FINTECH

Solution Brief

Revision 0.1

habana®

I.

AI in Finance

More than 70.3 billion real-time payment transactions were processed globally in 2020, an increase of 41 percent compared to the previous year [1].

This has been driven by the COVID-19 pandemic which accelerated trends away from physical transactions to relying on real-time and digital payments. Digital payments are expected to grow at a CAGR of 23% up to 2025. Social distancing due to pandemic driving virtual interactions and pandemic impact on the economy has spurred the deployment of AI solutions in banking and finance to streamline financial operations efficiently. AI is used in various financial sectors such as banking. insurance services, loans and credit management, personal finance, digital payments, wealth management, etc. [2].

II.

Applications of AI in Finance

Banks and financial institutions have been using AI to enable digitized efficient 24x7 customer experience with the use of virtual assistants, chatbots, AI-driven customer onboarding and Know Your Customer (KYC) procedures, and AI-driven fraud detection and Anti-Money Laundering (AML) operations.

The use of AI in the financial sector is becoming more widely adopted moving beyond traditional front-end solutions driving customer interactions to investment analysis, wealth management, credit risk management, regulatory compliance, etc. [3].

According to a Business Insider report [2], the estimated savings due to the deployment of AI solutions in banking and finance will be around $443 Billion in 2023. These savings are expected to accrue from the deployment of AI in front-end solutions driving customer interactions, middle-end solutions for fraud and risk management and compliance, and backend solutions for credit underwriting. AI enables the automation of many of the highly manual effort-driven processes such as credit risk management, KYC verification, fraud detection and AML, contract analysis, and regulatory compliance enforcement [3].

Challenges

Given the complexity and real-time needs of financial applications, AI deployment is key to ensuring a smooth and compliant financial process flow while ensuring that customer needs are met efficiently. Financial applications deal with huge volumes of structured and unstructured data. Complex AI models are needed to analyze and extract insights from this data.

Time is of the great essence for most finance applications. The faster the training cycles of AI models, the sooner we can achieve the desired results.

III.

Why is Gaudi a good fit for finance use cases?

Deep neural network-based models for financial data include a large amount of processing that can be parallelized and thus accelerated. Finance use cases benefit specifically from accelerators that can handle data parallelism when the training dataset is huge and model parallelism when the models are large.

The two primary considerations that come into play in employing AI processing—whether for computer vision or NLP applications-- are time-to-train models to the desired level of accuracy and cost-to-train. Habana's Gaudi Training Processors are expressly designed—in both hardware and software—to deliver high-efficiency cost- and time-to-train, making AI training more accessible to more organizations and for more applications. This helps to reduce development and validation costs and to enable rapid innovation and faster time to market.

Training with Gaudi clusters is available both in the cloud with AWS EC2 DL1 instances consisting of 8 Gaudis and on-premises with the Supermicro X12 Gaudi Training Server, also featuring 8 Gaudis.

The ideal equation for end users is to achieve desired AI price-performance, meaning that the cost and time to train each image or language sequence meets cost and time investment criteria. In other words, enabling more training at a low cost is the objective for data scientists and IT infrastructure management.

First-generation Gaudi, in fact, has proven delivery of up to 40% better price-performance than with comparable GPU-based solutions—for both the EC2 DL2 instance as well as for on-premise systems. And, there are customer cases that have proven even greater cost savings, which will be shared in the next section.

In addition, Gaudi2, which launched in May, offers substantial performance advances that enable significantly faster training of models, while preserving cost-efficiency. Gaudi2 systems will be available in 2H 2022 for on-premises implementation.

IV.

News and customer testimonials

In times of increased volatility and fragile economic conditions driven by global events such as pandemic, war and inflation, finance managers need to know the impacts of market conditions and external world events on a given instrument in real time as the day progresses to take appropriate actions. Unlike exchange-traded instruments, where values can be observed each time the instrument trades, values for derivatives need to be computed using complex financial models. One of the key areas of AI applications in finance is in providing real-time valuations and volatility analysis of various financial instruments such as derivatives [4].

One of our customers is a FinTech Startup providing real-time valuations and risk sensitivities throughout the trading day. They deployed complex AI models with RESNET-inspired architecture and trained on synthetic datasets (derived from complex slow solvers) to provide fast, timely information on valuations throughout the trading day. Models encoding symmetries in neural network architecture while being highly performant can adversely impact training time. Our customer benefitted from Gaudi-based DL1 instances to overcome this problem, and this enabled them to build high-quality models with lower training costs.

Riskfuel

"Our experiences with Gaudi give us confidence that we will be able to lower our training costs while improving model quality and translate that into even more powerful tools for our end-user."

Maxime Bergeron
R&D Director
Riskfuel

V.

References

[1] "Global Real-Time Payments Transactions Surge by 41 Percent", ACI Worldwide Report, available at https://investor.aciworldwide.com/news-releases/news-release-details/global-real-time-payments-transactions-surge-41-percent-2020

[2] "The impact of artificial intelligence in the banking sector & how AI is being used", Business Insider, available at https://www.businessinsider.in/finance/news/the-impact-of-artificial-intelligence-in-the-banking-sector-how-ai-is-being-used-in-2020/articleshow/72860899.cms

[3] Longbing Cao, "AI in Finance: Challenges, Techniques and Opportunities", 2021. https://arxiv.org/abs/2107.09051

[4] R. Ferguson and A.D. Green, "Deeply Learning Derivatives", 2018, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3244821