

研发效能 【稳定性】浅谈团队如何做好系统稳定性

京东云开发者 · 2024年04月09日 · 2616 次阅读



背景

稳定性建设需要一系列具体的建设活动推进和落地，这些建设活动涉及人员、机制和文化，全方位的建设活动才能更好地落实建设模式。

一、稳定性保障机制

稳定性涉及团队所有不同水平技术人员、所有系统、研发所有环节、线上时时刻刻，单个技术人员是无法保障好的，必须建立团队流程机制来可持续保障。

人为因素的根源一方面专业能力不足，经验不足，另一方面很多都是无心之失，所以需要通过流程、规范来保住“底线”，减少人为因素导致的故障。大家严格遵守咱们的各种规范即可

(CodeReview 规范、发布 xbp 流程、上线后 doublecheck 机制)。通过流程和 doublecheck 机制确保每个人发布不会太差，解决人的因素。永远要记住团队的力量是无穷的，要学会借力。

1、规范先行

稳定性关键还是要靠大家，如何靠大家呢？稳定性工作，规范先行.就是要落地一套稳定性的机制体系，用机制的严格执行来约束大家，不然无法展开。这套机制包括：

- 方案评审机制：在完成系统的建设或改造方案初稿后，需通过由业务、技术、测试、运维领域组成的团队进行方案评审，才能进一步对方案进行实施。
- 架构设计规范：概要设计、模块详细设计、API、Domain、数据缓存、容错设计、风险设计等。
- 代码编写规范：规范覆盖代码基础、日志、配置、多线程、数据库、异常使用等多层面，提升代码质量；
- 代码评审规范：changelist 描述、兼容性、性能、复杂性、团队评审文化等。
- 代码提测规范：Test 单测、代码编译构建、系统运行稳定性等、
- 代码测试规范：进入稳定性测试阶段，要严格审查系统是否达到测试准入条件，即满足测试实施的所有必要条件，如果未满足，则不开展稳定性测试。在稳定性测试实施结束后，严格检查所有测试准出条件是否满足，如：没有进行中的缺陷等，否则不予测试通过。
- 预发&引流压测规范：黄金链路必须进行 R2 引流验证。
- 发布上线规范：可灰度、可验证、可回滚等
- 验收规范：业务、产品验收规范
- 制定变更规范：提供变更级别、角色职责、活动阶段以及输入输出的详细规定
- 制定运维操作规范：针对公司日志标准，提供统一的日志排查命令及规范。
- 报警响应机制：针对运维相关的监控告警制定告警处理流程、告警升级机制
- 值班及责任判定机制：设置值班制度，每天有技术人员负责值班，值班周期内的所有问题由值班人员治理，不能及时完成的，添加到 BUG 定期跟踪并统计。在出现生产事件后，由专家团队对该问题进行详细分析，确定问题的发生原因、解决办法后，对该问题进行问责，明确责任团队、责任人、责任承担比例等内容。避免在稳定性治理中产生“囚徒困境”。
- 故障管理机制：故障管理机制包括规范管理故障响应流程、故障升级机制、故障复盘机制，规范技术人员在应对突发故障时的操作流程，明确职责边界，提升沟通效率，推动故障闭环，提升故障处理效率

2、开发和 SER 的区别

提到稳定性，先讲个概率 SRE（Site Reliability Engineering，站点可靠性/稳定性工程师）

一说到 Software Developer，人们脑子里就能反映出需求评审、编码、调试、测试、上线、修 bug 等具体工作内容。那 SRE 呢？SRE 与普通的开发工程师（Dev）不同，也与传统的运维工程师（Ops）不同，SRE 更接近是两者的结合，也就是 2008 年末提出的一个概念：DevOps，这个概念最近也越来越流行起来。SRE 模型是 Google 对 Dev+Ops 模型的一种实践和拓展（可以参考《Google 运维解密》一书），SRE 这个概念我比较喜欢，因为这个词不简单是两个概念的叠加，而是一种对系统稳定性、高可用、团队持续迭代和持续建设的体系化解决方案；

都是做技术的，很多开发刚刚转向稳定性方面时，有些弯转不过来。举个例子：对于“问题”，传统的开发人员更多的倾向于是“bug/错误”，而 SRE 倾向于是一种“风险/故障”，所以，两者对“问题”的处理方法是不一样的：

- 开发：了解业务 -> 定位问题 -> 排查问题 -> 解决问题
- SRE：了解业务归属 -> 快速定位问题范围 -> 协调相关人投入排查 -> 评估影响面 -> 决策恢复手段

可见，开发人员面对问题，会首先尝试去探究根因，研究解决方案；而 SRE 人员首先是评估影响，快速定位，快速止损恢复。目标和侧重点的不同，造成了 SRE 思考问题的特殊性。

所以，成为一名 SRE，就一定要从态度和方式上进行转变，切换到一个“团队稳定性负责人”的角度上去思考问题。

3、谈谈个人对 SRE 的几点要求

1.责任心、细心、耐心。

1.负责任是第一要素，主动承担，对报警、工单、线上问题、风险主动响应，不怕吃苦；一个不负责任的人，遇到问题与我无关的人，边界感太强的人，难以做好稳定性的工作；

2.及时、快速的响应，这是最关键的一点，作为一个 SRE，能够及时、快速的响应是第一要务，遇到报警、工单、线上问题，能够第一时间冲上去，不要去问是不是自己的，而是要问这个事情的影响



京东云开发者
@JCloud

拥抱技术，与开发者携手创造未来！



1 个赞



共收到 0 条回复



是什么，有没有坑，有没有需要优化的风险？

3.主动走到最前面、主动想优化的办法、主动出头解决问题、主动挖掘系统风险薄弱点。

2.不能只做当下，要看到未来的风险，善于总结。

3.把机制建立好，切实落地。作为一个 SRE，想做到“不出问题”这个基线，关键还是要靠大家。

二、稳定性建设方向

1、地基要打牢

稳定性建设工作重在预防，根据多年的工作经验，至少 70% 的线上故障都可以通过预防工作来消除。因此，在日常工作中，我们需要投入相应的精力来进行根基建设。所谓的根基建设，就是要把开发、测试和上线这三大流程做到透彻。包括：DesignReview、CodeReview、提测流程、上线流程、引流验证、性能测试等。

2、工作在日常

俗话说养兵一日，用兵一时。稳定性工作不是一蹴而就，而是日常的点点滴滴，一步一个脚印走出来的。

需要团队人人参与、持续完善监控告警、检查每一个告警是否配置、及时消灭线上小隐患。可参考每周的稳定性会议。

- 梳理：主动梳理团队的业务时序、核心链路流程、流量地图、依赖风险，通过这个过程明确链路风险，流量水位，时序冗余；
- 技术债务治理：主动组织技术债务的风险治理，将梳理出来的风险，以专项的形式治理掉，防患于未然。但需要注意别由于治理而导致线上问题，需要加强引流验证比对。
- 演练：把风险化成攻击，在没有故障时制造一些可控的故障点，通过演练来提高大家响应的能力和对风险点的认知。
- 报警：除了前面说过的主动响应之外，还要经常做报警保险和机制调整，保证报警的准确度和大家对报警的敏感度。同时也要做到不疏忽任何一个点，因为疏忽的点，就可能导致问题。

3、预案是关键

我们需要认识到预案的重要性，并投入相应的精力来进行预案的制定和更新。这样，我们才能更好地应对各种突发情况，保障项目的顺利进行。通过每周的稳定性去深入挖掘每个接口的隐患及不足，比如业务指标是否加上、业务指标是否能真实反馈该接口的特性等。

4、大促特殊场景

系统在大促的稳定性和日常稳定性的区别在哪呢？个人理解核心是两点：

- 1、【技术】高并发流量：大促流量峰值是日常的 N 倍（几十、几百倍），需要具备更高的并发流量处理能力，以保证系统的稳定性这方面。针对这评估好流量，做好容量规划即可。
- 2、【业务】业务场景多样化：大促会增加很多日常用不到的场景，很明显的比如预售场景、Promise 特殊时效控制、停运降级功能等。针对日常不用，大促才用的功能点。可整理功能点，在大促前 1 个月模拟大促，业务进行相关功能配置，演练全流程，类似每年大促都进行的预售场景演练。因为每年需求都在迭代增加，难免会影响之前的功能点。这样就可避免大促期间突然使用功能发现不好用的问题
- 5、执行是王道

其实听复盘会学东西是一方面，最主要是应该问问我们系统是不是也存在这种问题，我该怎么规避或解决这类风险问题，别人暴露的我也存在，应该第一时间去解决，而不是我知道但我不做。

三、前置：扁鹊三兄弟

与扁鹊三兄弟一样，如果想要让稳定性有价值，SRE 同学一定不能站到系统的屁股后面等着擦屁股，必须走到前面，看到未来的风险。

既要在发生时快速解决问题（做扁鹊）

也要把风险归纳总结，推动解决（做二哥）

还要在系统健康的时候评估链路，发现隐藏的问题（做大哥）；

1. 做扁鹊大哥：擅长的是“事前控制”，具有敏锐的洞察力和战略眼光，防患于未然，在系统健康时发现问题。
2. 做扁鹊二哥：擅长的是“事中控制”，具有出手迅速、果断、干练的特点，在系统有隐患时发现问题。
3. 做扁鹊： 擅长的是“事后控制”，是临危受命型的关键人物，在系统发生时快速解决问题。

根据《鶡冠子·卷下·世贤第十六》的记载，有一次，魏文王问扁鹊说：“你们家兄弟三人，都精于医术，到底哪一位最好呢？”扁鹊回答说：“长兄最好，中兄次之，我最差。”

”魏文王又问：“那么为什么你最出名呢？”扁鹊回答说：“长兄治病，是治病于病情发作之前，由于一般人不知道他事先能铲除病因，所以他的名气无法传出去；

中兄治病，是治病于病情初起时，一般人以为他只能治轻微的小病，所以他的名气只及本乡里；而我是治病于病情严重之时，一般人都看到我在经脉上穿针管放血，在皮肤上敷药等大手术，所以以为我的医术高明，名气因此响遍全国。”

参考：

公众号普惠出行产品技术：<https://mp.weixin.qq.com/s/R2qBQgJCueErBL35ld4KZQ>

♥ 1 个赞



暂无回复。

需要 [登录](#) 后方可回复, 如果你还没有账号请点击[这里](#) [注册](#)。



关于 活跃用户 中国移动互联网测试技术大会 反馈 Github API 帮助推广
TesterHome社区，测试之家，由众多测试工程师组织和维护的技术社区，致力于帮助新人成长，提高测试地位，推进质量发展。Inspired by RubyChina
友情链接 WeTest腾讯质量开放平台 / InfoQ / 掘金 / SegmentFault / 测试窝 / 百度测试吧 / IT大咖说
[简体中文](#) / [正體中文](#) / [English](#)

