

PDUUG

Kraków, 28-29 May 2025

EMPOWERING SOLUTIONS:

The Evolving Role of AI in Problem Solving

Nigel Slinger - Sr Product Development Architect

Wenjie Zhu - Sr Product Development Architect

BMC

POLISH DB2 USERS GROUP

» What is Machine Learning?

- » [Machine learning](#) is an application of AI. It's the process of using mathematical models of data to help a computer learn without direct instruction. This enables a computer system to continue learning and improving on its own, based on experience.

ARTIFICIAL INTELLIGENCE

» What is artificial intelligence (AI)?

- » [Artificial intelligence](#) is the capability of a computer system to mimic human cognitive functions such as learning and problem-solving. Through AI, a computer system uses math and logic to simulate the reasoning that people use to learn from new information and make decisions.

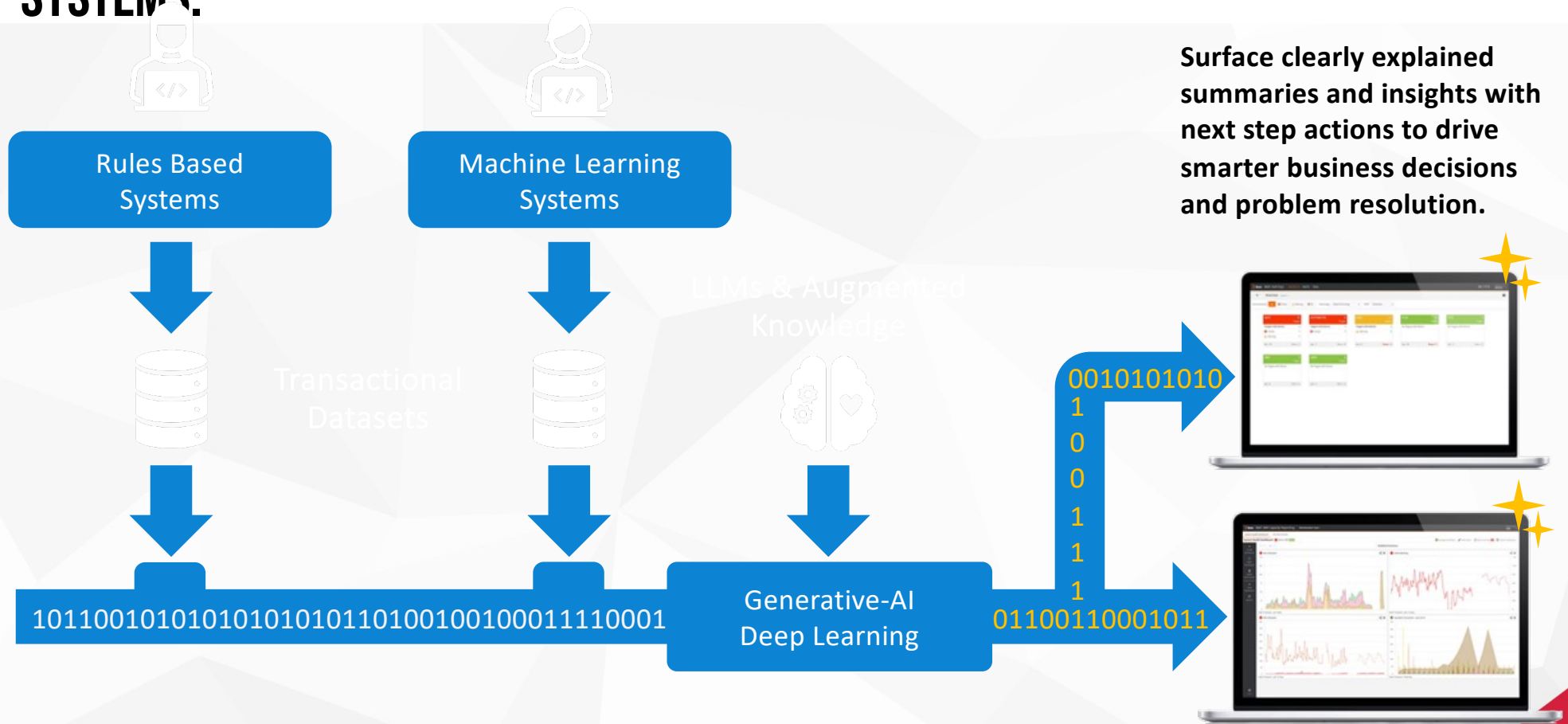
RULES BASED AI

- » A system designed to achieve artificial intelligence (AI) via a model solely based on predetermined rules is known as a *rule-based AI system*. The makeup of this simple system comprises a set of human-coded rules that result in pre-defined outcomes. These AI system models are defined by 'if-then' coding statements (i.e. if X performs Y, then Z is the result). Two important elements of rule-based AI models are "a set of rules" and "a set of facts" and by using these, developers can create a basic artificial intelligence model. These systems can be viewed as a more advanced form of [robotic process automation \(RPA\)](#).

UNSUPERVISED VS SUPERVISED LEARNING

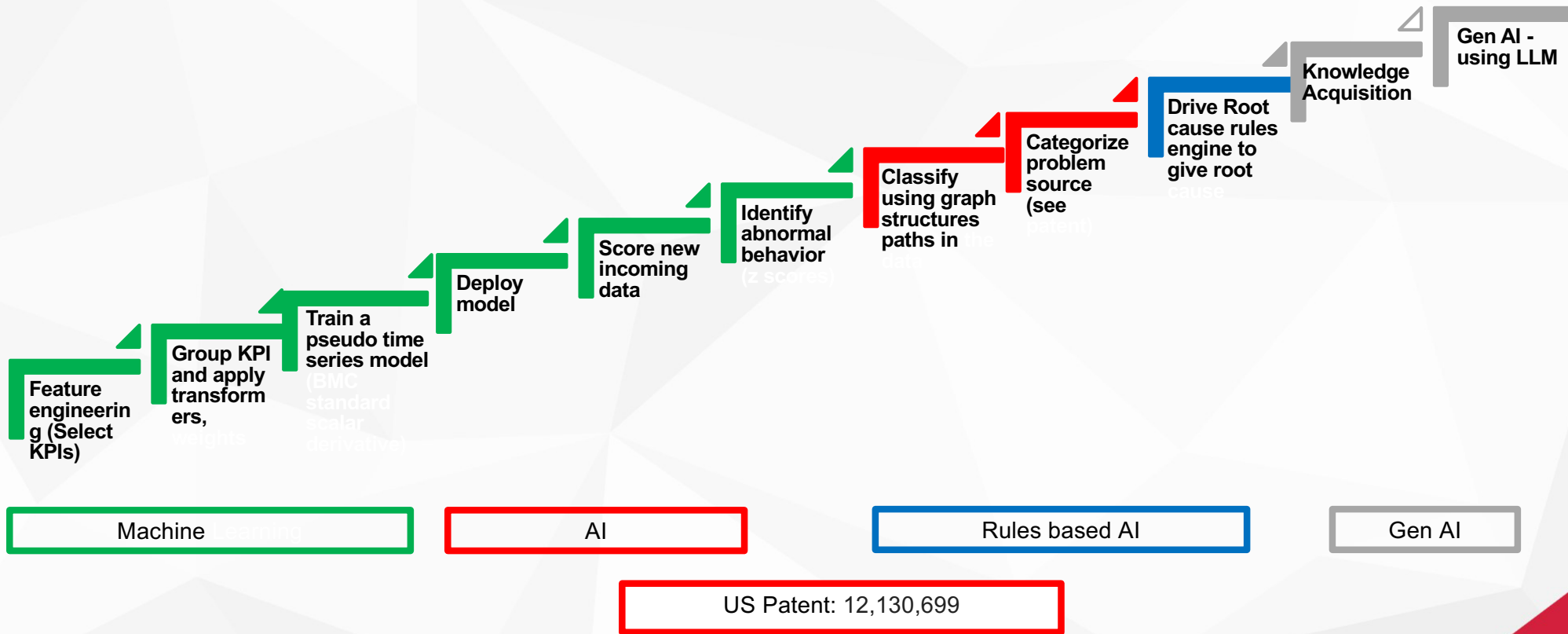
- » The main difference between supervised vs unsupervised learning is the need for labelled training data.
 - » **Supervised machine learning relies on labelled input and output training data,**
 - » **Unsupervised machine learning processes unlabeled or raw data**
- » No other way but to use unsupervised learning since we don't have 100000 z system problems to learn from. Unsupervised learning is by far the hardest implementation of Machine Learning to achieve with any degree of accuracy

HYBRID-AI: MACHINE LEARNING, DEEP LEARNING, AND RULE-BASED SYSTEMS.



A PROVEN MULTI LAYERED APPROACH

PDUG



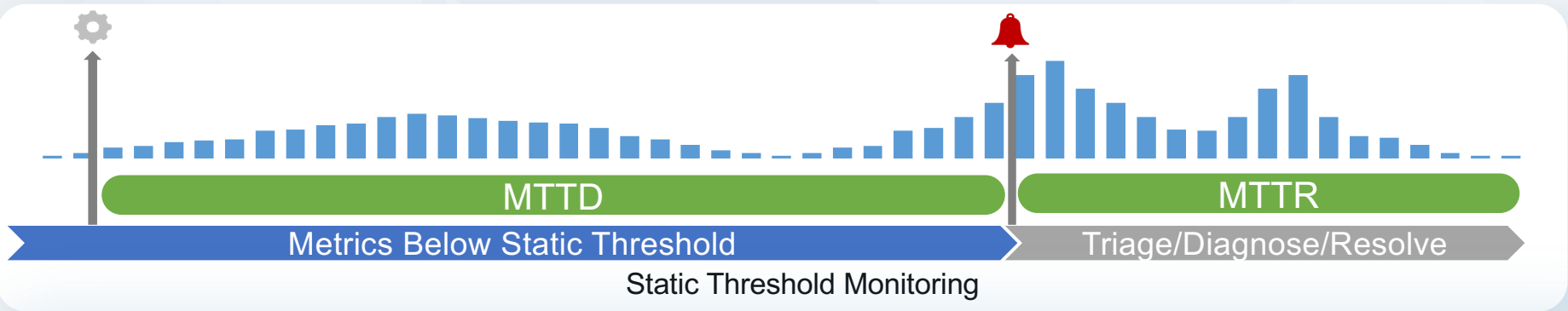
POLISH DB2 USERS GROUP

WHAT PEOPLE WANT FROM ML/AI BASED MONITORS

WHAT PEOPLE WANT FROM ML/AI

- » Lower MTTD (Mean time to detection)
- » Lower MTTR (Mean time to resolution)
- » Ease of use
- » Low cost
- » Easy button
 - » This is the biggest requested item
- » ML/AI should be explainable
- » Assumption : Something is always wrong somewhere, so False Positive suppression is a vital piece of the puzzle

GOAL: MINIMIZE DISRUPTIONS TO BUSINESS APPLICATIONS **PDUG**



Lower MTTD & MTTR

MINIMIZE MTTD – A PRACTICAL APPROACH



- Custom models
 - From historical data
 - Granular to a minute
- Equivalent to **11,000,000** plus alert points and counting

Understand Normal

- Score KPIs
 - In real time
 - At regular cadence
- Move data to preferred destination

Score in Realtime

- Multivariate Analysis
 - Allows early detection
 - Minimizes false positives
- Driven by embedded domain expertise

Identify Events

✓ Efficient Many-to-One Data Source

✓ Combine Domain Expertise and Data Science

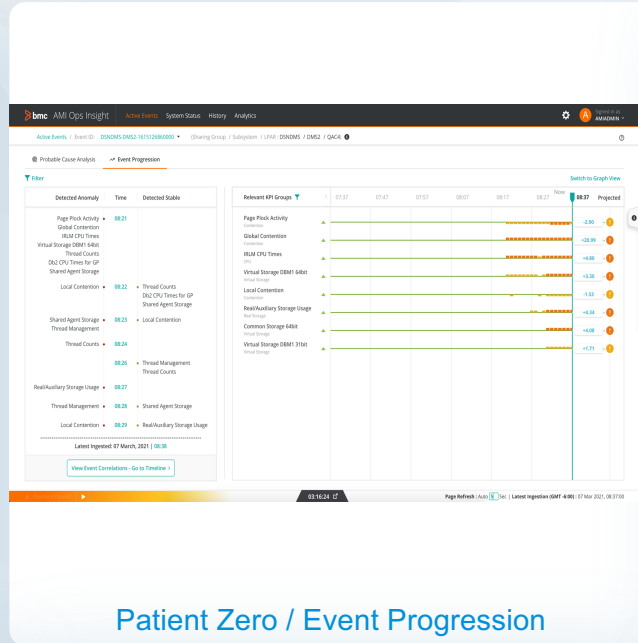
✓ Provide Actionable Insights

POLISH DB2 USERS GROUP

MINIMIZE MTTR



- Data driven cause-effect analysis
- Identifies hotspot(s) to investigate



- Start of event and cascading effect
- Event log shows problem areas



- In context graphs of problem areas
- Highlight KPI values and scores

EASY BUTTON – GIVE ME THE ANSWER BEFORE I HAVE A PROBLEM , THE PROMISE OF AI/ML SOLUTIONS



The screenshot displays the BMC AMI Ops Insight interface. The main window is titled "I/O - Details on 26 Sep 2021" and shows a "Probable Cause Analysis" for an event. The "OVERVIEW" section includes a table with performance metrics:

Metric	Value
Hit Ratio With Prefetch	100.0
Hit Ratio Without Prefetch	100.0

A "Why" button is visible below the table. The "ANALYSIS" section shows a dependency graph with nodes for "AMIDB2AA AMITSA1 Resource" and "CICV5507 CICS". A red box highlights the "Latest Conclusion" from "10 Jan 2022 | 03:13". A tooltip for the "SQL Statement" is shown, containing the following code:

```
DECLARE ECURSOR CURSOR FOR SELECT MAX ( NUMBER_OF_COUPONS )  
FROM AMI . DB2AA_TBL T1 WHERE EXISTS ( SELECT * FROM AMI .  
DB2AA_CODE T2 WHERE T1 . SERIAL_NUMBER = T2 . SERIAL_NUMBER ) AND ( T1 .  
SERIAL_NUMBER BETWEEN 396950 AND 1999999 ) AND T1 .  
TICKET_DESTINATION LIKE 'SV%' OR T1 . TICKET_DESTINATION LIKE 'UT%' OR  
T1 . TICKET_DESTINATION LIKE 'VO%' OR T1 . TICKET_DESTINATION LIKE  
'WE%' OR T1 . TICKET_DESTINATION LIKE 'PT%' OR T1 . TICKET_DESTINATION  
LIKE 'UF%' OR T1 . TICKET_DESTINATION LIKE 'GM%' OR T1 . TI
```

The interface also shows "Event Classification" for "I/O" (Bufferpool Read) and "Workload" (2). A status bar at the bottom indicates "Realtime & Playback Scoring Inactive".

Real Time Automated Deep Dive Into Problem Details

POLISH DB2 USERS GROUP

FALSE POSITIVE SUPPRESSION

- » Feature engineering
 - » Identifies appropriate and enough KPIs
- » Multi-variate analysis
 - » A single KPI does not trigger an event
- » Classification using Graph databases
 - » A valid cause-and-effect relationship is required
 - » Path inversion
- » Restart Detection
 - » Automatically switch off/on KPIs during system warmup following a restart
- » Business days
 - » Change model sensitivity
 - » Schedule it with calendaring
- » Insufficient or poor-quality training data
 - » Model health



GRAPH DATABASE

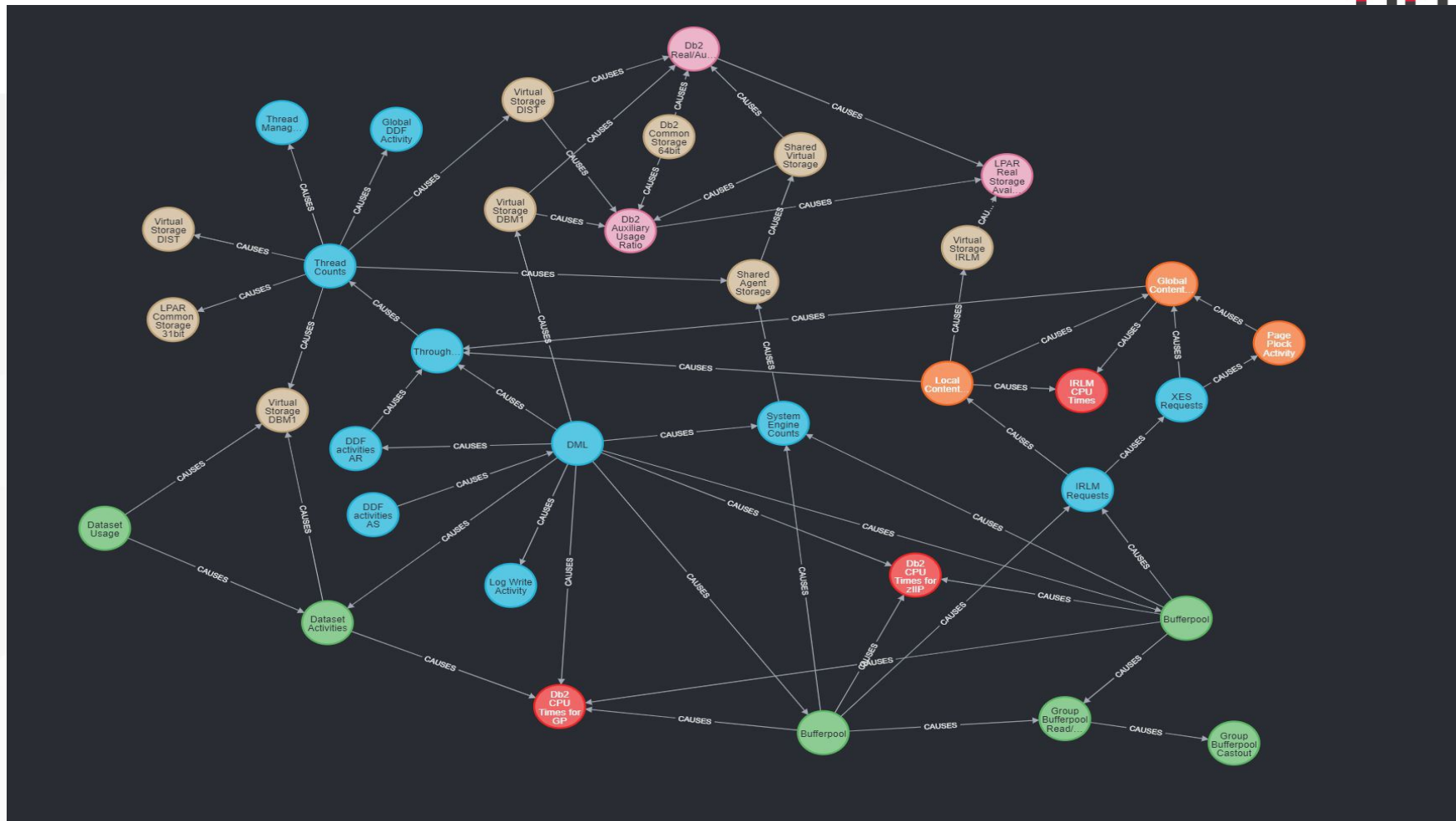
GRAPH DATABASE

- » The human brain is a living graph database
 - » Everything we do we look for a cause
 - » House too cold – Thermostat, too cold outside, heating broken

- » Graph database for cause and effect
 - » A causes B
 - » B cause C
 - » Therefore, A causes C, no need to worry about B

GRAPH DATABASE – EXAMPLE GROCERY

- » Issues / Problems
 - » Traffic Jam outside my house
 - » Traffic causes my driveway to be blocked
 - » Cannot leave the house
 - » Traffic light is defective
 - » Unable to buy groceries
- » Graph format cause and effect
 - » Traffic light is defective -> Traffic Jam outside my house -> Traffic causes my driveway to be blocked -> Cannot leave the house-> Unable to buy groceries



THE NEW WORLD OF GEN-AI

LLMS ARE NOT TRAINED FOR THE MAINFRAME ECO SYSTEM

- » The problem-solving data is not contained in the model
- » Models tend to be generic
- » Your results can vary by model
 - » Hallucinations are common
- » SLM would be a lot better
 - » Expensive to build and where is the training data
- » A good prompt helps

TEST – WHAT IS XES AND XCF

PDUUG

- » Let's try a prompt
- » Answer is expected to be terrible

WHAT IS XES AND XCF

The terms **XES** and **XCF** refer to specific file formats, and each serves a distinct purpose in different domains. Here's a breakdown of what each one represents:

1. XES (eXtensible Event Stream)

- Purpose:** XES is an open standard format used primarily for storing event logs, commonly used in the context of **process mining** and **business process analysis**. It is part of the ProM framework and is used to represent event data that is typically extracted from various IT systems such as enterprise resource planning (ERP), customer relationship management (CRM), or other business applications.

- Use Case:** XES files typically store data about the execution of business processes, including timestamps, event types, resources, and the activities within the processes. It is used to analyze workflows and detect inefficiencies, bottlenecks, and to improve operational processes.

- File Extension:** .xes

- Example:** A file containing records of sales transactions, where each event corresponds to a step in the sales process (e.g., order received, order processed, shipment sent).

WHAT IS XES AND XCF

2. XCF (eXperimental Computing Facility)

- Purpose:** XCF is the native file format used by **GIMP** (GNU Image Manipulation Program), an open-source graphics editor. XCF files store images and preserve all the layers, channels, paths, and other elements of the image that make up a GIMP project.
- Use Case:** XCF files are used to save the editable state of a project in GIMP so that users can continue editing it later with all of the original elements intact (such as layers and transparency). It is similar to Adobe Photoshop's .PSD format.
- File Extension:** .xcf
- Example:** A GIMP project file that contains an image with several layers (e.g., background, text, and illustrations), which can be edited in GIMP.

In summary:

- XES:** Primarily used in **process mining** to represent event logs.
 - XCF:** Used in **GIMP** for storing image projects with layers and other editable elements.
- Let me know if you need more details on either format!

TEST - EXPLAIN IBM SYSPLEX WITH REGARDS TO XES AND XCF

Summary of XES and XCF in IBM SYSPLEX:

- **XES (Cross-system Extensible Services)** provides **synchronization** and **communication** services for managing shared resources (e.g., locks, data sets) and ensuring data consistency across multiple systems in the SYSPLEX.
- **XCF (Extended Communication Facility)** handles **messaging** and **coordination** between systems in a SYSPLEX, enabling them to communicate about system health, workload distribution, and resource status.

Both XES and XCF work together to maintain **high availability**, **reliability**, and **scalability** in IBM mainframe environments by enabling systems within the SYSPLEX to share information, synchronize operations, and operate as a cohesive whole.

KNOWLEDGE ACQUISITION

PDUG

- » Allow experts to teach the model
- » Augment with tribal knowledge
- » "Expert on your shoulder"
- » Tooling required
 - » Works well with Graph Databases and Vector Databases

RAG TO THE RESCUE – WHAT IS IT

Retrieval-Augmented Generation - **RAG addresses this limitation:**

RAG allows LLMs to access and retrieve relevant information from external sources, effectively augmenting their knowledge.

How it works:

Retrieval: A user's query is first sent to a retrieval system, which searches for relevant information from external sources.

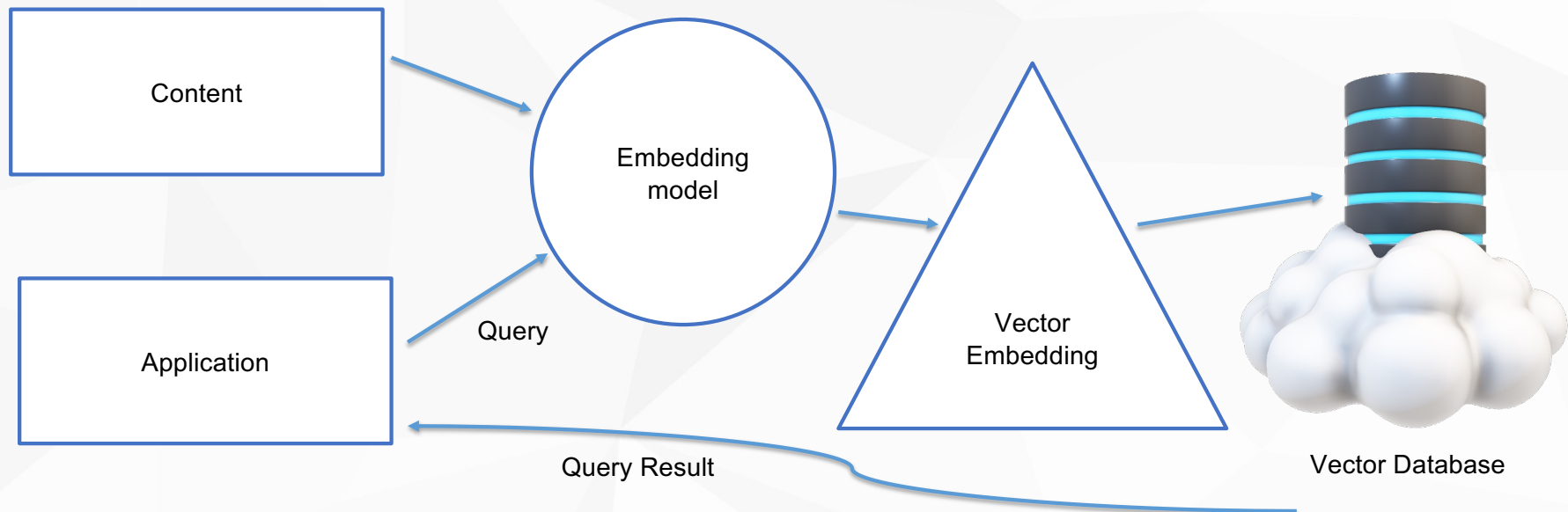
Integration: The retrieved information is then incorporated into the LLM's input, along with the original query.

Generation: The LLM generates a response based on the combined information, including the external knowledge.

ADDING ANOTHER DATABASE TYPE VECTOR

- » SQL databases (e.g. Postgres, HSQLDB, db2)
- » Graph databases (e.g. NEO4J)
- » Time series databases (e.g. Prometheus, Victoria Metrics)
- » Vector databases (e.g. Pinecone)

VECTOR DATABASE FLOW



VECTOR DATABASE

- » The content is fed into the embedding model
 - » A bit like an index
- » The vectors are added to the vector database
- » The query also uses the embedding model
 - » We're looking for similar content
- » The retrieved content and query are fed into the LLM

CONTENT AWARE STORAGE (IBM)

- » Interesting stuff from IBM to try and help with this problem
- » Improves AI assistant accuracy
- » Up to date data insights
 - » Recent events updated in real time
 - » Automatic pipelining to update the RAG (Vector) database
- » No need for ETL of data



GEN-AI - EXAMPLE

bmc AMI Ops | Events | System Status | Alarms | Views | Dashboards

← Back to Db2 Subsystem Timeline | You are in: Historical Events / DMS2 Event

DMS2 Event

Probable Cause Analysis | Event Progression

24 Nov, 2024 | 11:16 Event Start | 24 Nov, 2024 | 10:21 | 24 Nov, 2024 | 11:16 | 24 Nov, 2024 | 11:21 Event End

Event Classification

- Workload** 1
 - Thread Counts**
 - Z-score: + 1.89
 - BMC AMI Ops | THDACTV
 - View: Filter Causal Path | Classification Path
 - [View Detailed Analysis](#)

Thread Counts

Thread Counts shows as positive, it indicates more threads are created in the Db2 system than normal at this time of the day. The thread storage used by Db2 is in the **Shared Agent Storage**, when there are more threads, more **Shared Agent Storage** is used, so **Shared Agent Storage** shows as positive. As we know **Shared Agent Storage** is part of **Shared Virtual Storage 64bit**, so **Shared Virtual Storage 64bit** is also positive. The increased threads and thread storage usage might be because of the increased workload or extra threads created to support the current workload, which results in an increase in **Db2 Real/Auxiliary Storage Usage**, as more storage is needed to back up the increased workload.

Linear View Reset

Db2 Real/Auxiliary Storage Usage | Shared Virtual Storage 64bit | Shared Agent Storage | Thread Counts

GMT -6:00 | CURRENT TIME | GMT
Db2 Timezone 18:05:48 | 00:05:48 24hr clock

Latest Ingestion (GMT -6:00) : 26 Nov, 2024 | 18:04:00

GEN-AI - EXAMPLE

bmc AMI Ops | Events | System Status | Alarms | Views | Dashboards

← Back to Db2 Subsystem Timeline | You are in: Historical Events / DMS2 Event

DMS2 Event

Probable Cause Analysis | Event Progression

24 Nov, 2024 | 11:16 Event Start | 24 Nov, 2024 | 10:21 | 24 Nov, 2024 | 11:21 | 24 Nov, 2024 | 11:21 Event End

Event Classification

- Contention** 1
 - Local Contention**
Z-score: + 15.39
BMC AMI Ops View: STLOCK
 - Filter Causal Path
 - Classification Path 1
 - Classification Path 2
 - [View Detailed Analysis](#)
- Workload** 2
 - Thread Counts**
Z-score: + 2.35
BMC AMI Ops View: THDACTV
 - Filter Causal Path
- IRLM Requests**
Z-score: -1.62
BMC AMI Ops View: STLOCK

IRLM Requests

The Local Contention is positive, it indicates more Local Contention occurs. Local Contention happens when multiple processes or threads within the same Db2 subsystem compete for the same resource, so it slows down the workload processes and reduce the throughput, so the IRLM Requests is negative. However, IRLM CPU Times is positive, it implies that IRLM is spending more CPU time to solve the contention, which means there are more conflicts on locking requests between db2 members of data sharing group when those members are trying to serialize shared resources.

Virtual Storage IRLM | IRLM CPU Times | Local Contention | IRLM Requests

Db2 EXCEPTIONS: QTXATIM

GMT -6:00 | CURRENT TIME GMT | 18:09:12 | 00:09:12 | Latest Ingestion (GMT -6:00) : 26 Nov, 2024 | 18:04:00

OTHER HOT ADVANCES

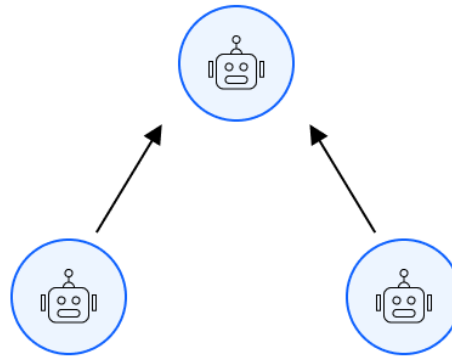
AGENTIC – AI AGENTS WORKING TOGETHER

Single agent architecture

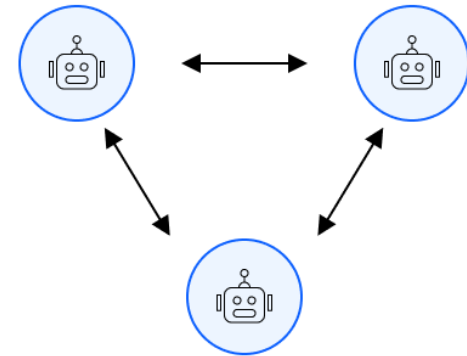


Multi agent architecture

Vertical architecture



Horizontal architecture



MCP SERVERS (RELEASED ABOUT 6 MONTHS AGO)

- » Model Context Protocol (MCP) is an open standard developed by Anthropic, the company behind Claude.
- » Very fast adoption
- » Light weight standard protocol adapters
- » Allows the AI data gathering to easily access other sources
- » Vendor agnostic open standard
 - » Allows you to develop without worry

QUESTIONS

?