



业界首款
面向任意服务器和各种云而打造的自适应加速器

赛灵思扩展 Alveo 产品组合

发布 Alveo U50

2019 年 8 月 6 日星期二，太平洋时间上午 6 点/东部时间上午 9 点，北京时间晚上 9 点

➤ 不断扩展壮大的 Alveo 平台

- 赛灵思宣布推出 Alveo U50 – 业界首款专为云和边缘数据中心而打造的自适应计算、网络、存储加速器
- 在一系列关键数据中心应用中，吞吐量、时延和功耗效率性能得到显著改善
- 广泛且不断壮大的 Alveo 软件合作伙伴生态系统，以及不断增强的开发工具，加速 Alveo 解决方案持续扩展

ALVEO™



快速

专为高吞吐量、超低时延而设计
加速计算、网络和存储



灵活应变

部署优化的域专用架构
适应不断变化的算法



方便易用

云端或本地皆可部署
丰富的加速应用集



➤ Alveo U50: 业界首款专为云和边缘数据中心而打造的自适应加速器

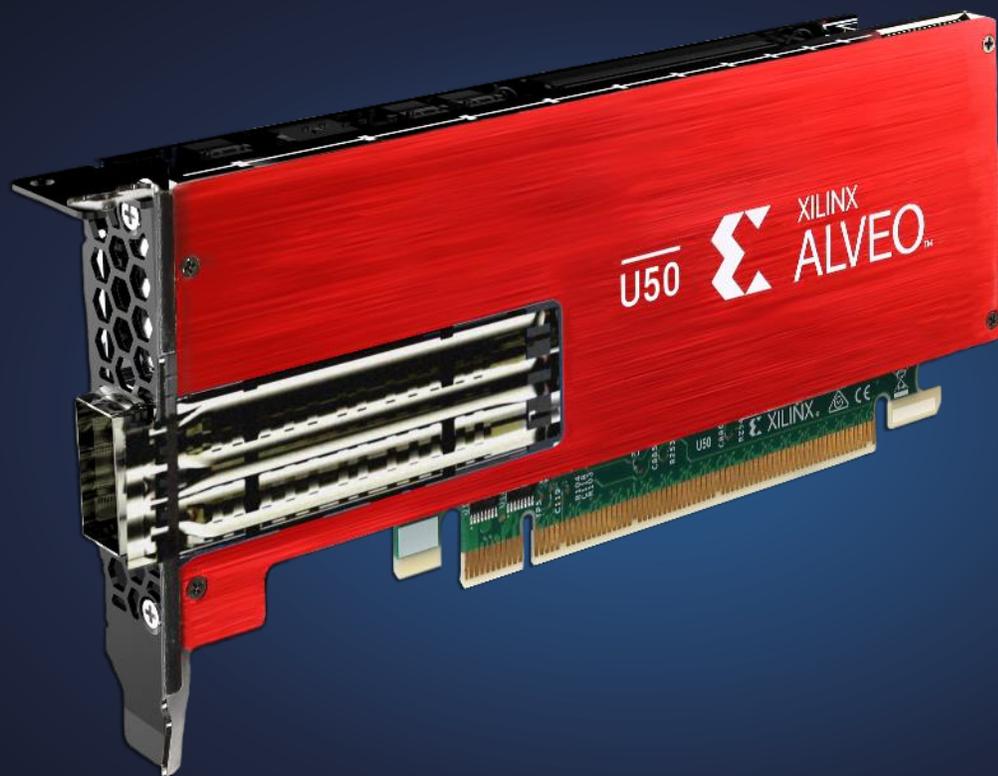
关键数据中心应用的性能与功耗效率提高 10 倍以上

最先进的自适应平台，用于加速计算、存储和网络

面向任何服务器、任何云而打造 (Built for any server, any cloud)



赛灵思 Alveo U50 的主要规格



UltraScale+ 架构

外形小巧

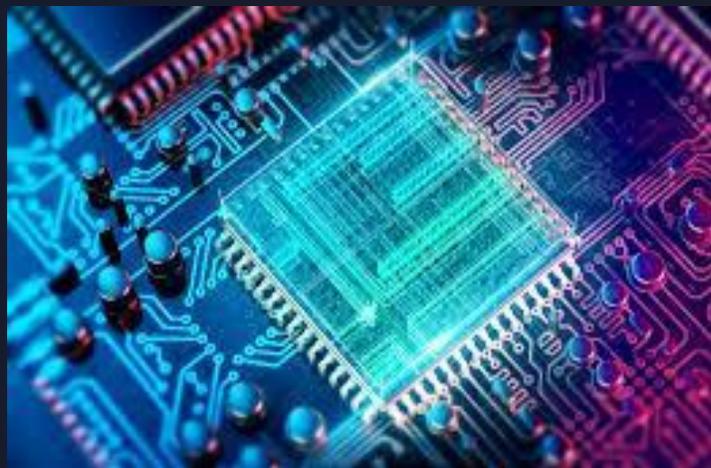
8GB HBM2 存储器, 460GB/s

PCIe Gen4、CCIX、PCIe Gen3

QSFP 28 (100GbE)

< 75W

加速数据中心



计算



网络



存储

领域专用架构

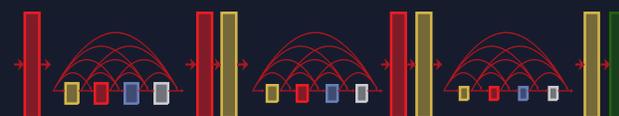
AlexNet



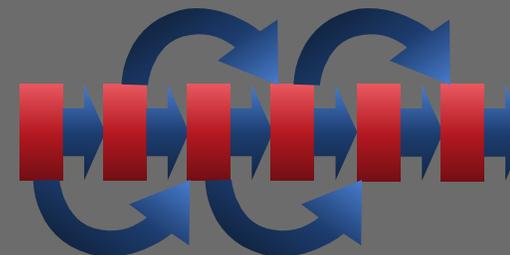
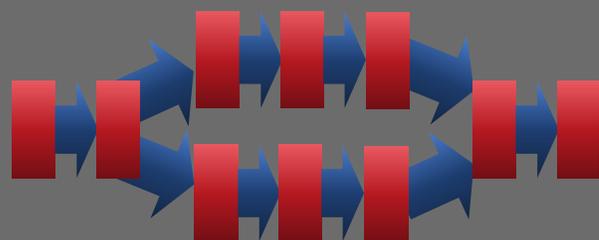
GoogLeNet



DenseNet



最高的吞吐量、低时延和高效能需要不同的硬件架构



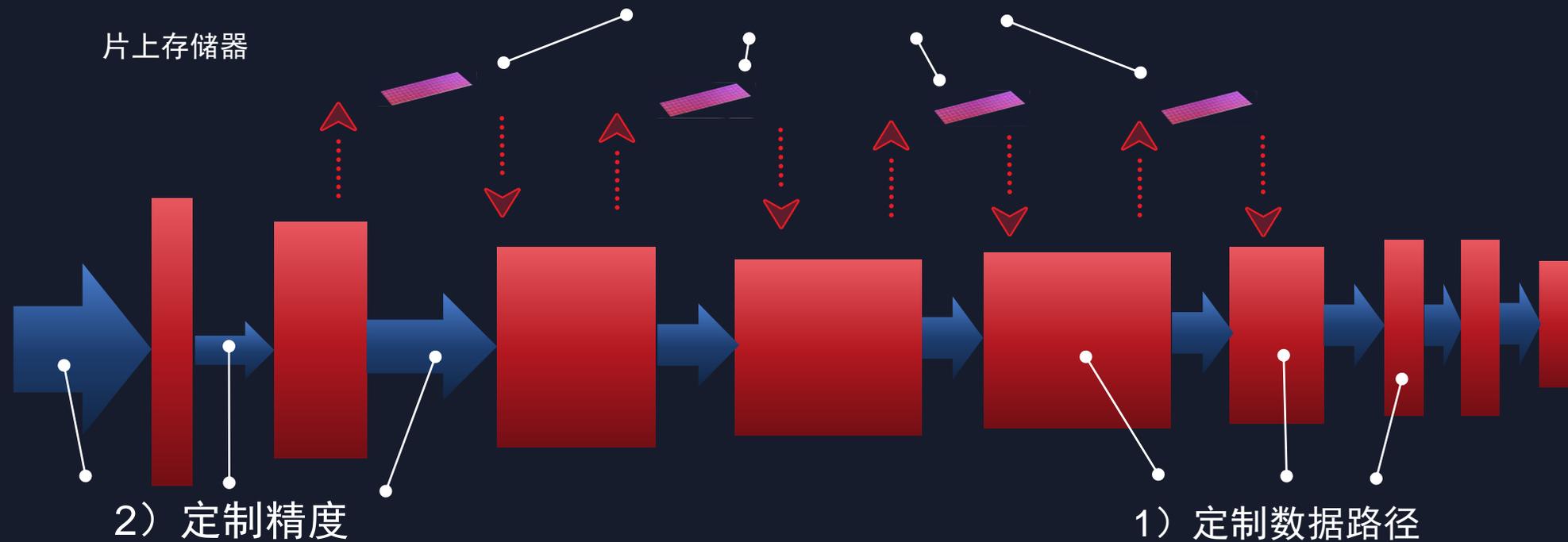
优化的性能

需要定制存储器与数据路径

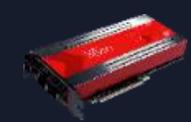
片外 DDR



3) 定制存储器层级



ALVEO 解决方案堆栈



扩大云的覆盖面



赛灵思为超大规模云数据中心提供支持

▶ 不断扩大的生态系统

数据分析



生命科学与 HPC



视频处理



机器学习



金融计算



图像处理



自 2018 年 10 月推出以来不断增长

数据分析

生命科学与 HPC

视频处理

2x

4x

机器学习 已发布的应用

金融计算 培训的开发者的

图像处理

大规模部署堆栈

ALVEO 容器化应用



KUBERNETES 编排



赛灵思 ALVEO



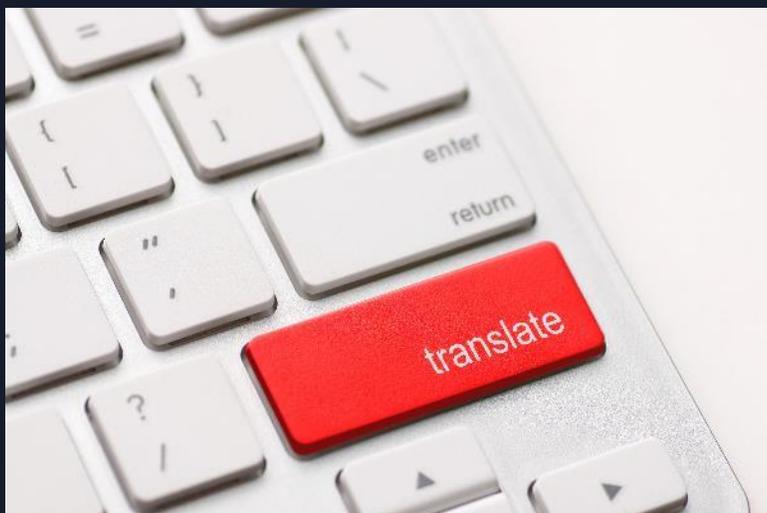
ALVEO 云应用实例

ALVEO 本地服务器

语音翻译

高吞吐量与低时延推断加速

CPU 与 GPU 无法实现的高吞吐量与低时延推断性能



估计数据: Alveo U50 (B=2, L=8), Tesla T4 (B=8, L=8)

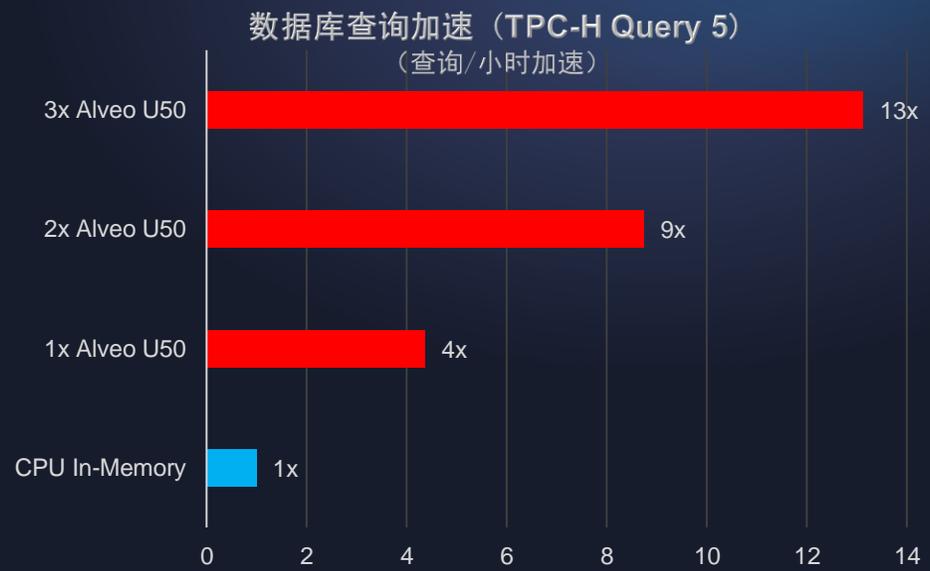
<https://medium.com/syncedreview/tsinghua-university-publishes-comprehensive-machine-translation-reading-list-c3f2df594218>

@ Copyright 2019 Xilinx

数据库分析

高吞吐量查询加速

更高的查询吞吐量与简化的基础设施



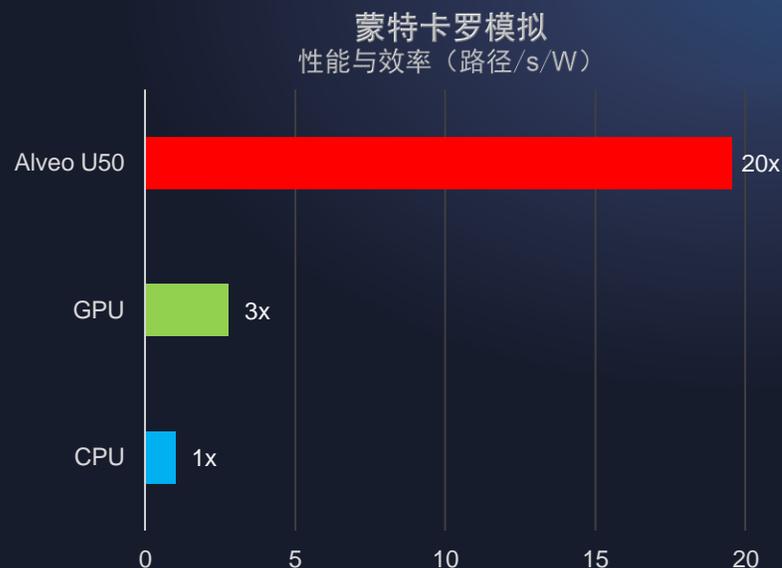
24 核 INTEL® XEON® PLATINUM 8260 处理器 (35.75M 高速缓存, 2.40GHz)

CPU 查询时间 = 210ms, 34,000 项查询/小时。Alveo U50 查询时间 = 24ms, 150,000 项查询/小时

金融市场建模

超高效衍生产品定价与风险建模

以小量成本实现更快、更具时间效益的洞察



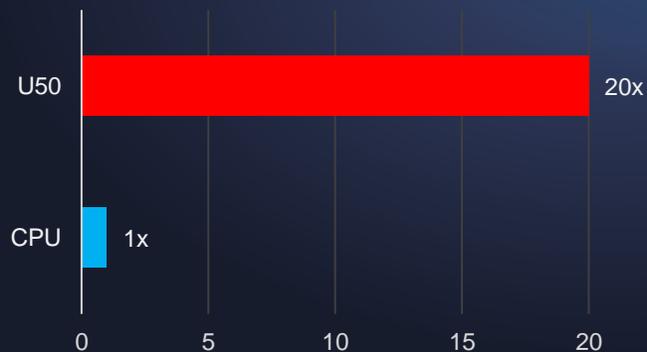
电子贸易业务

超低时延网络与计算加速

时延低于 500ns, 性能稳定



交易时间的加速
市场数据到 TCP 消息 (时延降低)



计算存储

线路速率数据压缩加速

压缩、解压缩、擦除编码、加密都在统一的平台上进行加速



Intel Skylake-SP 6152 @2.10GHz 22 核 CPU (Ubuntu 16.04), 每个CPU 核的 GB/s 压缩率=.0229。Alveo U50 = 10GB/s

Hadoop 加速

线路速率数据压缩加速

每节点吞吐量提高 20 倍

节点数减少 2 倍

总成本降低 40%

Alveo U50 加速



两个双 CPU 服务器
192TB，每节点压缩吞吐量 1GB/s

搭载两个 Alveo U50 的 Alveo 服务器
96TB（192TB 有效），每节点压缩
吞吐量 20GB/s

Intel Skylake-SP 6152 @2.10GHz CPU (Ubuntu 16.04)，每个 CPU 核的 GB/s 压缩率= .0229。Alveo U50 = 10GB/s，假设压缩比为 2: 1

今日要点总结

扩大市场机会

- 首款专为计算、网络和存储而打造的自适应加速器 – 面向任何服务器、任何云而打造

性能与 TCO 优势

- 将吞吐量、时延和功耗效率改善 10-20 倍
- 首款 PCIe Gen 4, 支持 HBM2 与 100Gbps 网络端口

简化编程与部署

- 不断发展的软件合作伙伴生态系统与加速解决方案
- 增强的开发工具与用于扩展的部署堆栈

附录

赛灵思 Alveo 产品阵容

ALVEO™ U50



UltraScale+ 架构

872k LUT

单槽、半高

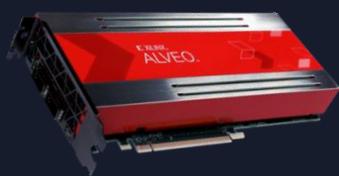
8GB HBM2, 460GB/s

PCIe Gen3、Gen4、CCIX

1x QSFP 28 (100GbE)

< 75W

ALVEO™ U200



UltraScale+ 架构

1,182k LUT

双槽、全高

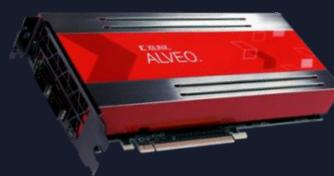
64GB DDR, 77GB/s

PCIe Gen3

2x QSFP 28 (100GbE)

< 225W

ALVEO™ U250



UltraScale+ 架构

1,728k LUT

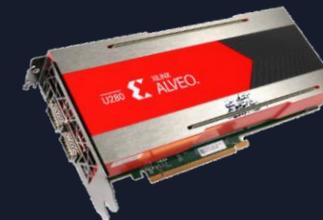
双槽、全高

64GB DDR, 77GB/s

PCIe Gen3

2x QSFP 28
(100GbE)
< 225W

ALVEO™ U280



UltraScale+ 架构

1,304k LUT

双槽、全高

8GB HBM2, 460GB/s

PCIe Gen3、Gen4、CCIX

2x QSFP 28 (100GbE)

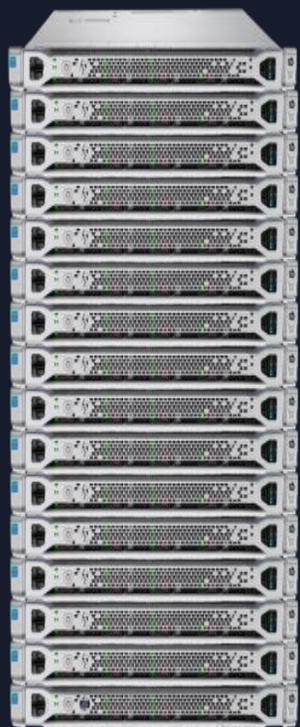
< 225W

一个平台. 最广泛的加速

	CPU (顺序)	GPU (并行)	Alveo (顺序 + 并行)
第三方应用	●	●	●
高级编码	●	●	●
复杂的存储器与数据路径			●
自适应硬件			●
AI 推断 + 处理前后			●
板载网络			●

实时视频转码

简化基础设施并降低成本



Alveo U50 HEVC 视频压缩



每节点吞吐量提高 20 倍

将硬件成本降低 8 倍

将功耗成本降低 8 倍



20 个双 CPU 服务器

40x Xeon Gold

超高质量的 H.265

20x 1080p30

单个 Alveo U50 服务器

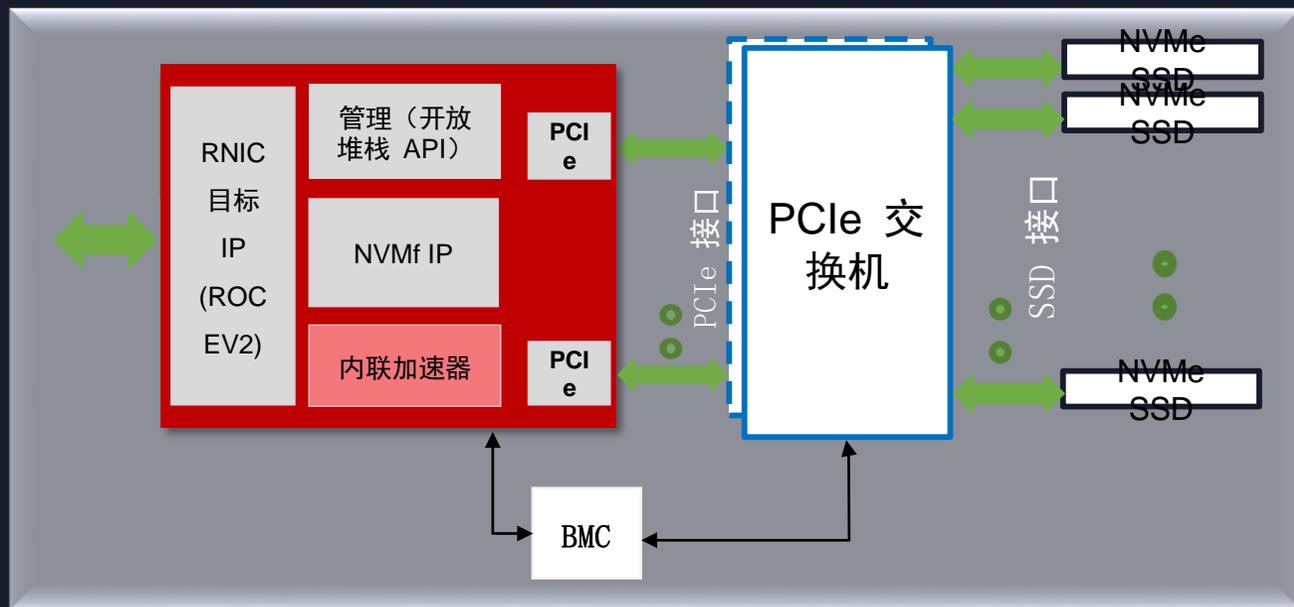
5 个 Alveo U50

极高质量的 NGCodec HEVC

20x 1080p30

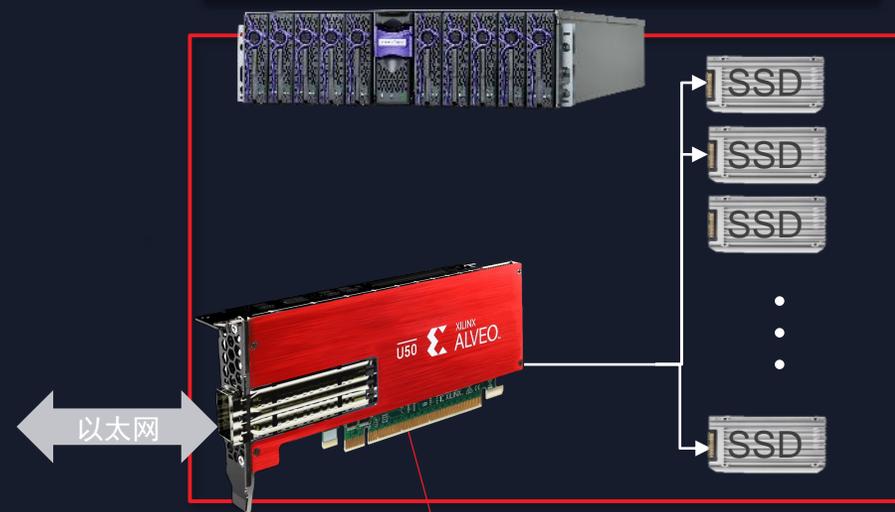
附加计算存储的架构

实现 NVMeoF 加速



- > 将架构连接到加速器前端的 SSD - 实现数据计算
- > 将 NVMeoF 目标卸载到 U50, 支持 250 万 IOPS
- > 可编程内联存储加速器的时延仅增加1微秒

技术预览：
附加架构
计算存储阵列



内联加速器示例：

存储服务：

- 压缩/解压缩
- 加密/解密
- 数据保护

数据库加速：

- 扫描
- 过滤
- 聚合



打造灵活应变，
万物智能的世界