

# 数据中心的性能优化

郭健美

2024年秋

# 本课内容

- **数据中心简介**
- 混部应用的性能干扰检修
- 数据中心的性能分析
- 数据中心的性能评价

# 数据中心简介

- 数据中心（data centers）是安置计算机系统及相关设备的基础设施
- 最早出现在1940年代初期，早期的计算机系统需要许多大型组件，操作员必须使用许多电缆连接这些组件。它们还会消耗大量电力，并且需要冷却以防止过热。为了管理这些称为**大型机（mainframes）**的计算机，公司通常将所有硬件放置在一个称为数据中心的房间中进行管理
- 近年来，随着云计算和虚拟化技术的引入，**云数据中心（cloud data centers）**开始出现并快速发展，第三方的云服务提供商开始管理和维护数据中心，并将基础设施以云服务的方式租用给其他公司或组织

# 数据中心简介

- 随着数据中心的规模越来越大，几十万甚至上百万台服务器的规模屡见不鲜，即便几个百分比的资源利用率提升也会带来相当可观的经济和社会效益
- 我国数据中心年用电量已经超过了三峡大坝年发电量



2020年中国数据中心耗电量为**2045亿千瓦时**  
预计2030年达到4000亿千瓦时

<https://www.163.com/dy/article/H8ITTKGG0512B07B.html>

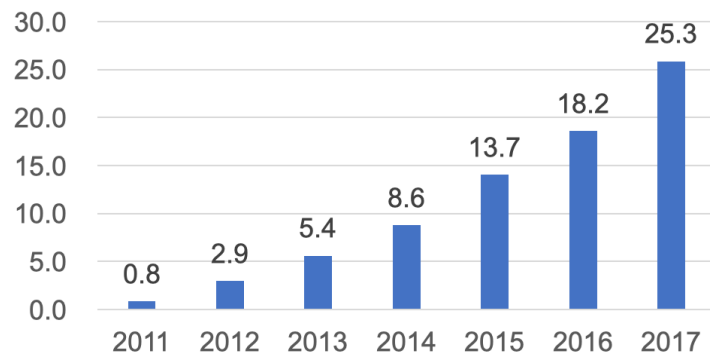


2020年三峡电站全年累计生产清洁电能**1118亿千瓦时**

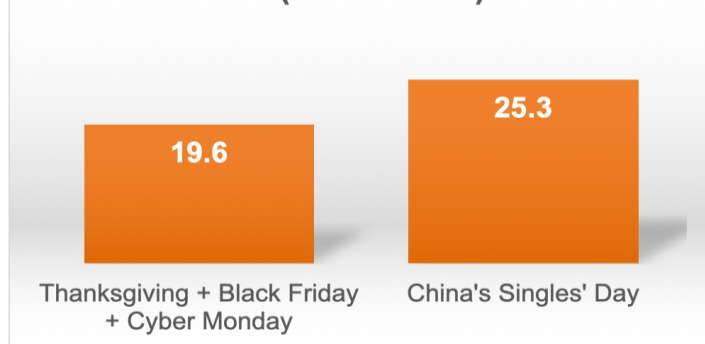
# 数据中心是关键的基础设施

- 数据中心是我国“新基建”和“卡脖子”工程的重要组成部分，也是电信、金融、教育、医疗等民生领域关键应用的基础设施

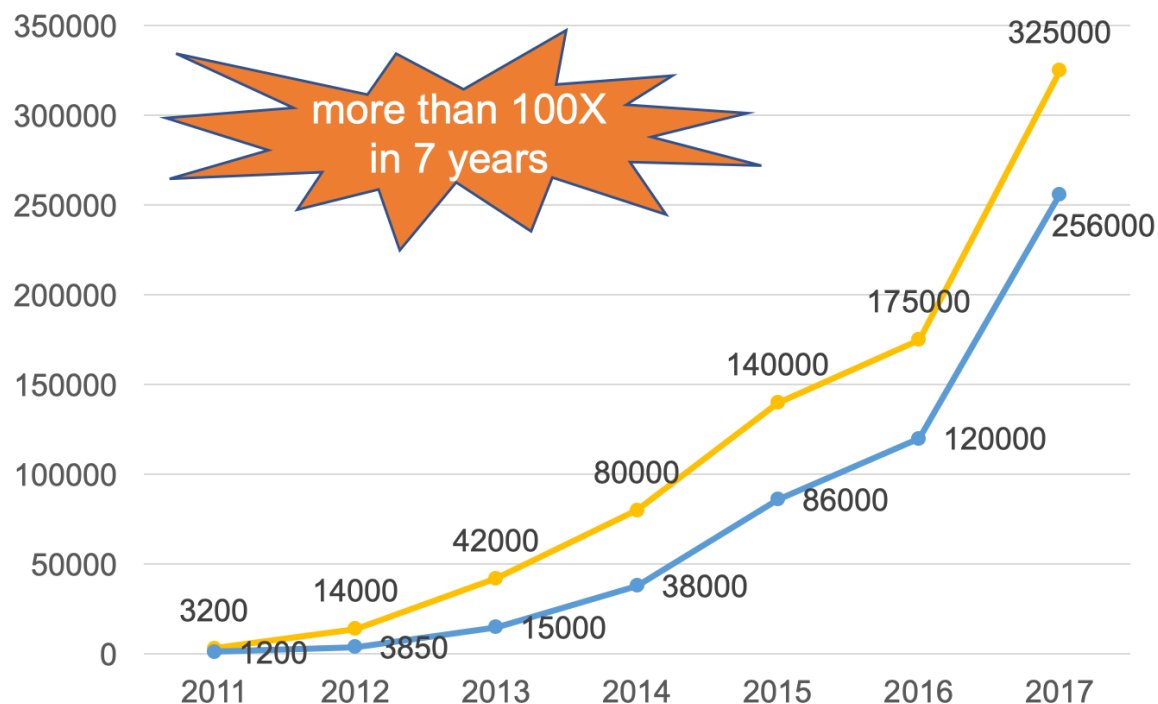
Total Sales (Billion USD)



Total Sale (Billion USD) in 2017



Peak Transactions / Second Peak Payments / Second



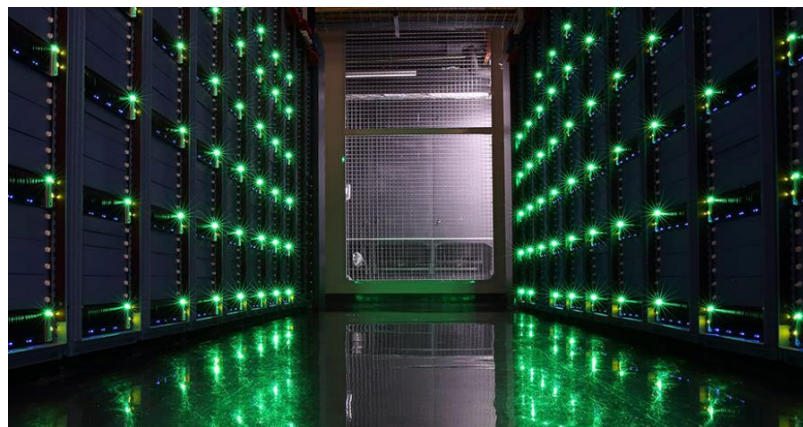
[GreenTea JUG Java Meetup, 上海 2018]





[http://www.xinhuanet.com/2020-07/14/c\\_1126233382.htm](http://www.xinhuanet.com/2020-07/14/c_1126233382.htm)

阿里巴巴张北数据中心总投资180亿元  
建设规模30万台服务器  
2017年阿里巴巴“双十一”的成交额1682亿元  
张北阿里云基地承担了40%的业务量  
支付峰值达到每秒25.6万笔



[https://blog.csdn.net/weixin\\_34174322/article/details/90590041](https://blog.csdn.net/weixin_34174322/article/details/90590041)

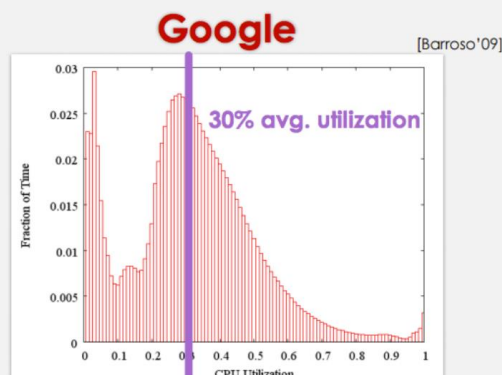
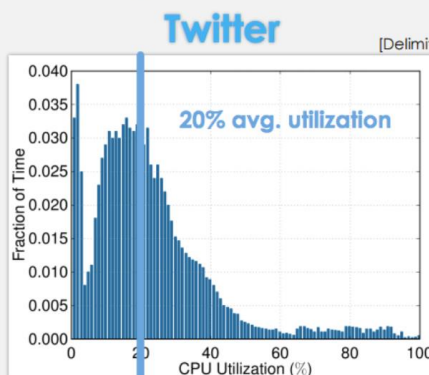
[http://www.gov.cn/xinwen/2018-02/28/content\\_5269382.htm](http://www.gov.cn/xinwen/2018-02/28/content_5269382.htm)



# 数据中心的两大任务

- **稳定性保障**：近年来全球各大云服务故障频发，大规模系统宕机和服务中断事件屡见不鲜
- **资源利用率提升**：数据中心资源利用率仍普遍较低，造成了资源和成本的巨大浪费

But the datacenters are poorly utilized!



Low utilization in large-scale clouds, even with automated management systems

David Lo Oral Defense April 7, 2015

8

阿里云香港机房故障致澳门多家网站无法访问！称已在紧急处理

南方都市报APP · 南都即时  
原创 2022-12-18 19:52



12月18日，南都记者从澳门特区司法警察局获悉，由于阿里云的香港机房节点发生故障，导致包括澳门多家网站及App自当日午时起无法访问使用。同日，阿里云就事件发布公告称，阿里云工程师已在紧急处理中。



澳門金融管理局  
AU TORIDADE MONETÁRIA DE MACAU

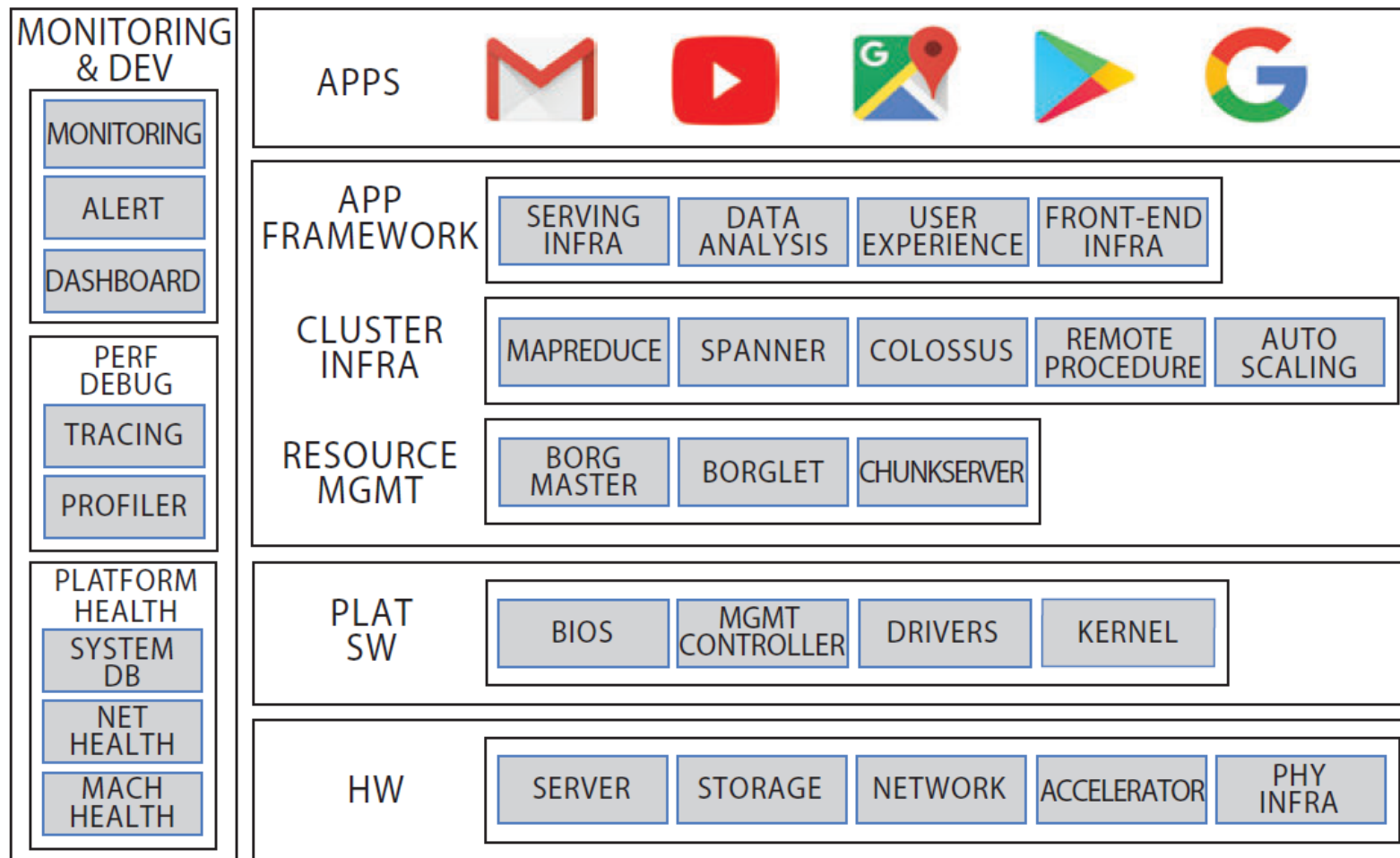


關於 ▾ 資訊發佈 ▾ 銀行業 ▾ 其他金融機構 ▾ 財經資訊 ▾ 保險業 ▾ 研究及統計 ▾ 貨幣 ▾



12月18日，澳门特区司法警察局发布公告指出，网络安全事故预警及应急中心当日较早前收到消息，由于阿里云的香港机房节点发生故障，导致澳门金融管理局、澳门银河、莲花卫视、澳门水泥厂等关键基础设施运营者的网站、澳觅和MFood等外卖平台、以及澳门日报等本地传媒应用程序，自当日中午开始暂时无法访问使用。网安中心已联系相关的关键基础设施运营者并跟进有关问题。

# 数据中心的软硬件栈



[L. A. Barroso, et al., The Datacenter as a Computer: Designing Warehouse-Scale Machines, Third Edition. 2019.]



# 本课内容

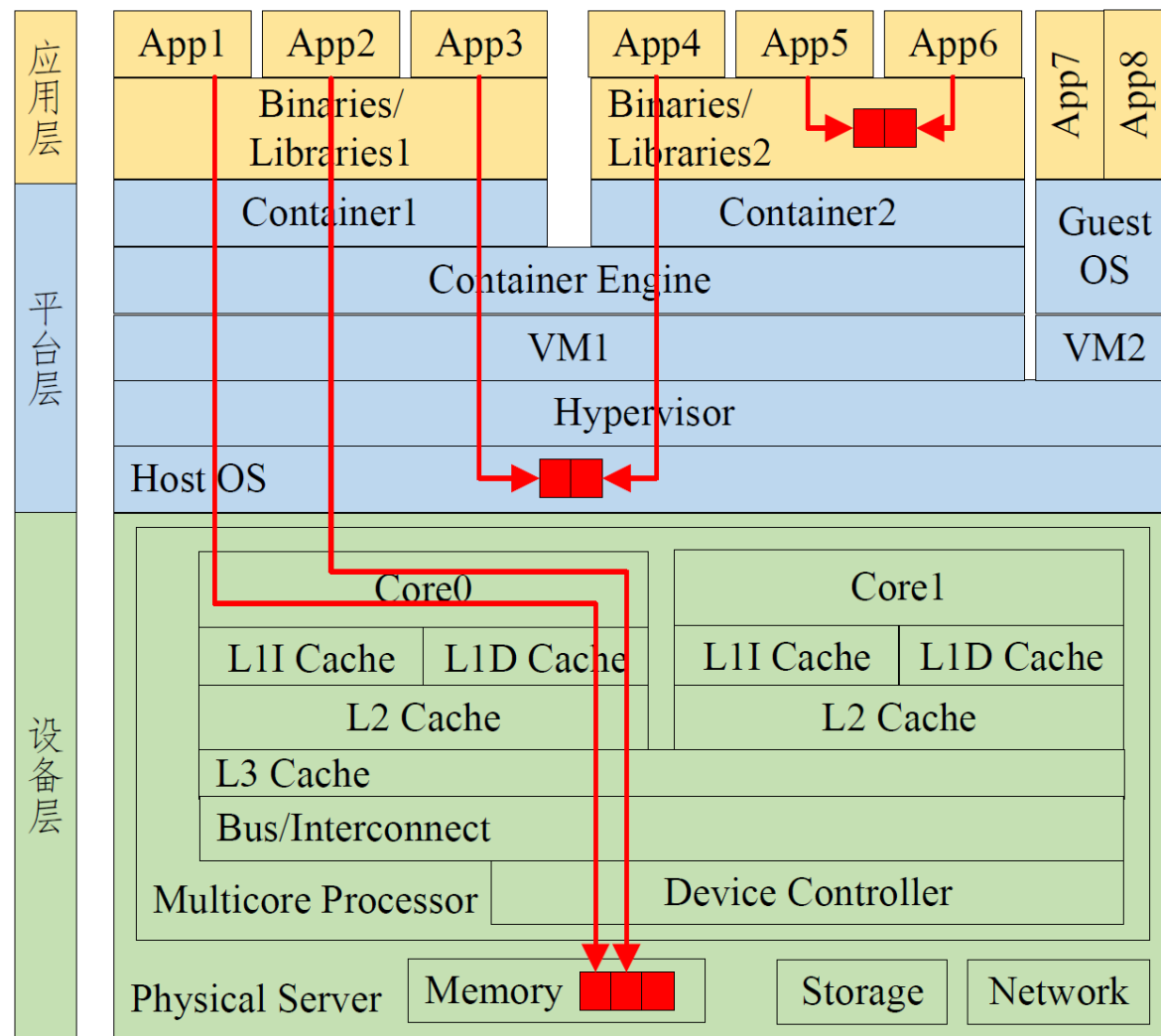
- 数据中心简介
- **混部应用的性能干扰检修**
- 数据中心的性能分析
- 数据中心的性能评价

# 数据中心内的应用混部

- 多应用混合部署，简称**混部 (colocation)**，是将多个软件应用混合部署在单个物理计算单元运行的模式。这里的物理计算单元，可以是处理器、服务器或集群等
- 数据中心通过混部为多个应用分配逻辑计算资源并共享物理计算资源，从而达到提高资源利用率并降低运营成本的目的
- 三个发展阶段
  - **宿主机**阶段，多应用直接部署在多核/众核服务器上
  - **虚拟化**阶段，多应用被封装在各自的虚拟机中
  - **容器化和微服务**阶段，多个应用（包括微服务）及其运行所依赖的二进制文件、库文件和环境配置等被打包到各自容器中

# 混部应用的性能干扰

- 多应用混部不可避免地带来了**性能干扰 (performance interference)**，即一个应用的性能由于另一个应用（或一组应用）的出现而受到影响
- 性能干扰通常会产生无法预期的性能行为和结果，引发性能异常、降低资源利用率、甚至导致系统失效和服务中断，严重影响应用的服务质量和稳定性
- 混部应用的性能干扰具有普遍性、复杂性、隐蔽性和动态性，为有效处理性能干扰带来了极大的技术挑战



# 性能干扰的检测

- 性能干扰属于一类较为复杂的性能异常
- 现有的性能异常检测技术相对比较成熟，主要基于统计分析和机器学习方法，通过分析大量性能相关历史数据，刻画和挖掘应用负载的正常工作模式并识别异常模式，从而检测性能异常
- 常见的可观测性能数据包括
  - 指标 (metrics)
  - 日志 (logs)
  - 链路追踪 (traces)



# 性能干扰的诊断

- 性能干扰的产生主要源于复杂软硬件栈上的资源竞争和大规模分布式应用的调用依赖，其诊断的关键在于根因分析和瓶颈定位
- 在资源竞争方面，目前的研究主要围绕处理器、内存、存储或网络资源争抢和性能瓶颈分析展开，主要方法包括微体系结构性能分析方法和 Brendan Gregg 提出的 USE 方法等
- 在大规模分布式应用的性能行为建模方面，主要通过分析各个软硬件的运行日志数据来学习和推理性能行为及其因果关系

# 性能干扰的修复

- 应用调度

- Kubernetes明确定义了不同优先级作业的接口和抢占机制，用户可以用YAML 语言定义应用间的亲和性（affinity）和反亲和性（anti-affinity）的规则描述，再通过命令式或声明式应用程序接口设定资源编排约束，实现应用、应用的容器、或放置容器的Pod 等单元的重新调度

- 资源隔离

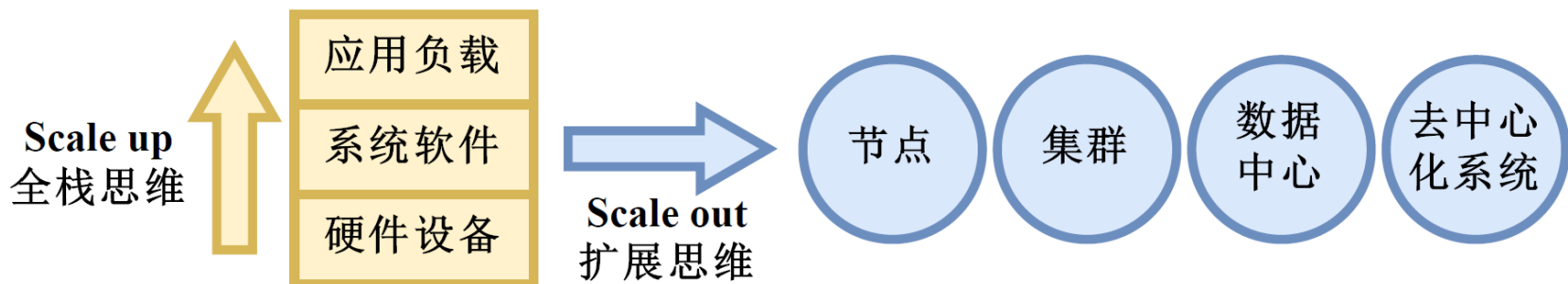
- 虚拟化和容器化本质上也是在多个应用之间建立资源隔离，防止性能干扰
- Linux cgroups 机制通过操作系统进程调度来限制应用单位时间内可用的CPU 时间，可以实现在线应用和离线应用的 CPU 资源分配
- Intel 资源调配技术（Resource Director Technology, RDT）和Arm 内存系统资源划分和监控（Memory system resource Partitioning And Monitoring, MPAM）技术也提供了多种监控和分配手段来控制最后一级高速缓存和内存带宽等硬件资源

# 本课内容

- 数据中心简介
- 混部应用的性能干扰检修
- **数据中心的性能分析**
- 数据中心的性能评价

# 性能分析方法论

- Scale up
  - **Iron Law of processor performance & CPI breakdown method** : J. L. Hennessy & D. A. Patterson, “Computer Architecture: A Quantitative Approach”, Morgan Kaufmann, 1990
  - **TMAM** : A. Yasin, “A Top-Down Method for Performance Analysis and Counters Architecture”, ISPASS 2014
  - **USE** : B. Gregg. “Thinking methodically about performance”. Commun. ACM 56(2): 45-51 (2013)
- Scale out
  - **GWP**: G. Ren, et al., “Google-Wide Profiling: a continuous profiling infrastructure for data centers”, IEEE Micro, 2010
  - **WSMeter** : J. Lee, et al., “WSMeter: A Performance Evaluation Methodology for Google's Production Warehouse-Scale Computers”, ASPLOS 2018
  - **RUE / RCU**: J. Guo. “From SPEC Benchmarking to Online Performance Evaluation in Data Centers”, LTB 2022 Keynote.





# 数据中心的性能分析

## 一个简单的例子

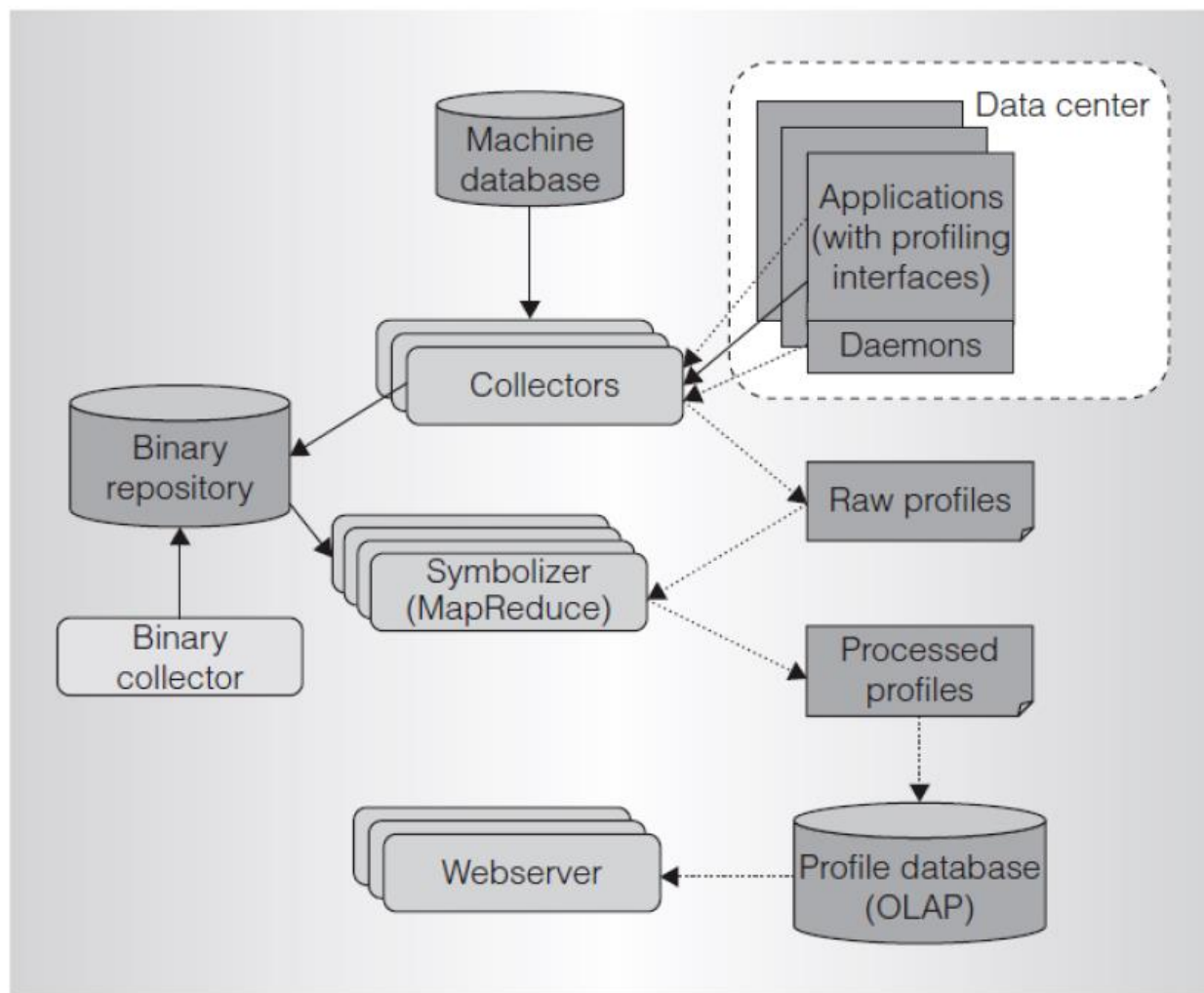
- 假设某数据中心有50 万台服务器，某个应用的优化手段X 可以降低单机CPU利用率20%，优化效果相当不错，但该应用及其优化手段X 只适用于整个数据center里1% 的服务器，即5 千台服务器
- 而另一个在某系统库上的优化手段Y 只能降低单机CPU 利用率3%，看上去优化效果一般，但该手段适用于整个数据中心80% 的服务器，即40 万台服务器
- 那么，从整体来看，X 对整个数据中心的优化效果为 $20\% \times 1\% = 0.2\%$ ，而Y 的优化效果为 $3\% \times 80\% = 2.4\%$ ，显然优化手段Y 对于数据中心的整体优化效果更明显

# GWP

- 一套在数据中心层面以较低开销进行持续剖析（continuous profiling）的工具框架
- 通过采样（sampling）以较低的开销在数据中心的不同机器和不同时间段上获取各种数据
- 根据Google 当时的报道，GWP 每天从数千台服务器上收集几千个应用的相关数据，每天收集到的数据经过压缩也会达到几个GB
- 相比于传统的单机性能分析，这样大规模的性能数据采集和分析本身极具技术挑战

[G. Ren, et al., "Google-Wide Profiling: a continuous profiling infrastructure for data centers", IEEE Micro, 2010]

# GWP 的基本实现框架



[G. Ren, et al., "Google-Wide Profiling: a continuous profiling infrastructure for data centers", IEEE Micro, 2010]

# GWP 的作用

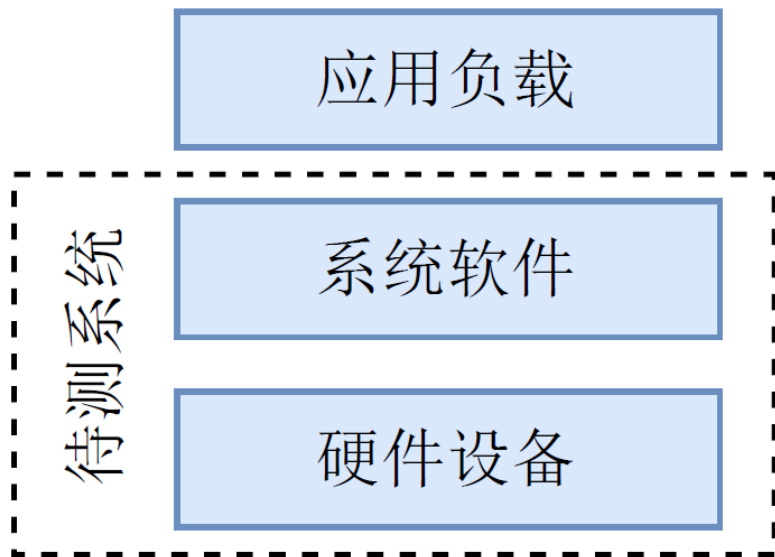
- 形成数据中心级别的全局性能画像
- 识别数据中心层面最热的共享代码，尤其那些在任意单个应用程序都不占很大比例但总体聚合后却消耗最多资源的热点程序
  - 例如，当时发现用于数据压缩的 zlib 函数库竟然消耗了整个数据中心约5% 的 CPU 资源
- 通过PMU等数据的分析，定位造成性能瓶颈的硬件部件
  - 例如，通过分析浮点计算占整个数据中心所有计算的比重，判断是否需要引入专门的协处理器来加速浮点计算
- 所采集的性能画像数据进一步支持编译优化
  - 剖析引导的优化 (Profile-Guided Optimization, PGO)
  - 反馈导向优化 (Feedback-Directed Optimization, FDO)



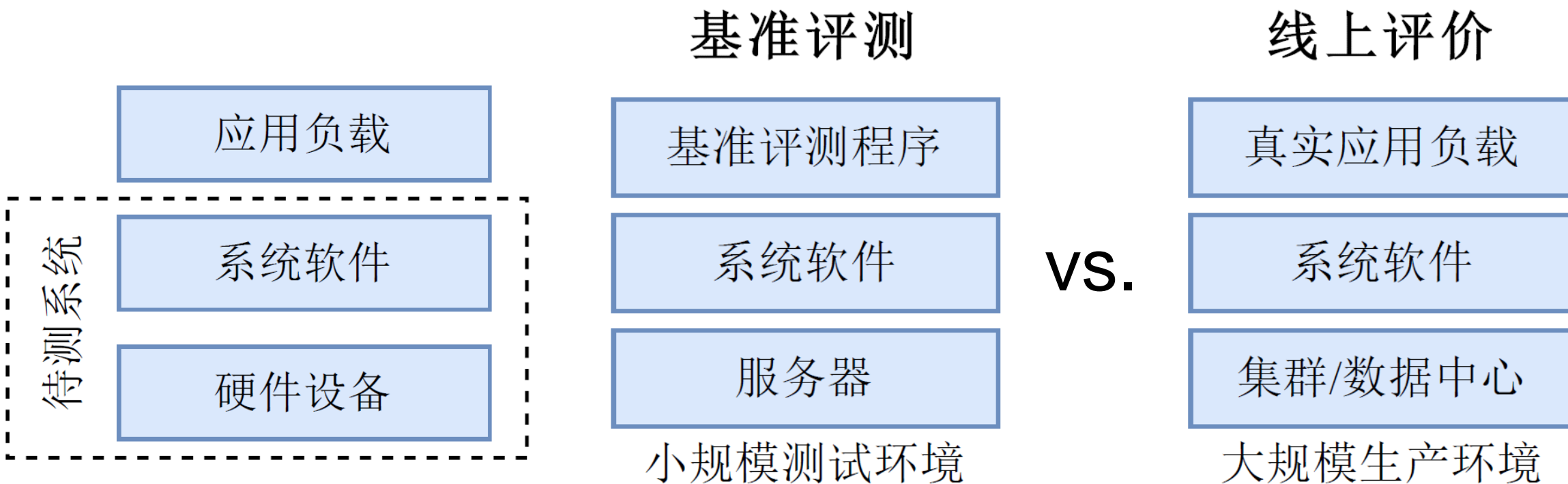
# 本课内容

- 数据中心简介
- 混部应用的性能干扰检修
- 数据中心的性能分析
- **数据中心的性能评价**

# 性能评价中待测系统的软硬件栈



# 性能评价的两种场景



# 数据中心的线上评价极具技术挑战

## 基准评测

- 运行和评价一组固定的基准评测程序
- 小规模测试环境 (几台或几十台服务器)
- 通常有固定的运行规则和BKC
- 采集和分析的开销可控



## 线上评价

- 每天运行和评价生产环境中各种真实应用
- 大规模生产环境 (上千到上百万台服务器)
- 部署环境复杂多样, 不同应用的开发运维规则各不相同, 还有混部造成的难以预期的性能干扰
- 采集和分析的开销可能很大, 需要考虑投入产出比



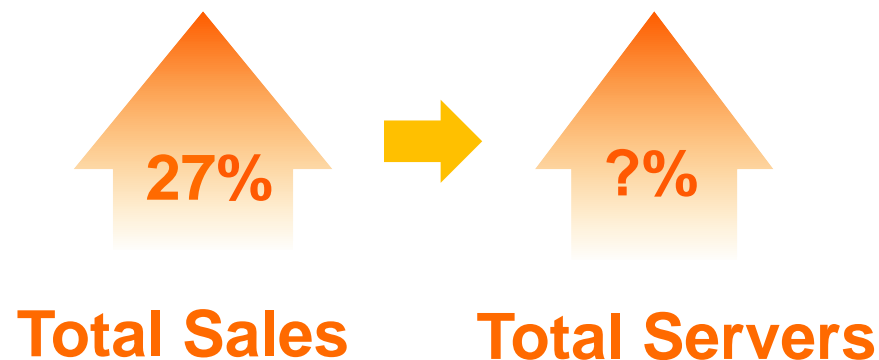
VS.



# 线上评价对数据中心的关键决策至关重要

- 容量规划 (capacity planning)

- 针对整个数据中心
- 针对单个应用

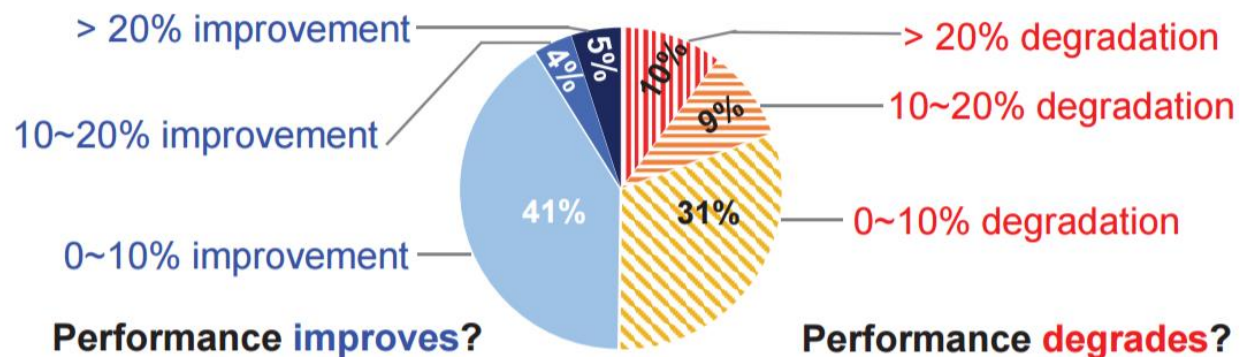


- 软硬件升级

- 处理器: x86 vs. Arm
- OS: RHEL vs. Ubuntu
- 编译器: GCC vs. LLVM
- JVM: Oracle JDK vs. OpenJDK
- ...

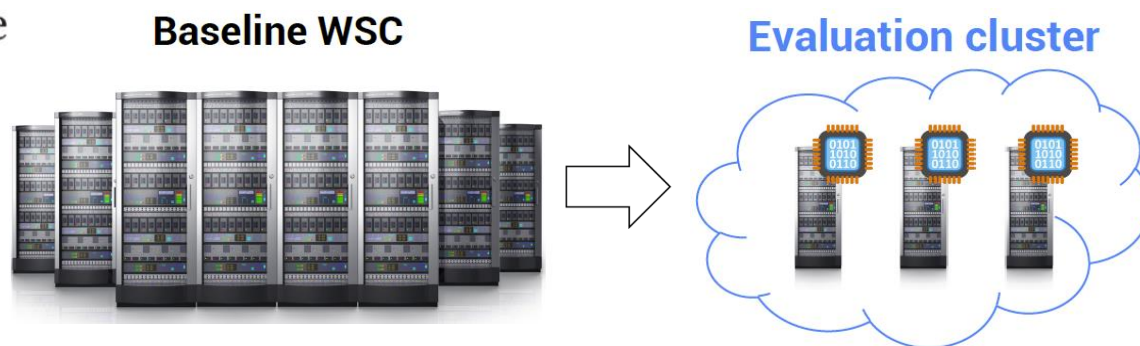


# Google WSMeter



**Figure 1.** Performance impact of a new DVFS policy on 1000+ production jobs. The pies represent the portion of the jobs falling into the category.

$$WSMeter = \sum_i Weight_i \times IPC_i$$



[J. Lee, et al., WSMeter: A Performance Evaluation Methodology for Google's Production Warehouse-Scale Computers. ASPLOS 2018: 549-563]

# Alibaba RUE

## System Performance Estimation, Evaluation and Decision (SPEED)

### Estimation

- What happens now?



### Evaluation

- What is the influence of SW/HW upgrade?



### Decision

- What happens in future?



### Validation

- What are the real effects?



## Resource Usage Effectiveness

System tools

$$RUE = \frac{\text{Resource Usage}}{\text{Work Done}}$$

Smaller is better

Workload specific

- Resource usage: CPU, Memory, Storage, Network
- Work done: Queries, Tasks

[郭健美. SPEED: 大规模数据中心的性能分析平台. 2019年杭州云栖大会阿里云系统软件开发者专场特邀报告]

[周经森, 郭健美, 李三红. Software Driven Cloud Optimization. 2020年 Arm DevSummit China (Arm中国开发者峰会)特邀主旨报告]

[Li Yi, Cong Li, Jianmei Guo: CPI for Runtime Performance Measurement: The Good, the Bad, and the Ugly. IISWC 2020: 106-113]

[Yang et al. Alibaba: Realizing Computing Power in Hyper-Scale Cloud Clusters. Intel White Paper 2021 <https://www.datacenterdynamics.com/en/whitepapers/alibaba-realizing-computing-power-in-hyper-scale-cloud-clusters/>]

# SOLE RCU (Resource Comprehensive Utilization) 资源综合利用率

技术能力	Google WSMeter	Alibaba RUE	SOLE RCU
单点/集群性能评价	√	√	√
系统资源使用量评价	√	√	√
业务负载工作量评价	X	√	√
非线性评价模型	X	X	√
层次化聚合方法	X	X	√

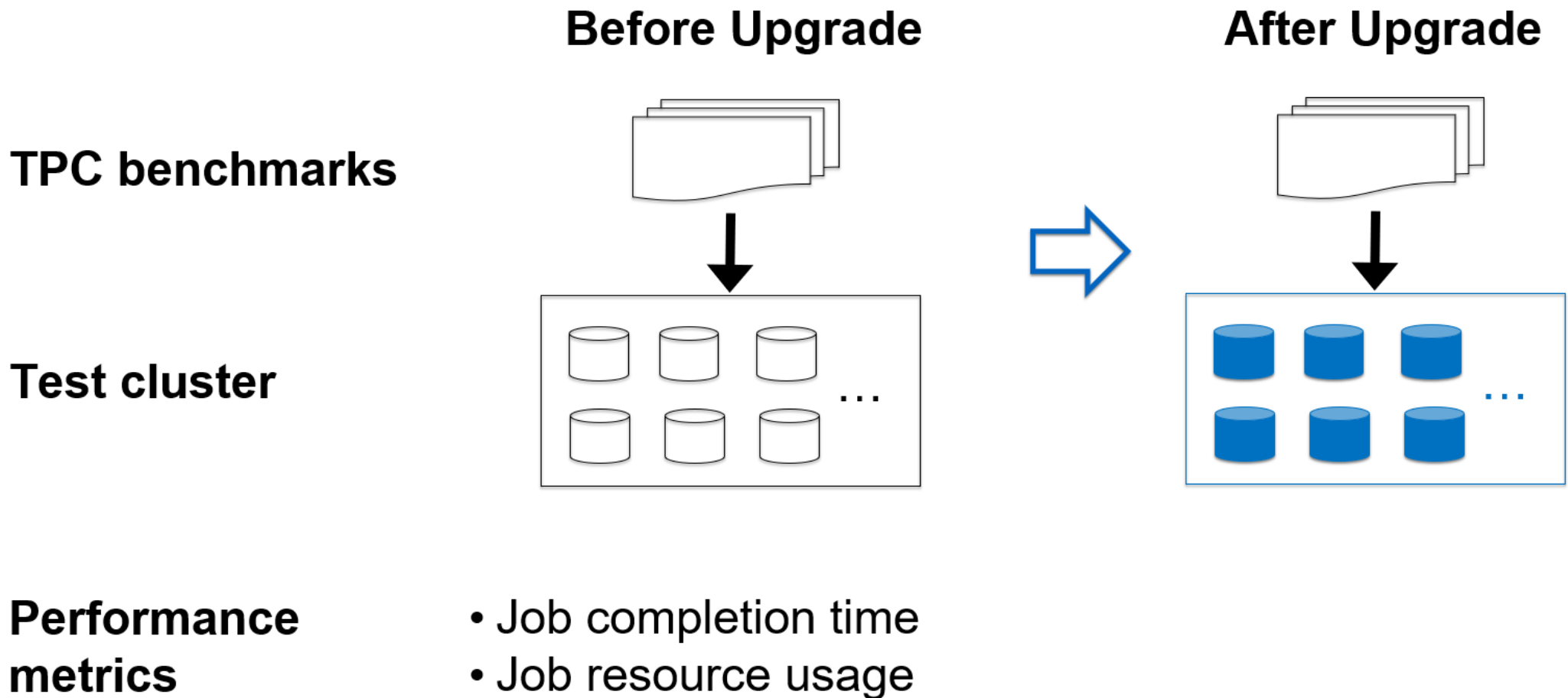
[郭健美, 黄波. 一种资源综合利用率的层次化分析方法、计算设备及存储介质. 国家发明专利申请 202110779100.9]

# 应用案例

- Case 1: CPU Upgrade for Big Data Processing Platform
- Case 2: JVM Upgrade for E-Commerce Applications
- Case 3: A Pitfall of Data Science: Simpson's Paradox

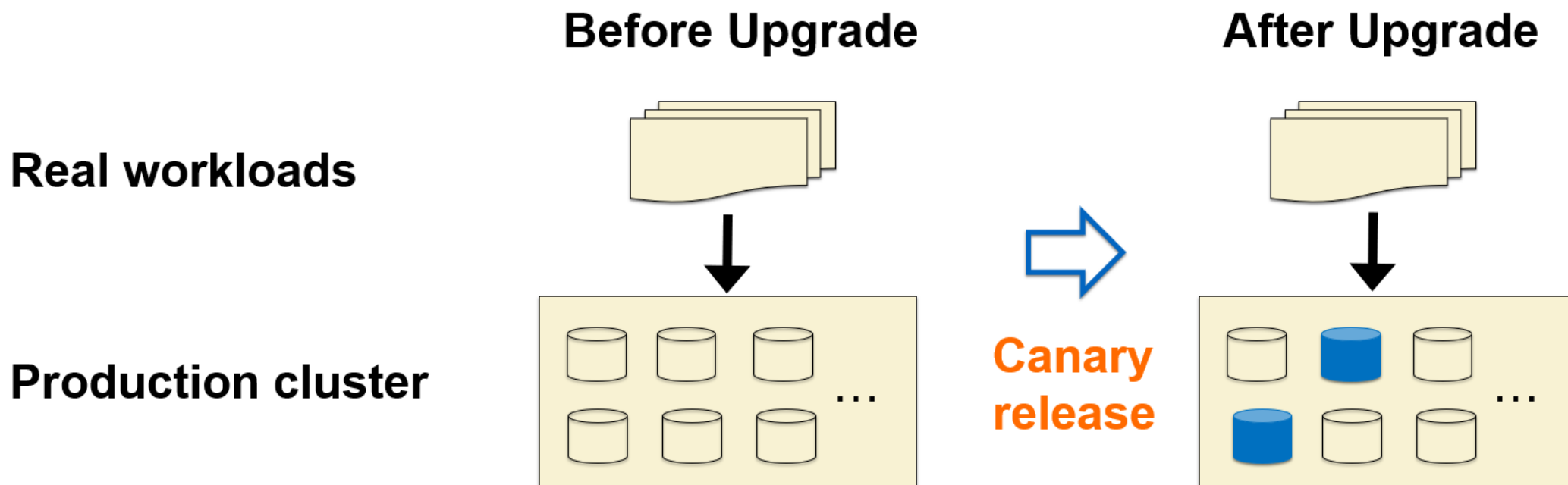
[郭健美. **SPEED: 大规模数据中心的性能分析平台**. 2019年杭州云栖大会阿里云系统软件开发专场特邀报告]

# Case 1 CPU Upgrade for Big Data Processing Platform

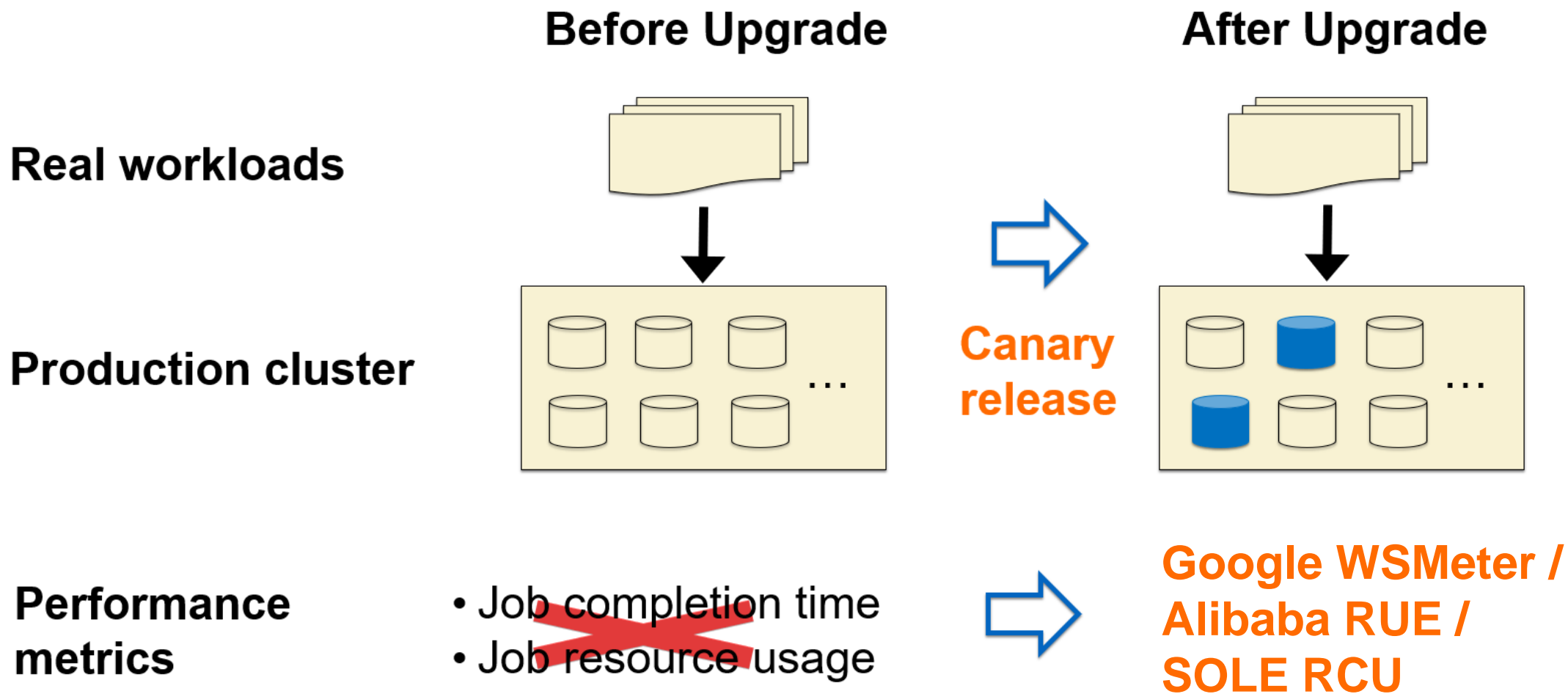




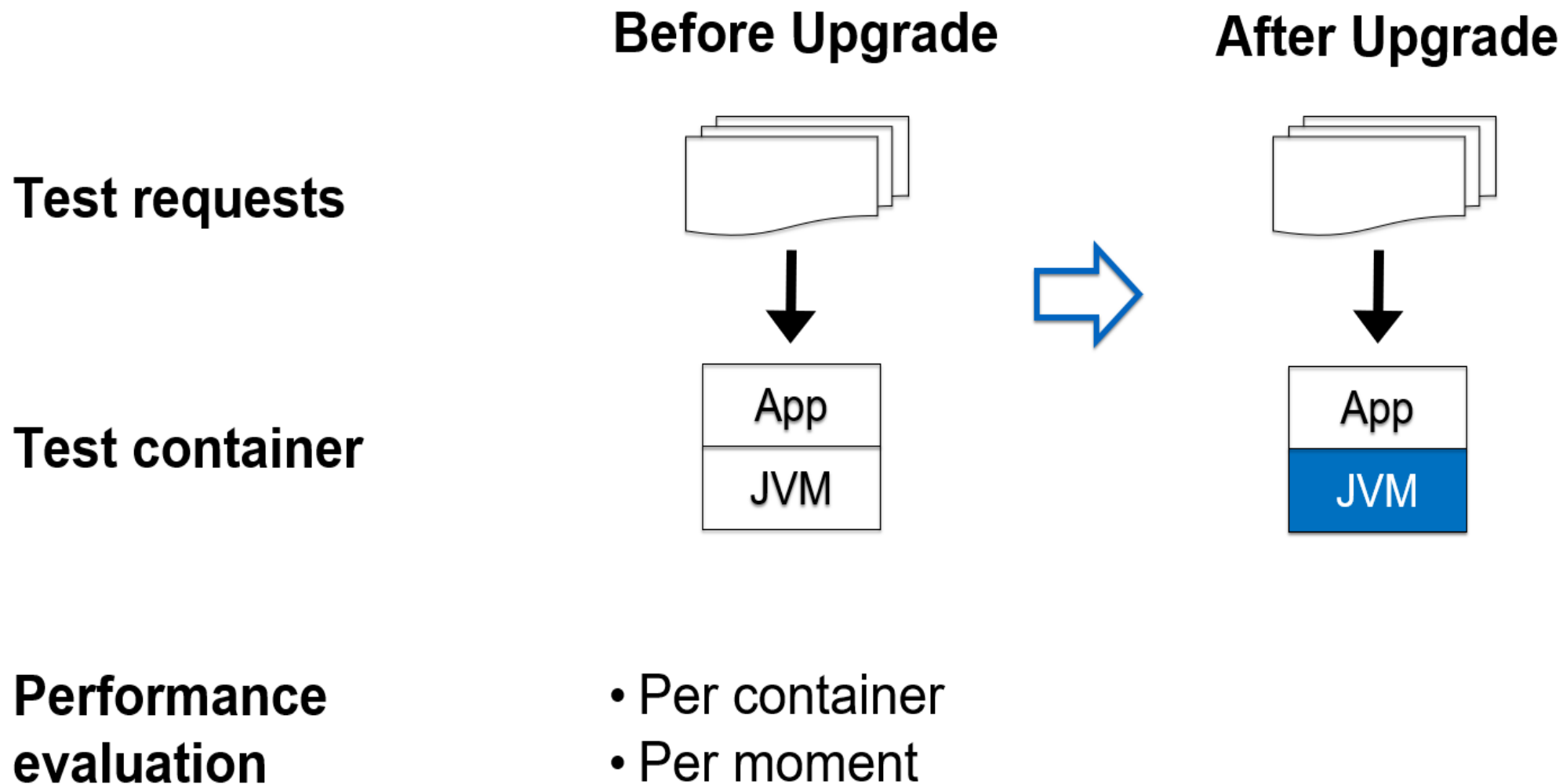
# Case 1 CPU Upgrade for Big Data Processing Platform



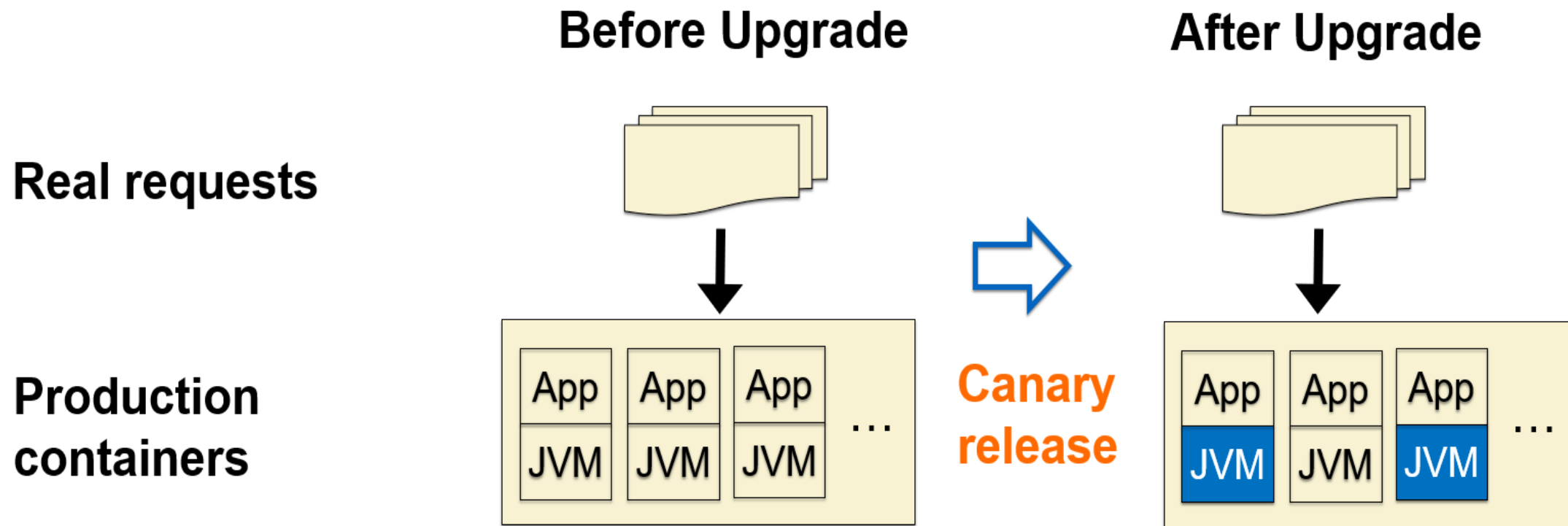
# Case 1 CPU Upgrade for Big Data Processing Platform



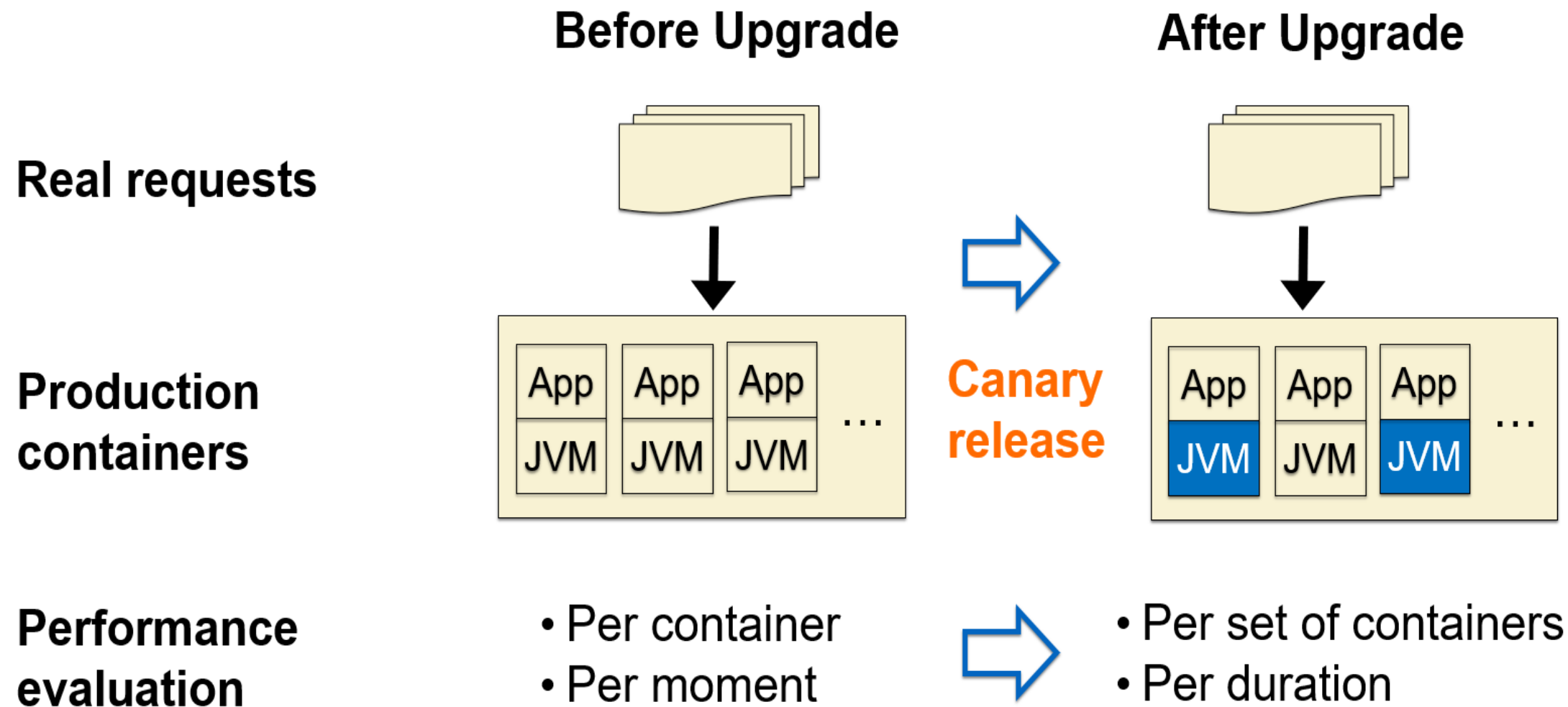
## Case 2 JVM Upgrade for E-Commerce Applications







## Case 2 JVM Upgrade for E-Commerce Applications



## Case 2 JVM Upgrade for E-Commerce Applications



## Case 3 A Pitfall of Data Science: Simpson's Paradox

	Before Upgrade		After Upgrade		$\frac{RUE}{RUE'}$	Evaluation Results
	# Containers	RUE	# Containers	RUE'		
App Total	99.0%	776	1.0%	716	1.08	 8%
App Group1	50.1%	1289	0.3%	1484	0.87	 13%
App Group2	31.5%	428	0.4%	434	0.99	 1%
App Group3	17.4%	550	0.3%	655	0.84	 16%



# 小结

- 数据中心及其云服务平台已经成为国民经济的关键基础设施，随着云数据中心的规模越来越大，几十万甚至上百万台服务器的规模屡见不鲜，稳定性保障和资源利用率提升是两个主要的研究任务
- 混部技术提高了数据中心的资源利用率，但也带来了复杂而隐蔽的性能干扰，这些干扰源于复杂软硬件栈上的资源竞争和大规模分布式应用的调用依赖
- 在数据center里，可扩展的性能优化更具吸引力。为此，需要深入理解数据center里各种应用的行为和性能特征，并且从数据center的整体对性能数据进行聚合分析
- 性能评价也需要从线下小规模测试环境中的基准评测迁移到大规模生产环境中的线上评价，这对于数据center里的容量规划、软硬件升级等关键决策至关重要。合理的性能分析和评价有助于数据center的优化升级和成本节约，而错误的分析可能误导决策、甚至造成巨大的成本损耗