

甘道夫 (Gandalf) : Azure 安全部署的“任意门”

一项为大规模云基础架构保驾护航的智能分析服务



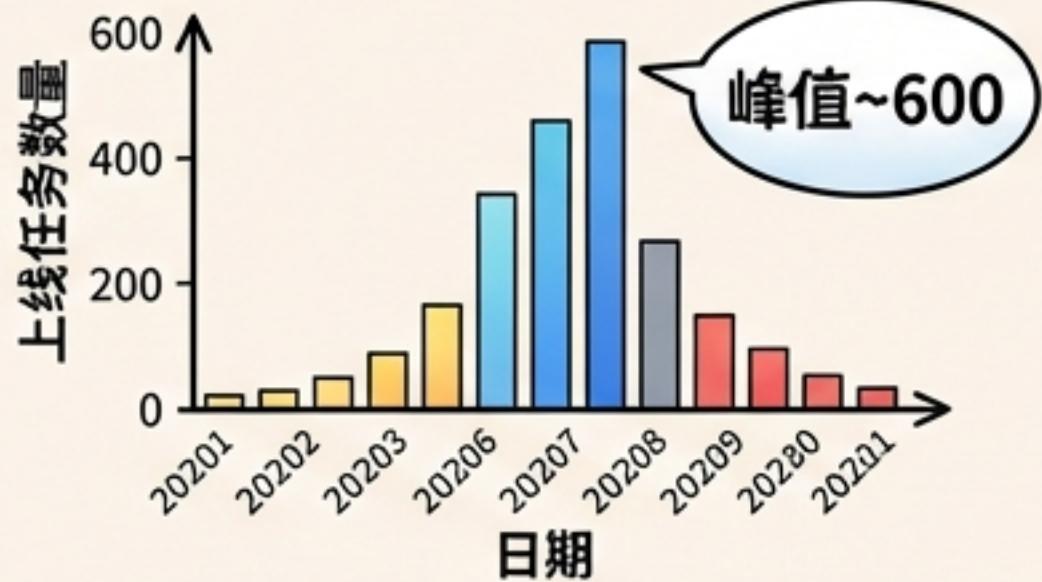
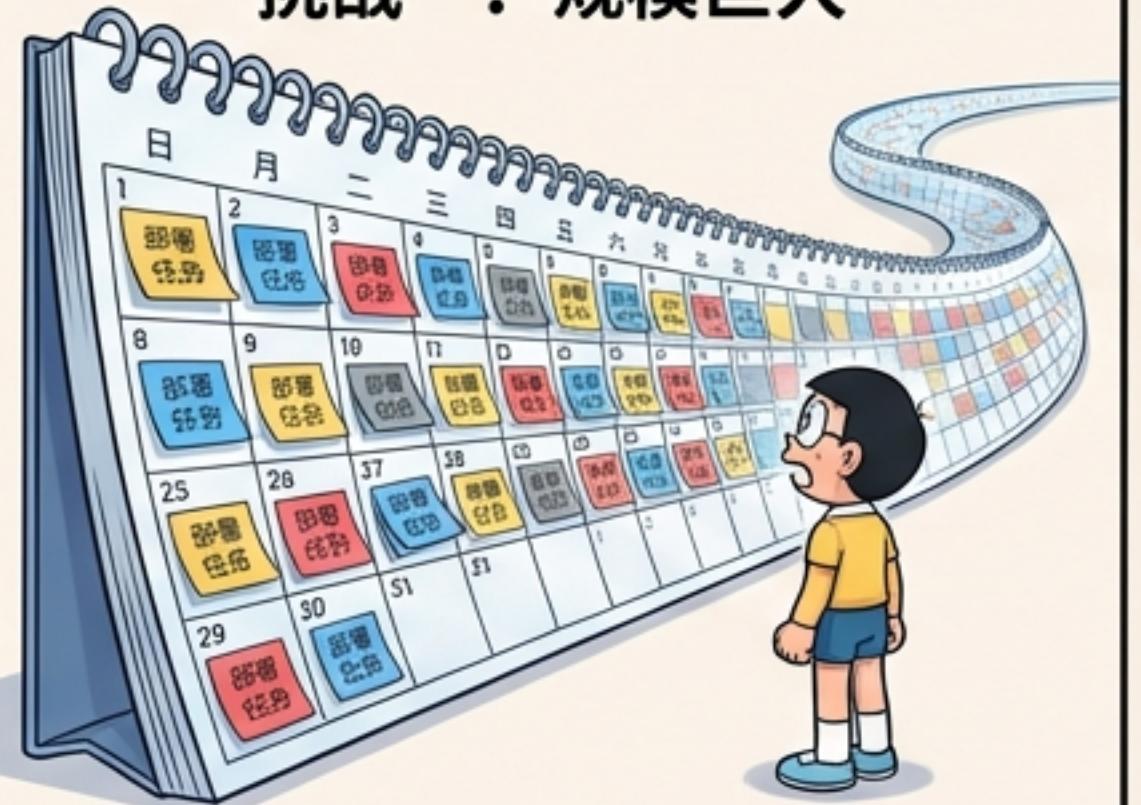
Azure开发者 大雄

哆啦A梦-甘道夫



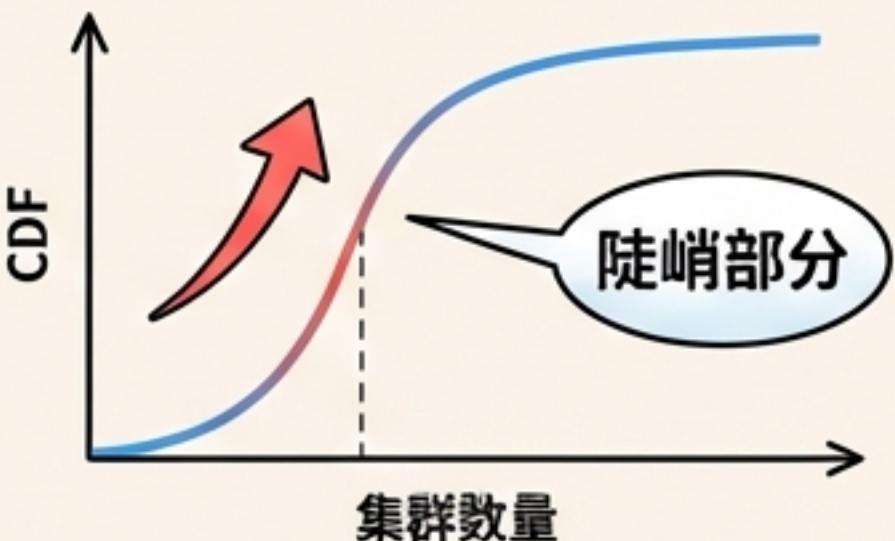
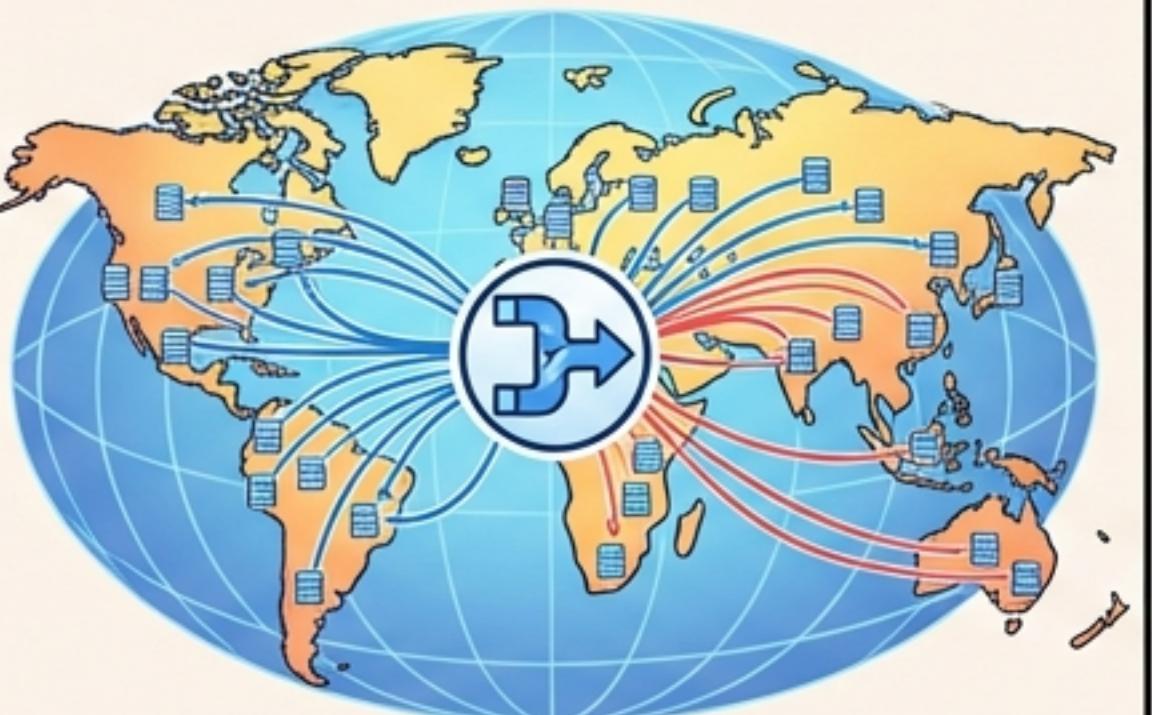
任何一个微小的失误，都可能引发全球性的服务中断…压力太大了！

挑战一：规模巨大



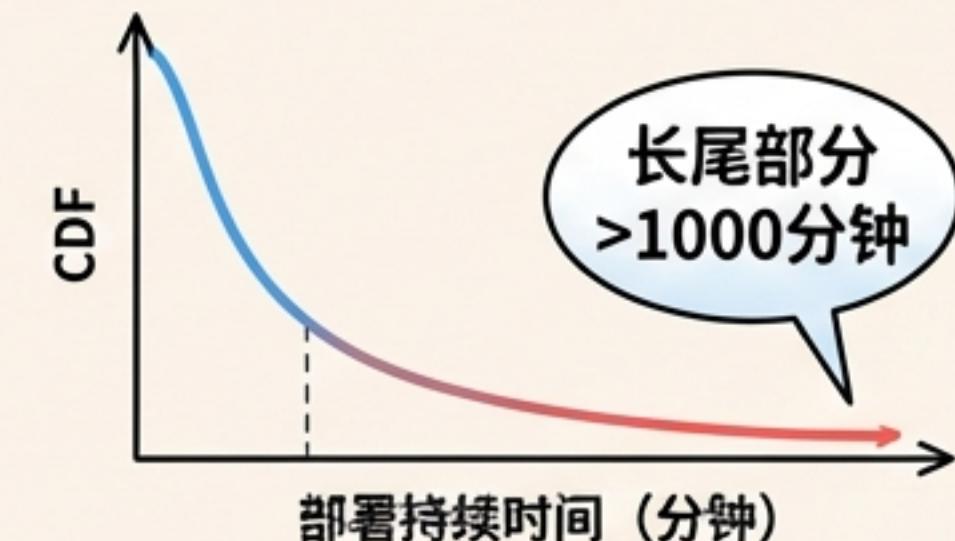
Azure 每天有数百个上线任务同时进行。

挑战二：范围广阔



超过70%的部署会影响多个集群。

挑战三：耗时漫长



超过20%的部署持续时间超过1000分钟。

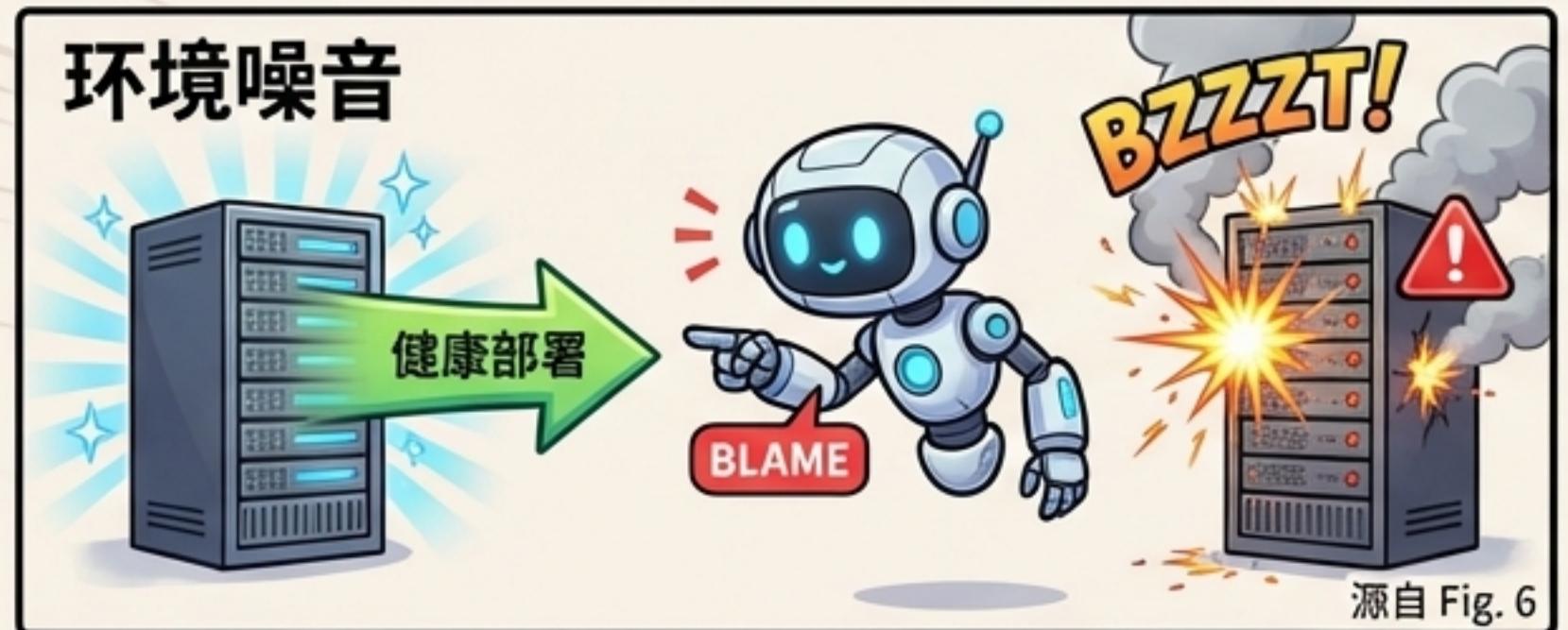
目录：深入了解我的神奇法宝！



甘道夫的法宝说明书

1. 棘手的难题：云规模部署的混乱与风险
2. 法宝登场：甘道夫系统概览
3. 法宝的秘密：核心算法深度解析
4. 大显神通：真实世界的惊人战绩
5. 全新的日常：更安全、更高效的部署新时代

为什么传统的监控方法不够用？



大量的环境噪音（如硬件故障、网络波动）常常会误导监控系统，导致无辜的部署被错误叫停。



有些问题（如内存泄漏）不会立即出现，而是在部署数小时甚至数天后才爆发，难以被快速察觉。



在同一集群上，每天可能有多个组件同时部署。当问题发生时，很难确定究竟是谁的责任。



一个组件的微小变更，可能不会导致自身失败，却会破坏与其他组件的API约定，引发连锁故障。

**这就是我的法宝——甘道夫！
一个端到端的智能分析服务，
它能从全局视角整体评估所有部署的影响！**

持续监控

实时监控基础设施遥测数据中的丰富信号。

智能分析

当检测到系统异常时，分析它是否由某个部署引起。

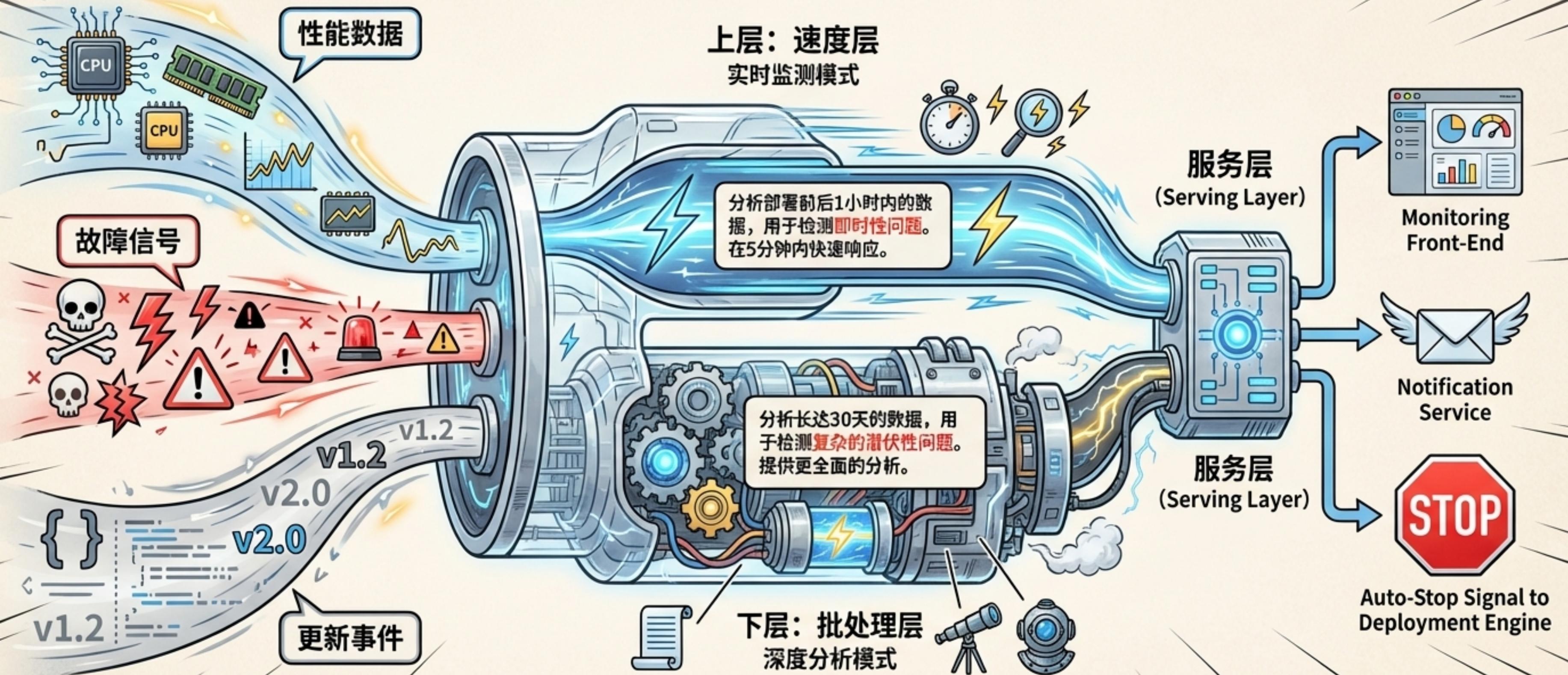
自动决策

如果识别出有问题的部署，甘道夫会自动发出“禁止通行(no-go)”决策来叫停它。

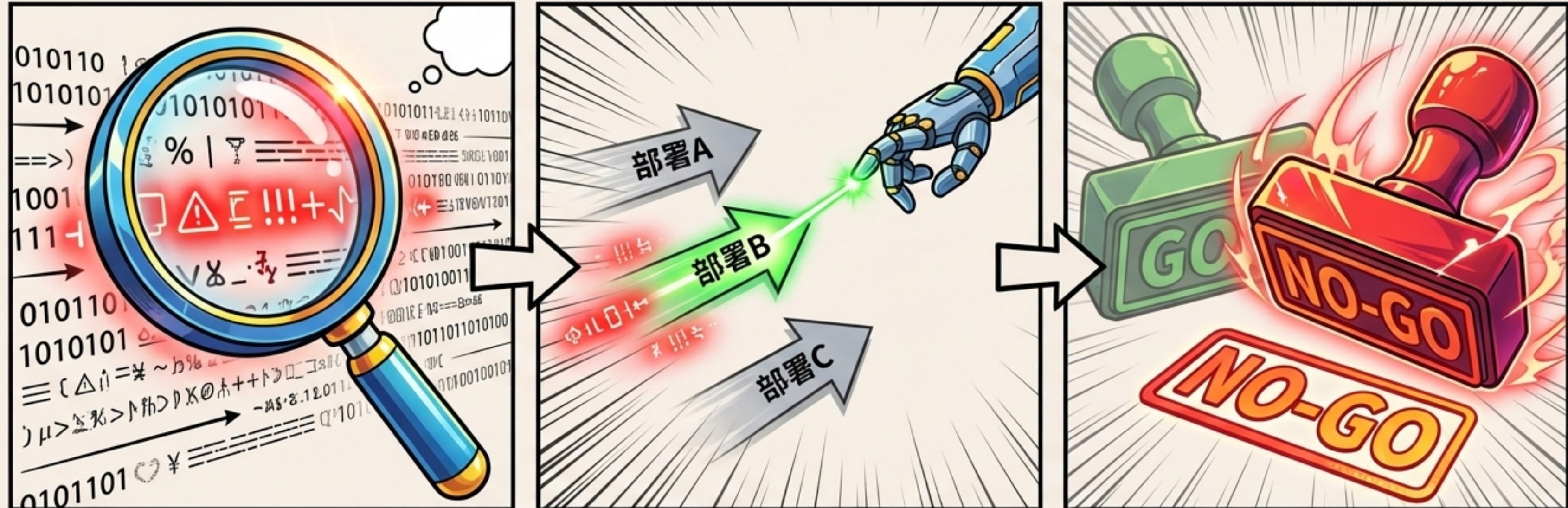
提供证据

提供详细的佐证和交互式前端，帮助工程师快速定位根因。

法宝的构造：实时与批处理的“双核引擎”



甘道夫的决策三部曲



第一步：异常检测

从海量原始遥测数据中，通过文本聚类和 Holt-Winters 预测模型，精确识别出有意义的故障信号和异常模式。

第二步：关联分析

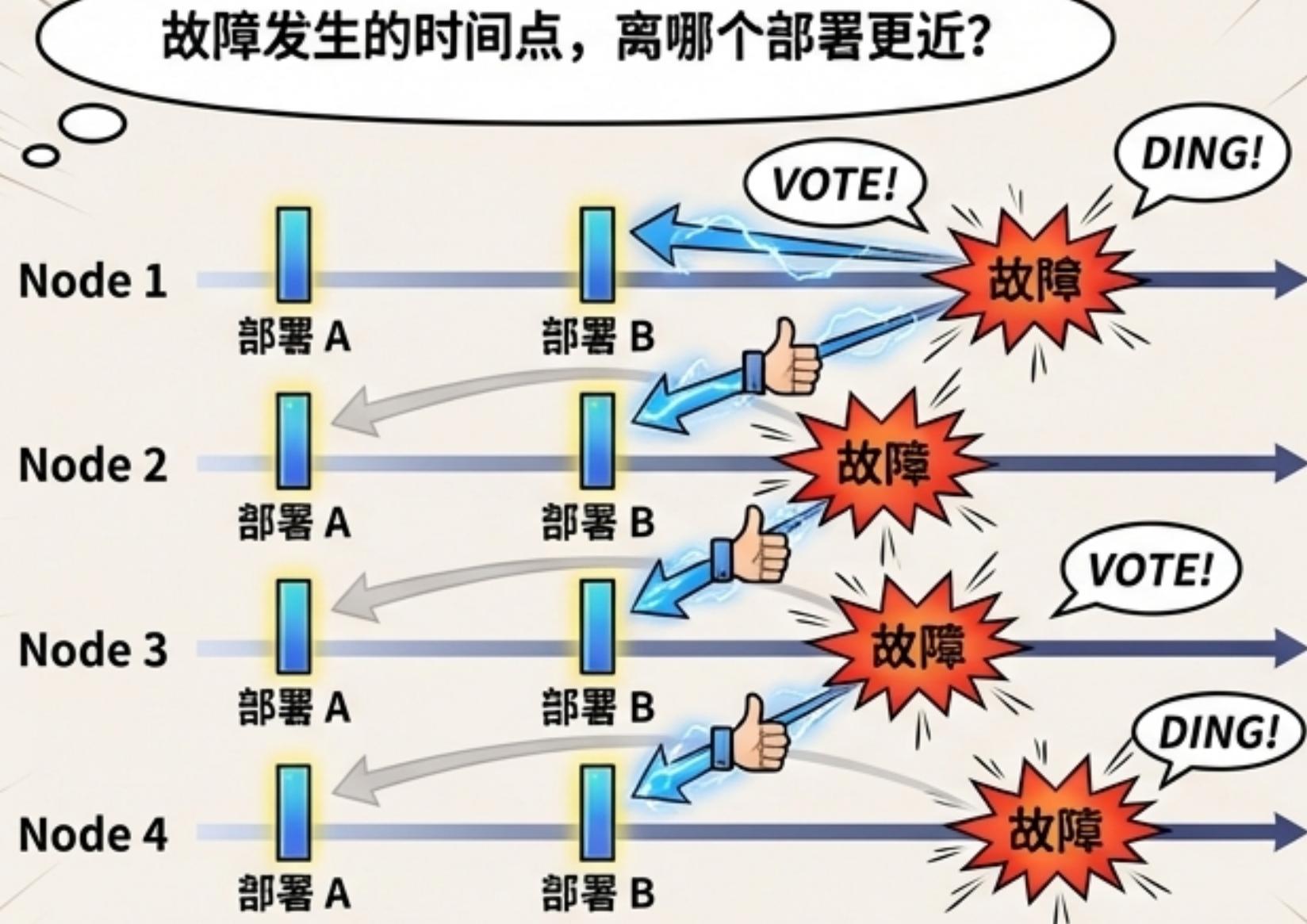
这是核心！通过时空关联和集成排序算法，在众多并发部署中，锁定最可疑的“元凶”。

第三步：影响评估与决策

使用高斯判别分类器，评估故障的影响范围（如影响的集群、节点、客户数量），最终做出是否叫停部署的决定。

核心揭秘(1): 时空锁定, 揪出元凶

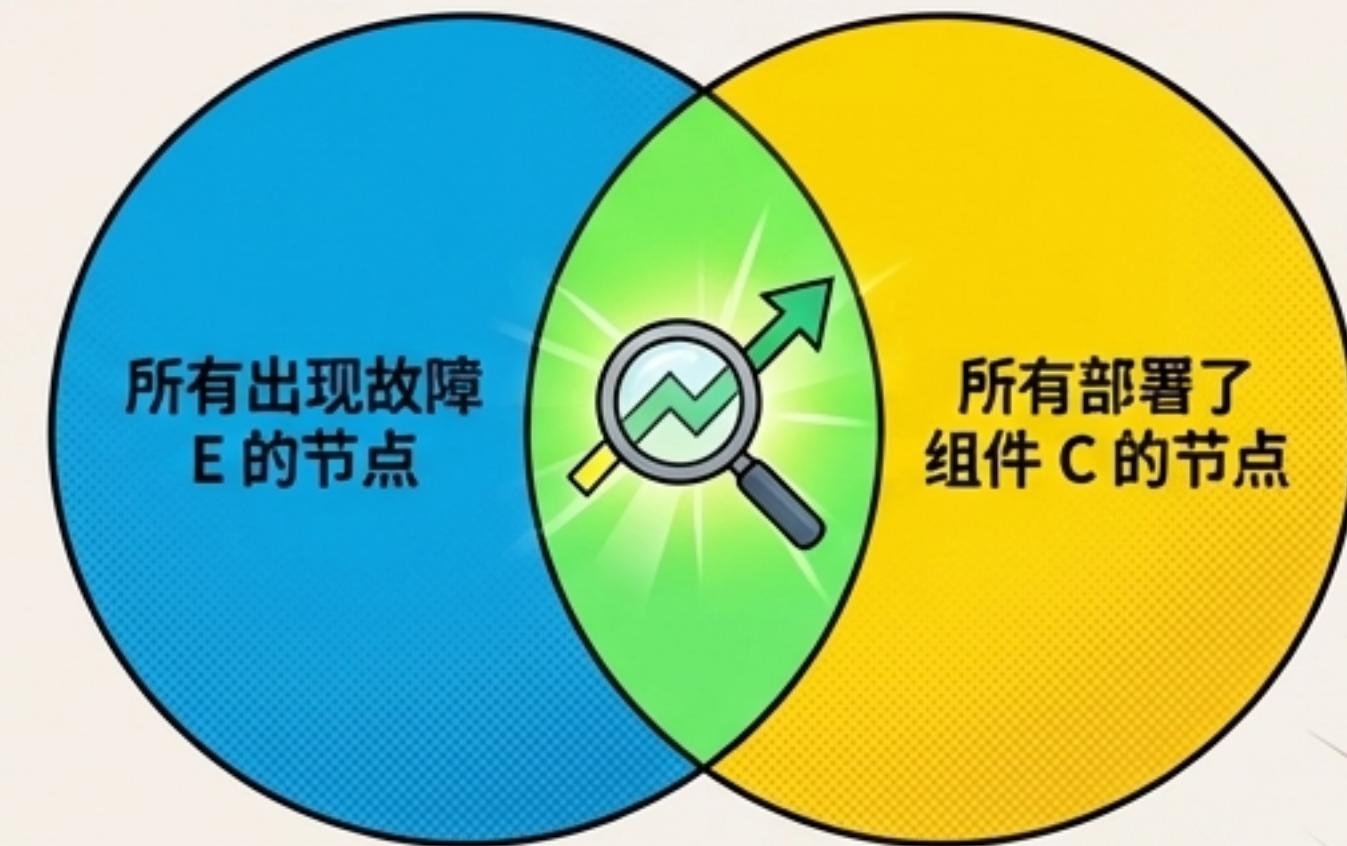
时序关联 (Temporal Correlation)



我们使用一种“投票-否决”机制。每个故障都会为它发生之前部署的组件“投票”。离得越近，权重越高（通过指数权重 w_i ）。

空间关联 (Spatial Correlation)

出现故障的节点，和部署了某组件的节点，重合度有多高？

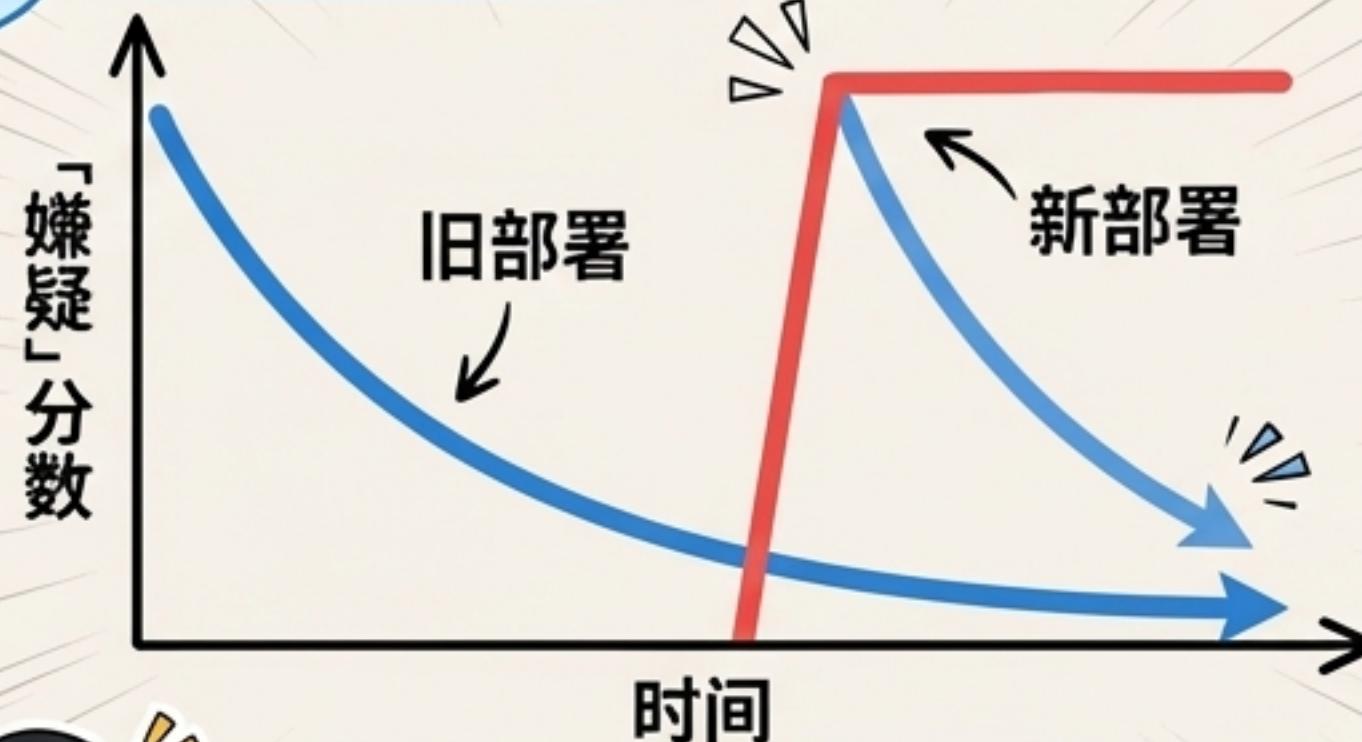


我们计算故障节点和部署节点的重合率 ($S_s = N_f / N_{df}$)。如果一个故障只发生在部署了某个组件的机器上，那么它们的关联性就非常强。

核心揭秘(2): 动态调整与专家智慧

时间衰减 (Time Decaying)

随着时间的推移，旧的部署引发新故障的可能性应该降低。



我们应用指数时间衰减因子，逐渐降低旧部署的“嫌疑”，确保系统能聚焦于新部署引入的问题。

$$\text{blame}(e) = \text{blame}(e) * (e^{-t} \dots)$$

领域知识 (Domain Knowledge)

并非所有故障信号都同等重要。专家的经验至关重要。



甘道夫允许开发人员根据经验自定义不同故障信号的权重。例如，将已知的良性故障权重设为0（加入白名单），或将高风险故障的权重调高，使系统更灵敏。



甘道夫的战绩：数据证明一切！

数据平面部署

92.4%

精准率 (Precision)

100%

召回率 (Recall)



控制平面部署

94.9%

精准率 (Precision)

99.8%

召回率 (Recall)

成功拦截155次严重故障，没有一次由
问题部署引发的重大服务中断 (Sev0-2)。

在1200+次区域级部署中，
仅有2次误报和2次漏报。

部署速度提升

部署时间
(甘道夫之前)



>50% 缩短!

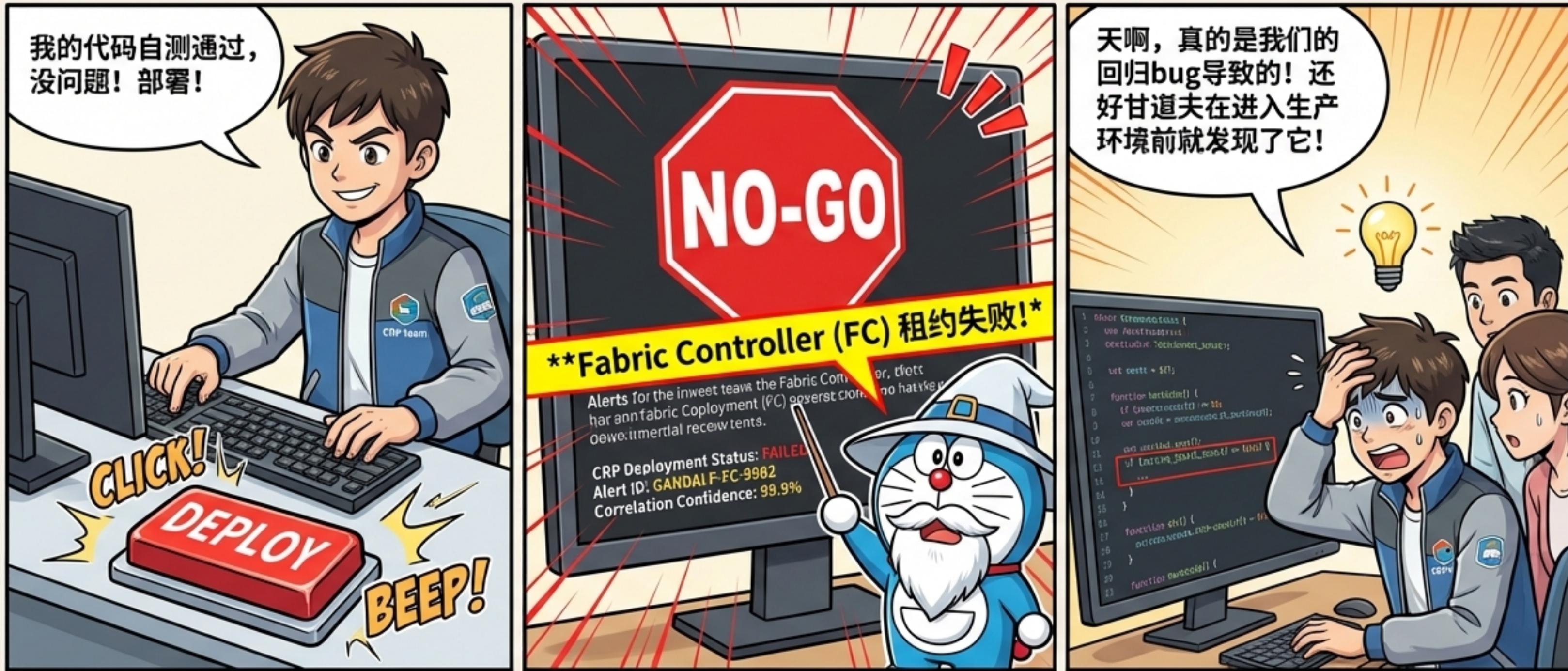


部署时间
(甘道夫之后)

部署到整个生产环境的时间缩短了超过一半！

实战案例 I：揪出“隐藏的敌人”

跨组件影响问题 (Cross-Component Impact Issue)



Lesson: 甘道夫的全局视角能发现人类专家基于经验也可能忽略的跨组件关联问题。

实战案例 II：唤醒“沉睡的恶龙”

潜伏性影响问题 (Latent Impact Issue)



Lesson: 甘道夫不仅能捕捉即时故障，还能通过长时间窗口分析，精准定位在特定负载下才触发的潜伏性问题。

从此，部署工作大变样



过去：证据分散，依赖邮件沟通和临时诊断，耗时耗力。



“甘道夫帮助我们的上线变得更好了。谢谢！”

—一位Azure工程师



现在：统一的可信信息源，自动化的决策，交互式的排错工具。



“从怀疑甘道夫的决策，到强制执行它的每一个‘No-Go’警报。”

—一位发布经理

甘道夫的成功秘诀



透明度和证据至关重要 (Transparency and Evidence are Crucial)

一个黑箱系统很难获得信任。甘道夫提供丰富的佐证，解释每一个决策背后的原因，从而建立起工程师的信任。



分析模型必须是适应性的 (Analytics Models Must Be Adaptive)

单纯依靠数据学习很难跟上系统演化的速度。我们与工程师紧密合作，持续将他们的领域知识融入甘道夫的决策模型中。



为不同需求量身定制 (Tailored for Different Needs)

不同团队对精准率和召回率的偏好不同。甘道夫允许定制，满足关键服务对100%召回率的严格要求，或帮助其他团队减少误报干扰。

有了甘道夫，
每一次部署，都通向更可靠的未来。

