

  2.用AIOPS，小范围试水，解决最迫切的需求　(频次高、影响大、解决难 的问题)

  3.逐步推广，降低成本，提升效率，减少痛苦

四.规划-AIOPS



四.规划-智能化故障处理

- 决策树规划
  1. 先按照系统架构，逐层建决策树，用于决策。
  2. 再建一个顶层决策树，做全局决策。

- 数据采集规划
  1. 将上线/代码发布 作为一个特征，收集到决策点中，用于判断是否故障是上线导致的。智能回滚。
  2. 对数据库执行计划，如explain输出，结合NLP，抽取特征。做sql 智能化审核。

- 计算规划
  1. 根据故障产生前的一段时间产生的数据，做故障预测。
  2. 通过算法得出组合特征数据做多值或多指标的智能化阈值告警，而不是单值阈值告警。减少报警。

五.Q&A

- 通过"4个三"，做定量定性分析。
- 三个步骤看流程，感知、分析、解决 是处理故障的三个步骤。我们从分析环节入手。
- 三个维度找方向，影响最大的、频率最高的、最难处理的 告警 找到‘痛点’。
- 三个集合做决策，告警+决策点+原因 三个集合，找到关联性。
- 三个10做定量，针对过去1年的10大类告警，以技术人力需要10分钟以上分析出告警原因，现在要系统10秒内分析出结果。

五.Q&A

- 谢谢　　陈永清@翼课网

Thanks