

Solving Challenges of Instant Payments by Using AI on IBM zSystems

Abid Alam Tabari Alexander Suhas Kashyap



IBM zSystems



Point-of-View

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

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

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

CICS®	IBM zSystems™	WebSphere®
IBM®	Redbooks®	z/OS®
IBM Watson®	Redbooks (logo) 🧬 🛽	z16™

Other company, product, or service names may be trademarks or service marks of others.



Highlights

- Industry-first, integrated, on-chip, Al accelerator designed for high-speed, latency-optimized inferencing with sustainability as a core principle
- Uses an on-chip AI accelerator to score 100% of transactions in real time
- Up to 300 billion deep learning inference requests per day with 1 ms response time
- Up to 19x higher throughput and 20x faster response time co-locating applications and inferencing

The cost of fraud on your business

Fraud poses a significant risk to businesses and end customers. Fraud can occur across various industries, including but not limited to credit cards, healthcare, payments, financial services, government, and insurance.

The cost of fraud continues to grow and impact businesses. Global credit card fraud is expected to reach \$35 billion by 2025. Since the beginning of the COVID-19 pandemic, 40% of financial services firms saw an increase in fraudulent activity. The COVID-19 pandemic also worsened the false positive cases for over two-thirds of the institutions.

In today's financial sector, some costs of fraud can be avoided:

- Major banks can avoid over \$1 B in fines because of inadequate anti-money laundering (AML) programs.
- 90% of false positive rates by legacy anti-fraud models can be prevented.

In an artificial intelligence (AI) optimized future, some losses can be significantly reduced:

- ▶ \$140 B in loss reduction can be achieved by US banks alone.
- \$100 M per bank loss reduction can be achieved through scoring of 100% of Tier 1 Bank's transactions.

AML efforts consist of laws, regulations, and procedures that are designed to prevent criminals from exchanging money obtained through illegal activities, sometimes called dirty money, into legitimate income, sometimes called clean money. Although money laundering is an international crime, many rules are local, and they can sometimes conflict with federal policies, making it difficult for financial institutions to remain compliant with rules and regulations. Some banks decided to suspend services in countries that make it difficult to stay compliant or have a reputation for facilitating money laundering. This IBM Redbooks Point-of-View publication describes how the IBM z16[™] can provide your enterprise with several advantages:

- An industry-first, integrated, on-chip, AI accelerator designed for high-speed, latency-optimized inferencing with sustainability as a core principle.
- ▶ Up to 300 billion deep learning inference requests per day with 1ms response time¹
- ► Up to 19x higher throughput and 20x faster response time co-locating applications and inferencing

This publication is written for those who play a strategic role within an organization; they hold senior positions and impact company-wide decisions. This publication is also useful for consultants or IT architects.

Real-time or instant payments

Real-time or instant payments are not a new concept and are not limited to large financial institutions. The concept of real-time payments has existed since the late 1970s. The first version of a real-time payments platform was started in Japan in 1973 but with limited capabilities. In the late 1980s, a true real-time payment solution was introduced in Switzerland. Since then, the availability of real-time payment networks and schemes increases around the world, but until recently, the rate of adoption was slow.

However, in the last 2 decades, the rate of adoption has accelerated because of the availability of newer technologies and more importantly, increased customer demand for seamless and frictionless payments.

Today, the value of the global market for real-time payments is estimated to be \$13.5 B and is growing at a compound annual growth rate of 30% from 2022-30. person-to-business (P2B) segment accounted for 64% of the share of global revenue that was driven by the COVID-19 pandemic, but the peer-to-peer (P2P) segment is anticipated to register the highest growth rate over the forecast period. With this tremendous growth forecast, the marketplace will yield more opportunities for fraud detection and monitoring solutions.

Different payment types are in use in today's world. Some of the more popular ones include the following types:

Automated clearing house

An automated clearing house (ACH) is an automated solution system that is the backbone for moving money between bank accounts in the US.

It is the direct evolution of the paper check transformed into a digital process to improve efficiency and reduce the need for human input. It is often referred to by functional nicknames such as direct deposit, direct debit, auto-pay, and the generic term bank transfer. It is also the underlying technology behind most peer-to-peer transfers that are made through services like Venmo, PayPal, Cash App, and Zelle.

¹ https://www.ibm.com/downloads/cas/DOXY3Q94

ACH transfers can occur between any two financial institutions on the network and can either push or pull money through the system as needed. Each transaction is submitted to the network as part of a batch that the network's operators repackage into new bundles for each receiving institution 5 times per business day. The party that initiates the request can either choose to pay for same-day service or default to one business day (for debits being pulled through) or two business days (for credits being pushed through). ACH involves the following processes:

- Authorization

A payment certificate that verifies the operation.

- Clearing

The process of updating the account.

Settlement

The actual movement and availability of funds. Money sits in central bank reserves in the interim.

► Faster payments

The funds clearing process is faster than the traditional ACH batch process, but funds are not settled instantly. The faster payments process still runs on an underlying, traditional ACH process.

This description leaves some room for interpretation, but workers in finance generally agree that faster payments include methods that issue confirmations within 15 seconds, with transactions completed within two hours.

Wire transfer

Direct point-to-point transfers between any two financial institutions that are typically used for higher-value items like property purchases or for settling large institutional transactions

Real-time payments and instant payments

Real-time payments are a new type of payment method where funds are cleared and settled instantly, with no delays.

This type of payment can be initiated through one of many channels:

- Smartphones
- Tablets
- Digital wallets
- The web

Real-time payment network infrastructures, also called rails, bring end-to-end, bidirectional communication. With older rails, communication flowed in one direction: from the payer to the payee. If the two parties wanted to exchange information in both directions, they did so outside of the payments system. Real-time payments connect the payment with payment data together in a single transaction. This newer, more seamless way to make financial transactions will affect consumers, businesses, and world governments socially, culturally, and economically.

Figure 1 provides an overview of the beneficiaries of real-time payments.



Figure 1 Beneficiaries of real-time payments

Figure 2 shows current real-time payment systems around the world.



Figure 2 Current real-time payment systems around the world

Figure 3 shows the components that make up a real-time-payment systems structure and how they interrelate.



Figure 3 Real-time payment systems structure

AI on IBM zSystems[™] can help prevent the challenge of fraud and money laundering in real-time payments by providing the following benefits:

- Being able to apply AML at every payment, resulting in more efficient screening and the ability to stop AML payments before they are released
- Improving accuracy of AML screening that meets stricter regulatory requirements without impacting service level agreements (SLAs)
- Augmentation of current business rules with AML, which can reduce the false positives that can result in high-cost investigations

Al offerings on IBM zSystems can work together to identify money laundering patterns in instant payments. The real-time solution can include the following components:

- ► Models that are trained on popular open-source frameworks can be used for inferencing.
- IBM Watson® Machine Learning for z/OS® (WMLz) to deploy an AML model trained anywhere on IBM zSystems for inference.

Fraud and money laundering in real-time payments

Various threats can target the instant payments systems. Because of the nature of instant payments with immediate clearing and settlement and the permanence of the transaction, the opportunity for fraudulent activities and schemes increases.

Adding a new rail inevitably introduces the risk of fraud, and organizations often act conservatively to manage it. Fraud affects all financial institutions, and real-time payment adds another layer of complexity to their existing fraud mitigation strategy. Most financial institutions do not have real-time decisioning for sent payments, which does not allow them to effectively monitor or manage the risk of a real-time payment service.

Business processes

Organizations that are not financial institutions, such as accounts payable (AP) and accounts receivable (AR) solution vendors, might find it difficult to keep up with the changing market and be forced to roll out advanced updates to support the new rail. AP and AR providers might have to completely reshape their business models, as real-time payments will change the structure, requirements, and communication flows of traditional B2B payments.

Fraud poses a significant risk to businesses and end customers. Fraud can occur across various industries, including but not limited to credit cards, healthcare, payments, financial services, government, and insurance. Using a scalable and consistent AI solution to detect, prevent and address fraud can help reduce this risk. Determine business insights during transactions by harnessing the scalability and speed of IBM zSystems to address these challenges.

The cost of fraud continues to grow and impact businesses. Global credit card fraud is expected to reach \$35 billion by 2025. Since the beginning of the COVID-19 pandemic, 40% of financial services firms saw an increase in fraudulent activity. The COVID-19 pandemic also worsened the false positive cases for over two-thirds of the institutions.

Anti-money laundering (AML) efforts consist of laws, regulations, and procedures that are designed to prevent criminals from exchanging the money obtained through illegal activities, dirty money, for legitimate, clean money, income. Although money laundering is an international crime, many rules are local, and local rules can sometimes conflict with federal policies, making it difficult for financial institutions to remain compliant with rules and regulations. Some banks decided to suspend services in countries that make it hard to stay compliant or in countries that have a reputation for facilitating money laundering.

Reducing the risk of fraudulent transactions in your enterprise with the IBM z16

The use of a scalable and consistent AI solution to detect, prevent, and address fraud can help reduce the risk of fraudulent transactions. Determine business insights during transactions by using IBM zSystems and harness the scalability and speed that is needed to address these challenges.

Al on IBM zSystems can help prevent fraud and money laundering in real-time payments by being able to apply AML techniques at every payment. This results in more efficient screening, which stops fraudulent payments before they are released.

Using on-chip AI accelerator to score 100% of transactions in real-time

One possible solution architecture for an IBM CICS® COBOL application to use AI and WMLz online scoring is by using ALNSCORE, shown in Figure 4. This has the following advantages:

- ► Enables simplified AI integration for CICS COBOL applications.
- ► CICS COBOL applications can start ONNX models by using standard CICS PUT/GET container commands.
- Provides features for optimal exploitation of Telum Integrated Accelerator for AI.
- WMLz manages deployments of model-scoring service.



Figure 4 Using ALNSCORE

Solution architecture for an IMS COBOL application

Figure 5 presents a possible solution architecture for an IMS COBOL application that leverages AI by using WMLz online scoring. This solution takes advantage of low-latency IBM® WebSphere® Optimized Local Adapter (WOLA) connectors to provide AI scoring capabilities to IMS COBOL applications. WMLz provides the ability to generate copybooks based on a deployed open neural network exchange (ONNX) model to define the appropriate interface for the application. Using this and other generated artifacts, a COBOL application that is running in a configured IMS message processing region can use existing WOLA APIs to pass scoring requests to WMLz to instantly score transactions.

This has the following advantages:

- ► Enables simplified AI Integration for IMS COBOL applications.
- ► IMS COBOL applications can score with ONNX models over WOLA APIs.
- Provides features for optimal exploitation of Telum Integrated Accelerator for AI.
- ► WMLz manages deployments of model scoring service.



Figure 5 Architecture for driving AI scoring with an IMS COBOL application

How IBM can help

IBM has various proprietary and open source AI solutions on IBM zSystems that can help you to attain your goals of fraud prevention. The following list includes some of the resources to solutions that IBM can provide. Your IBM client representative can provide you with specifics.

- IBM Cloud Pak for Data for zSystems
- ► IBM Z Deep Learning Compller
- IBM Z Optimized for TensorFlow
- Snap ML
- Watson Machine Learning for z/OS

Resources for more information

For more information about the concepts highlighted in the paper, see the following resources:

- Interact with the AI on IBM zSystems Team provides methods for you to discuss your AI on IBM zSystems and LinuxONE solutions.
- Resources for AI on IBM zSystems website provides additional resources for you to explore, including demo videos, solution briefs, and other IBM Redbooks® publications.



REDP-5698-00

ISBN 0738461148

Printed in U.S.A.



Get connected

