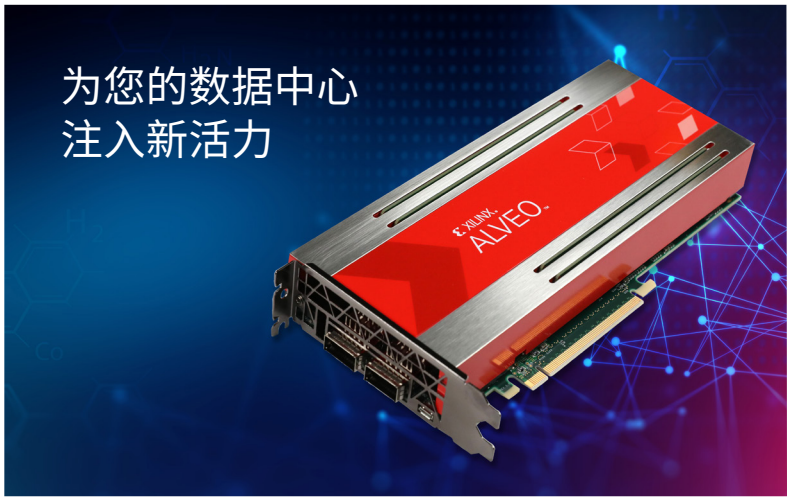
XILINX ALVEO™

U200 | U250

用于数据中心工作负载的自适应加速器卡



简介

赛灵思 Alveo™ 数据中心加速器卡专为满足现代数据中心变幻莫测的需求而设计。对于常规工作负载，与 CPU 相比性能提升高达 90 倍，这其中包括机器学习推断、视频转码和数据库搜索与分析。

由于复杂算法的发展速度快于半导体设计周期，因此功能固定的 GPU 和 ASIC 器件已经无法跟上发展步伐。基于赛灵思 16nm UltraScale™ 架构，Alveo U200 和 U250 加速器卡可提供能适应连续算法优化的可重配置加速，在降低总拥有成本的同时，可以支持任何类型的工作负载。

支持 Alveo 加速器卡的，是面向常见数据中心工作负载的不断壮大的赛灵思及合作伙伴应用生态系统。对于定制解决方案，赛灵思应用开发者工具（SDAccel™ 工具）和机器学习套件，为开发者提供了将差异化应用快速推向市场的开发工具。

主要特点

快速 – 最高性能

- 处理关键工作负载的性能最高可达 CPU[1] 的 90 倍，成本[2] 仅为 CPU 的 1/3
- 与基于 GPU 的解决方案[4]相比，具有推断吞吐量[3]高出 4 倍、时延低 3 倍的优势

灵活应变 – 加速任何工作负载

- 从机器学习推断到视频处理，再到任何工作负载，都可以使用这同一个加速器卡
- 随着工作负载算法的演进发展，使用可重配置硬件，适应速度快于固定功能加速器卡产品周期

可访问 – 云 ↔ 本地移动性

- 可在云端或本地互换部署解决方案，并可根据应用需求扩展
- 为常规工作负载提供应用，或使用应用开发者工具构建您自己的应用

可适应任何工作负载

- 数据库搜索与分析
- 金融计算
- 机器学习
- 存储压缩
- 视频处理/转码
- 基因组学

1：BlackLynx Elasticsearch (Alveo) 对比 EC2 c4.8xlarge
2：Alveo 相对于双插槽 Intel Xeon Platinum 服务器运行 DNN 推断所节省的资本支出和运营支出
3：使用 Alveo 加速器卡加速 DNN（白皮书）
4：测量 CNN + BLSTM 语音转文本 ML 推断,与 Nvidia P4 对比

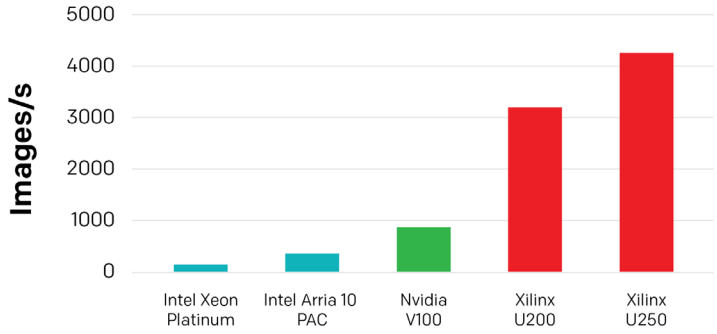
基准测试

自适应、加速任何工作负载

应用领域	合作伙伴工作负载	ALVEO 加速与 CPU 对比
数据库搜索与分析	BlackLynx 非结构化数据 Elasticsearch	90 倍
金融计算	Maxeler 风险价值 (VAR) 计算	89 倍
机器学习	赛灵思实时机器学习推断	20 倍
视频处理/转码	NGCodec HEVC 视频编码	12 倍
基因组学	Falcon Computing 基因组测序	10 倍

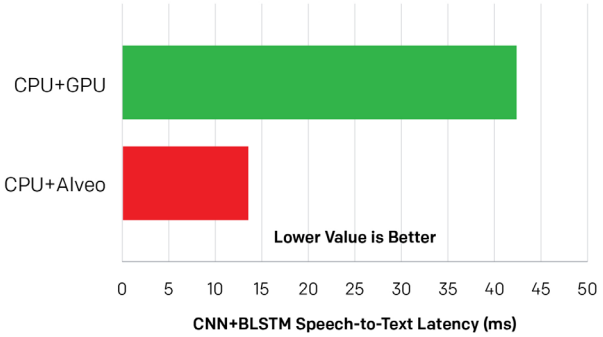
CPU 对比：Xeon c4.8xlarge AWS | Xeon E5-2643 v4 3.4GHz | Xeon Platinum c5.18xlarge AWS | 双插槽 E5-2680 v3 2.5GHz | Xeon f1.16xlarge

提高实时机器学习* 吞吐量 20 倍



*GoogLeNet V1：使用赛灵思 Alveo 加速器卡加速 DNN（白皮书）

降低 ML 推断时延 3 倍



CPU+GPU：Nvidia P4 + Xeon CPU E5-2690 v4 @2.60GHz（56 核）
CPU+Alveo：Alveo U200 或 U250 + Xeon CPU E5-2686 v4 @2.3GHz（8 核）

特性	ALVEO U200 加速器卡	ALVEO U250 加速器卡
峰值 INT8 TOP	18.6	33.3
DDR 存储器带宽	77 GB/s	77 GB/s
内部 SRAM 带宽	31 TB/s	38 TB/s
查找表 (LUT) 数量	892,000	1,341,000
散热选项	被动或主动	被动或主动

建议后续步骤

如需了解更多信息，请访问china.xilinx.com/alveo。立即体验，请访问 Nimbix 云服务，测试驱动数据中心工作负载，或购买 Alveo U200 或 U250 数据中心加速器卡用于本地部署。

企业总部
赛灵思公司
2100 Logic Drive
San Jose, CA 95124
USA
电话：408-559-7778
www.xilinx.com

赛灵思欧洲
One Logic Drive
Citywest Business Campus
Saggart, County Dublin
Ireland
电话：+353-1-464-0311
www.xilinx.com

日本
Xilinx K.K.
Art Village Osaki Central Tower 4F
1-2-2 Osaki, Shinagawa-ku
Tokyo 141-0032 Japan
电话：+81-3-6744-7777
japan.xilinx.com

Asia Pacific Pte. Ltd.
Xilinx, Asia Pacific
5 Changi Business Park
Singapore 486040
电话：+65-6407-3000
www.xilinx.com

印度
Meenakshi Tech Park
Block A, B, C, 8th & 13th floors,
Meenakshi Tech Park, Survey No. 39
Gachibowli(V), Seri Lingampally (M),
Hyderabad -500 084
电话：+91-40-6721-4747
www.xilinx.com

XILINX

在美国印刷。AC092018