Flashcat企业版套餐A

北极星

简介

北极星系统通过对业务核心指标的梳理、采集、计算、智能检测和报警，达到发现 真.故障，并驱动起故障处理流程的目的，是Flashcat故障发现和启动故障处理流程的入口。

电商系统北极星指标如：实时在线用户数、实时在线商品量、实时下单量、实时支付量、实时GMV等；出行系统北极星指标如：实时发单量、实时接单量、实时在线司机量、实时完单量、实时支付量等；各个行业的北极星指标均不尽相同，但都是该行业的核心业务指标。

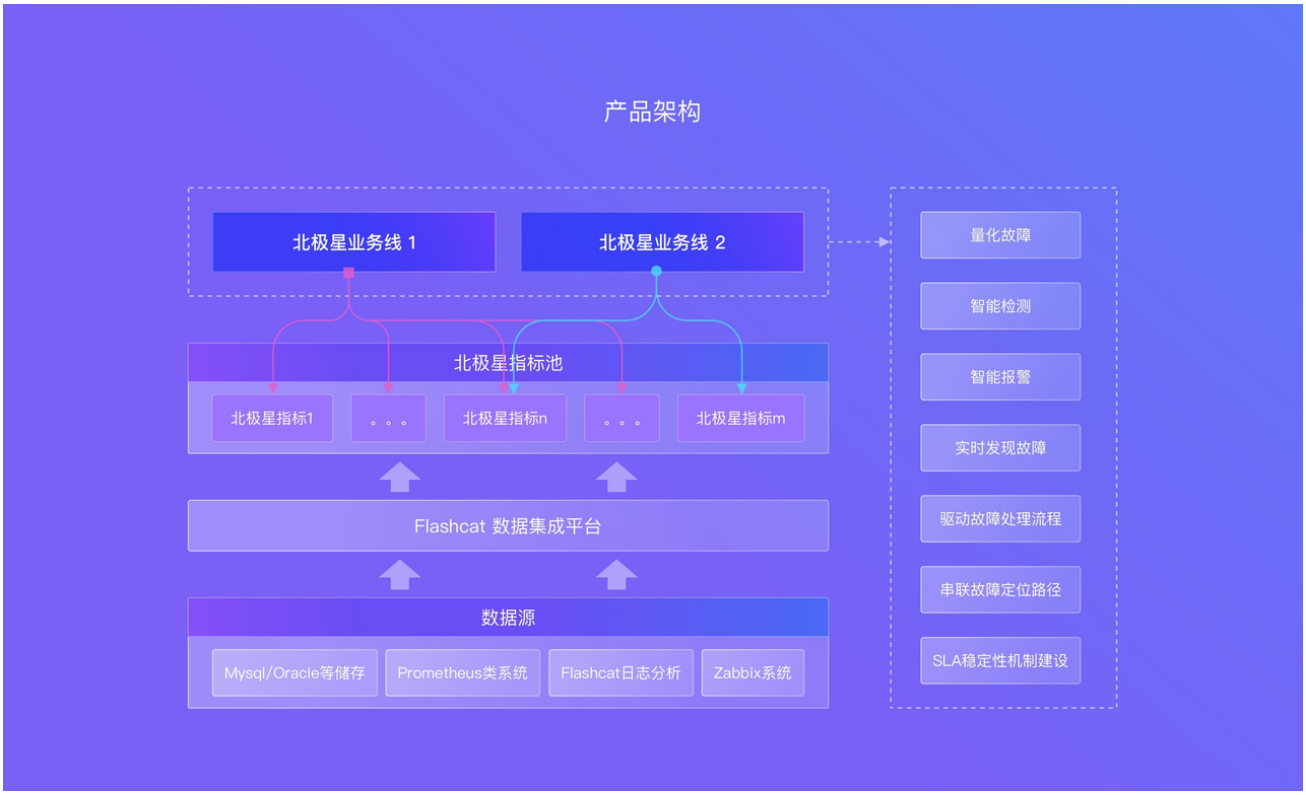
快猫云眼北极星系统通过产品设计和智能检测等功能，为这类指标提供VIP保障，用于量化整个业务系统的健康状态和稳定性保障工作的效果。

> 术语

概念	说明
北极星	该子系统的统称，主要由两层组成：业务线层和指标层。
业务线	通常对应企业内部的一个业务组织，如B2C业务线、B2B业务线、视频业务线、团购业务线、出行业务线等。
指标	北极星指标用于定义相应业务线的整体健康状态，由一个或多个时序数据组成。 北极星指标出现异常，则说明该业务线的核心价值受到了真正的影响。 北极星指标配置后是全局共享的，可以被关联进任意多个业务线，业务线和指标是多对多的关系。
智能报警	北极星智能报警分为智能检测和报警策略两部分。

智能检测	Flashcat支持检测4类异常： 越界异常 ： 系统会自动学习北极星指标趋势，生成预测模型，预测曲线未来实际值出现的区间，如实际值连续n次（默认值，可配置）超出预测的上界后下界，系统会输出越界异常事件； 同环比异常 ： 系统会将最新的数据和同比环比数据进行对比（默认为7天前和1天前，可配置），如偏差连续n次（默认值，可配置）均达到设定的幅度（默认值，可配置），则输出同环比异常事件； 数据中断异常 ： 系统周期性（通常是1分钟或30秒）对最新的数据进行检测，如连续n分钟（默认值，可配置）都没有检测到新数据，则输出数据中断异常事件； 绝对阈值异常 ： 用户可设置针对指标绝对值的异常阈值和条件（>/</=...）；
报警策略	智能检测系统输出的异常事件会自动输出给报警策略产生报警。报警策略配置的内容主要包括：报警时间区间、报警等级、报警渠道（支持电话、短信、邮件、IM等）、报警接收组等。报警策略可设置业务线共享的全局策略，和某个指标单独使用的个性化策略。
稳定性配额	北极星业务线可设置业务线全年的不可用时长配额，即允许业务故障出现的总时长，是业务稳定性管理的基础。 该功能属于进阶功能，启动时可暂不关注，待北极星指标完善、故障发现正常产生后再实施。

产品架构



落地步骤

目标：完成北极星指标的梳理、配置、报警，达到及时发现故障，驱动起故障处理流程的效果

步骤一：确定北极星业务线

北极星的业务线可根据企业内的业务组织来设置，一个业务线通常对应一个业务负责人（GM）。 如企业内的电商业务线、视频业务线、旅游业务线。也可视企业内的业务规模和划分情况做更细的拆分设置，如电商业务可能拆分为B2B业务线和B2C业务线。

参与方：北极星的落地需要由业务稳定性的负责方来lead推进。同时，北极星的核心工作之一是确定业务相关的北极星指标，因此建议业务负责人或业务人员参与进来。

重要提示：

- 建议先选取一条核心及成熟的北极星业务线进行试点，待该业务线的整个流程落地顺畅后再推进更多业务线的落地。
- 不建议选取新的、发展还不成熟或规模较小的业务线进行试点，这类业务线的北极星指标通常还不稳定、趋势还不明显，不利于对其进行观测和报警。选择这类业务线试点很可能无法充分发挥北极星系统的价值。

输出：企业内的北极星业务线，以及参与试点的业务线和相应的业务人员。

步骤二：梳理北极星指标

- 指标选取：北极星指标通常是一个“量”相关的指标，如实时订单量、实时GMV、实时支付量、实时在线用户数等。这类指标非技术人员也能够理解其含义，并“直接”知道其出现异常后对业务意味着什么，是业务负责人最关心的业务指标。
- 指标来源：北极星指标通常来源于业务的线上存储系统，如Mysql、Oracle等，也可选其它可靠来源，如prometheus、夜莺、Flashcat多维分析系统等。
- 北极星指标梳。

说明：北极星指标梳理可一次性梳理完，再集中配置。也可采取先完成部分甚至一个核心指标的梳理和配置，后续再由相关人员在平台自行新增和调整。

重要提示：

- 不建议选取如“订单系统请求成功率”这类指标作为北极星指标，主要有2点原因：1）请求成功率下降未必意味着用户请求的最终失败，可能系统重试后成功，或系统有其它容灾降级措施，甚至这个下降可能只是系统拒绝了攻击类或线下非预期连上来的请求；2）业务同学知道这个异常后仍然会问这个异常对业务的影响是什么这样的问题。这类指标和业务的直接感知仍然隔了一层；
- 系统级的成功率这类指标在Flashcat里建议配置到灭火图里，作用于故障定位环节，将在"灭火图"系统中介绍。
- 北极星指标有可能因现实的采集难度，导致暂时无法采集到最合适的指标，则可以考虑结合现实情况先选择一个替代的指标，待效果运行起来后再逐步优化为能够准确量化业务健康的北极星指标。

以上建议不是一成不变的，要视业务的差异和数据采集的难度来权衡，可遵循先落地后优化的原则。

输出：业务线的北极星指标及数据来源。

步骤三：配置北极星业务线和指标

- 北极星业务线：配置较为简单，其中的可用性目标和不可用时长管理启动时可不填写，在进阶实践部分介绍。
- 北极星指标：北极星提供了将数据库（Mysql、Oracle等）数据转换为北极星时序数据的配置功能。也支持来源于prometheus、Flashcat日志分析系统的指标。

输出：完成北极星业务线以及相应北极星指标的配置。

步骤四：开启北极星智能报警

指标梳理并配置完成后，下一步就是开启报警。北极星智能报警分为 `智能检测` 和 `报警策略` 两部分，智能检测输出异常事件，是报警策略的输入来源。

- 智能检测：
  - 北极星指标配置完成后，Flashcat会自动学习指标的趋势（通常在1周左右），并预置智能检测参数。这部分参数正常情况下无需用户关注或修改。这一步用户重点观察校验北极星指标的数据正确性和连续性。待数据和趋势稳定后Flashcat会自动输出数据抖动的异常事件；

> - 除了智能检测，Flashcat还支持开启同环比检测：通常采用Flashcat的默认值即可，作为兜底用；绝对阈值检测：针对类似0~100%的百分比指标适合使用；数据中断检测：核心指标建议开启该检测，默认为15分钟；

- 报警策略：

> - 业务线的报警策略分为共享策略和个性化策略。通常情况下只要配置一个共享策略，然后结合系统的推荐决定在该策略下开启哪些指标的报警即可。
> - 对于不适合使用共享策略的个别指标（如指标适合开启报警的时间和大部分其它北极星指标不一致、报警接收组不一致等情况）可以为其设置单独使用的报警策略。

重要提示：

- 部分指标不适合开启或配置报警策略，包括的情况有指标离散、趋势不固定、抖动明显，这类指标人工也很难判断异常与否，容易产生大量不必要的报警，对北极星报警造成狼来了的负面效应。这类指标可作为辅助观察，但不开启报警策略。

输出：完成业务线内报警策略配置，开启北极星指标报警开关，正常接收并响应指标报警。

最佳实践

北极星的核心目标是量化并发现故障，并驱动故障处理的流程。基于北极星的故障发现最佳实践举例如下。

实践一：建立完善的故障响应机制

1）北极星指标报警准确稳定后，建立业务线的稳定性保障群，加入业务负责人、技术负责人、SRE等稳定性保障相关人员；2）报警接收组里配置IM机器人，将报警事件发送到保障群中；3）可同时配置电话报警渠道，以便相关人员任何时段都能及时高优的感知和响应；4）北极星指标报警后，确保稳定性保障的相关人员都感知了故障的发生，进入故障定位状态；5）处理过程中，根据北极星指标的情况，向业务负责人和保障团队通报故障影响和趋势；6）处理完成，根据北极星指标整体受损情况，向业务负责人和保障团队通报故障整体影响；7）安排故障复盘，基于北极星指标的影响确定故障等级；

实践二：联动变更及时发现并控制故障

据统计，故障大部分是因变更引起的，而最多的变更就是线上程序发布。有了准确的北极星指标报警后，可以联动发布系统和北极星报警。一旦相应业务线的北极星指标产生报警，发布系统可通过查询或报警回调感知，立即阻断该业务线的所有变更进程，提示用户进行检查或立即执行回滚，以防止故障影响扩大并快速止损。

SLA管理实践

目标：基于北极星开展业务线级别的SLA运营管理

本进阶工作本质上是建立一套稳定性的运营机制，以达到让稳定性保障工作得到有效支持和可持续推进。能够完成北极星指标的梳理、配置、报警，达到及时发现故障，驱动起故障处理流程的效果，就实现了北极星系统的核心目标。该进阶工作建议在达成以上效果后再进行。

步骤一：概念统一 故障等级通常涉及几个类似和容易混淆的概念，如故障、异常、事故、事件、问题，最好在企业内部先统一这些概念的定义，以便降低沟通的成本。

> Flashcat中的概念：
>
> - 事件：是一个范围最大的概念，包括了异常、故障、事故、问题，以及服务正常情况下的一些关键事件，如大促导致的业务量正常突增等；
> - 异常：只要业务受损即为异常，轻量的异常可以是一次用户的正常请求失败，严重的异常可以是重大事故。其包括了“事件”中除服务正常情况下以外的事件；
> - 问题：轻量的异常，未达到触发北极星指标报警的条件；
> - 故障：达到一定严重程度的异常，触发了北极星指标的报警；
> - 事故：程度严重或较为严重的故障，可进一步量化区分；
>
> 基于以上定义：
>
> - 事件 = 日常事件 + 异常
> - 异常 = 问题 + 故障
> - 故障 = 一般故障 + 严重故障（事故）

其中 `故障` 是一个重要的分水岭，故障及以上的事件是需要紧急响应紧急处理的，即Flashcat重点针对的场景，而故障及以下的事件则不那么紧急，属于日常排障。

步骤二：基于SLI量化故障标准 1）SLI选取：

北极星指标即SLI（service level indicator）。但为便于观察，北极星业务线中采集的北极星指标通常可能会有一些重合，如总订单量指标和分地域或分城市的订单量指标。可依据业务的情况，从中选取部分指标来作为故障等级判断的依据。

选取原则：所选指标集合能够完整覆盖业务的核心流程；

2）等级量化：

基于选取的北极星指标（SLI）集合推荐一个异常等级量化的样例：

异常等级量化举例：

- P0 重大事故：任一SLI在故障期间累计量下降超过50%，且故障持续了30分钟以上；或累计下降量达到日总量的10%以上；
- P1 严重事故：任一SLI在故障期间累计量下降达到20%~50%，且故障持续30分钟以上；或累计下降量达到日总量的5%~10%；
- P2 重要事故：任一SLI在故障期间累计量下降达到10%~20%，且持续30分钟以上；或累计下降量达到日总量的1%~5%；
- P3 一般事故：任一SLI在故障期间累计量下降超过10%，但持续不到30分钟，且累计下降量不超过日总量的1%；
- P4 一般故障：任一SLI异常，触发了北极星报警，但不满足以上任一条件；
- P5 一般问题：服务出现一定的异常，但未达到触发北极星报警的程度，如日常bug等导致的局部问题。也可以叫做“一般异常”。

定义中的相关概念：

- 累计量：一定时间区间内时序数据每一个点对应值的加和；
- 故障期间：指标曲线开始异常到指标完全恢复到正常水平的时间区间；
- 下降幅度的参照值：计算故障开始前30分钟的累计量相对7天前同一时间区间的累计量的涨幅（正或负），7天前曲线对应故障区间时间内的累计量 x（1+涨幅），以此为参照值；

输出：明确量化的故障等级标准文稿。

步骤三：可用性配额（Error Budget）管理

可用性配额或错误配额，即全年允许故障持续的总时长，是可用性目标（SLO，service level objective）计算的基础，如可用性配额为262.8分钟，对应可用性目标则为99.95%。

其实也可以反过来，先确定全年可用性目标，录入目标值，Flahcat会自动计算出可用性配额。但配额值其实更具业务意义，因此建议采用先配额后目标，再微调的制定方案。

首先，确定需要扣减可用性配额的故障等级，如P3一般事故及以上的等级需要扣减配额。各个等级扣减配额的方案有多种，可选方案举例：

1）只要达到P3等级，无论严重程度，一律扣减异常持续时间的时长；2）按等级进行折算，如：

- P0：从配额中扣减 事故持续时长 x 100% 的时间；
- P1、P2、P3：从配额中扣减 事故持续时长 x 异常时间区间累计量下降幅度 的时间；

可以结合业务的情况，制定一个合理的算法。

扣减算法确定后，再基于业务线历史的故障情况，和未来的发展预期，大致推算一个可用性配额的量级，最后可能权衡微调为对应99.9%、99.95%、99.98%之类的可用性目标值。

输出：明确量化的可用性配额扣减标准，及各个业务线的可用性配额数值。

步骤四：可用性目标（SLO）管理

业务线全年的可用性配额确定后，可在Flashcat的北极星首页进行录入，Flashcat会根据录入的配额自动计算全年的可用性目标。

> 计算公式（单位：分钟）：
> 可用性目标 = （365 * 24 * 60 - 配额）/（365 * 24 * 60）* 100%如，不可用时长配额为262.8分钟，则计算的可用性目标即为99.95%。

后续每次故障复盘中，需要确认故障的等级和消耗的配额，并累加到已消耗的配额中，系统会根据已消耗的配额每天自动计算当前可用性。

> 计算公式：
> 当前可用性 = （1月1日到当前的天数 * 24 * 60 - 已消耗的时长）/（1月1日到当前的天数 * 24 * 60）* 100%

系统同时计算预估可用性，即假设从当前时间开始到年末都不出现消耗配额的事故，预期保持到年底时的可用性数值（可用于和可用性目标做对比）。

计算公式：

预估可用性 = （365 * 24 * 60 - 已消耗的配额）/（365 * 24 * 60）* 100%

重要提示：

1）可用性目标因为SLI的选取不同、故障等级的标准不同、配额扣减的标准不同，通常无法在不同标准体系下做横向比较。较有意义的是现在和过去、今年和往年进行对比。但前提是持续管理，并保持标准的一致性。

2）相关术语：SLA = Service Level Agreement = 服务水平协议SLO = Service Level Objective = 服务水平目标SLI = Services Level Indicator = 服务水平指标Error Budget：基于SLO计算的错误预算

输出：明确的可用性配额录入到北极星业务线中，被在每次事故后更新消耗的配额。

> SLA管理最佳实践

最佳实践：基于北极星的业务稳定性SLA管理最佳实践

1）成立公司级（至少是业务线级别）的稳定性联合保障组织，对公司/业务线的服务稳定性负责；2）定期总结各北极星业务线的稳定性指标达成情况、故障情况、改进项完成情况等数据，作为运营报告输出；3）对稳定性建设表现优秀的团队进行奖励，对可用性配额消耗大的团队进行复盘总结；

灭火图

简介

灭火图系统通过观测IT系统的核心功能和核心模块、服务组件、基础设施，快速收敛故障范围，并关联后续的多种故障定位渠道，引导用户下钻完成故障定位过程，是Flashcat里进入故障定位环节的入口。

术语

概念	说明
灭火图	该子系统的统称，主要包括系统层（包括功能和模块）、组件层、基础设施层。
系统	灭火图的系统，可以按企业内部技术团队的组织形式来设置，通常一个北极星业务线会对应多个灭火图系统。如北极星团购业务线可能对应灭火图团购业务系统、中台系统、地图系统、订单引擎系统等。 Flashcat提供将北极星业务线和灭火图进行关联的功能（系统级别和功能组/模块组级别）。 任一灭火图系统都有 功能 和 模块 两个观测维度。
功能	功能可以理解为IT系统的API或接口，表现形式通常为：域名/ip + path。如下单接口、支付接口、查询接口等。 灭火图观测功能的三个黄金指标(黄金指标SLO)：请求成功率、请求量、请求延迟（平均值及分位值）。 也支持只观测功能的某一个维度，指标由用户自定义（自定义SLO）。
模块	模块是研发编写并部署在线上运行的程序，通常一个模块由一个或多个实例组成，对应如kubernetes中的一个个service、deployment、statefulset等。 模块对外提供API或接口（即功能），是功能实现的物理支持。 灭火图观测模块的实例存活率（卡片异常飘红的依据）、cpu/mem/disk使用率指标。

功能组/模块组	相关的 功能 或 模块 可以在灭火图中加入到同一个功能组或模块组中，方便统一观测，并做异常相关性的判断。 一个功能组 或 模块组通常可对应一个相对独立的IT子系统，如券子系统、库存子系统、用户子系统等。
组件	即通用的服务组件，如mysql、kafka、elasticsearch、zookeeper等。
基础设施	即公共的基础设施，如外网运营商网络、内网网络、CDN、DNS等。
飘红	灭火图的核心目标是全局观测服务在IT层面的健康状态，并快速收敛范围，这个特点的主要实现途径就是飘红，灭火图可以针对观测对象设置组合的飘红条件，并层层上传这个飘红信息，以便在首页即可完成异常面的观测。
时间轴状态	灭火图会记录所有功能和模块在过去x小时内的飘红记录，飘红的卡片越多时间轴状态越红（类似地图路况图），点击时间轴可以回溯灭火图的历史状态。

产品架构

落地步骤

目标：选择一个重要系统，完成灭火图“功能”层核心信息的配置，使灭火图基本运行起来

步骤一：确定灭火图“系统”

灭火图首页的“系统”可以根据企业内部的技术团队组织情况来进行设置，如：B2C电商业务系统、B2B电商业务系统、出行业务系统、团购业务系统、订单引擎系统、公共中台系统、地图系统等。

参与方：灭火图的落地需要由业务稳定性的负责方来lead推进，灭火图配置的功能和模块主要是系统级的服务信息，因此需要研发/测试或SRE，或对线上信息很了解的人员参与进来。

重要提示：

- 灭火图是IT系统层面的信息，通常一个北极星业务线会有一个对应的技术团队为该业务研发业务系统，但支持该业务的可能还有如业务中台、订单引擎等公共技术团队。因此，北极星的“业务线”和灭火图的“系统”是一对多的关系。

输出：北极星业务线对应的灭火图“系统”，以及参与进一步工作的研发、测试或SRE人员。

步骤二：梳理灭火图“系统”指标

- 指标选取：指标梳理的关键实际是梳理核心功能和核心模块，即确定哪些功能和哪些模块需要进入灭火图。
- 指标来源：功能指标在平台上目前支持来源Flashcat的日志分析（网关或模块日志）、prometheus类数据源、zabbix数据源；模块指标在平台上目前支持prometheus类数据源；同时，功能和模块的信息都支持从API调用导入；

重要提示：

- 建议从某一核心“系统”开始梳理，并优先建设“功能”层面的灭火图信息，以后再逐步扩大功能的范围，并增加“模块”层面的信息。
- 建议只配置核心功能和核心模块，不建议把一个“系统”的所有功能和模块都配置到灭火图，这会干扰对真正的异常源的判断。

输出：灭火图各“系统”的核心功能及核心模块（可选），以及相应指标的来源。

步骤三：配置“系统”及指标

- 灭火图“系统”：配置较为简单，在首页新建即可。
- 功能和模块：在web平台支持单独新增和批量新增两种模式，建议采用批量新增的方式，更为准确高效。如相关信息已在其他系统中有维护，则建议通过API导入并保持动态更新。

配置建议：

- prometheus的指标可配置多种标签，如namespace、business、module、uri等，这些标签可以在模块的部署过程中（如注入k8s的pod anotation）或在指标采集时注入。
- 如果prometheus中的标签包含了系统、模块、接口、等级等信息，就可以在灭火图中的批量新增中，通过这些标签将功能（接口）或模块批量筛选出来，一次性添加到灭火图中。
- 这样无论是研发或SRE只要在源头或指标采集时管理好指标的标准和注入的流程，灭火图就实际成为了prometheus的“服务树”。
- 相应的筛选配置后续平台支持定期执行，则“服务树”就具备了动态更新的能力。

输出：完成灭火图“系统”以及相应功能和模块（可选）的指标配置。

步骤四：飘红设置和治理

灭火图的功能和模块配置过程中提供了默认的飘红阈值，目前该阈值是静态阈值，配置完成后需要对该阈值进行常态的治理调整，以达到最佳的状态。

重要提示：

- 灭火图的理想效果：如北极星指标出现报警，则相应的灭火图一定有飘红，否则，北极星指标正常，相应的灭火图都应该是绿色的。
- 灭火图的飘红阈值不建议设置的过于严格，否则可能导致飘红过多影响判断。对于容错率高的业务，功能成功率建议设置到90%及以下，模块存活率视实例的数量，可以设置为如80%。

输出：常态巡检灭火图，发现系统隐患的同时，对卡片飘红阈值进行调整治理。

最佳实践

灭火图使用最佳实践

- 确定支持某北极星业务线的“IT系统”，如xx业务线系统、中台系统、用户系统、引擎系统、地图系统等；
- 在灭火图首页创建这些IT系统（如该系统已存在则不需要重复创建，如中台系统、用户系统等公共系统可以在首页只创建一个卡片）；

- 完成对应IT系统的核心功能和核心模块的梳理；
- 在相应系统卡片内完成核心功能和核心模块的单独/批量录入，设置飘红阈值；
- 在灭火图“系统”内将相关的功能或相关的模块放到一个组里，如发单相关的功能或支付相关的模块等；
- 日常观察校验数据是否准确，治理飘红阈值；
- 北极星指标报警后，处理人员通过配置的关联查看对应灭火图的情况，或直接在灭火图首页查看全局“系统”的状态，收敛故障在IT系统层面的范围，确定需要重点跟进的团队；

完成选定IT系统的核心功能配置工作，灭火图基本可以启动并逐步完善起来。下面是进一步完善的进阶部分，可逐步推进。

> 进阶实践

目标：进一步完善灭火图数据，并打通灭火图和北极星等周边系统的关联，串联起故障定位的路径进阶部分包括完善灭火图的模块信息（如之前的步骤中未涉及）、服务组件信息、基础设施信息和关联信息。

步骤一：配置模块信息

功能最终是运行在线上的一个个程序实例以及他们之间的相互调用来实现的，完成相同功能的实例即构成模块。发现功能异常后一般进一步的定位思路是查看模块的情况，如模块的实例存活状态、模块的资源使用情况等。因此将模块信息配置到灭火图将非常有利于故障定位的进一步深入，甚至可以在灭火图一眼确定故障的“根源”。

相应的配置方案在前面已基本介绍，这里不再描述。

输出：灭火图模块卡片信息，可观测相应模块的实例存活率和各维度的资源使用情况。

步骤二：配置组件信息

组件服务的健康状态是定位服务故障的重要信息来源，组件灭火图的基本配置逻辑：

- 在Flashcat的仪表盘系统中，一键导入某类组件的大盘（前提是该组件的信息采集是通过标准的exporter/grafana-agent/categraf采集）；
- 新增组件灭火图时，首先选定相应的组件大盘；
- 选择组件的单元标签，如clustername等，系统将根据该标签识别组件单元；
- 设置每个单元的健康指标（可设置多个）以及异常的条件；
- 设置组件单元中每个实例的健康指标（可选）；
- 创建组件的灭火图，灭火图将根据提交的信息生成组件单元，并标识每个单元的状态，同

时每个单元可继续带参数下钻到选定的组件大盘，以便问题定位时深入分析；

输出：每类组件都完成标准的信息采集，生成组件大盘，并按管理单元生成组件灭火图信息。

步骤三：配置基础设施信息

基础设施包括内外网网络、CDN、DNS、主机、容器、容器平台等，基础设施如出现异常，通常就是故障的直接“根源”。

基础设施的健康状态配置方式和组件灭火图的基本逻辑类似：

- 在Flashcat的仪表盘系统中，一键导入某类基础设施的大盘（前提是该组件的信息采集是通过标准的exporter/grafana-agent/categraf采集）。或者针对基础设施进行专项数据采集，并生成prometheus指标。
- 创建基础设施的大盘信息。
- 新增基础设施灭火图时，首先选定相应的组件大盘；
- 选择基础设施的单元标签，如region等，系统将根据该标签识别基础设施单元；
- 设置每个单元的健康指标（可设置多个）以及异常的条件；
- 基础设施一般不存在实例概念，因此不需要指定实例指标；
- 创建基础设施的灭火图，灭火图将根据提交的信息生成基础设施单元，并标识每个单元的状态，同时每个单元可继续带参数下钻到选定的基础设施大盘，以便问题定位时深入分析；

输出：每类基础设施都完成标准的信息采集，生成基础设施大盘，并按管理单元生成基础设施灭火图信息。

步骤四：配置关联信息

典型的故障定位过程是从业务、到功能、到模块、到组件、到基础设施。当然，也可能直接从基础设等底层对象的健康状态得出一次北极星异常的原因。如能将这些信息从上到下关联起来，特别是北极星、功能、模块、组件、大盘、事件这些信息的关联和最佳实践路径的引导，将非常有助于故障定位效率的提升和故障定位门槛的降低。因此，北极星、灭火图都提供了串联各相关信息的功能，将故障定位的最佳实践和老司机的经验沉淀到系统。

北极星的业务线内和灭火图各层卡片上都提供了关联配置的入口，可以根据系统内各元素的实际关联和定位经验进行配置。

重要提示：

- 关联的配置是一个逐步累积的过程，不需要一次性完成全部的配置，可以在日常故障定位的过程中和故障的复盘后逐步迭代完善，并进入一个迭代和价值发挥的良性循环。

输出：逐步在系统中沉淀故障定位的最佳实践和最佳路径。

重要提示：

- 灭火图的配置及关联的配置都支持api接口调用，可批量操作，或通过该方式联动如cmdb等系统。

通过以上配置，整个IT系统的健康状态和故障定位信息将逐步完善。在确定故障范围和责任团队的基础上，故障定位人员将可以在相应的灭火图对象上直接点击调出故障定位所需的下一步信息，引导完成整个故障定位过程，进一步提升故障定位的效率。

智能告警

时序数据异常检测简介

对于所有的在线业务，都会随着时间产生一些数据，这些数据我们称为时序数据，在服务正常的时候，这些时序数据的变化会符合一定的模式，我们可以根据这些时序数据的变化，来判断我们服务是否出现了异常。业内目前主要有三个方式来判断时序数据是否异常：

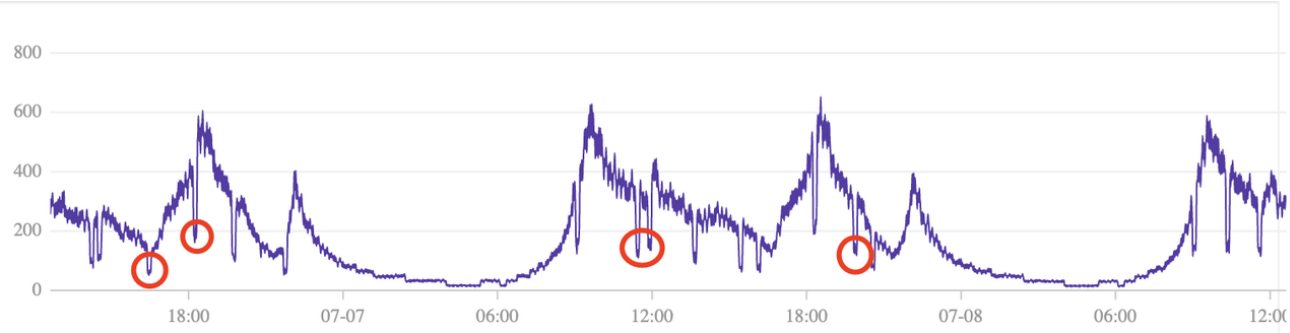
- 第一种是有值班人员实时盯着重要的时序数据，根据经验来判断时序数据是否出现了异常
- 第二种是使用监控产品，给关注的时序数据配置一个静态的阈值，如果超过阈值就表示时序数据出现异常
- 第三种是近几年出现的新的方式，使用机器学习的能力，动态学习时序数据的规律，实时计算动态的阈值，识别是否异常。

目前业界主流的方式是使用配置静态阈值来判断，但随着业务发展，这个方式也开始出现一些问题，下面介绍下传统静态阈值告警遇到的问题。

静态阈值可能遇到的问题

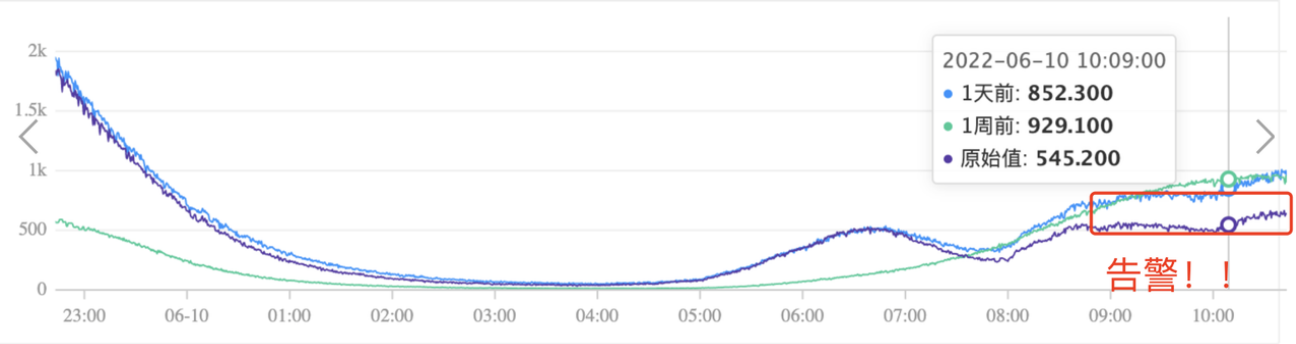
01.静态阈值覆盖场景有限，业务类监控数据不适用

业务类监控数据，使用静态阈值很多情况下不能很好的标识是否异常，比如下图的曲线，常见的业务数据都有这些特点，峰值和谷值差距很大，如果上限阈值配置是600，下降阈值配置10，那图中红圈标记的异常就会出现漏报。



02.阈值会由于特殊日或业务发展产生变化

业务监控指标经常会由于“特殊日”（节假日、营销活动日）或者业务发展影响而产生变化，传统的静态阈值或同环比策略在这种场景下，会产生多次误报，给负责稳定性的同学造成不必要的打扰，像下图的情况，紫色曲线是当天的监控数据同比1天和7天都低很多，但属于正常情况，这个在静态阈值的同环比策略下则会发出误报。



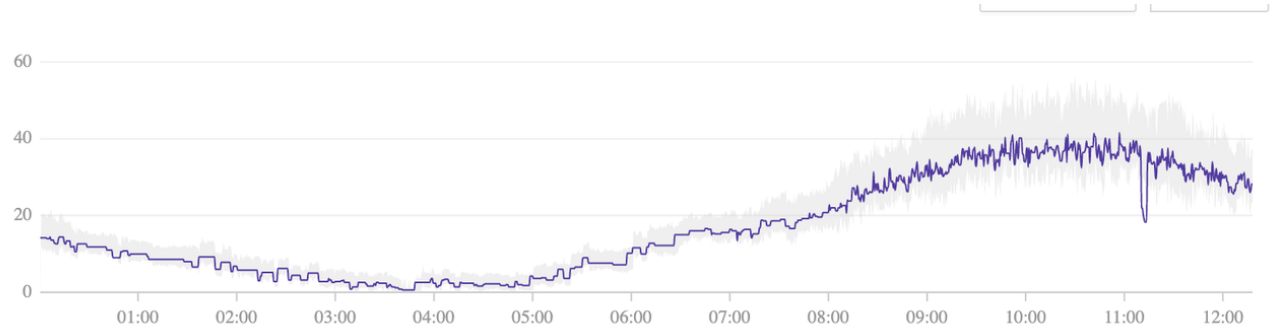
03.传统静态阈值的设置，依赖专家经验，人力维护成本高

下图是静态阈值告警配置常见的流程，经过几轮调整之后，才可正常使用，而随着业务增长，仍然需要不定期调整阈值，人力维护成本高。

智能异常检测的优势

智能异常检测基于机器学习算法模型实时生成动态基线，可以有效避免传统阈值方式造成的误报问题，也摆脱了对专家经验的依赖，提升了告警准确率，也提升了值班同学的幸福感：）

下图是智能异常算法实时计算出来的动态基线，会随着业务增长动态变化



下图总结了静态阈值和智能算法的区别：

对比项	静态阈值	智能告警
准确率	一般	高
维护成本	高	低
实现成本	低	高

哪些场景适合智能异常检测?

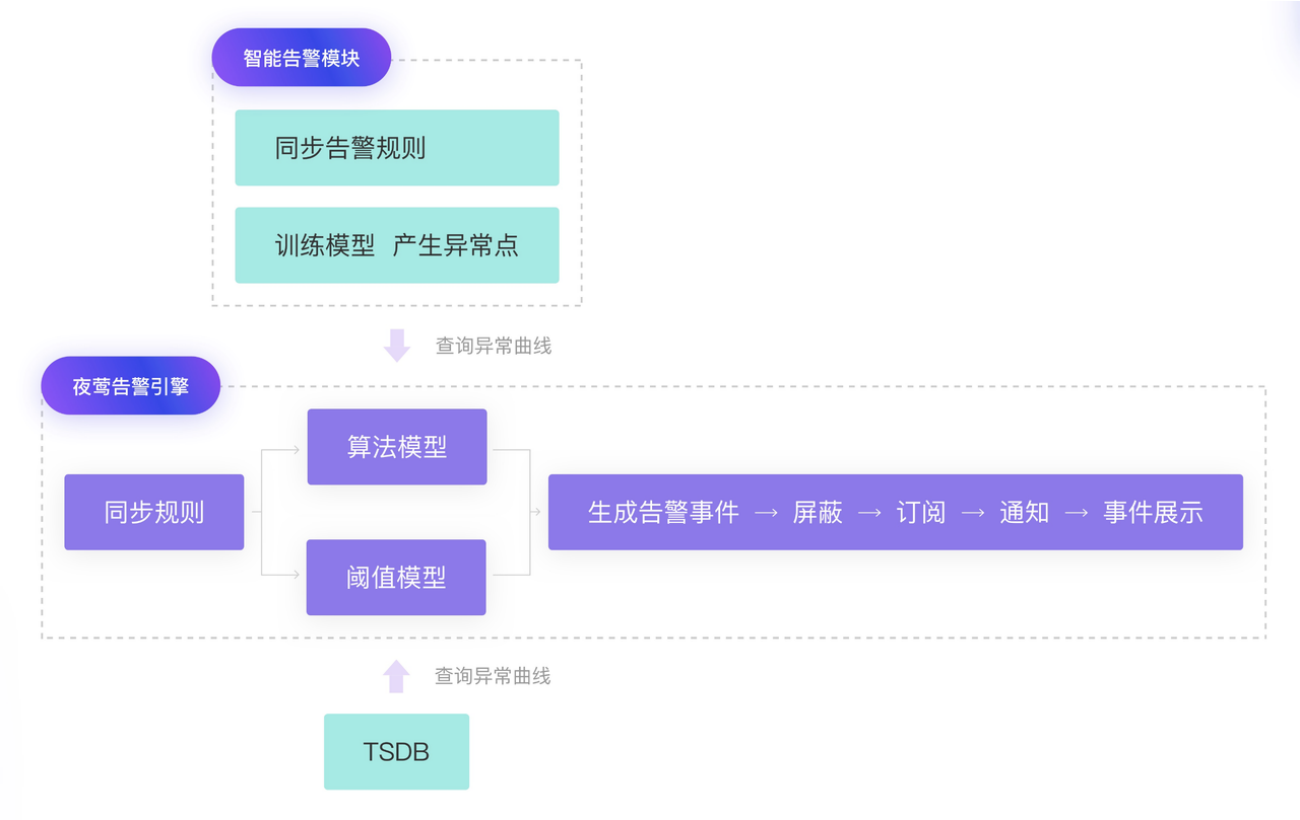
智能异常检测相比静态阈值的规则，有很多优势，对于有周期性的时序数据尤其合适，以下列举了一些常见的场景：

- 网页浏览量
- 活跃用户数
- 应用下载量
- 购物下单量
- 证券交易量
- 打车呼叫量
- ……

前面介绍了智能异常检测的优势和适用场景，那如何落地呢？下面介绍下夜莺的落地方案。

夜莺的智能告警落地方案

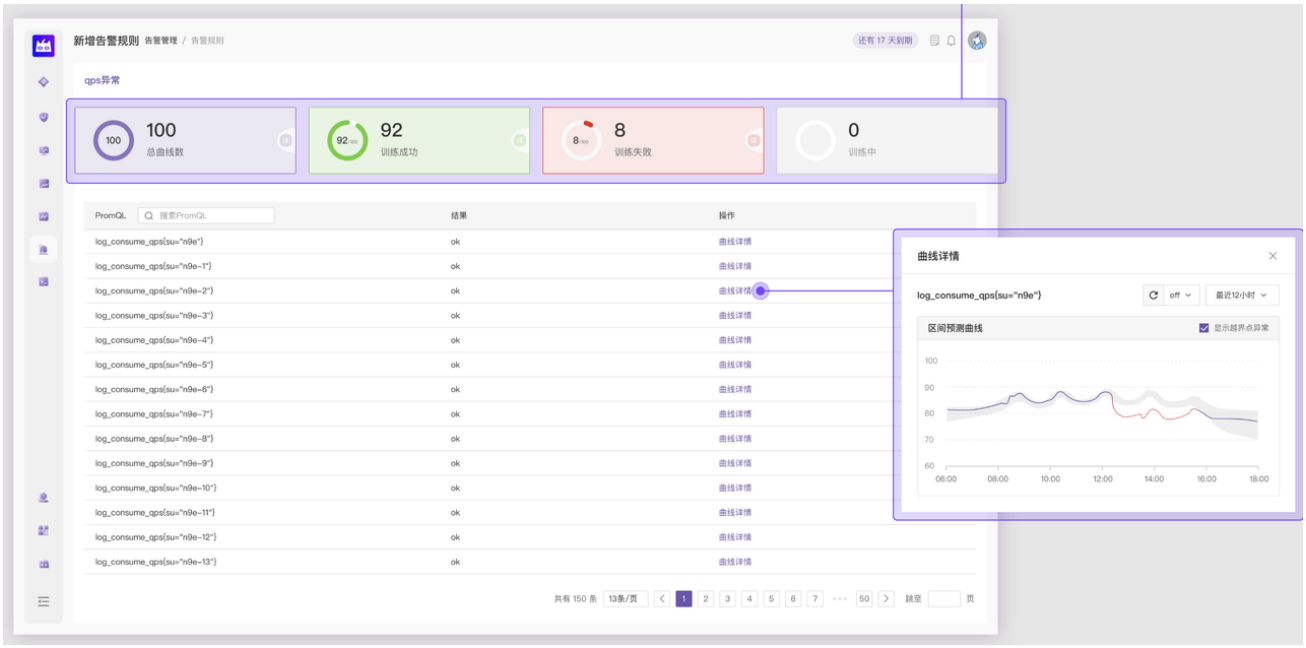
如果之前使用了夜莺，再部署一个智能异常检测模块即可，可以和开源的夜莺监控无缝集成，整体架构如下图



智能异常检测模块完成安装之后，在夜莺告警规则配置页面，会多出一个智能告警的选项，如下图所示：

选择智能告警之后，只需填写要监控的指标，不需要填写阈值，点击保存即可，之后在告警规则列表页，智能告警的规则右侧会有一个“训练结果”的按钮

点击“训练结果”，可以进入训练结果详情页，点击曲线详情，可以看到曲线学习出来的动态基线。如果曲线偏离到基线之外，夜莺的告警引擎会发出告警通知。


新增告警规则
告警管理 / 告警规则
还有 17 天到期
qps异常
100
总曲线数
92
训练成功
8
训练失败
0
训练中
PromQL
搜索PromQL
结果
操作
log_consume_qps{su="n9e"}
ok
曲线详情
曲线详情
log_consume_qps{su="n9e"}
off
最近12小时
区间预测曲线
显示越界点异常
100
90
80
70
60
06:00
08:00
10:00
12:00
14:00
16:00
18:00
共有 150 条
13条/页
跳至
页