UtilitasMathematica

ISSN 0315-3681 Volume 122 (2), 2025

Choosing the Right Data Deployment Architecture in Industry 4.0: From Sensors to Decisions

Rohith Kumar Punithavel, Graduate Student Researcher, University of the Cumberlands, Kentucky, USA, rohithkumar.punithavel@gmail.com

Thiagarajan Chidambareswaran, Researcher, Arizona State University, Arizona, USA, tchidam2@asu.edu

Abstract:
The future of manufacturing is not about producing more; it is about producing smarter and more efficiently. Industry 4.0 combines advanced technologies such as the Internet of Things (IoT), artificial intelligence (AI), and big data into manufacturing and supply chain practices. Modern machines come with smart sensors and connected systems that enable the acquisition of data that can be used to inform a wide variety of data-driven decisions, such as predictive maintenance, quality assurance, optimizing energy use, and production governing. However, the value of this data is not apparent until it has been properly processed and analyzed. Analytics, machine learning, and real-time processing frameworks transform the data into information and provide actionable decision-making ability so that firms can be forewarned of future failure, optimize resource consumption, and respond proactively to dynamic challenges of an industrial environment. This paper presents a comparative study of four primary approaches to data deployment architecture in Industry 4.0: on-prem, near-prem (edge), cloud, and hybrid, highlighting their strengths and weaknesses. Latency requirements, security, cost, scalability, and regulatory compliance are often determining constraints to help the industry choose one of these approaches. The study contributes by offering a structured evaluation of data deployment architectures and providing practical guidelines to align architectural choices with industrial needs, thereby supporting more effective adoption of Industry 4.0 strategies.

Keywords: Industry 4.0, Data deployment architectures, Data processing, Data-driven decision making, Bigdata analytics, Smart factories, Smart manufacturing.

1. Introduction:

Industry 4.0, or the fourth industrial revolution, is an important shift in industrial production and value creation systems through the convergence of digitalization, automation, and advanced connectivity technologies (Lasi et al., 2014). Industry 4.0 was first introduced in 2011 within the German federal government to establish a strategic position to enhance competitiveness through flexible and resource-efficient manufacturing systems (Frank et al., 2019). Building upon the legacy of past industrial revolutions like the 18th-century mechanization, 19th-century electrification, and the 20th-century digital automation, this Industry 4.0 revolution incorporates cyber-physical systems, the Internet of Things (IoT), cloud services, and big data analytics, allowing for real-time monitoring, decisions, and optimization across the product lifecycle (Ghobakhloo, 2019).

Industry 4.0 focuses on interoperability and extensive systemic integration, leading to smart factories composed of intelligent machines, smart human input, and innovative products communicating autonomously. Industry 4.0 comprises a layered technological architecture, which includes the base technologies of IoT, Cloud computing, and data analytics in relation to the front-end technologies of Smart Manufacturing, Smart Products, Smart Supply Chains, and Smart working practices. Smart Manufacturing leverages machine-to-machine communication, automation technology, and predictive maintenance to improve efficiency and flexibility. Smart products leverage connectivity and sensors to facilitate service-based business models. (Lasi et al., 2014; Frank et al., 2019; Ghobakhloo, 2019).

Industry 4.0 is characterized as a digitalized and networked manufacturing environment. Data is the most central idea of this paradigm and a significant resource in this industrial revolution due to its abundance, non-rivalrous, and reusable nature. Data is used in manufacturing optimization, predictive maintenance, product development, and novel business models. Industry 4.0 technologies prepare and structure data for use, while value-creating technologies apply processed data to improve decision-making, enhance efficiency, and enable new business models (Klingenberg et al., 2019). With this centrality of data, Industry 4.0 has opened doors to a diverse range of applications that extend beyond efficiency improvements toward transformative industrial practices.

Smart factories utilize the Industrial Internet of Things (IIoT) to connect machines, sensors, and human operators and enable real-time monitoring, autonomous decisions, and adaptive scheduling of resources. Data-driven manufacturing employs machine-learning algorithms to predict equipment failures and minimize downtime to extend machines longevity. Process optimization utilizes real-time analytics to reconfigure parameters, where waste is minimized, and productivity is enhanced. Digital twins are computer models that represent real-time replicas of physical systems for use in simulation, monitoring, and decision-making throughout all phases of a product lifecycle. Advanced analytics and artificial intelligence allow for self-configuration, adaptive resource scheduling, and mass customization as enabled technologies through additive manufacturing and collaborative robotics. Industry 4.0 technologies have begun to access warehousing and logistics, energy management, healthcare, agriculture, and autonomous delivery drones. These applications demonstrate that moving from enabling technologies to value-creating technologies generates improved operational efficiencies, new business models, and increased sustainability as waste and resources are used more efficiently (Contreras et al., 2017; Tikwayo & Mathaba, 2023; Zheng et al., 2020).

Industry 4.0 primarily involves aligning physical machinery, digital infrastructures, and decision support systems in a layered structure. Smart factories have four layers: physical resource, network, data application, and terminal. The physical resource layer represents intelligent sensors, reconfigurable controllers, and modular robotic systems for real-time video monitoring and adaptive changes during production. The network layer of the smart facility connects various IIoT, IWSNs, and SDNs. The data application layer uses cloud computing, edge computing, big data analytics, domain ontology models, and local context models. These models can convert manufacturing data into actionable items. Terminal systems allow operators, managers, and customers to access, monitor, and control aspects of production, resulting in a manufacturing facility that can quickly react to change while providing corporate transparency (B. Chen et al., 2017). This paper discusses system implementation strategies, including on-premises, near-premises (edge), cloud-based, and hybrid models, and their advantages and disadvantages. The diagram below is inspired from Chen et al. (2017) research.

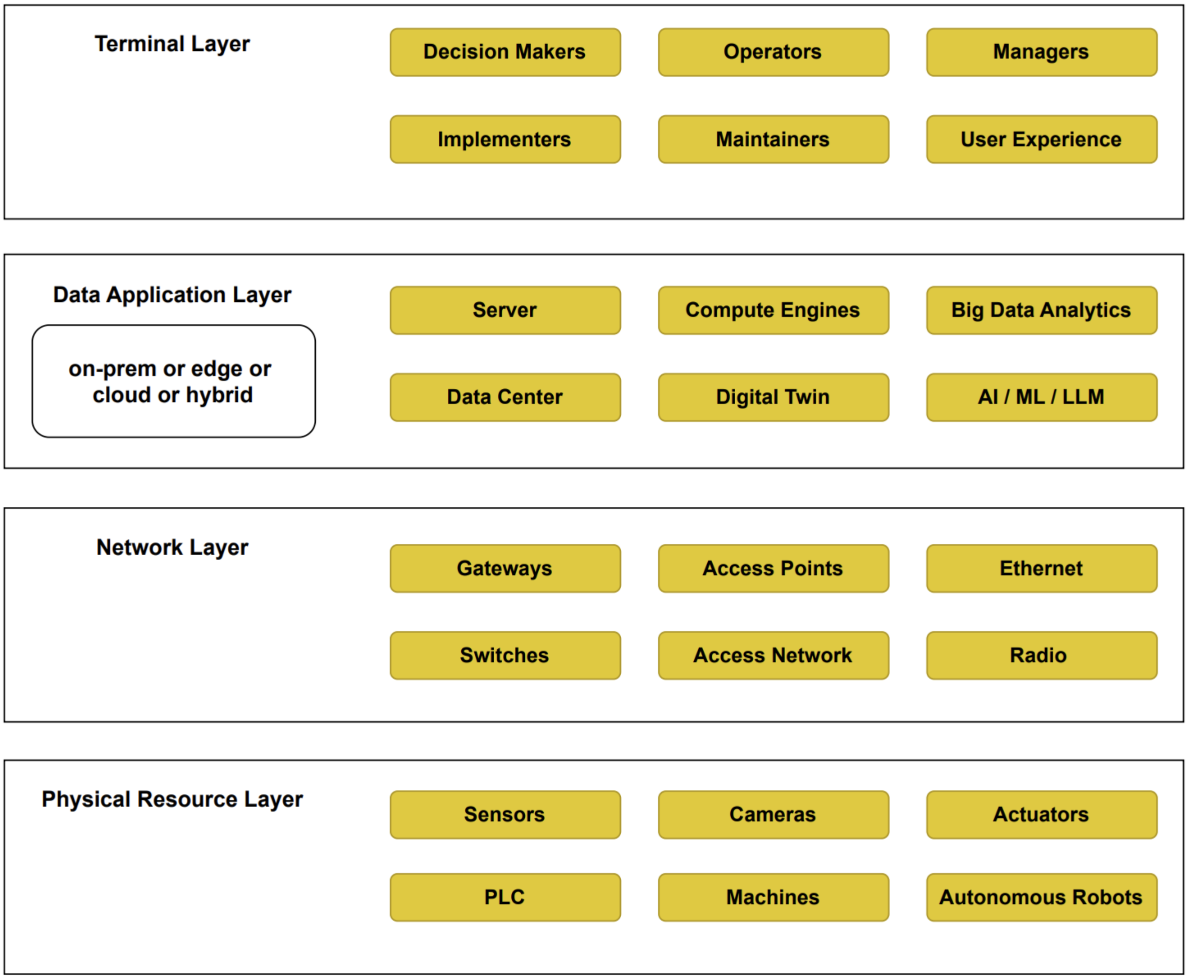

Figure 1 Hierarchical Architecture of Industry 4.0

The research areas of Industry 4.0 are vast, and most focus on one type of architecture deployment. The decision makers know of various deployment options but lack a structured framework to understand their alignment, overlap, or divergence in practice. This paper systematically evaluates deployment architectures by positioning them against one another for direct comparison and clarity. The goal is to examine each model individually and create a cohesive framework that supports decision-makers in making informed decisions.

2. Data Deployment Architectures

2.1. On-prem Deployment

On-premises deployment in Industry 4.0 means hosting data centers, processing servers, and analytic tools physically inside the facilities of a company. In this architecture, all data collection, processing, and storage occur inside the organizational boundary and do not pass over the Internet to the outside world. This architecture makes sense in a manufacturing enclosure, where lots of operational data needs to be ingested in real-time from sensors, machines, and control systems. So, on-prem solutions allow organizations to manage latency-sensitive applications and increase security controls.

Compared to other deployments, the primary advantages of on-premises are low latency, security, and regulatory oversight. By processing the data in the same location, round-trip delays in communication are minimized, which is ideal for real-time needs such as security monitoring applications, etc. In on-premises deployments, sensitive manufacturing data does not necessarily need to transit over the Internet, which reduces exposure to external threats. Organizations can also better implement strong cybersecurity regimes, based on monitored perimeters and hardened edge calling between control systems. Compliance is also a consideration, especially in industries such as pharmaceuticals, defense, and critical infrastructure, as many regulatory regimes are prescriptive in providing data sovereignty, where audit controls are generally strict (Rodenbeck et al., 2025).

On-premises deployment, while offering strong security and control, can have scalability, costs, and maintenance challenges. Industry 4.0 increases data volume and processing needs, so when scaling, those needs must include purchasing new servers, adding storage, and increasing power and cooling expenses. New infrastructure for on-premises deployment is costly, but there are also limits on physical space within facilities. On-premises systems must go through careful planning and procurement cycles, be installed, and add more capacity to production environments. Therefore, logistically, on-premises systems cannot scale elastically in near real-time. As the urgency increases between the capabilities scaled in the on-premises model, it becomes increasingly complex for organizations to add capacity and scale quickly to keep pace with their growing data and computational requirements. A study completed by Medina and Mindoro, 2023 found that the total cost of ownership of the on-prem infrastructure can increase over the years due to staffing, electricity, and hardware refresh cycles. Latency may also become an issue with the on-prem solution when remote workers or multiple users try to access the on-prem server concurrently. A significant failure of on-prem infrastructure is the possible interruption of staff who have become too reliant on the in-house IT team to apply updates, patching, and respond to incidents. Consequently, while the on-prem approach offers top security and control over resource commitments and delivery, the cost-equivalent incurred usually produces additional short- and long-term flexibility and operational expenditure disadvantages (Medina & Mindoro, 2023).

Given the strict regulatory and compliance requirements, on-premises deployment is important for Industry 4.0 within the pharmaceutical industry. This is clearly due to the requirements established by the pharmaceutical manufacturers, where data integrity, traceability, and system validation are paramount. The manufacturer can monitor their production lines in real-time with IoT-enabled devices by deploying on-premises data centers. To process sensor data without latency, many biopharmaceutical companies have deployed servers inside cleanrooms and controlled environments to allow predictive maintenance activity, less downtime, and quicker identification of quality deviations. Since sensitive clinical and production data remain inside the organization, the reduced risk of data exposure to a third-party provider can also reduce risk during compliance audits.

2.2. Near-Premises Deployment

Near-premises deployment, often called edge deployment, is where IT resources for processing and storage are in physical proximity to the industrial environments but not entirely within the premises of factories, plants, and manufacturing environments. Instead of sending all raw data to distant cloud servers, edge or near-premise data centers are established in nearby hubs. These edge computing nodes pre-process the data streams generated directly from machines, sensors, and industrial IoT devices and return decisions or filtered datasets to the production environment. In this computing environment, edge computing decision nodes provide much lower latencies than cloud architectures (Sittón-Candanedo et al., 2019; Willner & Gowtham, 2020).

A related concept within near-premises deployment is fog computing. Fog computing is a distributed computation layer that allows for the real-time filtering, aggregation, and analysis of data streams generated by an industry. Fog computing is particularly important with mission-critical operations and can be executed locally with minimal latency. Fog computing can help to conserve bandwidth and migration costs, as well as help make organizations more compliant with data-protection regulations that require keeping sensitive information within local or regional bounds. Fog computing is a valuable enhancement for near-premises edge deployments in Industry 4.0. It serves as an additional abstraction layer, allowing organizations to experience the benefits of speed, cost-effective operation, and due diligence with regulations (Aazam et al., 2018).

Near-premises architectures are specifically designed for industrial use cases that require quick responses. For example, within a manufacturing line, sensor data could be on temperature or the vibration of machinery, and the fastest reaction time could be required to ensure safety interlocks, robotic coordination, or real-time quality checks. Furthermore, the use of near-premises servers as an architecture allows for processing of data at times and locations where the data is processed, such as minimizing the amount of sensitive, operational data sent across a wide-area network, lowering risks of interception, and helping enterprises comply with strict data-protection regulations. Near-premises architectures are also cost-efficient and can be achieved through managing bandwidth and less reliance on centralized management, which means the overall solution is cheaper over time. Additionally, near-prem nodes reduce migration costs by managing and filtering redundant or irrelevant data over the network. Near-premises deployments have a modular feature that allows adding more edge nodes in scenarios where there is a need to support higher production capacity or an increased number of devices. This capability enables the near-premise nodes to adapt to workloads that emerge from fluctuating production or increased demand for a product or service. Among several potential benefits of near-premises architectures for industrial usage, regulatory compliance can be another benefit. It offers a technical mechanism to ensure local data residency by allowing only sensitive data to potentially be processed and stored at the edge, regardless of the other issues associated with cross-border data transfers and data laws (Trinks & Felden, 2018; Qiu et al., 2020; Willner & Gowtham, 2020).

Near-premise or edge solutions have several drawbacks, including the expense of setting them up and keeping them going, security vulnerabilities at the edge, limited scalability compared to cloud, and issues with regulatory environments concerning a distributed architecture. Near-premise deployments require significant capital expense for the setup or to rent resources like the servers, power, network infrastructure, and maintenance staff, which will only increase the cost of operating the edge. Near-premises data centers are great for localized, time-sensitive work. However, more computationally demanding work requires finding an alternative path. It is also considerably complicated to maintain consistent compliance across multiple provinces. Depending on the number of edge deployments, each must comply with the local data-handling regulations in each jurisdiction of the edge node. This shortcoming will require a governance framework for the different regions and an understanding of cross-border legal issues (Trinks & Felden, 2018; Qiu et al., 2020; Willner & Gowtham, 2020).

Smart factories leverage high-speed cameras and sensors to find micro-defects in products, creating large and time-sensitive data streams that could introduce unacceptable latency. The cost, power, and solution management requirements for setting up systems to identify product micro-defects become challenging within the factory. Near-premises edge deployment can mitigate latency by processing the data locally at edge hubs near the function. Quality-control algorithms on the edge servers can process sensor inputs in milliseconds and trigger immediate action, such as removing defective products from the production line. This highlights the importance of near-premises architecture in enabling safe and efficient processing, ensuring data privacy, and meeting regulatory requirements for location, storage, and handling industrial data (Trinks & Felden, 2018; Qiu et al., 2020).

2.3. Cloud Deployment

Cloud deployment in Industry 4.0 perspective refers to utilizing centralized cloud infrastructures for the storage, processing and analysis of industrial data. In this architecture, information or data generated from factory machines, IIoT devices, and production systems is transmitted over wide-area networks to remote vendor-hosted data centers for further processing. Cloud deployment as a concept, is particularly compelling in Industry 4.0 as it offers virtually unlimited scalability, global accessibility, and access to advanced data services, such as digital twins, machine-learning frameworks, predictive maintenance and AI model training capabilities that are otherwise difficult to implement locally (Dritsas & Trigka, 2025).

The primary advantages of cloud deployment include scalability, cost efficiency, and collaboration across distributed facilities. Contrasting from on-premises architectures, Cloud platforms provide the organizations to dynamically operate on a pay-as-you-go model, enabling them to avoid the capital expense of building and maintaining local data centers, while still being able to elastically increase or decrease computing resources as needed. With help of centralized data environment, cloud architecture facilitates collaboration between geographically separated sites, where global predictive maintenance, enterprise-wide production planning, analysis and decision-making can be achieved (Routray, 2025). It also provides the organizations with IT management services such as automated backup, disaster recovery, and high-availability frameworks. In addition to single-cloud solutions, several organizations have successfully implemented multi-cloud deployment architecture which provides flexibility by distributing workloads across different providers.

Though Cloud deployment has wide range of advantages over other methods, it also carries significant challenges limiting its applicability in certain industrial contexts. Latency is a major challenge of cloud deployment also presents key challenges that restricts its applicability in certain industrial contexts where data is required to be transmitted to remote servers for processing. It’s also not suitable in case of real-time responsive required applications such as robotic coordination, motion control, or multilevel safety interlocks (Jamil et al., 2024). Unlike On and near premises deployment architectures, Cloud deployment does not meet regulatory requirements for data sovereignty and traceability in industries like defense, healthcare, and pharmaceuticals. The movement of sensitive production data outside national or organizational boundaries, leads to legal challenges when information must be transferred or stored beyond designated jurisdictions (Dritsas & Trigka, 2025). Since a large volume of data is transmitted and stored off-site, it increases exposure to cyber threats. Hence data security is always a concern. A solution requires a strong encryption and monitoring frameworks. Another drawback is dependence on third-party providers, which creates the risk of vendor lock-in. Where switching providers will become costly, complex, and disrupt operations.

Cloud deployment has proven valuable in large-scale industrial analytics and enterprise-level applications within Industry 4.0. A well-known example is predictive maintenance, where sensor data such as vibration, temperature, acoustic signals, and even computer vision outputs from multiple production sites is aggregated in the cloud. This centralized data aids in training of more accurate fault-detection models, which can be deployed back to factory floors to optimize maintenance scheduling, in-turn reducing downtime, and possibly extend equipment lifespan (Dritsas & Trigka, 2025). Another area where cloud deployment plays strong impact is in supply-chain optimization. It allows the organizations to forecast disruptions, adjust scheduling and improve overall responsiveness by integrating production data with logistics and market information across geographically dispersed facilities (Jamil et al., 2024).

In response to the previously mentioned disadvantage of vendor lock-in, some industrial platforms have successfully adopted multi-cloud strategies, by distributing workloads across multiple providers. This approach not only reduces dependency on a single vendor but also supports regional data residency requirements. Cloud-based systems further enhance collaborative decision-making with help of unified datasets and analytics across the enterprise. This improves information sharing, cost efficiency, and long-term productivity, even though scalability and privacy concerns remain (Atobishi et al., 2018). These examples illustrate that cloud deployment is ideal for compute-intensive, collaborative and data-driven applications where ultra-low latency is not primary requirement.

2.4. Hybrid Deployment

Hybrid deployment in Industry 4.0 refers to an architectural design where industrial data processing is distributed across local (on-prem or near-prem) systems and centralized cloud platforms. In such an arrangement, industrial sensors, machines, and controllers send data in continuous streams, which are effectively processed at local systems, limiting data latency, enabling near real-time responses to events such as equipment failures, safety issues, and quality deviations. At the same time, the cloud layer also presents a large amount of virtual space for storing data, an ability to aggregate and integrate cross-factory data, and analytical capabilities such as machine learning and optimization. Split use cases of processing and storage, such as this, is an exchange between low-latency control systems and those associated with high-level strategic decision-making. Hybrid deployment is particularly relevant in Industry 4.0 as new-age factories require deterministic real-time control, yet they must be flexible and scalable. Local systems are essential to implement time-sensitive tasks such as safety interlocks, process adjustments, or predictive warnings. At the same time, clouds aggregate data to make decisions for global optimization, workforce scheduling, supply chain forecasting, and regulatory reporting. This dual-method system allows factories to be action-oriented in real time while adapting to the future. With hybrid approaches, factories can respond to short-term needs while considering serious issues such as security, operations cost, scaling, and regulations (Salis et al., 2023; Yang et al., 2020).

Hybrid deployment in Industry 4.0 presents several benefits. By processing sensitive data locally rather than using an external cloud service, hybrid deployment also lowers the risk and probability of risk exposure because it allows for more effective response times to important events. Local systems (on-premise or near-premise) devices filter the raw sensor streams that may encounter latency in external cloud services. This significantly lowers the raw stream transfer costs needed for cloud services. Cloud deployment offers elasticity that supports scaling in a multi-factory or global setup. Hybrid implementations offer the dual benefits of flexibility with cloud solutions and ensure local data compliance and regulatory policies are followed for local system-based implementations (Salis et al., 2023; Yang et al., 2020).

Hybrid deployment in Industry 4.0 has several disadvantages, including latency variability, security complexity due to endpoints increased attack surface, dual infrastructure cost, and compliance management overhead. Edge resources can introduce unpredictable delays, while maintaining local hardware and cloud services can increase operational complexity and lifecycle costs. Deciding which data to keep local or centralized can also increase administrative burden (Verba et al., 2018).

A food packaging plant gathers and processes sensor data in real-time using local computing nodes, either on-premises or near-premises, to ensure and verify high-quality seals for food safety and regulatory compliance. The edge system meets strict latency requirements and works according to compliance and regulatory policies. The industry needs to store data and train its decision model to make the process more accurate. With bigger datasets, cloud-based systems can help retrain quality detection models, gradually increasing their accuracy. This scalability lowers costs and saves time as each factory no longer needs to host its own high-performance computing and storage systems. The hybrid system ensures that the industry complies with food safety regulations by guaranteeing prompt safety measures, preserving long-term audit trails, and sharing data. The hybrid approach also ensures that immediate safety-critical actions remain under local, low-latency control while transforming quality control into a global process (Salis et al., 2023; Yang et al., 2020).

Table 1 Comparison of Data Deployment Architectures in Industry 4.0

Parameter	On-prem	Near-prem	Cloud	Hybrid
Location of Processing	Inside company facilities	Local edge nodes / regional hubs near facilities	Remote vendor data centers	Split between local systems (on/near-prem) and cloud
Latency	Very low: real-time capable	Low: in milliseconds, suitable for time-critical tasks)	Higher: WAN transmission delays	Variable: local ensures low latency, cloud adds some delay
Scalability	Limited: requires new hardware & space	Moderate: modular edge nodes, but still resource-bound	High: elastic, virtually unlimited	High: cloud scalability + local control
Cost Model	High CapEx: hardware, power, staff	Medium: High CapEx + OpEx for edge setup & maintenance	Pay-as-you-go: OpEx cost efficient for scaling	Dual cost: local + cloud infrastructure
Security	Strong local control, reduced exposure	Vulnerable at distributed edge nodes	Dependent on provider, higher exposure	Complex: must secure both cloud & local endpoints
Regulatory Compliance	Strong: data stays in-house, sovereignty ensured	Good: local residency possible, but compliance across multiple regions is complex	Weak: data often crosses borders, sovereignty issues	Balanced: local for compliance, cloud for aggregation

3. Bridging Industry 4.0 and 5.0 through Data Deployment Models

Industry 5.0 is a transition towards a human-centric, sustainable, and resilient path that can be supported through production systems. In simpler terms, Industry 5.0 stresses collaboration between humans and intelligent machines. To make Industry 5.0 effective, scalable data infrastructures must process vast amounts of information while ensuring transparency, trust, and adaptability. Industry 5.0 focuses on the transparency and accountability of automated systems (Gadekallu et al., 2024).

On-premises deployment is a dominant model and gives organizations a tighter rein on data sovereignty and compliance in regulated industries such as pharmaceuticals, defense, and healthcare. This nature of deployment builds trust. On-premises deployment aligns well with Industry 5.0's vision of augmented human capital, rather than displacing human capital altogether. Near-premises or edge deployment becomes more important in Industry 5.0 because it facilitates real-time human and machine collaboration. Edge computing can support time-sensitive applications like robotic coordination, safety interlocks, and customized assembly tasks. The integration of fog computing further strengthens this approach, allowing for localized filtering and analysis while still complying with local regulations. However, Industry 5.0 presents new issues, such as continued governance, interoperability, and transparency regarding legal and technical environments (Gadekallu et al., 2024; Verba et al., 2018).

In Industry 5.0, cloud deployment will now have a different role, based on its scalable nature, ability to provide access to a global community, and ability to capture large volumes of data from numerous disparate sectors. The cloud can improve production processes through more sophisticated and enhanced algorithm training, with the long-term implications of waste reduction, energy conservation, and achieving sustainability objectives. However, latency and security issues continue to constrain real-time control tasks. The intended role of cloud deployment in Industry 5.0 is more towards optimizing long-term approaches, knowledge exchange, and development, supporting innovation across ecosystems that create and utilize technology across geographical borders. Hybrid architectures capture the essence of Industry 5.0, striking a balance between real-time reactive capability (at the edge) and long-term intelligence. This duality allows the organization to meet its immediate human safety and compliance requirements by operating locally. At the same time, globally based intelligence provides knowledge of what is happening elsewhere as it drives strategic decisions. At the same time, hybrid architectures embody the resilience and adaptability of Industry 5.0 by allowing an organization to continue pursuing innovation even when it cannot conduct business as usual, allowing it to maintain operations through disruptions (Gadekallu et al., 2024; Verba et al., 2018).

4. Conclusion:

Industry 4.0 has made data a crucial resource for production, influencing organization's effectiveness, resilience, and competitiveness. This paper investigated four primary data deployment architectures: on-premises, near-premises, cloud, and hybrid. On-premises deployment models provide high levels of control and regulatory safety but insufficient flexibility and long-term scalability to build resilient operations. Near-premises deployed data permits low-latency decisions but introduces additional challenges with governance and general maintenance. Cloud provides elasticity, extensibility to advanced analytics, and collaboration, but latency sensitivity and domestic sovereignty issues constrain it. Hybrid deployment allows organizations to combine local capabilities to respond swiftly and perform benchmarking with global intelligence. Still, it also means that organizations would incur dual infrastructure costs and an increased administrative burden. The comparative study demonstrates that no architecture is best by default. The data deployment architecture choice must align with the primary objective and context of the industry, regulatory environment, and organizational priorities. On-premises functions well in highly regulated situations, time-sensitive environments benefit from edge deployments, cloud works well in more data-driven ecosystems globally, and hybrid will work in dynamic industries. The study provides insight for industrial managers and technology strategists in aligning their deployment architectures with organizational goals. It emphasizes that the deployment architecture may need to evolve alongside organizational needs, regulatory landscape, or the transition to Industry 5.0.

Looking ahead, Industry 5.0 reframes these architectures within a human-centric, sustainable, and resilient paradigm. The rapid emergence of human and machine collaboration increasingly requires data deployment models to support transparency, trust, and adaptability. Hybrid and edge solutions appear well-suited to facilitate highly interactive real-time human and machine interaction. In contrast, cloud solutions continue to support sustained and long-term innovation and optimization at scale. Ultimately, the evolution in deployment architectures will determine the efficiency of the industrial system and to what extent Industry 4.0 and Industry 5.0 will be able to deliver on their promises of smarter, safer, and more sustainable manufacturing.

References:

Aazam, M., Zeadally, S., & Harras, K. A. (2018). Deploying Fog computing in industrial internet of things and industry 4.0. *IEEE Transactions on Industrial Informatics*, *14*(10), 4674–4682. https://doi.org/10.1109/tii.2018.2855198

Atobishi, T., Gábor, S. Z., & Podruzsik, S. (2018, March 5). Cloud Computing and Big Data in the Context of Industry 4.0: Opportunities and Challenges. *IISES Annual Conference, Sevilla, Spain*. ISBN 978-80-87927-45-8. https://doi.org/10.20472/IAC.2018.035.004

Chen, B., Wan, J., Shu, L., Li, P., Mukherjee, M., & Yin, B. (2017). Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges. *IEEE Access*, *6*, 6505–6519. https://doi.org/10.1109/access.2017.2783682

Contreras, J. D., Garcia, J. I., & Diaz, J. D. (2017). Developing of industry 4.0 applications. *International Journal of Online and Biomedical Engineering (iJOE)*, *13*(10), 30–47. https://doi.org/10.3991/ijoe.v13i10.7331

Dritsas, E., & Trigka, M. (2025). A survey on the applications of cloud computing in the industrial internet of things. *Big Data and Cognitive Computing*, *9*(2), 44. https://doi.org/10.3390/bdcc9020044

Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, *210*, 15–26. https://doi.org/10.1016/j.ijpe.2019.01.004

Gadekallu, T. R., Maddikunta, P. K. R., Boopathy, P., Deepa, N., Chengoden, R., Victor, N., Wang, W., Wang, W., Zhu, Y., & Dev, K. (2024). XAI for Industry 5.0 -Concepts, Opportunities, Challenges and Future Directions. *IEEE Open Journal of the Communications Society*, 1. https://doi.org/10.1109/ojcoms.2024.3473891

Ghobakhloo, M. (2019). Industry 4.0, digitization, and opportunities for sustainability. *Journal of Cleaner Production*, *252*, 119869. https://doi.org/10.1016/j.jclepro.2019.119869

Jamil, M. N., Schelén, O., Monrat, A. A., & Andersson, K. (2024). Enabling Industrial Internet of Things by Leveraging Distributed Edge-to-Cloud Computing: Challenges and Opportunities. *IEEE Access*, *12*, 127294–127308. https://doi.org/10.1109/access.2024.3454812

Klingenberg, C. O., Borges, M. a. V., & Antunes, J. a. V., Jr. (2019). Industry 4.0 as a data-driven paradigm: a systematic literature review on technologies. *Journal of Manufacturing Technology Management*, *32*(3), 570–592. https://doi.org/10.1108/jmtm-09-2018-0325

Lasi, H., Fettke, P., Kemper, H., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering*, *6*(4), 239–242. https://doi.org/10.1007/s12599-014-0334-4

Medina, J. R., & Mindoro, J. (2023). On-premise File Server Vs Cloud Storage With Incident Management: a Comparative Study. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4645107

Özdemir, V., & Hekim, N. (2018). Birth of Industry 5.0: Making Sense of Big Data with Artificial Intelligence, "The Internet of Things" and Next-Generation Technology Policy. *OMICS a Journal of Integrative Biology*, *22*(1), 65–76. https://doi.org/10.1089/omi.2017.0194

Qiu, T., Chi, J., Zhou, X., Ning, Z., Atiquzzaman, M., & Wu, D. O. (2020). Edge Computing in Industrial Internet of Things: architecture, advances and challenges. *IEEE Communications Surveys & Tutorials*, *22*(4), 2462–2488. https://doi.org/10.1109/comst.2020.3009103

Rodenbeck, S., Gough, E., Sangeetha, A. M. K., Ashish, Ahlawat, M., Ragavan, V. K. K., Muthukumar, A., & Ahmad, A. (2025). Providing On-Prem GenAI Inference Services to a Campus Community. In *Practice and Experience in Advanced Research Computing 2025: The Power of Collaboration* (pp. 1-4). https://doi.org/10.1145/3708035.3736039

Routray, S. K. (2025). Cloud and edge computing for industry 4.0. *IT Professional*, *27*(4), 48–53. https://doi.org/10.1109/mitp.2024.3523697

Salis, A., Marguglio, A., De Luca, G., Razzetti, S., Quadrini, W., & Gusmeroli, S. (2023). An Edge-Cloud based Reference Architecture to support cognitive solutions in Process Industry. *Procedia Computer Science*, *217*, 20–30. https://doi.org/10.1016/j.procs.2022.12.198

Sittón-Candanedo, I., Alonso, R. S., Rodríguez-González, S., Coria, J. a. G., & De La Prieta, F. (2019). Edge Computing Architectures in Industry 4.0: A General Survey and comparison. In *Advances in intelligent systems and computing* (pp. 121–131). https://doi.org/10.1007/978-3-030-20055-8_12

Tikwayo, L. N., & Mathaba, T. N. D. (2023). Applications of Industry 4.0 Technologies in Warehouse Management: A Systematic Literature Review. *Logistics*, *7*(2), 24. https://doi.org/10.3390/logistics7020024

Trinks, S., & Felden, C. (2018). Edge Computing architecture to support Real Time Analytic applications : A State-of-the-art within the application area of Smart Factory and Industry 4.0. *2021 IEEE International Conference on Big Data (Big Data)*, 2930–2939. https://doi.org/10.1109/bigdata.2018.8622649

Verba, N., Chao, K., Lewandowski, J., Shah, N., James, A., & Tian, F. (2018). Modeling industry 4.0 based fog computing environments for application analysis and deployment. *Future Generation Computer Systems*, *91*, 48–60. https://doi.org/10.1016/j.future.2018.08.043

Willner, A., & Gowtham, V. (2020). Toward a reference architecture model for industrial edge computing. *IEEE Communications Standards Magazine*, *4*(4), 42–48. https://doi.org/10.1109/mcomstd.001.2000007

Yang, C., Lan, S., Shen, W., Wang, L., & Huang, G. Q. (2020). Software-defined Cloud Manufacturing with Edge Computing for Industry 4.0. *2022 International Wireless Communications and Mobile Computing (IWCMC)*, 1618–1623. https://doi.org/10.1109/iwcmc48107.2020.9148467

Zheng, T., Ardolino, M., Bacchetti, A., & Perona, M. (2020). The applications of Industry 4.0 technologies in manufacturing context: a systematic literature review. *International Journal of Production Research*, *59*(6), 1922–1954. https://doi.org/10.1080/00207543.2020.1824085