

大数据平台质量保证 最后一公里探索与实践

沈立彬
Splunk 上海研发中心

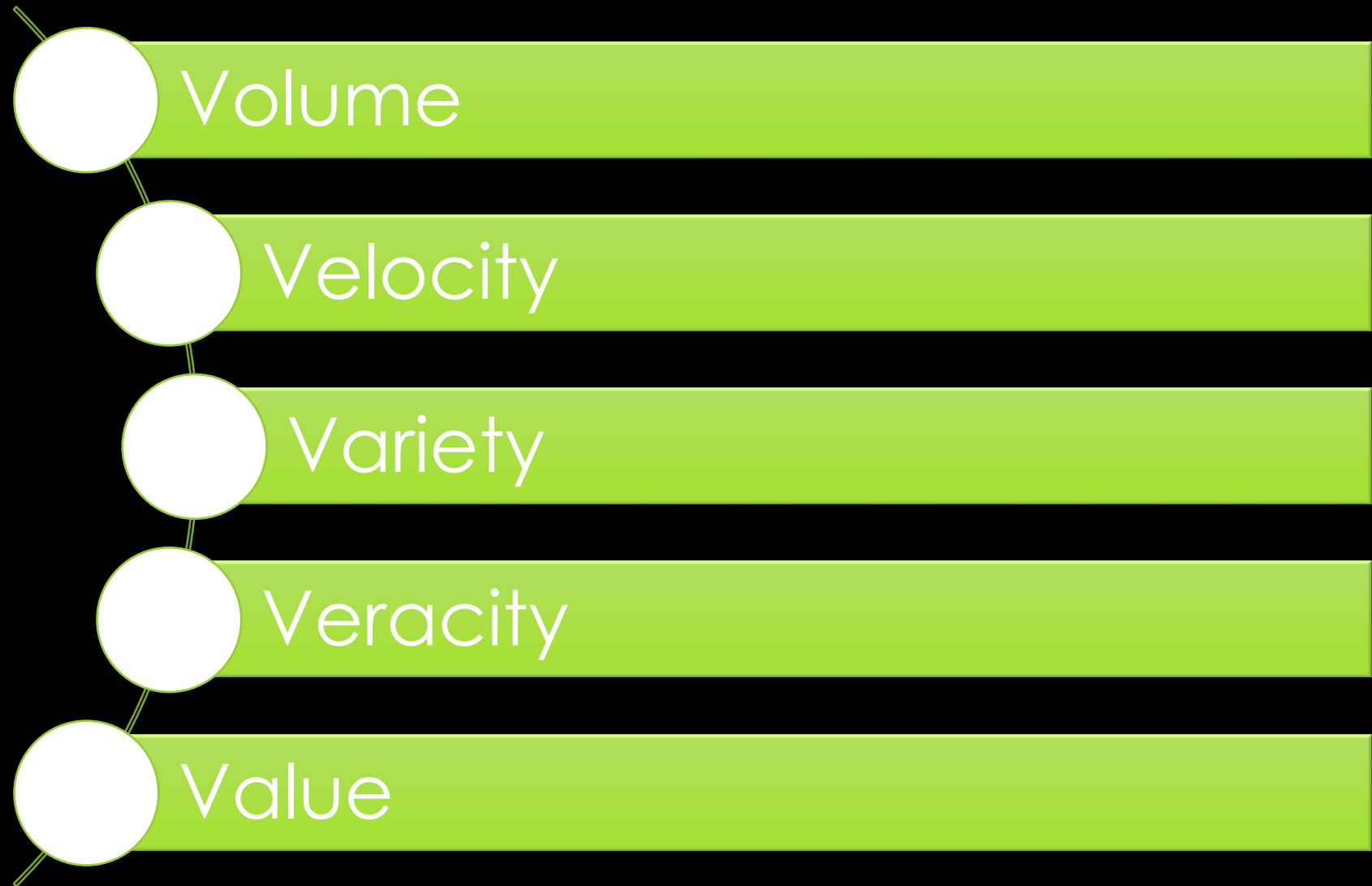
Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ material. For important factors that may cause actual results to differ from those captured in our forward-looking statements, please review our filings with the SEC.

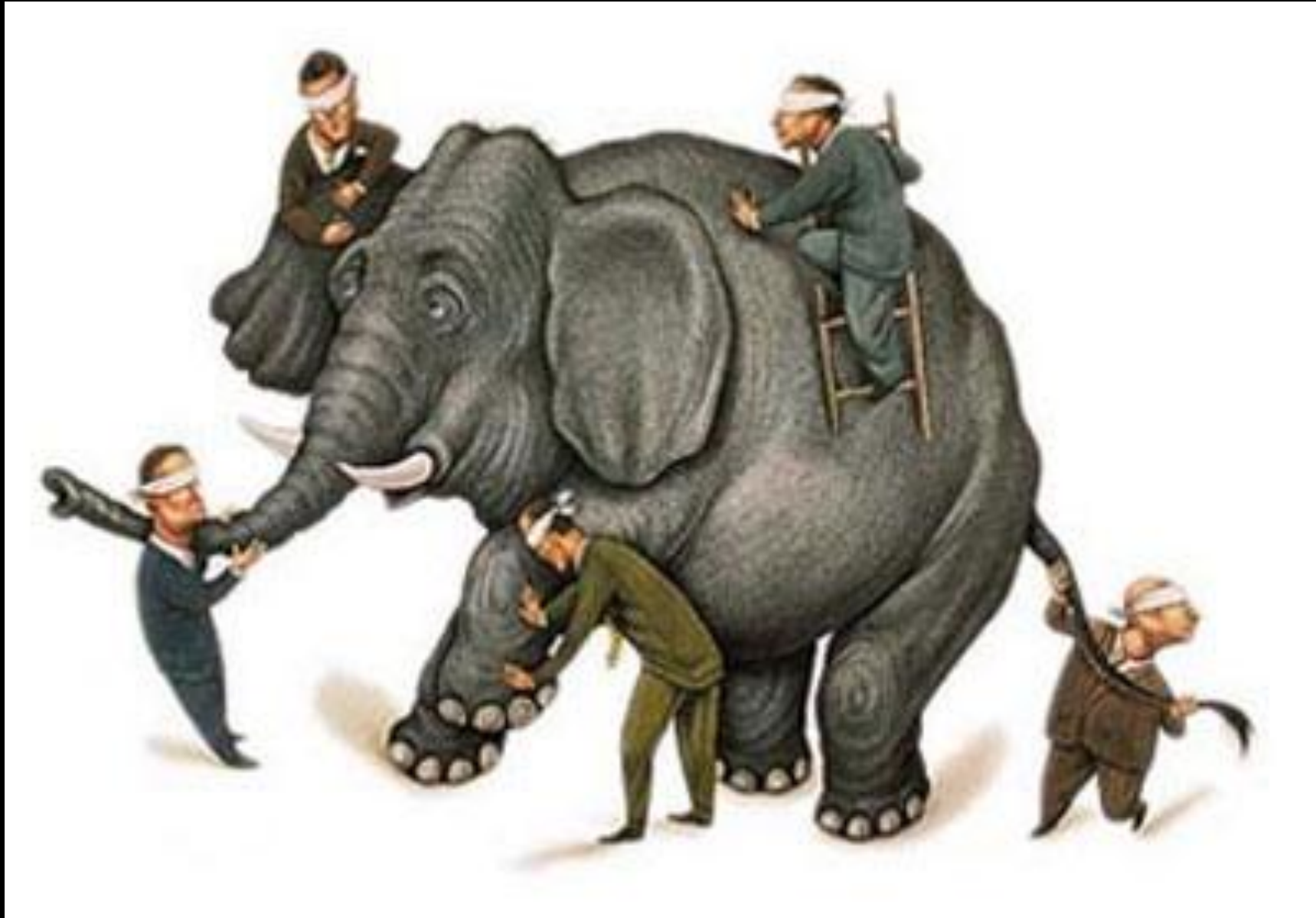
The forward-looking statements made in this presentation are being made as of the time and data of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2017 Splunk Inc. All rights reserved.

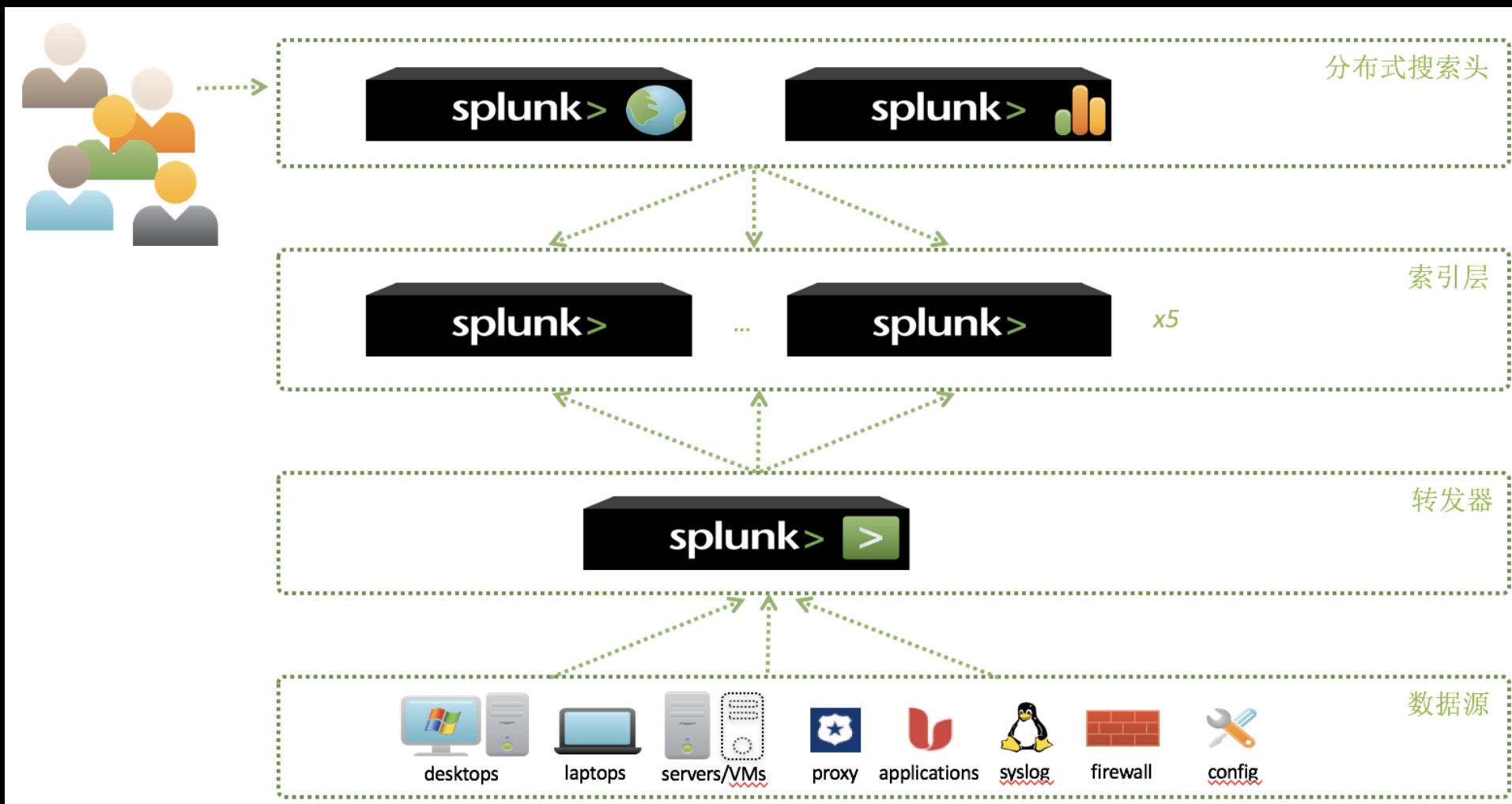
大数据



大数据测试



SPLUNK



功能测试覆盖已经很高了，为什么质量还是提高不上去

a) 遇到的问题

我们有数以万计的测试用例在执行，可是每次发布之后客户上报的各种问题依旧层出不穷

为什么

- Splunk产品的部署灵活多变，如何用有限的资源来保证最终的产品质量，让产品团队放心我们的产品可以很好的运行在每一个客户的部署里
- 用户日处理数据量从MB到PB，拓扑从一个节点到几千个节点，如何能更好的模拟用户场景进行测试

b) 前人的经验



我们做了什么

01 根本原因分析

- 整理，分析客户的问题，分析根本原因
- 反馈到我们的测试设计当中，逐渐削弱测试设计和实际使用中的缺口

03 海量数据与部署

- 数据构建与生成
- 容器技术的应用
- 部署配置生成与推送

02 系统测试优化

- 用户案例驱动的探索性测试
- 灾难恢复测试
- 迁移测试
- 多版本互操作性测试

04 监控工具开发

- 性能监控
- 崩溃分析
- 环境健康指数监控

100% 自动化测试?

纯手动

- 测试执行开销大
- 不能支持频繁的发布

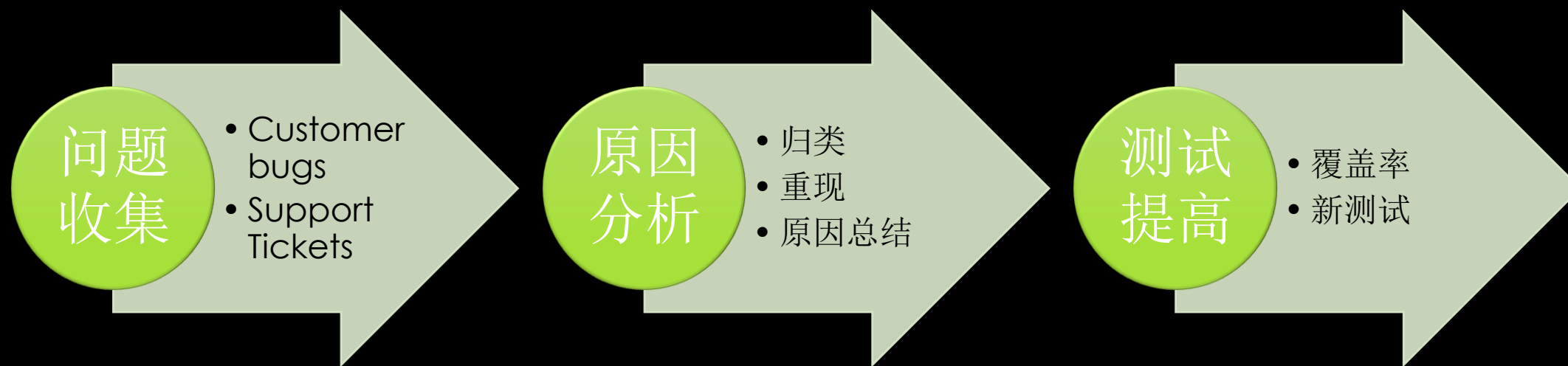
全自动

- 执行成本低
- 纯回归测试
- 场景测试维护成本高

“聪明”的自动化测试+探索性测试

- 执行成本较低
- 维护容易
- 大负载支持

根本原因分析(Root Cause Analysis)



我们发现了什么？

负载

故障

大数据产品质量
保证的特殊性

升级

多版本

我们发现了



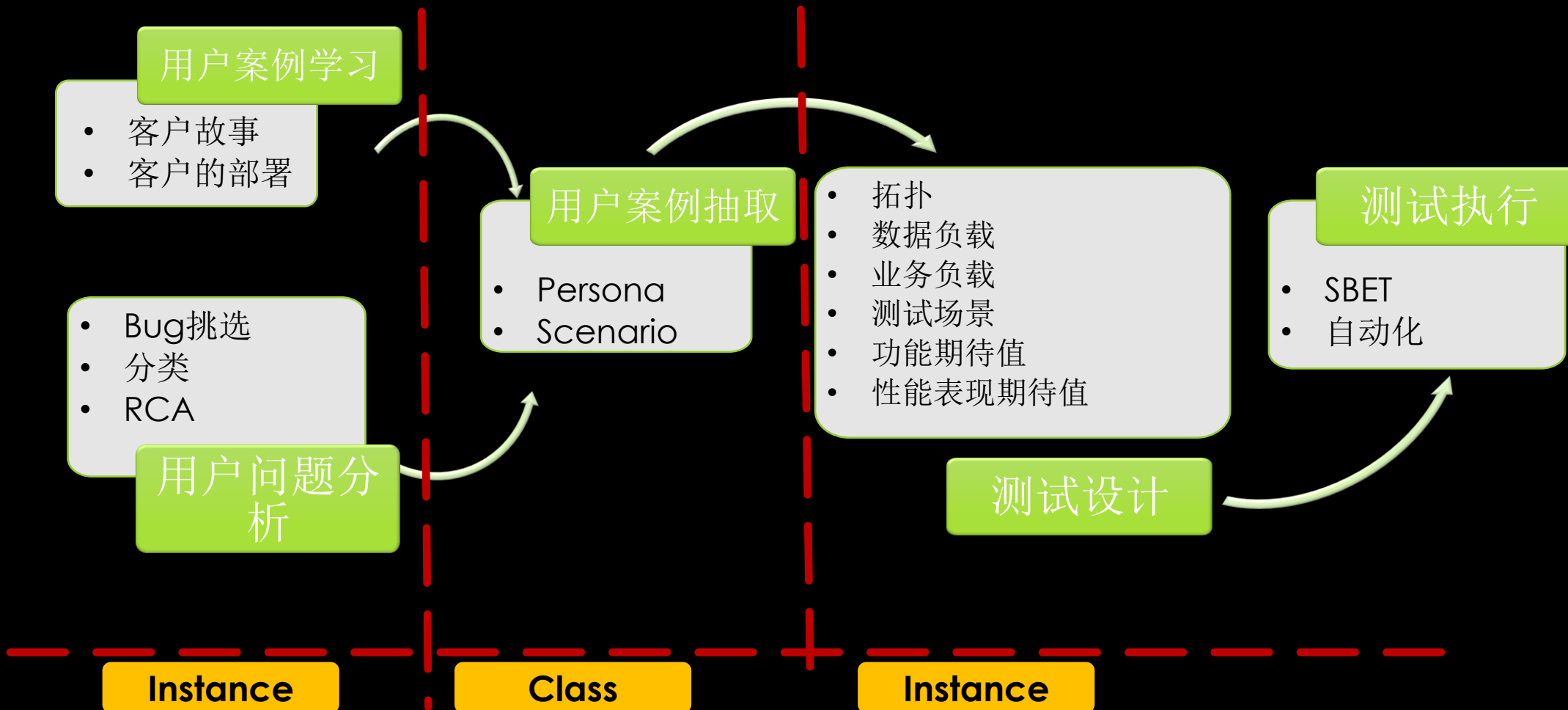
章

本

大负载下的系统测试



用户案例驱动的探索性测试



基于session的探索性测试



灾难恢复测试



灾难恢复测试

- 内存不足
- CPU过载
- IO繁忙
- 网络延迟
- 节点故障
- 站点故障

制造麻烦

恢复生产

- 功能验证
- 性能监控

迁移/升级测试

离线升级

在线升级

数据迁移

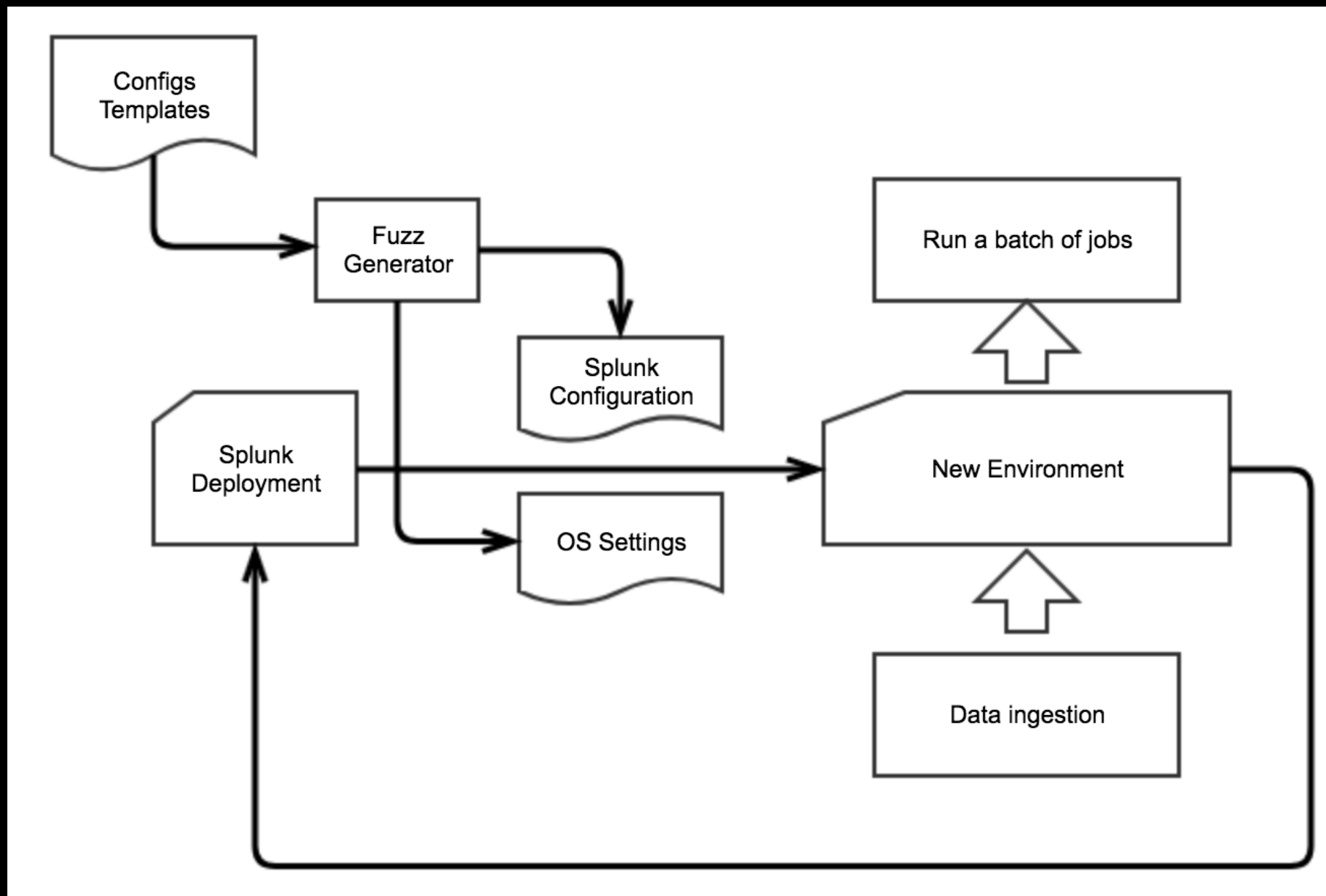
节点迁移

多版本互操作性测试

1. 多版本部署
2. 边界功能分析

Release (> Feature (V)	6.0	6.1	6.2	6.3
A	+	X	X	-
B	NA	+	X	-
C	NA	NA	+	X

部署FUZZ测试



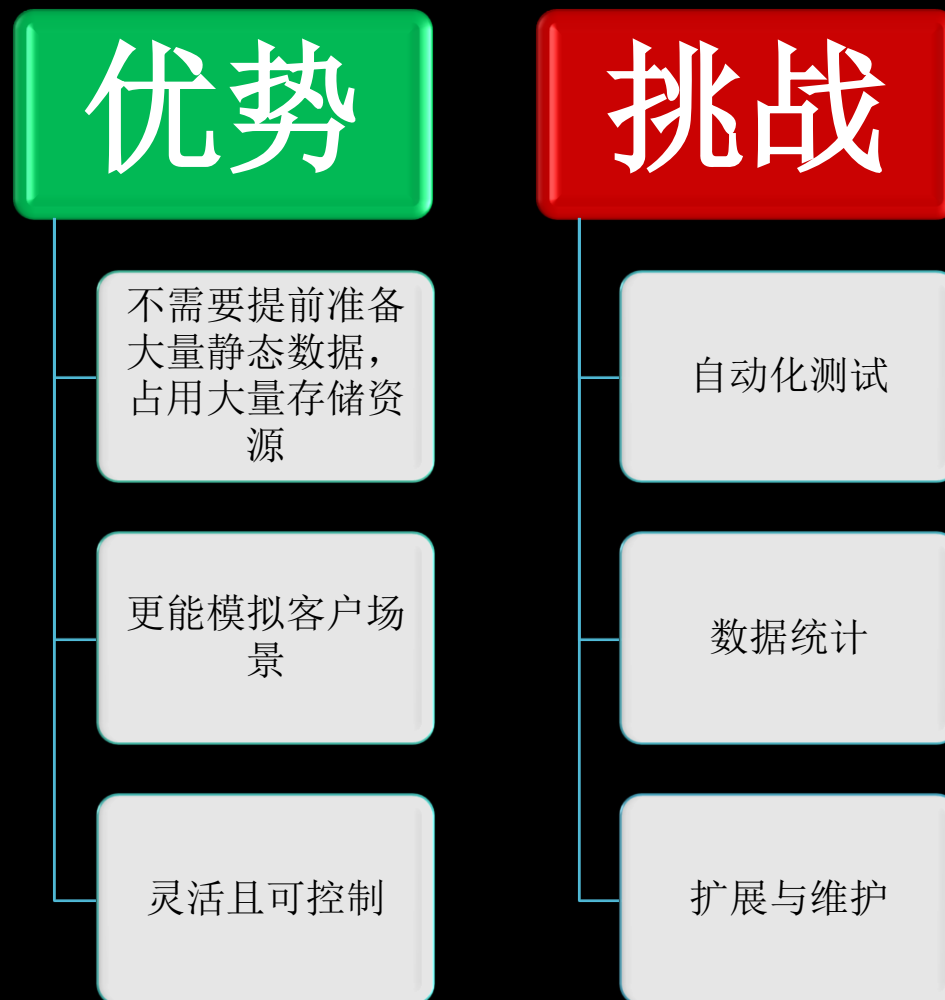
测试数据



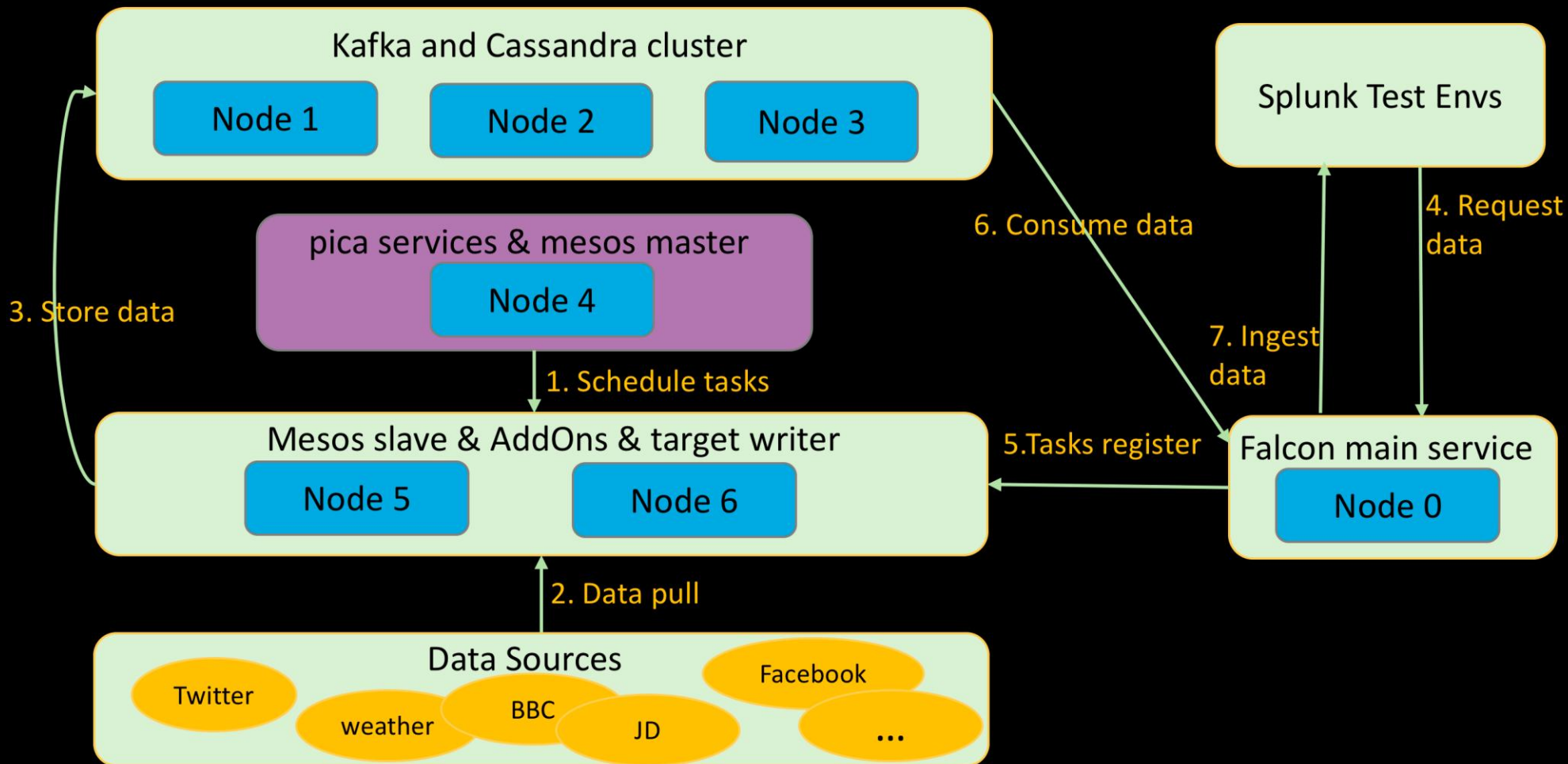
VS



实时数据



测试数据生成-数据爬取

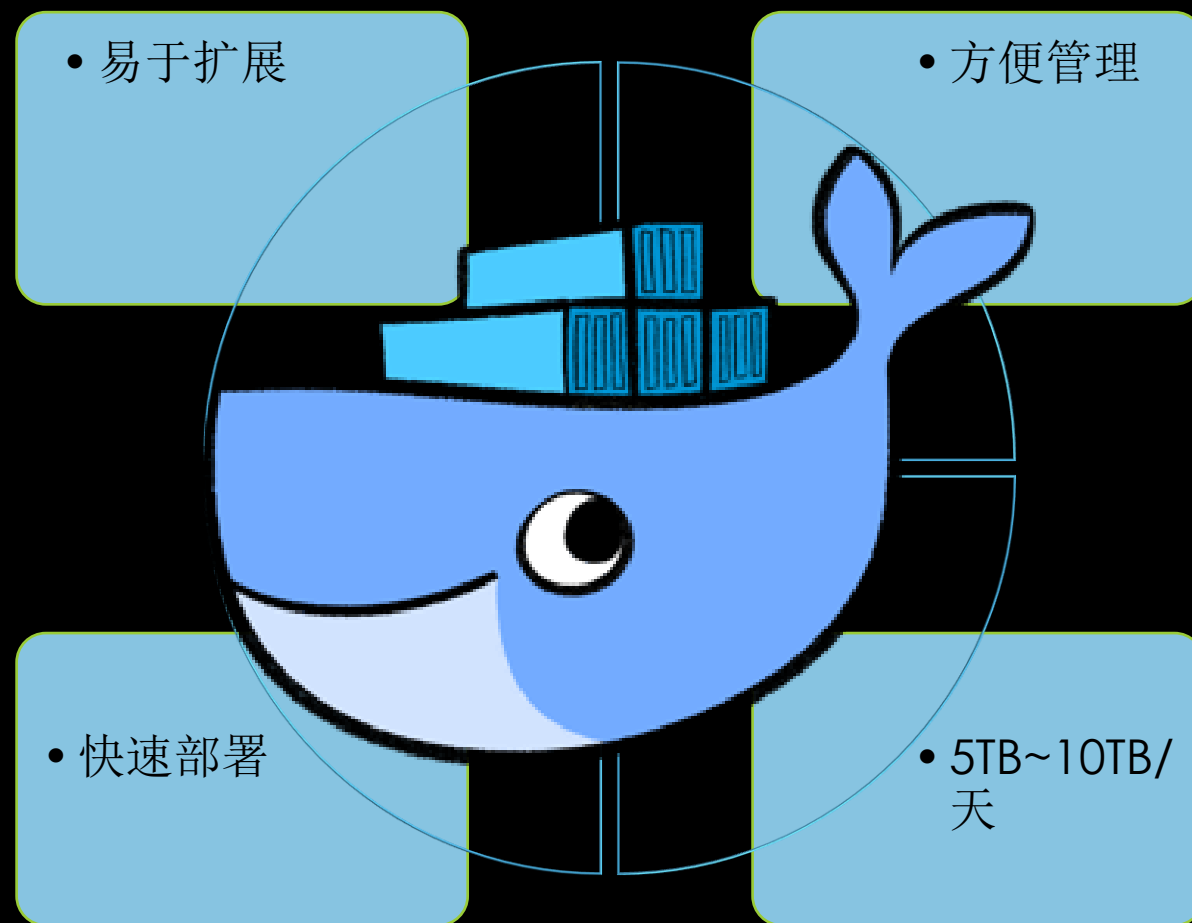


测试数据生成-基于样本生成

- 录制回放
- 多种输出支持
- 速度调节
- 高效

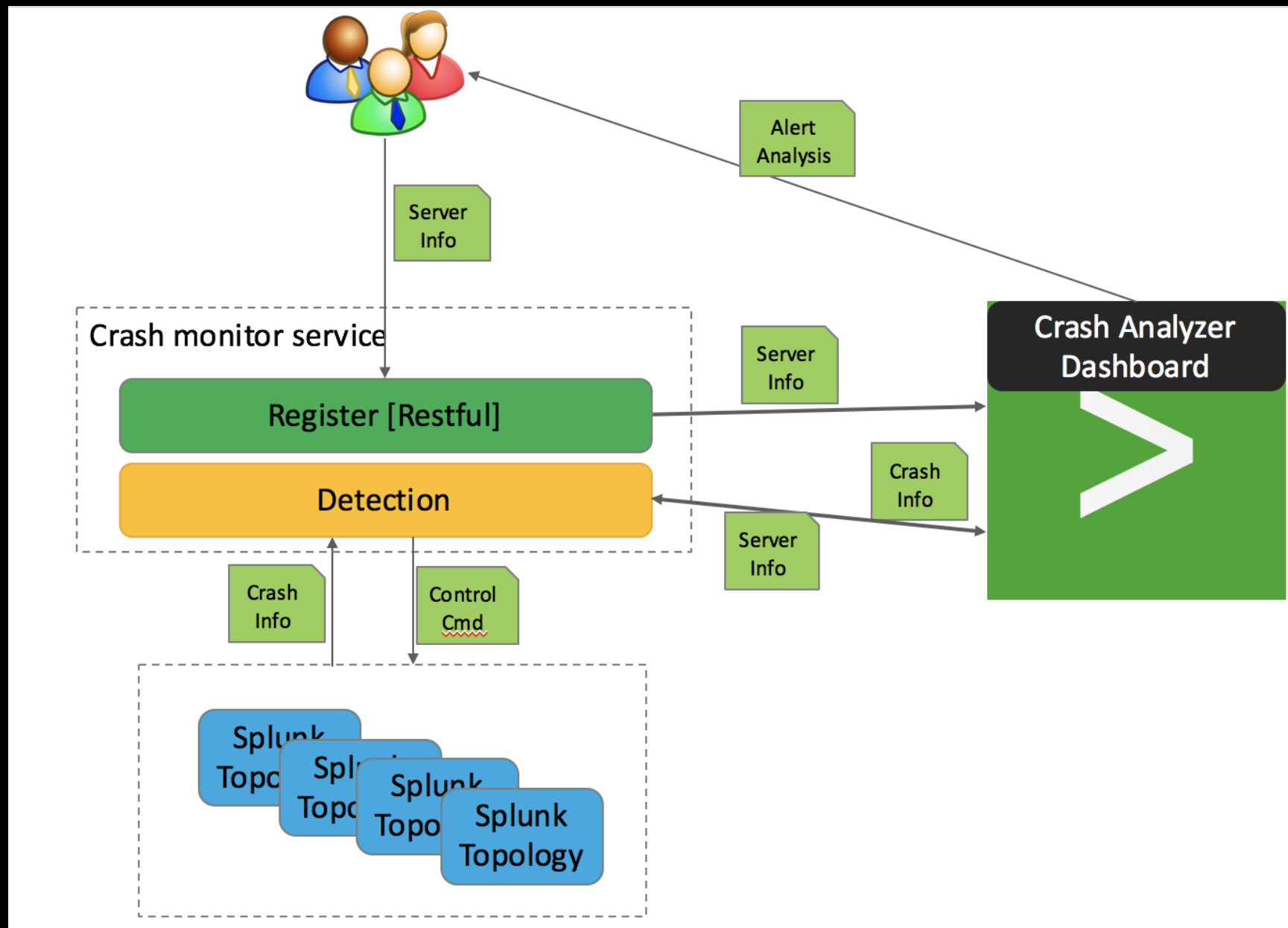
<https://github.com/splunk/eventgen>

测试数据生成-容器技术的应用

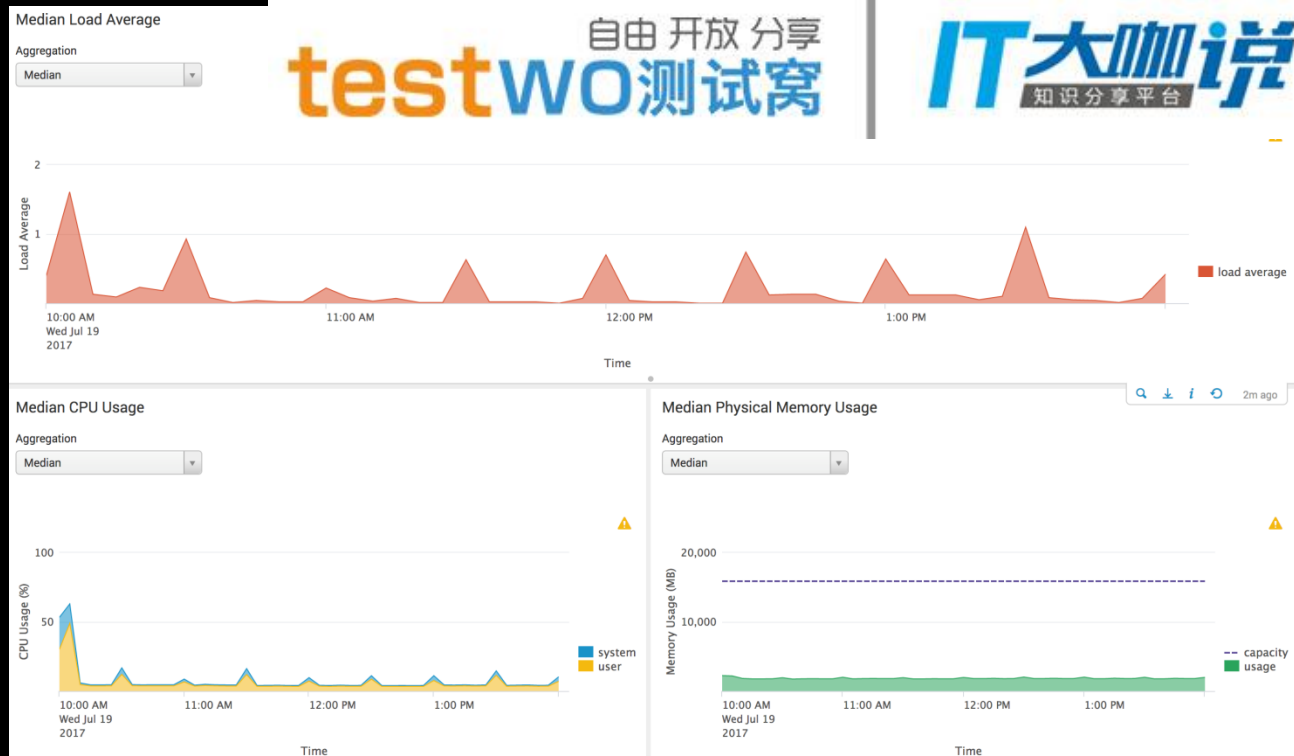


崩溃监控与分析

- Pull模式
- 实时indexing
- 亲缘度计算，归类

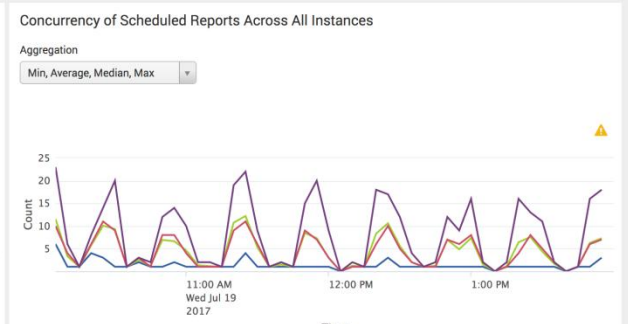
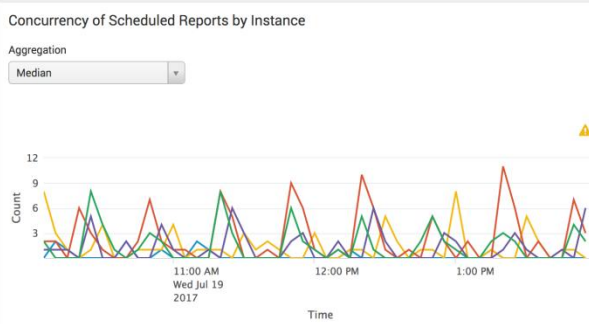
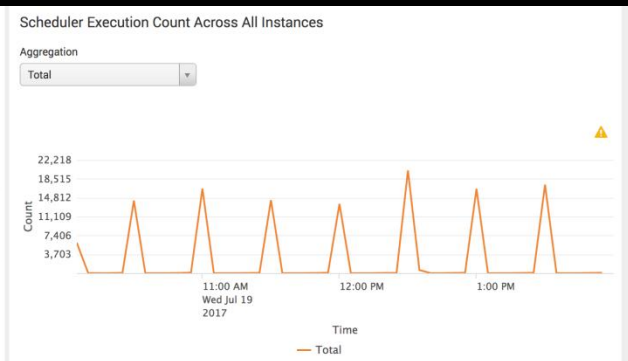


健康状况监控



自由 开放 分享
testwo 测试窝

IT大咖说
知识分享平台



splunk> Administrator Messages Settings Activity Help Find

Overview Health Check Instances Indexing Search Resource Usage Forwarders Settings Run a Search Monitoring Console

Health Check

Comprehensive health check for Splunk Enterprise Instances. To add additional items to this list go to: [Health Check Items](#)

Group: All App: All 0% complete

Tags: Category:

Check	Category	Tags	Results
Event-processing issues	Data Collection	event_breaking, indexing, timestamp_extraction	
Expiring or expired licenses	Data Indexing	licensing	
Indexing status	Data Indexing	indexing	
License warnings and violations	Data Indexing	indexing, licensing	
Local indexing on non-indexer instances	Data Indexing	best_practices, forwarding, indexing	
Missing forwarders	Data Indexing	forwarding	
Saturation of event-processing queues	Data Indexing	indexing, queues	
Distributed search health assessment	Data Search	distributed_search	
Search scheduler skip ratio	Data Search	scheduler	
Excessive physical memory usage	Splunk Miscellaneous	resource_usage	
Integrity check of installed files	Splunk Miscellaneous	configuration, installation	
KV Store status	Splunk Miscellaneous	kv_store	
Orphaned scheduled searches	Splunk Miscellaneous	configuration, search	
Upgrade opportunity from search head pooling to search head clustering	Splunk Miscellaneous	best_practices, configuration	
Assessment of server ulimits	System and Environment	best_practices, operating_system	
Linux kernel transparent huge pages	System and Environment	best_practices, operating_system	
Near-critical disk usage	System and Environment	capacity, storage	
System hardware provisioning assessment	System and Environment	best_practices, capacity, scalability	

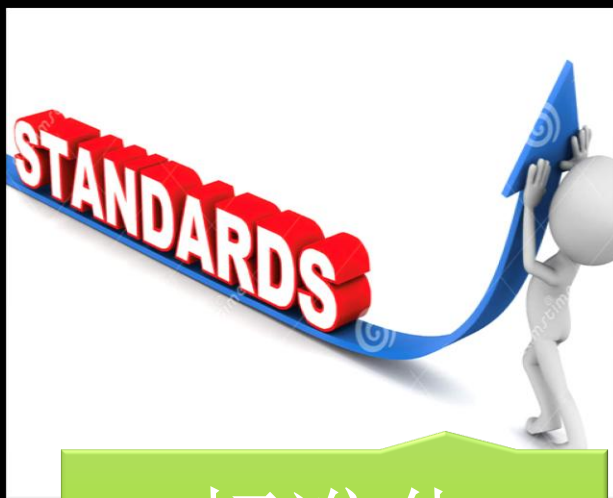
回顾

- 大数据测试有别于传统软件测试之一是数据量大，如果只拿样本数据不能有效的把潜在的用户问题在测试阶段探测出来，所以如何生成有价值的测试数据是至关重要的。有价值的测试数据应该表现在如下方面：
 - 效率性
 - 多样性
 - 合理性
- 基于复杂拓扑的大负载下的系统级测试，要能够做到从用户中来到用户中去
 - 案例来自于用户的真实需求和问题
 - 测试结果反馈到best practice
- 自动化测试

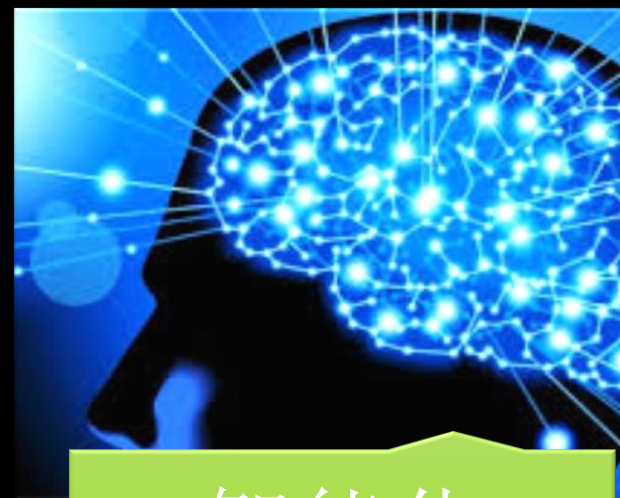
路漫漫其修远兮，吾将上下南北东西中发白而求索



系统化



标准化



智能化

Thank you, Q&A