# Myrtle Software Limited

16.12.2022 – 14:26 Uhr

## Myrtle.ai erzielt mit seinem Produkt VOLLO konkurrenzlose Latenzen in STAC-ML™-Benchmarks

*Cambridge, England (ots/PRNewswire) -*

Myrtle.ai, ein anerkannter Marktführer bei der Beschleunigung von Machine-Learning-Inferenzen, gab heute bekannt, dass sein Produkt VOLLO™ die bisher niedrigsten Latenzen für die Benchmarks STAC-ML™ Markets (Inference) erreicht hat. Diese Benchmarks wurden vom STAC Benchmark Council™ eingeführt, dem die weltweit größten globalen Banken, Broker, Börsen, Hedgefonds, proprietäre Handelsabteilungen und Vermögensverwalter sowie mehr als 50 führende Technologieanbieter angehören. Sie sind ein Instrument, um verschiedene Plattformen hinsichtlich Latenz, Durchsatz, Effizienz und Qualität bei der Inferenz durch maschinelles Lernen (ML) zu vergleichen.

VOLLO erreichte Latenzen von nur 24 Mikrosekunden bei einem Durchsatz von mehr als 170.000 Inferenzen pro Sekunde1. Diese unerreichte Leistung bei Latenz und Durchsatz ermöglicht es Benutzern, schneller als in der Vergangenheit intelligentere Entscheidungen mit komplexeren Modellen zu treffen, was ihnen einen Wettbewerbsvorteil in Bezug auf Trading, Risikoanalyse, Angebote und viele andere handelsbezogene Aktivitäten verschafft. VOLLO wurde für eine einfache und schnelle Installation auf Servern an einem gemeinsamen Standort entwickelt und ermöglicht außerdem eine Reduzierung der Rack-Fläche und des Energieverbrauchs, zwei kostbare Ressourcen an solchen Standorten.

VOLLO läuft auf einer Intel FPGA-basierten PCIe-Beschleunigerkarte mit Standardformfaktor, von der bis zu vier in einem 1U-Server installiert werden können. VOLLO kann mit vom Kunden trainierten Modellen konfiguriert werden, indem Modellarchitekturen aus dem LSTM-Modellzoo verwendet werden, was es Benutzern ermöglicht, eine Reihe von Workloads bereitzustellen, die auf ihre Anwendungsanforderungen zugeschnitten sind, und zwar über einen Exportprozess aus standardmäßigen ML-Tool-Flüssen. Es kann bis zu 12 parallele Modelle pro FPGA-Beschleunigerkarte unterstützen, die im System installiert sind, und ermöglicht maximal 48 parallele Modelle für ein System mit vier Karten.

Eine Schlüsseltechnologie ist das FPGA der Intel Agilex F-Serie. „Intel Agilex FPGA sind eine hervorragende Bereitstellungsplattform für unsere neue Fintech-Inferenzlösung", sagte Peter Baldwin, CEO von myrtle.ai. „Agilex FPGA haben es uns ermöglicht, bei diesen neuen Benchmarks für maschinelles Lernen von STAC schnell eine hervorragende Leistung zu erzielen. Durch die gehärtete bfloat16-Unterstützung und das ausreichende On-Chip-RAM sind sie ideal geeignet, um die wiederkehrenden Netzwerke in den STAC-Benchmarks zu beschleunigen, wo immer sie eingesetzt werden, während sich diese Workloads entwickeln und skalieren."

Als Titan-Partner in der Intel Partner Alliance wurde Myrtle.ai von Intel bei der Entwicklung von VOLLO unterstützt. „Diese Ergebnisse zeigen, was erreicht werden kann, wenn man die führenden FPGA von Intel mit der Kompetenz von Myrtle im Bereich KI/ML-Beschleuniger kombiniert", sagte Jim Dworkin, Vice President und General Manager der Cloud & Enterprise Acceleration Division in der Programmable Solutions Group von Intel. „Wir freuen uns, mit Myrtle zusammenzuarbeiten, um seine bemerkenswerte Technologie mit Intel Agilex FPGA zum Vorteil unserer Kunden zu kombinieren, und wir danken STAC für seine Führungsrolle bei der Festlegung relevanter Benchmarks für die Branche."

„STAC-Benchmarks werden von Finanzunternehmen basierend auf ihren Geschäftsanforderungen festgelegt, und sie haben diese Benchmark entwickelt, um die Inferenzleistung über mehrere Architekturen hinweg zu vergleichen", kommentierte STAC® President Peter Nabicht.„Myrtle.ai ist das erste Unternehmen, das geprüfte STAC-ML-Benchmark-Ergebnisse mit FPGA-Technologie erzielt und Finanzunternehmen, die nach der besten Lösung für ihren Bedarf suchen, wertvolle Datenpunkte bietet."

VOLLO ist ab heute erhältlich und es können Bewertungen arrangiert werden. Weitere Informationen finden Sie auf myrtle.ai/fintech, oder kontaktieren Sie Myrtle.ai per E-Mail an fintech@myrtle.ai.

Weitere Informationen über STAC und das STAC Benchmark Council finden Sie auf www.STACresearch.com. Die vollständigen Benchmark-Ergebnisse sind im STAC-Bericht (SUT ID MRTL221125) auf www.STACresearch.com/MRTL221125 verfügbar.

1 STAC-ML.Markets.Inf.S.LSTM_A.4.LAT.v1 and STAC-ML.Markets.Inf.S.LSTM_A. 4.TPUT.v1

**Informationen zu Myrtle.ai**

Myrtle.ai ist ein KI/ML-Softwareunternehmen, das erstklassige Inferenzbeschleuniger auf FPGA-basierten Plattformen von allen führenden FPGA-Anbietern liefert. Mit Fachwissen im Bereich neuronale Netzwerke über das gesamte Spektrum von ML-Netzwerken hinweg hat Myrtle Beschleuniger für FinTech, Sprachverarbeitung und Empfehlungen bereitgestellt.



Foto – https://mma.prnewswire.com/media/1969390/Myrtle_Software_Vollo.jpg

Logo – https://mma.prnewswire.com/media/1969391/Myrtle_Software_Limited_Logo.jpg

View original content:https://www.prnewswire.com/news-releases/myrtleai-erzielt-mit-seinem-produkt-vollo-konkurrenzlose-latenzen-in-stac-ml-benchmarks-301705223.html

Pressekontakt:

Giles Peckham+ 44 7785 278478,giles@myrtle.ai