IBM Newsroom

IBM Unveils On-Chip Accelerated Artificial Intelligence Processor

New chip design unlocks the ability to leverage deep learning inferencing on high-value transactions, designed to greatly improve the ability to intercept fraud, among other use cases.

ARMONK, N.Y., Aug. 23, 2021 /PRNewswire/ -- At the annual Hot Chips conference, IBM (NYSE: IBM) today unveiled details of the upcoming new IBM Telum Processor, designed to bring deep learning inference to enterprise workloads to help address fraud in real-time. Telum is IBM’s first processor that contains on-chip acceleration for AI inferencing while a transaction is taking place. Three years in development, the breakthrough of this new on-chip hardware acceleration is designed to help customers achieve business insights at scale across banking, finance, trading, insurance applications and customer interactions. A Telum-based system is planned for the first half of 2022.
Innovations across banking, finance, trading, insurance

Today, businesses typically apply detection techniques to catch fraud after it occurs, a process that can be time consuming and compute-intensive due to the limitations of today's technology, particularly when fraud analysis and detection is conducted far away from mission critical transactions and data. Due to latency requirements, complex fraud detection often cannot be completed in real-time - meaning a bad actor could have already successfully purchased goods with a stolen credit card before the retailer is aware fraud has taken place.

According to the Federal Trade Commission's 2020 Consumer Sentinel Network Databook, consumers reported losing more than $3.3 billion to fraud in 2020, up from $1.8 billion in 2019\(^2\). Telum can help clients to move their thinking from a *fraud detection* posture to a *fraud prevention* posture, evolving from catching many cases of fraud today, to a potentially new era of prevention of fraud at scale, without impacting service level agreements (SLAs), before the transaction is completed.

The new chip features an innovative centralized design, which allows clients to leverage the full power of the AI processor for AI-specific workloads, making it ideal for financial services workloads like fraud detection, loan processing, clearing and settlement of trades, anti-money laundering and risk analysis. With these new innovations, clients will be positioned to enhance existing rules-based fraud detection or use machine learning, accelerate credit approval processes, improve customer service and profitability, identify which trades or transactions may fail, and propose solutions to create a more efficient settlement process.
Telum and IBM's Full Stack Approach to Chip Design

Telum follows IBM's long heritage of innovative design and engineering that includes hardware and software co-creation and integration that spans silicon, system, firmware, operating systems and leading software frameworks.

The chip contains 8 processor cores with a deep super-scalar out-of-order instruction pipeline, running with more than 5GHz clock frequency, optimized for the demands of heterogenous enterprise class workloads. The completely redesigned cache and chip-interconnection infrastructure provides 32MB cache per core, and can scale to 32 Telum chips. The dual-chip module design contains 22 billion transistors and 19 miles of wire on 17 metal layers.

Leadership in semiconductors

Telum is the first IBM chip with technology created by the IBM Research AI Hardware Center. In addition, Samsung is IBM's technology development partner for the Telum processor, developed in 7nm EUV technology node.

Telum is another example of IBM's leadership in hardware technology. IBM Research, among the world's largest industrial research organizations, recently announced scaling to the 2 nm node, the latest benchmark in IBM's legacy of contributions to silicon and semiconductor innovation. In Albany, NY, home to the IBM AI Hardware Center and Albany Nanotech Complex, IBM Research has created a leading collaborative ecosystem with public-private industry players to fuel advances in semiconductor research, helping to address global manufacturing
demands and accelerate the growth of the chip industry.

For more information please visit www.ibm.com/it-infrastructure/z/capabilities/real-time-analytics.

Statements regarding IBM’s future direction and intent are subject to change or withdrawal without notice and represent goals and objectives only.

About IBM

IBM is a leading global hybrid cloud and AI, and business services provider, helping clients in more than 175 countries capitalize on insights from their data, streamline business processes, reduce costs and gain the competitive edge in their industries. Nearly 3,000 government and corporate entities in critical infrastructure areas such as financial services, telecommunications and healthcare rely on IBM's hybrid cloud platform and Red Hat OpenShift to affect their digital transformations quickly, efficiently, and securely. IBM's breakthrough innovations in AI, quantum computing, industry-specific cloud solutions and business services deliver open and flexible options to our clients. All of this is backed by IBM's legendary commitment to trust, transparency, responsibility, inclusivity, and service.

For more information, visit www.ibm.com

Media Contact:
Elizabeth Banta
IBM Systems
Elizabeth.Banta@ibm.com
732-996-4159

1 Morning Consult – Global AI Adoption Index 2021

SOURCE IBM

Additional assets available online: Photos (14)  Video (1)