

QCon全球软件开发大会

上海·2020

International Software Development Conference

上海·2020

Brought by **InfoQ**



创建高可靠性系统的工程实践

宋涛 - 携程集团 商旅事业部 CTO

简单介绍

- 携程商旅CTO，2017年加入携程，领导研发新一代机票引擎
- 此前在微软Windows团队，亚马逊云计算团队从事技术和管理工作
- 北京大学计算机毕业，加州大学计算机博士

系统可靠性：互联网时代的挑战

CNN
Gmail suffers another outage
Google suffered its second outage in as many days, with its email service Gmail going down on Tuesday for more than two hours.
22 hours ago

BBC.com
Google outage: YouTube, Docs and Gmail knocked offline
Today, at 3:47AM PT Google experienced an authentication system outage for approximately 45 minutes due to an internal storage quota issue...
2 days ago

Bloomberg
Google Services Including Gmail, YouTube Suffer Major Outage
Services from Alphabet Inc.'s Google experienced widespread outages around the world Monday, temporarily preventing people from...
2 days ago

2020/12/14

45

Global

The Verge
Prolonged AWS outage takes down a big chunk of the internet
Amazon Web Services (AWS), Amazon's internet infrastructure service that is the backbone of many websites and apps, experienced a ...
3 weeks ago

ZDNet
Amazon: Here's what caused the major AWS outage last week
AWS explains how adding a small amount of capacity to Kinesis servers knocked out dozens of services for hours.
2 weeks ago

GeekWire
Amazon details cause of AWS outage that hobbled thousands of online sites and services
A "relatively small addition of capacity" to the Amazon Kinesis real-time data processing service triggered a widespread Amazon Web Services ...
2 weeks ago

2020/11/27

308

1 Region

ZDNet
Microsoft's Azure AD authentication outage: What went wrong
Microsoft's Azure AD authentication outage: What went wrong. It's been a rough week for Microsoft users who have first- and third-party apps that ...
Oct 1, 2020

DatacenterDynamics
Microsoft outage brings down Azure, Office 365 and Teams
Between 2125 UTC and 0023 UTC, Microsoft's Azure Active Directory (Azure AD) suffered an outage, which affected authentication services, ...
Sep 29, 2020

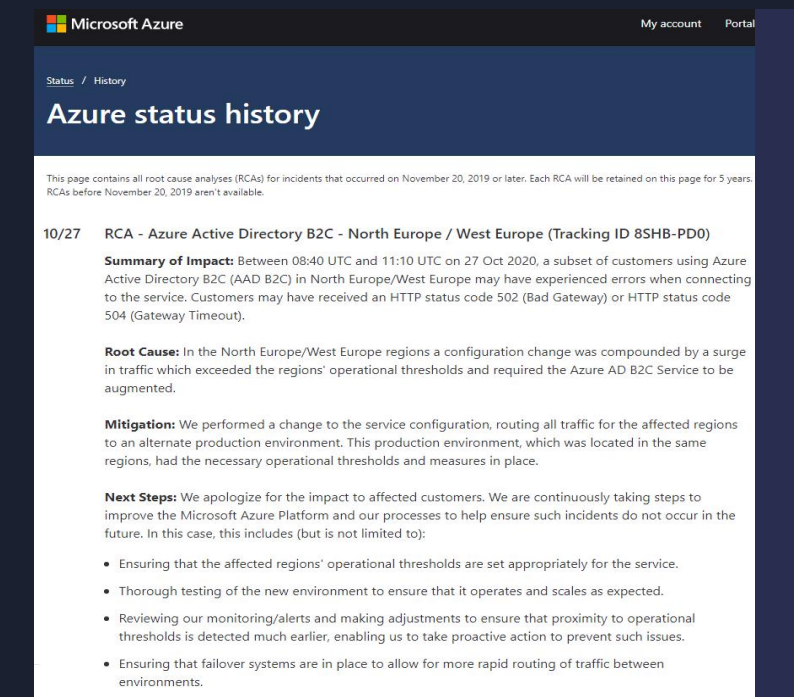
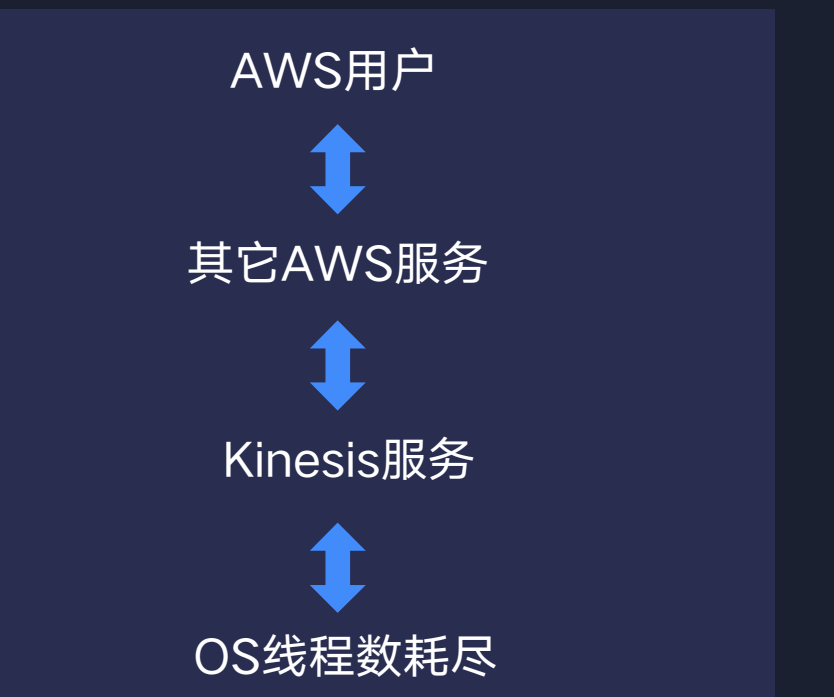
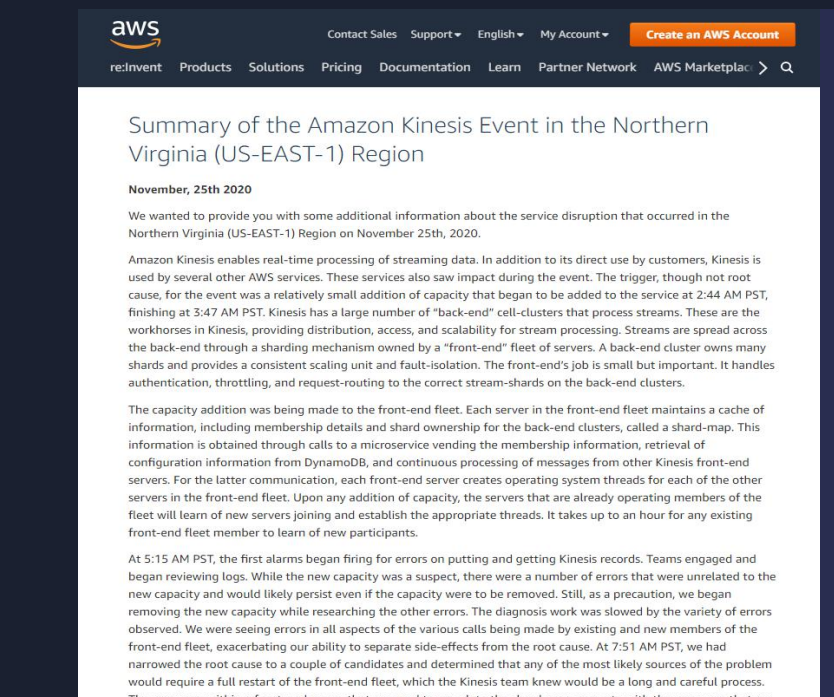
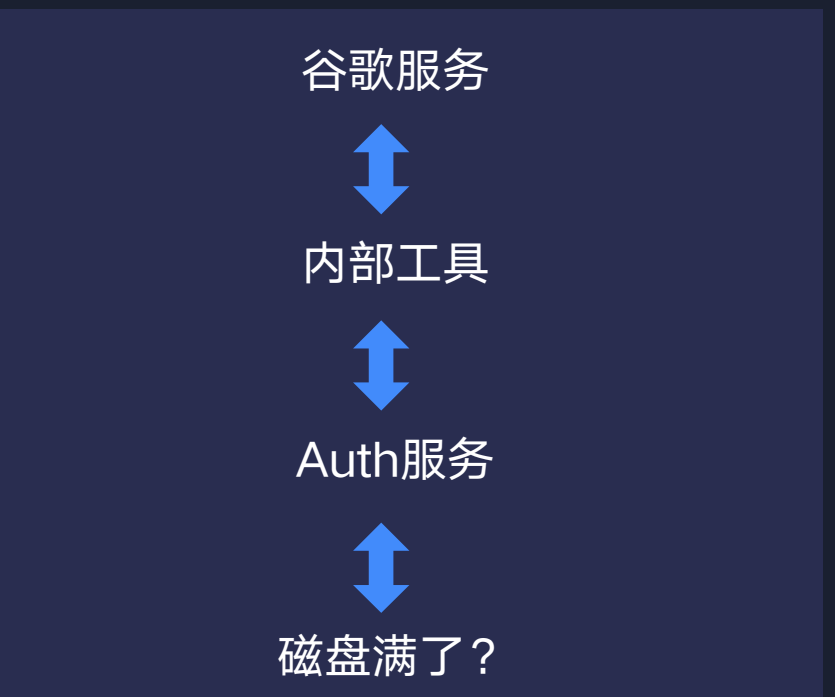
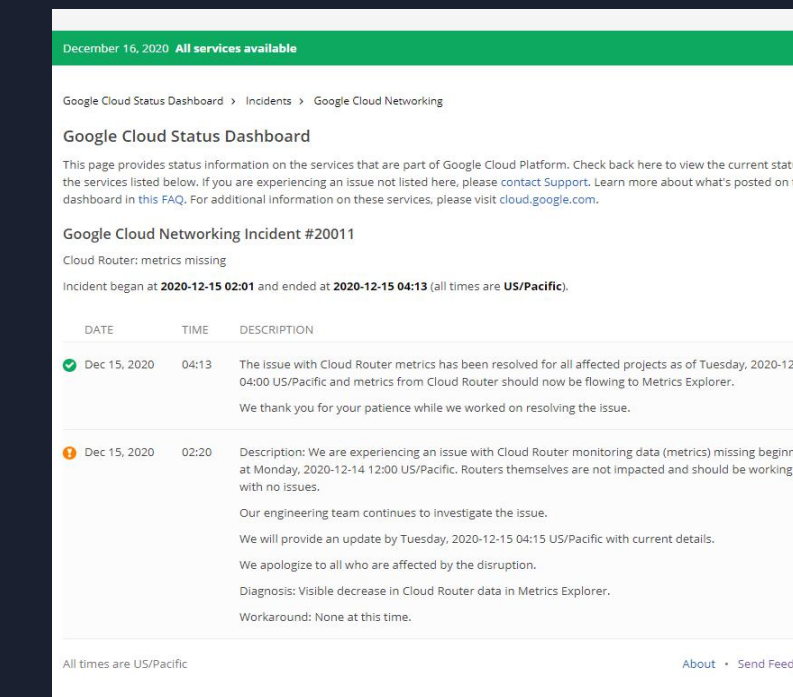
Computer Business Review
Microsoft Wobbles Again: Do Azure Staging Procedures Need a Rethink?
UPDATED: Azure said in a root cause analysis: "A service update ... The issue comes a fortnight after a protracted outage in Microsoft's UK ...
Sep 29, 2020

2020/10/27

150

Europe

系统可靠性：互联网时代的挑战



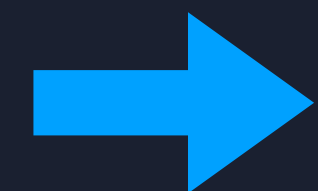
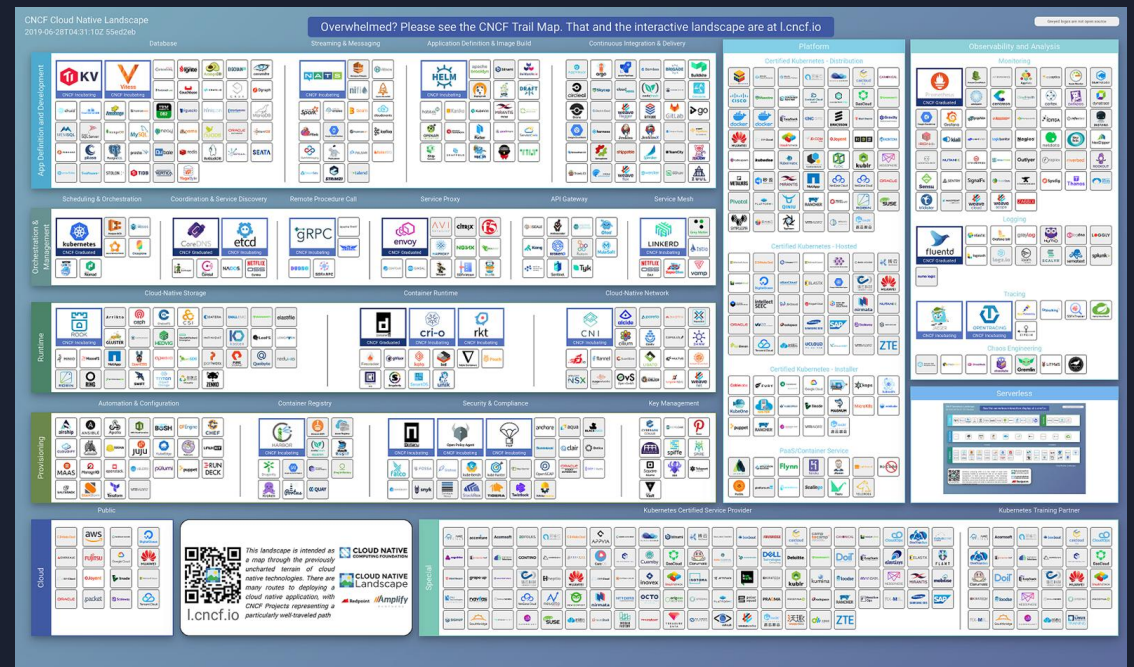
系统可靠性：互联网时代的挑战

高流量

低延时

数据一致性

快速上线



自研系统

Cloud Native

云计算



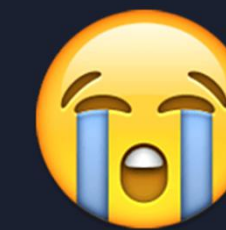
架构



开发



测试



运维

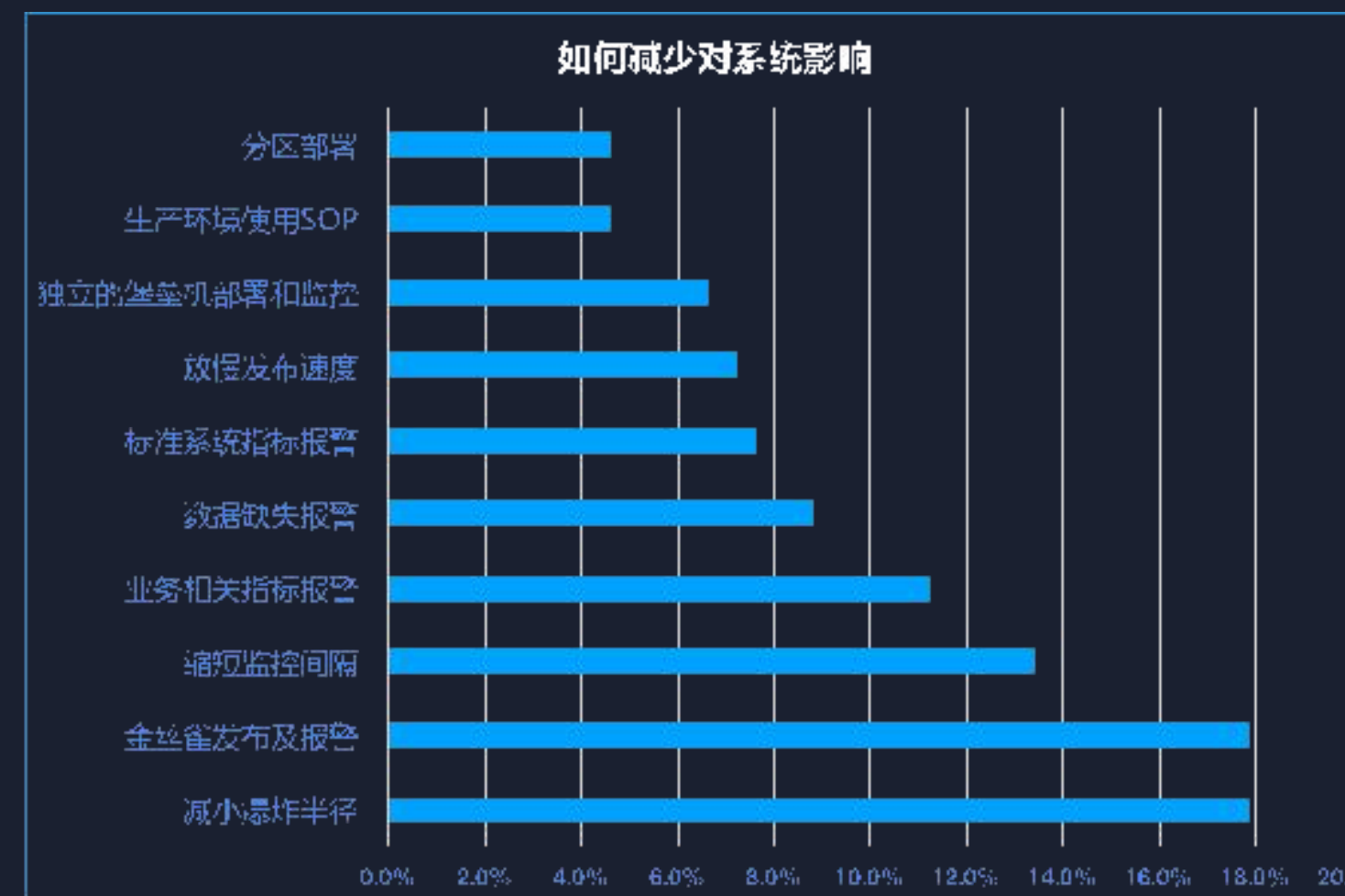
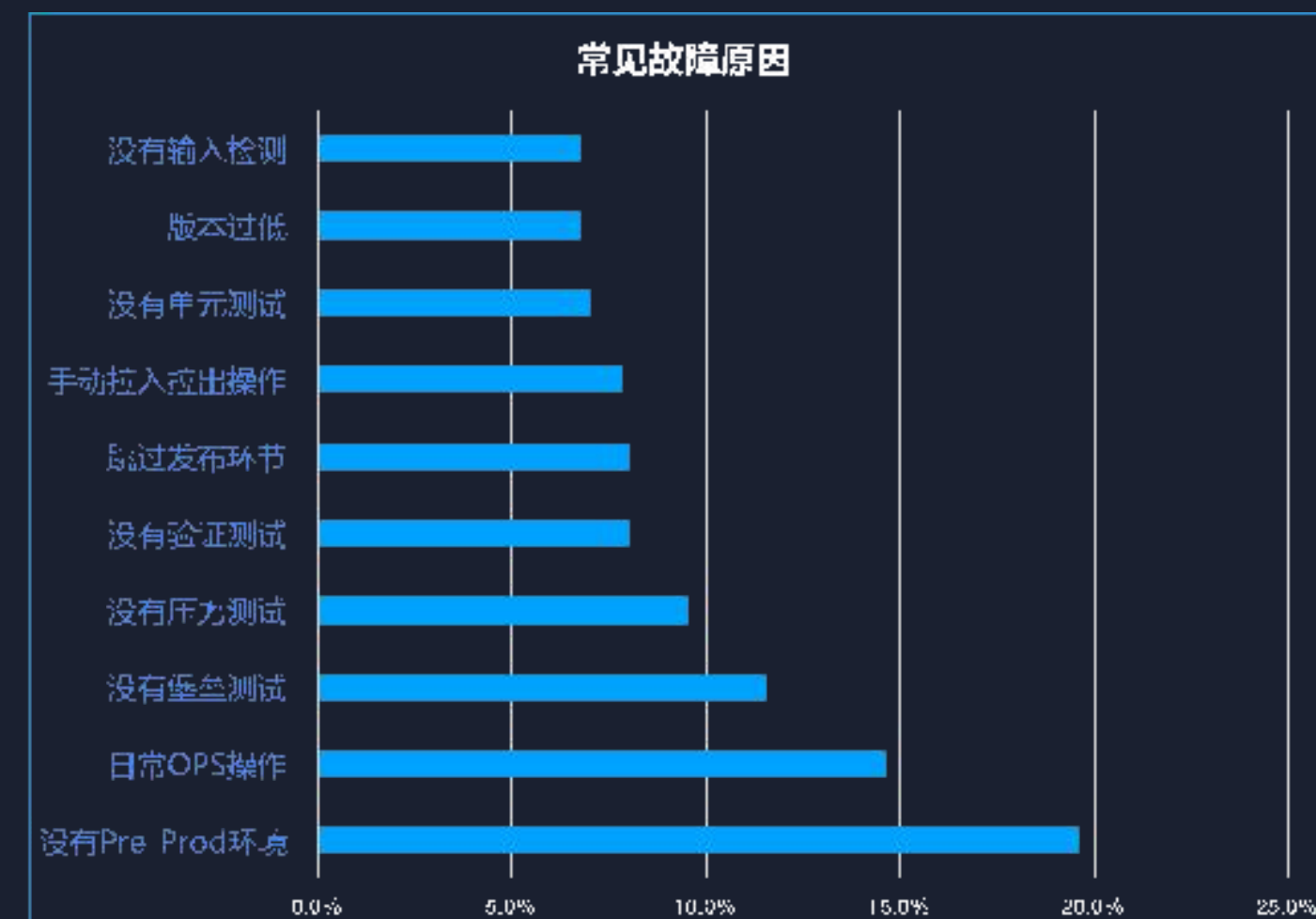
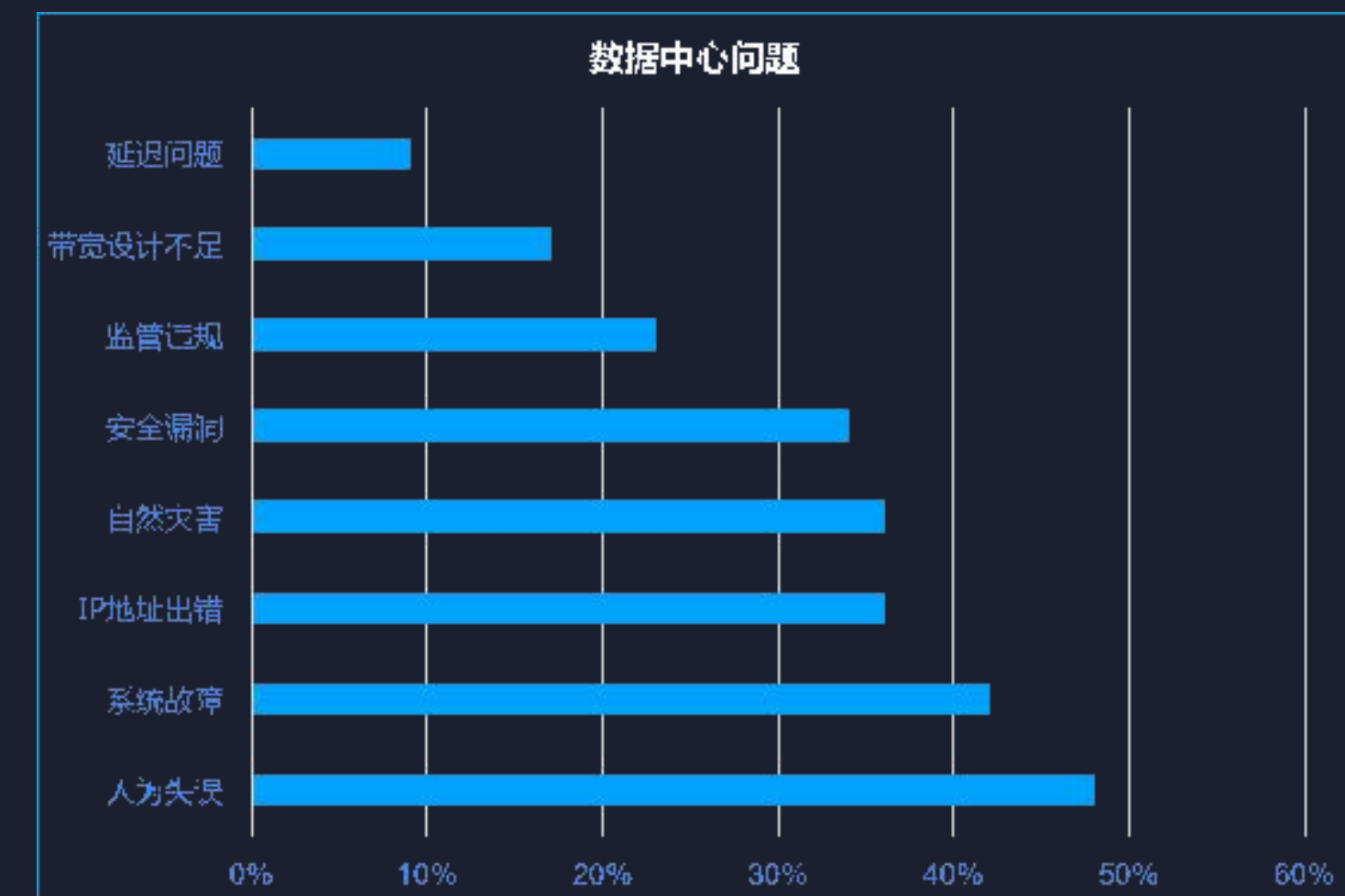
平生不识CNCF，便称架构师也枉然。

系统可靠性：互联网时代的挑战

系统可靠性一般是指在规定的时间内和规定的工况下，系统完成规定功能的能力/概率。

可靠性 = f(发生概率, 持续时间, 影响半径)

系统可靠性：影响因素



系统可靠性：质量提升&快速响应

设计

- 面向失败的设计
- 冗余（副本，主从，灾备）
- 隔离（DC，集群，渠道）
- Design Review

开发

- 编程规范
- 内部封装开源
- Code Review & Group Code Review

测试

- UT, FT, LT, Fuzzy
- Chaos

发布&OPS

- 监控
- 报警
- SOP
- 灰度发布
- 金丝雀发布

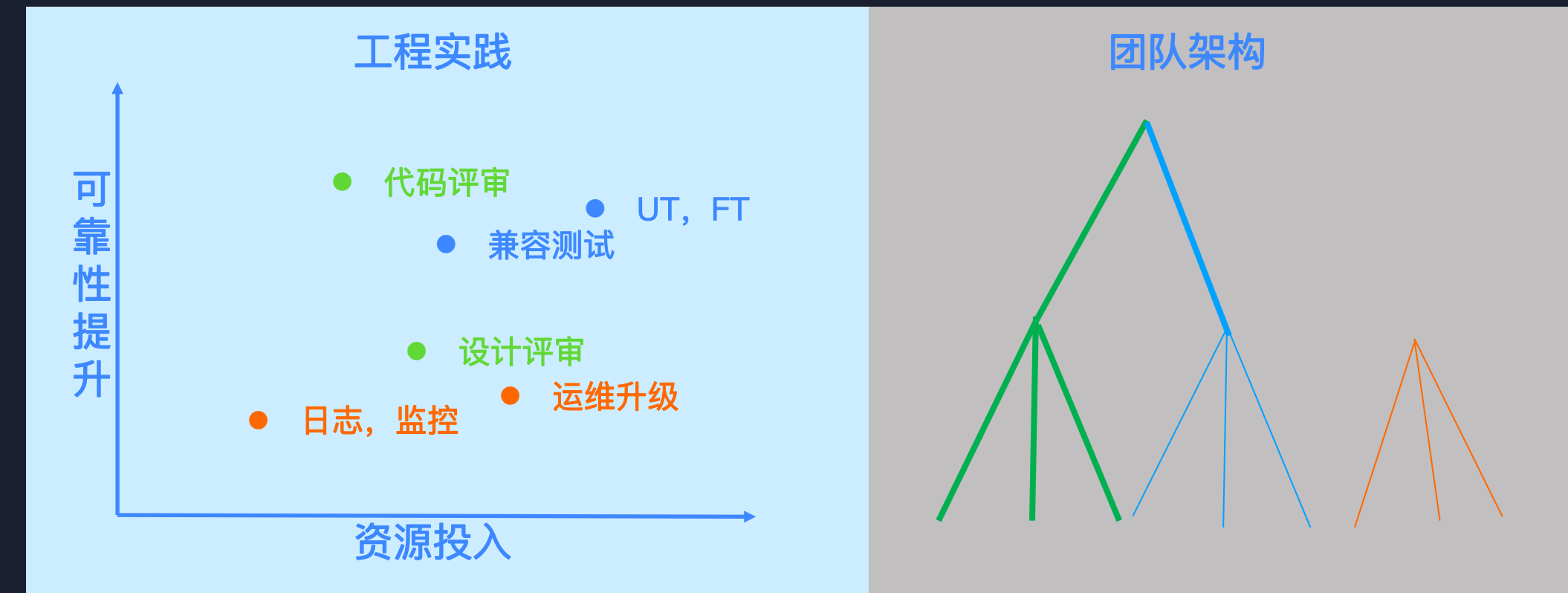
SRE

- Oncall
- Escalation
- 复盘

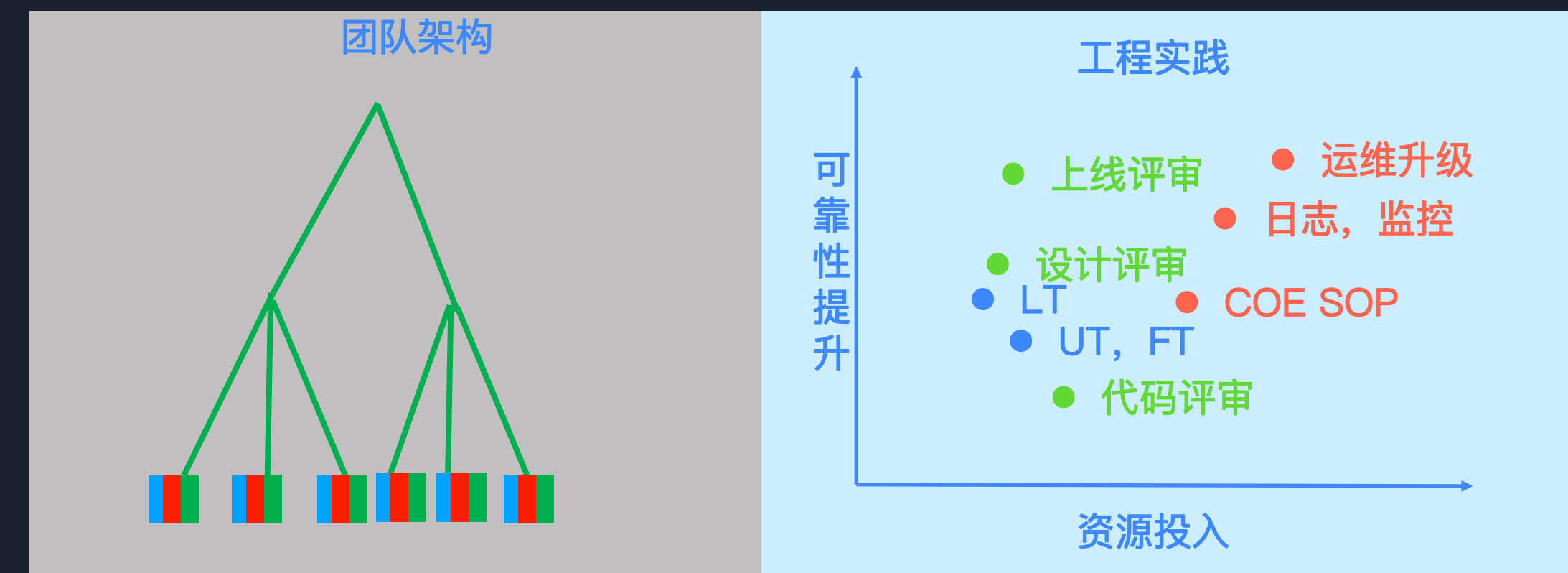
当质量提升陷入瓶颈或ROI变低时，
通过**增强监控和快速响应**能持续提升系统可靠性

系统可靠性：质量提升&快速响应

Microsoft

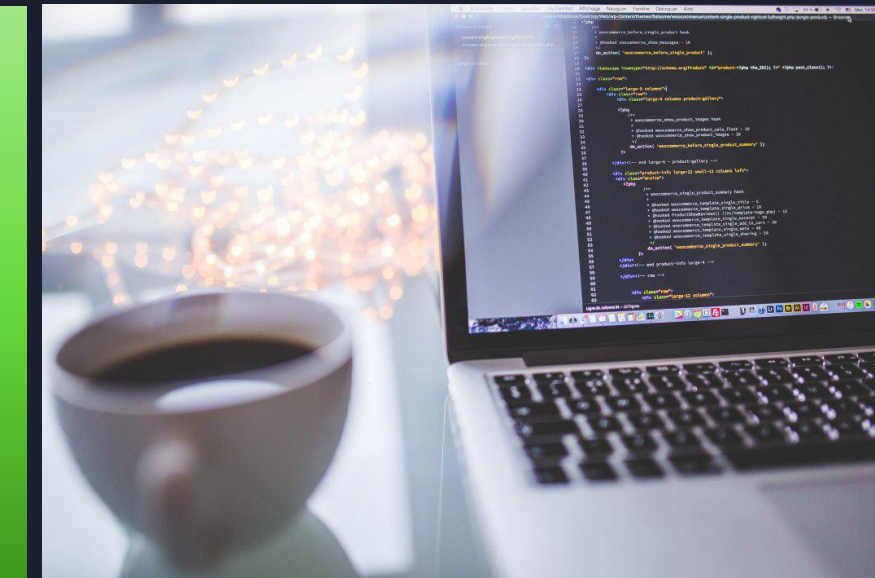


Amazon



系统可靠性：RCA分享

09:37



TTD (事故监测时间)

09:39



09:41



TTR (事故恢复时间)

- 09:48 有人误删...
- 09:57 无法冷启动
失败, 重试, 再失败
- 10:30 上CNN了
- 13:54 完全恢复

事件报警及时
响应升级迅速
数据备份恢复正常
团队应急有序

生产环境误操作
冷启动多年未试
失败重试缺少规范

复盘整改

系统可靠性：携程商旅实践分享



Quality is not an act, it is a habit."
-Aristotle

“品质不是一时的表现，是长久的习惯。”
- 亚里士多德

以自动机制培养良好的工程习惯
(Engineering Excellence)



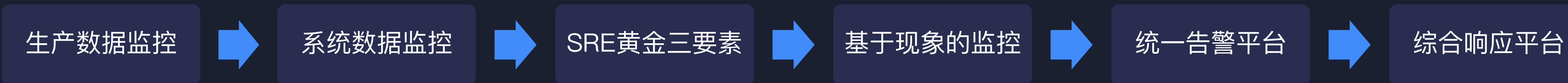
Sonar, Doc, COE



“Good intentions don't work,
mechanisms do.”
-Jeff Bezos

“良好的意愿是没有用的，建立机制才是关键！”
- 贝索斯

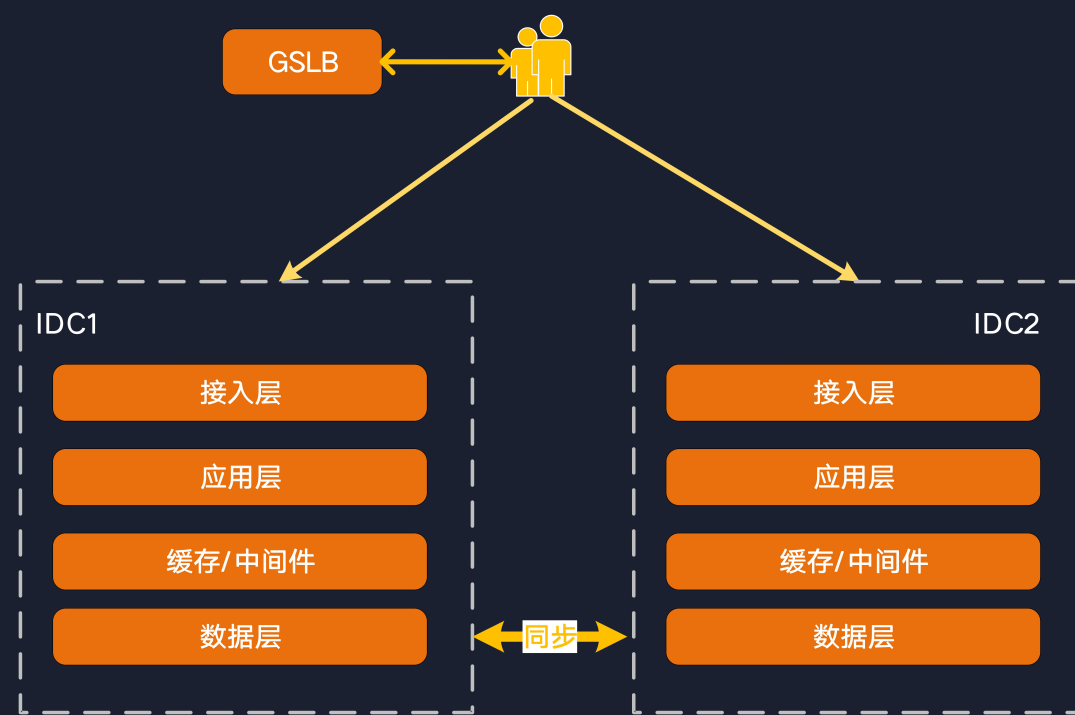
系统可靠性：携程商旅实践分享



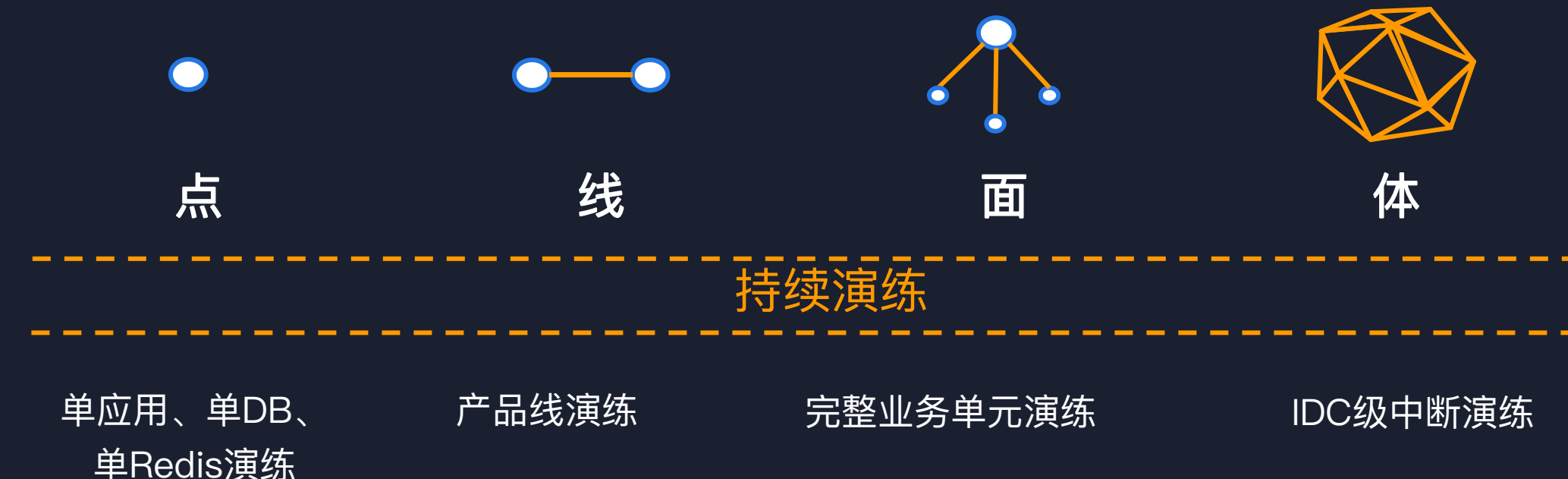
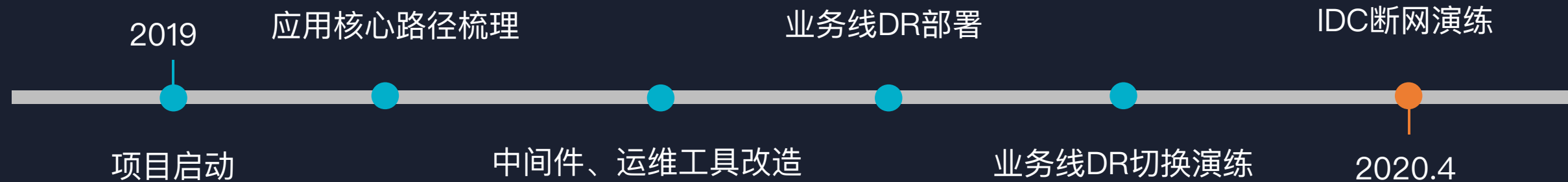
目标：5分钟检测，10分钟恢复



流量地球项目：“当一个太阳（IDC）发生不可用故障的时候，我们地球（业务）还可以继续活下去！”



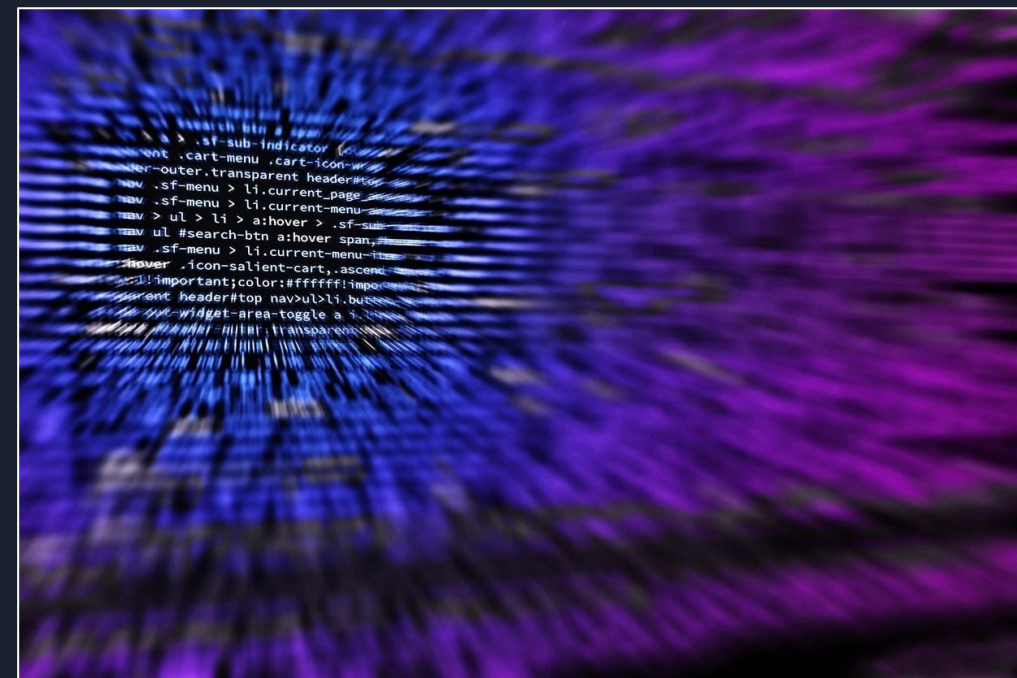
同城双活架构



技术展望



系统架构会更复杂，体量会更大



常规开发测试对可靠性的提升出现边界效应



快速反应 & 控制影响半径会成为关键

上海·2020

THANKS

上海·2020