

# AI on IBM Z and LinuxONE Strategy for 2026 and Beyond

**February 23, 2026**  
**2:30-3:30, Salon 22**



Elpida Tzortzatos  
IBM Fellow, CTO AI on IBM Z & LinuxONE  
[elpida@us.ibm.com](mailto:elpida@us.ibm.com)

Steve Warren  
STSM, Principal Solution Architect  
Worldwide System Center  
[swarren@us.ibm.com](mailto:swarren@us.ibm.com)

# AI Strategy & Use Cases



# IBM Z accelerates enterprise agentic AI transformation

- Full “agentic AI stack” enables client-built agents and powers IBM Z software innovation
- Access to systems, applications and data for insight and decisioning, with built-in security by design
- Multi-model and hybrid architecture optimizes cost and performance with “fit-for-purpose” approach



## *Secure transaction processing with agentic AI on IBM Z*

### Z Software and Applications

Data • Tools • Agents

Concert for Z • SQL Data Insights Pro • Client Apps (BYO)

### Common Agentic Services

Agentic Framework • Security • Orchestration

watsonx Assistant for Z



RHOAI

### AI Inferencing Services

Multi-model routing • performance optimization • monitoring

AI Optimizer for Z



RHOAI

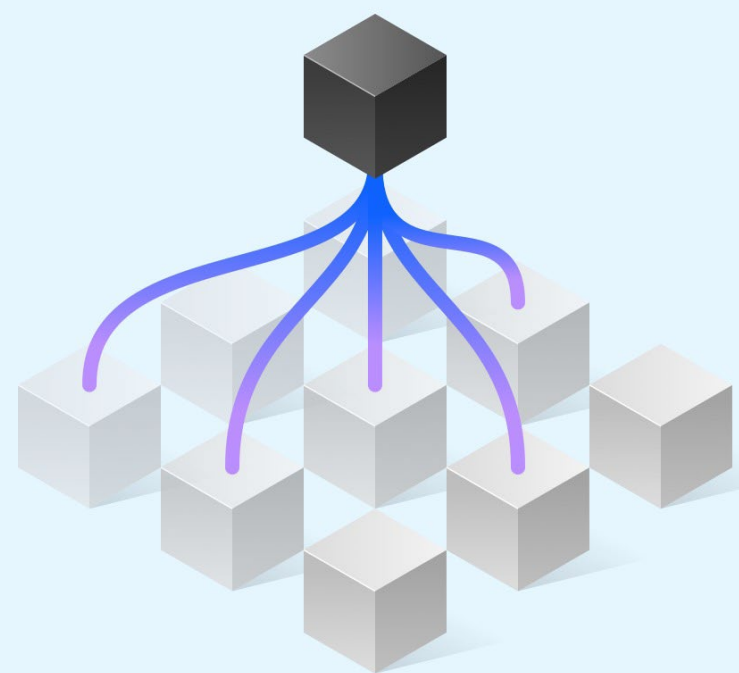
### AI Accelerators for SLMs

Telum II Processor • Spyre Accelerator

IBM z17

SLMs, LLMs,  
Frontier Models

## Delivering broad spectrum of key AI capabilities across the z platform stack Compute Differences



# Predictive AI vs. Generative AI vs. Agentic AI

## Predictive AI

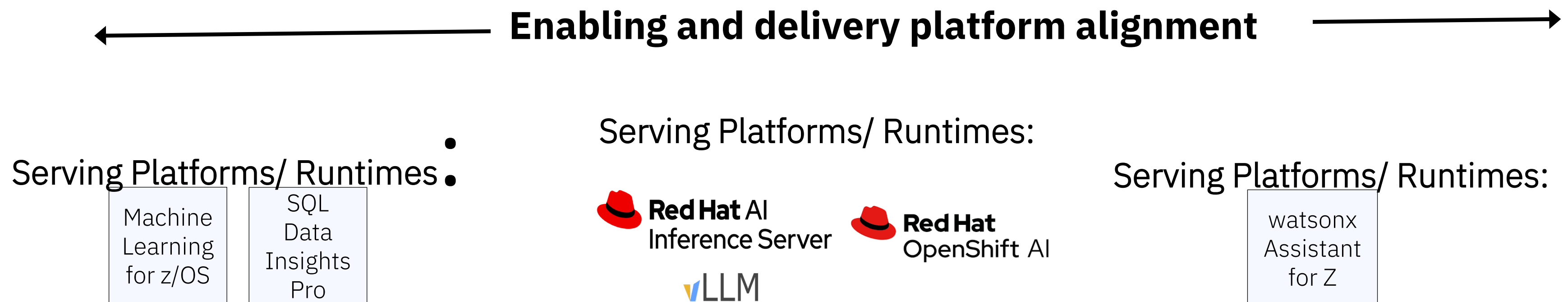
- ✓ Enables advanced business transaction capabilities
- ✓ Speed and scale are critical.
- ✓ Low latency and high throughput are absolute requirements.
- ✓ Today: smaller models less compute needed (ML/DL model and encoder LLMs).

## Generative AI (GenAI):

- ✓ The focus is on acceptable performance, measured in tokens per second, rather than the best performance in the industry.
- ✓ Larger Models requiring a lot more compute (GenAI decoder LLMs).
- ✓ Security is a priority, ensuring the protection of client data and models.

## Agentic AI

- ✓ Agentic AI is a framework for accomplishing goals with limited supervision that consists of AI agents.
- ✓ AI Agents use LLMs (LLM in a loop) to reason and can interface with tools, other models, and other IT systems to fulfill user goals.
- ✓ Capacity planning shifts from single-model sizing to workflow-level concurrency.

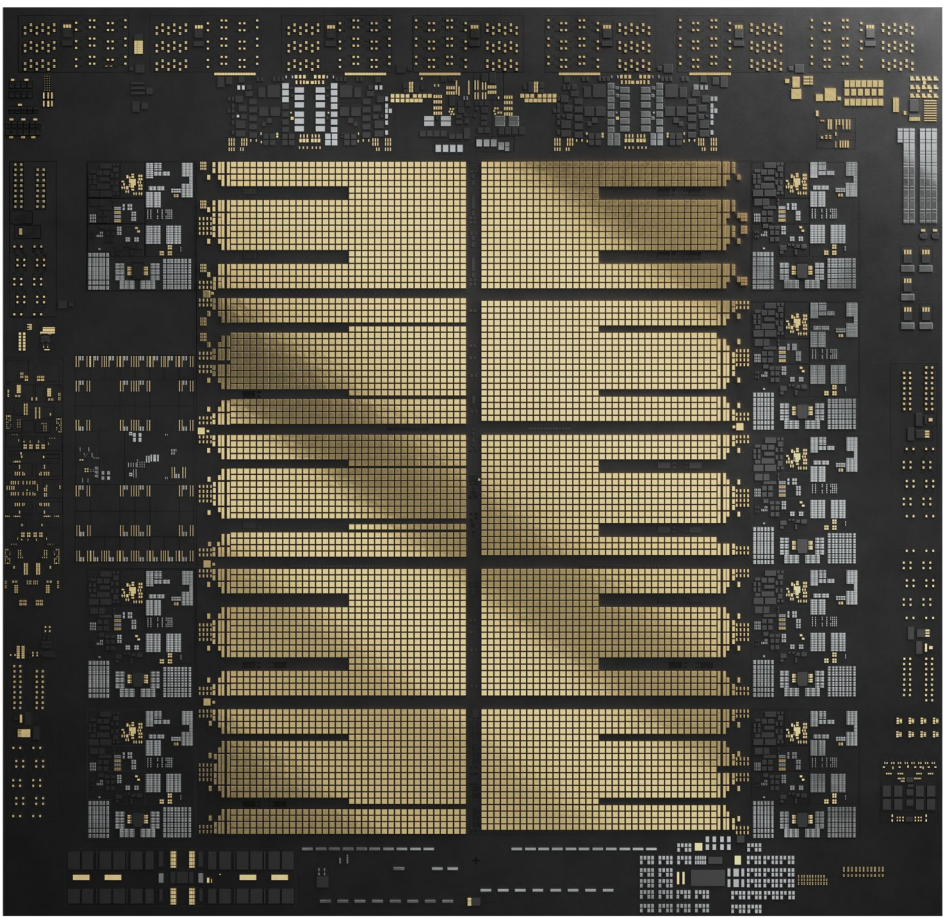


# IBM Announced New Telum II and IBM Spyre Accelerator

## Expanding AI on IBM Z

- 70% of all financial transactions by value go through an IBM Z System
- 99.999999% availability
- In-transaction AI for millions of transactions per second

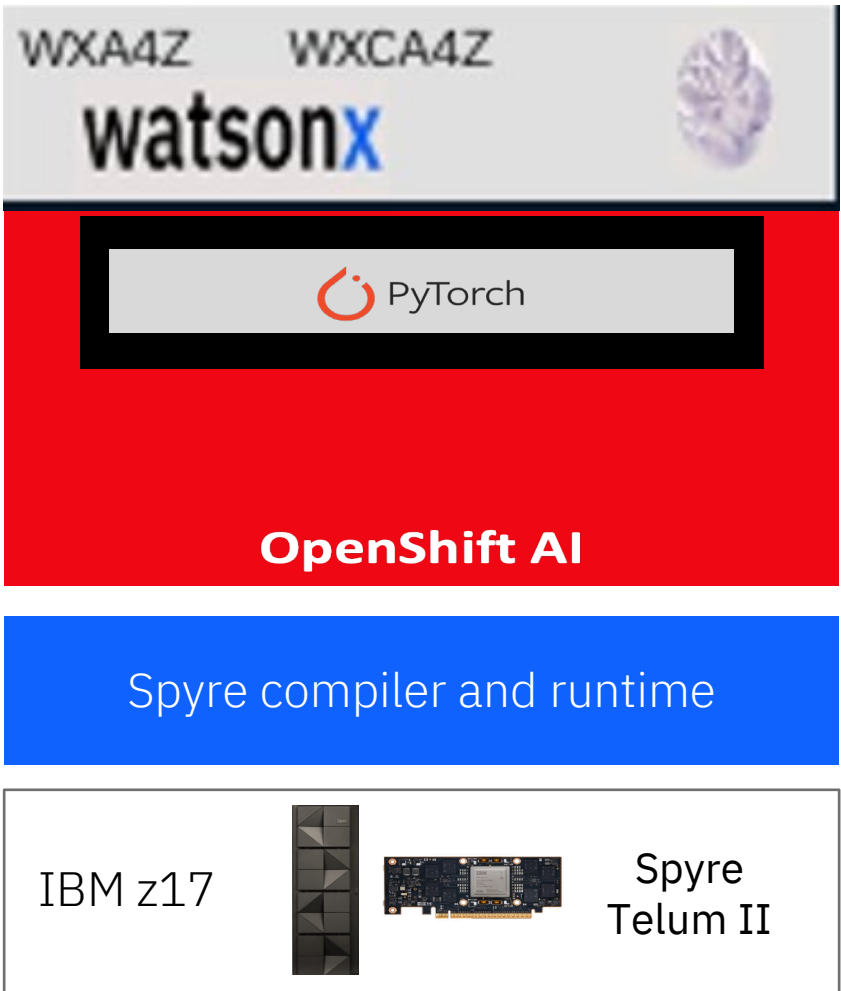
Telum™ II



Spyre™ Accelerator



GenAI Software Stack



5nm logic technology

Micro-processor

43 billion transistors  
600 sqmm  
24 miles of wire

AI Accelerator Chip

26 billion transistors  
330 sqmm  
14 miles of wire

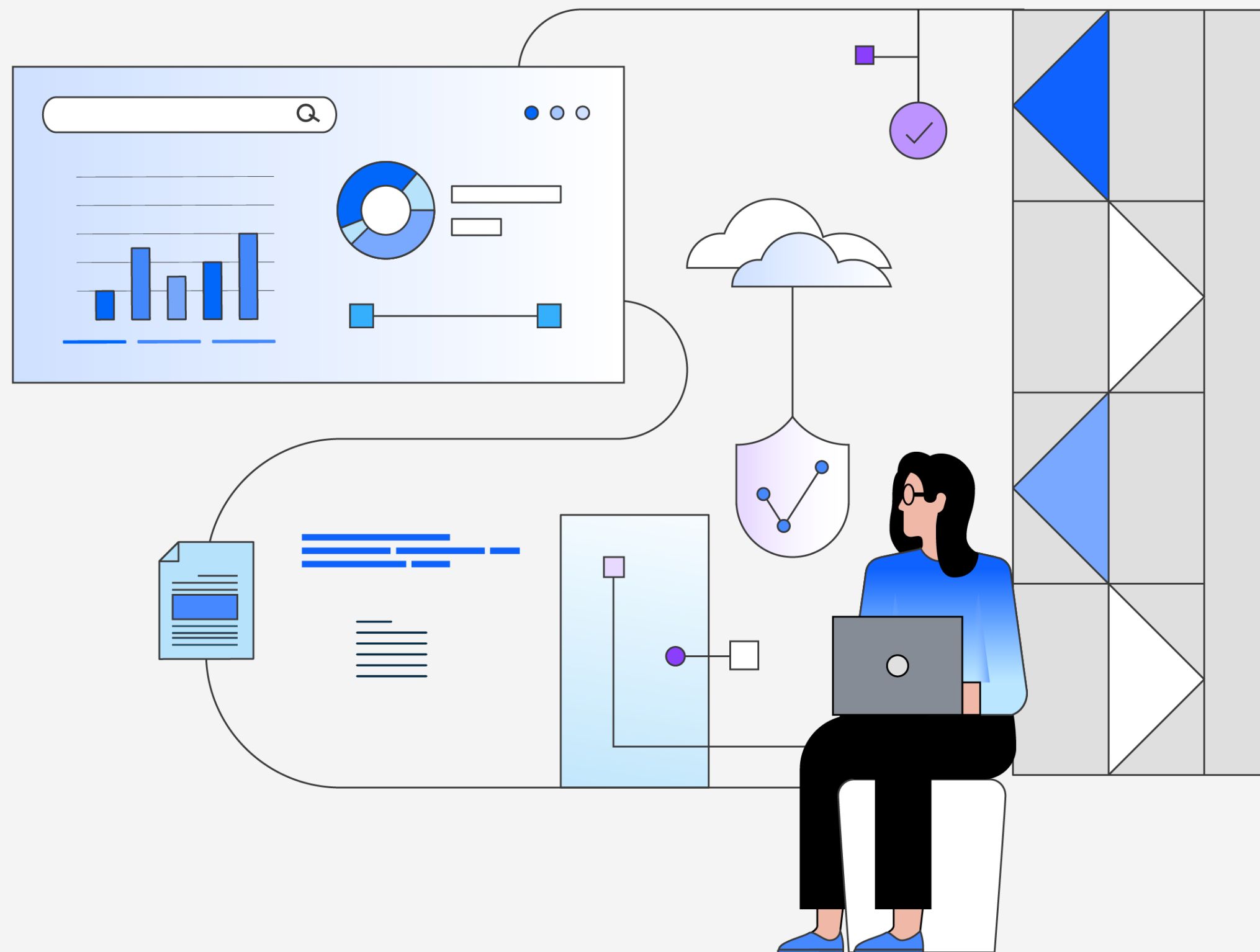
Optimized Integrated Stack

Maximum convergence of AI software, compilers, open source

Drive optimizations across all layers of the stack

IBM mainframes run an estimated 70% of all world transactions by value.  
Source: Operationalizing Fraud Prevention on IBM z16, Celent, 2022. [ibm.com/downloads/cas/DOXY3Q94](https://ibm.com/downloads/cas/DOXY3Q94)

# AI inside transactions, where milliseconds matter



Transactional AI embeds **machine learning models** directly into your *transactional systems*

Instead of waiting for off-platform AI, insights are injected right into the flow:

- ✓ Fraud flagged before approval
- ✓ Payments rerouted before failure
- ✓ Claims adjusted before submission errors
- ✓ Customer offers are optimized while they're interacting

# IBM z17

*Fully engineered stack for AI where it matters most*



Transaction  
processing platform

Operating systems  
& firmware

IBM Z  
infrastructure

Built on a foundation of security,  
resiliency, and high availability.

450 billion

Inference operations per  
day with 1ms response  
time vs. 300B on IBM  
z16<sup>8</sup>

7.5x  
AI throughput

Utilizing 8 AI processing  
units vs. one on IBM z16<sup>9</sup>

83%  
less power

Moving AI-infused OLTP  
workloads from compared  
x86 servers<sup>10</sup>

# Fraud moves fast!!

## Avoid latency and limited coverage to prevent weak links



Fraud is rising fast — **\$533B lost globally** across banking and payments in 2024<sup>†</sup>.

- Cards and payments fraud is growing at 14.2% annually.
- External fraud solutions often analyze **only 1 in 5 transactions**, often post-authorization — limited by **latency, architecture, and SLA constraints**.
- This leaves **millions of dollars unprotected — every single day**.

### Business Case – Card Payments

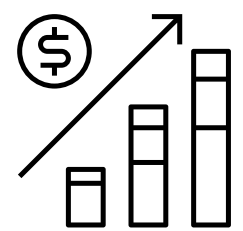
- 5M Transactions/day at \$85 = \$425M
- 1.5% fraud rate = \$6.38M

| Approach              | Coverage       | Daily Fraud Prevented | Annual Impact     |
|-----------------------|----------------|-----------------------|-------------------|
| Off-platform AI       | 20% sampled    | \$1.28M               | \$466M            |
| Transactional AI on Z | 100% real-time | \$6.38M               | \$2.33B           |
| Value Creation        |                | +\$5.10M daily        | +\$1.86B annually |

AI on IBM Z enables 100% transaction scoring at the moment of decision — not after the fact. That’s how you move from **\$466M saved to \$2.33B protected**.

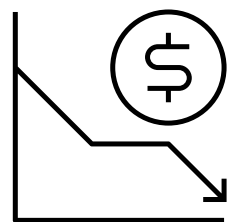
1. Mitigating Fraud in The AI Age: Understanding the Challenge, Celent, March 2025.

# Transactional AI enhances the business drivers that matter most – revenue, cost, and risk



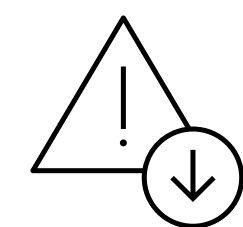
## Revenue Growth

- Intelligent decisioning
- Real-time personalization
- Dynamic pricing



## Cost Reduction

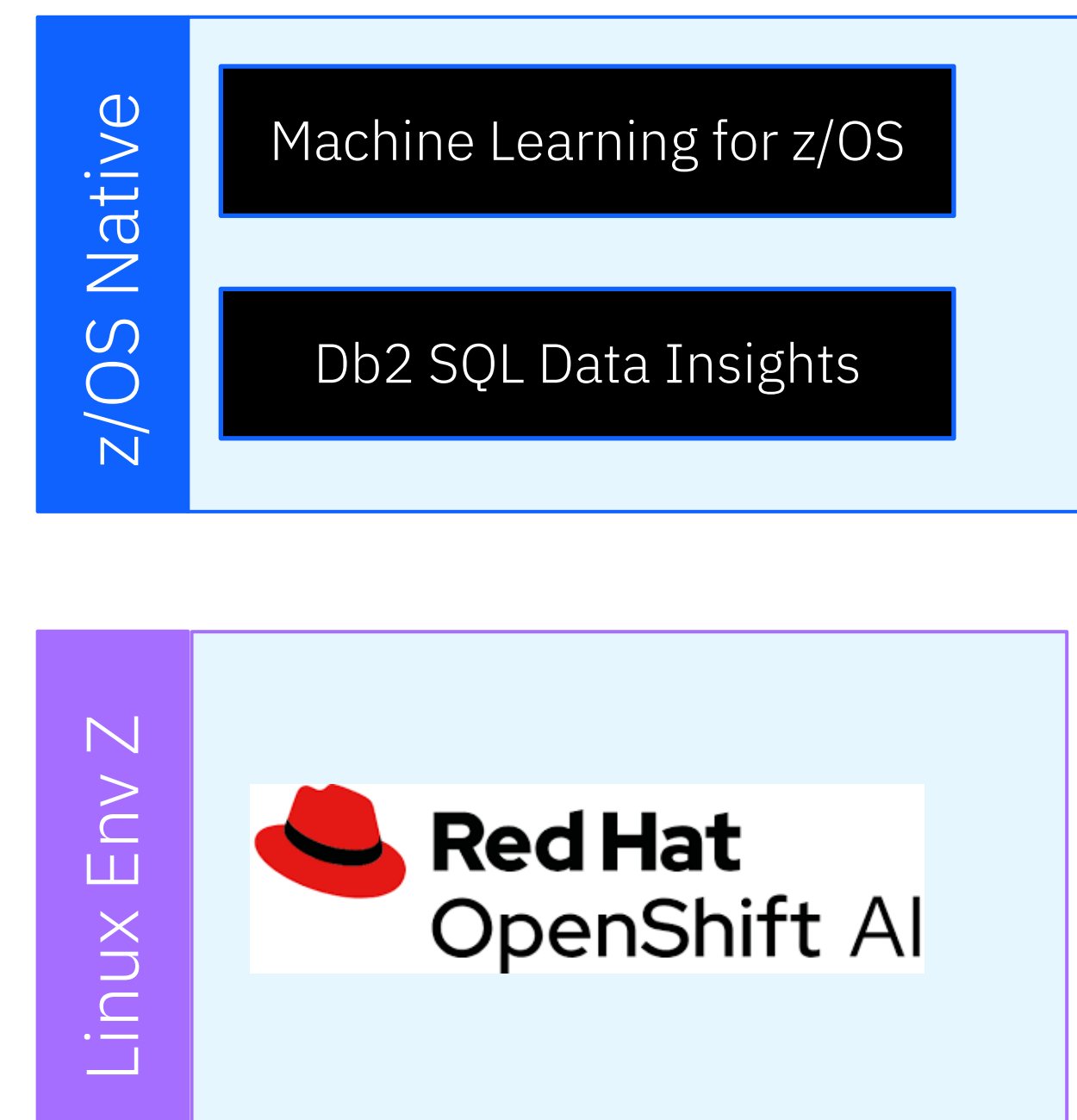
- Process automation
- Exception handling
- Resource optimization



## Risk Management

- Fraud prevention
- Compliance monitoring
- Operational resilience

## Transactional AI Stack

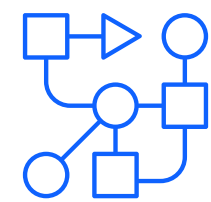


# Machine Learning for IBM z/OS

*Accelerated inferencing at scale for Transactional AI workloads on z/OS*

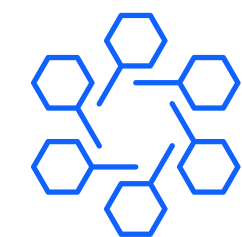
The full-featured machine learning platform on z/OS designed to infuse AI into mission critical IBM Z enterprise applications by leveraging AI models trained anywhere or on IBM Z and deploying them on z/OS including a multi-model architecture, co-located with enterprise applications, transaction data, and business logic for high throughput and extremely low latency to modernize applications and drive business insights at scale

## Benefits



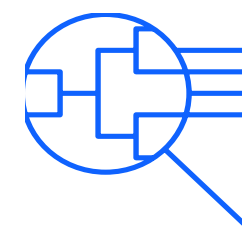
### End-end platform

GUI and script driven features for end-end AI model lifecycle management



### Scalable AI

Leverage the latest accelerator to score transactions natively in CICS, IMS and batch applications at scale



### Easy import and deploy

Leverage predictive models and Encoder LLMs trained anywhere or with watsonx.ai and deploy with MLz

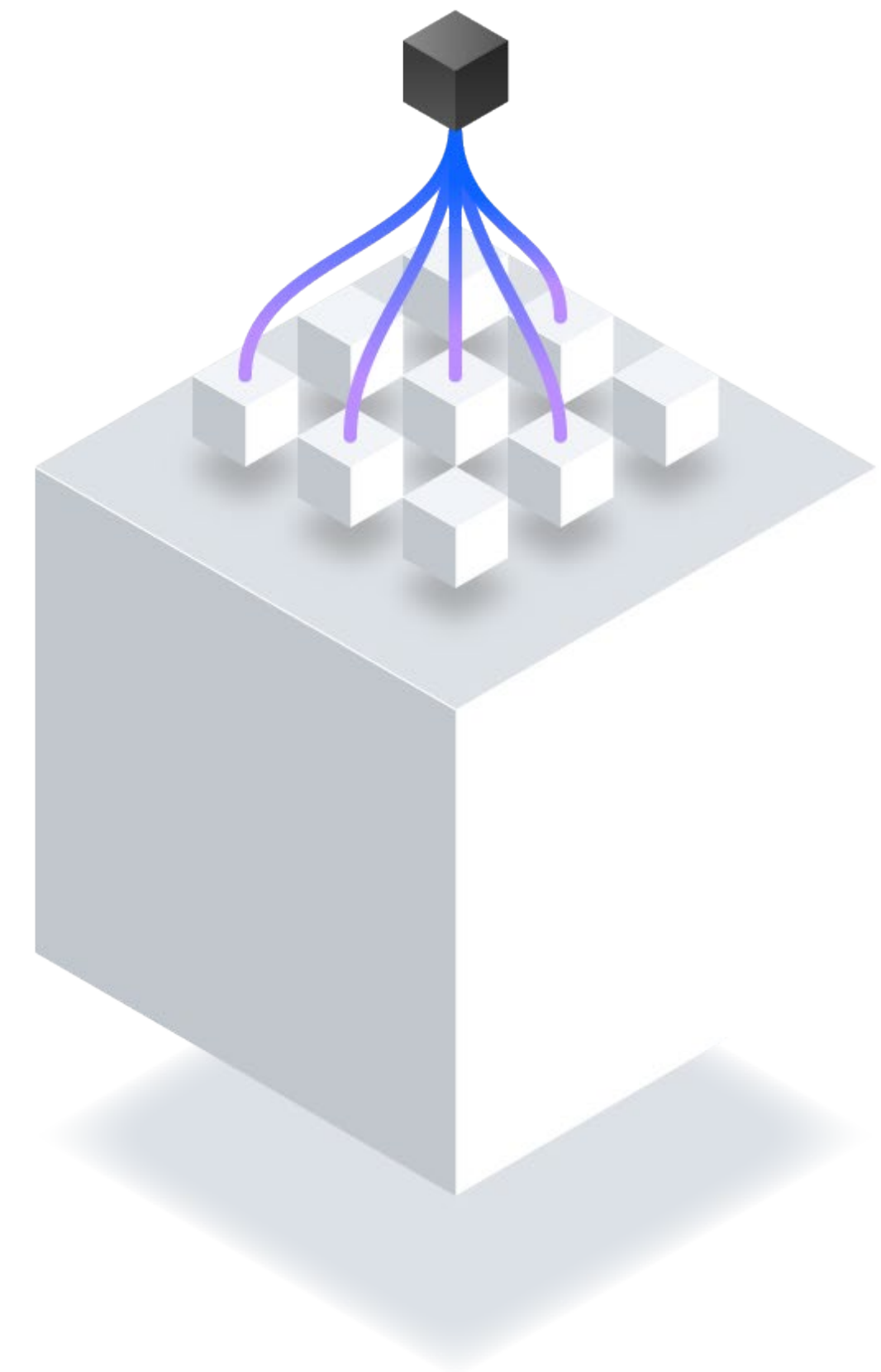


### Trustworthy features

Leverage built in explainability and drift detection features or enterprise AI governance with watsonx.governance

## Transactional AI

Infuse AI into every transaction



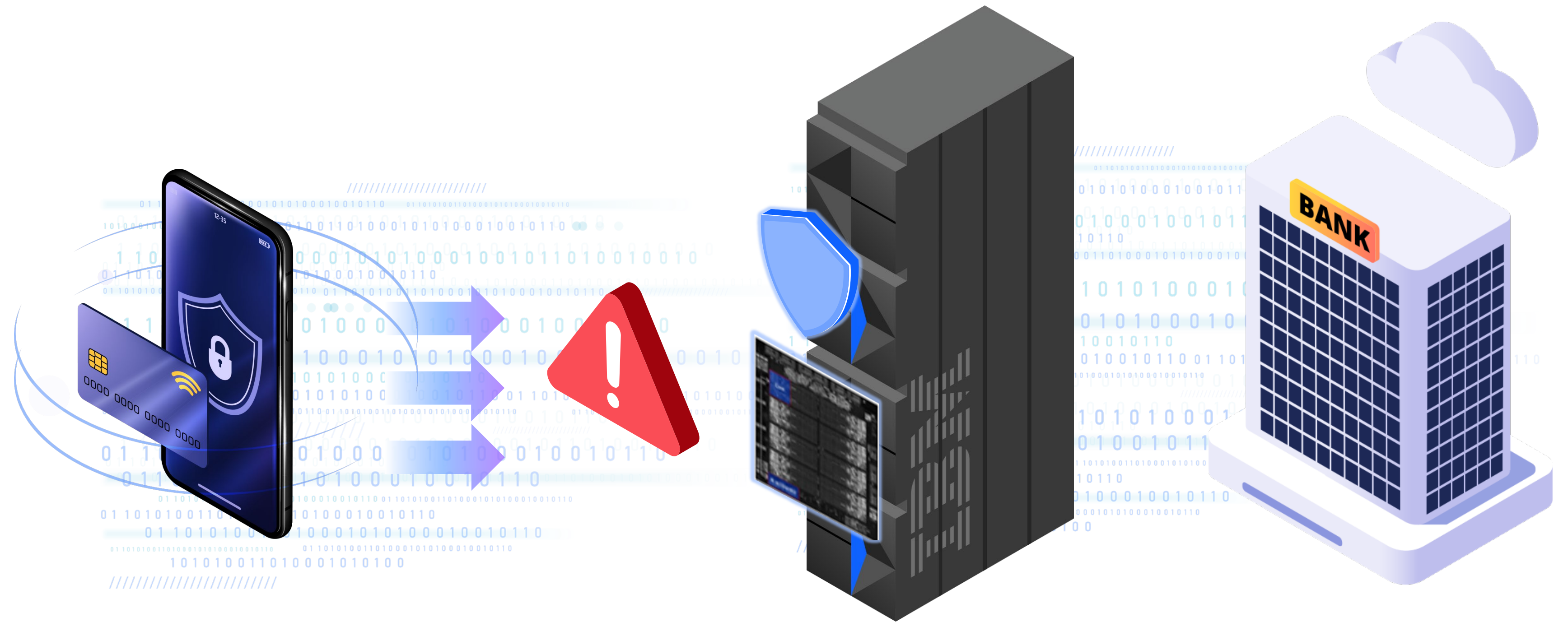
# Building on Experience: Value of real-time payments fraud detection with IBM Z<sup>®</sup>

A global financial company wanted to capture fraud before it happened, before payment was made, without impacting resources and response time on their systems.



MACHINE LEARNING  
FOR IBM z/OS

Deliver AI solutions at an  
unprecedented speed



**With AI on the mainframe, this client went from 80ms to 1ms inferencing.**  
Prior to exploiting AI on the system, only 10-20% of fraud was detected before transaction complete, and now they can do 100%.

**Upwards of \$20M savings per year in fraud they do not have to try to claw back!**

# Getting ready for EU Instant Payments' Regulation Pt 1

## The Client Situation

A European large IT provider for banks serving hundreds of FSS institutions needs to address EU regulation (IPR) to enable instant payments with real-time payee verification and sanctions screening at the point of transaction

## The Business Challenge Faced

- The **fraud rate in instant payments is disproportionately higher** than in other payment transactions. Additionally, from 2025 onward, the **use of their instant payments will increase sharply**.
- Banks required that the fraud fees for instant payments must be identical to the existing. Unfortunately, **fraud fees were higher for instant payments**, meaning costs would be passed to customers.
- The current off-prem AI solution **cannot intervene in real-time payments**, as there is no downstream processing in the check dialog for instant payments. **Fraud was being missed and was too costly**.



# Takeaway:

## Transactional AI on IBM Z delivers real-time inferencing in 100% of transactions compared to off-prem AI

### Co-location Benefits

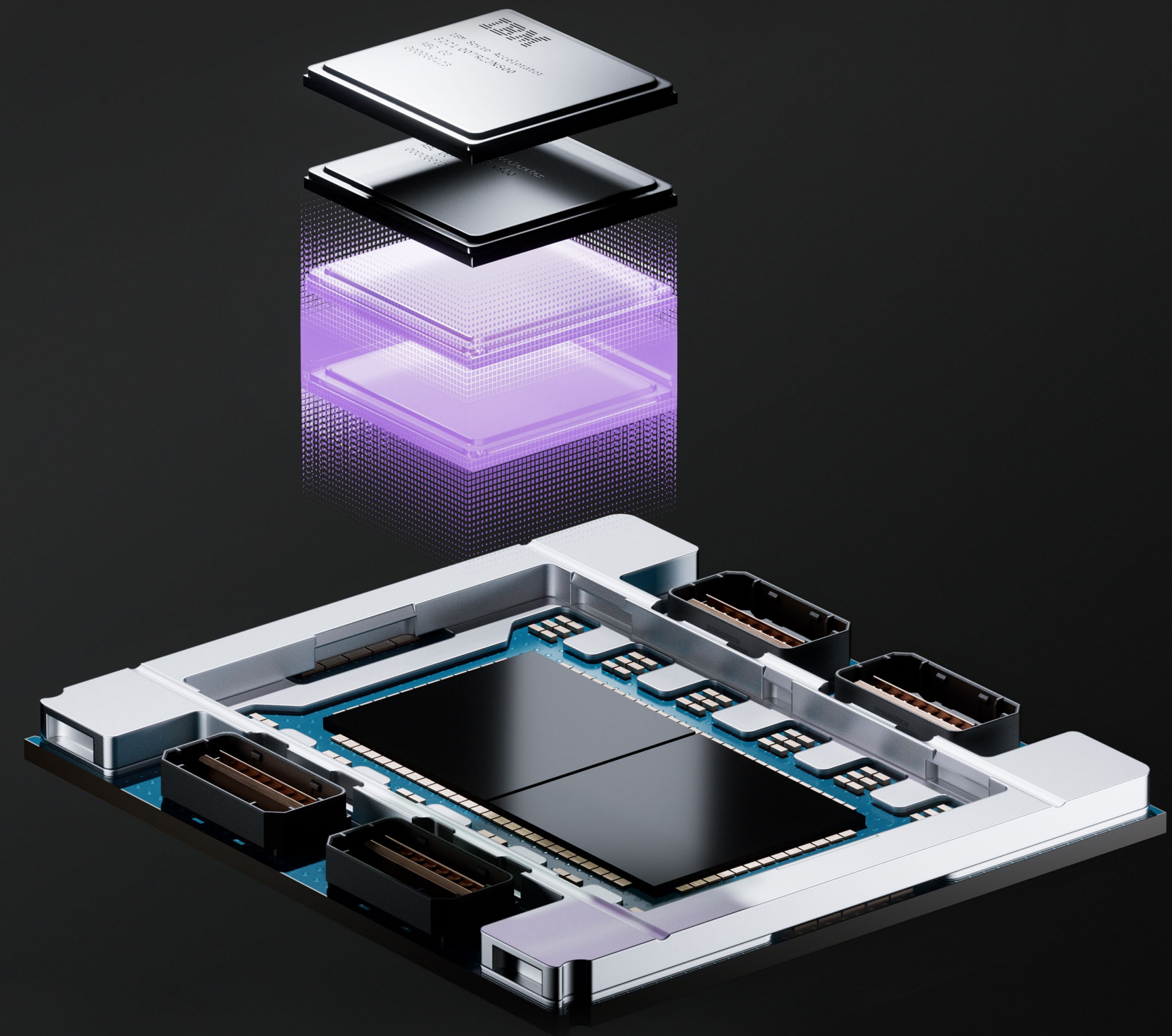
- Inference runs next to the data
- No latency
- No transfer risk

### Hardware Acceleration

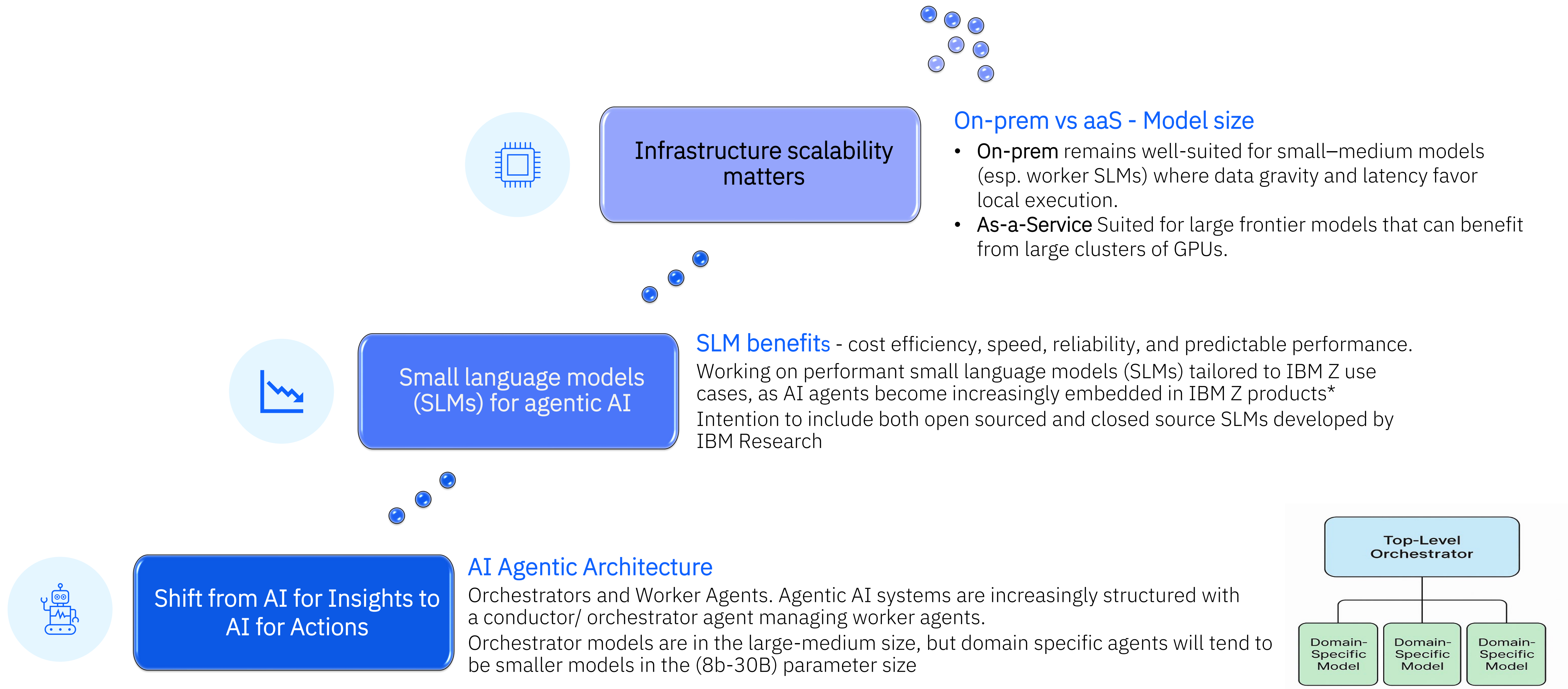
- Telum & Telum II delivers inferencing acceleration for low-latency scoring

### Qualities of Service

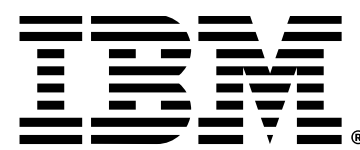
- Models operate on complete, trusted transactional data
- Availability
- Security
- Resiliency



# GenAI and Agentic AI trends – favors hybrid architecture with ‘fit-for-purpose’ approach



\*in collab. with IBM Research



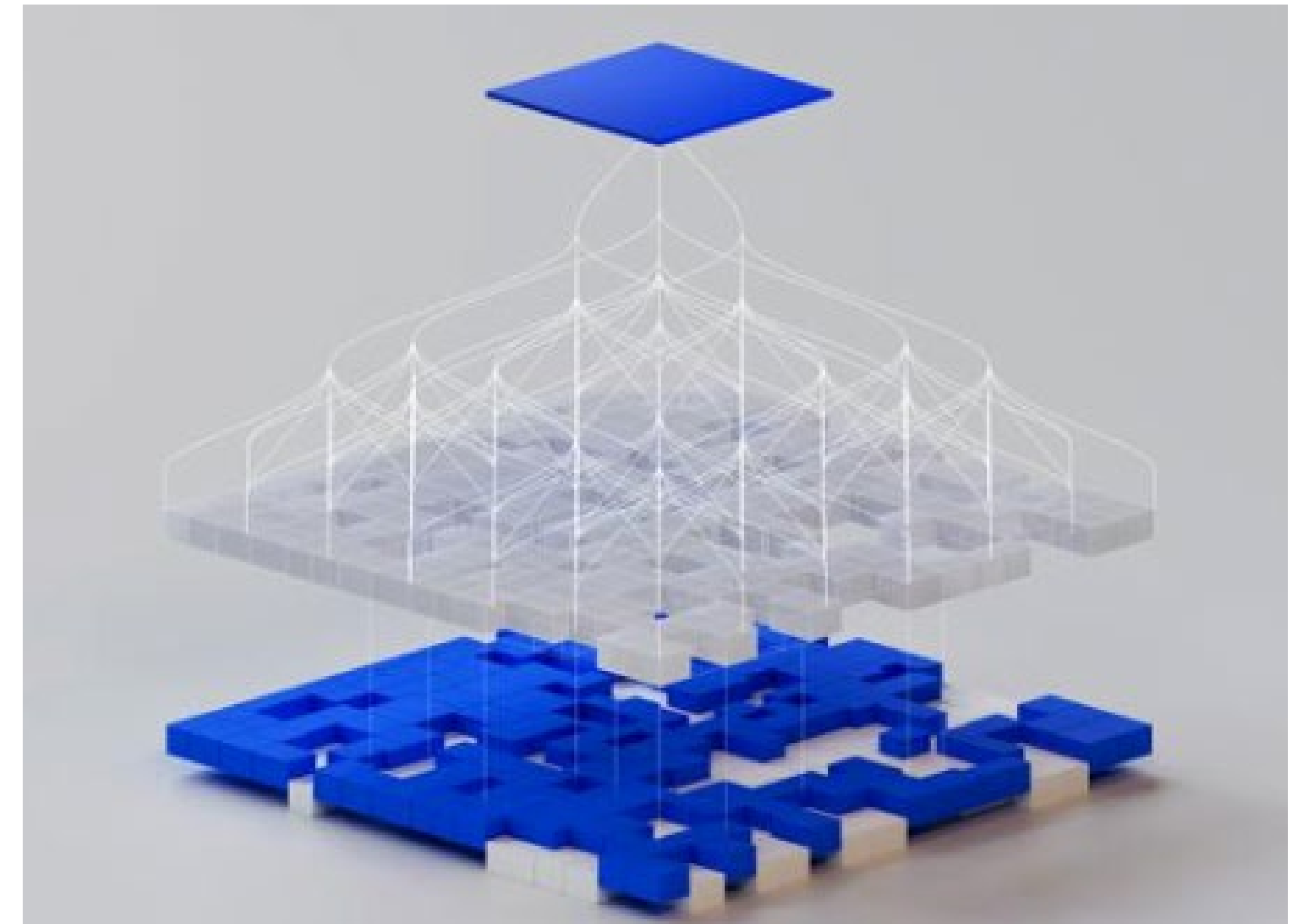
AI Intelligence (leveraging LLMs)  
built into Db2 for z/OS

# Db2 SQL Data Insights – An industry-first relational database capability

*Transparent, out-of-the-box, cost-effective and secure exploitation of in-database AI*

## Why is it useful:

- Enables novel data insights that can be extracted directly from Db2 tables
- Works on raw Db2 tables, native training, no need to move data out for labeling/processing or data science expertise
- Provides new, AI-enhanced, semantic queries which capture hidden relationships using standard SQL interfaces
- Integrates neural network training and inferencing into existing database infrastructure securely near the data source



<https://www.forbes.com/sites/ericsiegel/2025/01/13/the-rise-of-large-database-models/>

Semantic AI functions  
encode complex  
semantic operations  
between relational  
values while hiding  
internal vector  
operations from users.

Semantic AI functions  
support both intra- and  
inter-column  
operations, irrespective  
of the SQL types

Db2 SQL Data Insights Built-in Semantic Scalar Functions

| Semantic Functions                      | Functional Classification      | Functional Description                                                                                                                                                                        | Function Names      |
|-----------------------------------------|--------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| Semantic similarity and dissimilarities | Entity Matching Recommendation | <ul style="list-style-type: none"><li>•Matching rows/entities based on overall meaning (similarity/dissimilarity)</li><li>•Suggest choices for incorrect or missing entities</li></ul>        | AI_SIMILARITY       |
| Semantic Clustering                     | Recommendation                 | <ul style="list-style-type: none"><li>•Find entities/rows based on relationships between attributes in a given set</li><li>•Example: Find animals similar to (lion, tiger, panther)</li></ul> | AI_SEMANTIC_CLUSTER |
| Reasoning Analogy                       | Recommendation                 | <ul style="list-style-type: none"><li>•Find entities/rows based on relationships between attributes</li><li>•Example: Newspaper:Press :: Cloth:?</li></ul>                                    | AI_ANALOGY          |
| Pattern Queries                         | Detect global patterns         | <ul style="list-style-type: none"><li>•Identify entities that are outliers or with most common behavior</li></ul>                                                                             | AI_COMMONALITY      |

# Db2 SQL Data Insights can expose that hidden information

Input record

| Customer ID | Gender | Senior Citizen | Dependents | Tenure | Phone Service | Multiple Lines | Internet service | Contract       | Paperless billing | Payment method | Charges | Churn |
|-------------|--------|----------------|------------|--------|---------------|----------------|------------------|----------------|-------------------|----------------|---------|-------|
| 001         | Male   | 0              | NO         | 2      | YES           | NO             | DSL              | Month-to-Month | YES               | Mailed Check   | 120.22  | YES   |

```
SELECT * ,
AI_SIMILARITY (Customer_ID, '001' )
AS SimScore
FROM TABLE WHERE ....
```

Ranked Similarity Results (Most to least similar)

| SimScore | Customer ID | Gender | Senior Citizen | Dependents | Tenure | Phone Service | Multiple Lines | Internet service | Contract       | Paperless billing | Payment method | Charges | Churn |
|----------|-------------|--------|----------------|------------|--------|---------------|----------------|------------------|----------------|-------------------|----------------|---------|-------|
| 0.80     | 004         | Male   | 0              | NO         | 1      | YES           | NO             | DSL              | Month-to-Month | YES               | Mailed Check   | 48.55   | YES   |
| 0.75     | 002         | Male   | 0              | NO         | 7      | NO            | NO             | DSL              | Month-to-Month | YES               | Mailed Check   | 51.00   | YES   |
| 0.70     | 006         | Male   | 0              | NO         | 3      | NO            | NO             | DSL              | Month-to-Month | YES               | Mailed Check   | 49.80   | YES   |
| 0.55     | 003         | Female | 0              | NO         | 4      | NO            | NO             | DSL              | Month-to-Month | YES               | Mailed Check   | 60.40   | YES   |
| 0.35     | 005         | Female | 1              | NO         | 1      | NO            | NO             | DSL              | Month-to-Month | YES               | Credit Card    | 55.10   | YES   |

The logo for die Mobiliar, featuring the text "die Mobiliar" in white lowercase letters on a red rectangular background.

# Swiss Mobiliar

"In the insurance sector, customers look for the most personalized service possible to get the assurance they need to feel protected. As a leading provider in Switzerland, our goal is to use the latest technologies available in order to deliver on this promise to our customers. **We worked together with IBM to apply the AI capabilities on their trusted z16 systems in order to process insurance offer recommendations faster and more accurately. By unlocking hidden data patterns with NLP-based AI functions in near-real time while ensuring privacy and security, we saw an 94% accuracy in prediction results. These very promising results have motivated us to integrate this technology into our underwriting processes in the near future.**"

- Thomas Baumann, La Mobilière's IT architect

## AI-powered insurance product recommendation

### Business Challenge:

The client wants to predict the "conversion probability" of an insurance offer to be converted into a contract immediately after the offer is created either in-person by a sales representative or online by the end-user.

### Business Impact:

Leveraging SQL Data Insights, achieved near real time AI inferencing with 94% prediction accuracy. They can now unlock hidden patterns in their Db2 data with NLP-based AI functions, query and training enhancements along with outlier detection to reduce cost and improve their prediction results.

### Technologies leveraged:



SQL Data Insights  
on Db2 z/OS



# Drive revenue growth through product recommendations with Transactional AI

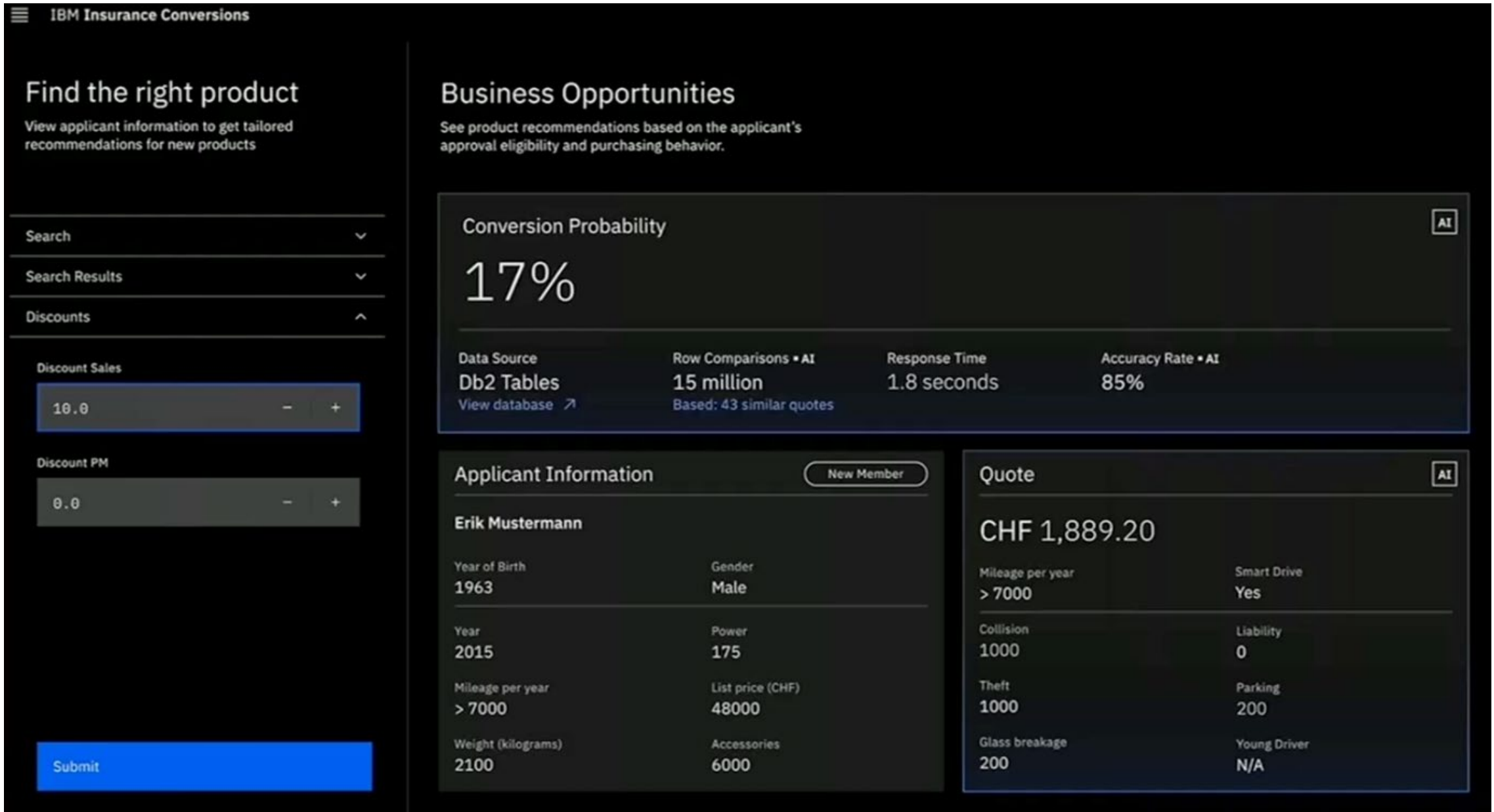
## Business Scenario:

An insurance client wants to predict the success probability of an insurance offer to be converted into a contract immediately after the offer is created either in-person by a sales representative or online by the end-user. The client wants to minimize data movement and avoid costly AI training off-platform

## Business Impact:

- 4 hours from ideation to production
- Increased quote to conversion rate
- 10% increase in sales
- 2 second response time

Client leveraged the built in AI capabilities of Db2 v13 SQL Data Insights to run semantic queries such as similarity and clustering with simple SQL statements to compare in-progress quotes with similar historical quotes give the user a conversion probability score.



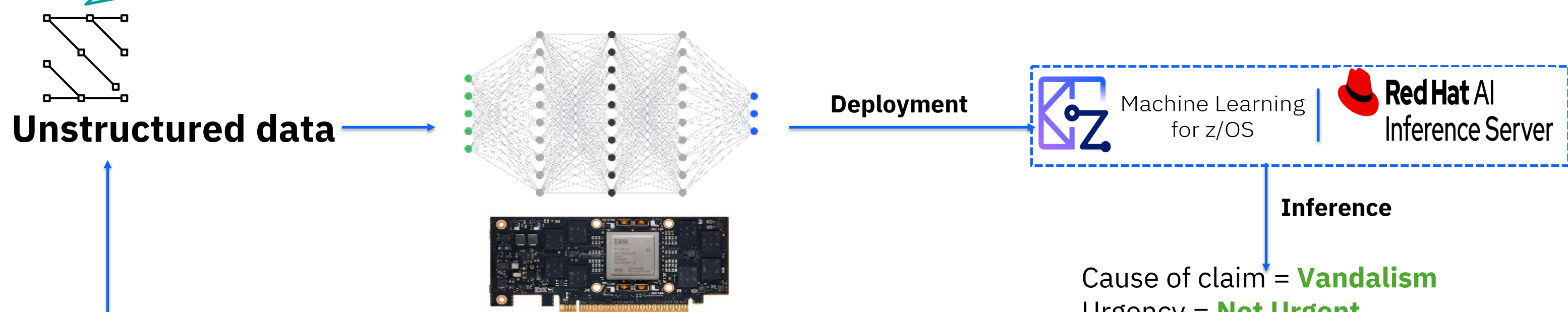
# Multi-Model AI - Home Insurance Claim Fraud Use Case

### Claim Narrative

Two days ago, there was **malicious damage**. Here is my initial summary of items for you to cover: the dormers, the outside walls, and the decorative benches. Probably **not very urgent**. My losses were \$3,920

### Claims Image Data





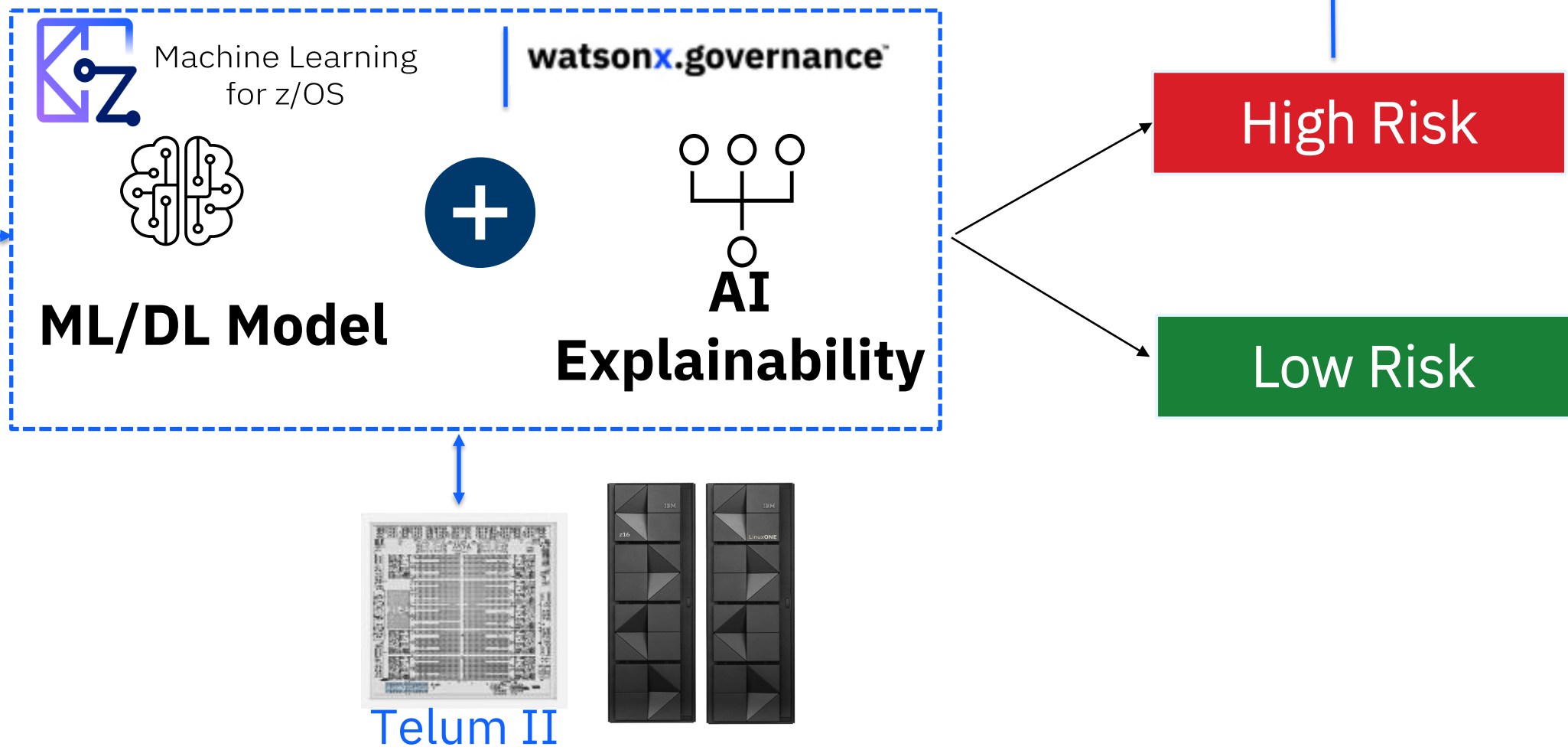
IBM Synthetic Datasets Product

Structured data

| Policy ID | Date   | Total Paid | ...     | Claim Amount |
|-----------|--------|------------|---------|--------------|
| 6AA823A34 | 1/2/23 | \$2500     | <value> | \$1500       |

| Cause of Claim | Urgency    | Damage Severity |
|----------------|------------|-----------------|
| Vandalism      | Not Urgent | Low             |

New features generated in real-time integrate with structured data



# Agentic AI Applications for Z

Drive productivity, streamline operations, and reduce costs



## Business Operations

Automate core business workflows with domain-aware agents to boost efficiency, accuracy & throughput.

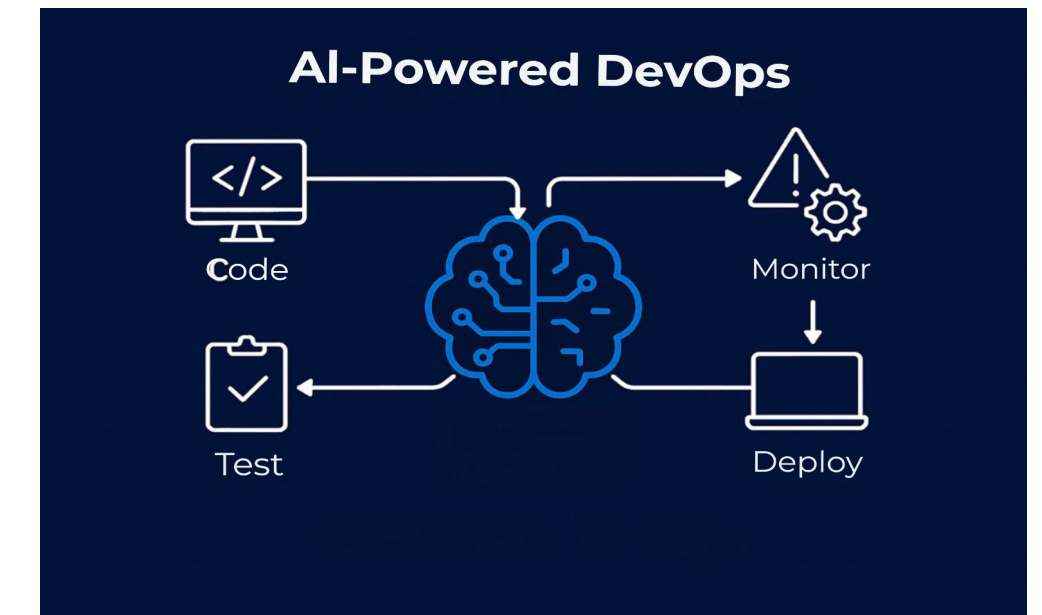
- Apply AI agents automate high-value operations in industry such as insurance, banking, retail.
- Automate tasks like fraud response, claims adjudication, and payment processing.



## ITOps Intelligent Infrastructure

Optimize and automate IT operations with intelligent agents that enhance reliability, performance, and resource efficiency.

- AI-powered automation for install, configuration, system management.
- Intelligent resource optimization and workload placement.
- Operational simplification with trusted, efficient AI.



## AI-Powered DevOps

Accelerate software delivery by automating key development and testing processes with AI agents.

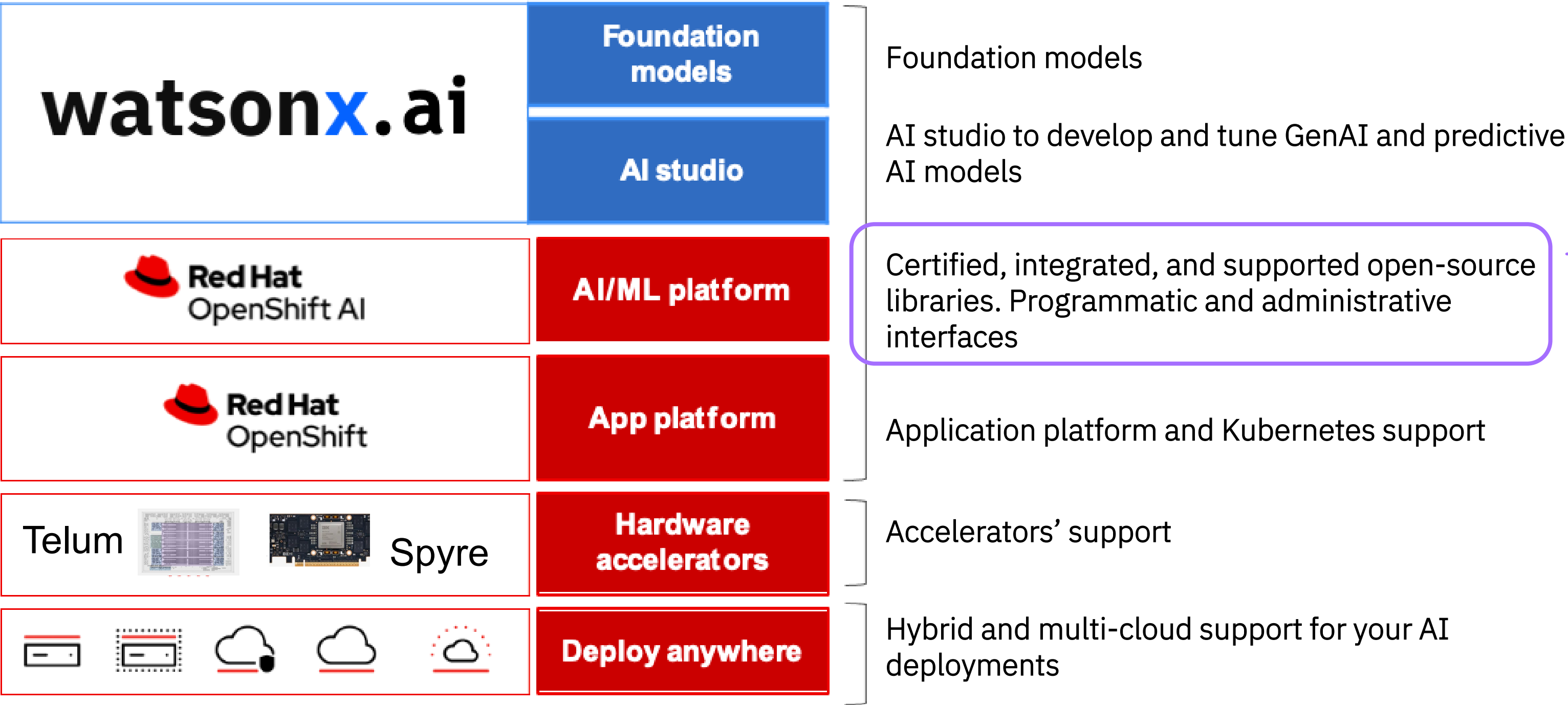
- High-quality releases with AI generated code and continuous test coverage.
- Improved developer productivity and throughput.
- Shorter time-to-value by streamlining complex engineering workflows end-to-end.

# Agentic AI for IT operations

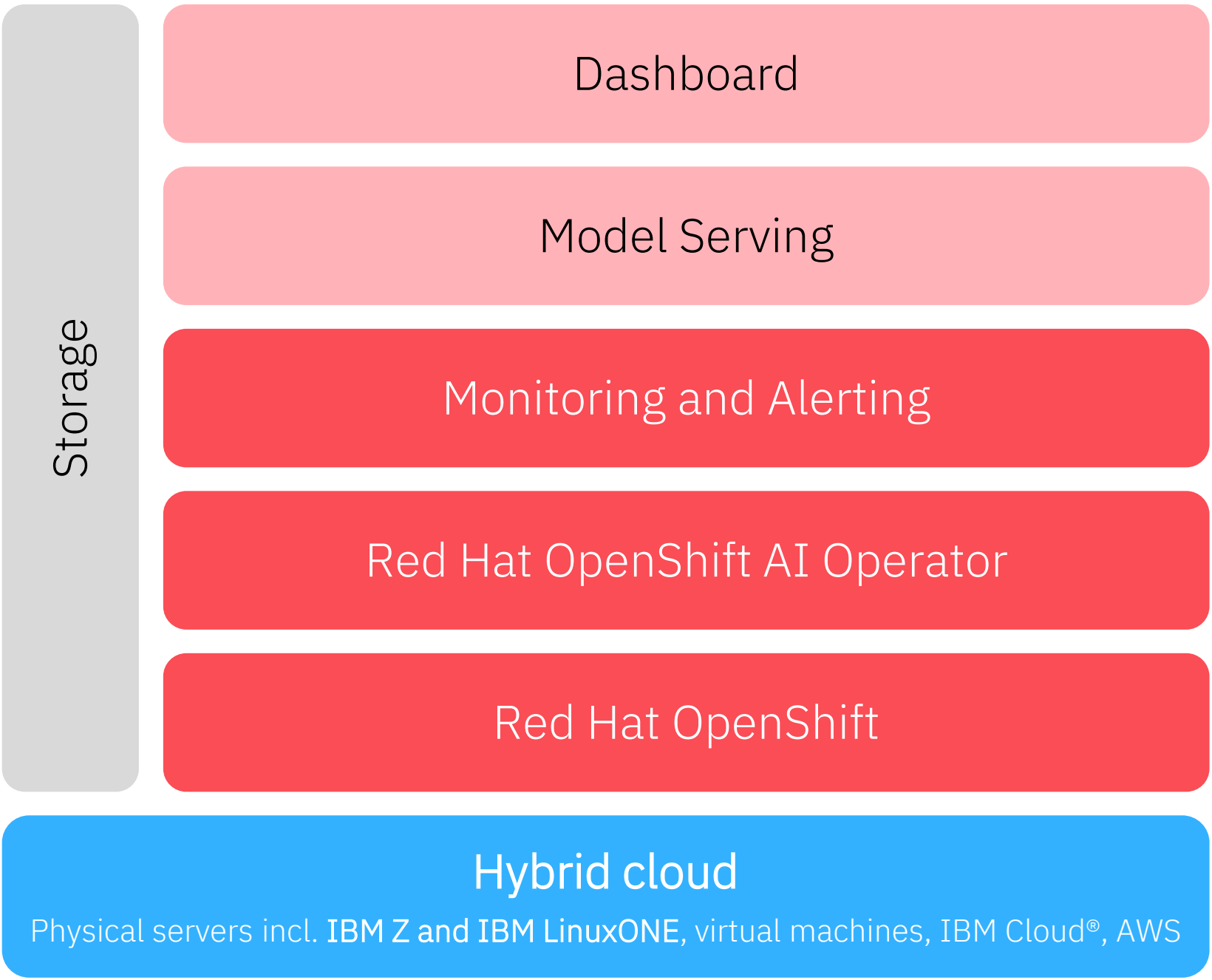
# IBM Z & LinuxONE GenAI Stack - Red Hat OpenShift AI

*Build, serve, and monitor your AI model, and manage data science pipelines*

## End-to-end Stack



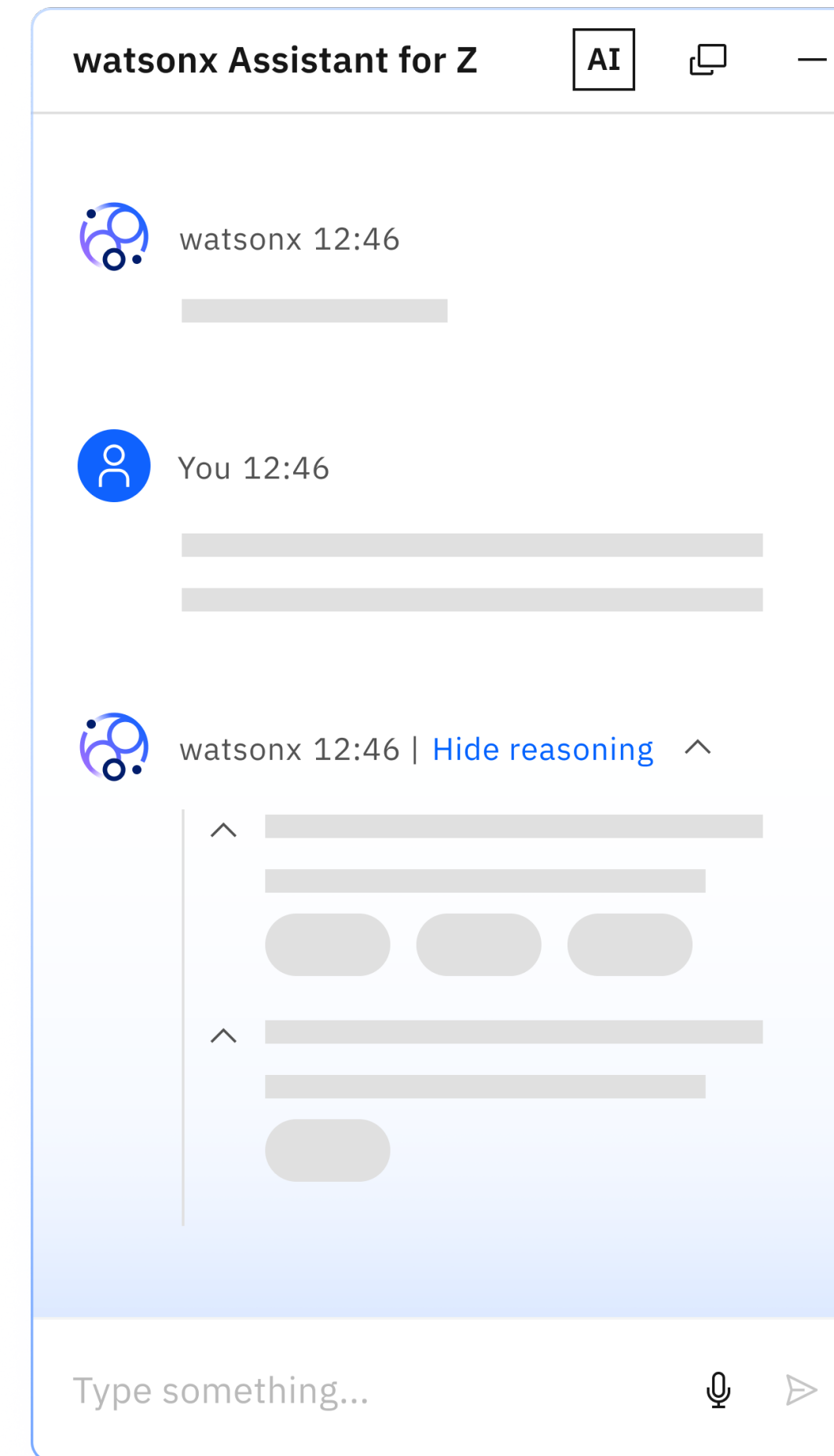
## Features and capabilities



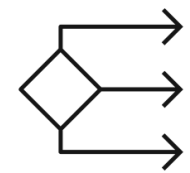
# IBM watsonx Assistant for Z

Simplifying and transforming the IBM Z experience

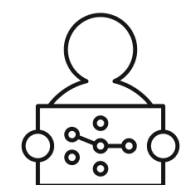
- Conversational AI for IBM Z
- AI-Driven Automation
- Foundational Agentic Framework



Generative AI  
responses for IBM Z



Integrated  
Automation



Agentic AI

Benefits:

Reduce  
learning curve

Increase  
productivity

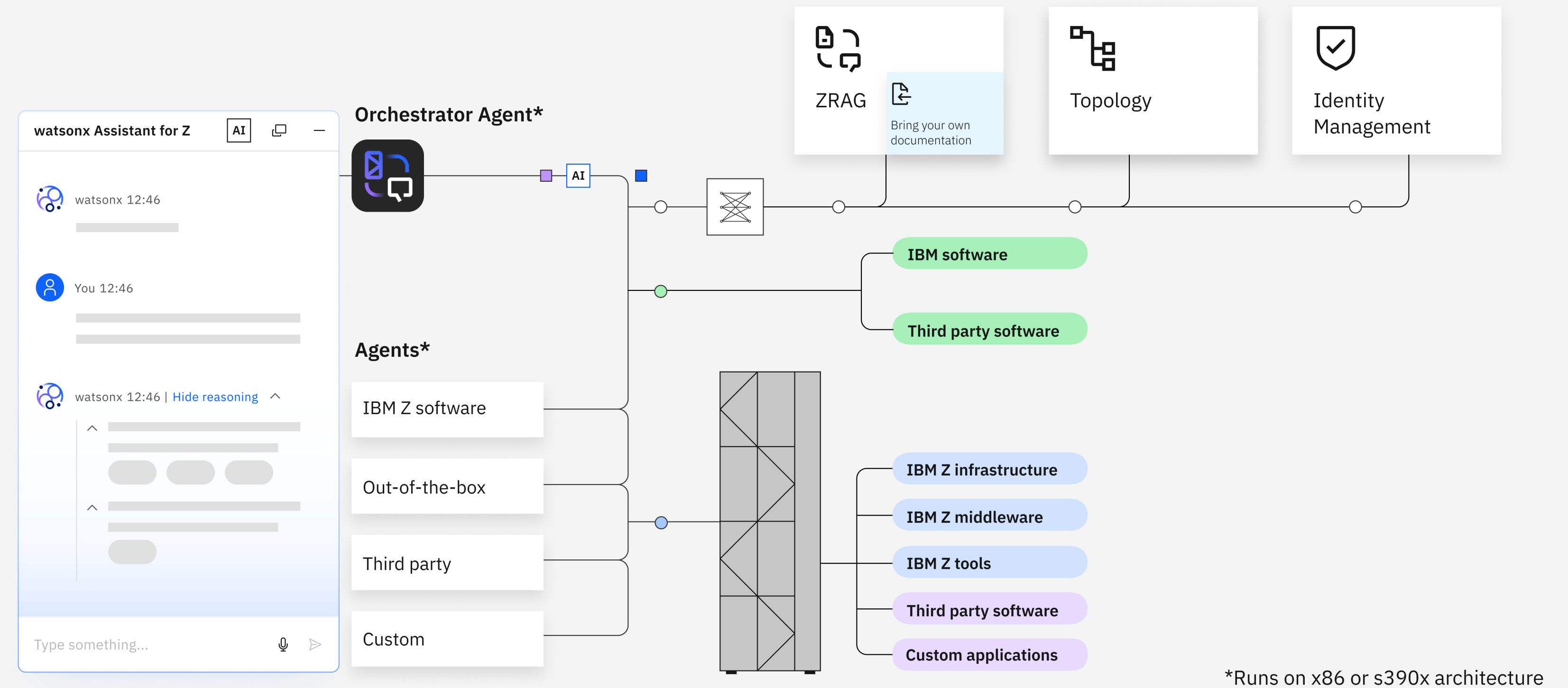
Increase  
efficiency

Reduce  
errors

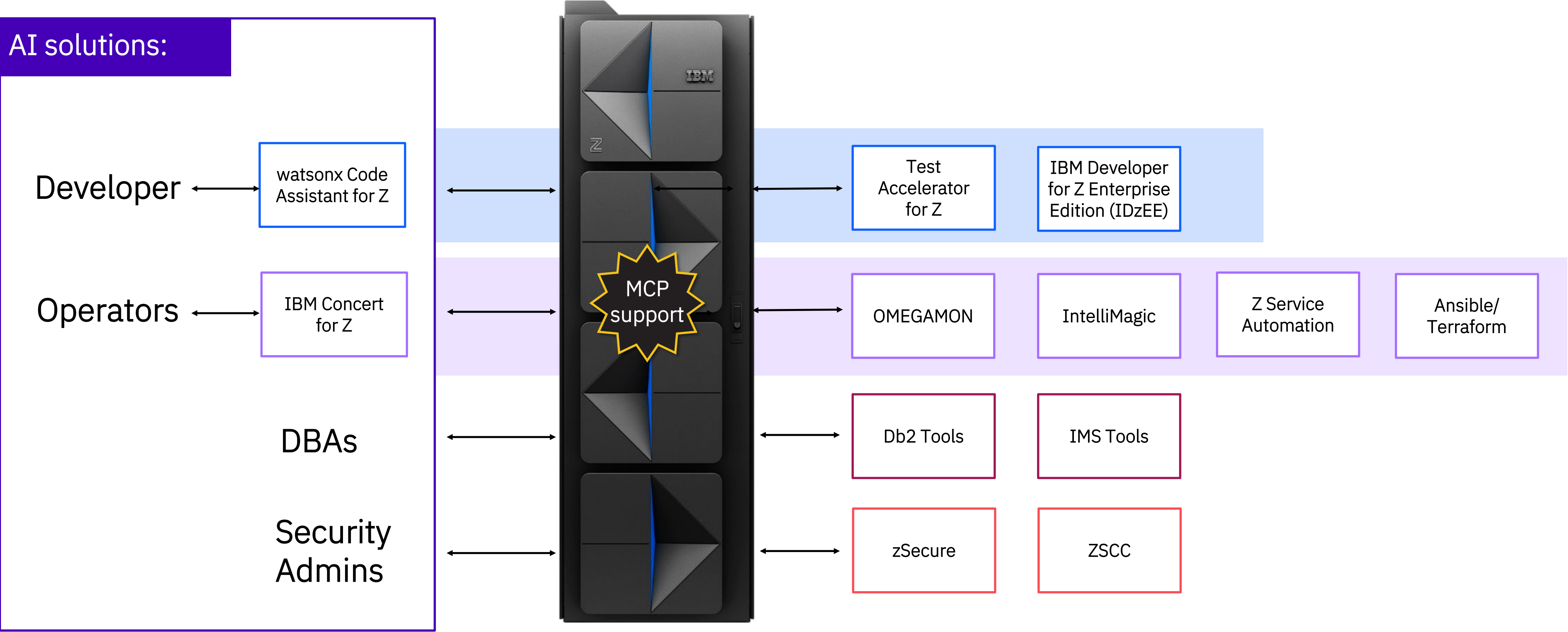
# IBM watsonx Assistant for Z

## Agentic vision

- Agents working together on tasks in development, operations, performance, security, compliance, and other areas.
- Allow for integrations with 3<sup>rd</sup> party provided agents.
- Easily navigate and discover agents and tools in the catalog.
- With managed enterprise security



# Persona-based AI solutions



# AI on z/OS

## Predictive AI

Analyzes historical and real-time [operational data](#) to forecast patterns, anticipate events and detect anomalies.

Enabled through the [AI Framework for IBM z/OS](#) that provides the infrastructure for AI integration.

Leverages [zIIPs](#), [Telum I](#) and [Telum II](#).



Predicts

### Predict WLM Initiator Demand

Use historical and real-time data to forecast optimal WLM initiator number for batch workloads.

### Auto-Tune LE Job Parameters

Detect and adjust LE job parameters to boost performance and resource efficiency.

## Agentic AI examples for future z/OS use cases

[Acts autonomously](#) to understand, plan, and execute a request.

Uses [Generative AI](#) to determine which tools to use, analyze data, and interact with the user.

Leverages [IBM Spyre Accelerator](#).



Does

### Problem determination

Assist in problem determination; automatically gather key diagnostic data for error analysis.

### Analyze WLM Policy and provide guidance

Highlight deviations from WLM Policy definition best practices.

### Gather live system information and provide feedback

Gather system information to ease monitoring and answer Q's.

**Predictive AI** enabled by AI Framework for z/OS [anticipates issues](#).

**Agents** embedded in z/OS will [act](#).

**Together** – enable [intelligent system management](#).

# z/OS Agentic Strategy

## Deliver MCP Servers to Enable the Agentic Ecosystem

Provide native MCP servers on z/OS to enable AI powered tooling to access live system services.

- Expose Commands, APIs, and Tools via MCP

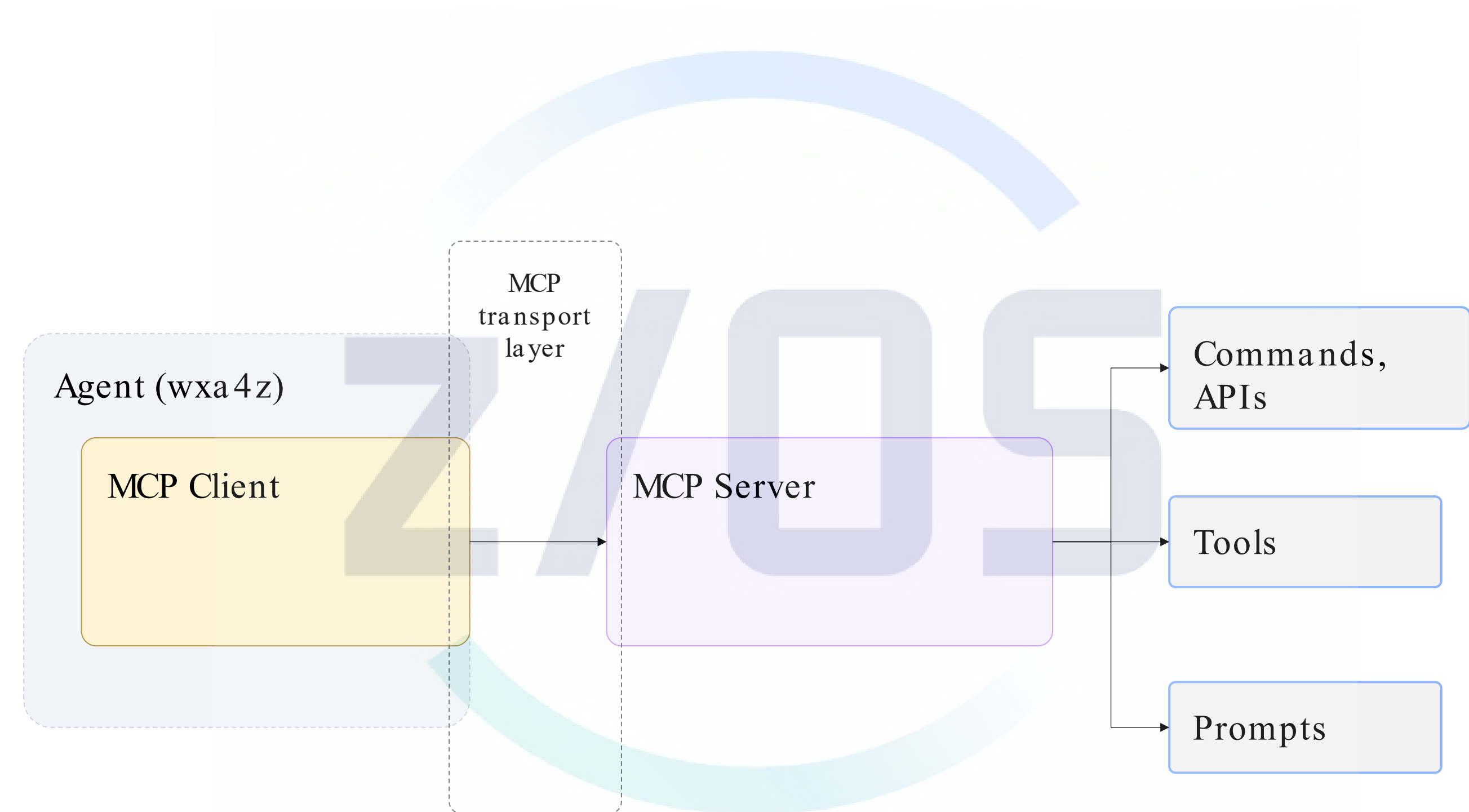
## Accelerate Platform Simplification for New-to-Z Users

Deliver agents focused on use cases that streamline sysprog tasks and enabling faster adoption of z/OS.

- Simplify installation and configuration processes for clients
- Optimize system resource usage for peak performance

## Aligned with watsonx Assistant for Z

Deliver foundational z/OS APIs that strengthen Watsonx Assistant and align with IBM Z platform strategy.



# Agentic AI for business operations

# Agentic AI Business Value Examples

**Gartner** predicts, agentic AI will resolve 80% of common customer service issues without human intervention by 2029.

## Customer Service

### AI agent Example

- When a customer calls about a delayed shipment, while traditional chatbot may provide tracking or escalate to human, an agentic AI system could:
  - Access & analyze the shipping data in real time to identify root cause of delay instantly without a human intervention.
  - Provide immediate resolution options, such as replacement or partial refunds<sup>34</sup>
  - Execute the selected action end-to-end, including record updates and notify customer.

## Financial Services

### Vision

- Scale personalized wealth management without human oversight, improving access & efficiency.
- Automates full financial processes – such as portfolio management & investment operations.
- AI agents analyze market data, assess risk, and execute investment decisions (buy, sell, rebalance) in real time.
- Dynamically adapt strategies to market conditions and individual financial goals and tolerable risk rate.

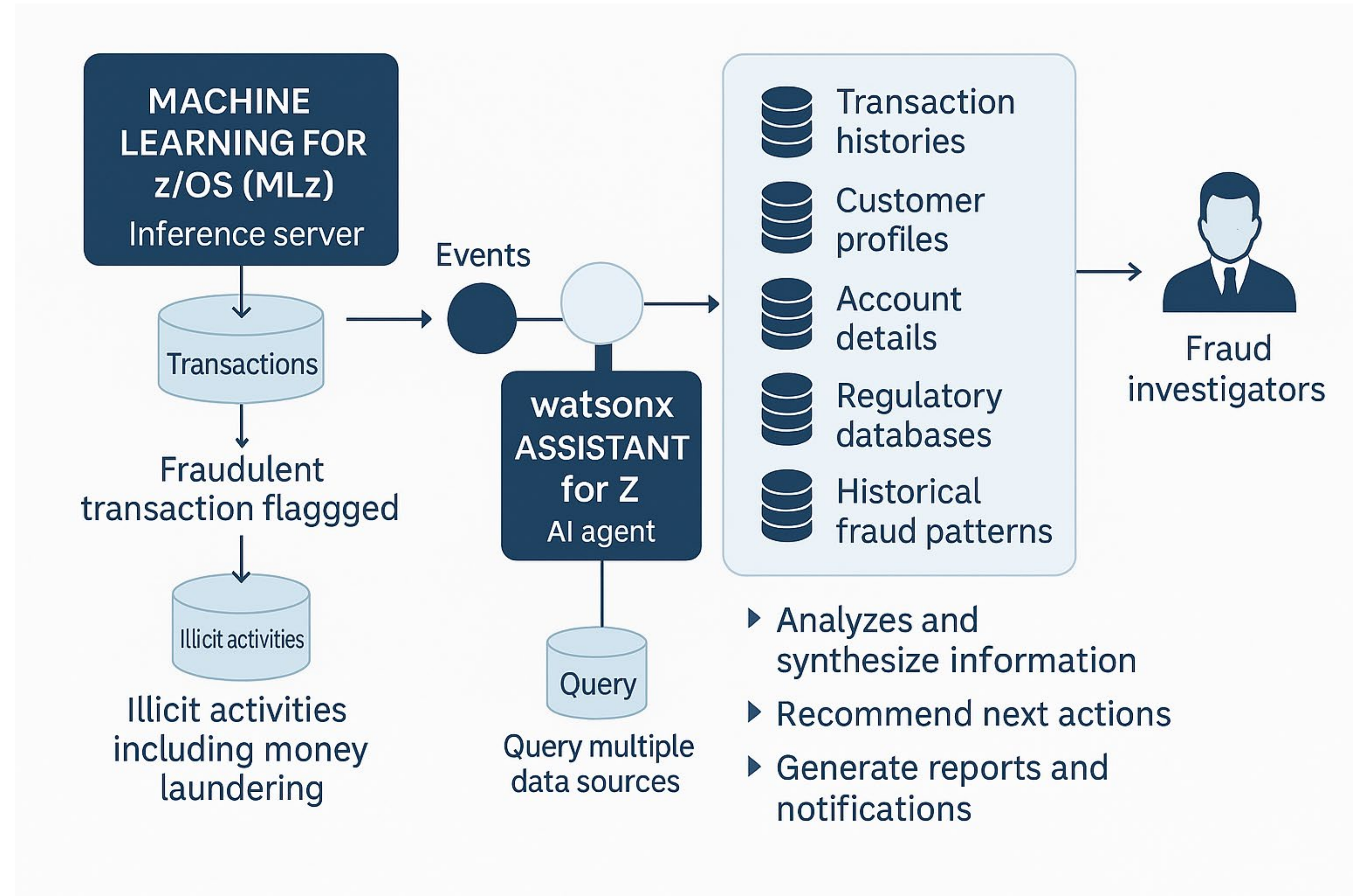
## Healthcare

### Streamlined Office tasks

- Automates routine administrative tasks – scheduling appointments, sending reminders, and record management.
- AI agents for suspended claims processing
- Frees staff to focus on high-value patient care
- Supports patients and clinicians with timely guidance to enhance care coordination and outcomes.

# Intelligent Fraud & AML Automation on IBM Z Powered by Transactional AI & Agentic AI

## Event Driven Agents for Fraud Detection & AML Response on IBM Z MLz + WXA4Z

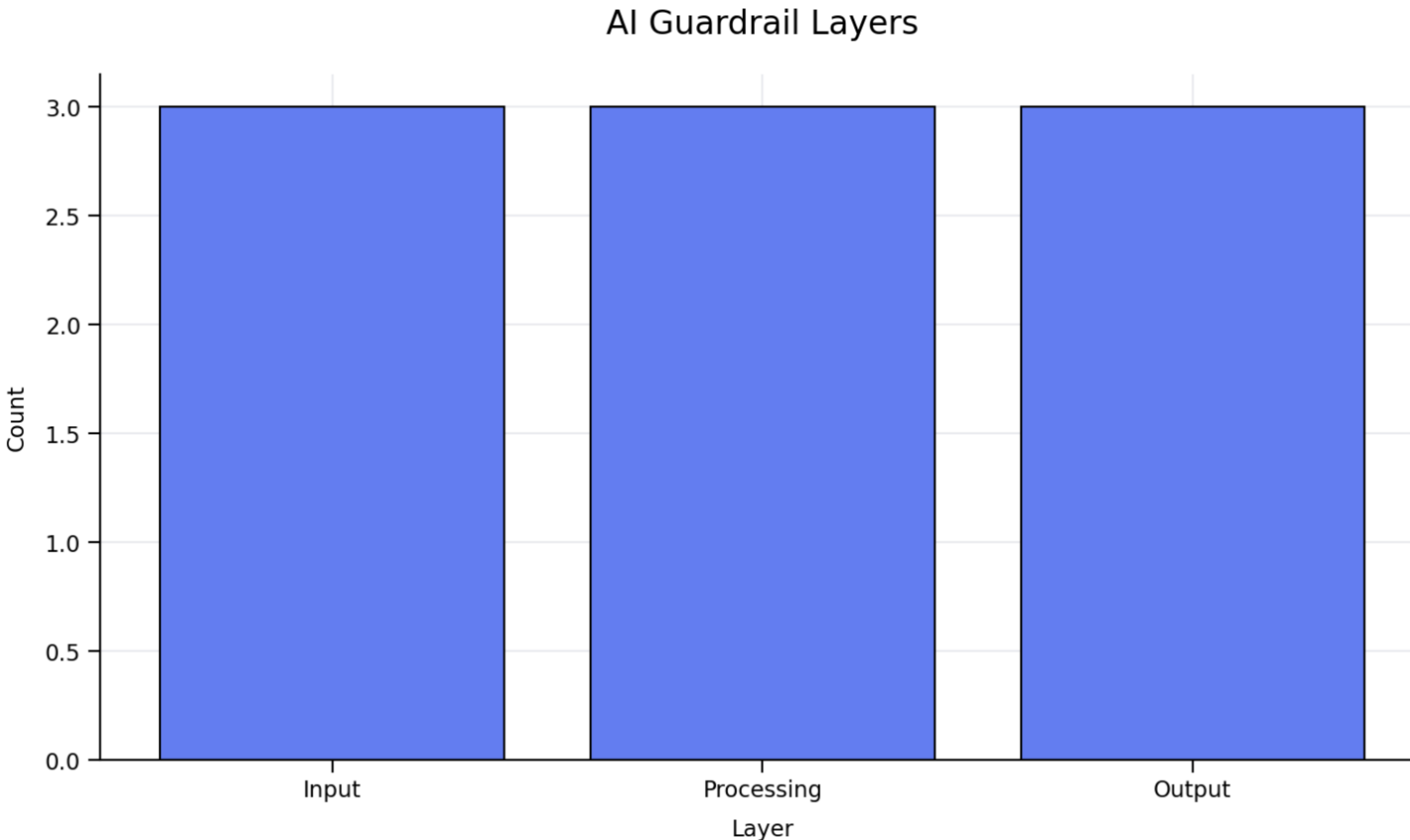


Transformation roadmap:  
AI on IBM Z

Drive more value from your applications and data and  
fundamentally transform the experience on IBM Z

| 2025                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 2026                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 2027                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Deliver foundational agentic AI stack, with common services and AI accelerators</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | <i>Enable AI-first operating model with agentic workflows and platform simplification</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | <i>Maximize full-stack AI value with goal-driven, autonomous agents.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <div>Applications and Data<ul style="list-style-type: none"><li>In-transaction predictive AI augments core workloads with real-time scoring.</li><li>AI insights on structured data to discover hidden patterns</li></ul></div> <div>Platform Simplification<ul style="list-style-type: none"><li>AI agents access real-time operational data, domain-specific expertise and system-wide information to synthesize accurate, context-aware system insights</li><li>AI-infused operating system, including early threat detection and intelligent workload management</li></ul></div> <div>Infrastructure<ul style="list-style-type: none"><li>Telum II processor for real-time, low latency, high throughput AI inference</li><li>Spyre accelerator for agentic and gen-AI, optimized for small-to-medium models</li><li>Hybrid AI and multi-model support</li></ul></div> | <div><ul style="list-style-type: none"><li>Transactions trigger agentic retrieval and decision workflows, <b>reducing effort up to 50%</b></li><li>AI insights expand to support unstructured data, <b>unlocking greater value</b></li></ul></div> <div><ul style="list-style-type: none"><li>AI agents exploit infrastructure and configuration automation, reducing <b>time-to-proficiency for system management from years to months</b></li><li>AI discovery makes relationships between systems, subsystems, and configurations visible and navigable, <b>cutting analysis from hours to minutes</b></li></ul></div> <div><ul style="list-style-type: none"><li>Expand on-platform model support, <b>enabling richer GenAI workflows on Spyre</b></li></ul></div> | <div><ul style="list-style-type: none"><li>Pre-built domain agents provide agentic business workflows across Z Data and hybrid systems, <b>shortening time-to-value from months to weeks.</b></li></ul></div> <div><ul style="list-style-type: none"><li>AI agents continuously assess workload behavior and resource utilization, autonomously recommending or executing tuning actions, <b>dramatically reducing operational effort to stay within SLOs, from months to hours</b></li></ul></div> <div><ul style="list-style-type: none"><li>Next-Gen AI Accelerator <b>unlocks new high-fidelity agentic AI use cases</b></li></ul></div> |

# AI Guardrails & Safety Overview



## Types of Guardrails Used in Production:

### 1. Input Guardrails

Prevent harmful or malformed prompts from ever reaching the model.

### 2. Processing Guardrails

Control what data, tools, and internal reasoning steps the model can access.

### 3. Output Guardrails

Validate model responses before they reach a user or downstream system.

What are your biggest concerns or **blockers** about using AI, and how can we **help mitigate** them?

What **AI and agentic AI** use cases are you pursuing today both in **business domains and IT Operations**?

What **governance or audit capabilities** do you need for AI agents operating on core systems?

# Start Your AI Journey

## AI Discovery Workshop



# AI Business Value Discovery Workshop on IBM Z & LinuxONE

Making transaction processing smarter and more valuable



By applying AI decisioning directly within your high-volume, mission critical workloads and applications already running your business, this free-of-charge, 1-2 day workshop will identify and demonstrate high-value that matters to your business: revenue growth, operating margin, customer experience, risk reduction, and more.

Working with IBM experts, selected opportunities are rapidly validated without disrupting operations through a focused Proof of Concept, demonstrating measurable improvements in decision velocity, consistency, and business outcomes on your core transactional systems, IBM Z<sup>®</sup> and LinuxONE<sup>™</sup>.

Contact us at [purvi@us.ibm.com](mailto:purvi@us.ibm.com), [swarren@us.ibm.com](mailto:swarren@us.ibm.com)



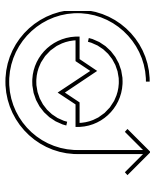
## Required Client Personas for Workshop

- Line of Business and/or Business Analyst
- Chief Data Officer
- AI Leader | Data Scientist
- Application Architect
- Infrastructure Architect



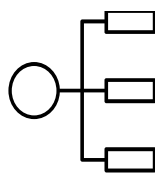
### Identify and Assess Business Value

Pinpoint high-impact AI opportunities within your mission-critical transaction flows to drive growth, reduce risk, and unlock deeper operational and market insights from your core systems.



### Observe Transactional AI in Action

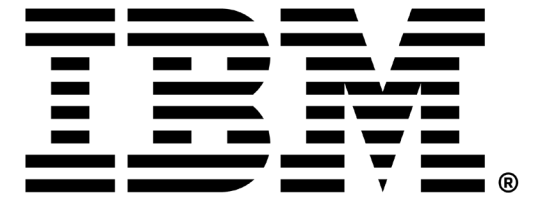
Witness intelligence elevate at the transaction level as AI infusion transforms routine business decisions into valuable insights that unlock competitive differentiation and untapped business potential.



### Design Your Proof of Concept

Co-create your Transactional AI proof of concept with IBM and your key leaders to move beyond ideas, proving value early and creating momentum toward confidently deploying intelligence into everyday business execution.

# Experience more with IBM



## Visit us at the IBM Booth #113

After a full day of technical sessions, take a break with us!

Connect with our experts, snap a photo with the z17 Plexi or the latest Telum II, and get an up-close look at our Spyre Accelerator.

Come back each day for fresh topics and demos at our expert stations.

## Think 2026

Join 5000+ senior business and technology leaders who are seizing the AI revolution to unlock unprecedented growth and productivity at **Think 2026**.

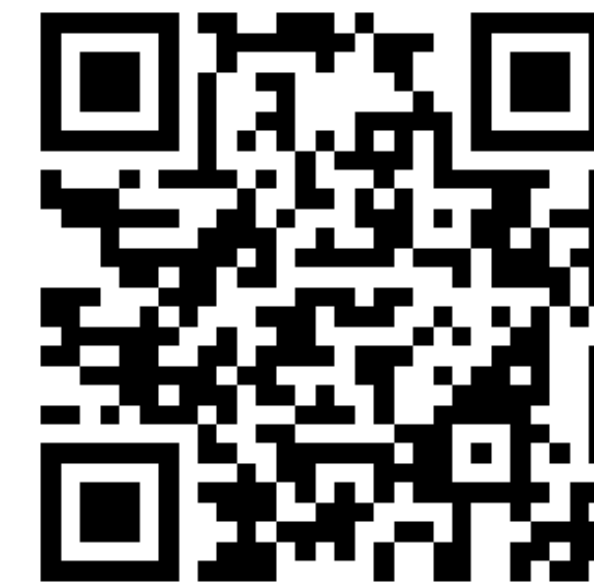
Find out more information using the QR code below.



## IBM Digital Asset Haven

IBM Digital Asset Haven is the operational backbone for financial institutions and regulated enterprises entering the digital asset economy.

Find out more information using the QR code below.



# Additional AI on IBM Z AI Practitioner Resources

## Key Resources

### [AI on IBM Z Seismic Page](#)

One stop shop for all AI on IBM Z sales kits, enablement information and others...

### [AI on Z Github 101](#)

Repository of information on AI tooling and frameworks available in the IBM Z and IBM® LinuxONE ecosystem

### [Journey to AI Content Solution](#)

Hands on guides to get started with top use cases

### [Solution Templates](#)

Hands on guides to get started with top use cases

### [IBM Z and IBM® LinuxONE Container Registry](#)

Download open-source frameworks and tooling vetted for security from this registry

## Redbook Papers

- [What AI Can Do for You: Use Cases for AI on IBM Z](#)
- [Optimized Inferencing and Integration with AI on IBM zSystems: Introduction, Methodology, and Use Cases](#)
- [Enriching Linux on IBM Z Workloads with AI](#)
- [Finding an On-Ramp to Your AI on IBM Z Journey](#)
- [IBM Synthetic Data Sets](#)
- [Turning Data into Insight with Machine Learning for IBM z/OS](#)
- [Securely Leverage Open-Source Software with Python AI Toolkit for IBM z/OS](#)
- [IBM Cloud Pak for Data on IBM Z](#)

# AI on Z and LinuxONE Sessions at SHARE



## Generative AI, Assistants and Agents

**watsonx Assistant for Z - for Db2 Systems Programmers and DBAs**  
Salon 19  
Tues, Feb 24, 10:30am - 11:30am  
Speaker: Maryela Weihrauch

**IBM Concert for Z, an AI-Powered Mainframe Resilience Platform: Solutions vs. Alerts**  
Salon 22  
Wed, Feb 25, 9:15am - 10:15am  
Speakers: Domenico D’Alterio, Fabricio Miatto

**Using IBM AI Generative Tooling to Assist with Code Development**  
Salon 22  
Thurs, Feb 26, 10:30am - 11:30am  
Speakers: Ben Hicks, Mary Julian

**MQ Meets Agentic AI: Intelligent Workflows on z/OS**  
Salon 19  
Wed, Feb 25, 3:45pm - 4:45pm  
Speaker: Toby Keegan

**An Introduction to Leveraging AI to Increase Productivity in CICS**  
Salon 17  
Wed, Feb 25, 9:15am - 10:15am  
Speakers: Kye Maloy, James Davies

**CICS 6.3 is AI Ready!**  
Salon 17  
Mon, Feb 23, 3:45pm - 4:45pm  
Speaker: Kye Maloy

**Data Center Automation - Z System Automation and Agentic AI**  
Salon 21  
Mon, Feb 23, 9:45am - 10:05am  
Speakers: Johannes Hausch, Ash Mahay

## AI on Z in General

**AI on IBM Z and LinuxONE Strategy for 2026 and Beyond!**  
Salon 22  
Mon, Feb 23 2:30pm - 3:30pm  
Speaker: Elpida Tzortzatos

**Unlocking Business Insights: Deep Dive Into AI on IBM Z and LinuxONE**  
Salon 22  
Tues, Feb 24 9:15am - 10:15am  
Speakers: Purvi Patel, Steve Warren

**Securing AI on Z: Addressing Emerging Threats and Building Trustworthy Systems**  
Salon 20  
Wed, Feb 25, 2:30pm - 3:30pm  
Speakers: Gregg Arquero, Elijah Swift

## AI-Infused z/OS

**AI Infused z/OS: Overview and Updates**  
Salon 14  
Thurs, Feb 26, 9:15am - 10:15am  
Speaker: Anastasiia Didkovska

**The 'Next-Generation z/OS Experience' Overview**  
Salon 13  
Mon, Feb 23, 2:30pm - 3:30pm  
Speaker: Roger Bales

## ML/AI Business Insights

**Accelerated Business Decisions with Machine Learning for z/OS**  
Salon 22  
Tues, Feb 24, 10:30am – 11:30am  
Speakers: Steve Warren, Purvi Patel

## Hands-on Labs

**BYOD Lab: Advanced AI Insights Leveraging Hugging Face Large Language Models as Part of a Multi-Model Approach on IBM Z and LinuxONE**  
STE Lab  
Tues, Feb 24, 3:45pm - 4:45pm  
Speakers: Steve Warren, Purvi Patel

**BYOD Lab: WXA4Z Agentic Hands-on Workshop**  
Salon 1  
Mon, Feb 23, 1:15pm - 2:15pm  
Speaker: Dan Snyder

**BYOD Lab: AI Enabled Proactive Monitoring to Get the Most From Your System With IBM Concert for Z**  
Salon 22  
Wed, Feb 25, 1:15pm - 2:15pm  
Speakers: Domenico D’Alterio, Fabricio Miatto

**BYOD Lab: Using the Spyre Accelerator (and Telum II) to Operationalize AI at Scale on IBM z17**  
STE Lab  
Tues, Feb 24, 10:30am - 11:30am  
Speaker: Artem Minin

# Your feedback is important!

## Submit a session evaluation for each session you attend:

[www.share.org/evaluation](http://www.share.org/evaluation)



