中国移动CMChaos

混沌工程企业实践

中国移动
China Mobile

目录

CONTENT

5G+

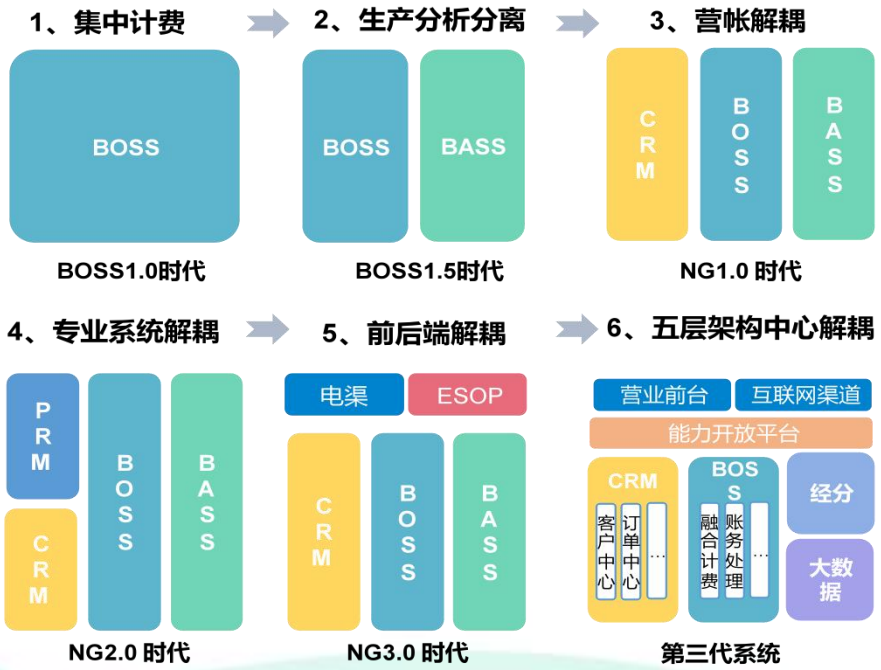
一、建设背景

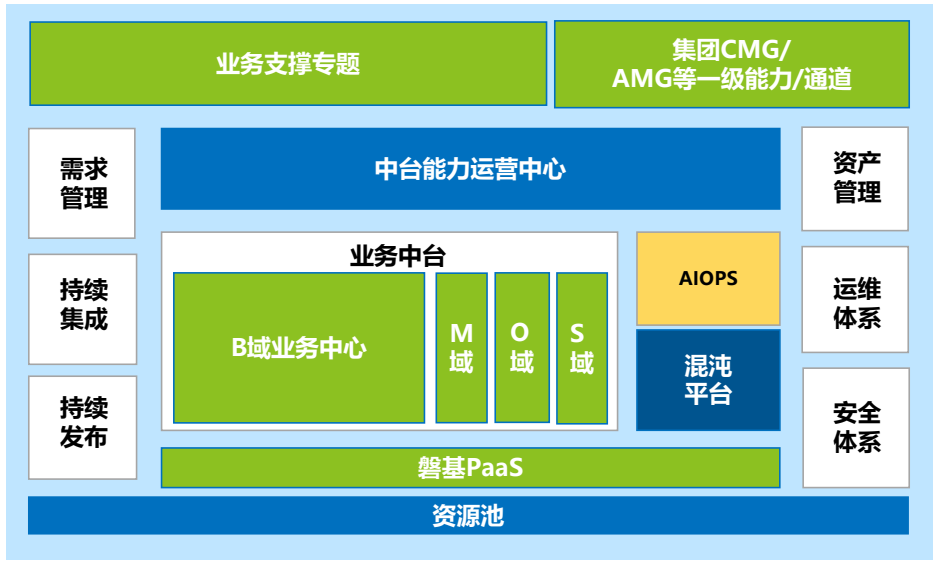
故障点增加	IT系统容器化、微服务化演进及大量开源引入，故障点增多，事后处置风险增加。
业务系统复杂	业务系统多样，服务调用关系复杂，导致故障定位难度加大，排查效率显著降低。
信创产品上线	新系统或信创产品上线缺乏科学验证，导致系统可靠性、稳定性和兼容性问题频发。
集群规模扩大	集群规模增长，对基础设施稳定性要求提升，需演练每个迁移系统，确保平台可靠性。
传统演练局限	宕机、停服演练方式单一，无法全面覆盖复杂故障场景，演练价值有限。
人工演练不足	人工演练耗时长、效率低，难适应复杂架构运维需求，影响快速响应能力。
运维挑战升级	庞大IT系统和复杂服务架构，应用升级和迭代频繁，生产环境安全性受到威胁。

一、建设背景

业务运营支撑系统经过二十多年的升级迭代，目前依托磐基PaaS平台为技术底座，已基本完成云化、微服务化、容器化等技术架构升级改造，具备了较先进的业务支撑能力，可实现资源弹性伸缩、应用快速扩缩容、需求快速响应。

IT系统的演进历程





基础设施标准化通用化、架构分布化、能力服务化、交付敏捷化、运营智能化

一、建设背景

CMChaos是一套专为运维和运营团队量身打造的云原生混沌工程实践平台，基于**中国移动磐基PaaS平台底层架构**和**丰富的实践经验**，结合实际业务场景倾力打造而成，平台集**主机类、应用类、网络类、存储类、安全类、信创兼容类**以及**中间件**等较复杂的演练事件和混沌实验场景，底层采用微服务部署架构，确保平台具备**高可用性**和**稳定性**。

持续思考中……

- 随着数字化转型的加速推进，如何有效确保生产环境具备较强的韧性成为关键课题；
- 云服务的稳定与安全运行是保障云上生产安全和业务连续性的基本前提；
- 云服务事故因其不可预测、不可控及高复杂性，常导致重大安全事件和经济损失；
- 网络安全应急响应作为网络安全体系的重要组成部分，已被纳入国家网络安全顶层设计中，成为维护国家和企业数字安全的重要手段。

混沌工程可以模拟各种可能的故障场景，帮助团队提前发现并修复潜在的问题，从而在不影响用户的情况下提高系统的整体稳健性和可靠性。

中国移动
China Mobile

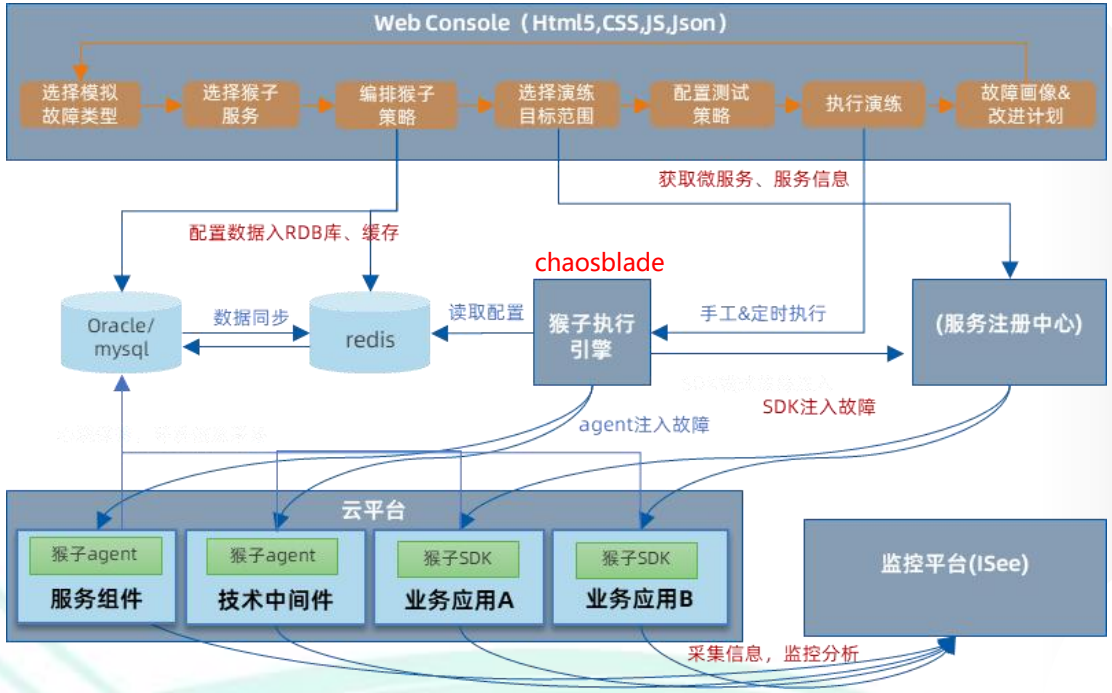
目录

CONTENT

5G+

二、创新实践-案例1：持续混沌演练，洞悉潜在风险

基于chaosblade完成混沌实验工程基本能力建设，主动注入故障，不断演练和复盘、迭代改进和升级，提前找到系统缺陷、尽可能多识别风险，防止其演变成重大故障，提高系统在生产环境的弹性能力和韧性；同时采用小步快跑的方式，从**微服务、应用系统、主机资源3类典型故障场景**入手，直接面向生产环境的实际流量进行破坏性实验。推动复杂架构下系统健壮能力的不断完善，强化应急抢修实战能力的不断提升。



1 模拟杀服务Pod、杀节点、增大Pod资源负载，验证微服务的容错能力

2 模拟应用不可用，验证应用层的降级熔断、故障转移、隔离、自愈能力

3 模拟主机资源异常，验证故障节点或实例是否被自动隔离、下线

二、创新实践-案例2：高可用性测试

提高了业务系统的**高可用性**，特别是在流量高峰期间，显著降低了因故障导致的用户交易失败率，确保了用户体验的稳定性。

案例解析： 场景 **业务系统的混沌工程高可用性测试**

明确目标：

验证在数据库故障、网络延迟、服务器宕机等情况下，平台能否继续处理用户请求，确保结算的顺利进行和用户体验不受影响。

结果与优化：

尽管在网络延迟和部分服务器宕机的情况下，平台整体仍能保持较高的交易成功率和响应速度，但在数据库故障的场景中，切换到备用数据库的时间较长，导致部分交易失败。根据这一发现，**团队优化了数据库切换机制，并增加了更为灵活的负载均衡策略。**

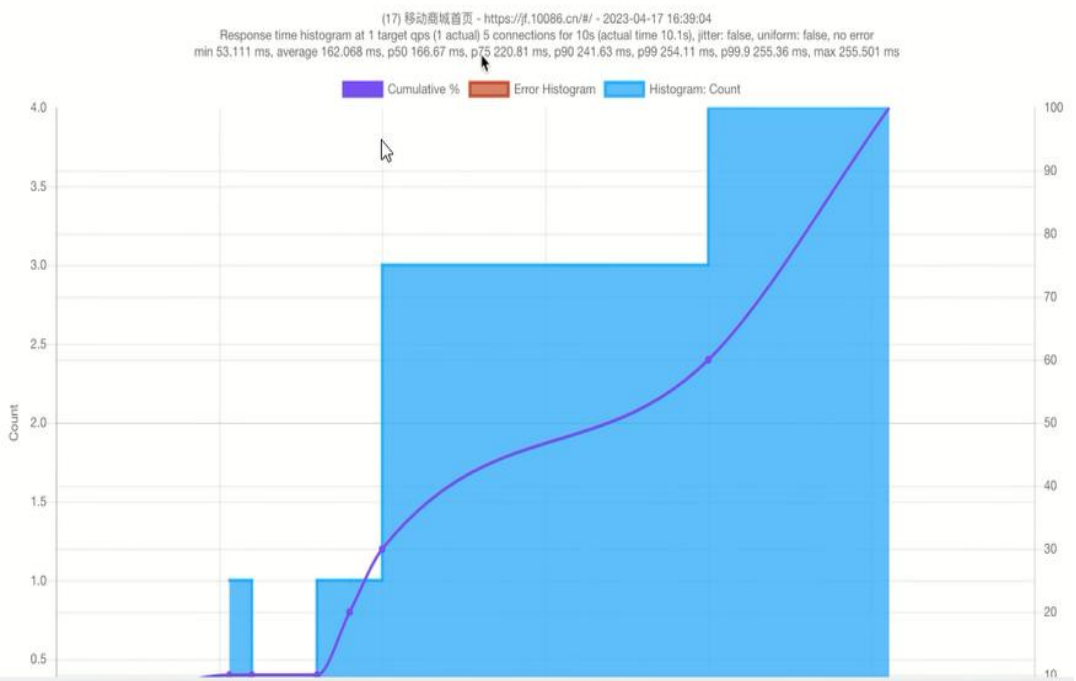
实验设计

- 定义正常状态：记录无故障时的响应时间、交易成功率等关键指标。
- 假设稳定性：在故障情况下，确保结算交易成功率≥95%，响应时间≤正常的两倍。
- 数据库故障注入：模拟主数据库宕机，验证备用数据库切换和交易完整性。
- 网络延迟注入：对关键服务随机注入延迟，监测响应时间和用户体验。
- 服务器宕机注入：关闭部分服务器，验证负载均衡的流量分配效果。
- 监控配置：实时监控交易成功率、响应时间及服务器资源利用率。
- 执行策略：在模拟高峰期注入故障，确保测试环境接近生产环境。
- 防护策略：配置自动故障恢复，确保实验不影响真实用户交易。
- 分析报告：生成报告，评估故障影响，识别系统瓶颈并优化。

二、创新实践-案例3：业务仿真模拟

进行**业务仿真模拟**，快速洞察系统性能，多维度分析和异常监控，可视化展现，为业务上线提供可行性验证，使业务更加敏捷和韧性。

案例解析： 场景 **业务系统上线前测试，高并发用户访问**



用户行为模式：

- 用户访问首页： 用户以平均每5秒的频率访问首页，查看最新上架的应用和服务信息；
- 用户浏览商品： 用户以平均每10秒的频率随机浏览网站上的应用；
- 用户下单购买： 用户以平均每30秒的频率随机选择一种应用下单购买；
- 查询订单状态： 用户以平均每60秒的频率查询一次他们最近的订单状态。

业务仿真模拟：

模拟用户的上述行为，模拟10000个用户在1小时内的同时访问，

每个用户行为的时间间隔将根据指定的平均频率进行随机调整，以模拟用户行为的随机性，

在模拟过程中，平台将记录**网站的响应时间**、**服务器负载**等性能指标。

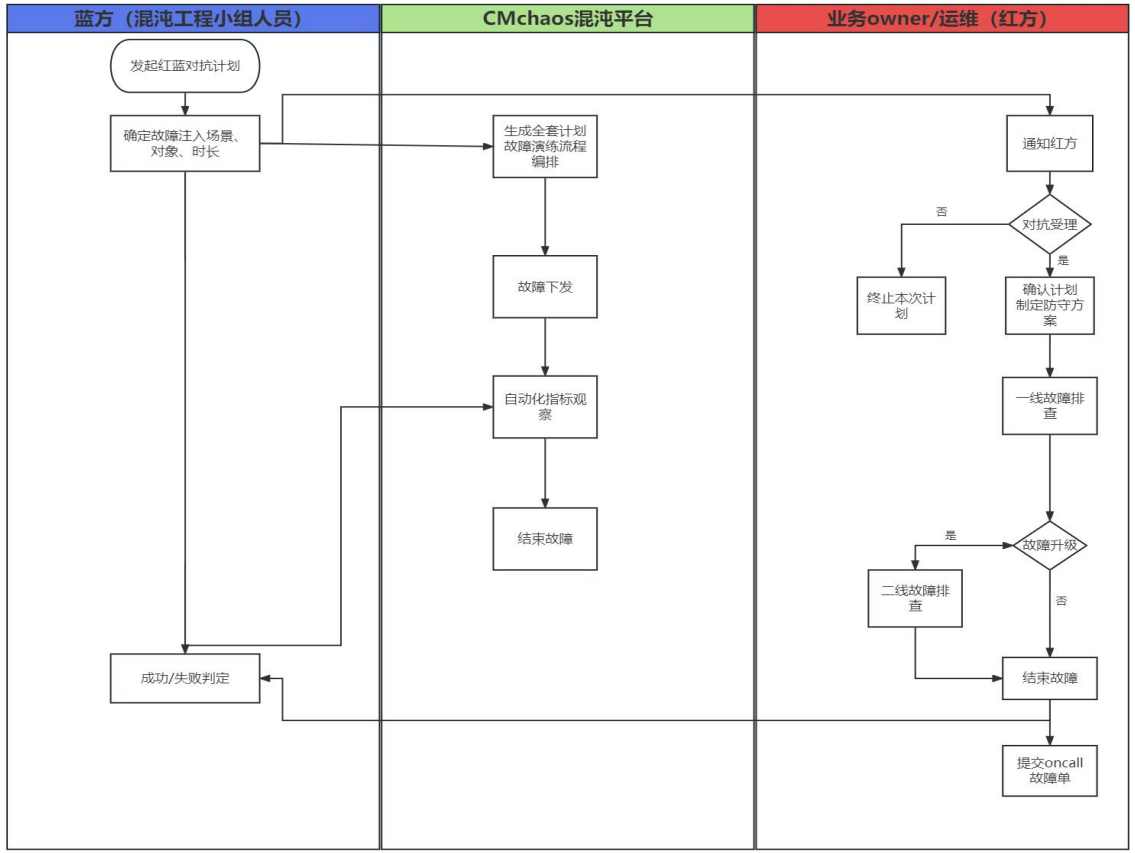
模拟结果分析：

评估网站在高并发访问情况下的性能表现，如响应时间、服务器负载等；

根据生成报告分析结果，确定是否存在性能瓶颈，并制定改进策略。

二、创新实践-案例4：红蓝对抗实战演练

坚持“**实战化、常态化**”原则，通过制定策略、编排方案、执行演练、协同对抗和形成报告，深入挖掘问题隐患，检验并增强应急响应应能力。同时，**提升SRE意识**，强化应急流程与团队协作，从而有效提升**整体运营管理能力。**



二、创新实践-案例5：网络和信息安全演练

通过安全演练模拟，增强运维团队的**网络和信息安全防范意识**，发现安全类潜在问题。

场景概述

在混沌工程体系建设中进行安全演练模拟，结合安全原子能力场景库中的场景，自定义构建安全实战学习环境，提供一个灵活且场景覆盖全面的运维及安全人员培训环境。

应用效益

- 安全原子能力脚本来源于国内外安全专家验证过的公开漏洞库以及移动安全专家的研究总结形成场景库，提供给供新手进行实战演练学习，大大提高安全人员培训质量;
- 场景脚本的调用、实现简单且大部分操作由平台自动完成，对于新手较为友好，能够提高用户体验，提高学习效率;
- 安全人员能力的提升，直接带来组织安全运维效率、安全加固效率的提高。

二、创新实践-案例6：混沌工程测试场景

中国移动 China Mobile

网络故障场景

网络延迟：模拟网络传输延迟，检查系统在网络拥堵或长距离传输时的表现。
网络丢包：模拟数据包丢失，测试系统在数据不完整情况下的容错机制。
网络分区：模拟网络分割，检验分布式系统在部分节点间通信中断时的数据一致性与可用性。

计算资源故障场景

CPU压力测试：模拟CPU使用率激增，检查系统在资源受限时的性能表现和自我保护机制。
内存泄漏：模拟应用程序内存占用过高，测试系统如何防止内存泄露带来的稳定性问题。
磁盘空间不足：模拟存储资源耗尽，验证系统对存储空间告警和恢复机制的有效性。

服务层故障场景

服务崩溃：模拟单个服务进程崩溃或停止响应，测试服务重启和负载均衡策略。
服务降级：模拟服务进入降级模式，检查系统能否按预定的方式降低功能级别以保持核心功能可用。
依赖服务故障：模拟依赖的外部服务不可用，检验系统是否有备用方案或容错机制。

数据库故障场景

数据库主从切换：模拟主数据库发生故障，测试备库接管后数据的一致性和服务连续性。
数据丢失或损坏：模拟数据写入失败或已存储数据损坏，验证数据备份与恢复方案。
事务冲突与并发问题：模拟并发访问导致的事务冲突，测试数据库的并发控制策略。

系统负载与性能场景

突发流量：模拟短时间内大量请求涌入，测试系统的弹性扩容能力和处理高负载时的性能表现。
请求超时：设置服务端或客户端请求超时，观察系统如何处理超时后的重试策略和资源释放。

配置错误与变更场景

配置不当：故意更改错误的系统配置，测试系统对无效配置的识别和处理能力。
配置热更新：模拟在运行时动态修改系统配置，验证系统无中断切换配置的能力。

二、创新实践-案例7：混沌工程测试场景-非功能性测试

中国移动 China Mobile

接口调用测试，通过模拟各种异常场景，以验证系统接口在面对不同故障情况时的稳定性和可靠性。

随机延迟注入

引入随机或预设的网络延迟，模拟网络不稳定或拥塞的情况，测试接口在延迟场景下能否正确处理请求并返回响应。

请求失败模拟

模拟接口调用失败，如返回HTTP 5XX错误码、连接超时、请求取消等，以此检查系统的重试策略、熔断机制和错误处理逻辑。

并发压力测试

同时发起大量并发请求，模拟高峰期的系统负载，检验接口在高并发场景下的性能和并发控制能力。

资源限制

限制服务器资源（如CPU、内存），当接口在资源紧张的环境下运行时，验证接口是否仍然能够提供稳定的服务。

依赖服务故障

如果接口依赖其他服务，可以模拟这些依赖服务的故障，如返回错误数据、响应超时或完全不可达，考察接口对上游服务故障的容错处理。

数据异常模拟

在接口处理过程中，人为制造数据异常，如传入非法参数、插入错误数据等，测试接口的数据校验和异常处理能力。

权限与认证异常

模拟认证失败、权限不足等情况，检查接口在权限异常时的处理流程是否符合预期。

API版本更迭

当接口版本发生变化时，模拟旧版本客户端向新版本接口发送请求，确认兼容性和降级处理是否合理。

中国移动
China Mobile

目录

CONTENT

5G+

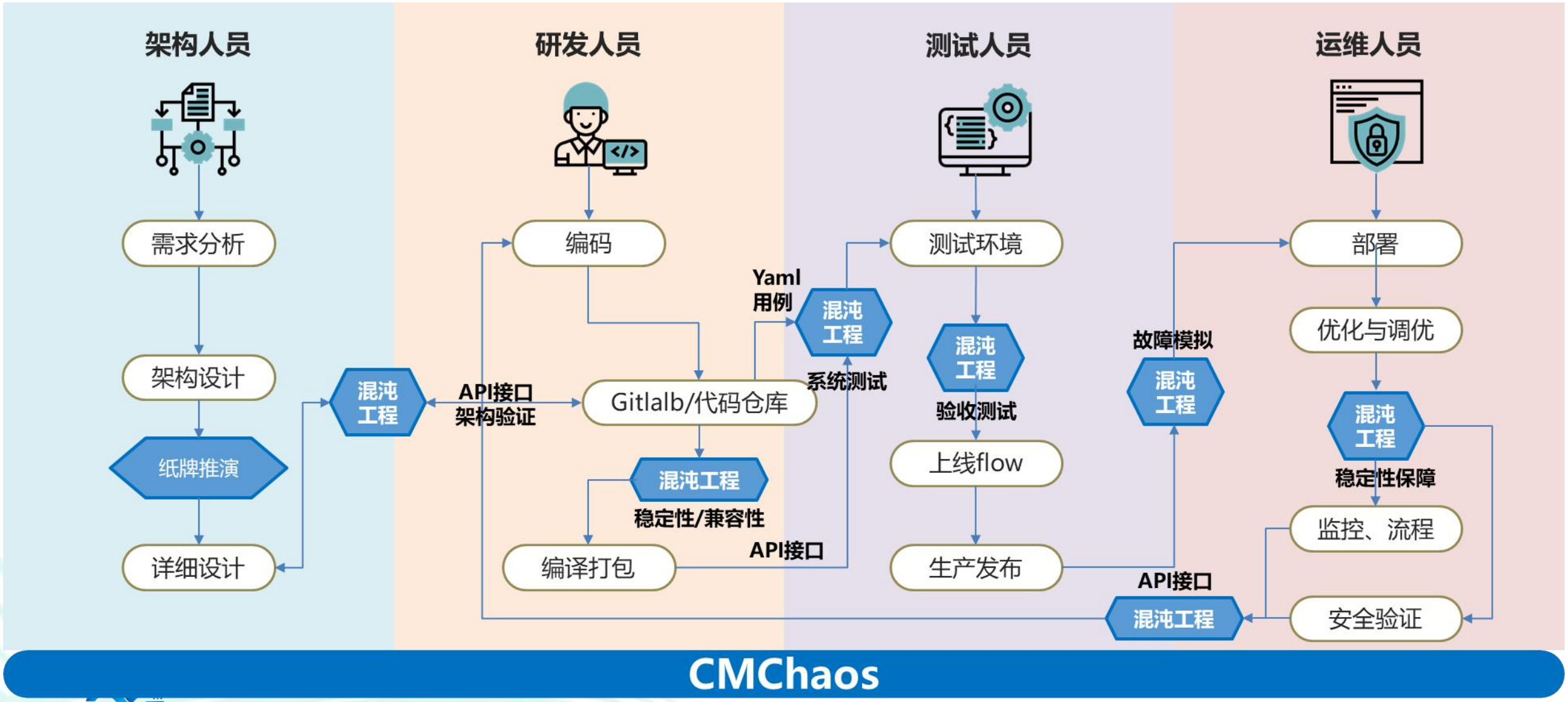
三、CMChaos混沌工程价值意义

CMChaos混沌工程通过故障实战为工程师构建了一个非确定性、非周期性的故障环境，剥离工程师对初始条件的敏感依赖，进而提升了组织对故障防御的设计能力、故障事件的构建能力、故障问题的描述能力以及故障应对的组织协调能力。

架构团队	研发团队	测试团队	运维团队
规划设计或采购需求梳理 **性能** 特指对延迟或资源成本的最小化。 **可用性** 系统正常响应和避免停机的能力。 **容错能力** 系统从非正常状态中恢复的能力。	通过混沌工程积累案例场景，从开发之初或开发时将由于开发或引入的组件导致的故障场景考虑进去，从故障场景分析**如何增强问题的快速定位、防护、跟踪等能力。**	单元测试 > 集成测试 > 生产环境 debug传统软件测试验证代码的逻辑正确性。 **混沌测试验证系统的健壮性。**	通过注入故障快速模拟故障的发生，观察系统的反应、对业务的影响、以及当前的应急预案完备性，**帮助运维人员对系统建立起稳定性信心，反向管理研发团队。**

“被动响应”到“主动防御”

三、CMChaos混沌工程是职能部门有效协同工具



三、CMChaos混沌工程是管理团队有效的运维管理和决策工具

强大的混沌管理能力

- 管理团队通过整合SRE和混沌工程方法，能够在复杂和动态的环境中增强系统的稳定性和适应性；
- 实时监控和自动化工具确保管理团队能够迅速做出决策，优化运维流程和应急响应效率；
- 定期的应急和实战演练，管理团队能够不断完善应急预案，增强团队的协同作战能力，确保部门高效合作。

强大的智能管理工具

- 全面的可视化工具，确保运维人员能够实时监控系统的健康状况和性能指标；
- 性能分析和诊断工具，使运维和开发团队能够迅速识别性能瓶颈和故障根源；
- 全面的监控数据和详细的分析报告，提供资源分配和优化决策，提高整体运维和管理的水平。

强大的自动化能力

- 高度自动化的运维流程，大幅减少了手动操作和人为错误，提高了运维效率和系统可靠性；
- 灵活的扩缩容机制，能够根据负载变化自动调整资源配置，避免资源浪费和性能瓶颈；
- 运维资源管理策略，实现资源的最优配置和成本控制，降低运维成本，提高投资回报率。

健全的安全管理能力

- 支持管理基于ARM和X86架构的虚拟化混合部署，适配国产操作系统，异构数据库适配，安全自主可控；
- 租户数据安全隔离，多层次的安全防护策略，确保系统在面对各种安全威胁时能够保持稳定和安全；
- 平台具有管理团队定期组织安全审计和安全演练能力，随时评估和改进现有的安全策略和措施；

三、CMChaos混沌工程可以构建SRE运维保障体系

CMChaos混沌工程通过**自动化故障注入**和**实时监控**，验证并提升系统弹性和可靠性，确保在突发故障和高负载情况下系统满足**服务级别目标（SLO）**，快速恢复并推动**持续改进**和**跨团队协作**。

验证系统能否满足SLO

平台通过模拟故障来验证系统是否能够满足既定的SLO，并且帮助识别和消耗错误预算，确保了系统在预设的错误预算范围内运行，并帮助团队了解系统的真实表现和改进需求。

性能监控和实时告警

平台集成了多种监控工具，提供全面的系统性能监控和告警功能，通过实时数据分析和智能报警，确保问题能够被快速发现和响应，帮助开发和运维团队快速识别和解决性能瓶颈。

自动化故障注入和实验能力

平台采用先进的自动化运维工具，实现自动部署、自动扩缩容和自动恢复。通过自动化故障注入和监控能够快速识别系统薄弱环节，并验证监控工具的有效性和覆盖范围。

故障管理和应急响应能力

平台积累300+故障事件，1000+故障场景，严格执行故障管理流程，确保实验的安全性，应急演练深度挖掘问题隐患，检验和增强应急有效性，增强SRE意识，提升应急、人员流程的协作意识。

不仅是运维工具还是有效管理工具

平台建立了完善的运维流程和制度，确保各项操作有章可循，通过持续改进和学习，不断优化流程，提高运维效率。本着“实战化、常态化”的思想进行实战演练，有效提升运营管理能力。

中国移动
China Mobile

目录

CONTENT

5G+

四、总体收益

积累了1000+事件场景，完成300+实战演练，输出专题报告1000+份，检测发现漏洞200+个。近两年，无重大故障事件数量，全年异常事件总量从52次减少到35次，累计减少32.69%，全年平均异常持续时长由***分钟缩短为**分钟，累计缩短**。**%，异常事件、平均异常持续时长双减。

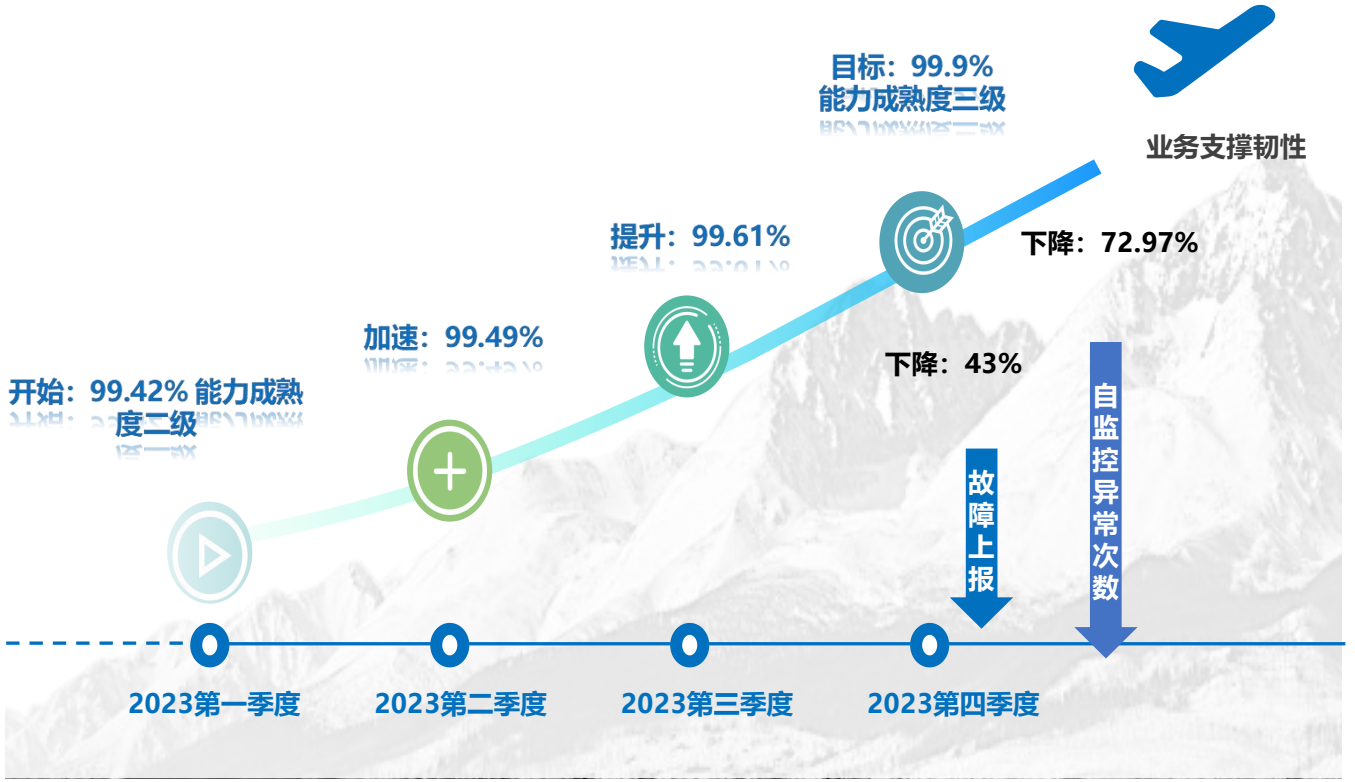
混沌工程演练情况

落地场景（1000+）	问题及漏洞检测200+
实战演练（300+）	高可用提升40%↑
报告输出（1000+）	稳定性提升20%↑

发现：1分钟故障发现数量从40%提升到**70%**

定位：5分钟故障定位数量从51%提升到**65%**

恢复：10分钟可恢复故障数量从46%提升到**70%**



四、总体收益

降本增效类型	之前数值	应用混沌工程后数值	效益价值
自研可控	非自研 不适配国产化信创产品	自研可控 适配国产化信创产品	满足公司自主可控要求，自研原子事件+个性化脚本能力，国产化信创产品兼容适配
节约人工		使用该成果之后只需要1-2人即可	人员投入节省30%，整体项目上线效率提升50%
节约成本	测试费用***万元/每年（根据AWS的FIS报价计算所得）	根据基于混沌工程的韧性测试框架开展韧性测试，无测试行动服务成本。	成本节省***万元/每年
提质增速	异常总数52次， 平均持续时长***分钟	异常总数35次， 平均持续时长***分钟	提前消除应急架构先天不足隐患，提升应急效能，异常事件减少32.69%；平均异常持续时长缩短43.22%
增强能力	传统测试工具，各工具独立，缺少自主编排能力和可视化等能力	测试平台支持自动化、流程化、标准化、可视化，支撑韧性测试的高效有序执行	提升公司数智化水平，助力高质量发展

中国移动 China Mobile

谢谢！