Multi-Agent System For Generic Anomaly Detection And Power BI Reporting

PREPARED BY Chammami Yahya

SUPERVISED BY

Dr. Abdeljaoued Tej Ines

Mr. Sahli Chekir

Ms. Ben Farhat Emna

Introduction

- Modern IT systems generate massive logs across infrastructure.
- 22% of organizations create 1TB or more worth of logs each day
- Log data grew 250% year-over-year on average

chronosphere

There is a growing necessity to explore these logs to make sure everything is working well and safe, with no problems.

TALAN

Typical Anomaly Rates

Anomaly Demographics

5%

Cloud servers
load balancing issues,
or hardware failures.

10%

Virtual Machine
system events. resource
spikes, unauthorized
access

8%

OS logs
CPU surges, memory
leaks

15%

Physical computers
hardware issues, software
bugs, or external threats

From cloud servers to physical machines, anomaly rates can range from 5% to 15%. This can include everything from hardware failures to unauthorized access attempts or sudden CPU spikes.

Problem

09:01:06 Jun ERROR [main] - Failed to authenticate user: Invalid credentials provided

Jun 14 15:16:02 combo sshd(pam_unix)[19937]: check pass; user unknown

17/06/09 20:10:40 INFO - root - Unhandled exception occurred while processing file 'report.csv'

Quantity & Complexity of Logs

The complex structure of log data varies from one type to another

Poor Detection Results

Traditional approaches and rule-based methods struggle to keep up with patterns changing.

Limited Insight & Reporting

Most systems detect issues but fail to clearly explain or summarize them, delaying effective response.

Solution

an intelligent system capable of tracking and detecting anomalies in real time, using a multi-agent architecture that coordinates anomaly detection, insight discovery through web search, and automated report generation

Real-time

To handle the large amount of logs, we need real-time data ingestion and detection.

Multi-Agent

We implement a multi-agent system that enables parallel processing, intelligent task delegation, and continuous monitoring .

Anomaly Detection

The system leverages adaptive models that analyze, explain, and prioritize anomalies based on context and severity.

Power BI

We redesign the Power BI structure to accommodate the diversity of log data and enable automated report generation.

TODAY'S AGENDA

01 OVERALL SYSTEM ARCHITECTURE

02 AGENTIC SYSTEM ARCHITECTURE

03 ANOMALY DETECTION AGENT

04 CRITICALITY & WEB SEARCH AGENT

06 POWER BI REPORTING AGENT

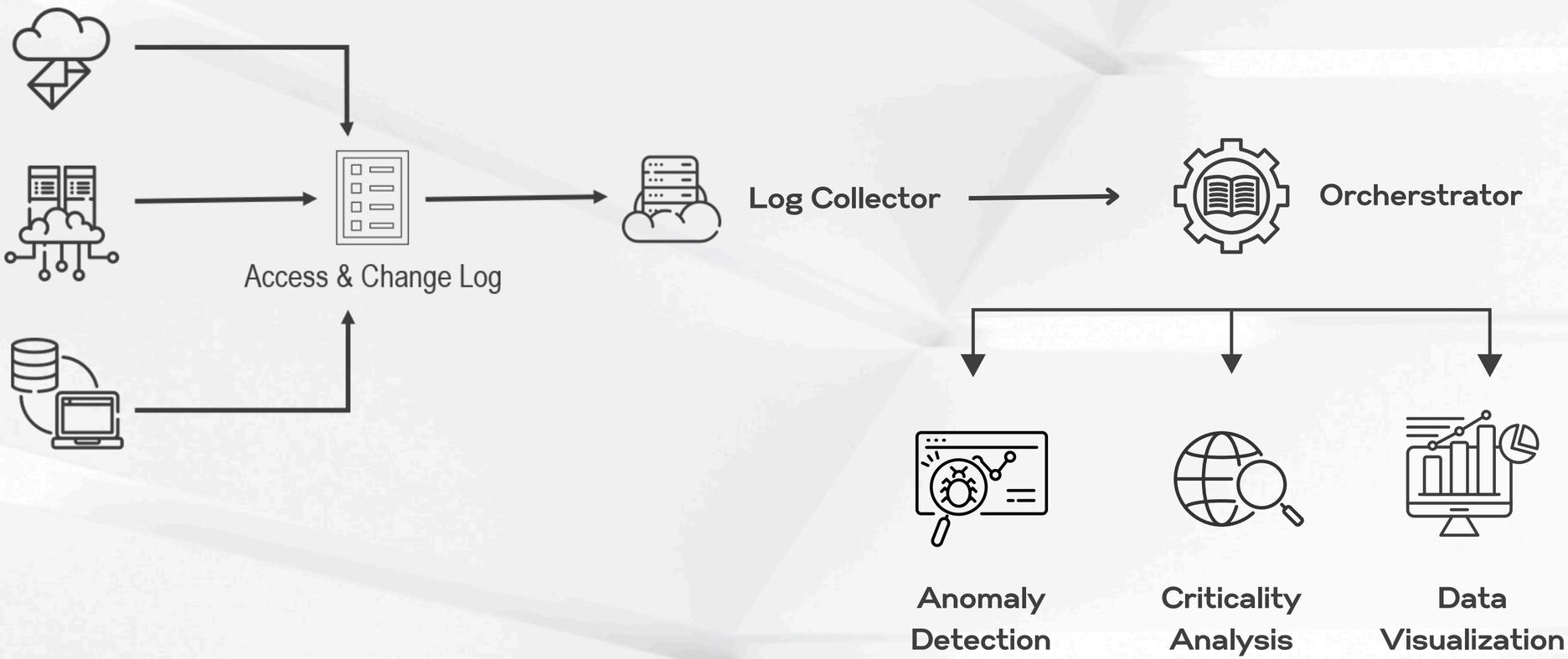
07 IMPACT OVERVIEW

08 GENERAL CONCLUSION

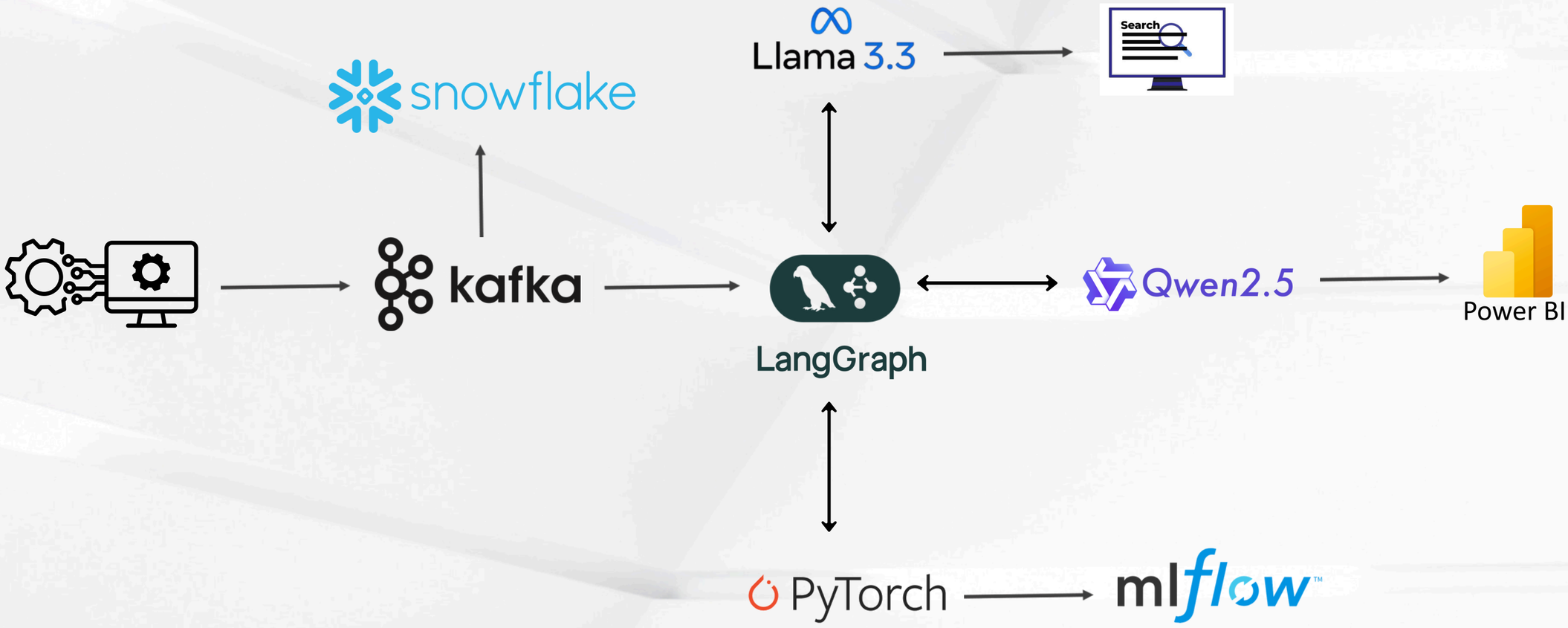
09 LIMITES & PERSPECTIVES

TALAN

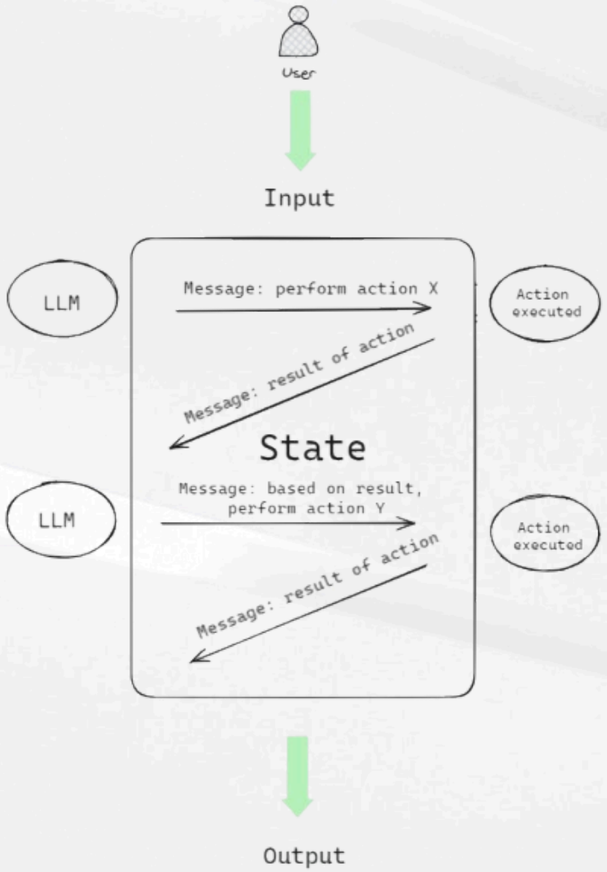
Overall System Architecture

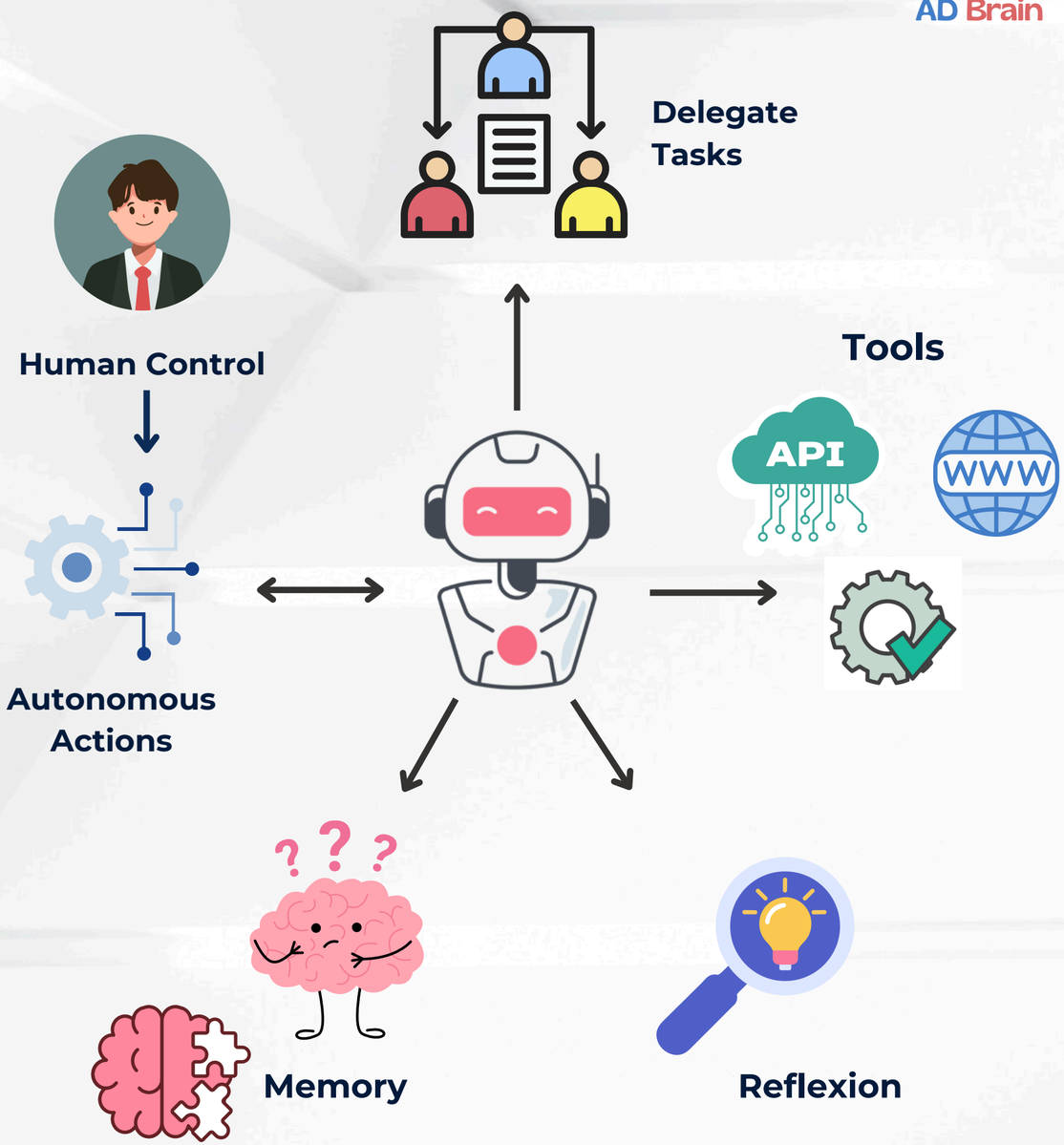


Overall System Architecture



Agentic System Architecture





$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Row Logs Datasets

BGL (Supercomputer)

7.34% anomalies 92% Normal

HDFS (Distributed File System)

2.51% anomalies 97.5% Normal

Openstack (Cloud)

8.87% anomalies 91.1% Normal

Log Parsing

Drain Algorithm

Feature Extraction

Code-BERT

Data Preprocessing

SimCLR

Anomaly Detection Models

2025-06-19 11:50:12 INFO [server] Received request from IP 192.168.1.10 on port 443

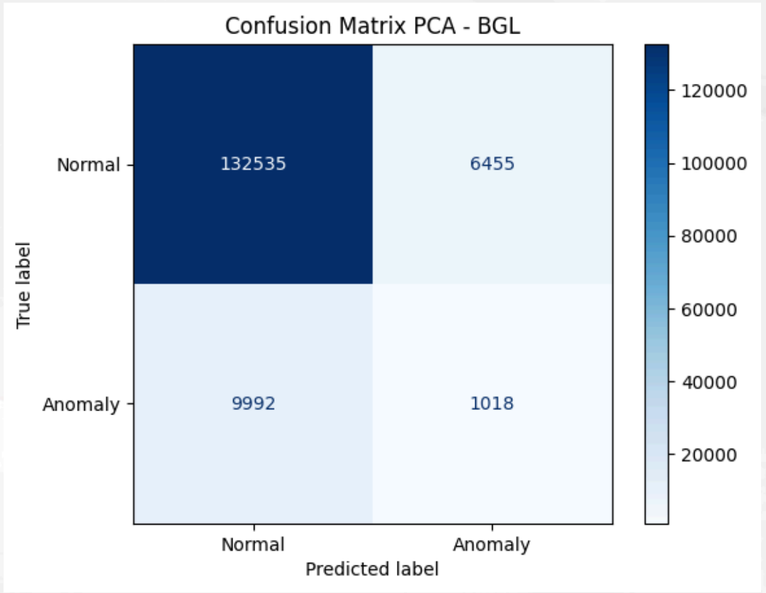
Received request from IP <**> on port <**> ["192.168.1.10", "443"]

TALAN

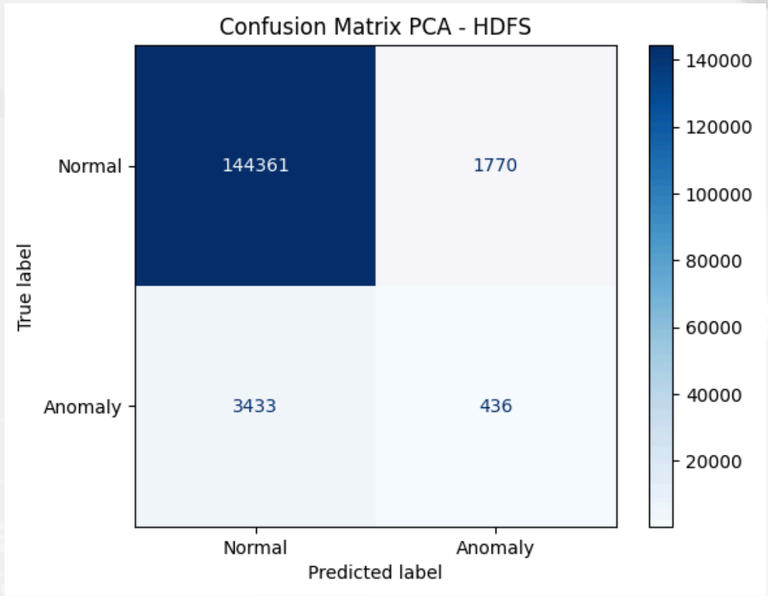
Anomaly Detection Agent

PCA

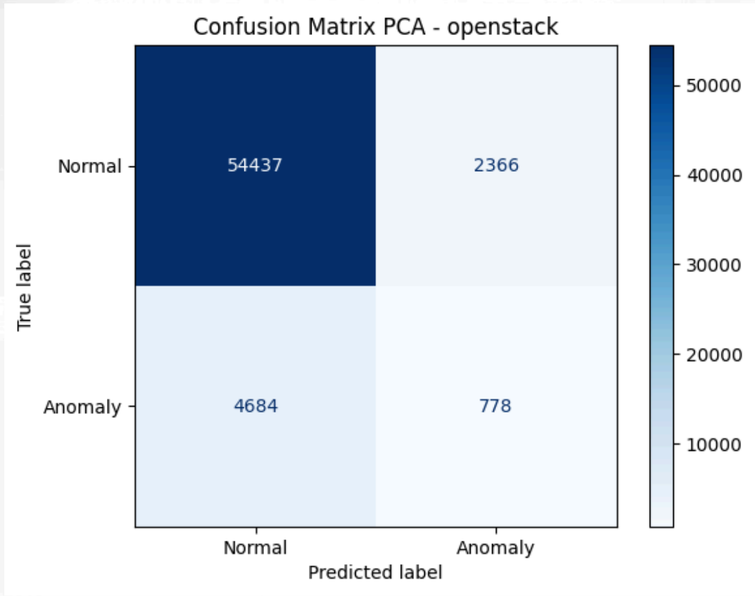
13.62% | 9.25% | 11.02%



19.76% | 11.27% | 14.35%



24.75% | 14.24% | 18.08%

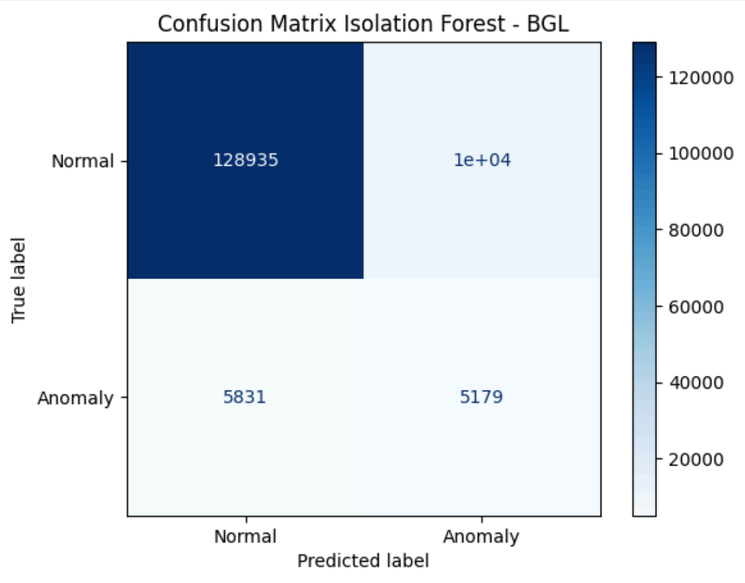


PCA, which is a dimensionality reduction technique, detects anomalies based on how far points deviate from the main components, and it showed poor results on all datasets.

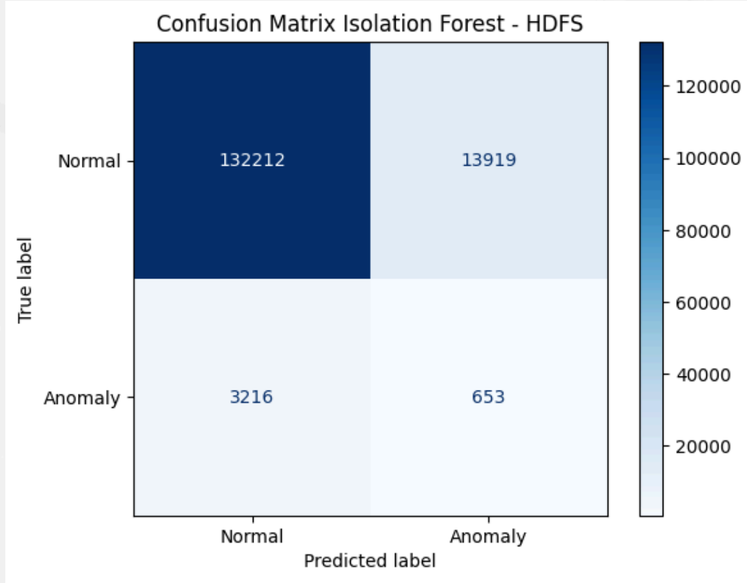
Anomaly Detection Agent

Isolation-Forest

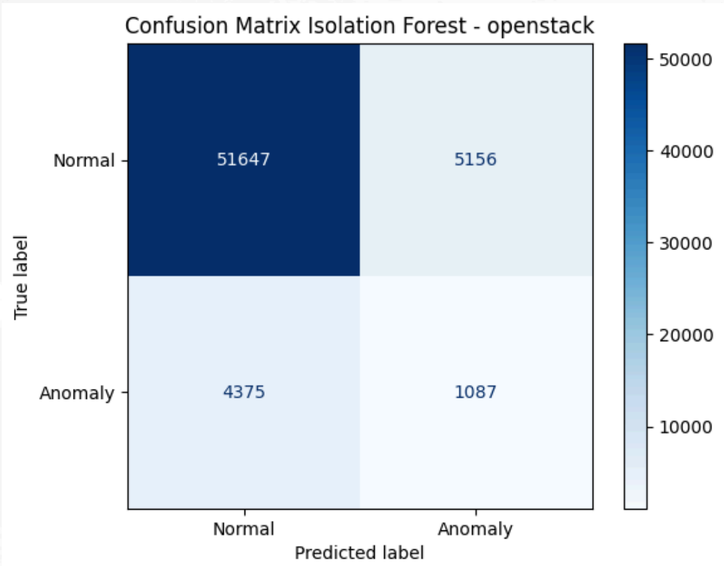
34% | 47.04% | 39.47%



4.48% | 16.88% | 7.08%



17.41% | 19.9% | 18.57%

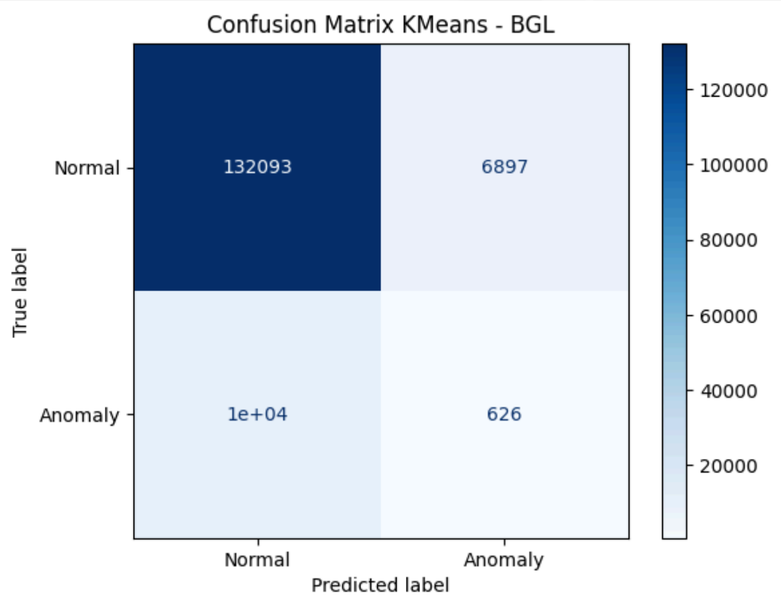


Isolation Forest is a tree-based algorithm that detects anomalies by isolating data points that behave differently. It also showed poor result on the three log dataset.

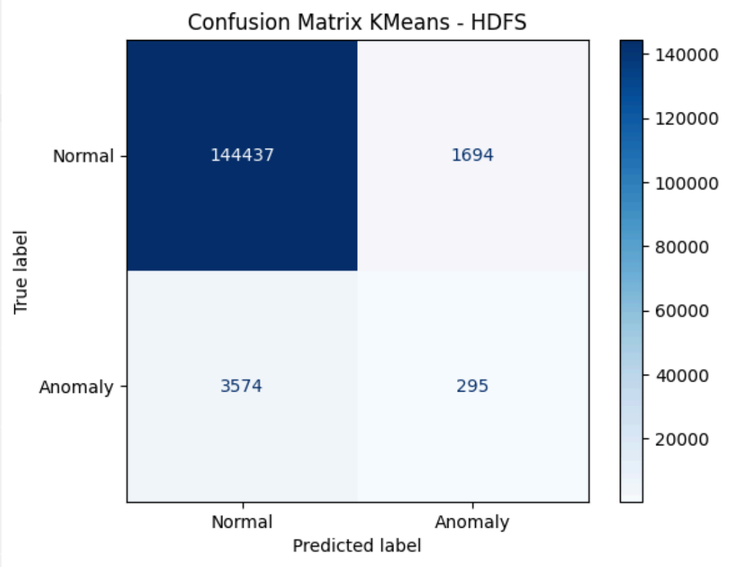
Anomaly Detection Agent

K-means

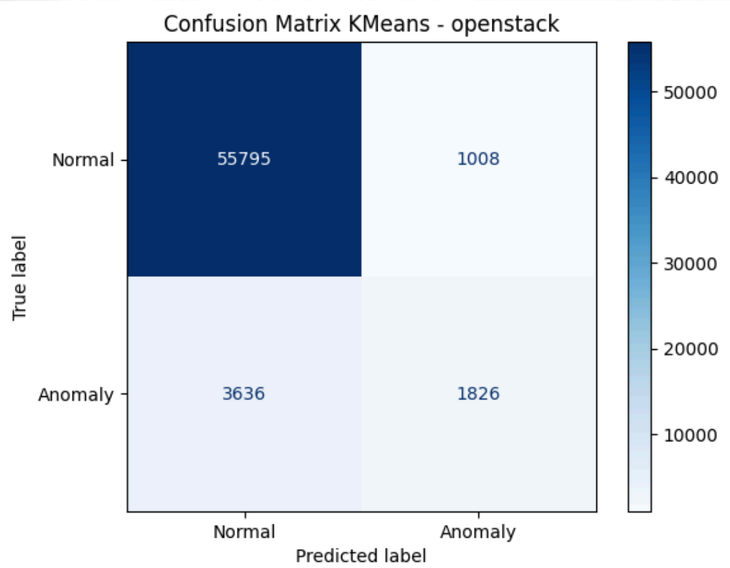
8.32% | 5.69% | 6.76%



14.86% | 7.62 %| 10.8%



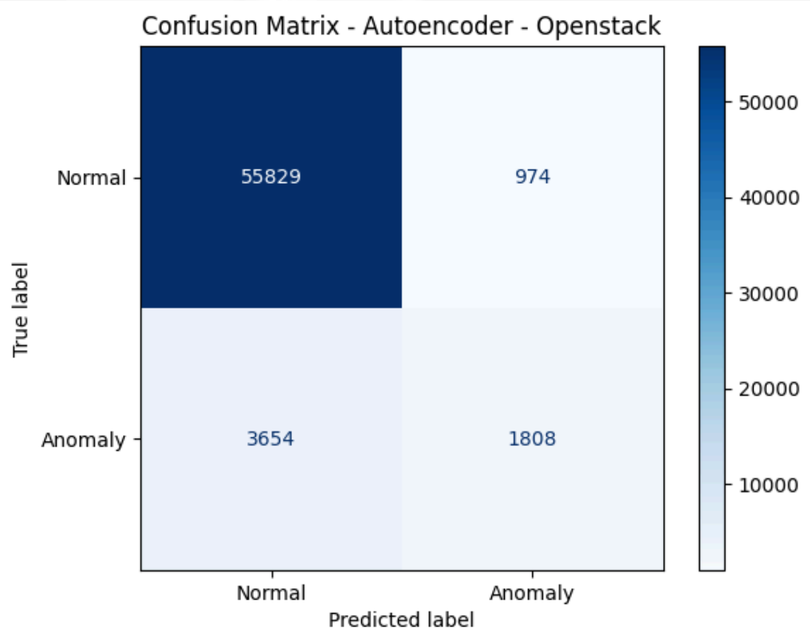
64.43% | 33.43% | 44.02%

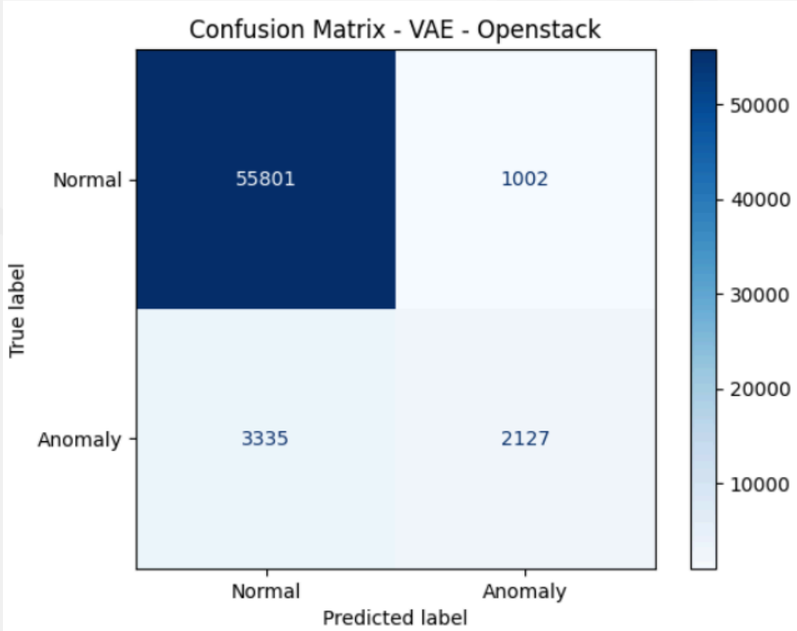


K-Means is a clustering algorithm that groups data to find anomalies, It showed slight improvement in results but lacked generalization.

Anomaly Detection Agent

Autoencoder & PC-DARTS Variational Autoencoder





AEs

64.98% | 33.1% | 43.86%

PC DARTS - VAEs

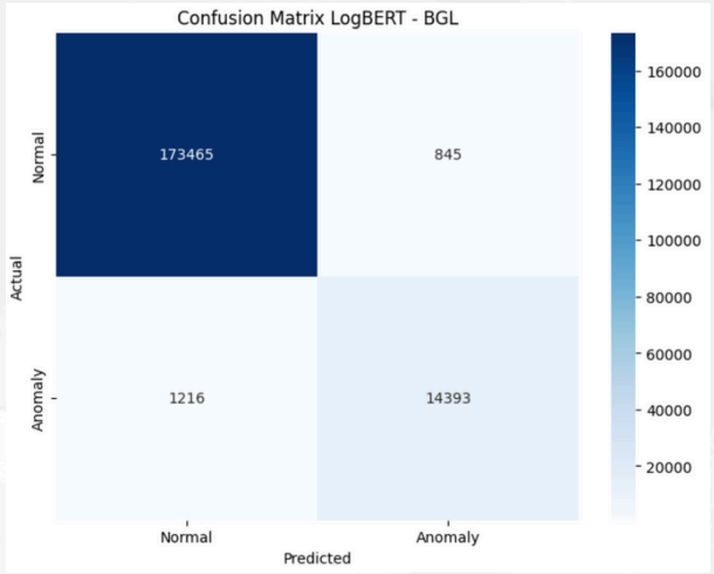
67.9% | 38.94% | 49.51%

Autoencoder is a neural network that learns to compress and reconstruct data. It detects anomalies by measuring reconstruction errors.

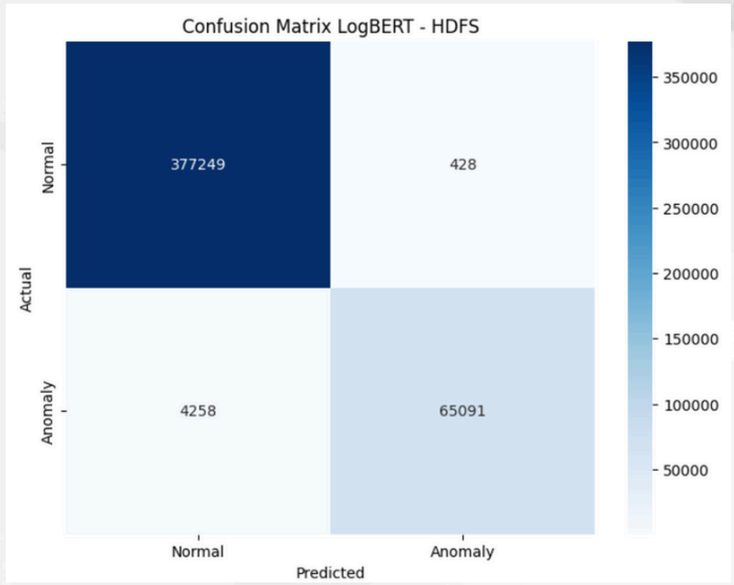
Anomaly Detection Agent

Log-BERT

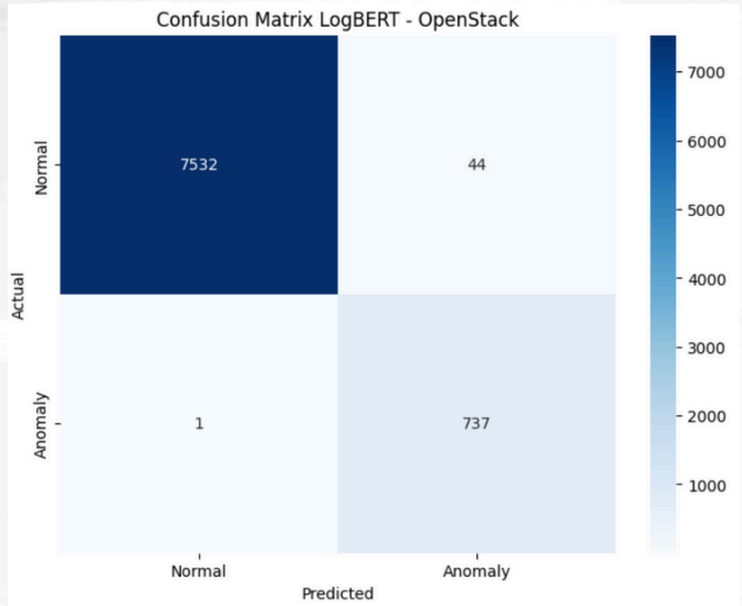
0.94% | 0.92% | 0.93%



0.99% | 0.93% | 0.96%



0.94% | 0.99% | 0.97%



LogBERT is a transformer-based model trained to understand the structure and meaning of log sequences. It detects anomalies by identifying unexpected patterns in the log context.

Criticality & Web Search Agent

Enter Index of Anomaly to Analyze

17

Analyze Selected Anomaly

Analysis Result

This log entry is related to a hardware error and is also known as a WHEA-Logger Event ID 17. The error occurs due to a corrected hardware error and may be caused by various reasons such as outdated drivers, corrupted system files, or malfunctioning hardware components.

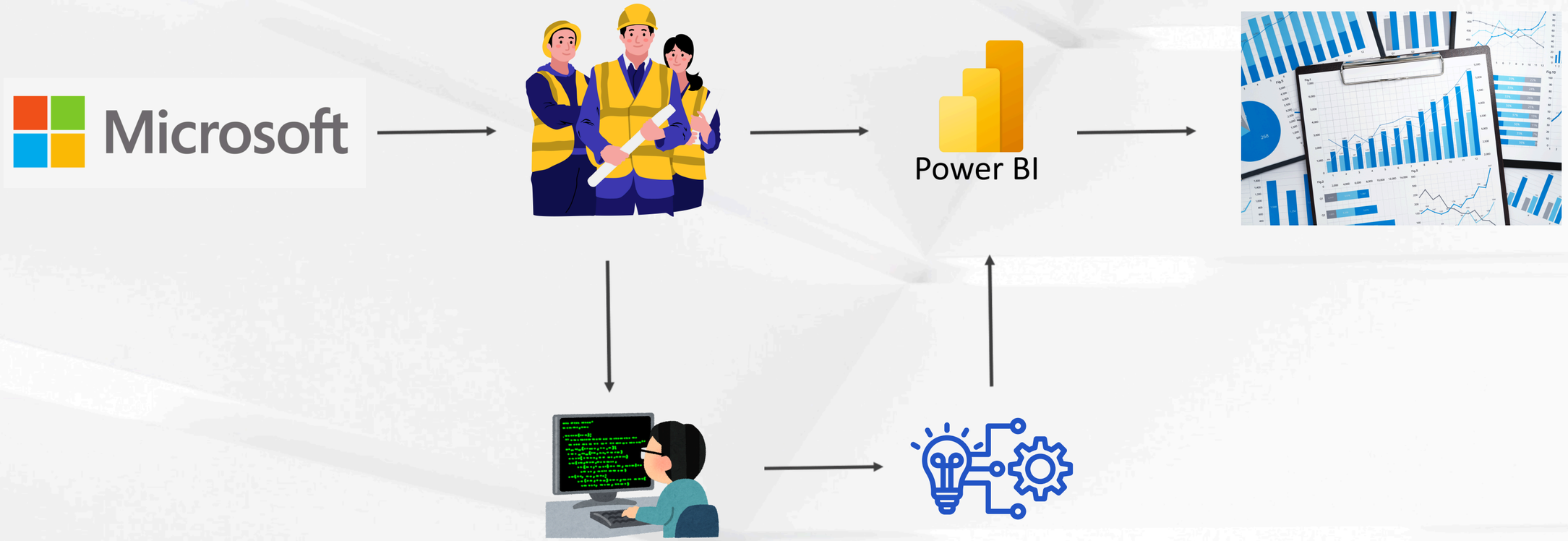
To fix this error, you can try the following steps:

1. Update your drivers to the latest version.
2. Run a System File Checker scan to detect and replace corrupted system files.
3. Disable the WHEA-Logger service.
4. Perform a clean boot to start Windows with a minimal set of drivers and startup programs.
5. Check the Event Viewer logs for more event 17 errors and clear them.

It is essential to note that if the issue persists, it may be related to a hardware problem, and you should contact the manufacturer for further assistance.

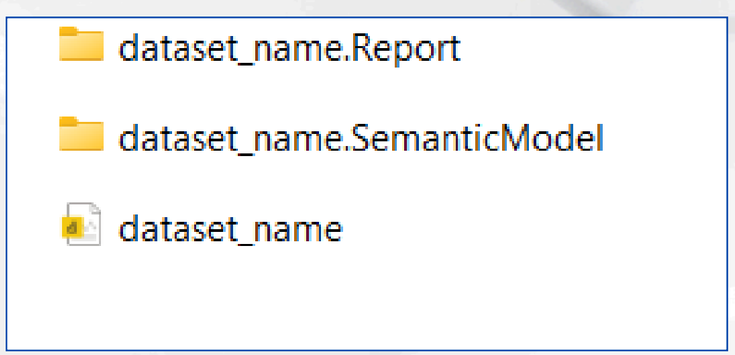
Sources:
1. https://answers.microsoft.com/en-us/windows/forum/all/whea-logger-event-id17-a-corrected-hardware-error/e0079af5-68f1-4b7d-b3f2-d56e5a078cfe
2. https://www.thewindowsclub.com/fix-whea-logger-fatal-hardware-and-event-id-errors
3. https://support.lenovo.com/us/en/solutions/ht500599-many-whea-logger-event-17-error-messages-shown-in-windows-event-viewer-system-log-thinkcentre-m700-m800-m900-m700z-m800z-x1-thinkstation-p310
4. https://h30434.www3.hp.com/t5/Business-PCs-Workstations-and-Point-of-Sale-Systems/WHEA-Logger-errors-in-event-viewer-several-pet-minute/td-p/8359533

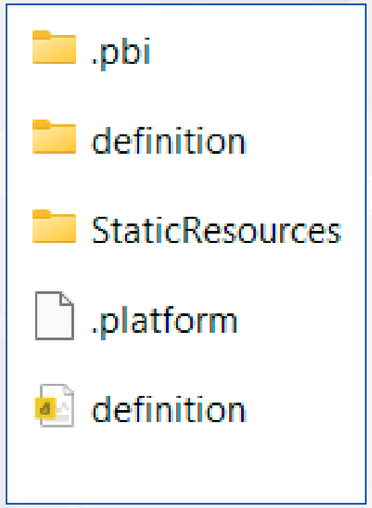
PowerBI Reporting Agent

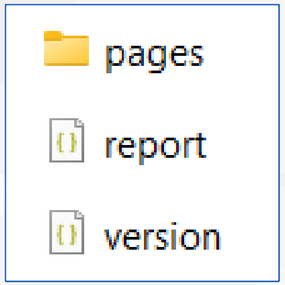


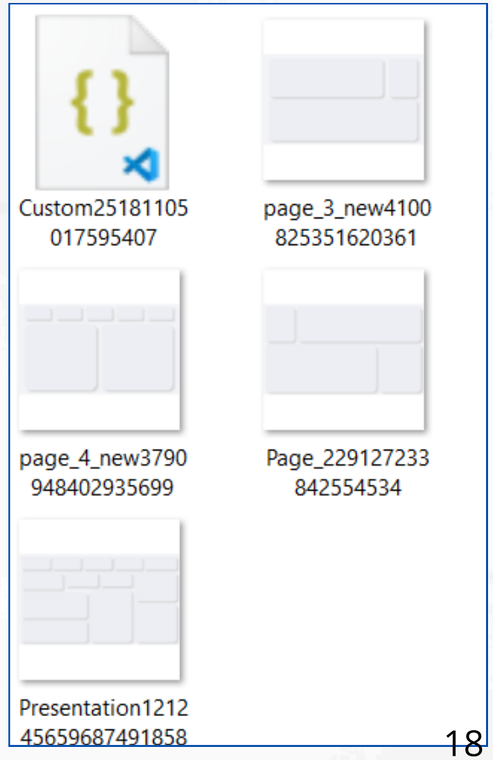
PowerBI Reporting Agent

```
output/
    |--- dataset_name.Report/
        |--- definition/
        |--- StaticResources/
        |--- definition.pbir
    |--- dataset_name.SemanticModel/
        |--- definition.pbism
        |--- diagramLayout
        |--- model.bim
    |--- dataset_name.pbip
```





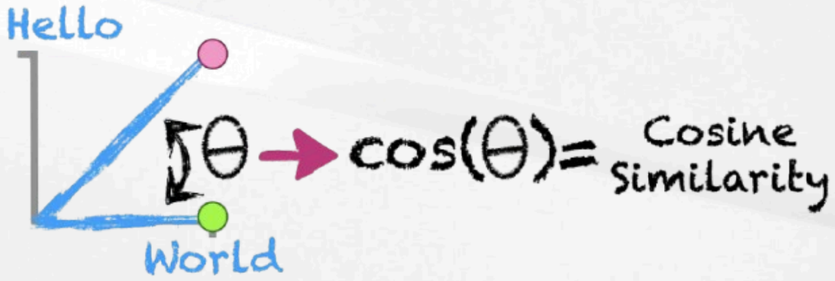




Template Extraction

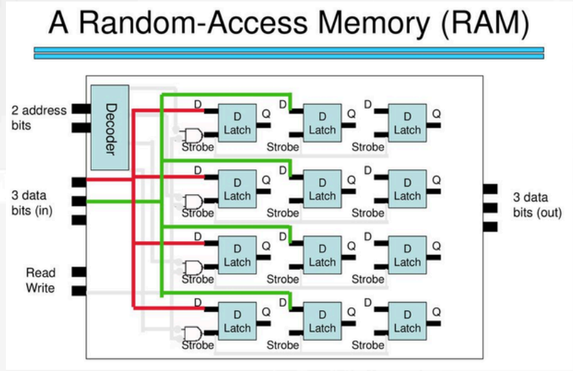
RAG & MemoRAG

RAG relies on indexing the input and retrieving relevant content based on similarity search. Memory-augmented models enhance this by adding memory components, like LSTMs, which help us better understand the structure of configuration files.

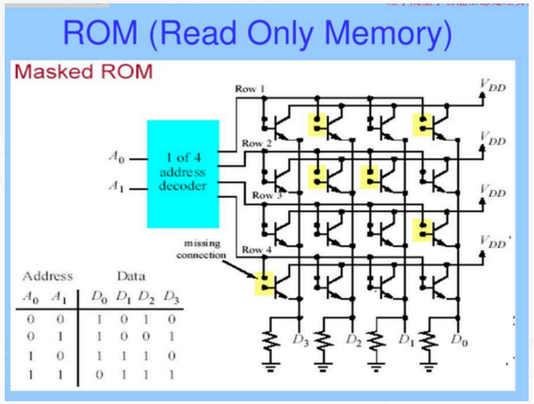


It's similar to how computers distinguish between RAM and ROM — one acts as global memory, while the other functions more like a cache.

Global Memory



Cached Memory



- report.json (a JSON file) containing:


```
{
  "config": "{\"version\":\"5.59\",\"themeCollection\":{\"baseTh
  \"defaultDrillFilterOtherVisuals\":true,\"linguisticSchemaSync
  \"useStylableVisualContainerHeader\":true,\"queryLimitOption\"
  \"objects\":{\"section\":[{\"properties\":{\"verticalAlignment
  "layoutOptimization": 0,
  "resourcePackages": [
    {
      "resourcePackage": {
        "disabled": false,
        "items": [
          {
            "name": "CY24SU10",
            "path": "BaseThemes/CY24SU10.json",
            "type": 202
          }
        ],
        "name": "SharedResources",
        "type": 2
      }
    }
  ],
  "sections": [
    {
      "config": "{}",
      "displayName": "Page 1",
      "displayOption": 1,
      "filters": "[]",
      "height": 720.00,
      "name": "e5c405b41d47b615abf6",
      "visualContainers": [
        {
```



```
"barChart": {
  "*": {
    "general": [
      {
        "responsive": true
      }
    ],
    "legend": [
      {
        "showGradientLegend": true
      }
    ],
    "smallMultiplesLayout": [
      {
        "backgroundTransparency": 0,
        "gridLineType": "inner"
      }
    ]
  }
},
"clusteredBarChart": {
  "*": {
    "general": [
      {
        "responsive": true
      }
    ],
    "legend": [
      {
        "showGradientLegend": true
      }
    ],
    "smallMultiplesLayout": [
      {
        "backgroundTransparency": 0,
        "gridLineType": "inner"
      }
```



```
"image": {
  "*": {
    "background": [
      {
        "show": false
      }
    ],
    "general": [
      {
        "keepLayerOrder": true
      }
    ],
    "visualHeader": [
      {
        "show": false
      }
    ],
    "lockAspect": [
      {
        "show": true
      }
    ]
  }
},
"actionButton": {
  "*": {
    "background": [
      {
        "show": false
      }
    ],
    "visualHeader": [
      {
        "show": false
      }
```

Report Generation

QwenCoder 2.5 & Power BI

• Last saved: 5/6/2025 at 3:46 PM (Power BI Project)

Search

Insert Modeling View Optimize Help

Get data | Excel workbook | OneLake catalog | SQL Server | Enter data | Dataverse | Recent sources — Data

Transform data | Refresh — Queries

New visual | Text box | More visuals — Insert

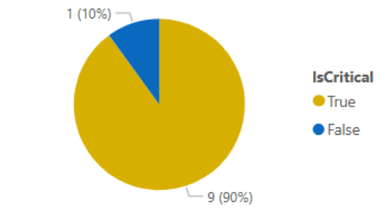
New visual calculation | New measure | Quick measure — Calculations

LogMessage	IsCritical	Frequency (%)	Explanation	OccurrenceCount	Sources
Backup failed	False	19.0	Scheduled backup did not complete successfully.	19	https://docs.backupsolutions.com/; https://stack
Configuration file missing	True	11.0	The system couldn't locate a required configuration file.	11	https://linuxize.com/post/how-to-fix-missing-co
Disk space low	True	12.0	The server is running low on disk space, potentially affecting performance.	12	https://serverfault.com/questions/1; https://docs
Error connecting to database	True	8.0	The system failed to connect to the database due to invalid credentials or a timeout.	8	https://docs.microsoft.com/en-us/sql/; https://st
Memory usage high	True	8.0	High memory usage detected, which could lead to system instability.	8	https://docs.oracle.com/en/memory/; https://sta
Service restarted	True	18.0	A system service was restarted manually or due to a crash.	18	https://superuser.com/questions/1; https://docs.
System rebooted	True	18.0	The system rebooted unexpectedly, possibly due to a critical error.	18	https://docs.microsoft.com/en-us/windows/rebc
Timeout while processing request	True	5.0	The system timed out while processing a request.	5	https://httpstatuses.com/408; https://stackoverfl
Unauthorized access attempt	True	5.0	An unauthorized access attempt was detected.	5	https://owasp.org/www-project-top-ten/; https:/
User login failed	True	13.0	User login failed due to incorrect username or password.	13	https://auth0.com/docs/; https://stackoverflow.c

92

Number of anomalies detected

Count of IsCritical by IsCritical



LogMessage	Frequency
User login failed	11,11%
Unauthorized access attempt	4,27%
Timeout while processing request	4,27%
System rebooted	15,38%
Service restarted	15,38%
Memory usage high	6,84%
Error connecting to database	6,84%
Disk space low	10,26%
Configuration file missing	9,40%
Backup failed	16,24%
Total	**100,00%**

Most commun anomaly

Backup failed

Frequency percentage

11.0

TALAN

Impact Overview

Metric	Manual BI	LLM-Powered BI
Dashboard Creation Time	2–5 days	2–5 minutes
Number of Users Enabled	Analysts only	Entire organization
Log Insight to Visualization	Hours/days	Real-time
Business Decision Velocity	Slow	Immediate
Cost per Dashboard	High (time + labor)	Near zero

General Conclusion

This project provides an intelligent, and explainable system for anomaly detection and reporting in log data. We tackle the real pain points of IT monitoring with a market-aligned and technological advancement.

By combining state-of-the-art machine learning, multi-agent orchestration, and business intelligence reporting, the system is designed for both production deployment and commercialization.

Limites & perspectives

1

LLM Cost & Latency for Real-Time Use

Running large language models (LLMs) can be expensive and introduce latency.

Improve LLM inference with distillation, caching, quantization, or better GPU utilization.

2

Explainability in Complex Anomaly Contexts

LLMs may struggle with deep technical root causes in highly noisy environments.

Combine LLMs with knowledge graphs or domain-specific fine-tuning.

3

Dependency on Power BI Ecosystem

The system currently focuses on automating dashboard generation for Power BI

Develop the prototype to handle video games or applications with graphical interfaces

Thank you

FOR YOUR TIME AND ATTENTION

E-mail yahiachammemi@gmail.com

Github @yahyachammami

Phone +216-20688523

Address Charguia2

Positive innovation