云计算数据中心当下问题与基础软件关键技术

刘峥

阿里达摩院操作系统实验室，wenqing.lz@alibaba-inc.com

2022/10/14

目录

趋势背景：摩尔定律失效，传统算力提升缓慢，资源成本上升

Figure 6. Growth of computer performance using integer programs (SPECintCPU).

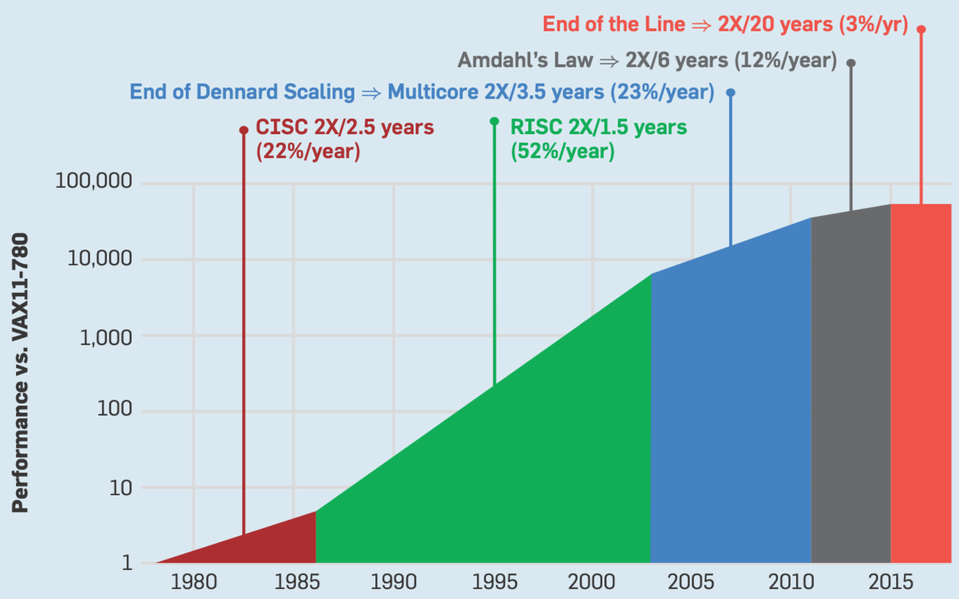
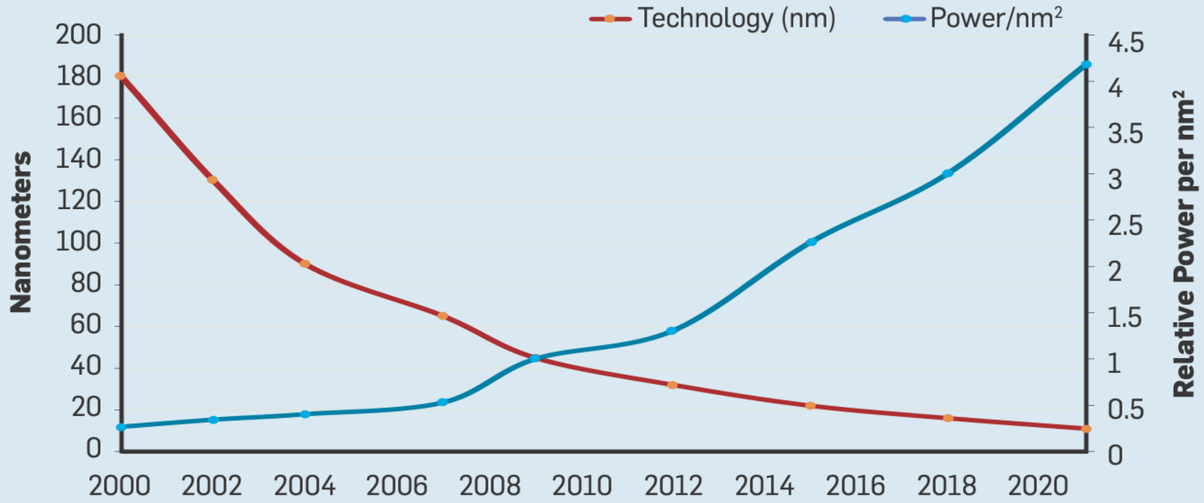
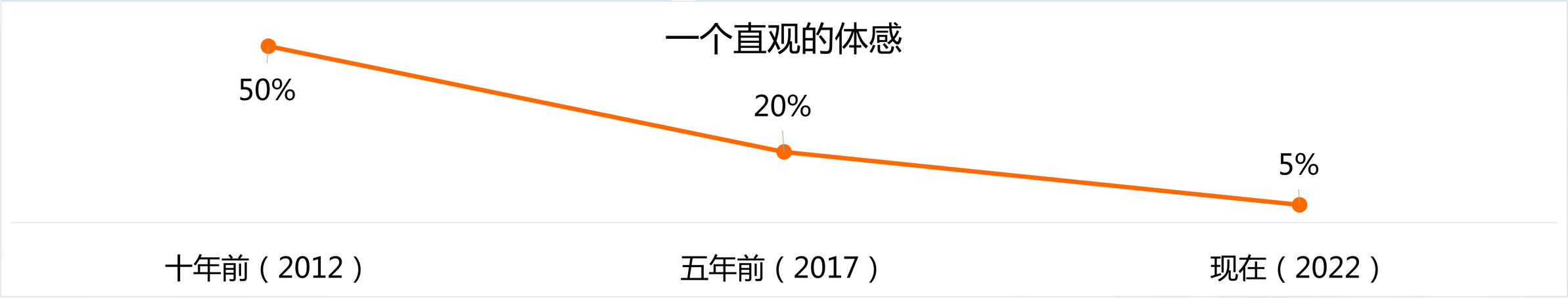
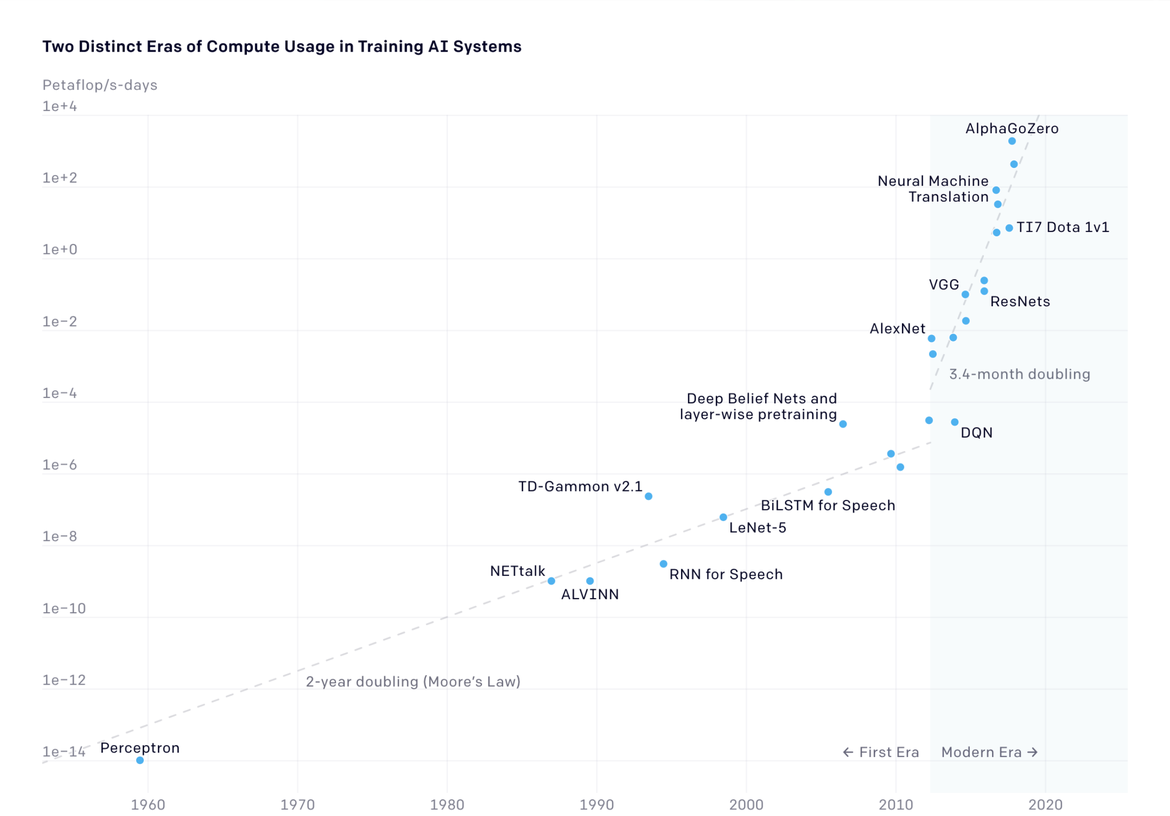


Figure 3. Transistors per chip and power per mm².





趋势背景：大数据/机器学习快速发展，对算力、存储、安全提出诸多诉求



2016
Tensorflow, OSDI 16

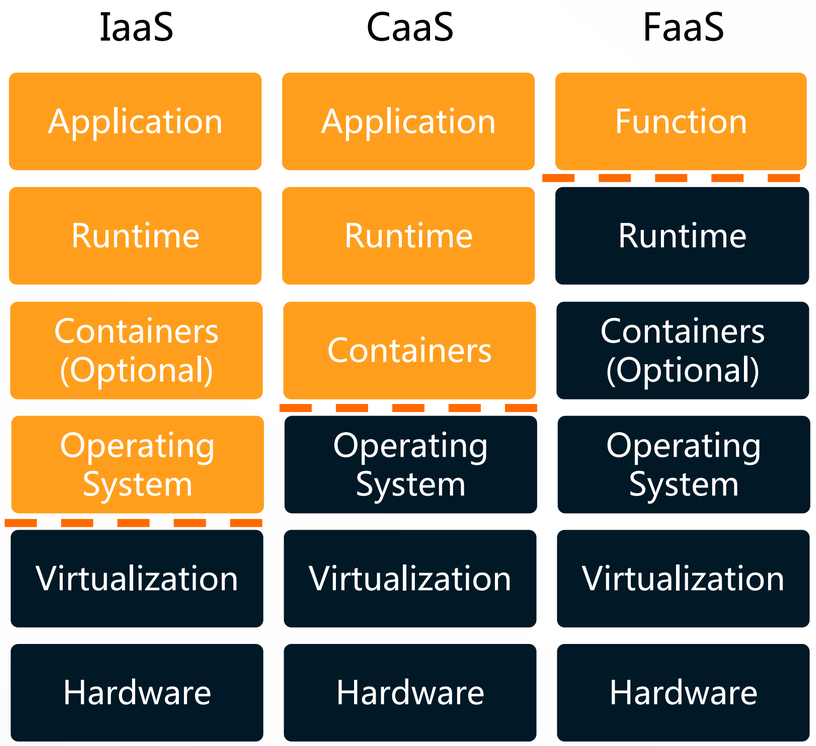
2017
TPU, ISCA 17

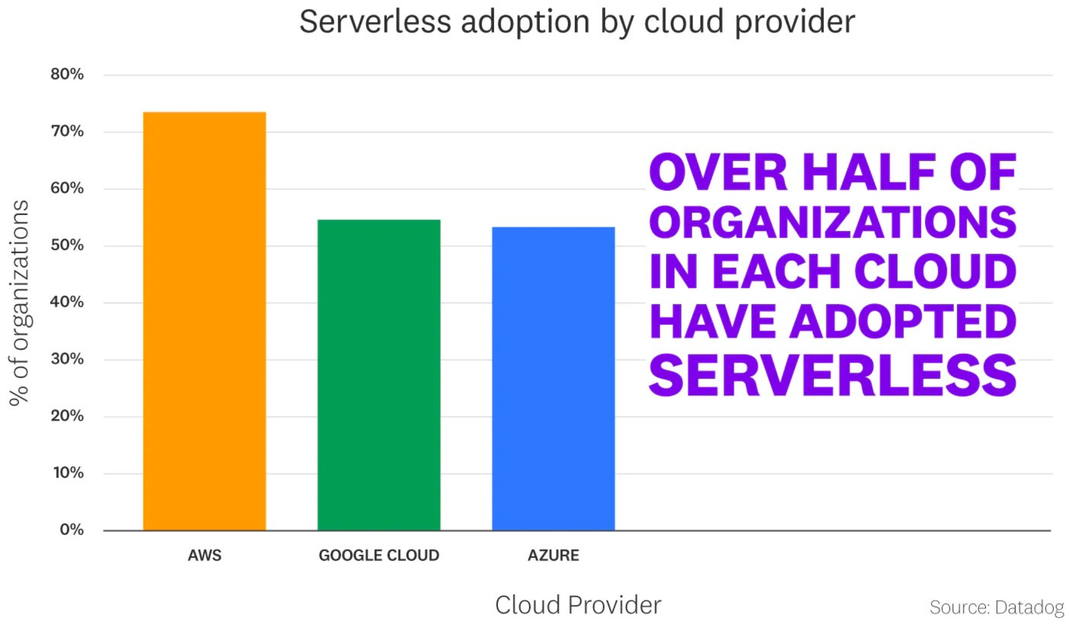
2018
TVM, OSDI 18

2019
AWS Inferentia

2021
AWS Trainium

趋势背景：云计算催生新的云原生计算形态

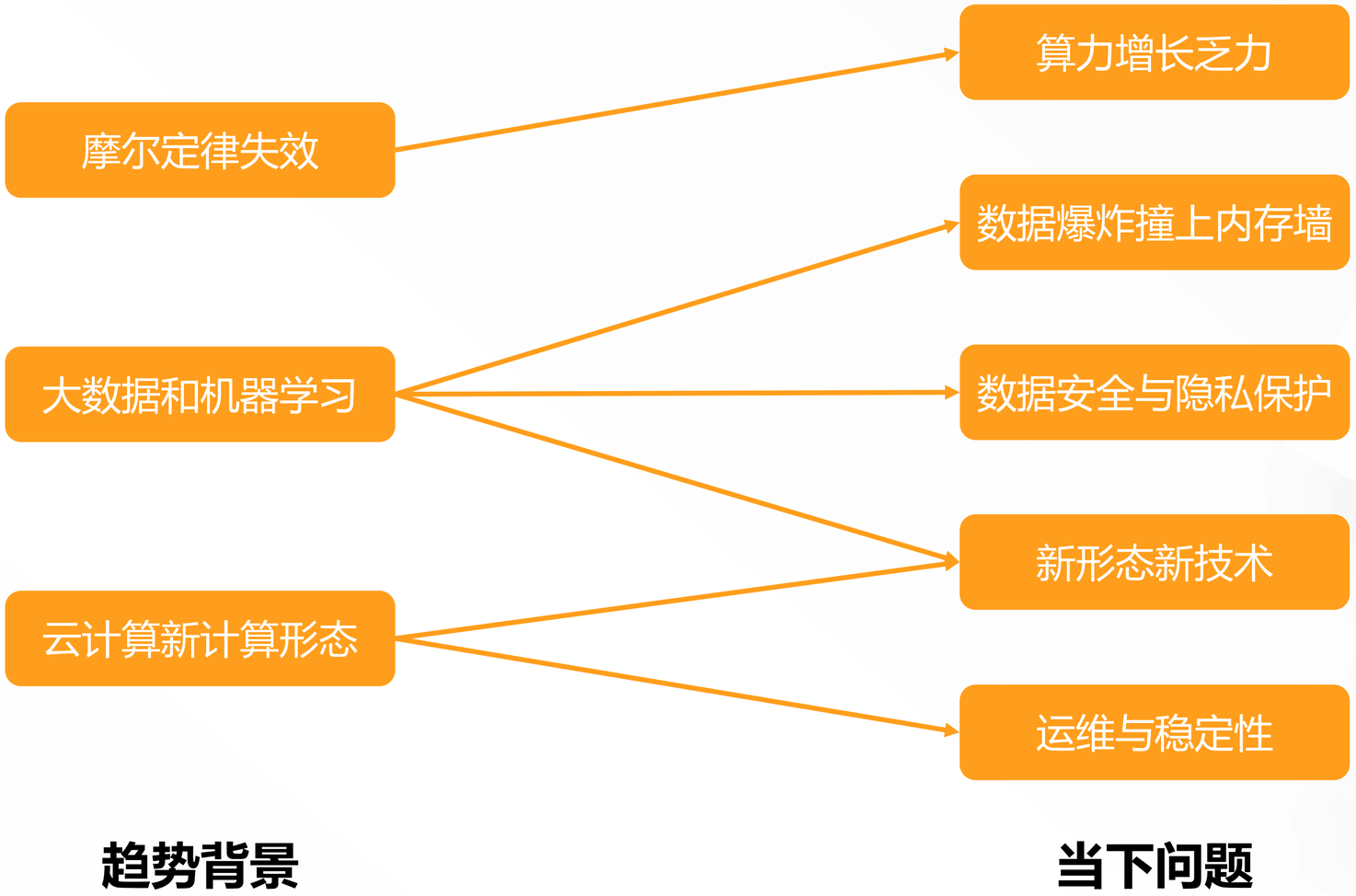




国外超过一半的组织已经开始采用 Serverless 计算形态

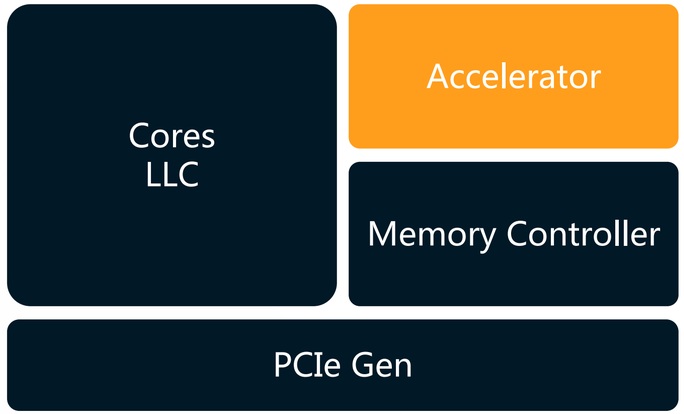
目录

当下问题与关键技术


摩尔定律失效
大数据和机器学习
云计算新计算形态
算力增长乏力
数据爆炸撞上内存墙
数据安全与隐私保护
新形态新技术
运维与稳定性
趋势背景
当下问题

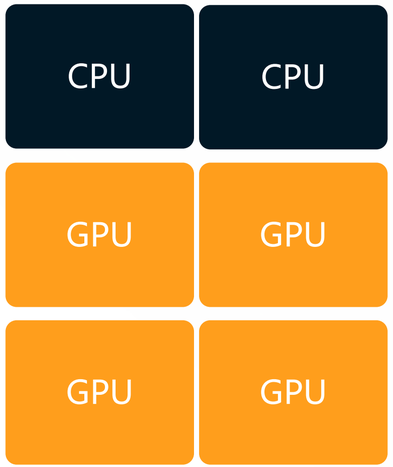
算力增长乏力：当下问题

CPU 内置加速器



- 如何充分利用内置加速器

各种 XPU



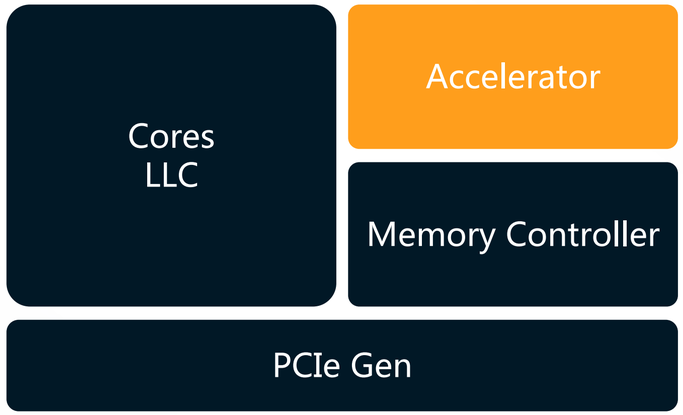
- 代码在“正确”的 XPU 上运行

不同体系结构

- 跨体系结构迁移应用

算力增长乏力：关键技术（异构算力支持）

CPU 内置加速器



操作系统内核内存操作

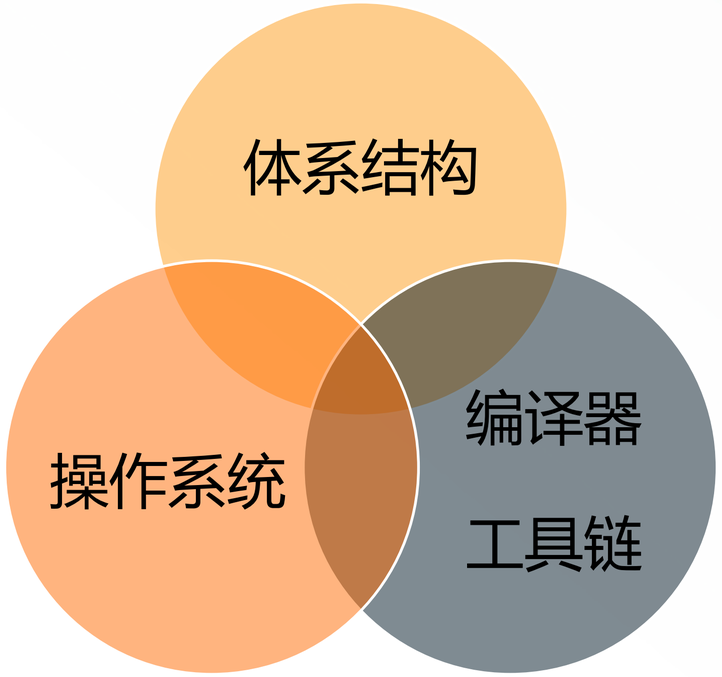
- Page Copy
- Compaction

不同体系结构

重新编译 & 二进制翻译

- Memory Models

算力增长乏力：未来机会



算力提升

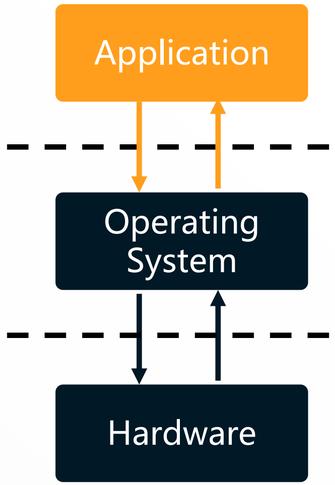
软硬件协同设计

缺少同时了解软硬件的人才

数据爆炸撞上内存墙：当下问题（数据频繁移动）

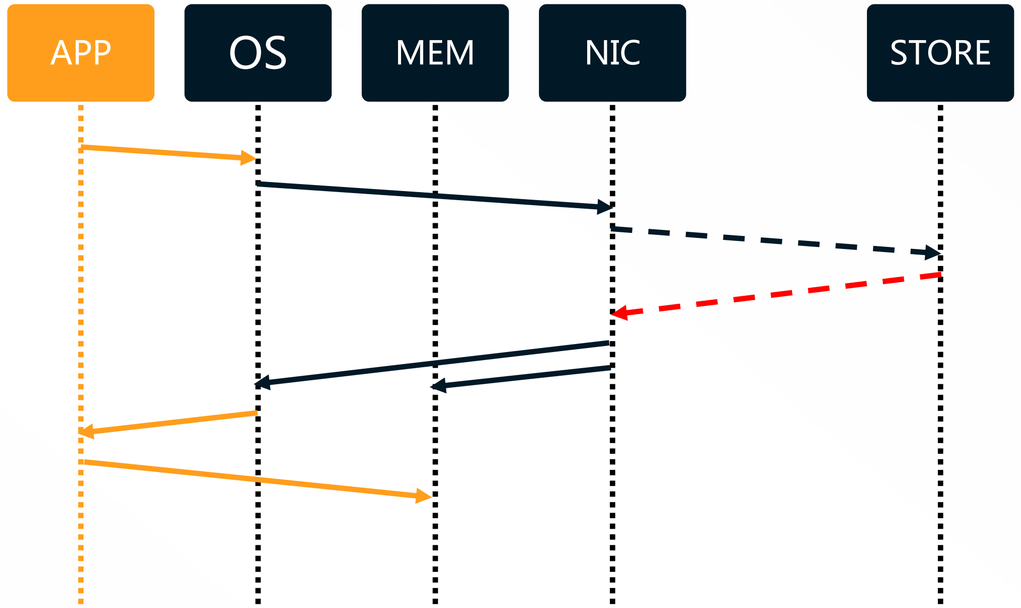
用户视角

程序处理数据



实际流程

程序处理数据



当下问题

表象：需要频繁移动数据

本质：CPU-Centric 的软件架构

数据爆炸撞上内存墙：关键技术

关键技术1：更优秀的分层存储

- 操作系统实现分层内存，更好识别冷热数据
- 针对大数据、AI 场景提供弹性数据抽象和加速服务（*Fluid*）

关键技术2：改变已有软硬件架构

- eBPF 卸载执行（存储 [ZCSD]、网络 [hXDP]）
- Zero-CPU DPU [Hyperion]
- 如何更好的移动计算？(Ray, Spark, PyTorch, ...)

数据安全与隐私保护日趋严格：当下问题（提供更易用的隐私保护技术）

Sandbox

过去

云厂商要确保自身安全

避免用户间相互攻击

e.g. Noisy Neighbor

TEE

(Trusted Execution Environment)

未来

确保自身安全的同时

让用户保护自身隐私数据

e.g. ML Models

Intel SGX

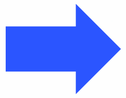
AMD SEV

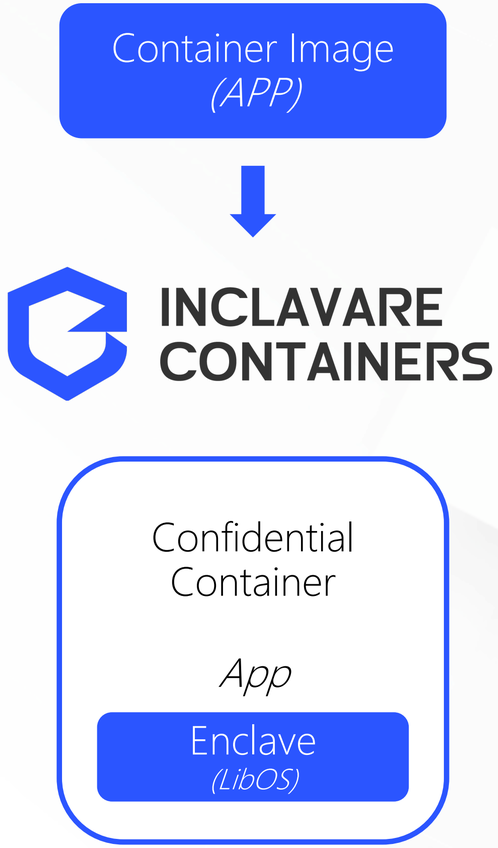
ARM TrustZone

<u>需要用户修改应用来配合</u>

数据安全与隐私保护日趋严格：关键技术

- **申请密钥**
  - 申请 SGX 相关的密钥
- **配置环境**
  - 安装 SGX 驱动和 SDK
- **应用适配**
  - 修改应用添加 ecall/ocall 调用
- **编译构建**
  - 编译应用，打包镜像
- **部署运行**
  - docker run





更多信息
https://inclavare-containers.io

数据安全与隐私保护日趋严格：未来挑战

挑战1：新的硬件技术支持

- TDX/SEV/CCA
- 通过 VMM 来支持 TEE，避免用户修改应用

更多信息

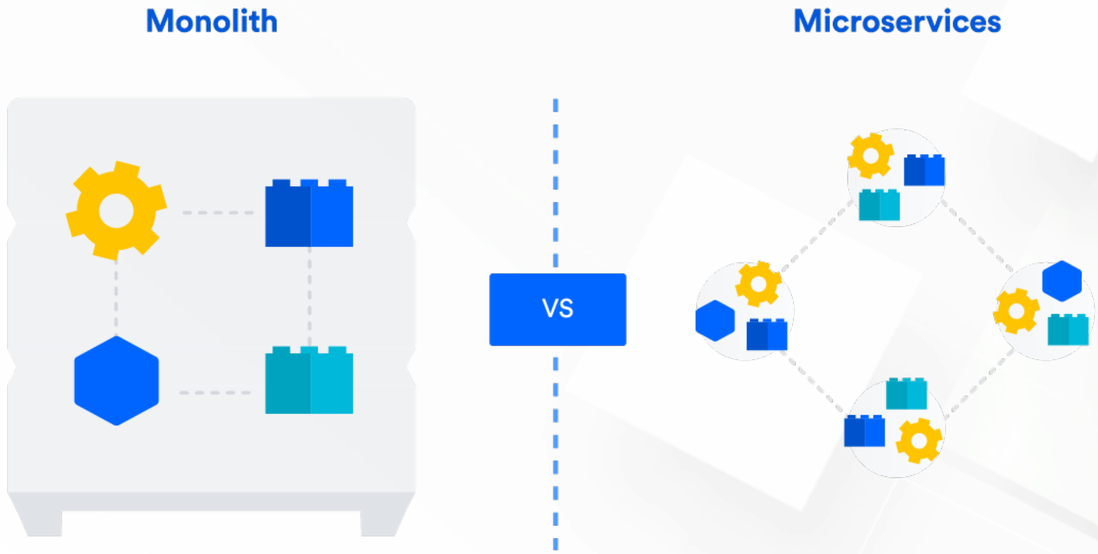
https://github.com/confidential-containers

挑战2：远程证明基础设施

- 远程证明技术仅提供了认证和授权TEE能否访问敏感数据的技术手段，需要引入其他技术（如软件供应链安全）解决可自证性的问题
- 远程证明体系需要第三方服务基础设施，降低用户自己部署此类服务开销
- 远程证明会从CPU扩展到外设，技术架构需要进行扩展：点（单一TEE的远程证明能力）->线（跨TEE的远程证明能力）->面（提供完整的远程证明体系）->层（提供多层级的远程证明体系）

新的计算形态要求新的底层技术：当下问题

IaaS	CaaS	FaaS
Application	Application	Function
Runtime	Runtime	Runtime
Containers (Optional)	Containers	Containers (Optional)
Operating System	Operating System	Operating System
Virtualization	Virtualization	Virtualization
Hardware	Hardware	Hardware



从资源角度看

- 资源颗粒度逐渐变小，资源弹性要求更高

从应用架构看

- 应用颗粒度更小，单机部署密度要求更高

新的计算形态要求新的底层技术：关键技术（资源弹性扩展能力）

大部分是小规格实例

- *47%的函数实例 < 128MB*

实际内存使用量不高

- *90% Azure实例内存用量 < 400MB*

短时间创建量大

- *1s并发创建请求 > 200*

单机同时运行的实例多

- *256GB→max 256*1024/128=2K*

基本要求

低响应延迟

两个需求

高并发创建

高部署密度

RunD

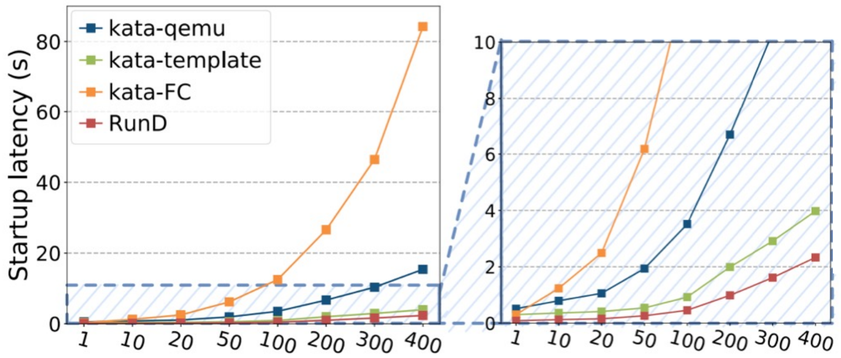
OS/Runtime 协同优化

- *控制平面优化：containerd->rund*
- *安全容器rootfs优化*
- *全栈模版启动*
- *优化cgroup路径*

新的计算形态要求新的底层技术：关键技术（资源弹性扩展能力）

Avg 88ms

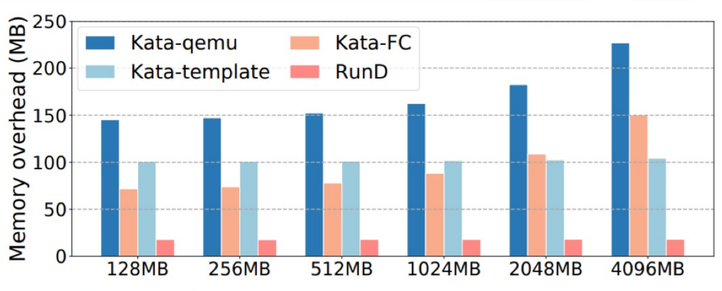
Reduced cold startup latency for a single sandbox



(a) End-to-end startup latency with different concurrency

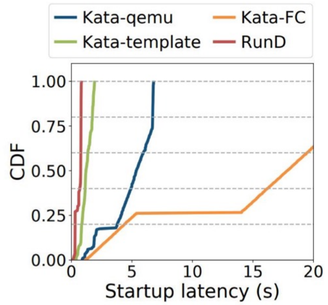
20MB–

The memory overhead is less than 20MB per sandbox with RunD.

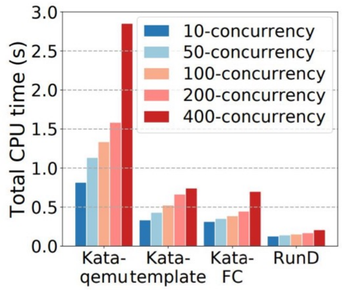


Max 200/s

launch 200 sandboxes simultaneously within 1s, with minor fluctuation and CPU overhead.



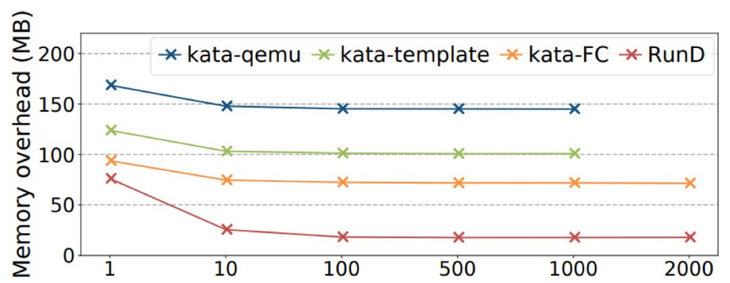
(b) Latency distribution



(c) CPU time

2500density

deploy over 2,500 sandboxes of 128MB memory specification on the node with 384GB memory



更多细节

- RunD: A Lightweight Secure Container Runtime for High-density Deployment and High-concurrency Startup in Serverless Computing [ATC 22]
- Help Rather Than Recycle: Alleviating Cold Startup in Serverless Computing Through Inter-Function Container Sharing [ATC 22]

新的计算形态要求新的底层技术：未来机会

挑战1：持续改进隔离技术

- 运行环境 e.g. Unikernel ?
- 当前 Serverless 形态对 Unikernel 很不友好

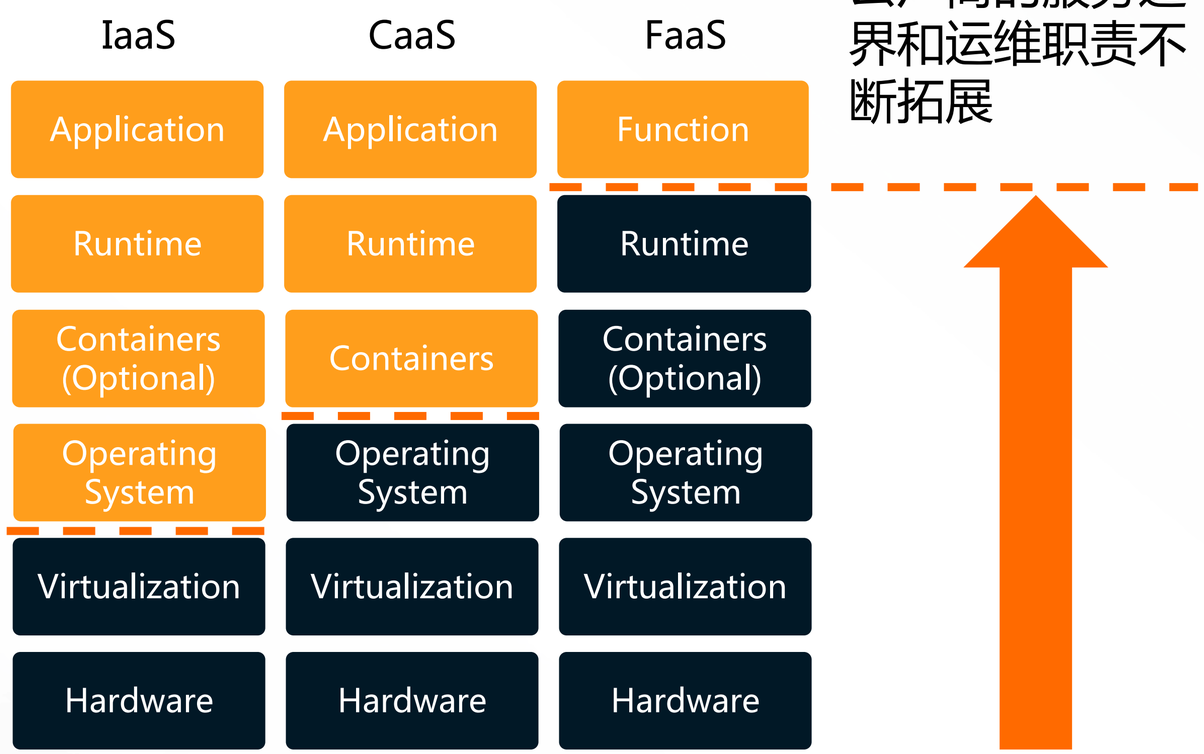
挑战2：未来的隔离技术

- WASM ?
- 非浏览器场景进展缓慢

挑战3：如何改进 Runtime

- 软硬件协同 e.g. RISC-V J extension
- 整体规范建设中

运维与稳定性：当下问题



云厂商的服务边界和运维职责不断拓展

可观测性能力建设

智能运维、故障定位

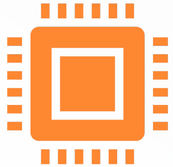
基础能力提升

运维与稳定性：关键技术（基础能力）

内存故障纠错

- 内存在数据中心硬件故障占比高
- CU/UE 有很大相关性

- 故障预测



芯片 RAS 能力支持

- 针对 ARM 芯片的故障注入

热升级能力

- 操作系统热补丁
- 虚拟化“三热”
- 调度器热升级

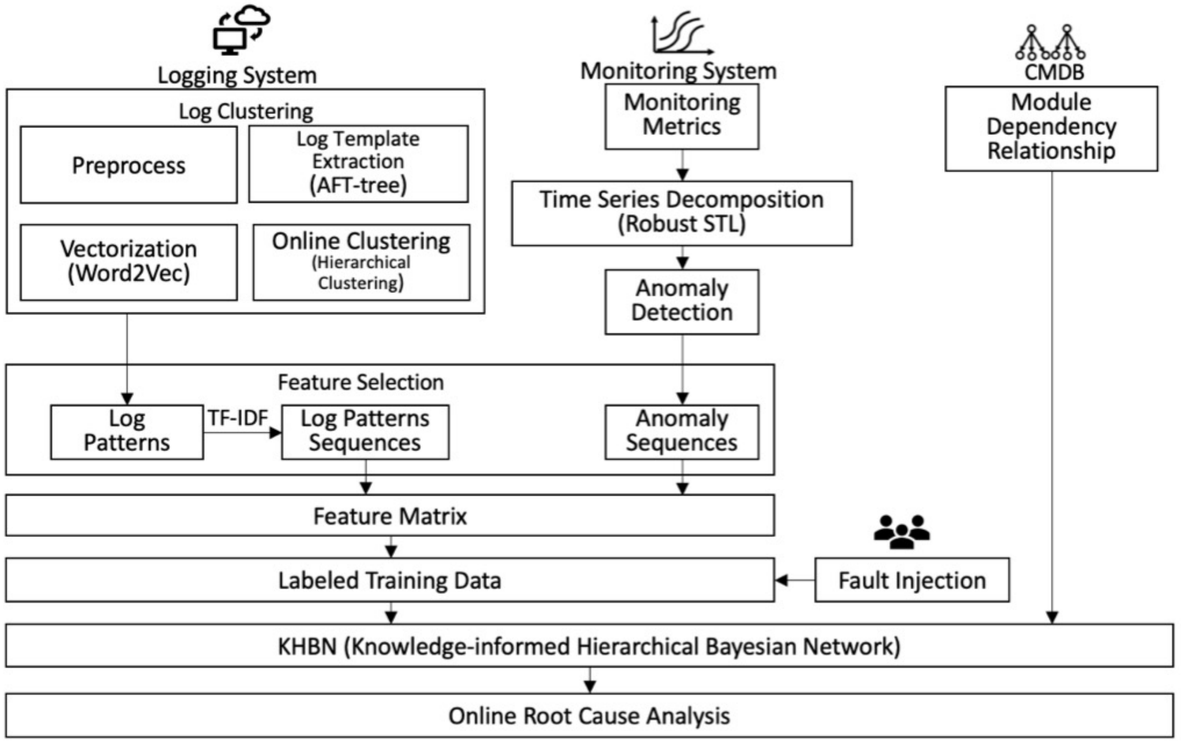
更多信息

- RAS 故障注入：https://gitee.com/anolis/ras-tools
- 调度器热升级：https://gitee.com/anolis/plugsched

运维与稳定性：关键技术（AIOps）

CloudRCA: A Root Cause Analysis Framework for Cloud Computing Platforms

Yingying Zhang, Zhengxiong Guan, Huajie Qian, Leili Xu, Hengbo Liu, Qingsong Wen, Liang Sun, Junwei Jiang, Lunting Fan, Min Ke
Alibaba Group
Hangzhou, China



Industry Practice of Configuration Auto-tuning for Cloud Applications and Services

Runzhe Wang*
Alibaba Group
China

Qinglong Wang*
Alibaba Group
China

Yuxi Hu
Alibaba Group
China

Heyuan Shi†
Central South University
China

Yuheng Shen
Tsinghua University
China

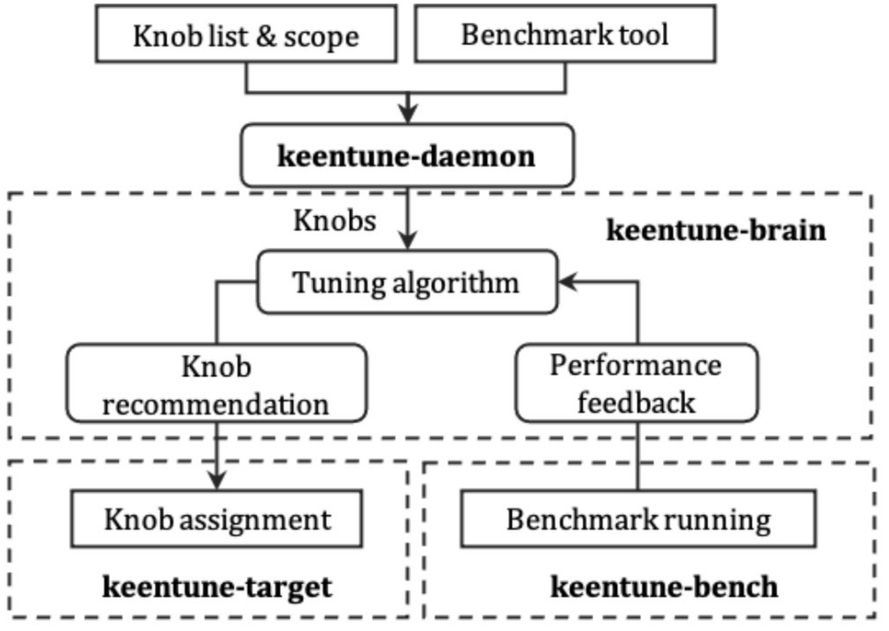
Yu Zhan
Central South University
China

Ying Fu
Ant Group
China

Zheng Liu
Alibaba Group&Zhejiang University
China

Xiaohai Shi
Alibaba Group
China

Yu Jiang
Tsinghua University
Chinal



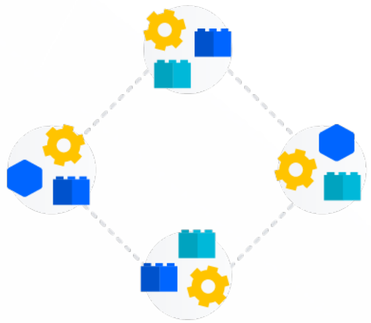
更多信息
http://keentune.io

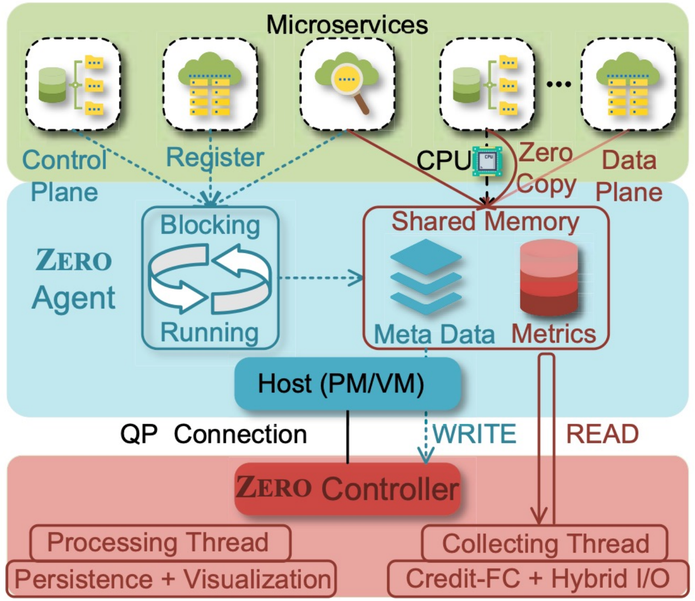
运维与稳定性：关键技术（监控与可观测性）

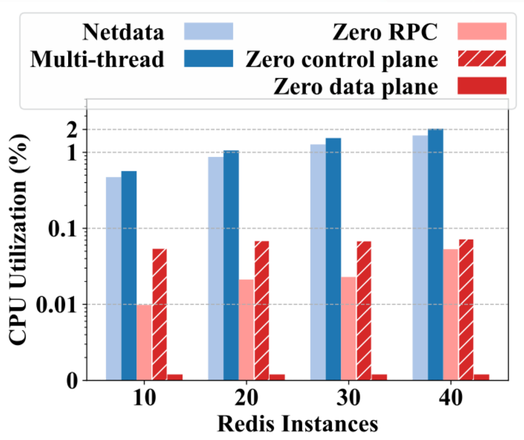
- 高部署密度
- 生命周期短
- 应用类型丰富

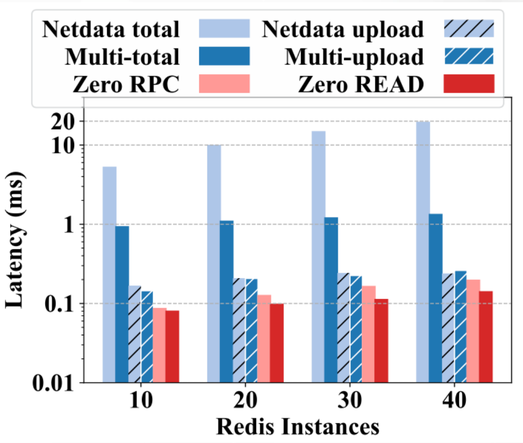
- 资源更紧张
- 观测指标数量多

- 实现0开销监控









核心方案

- *通过 RDMA one-side 操作实现 CPU& Kernel Bypass*
- *数据收集端进行数据处理*

更多细节

- Zero Overhead Monitoring for Cloud-native Infrastructure using RDMA [ATC 22]

当下问题与关键技术：总结

云 & 科技

Cloud & Technology