AN INTEGRATED AI-POWERED FRAMEWORK FOR NETWORKING AND PROCESSING IN INDUSTRIAL IOT APPLICATIONS

Dr. Mahesh Sharma

ABSTRACT

In the era of Industry 4.0, where digitalized production facilities heavily depend on intricate sensor networks, optimizing data utilization is imperative for bolstering production sustainability. This entails refining processes, minimizing downtime, curbing material wastage, and more. However, achieving intelligent, data-driven decisions within stringent time constraints necessitates the seamless integration of time-sensitive networks with robust data ingestion and processing infrastructure, complemented by versatile support for Machine Learning (ML) pipelines. Unfortunately, existing frameworks often grapple with the challenge of harmonizing and programming both networking and computing infrastructures while accommodating real-time ML decision-making based on collected data. To surmount this obstacle, this paper introduces AIDA, a cutting-edge holistic AI-driven network and processing framework meticulously crafted for real-time, reliable industrial IoT applications. AIDA adeptly orchestrates Time-Sensitive Networks (TSN) to facilitate instantaneous data ingestion across an observable AI-powered edge/cloud continuum. Furthermore, AIDA integrates adaptable and dependable ML components capable of rendering timely decisions tailored to the diverse needs of industrial IoT applications. This paper meticulously delineates the AIDA architecture, elucidates key components of the framework, and offers insights through the illustration of two distinct use cases.

Index Terms *Internet of Things (IoT), Edge/cloud computing, Time-Sensitive Networks (TSN), Machine Learning.*

Reference *to this paper should be made as follows**:** Dr. Mahesh Sharma, (2023), "An Integrated AI-Powered Framework for Networking and Processing in Industrial IoT Applications" Int. J. Electronics Engineering and Applications, Vol. 11, Issue I, pp. 1-11.*

Biographical notes:

Prof. (Dr.) Mahesh Sharma *is Professor at Department of Computer Science, Institute of Innovation in Technology & Management affiliated to Guru Gobind Singh Indraprastha University, Delhi, India. He holds Master's Degree in Computer Science & Computer Applications. He is Ph.D. in Computer Science. Currently he is pursuing LLB. He has more than 16 years of experience of Industry & Academia. He has presented/published 19 Research Papers (UGC listed & Peer Reviewed Journals – National & International). Also, he has 5 Books & 5 Patents to his credit. 5 books & 5 Research Papers are pipelined. He has served as Vice Principal, Dean and IQAC Coordinator with Ideal Institute of Management and Technology, GGSPU, Delhi.*

1. INTRODUCTION

The AIDA (Adaptable, Intelligent, and Dependable Architecture) framework is introduced in response to the challenges inherent in developing real-time and dependable applications for the Industrial Internet of Things (IoT), particularly within the context of Industry 4.0. In Industry 4.0, the foundation of the Industrial IoT lies in the data collected from sensors on the shop floor, with the overarching objective of enhancing manufacturing processes through intelligent decision-making.

The paper underscores the efficacy of leveraging Artificial Intelligence (AI) and Machine Learning (ML) for real-time event detection and decision-making in industrial settings. However, it also emphasizes the associated challenges, specifically in guaranteeing real-time performance and maintaining overall system reliability. AIDA addresses these challenges through a multi-faceted approach.

One key aspect of AIDA is the integration of a converged real-time network infrastructure, facilitated by Time-Sensitive Network Configuration Management. This configuration management plays a crucial role in ensuring that the network can meet the stringent real-time requirements of industrial applications. Additionally, AIDA establishes an industrial edge–cloud continuum that incorporates adaptable observability components. This continuum is designed to seamlessly integrate computing resources at the edge with cloud-based services, enhancing flexibility and scalability.

Furthermore, AIDA incorporates agile and trustworthy ML/AI algorithms, acknowledging the need for adaptability and reliability in decision-making processes. The architecture is shaped by industrial use cases, prioritizing flexibility, scalability, and robust support for real-time applications. The paper delves into various aspects of AIDA, including TSN network configuration management, real-time performance monitoring for edge microservices, and the design of the edge/cloud software continuum.

Experimental results presented in the paper showcase AIDA's capabilities in network reconfiguration, fault detection, monitoring overhead, and dynamic ML pipeline adaptation for industrial processes. As a result, the AIDA architecture emerges as a comprehensive solution, seamlessly connecting networking, computing, and AI components to ensure reliable and efficient industrial IoT applications.

2. BACKGROUND AND RELATED WORK

In the pursuit of establishing a reliable AI/ML-enabled networked-compute fabric for Industrial IoT (I-IoT), three critical components must seamlessly collaborate to meet the required timeliness and reliability constraints. The following overview delves into each component, highlighting their challenges and the state-of-the-art approaches in the existing literature.

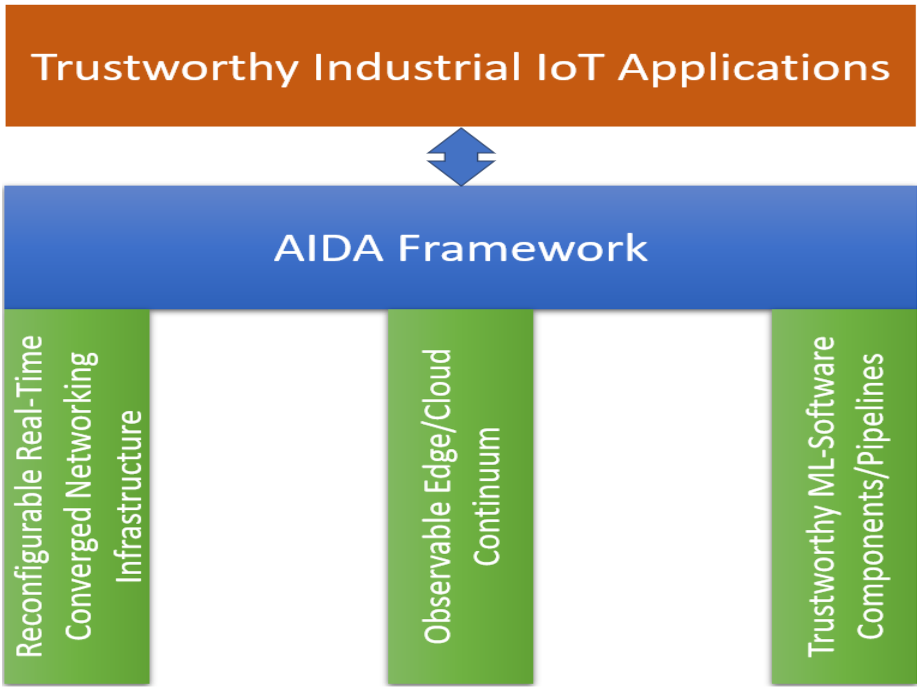

Fig. 1. Core Components of the AIDA Architectural Framework

1. Flexible and Reconfigurable Converged Real-Time Network Infrastructure (TSN):

Traditional Ethernet-based networks face challenges in delivering necessary timeliness guarantees for industrial applications. Time-Sensitive Networking (TSN) emerges as a solution, extending the IEEE 802.1 Ethernet standard with real-time capabilities. TSN offers algorithms and components for real-time traffic delivery by categorizing traffic into time-aware streams, regular streams, traffic-engineered non-streams, and non-streams. However, configuring TSN entities poses a challenge, and optimization studies leverage techniques such as array theory, Optimization Modulo Theories (OMT), Satisfiability Modulo Theories (SMT) solvers, Integer Linear Programming (ILP), heuristics, and metaheuristics. These approaches aim to streamline the configuration process and enhance the efficiency of TSN for meeting stringent industrial timeliness requirements.

2. Monitoring the Network-Compute Fabric:

Effective monitoring is crucial for ensuring the health of the system, with a distinction between System Monitoring (SM) and Application Performance Monitoring (APM). SM focuses on collecting data regarding infrastructure resources, while APM measures application metrics and performance indicators. Traditional statistical approaches, container monitoring, resource allocation, unified monitoring for physical and virtual entities, and coordinated monitoring agents contribute to monitoring solutions. However, existing solutions may lack support for containerized infrastructures, multi-tenancy, or consideration of performance requirements across microservices chains. Addressing these gaps, ongoing research explores advancements in monitoring techniques to cater to the evolving complexities of modern industrial systems, encompassing diverse architectures and deployment models.

3. Big Data Stream Processing (DSP):

Handling big data streams under timeliness constraints is imperative for I-IoT applications. Two prominent types of big data processing, batch and stream computing, coexist, with stream computing emphasizing immediate data analysis. Data Stream Processing (DSP) becomes crucial for the real-time analysis of continuous data streams in the industrial context. Research in this domain explores

DSP applications in fog/edge computing, fog-oriented run-time systems, and high-level interfaces for fog stream processing. Various frameworks, including Storm, Flink, Kafka Streams, and IBM Streams, have been investigated, considering low latency and real-time data analysis at the network edge. However, the choice of suitable software packages for building a stream processing infrastructure remains a challenge. Research efforts have focused on developing multilayer streaming analytics architectures, taxonomies of technologies, and scalable edge–cloud infrastructures tailored for large-scale IoT applications. These endeavors aim to provide practitioners with a comprehensive toolkit for implementing efficient and scalable big data stream processing solutions in the context of industrial IoT.

3. CHALLENGES AND MOTIVATIONS :

The proposed architecture is designed to overcome key challenges in developing a reliable and efficient system for Industrial Internet of Things (I-IoT) applications. The primary challenges addressed by the architecture revolve around the lack of orchestrated configuration, monitoring and decision-making systems for edge platforms, and the complexities associated with data-driven decision-making in industrial IoT environments.

3.1. Lack of Orchestrated Configuration:

The challenge in allocating proper resources to the compute infrastructure while configuring the network for real-time streams is identified as a crucial aspect. Coordinated efforts in both network and compute resources are deemed essential for optimal performance. The proposed solution acknowledges that while this problem is well-understood in virtualized environments, achieving coordination remains challenging. To address this, the architecture advocates for the development of interfaces and frameworks that enable coherent configuration across compute and networks. The complex nature of Time-Sensitive Networking (TSN) device configuration, involving time synchronization and packet scheduling, is recognized, and the need for robust solutions to minimize errors is emphasized.

3.2. Lack of Monitoring and Decision-Making Systems for Edge Platforms:

The transformation of industrial control applications into containerized microservices introduces challenges in meeting real-time guarantees. Monitoring and analytics become crucial for gaining insights into performance and facilitating system (re)configurations. The proposed solution involves leveraging lightweight virtualization technologies and container orchestration frameworks, such as Kubernetes, to manage edge infrastructures. However, it recognizes the limitations of default monitoring and emphasizes the need for adaptability in measurement intervals, filtering, aggregation, and redirection flexibility in monitoring data. This approach ensures that monitoring and decision-making systems are well-equipped to handle the unique demands of edge platforms.

3.3. Data-Driven Decision Making for Industrial IoT:

Challenges related to ML software system life-cycle changes, up-scaling of services, and detecting anomalies such as sensor faults, network attacks, and corrupt data are identified. Limited data availability further complicates the task of anomaly detection. The proposed solution introduces MLOps as a strategy to dynamically adapt ML models to changes in the software system life cycle. Detecting model degradation using a digital twin, which relies on both historical and real-time data, is considered crucial. The architecture acknowledges challenges associated with training data drift, target or concept drift, and quality degradation in incoming data. Dynamically adapting the ML software system is identified as essential for achieving scalability and cost savings in industrial processes.

In essence, the proposed architecture tackles these challenges through a holistic approach, emphasizing the importance of coordination, adaptability, and dynamic adaptation in ensuring the reliability and efficiency of industrial IoT applications. By addressing these challenges, the architecture aims to pave the way for scalable and resilient systems that can meet the demands of the evolving industrial landscape.

4. USE CASES:

4.1. Use Case 1 - Converged Infrastructure for Two Industrial Control Processes:

Overview: In this use case, AIDA addresses the challenge of managing two distinct industrial control processes, traditionally handled through separate networks and Programmable Logic Controllers (PLCs). The primary objective is to consolidate and enhance these processes within the edge–cloud continuum.

Process 1 (P1-VU8 - Burner Management): This process involves overseeing burners in an industrial furnace. The existing setup includes PLCs, IO modules, and devices connected via Profinet switches. AIDA explores implications related to networking configuration challenges in this scenario.

Process 2 (CNC Milling Machine): This process utilizes an NC-Controller unit as a PLC, managing real-time drive control. It involves diagnosis traffic using OPC UA and monitoring data through a web browser. AIDA focuses on addressing networking and communication aspects within the CNC milling machine setup.

AIDA Implementation: The implementation involves deploying virtualized PLCs in the edge–cloud continuum. The architecture incorporates agile AI-/ML pipelines to process diverse sensor data from both processes, effectively addressing issues related to data quality and model drift. By virtualizing PLCs and leveraging AI/ML, AIDA enhances the adaptability and efficiency of industrial control processes, ensuring seamless communication and networking configurations.

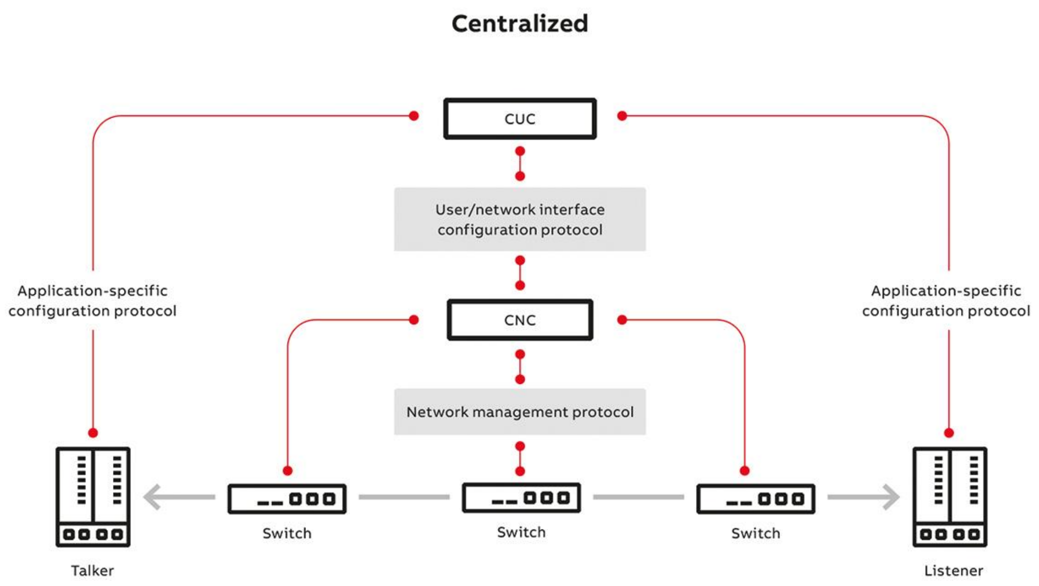

Fig. 2. Use Case Example: Dynamic Configuration Updates for TSN Streams in CNC Machine Control

4.2. Use Case 2 - Vacuum Pump Control:

Overview: This use case underscores the importance of trustworthiness in I-IoT applications, specifically focusing on AI-/ML and data quality aspects within the Electroslag Remelting (ESR) process for the steel industry.

Process: The ESR process requires vacuum pumping to create a contaminant-free environment. AIDA introduces ML forecasting software on the edge node to predict the minimum pressure within the first minute of pumping. This eliminates the 20-minute wait for quality inspection, enhancing the efficiency of the steel manufacturing process.

Data Collection: Pressure data from vacuum chamber sensors is collected every second. The ML model analyzes the initial minute of pumping data to swiftly forecast the minimum pressure. This proactive approach allows for early detection of potential issues.

AIDA Benefits: The early detection of pumping issues facilitated by the ML model leads to valuable time savings in the steel manufacturing process. If the ML model predicts pressure above the required threshold, pumping is stopped early, triggering the inspection procedure. This not only improves the overall efficiency of the ESR process but also contributes to cost savings and quality control.

5. The Architecture for Trustworthy Data-Driven I-iot Applications

In the rapidly evolving landscape of Industry 4.0, the convergence of Artificial Intelligence (AI) and the Industrial Internet of Things (I-IoT) has become paramount for organizations seeking to optimize their operations. The Architecture for Industrial Data Applications (AIDA) is a meticulously designed solution tailored specifically for data-driven I-IoT applications, providing a unified network architecture with real-time capabilities. This article delves into the intricate details of AIDA, highlighting its three core components: the converged network infrastructure, the edge–cloud continuum, and the AI-driven framework.

1. Converged Network Infrastructure:

1.1 AIDA's Utilization of Time-Sensitive Network (TSN) Elements:

AIDA strategically employs TSN elements to establish reliable connectivity between edge–compute nodes and industrial IoT sensors. At the heart of this infrastructure lies the Centralized Network Controller (CNC), leveraging Software-Defined Networking principles to automate global network configurations. The Configuration and Usage Collector (CUC) plays a pivotal role in collecting relevant information, empowering the CNC to seamlessly synthesize and deploy configurations.

1.2 Edge Nodes as Dynamic Compute Fabric:

The edge nodes within AIDA function as a dynamic compute fabric, deploying a diverse set of microservices encompassing both industrial and networking applications. Orchestrated by Kubernetes, the Centralized Edge Node Controller (ENC) manages the control plane, ensuring the smooth operation of the cluster. Fine-grained monitoring mechanisms are implemented to guarantee persistent and efficient operation across the edge nodes.

2. Software-Defined Network Configuration for TSNs:

2.1 CNC Architecture:

The CNC architecture comprises three main building blocks: interfaces, a configuration database, and internal subsystems (TSN and Operational). Centralized User Configuration facilitates interactions with CNC and TSN end-points, while the Network Optimizer optimizes forwarding configurations. Other Control-Plane Entities manage inter-controller communication.

2.2 Centralized Network Configuration:

The Operational Sub-System includes topology discovery, sync tree, network provisioning, and network monitor entities. The TSN Subsystem encompasses path, resource allocation, and main entities. CNC utilizes central storage for information, and microservices communicate using gRPC-based APIs, ensuring seamless coordination in the network configuration.

3. Edge Compute Monitoring Architecture:

AIDA's monitoring architecture supports end-to-end runtime performance monitoring within the edge compute cluster. This includes Measurement, Delivery, Fusion, Storage, Visualization, and Provisioning/Orchestration services. Telegraf handles metric collection, Apache Kafka ensures reliable data transfer, and Grafana is employed for visualization. The Provisioning service manages deployed microservices and triggers network configuration through CUC.

4. Cloud and Edge-Centric Software Architecture for Industrial IoT:

4.1 Adaptable ML Pipelines:

AIDA's adaptable ML pipelines efficiently manage two distinct types of data: application data sourced from IoT sensors and monitoring/telemetry data originating from the edge infrastructure and networking devices.

4.2 Data Flow within TSN Network to Edge Node:

Orchestrated microservices, covering Use Cases, Delivery and Collection Services, and Action Services, operate on edge nodes, facilitating real-time decision-making. ML pipelines play a crucial role in evaluating data quality, conducting predictive analyses, and executing actions based on the generated output.

4.3 Data Flow within Edge Node to Cloud:

Cloud services within AIDA handle prolonged processes, including automated retraining of ML models. Critical tasks such as real-time data quality assessment, ML software testing, and automated retraining and redeployment of ML components transpire within the cloud infrastructure.

In conclusion, the Architecture for Industrial Data Applications (AIDA) offers a comprehensive and integrated solution for data-driven I-IoT applications. Through its converged network infrastructure, software-defined network configuration for TSNs, edge compute monitoring architecture, and cloud-edge-centric software architecture, AIDA aims to streamline automation processes, enhance decision-making, and pave the way for a trustworthy and efficient industrial ecosystem. As industries continue to embrace digital transformation, architectures like AIDA play a pivotal role in shaping the future of data-driven industrial applications.

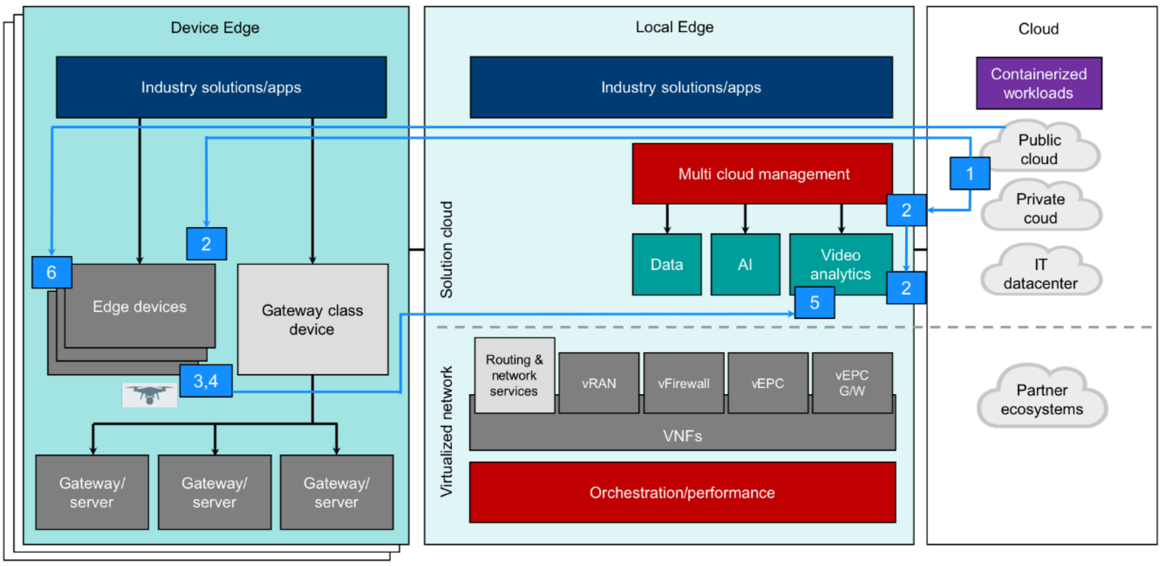

Fig. 3. Detailed Architecture for Cloud and Edge-Based Software, Highlighting Containerized Components

6. PRELIMINARY FINDINGS

In the preliminary findings section, the paper presents a comprehensive assessment of the AIDA architecture through experiments focusing on network reconfiguration, edge observability, and machine learning (ML) adaptation for scalable industrial processes.

6.1. Network Reconfiguration:

The evaluation of AIDA's capacity to dynamically reconfigure Time-Sensitive Networking (TSN) elements for accommodating new real-time streams involves fetching application requirements, constructing configuration requests, and utilizing the Control Node Controller (CNC). The experiment onboards a new CNC milling machine onto an existing converged network infrastructure. Network reconfiguration, initiated by the Configuration Update Controller (CUC), is based on edge node configuration engine requirements. Latency and jitter histograms before and after reconfiguration reveal a slight increase in latency for some frames while maintaining overall performance. This demonstrates AIDA's ability to adapt and reconfigure the network dynamically, a critical aspect for accommodating evolving industrial processes.

6.2. Edge Observability:

AIDA's Container Monitoring System (CMS) and Container Monitoring Agent (CMA) are highlighted in the experiment for monitoring edge-based applications. The focus is on proactive fault detection beyond default Kubernetes monitoring policies. AIDA extends native Kubernetes monitoring, actively monitoring the AIDA Cluster Coordinator (AIDA-CC) for fault detection and recovery. A case study illustrates the detection and recovery of an unnoticed application fault (App-1) by default Kubernetes monitoring. The low overhead performance monitoring of CMA is assessed, showing minimal impact on CPU and memory resources. This emphasizes the effectiveness of AIDA's edge observability components in ensuring robust and proactive fault detection while maintaining resource efficiency.

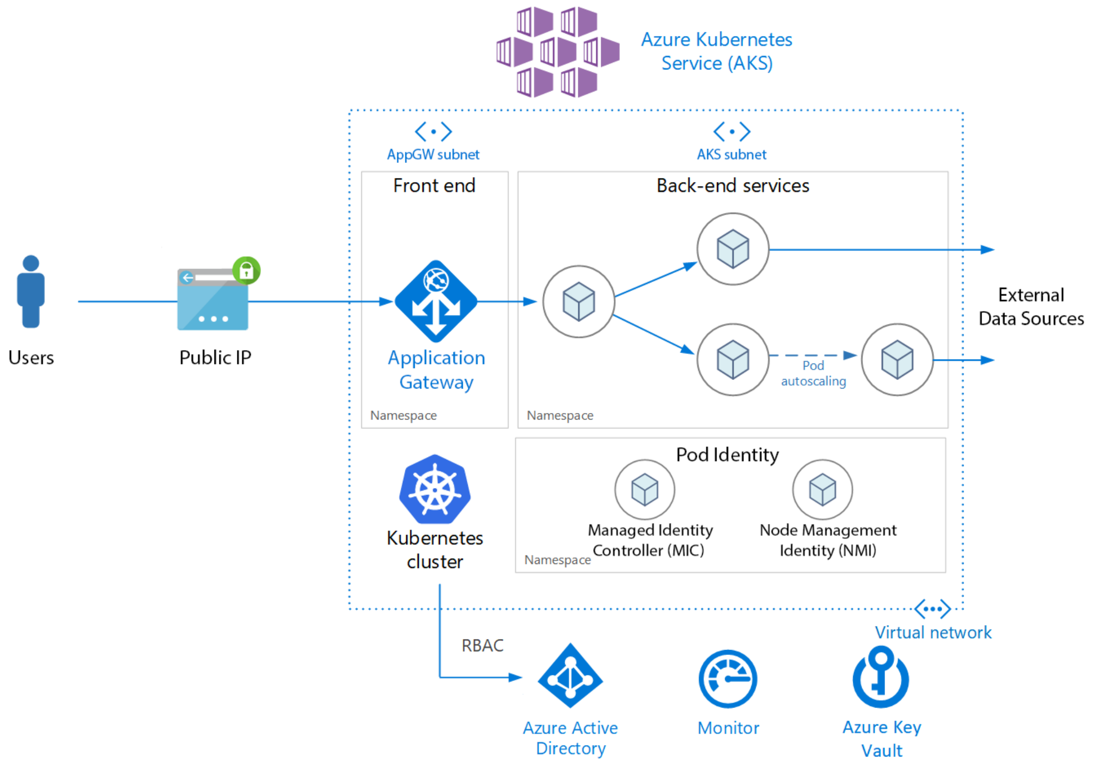

Fig. 4. Setup for Verifying Monitoring Architecture Using a Simplified AIDA Version to Address Use Case Application Faults Unhandled by Kubernetes

6.3. ML Adaptation for Scalable Industrial Processes:

This section delves into ML adaptation with the addition of a new furnace to an industrial process using AIDA. An adaptive ML strategy is implemented to address the challenge of limited historical data for the new furnace. The drift handling approach, specifically importance weighting (IW), is applied to tackle distribution shift between domains. Two decision tree algorithms, XGBoost and random forests, are employed, showcasing consistent improvement in the Mean Absolute Percentage Error (MAPE) rate. The results indicate that employing the importance weighting technique leads to a significant reduction in error rates for both the existing and new furnaces. This underscores AIDA's effectiveness in handling ML adaptation challenges and improving the reliability of decision-making in the face of evolving industrial processes.

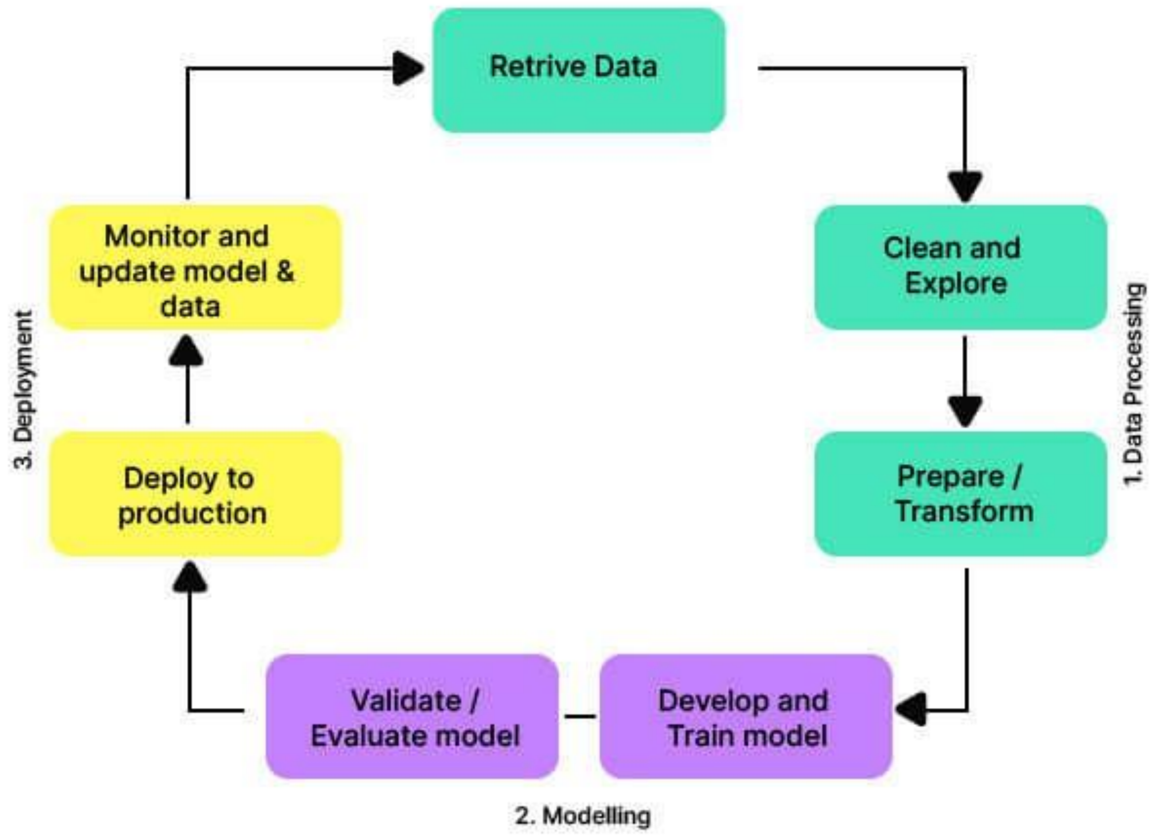

Fig. 5. Workflow of Adaptive Machine Learning (Adaptive-ML) Service

7. SUMMARY AND CONCLUSIONS

The AIDA architecture addresses the core requirements of industrial IoT applications, emphasizing robustness through the integration of three key components: a converged real-time network infrastructure, an industrial edge–cloud continuum, and flexible, reliable ML components. The architecture leverages a Time-Sensitive Networking (TSN)-based network, providing a foundation for real-time data streams. The software-defined control plane ensures seamless and optimized reconfiguration, facilitating dynamic adjustments to the network structure. This adaptability is essential for meeting the stringent demands of industrial environments, where real-time responsiveness and fault tolerance are paramount.

In experimental use cases, AIDA demonstrates its versatility by supporting real-time data streams and showcasing dynamic network reconfiguration capabilities. Notably, the architecture surpasses standard Kubernetes monitoring in fault detection, highlighting its superiority in identifying and addressing potential issues promptly. This is a critical aspect in industrial settings where any downtime or disruption can have significant consequences. The convergence of real-time networking and ML components within the AIDA architecture positions it as a comprehensive solution for industrial IoT applications.

The adaptive ML system integrated into AIDA plays a pivotal role in managing missing data within scalable industrial processes. This capability enhances the reliability and efficiency of data-driven decision-making in complex industrial environments. As the architecture is put to the test through practical use cases, its adaptability and robustness in handling diverse challenges become apparent, reinforcing its suitability for industrial IoT applications.

In the realm of future developments, the AIDA architecture aims to further refine its capabilities. This includes modeling and optimizing the cost associated with network reconfiguration, enhancing fault detection metrics within the monitoring architecture, and fine-tuning ML-based algorithms for data quality. Additionally, the architecture seeks to achieve a more comprehensive integration of its three pillars – real-time networking, edge–cloud continuum, and machine learning – to create a holistic solution that addresses the evolving needs of robust industrial IoT applications. As AIDA continues to evolve, it holds the potential to set new standards for reliability and adaptability in industrial IoT infrastructures.

REFERENCES

[1] Muhammad Usman, Simone Ferlin, Anna Brunstrom, Javid Taheri, A survey on observability of distributed edge & container-based microservices, IEEE Access (2022) 86904–86919.

[2] Buvaneswari Ramanan, Lawrence Drabeck, Thomas Woo, Troy Cauble, Anil Rana, pb&j - easy automation of data science/machine learning workflows, in: 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 361–371, http://dx.doi.org/10.1109/BigData50022.2020.9378128.

[3] A. Villa-Henriksen, G.T. Edwards, L.A. Pesonen, O. Green, C.A.G. Sørensen, Internet of Things in arable farming: Implementation, applications, challenges and potential, Biosyst. Eng. 191 (2020) 60–84.

[4] D. Glaroudis, A. Iossifides, P. Chatzimisios, Survey, comparison and research challenges of IoT application protocols for smart farming, Comput. Netw. 168 (2020) 107037.

[5] X.E. Lee, Made in China 2025: A new era for Chinese manufacturing, CKGSB Knowl. (2015).

[6] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, M. Gidlund, Industrial internet of things: Challenges, opportunities, and directions, IEEE Trans. Ind. Inform. 14 (11) (2018) 4724–4734.

[7] M.S. Islam, H. Verma, L. Khan, M. Kantarcioglu, Secure real-time heterogeneous iot data management system, in: 2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications, TPS-ISA, IEEE, 2019, pp. 228–235.

[8] S. Zhou, Y. Du, B. Chen, Y. Li, X. Luan, An intelligent IoT sensing system for rail vehicle running states based on TinyML, IEEE Access 10 (2022) 98860–98871, http://dx.doi.org/10.1109/ACCESS.2022.3206954.

[9] Mamoona Humayun, Mohammed Saleh Alsaqer, Nz Jhanjhi, Energy optimization for smart cities using IoT, Appl. Artif. Intell. (2022) 1–17.

[10] Mohamad Razwan Abdul Malek, et al., Comfort and energy consumption optimization in smart homes using bat algorithm with inertia weight, J. Build. Eng. 47 (2022) 103848.

[11] A. Chatterjee, B.S. Ahmed, IoT anomaly detection methods and applications: A survey, Internet Things 19 (2022) 100568, Publisher: Elsevier.

[12] L. Jiang, D.-Y. Liu, B. Yang, Smart home research, in: Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826), Vol. 2, 2004, pp. 659–663, http://dx.doi.org/10.1109/ICMLC.2004.1382266.

[13] L. Hughes, Y.K. Dwivedi, N.P. Rana, M.D. Williams, V. Raghavan, Perspectives on the future of manufacturing within the industry 4.0 era, Prod. Plan. Control 33 (2–3) (2022) 138–158, http://dx.doi.org/10.1080/09537287.2020.1810762.

[14] M. Hosseini Shirvani, M. Masdari, A survey study on trust-based security in internet of things: Challenges and issues, Internet Things 21 (2023) 100640, http://dx.doi.org/10.1016/j.iot.2022.100640.

[15] R. Vuorikari, S. Kluzer, Y. Punie, DigComp 2.2: The Digital Competence Framework for Citizens - with New Examples of Knowledge, Skills and Attitudes, Scientific analysis or review KJ-NA-31006-EN-N (online),KJ-NA-31006-EN-C (print), Joint Research Council, Luxembourg (Luxembourg), 2022, http://dx.doi.org/10.2760/115376(online), 10.2760/490274(print).

[16] C. Redecker, et al., European Framework for the Digital Competence of Educators: DigCompEdu, Technical Report, Joint Research Centre (Seville site), 2017.