pelion

The power of unity: How the Edge Gateway can deliver an applications enablement ecosystem

The need to increase efficiency and performance of and return on large-scale industrial and manufacturing production processes has never been more evident.

Global competition is only exacerbating traditional pressure points, and the events of 2020 have proven just how volatile overly-intricate supply chains can be. Meaningful, actionable data-driven insights are the bedrock of progressive strategies to optimize existing assets, extract maximum value, and integrate emerging technologies.

One sure way to add value is to reinvigorate existing systems with the judicious addition of complementary solutions. No organization relishes a bold rip-and-replace approach; it's invariably disruptive, depreciates the equity of existing assets, and rarely delivers entirely on the promise. Therefore, selectively adding, rather than taking away, will provide a more predictable and deterministic result.

Just as the technology itself is evolving, so too is the service space. Increasingly, organizations seek to redefine what it takes to create and deliver a product. Vertical integration empowers businesses to exercise significantly greater control over the end-to-end development and manufacturing process; an integrated supply chain is an essential part of the transition, delivering all-important continuity.

However, faced with the ever-present reality of finite resources, a shift in focus to greater integration and verticalization means that companies may longer be able to resource non-core functions as was once possible.

In the industrial context, all indications are that computing services will continue to migrate from the cloud to the on-site edge, driving a corresponding consolidation around a unified, intelligent edge gateway solution. With the vast majority of existing downstream devices lacking direct connectivity, added to the security concerns of such a large potential attack surface making direct cloud connectivity too risky, the edge gateway plays an essential intermediary role. Additionally, an ever-increasing amount of device data will be locally gathered, ingested (or redacted), processed, with inferences gleaned and actioned; sources put this figure at as much as 75% by the year 2025, delivering on the many drivers for enhanced autonomous operations.

The Dark Side of Tech

It's the less glamorous aspect of the technology industry: the fact that organizations, large and small, operate mission-critical systems that are hanging on by a thread. The obligation that many businesses face to sweat their technology assets means that even when viable, state-of-the-art replacements become available, prevailing economic realities mean that many will not be in a position to upgrade to the alternatives.

Legacy systems abound: Most IT departments will have at least one of these aberrations in their portfolio, a system too vital to decommission yet too fragile to maintain and modernize proactively.

And then there's the added factor of legacy systems, where no one dares speak too loudly in their presence, let alone contemplate patching or updating them.

Most IT departments will have at least one of these aberrations in their portfolio, a system too vital to decommission yet too fragile to maintain and modernize proactively. Everyone turns a blind eye and hopes that nothing goes wrong, at least not on their watch. It might be the obscure yet crucial order-picking system in a logistics environment or a vital OSS component for a retail enterprise. They are tech's equivalent of the Sword of Damocles.

Such is the reality of the average enterprise's complex and multi-faceted technology infrastructure that's evolved over time. No environment starts from zero, from a clean sheet, and very rarely do we see genuine greenfield. All are a hybrid combination of various technologies, acquired over time, across generations, and stitched together in the context of budget constraints and changes in policy and personnel. Brown(field) is indeed the new black. It was forever thus, and maybe it will always be so.

And, when viewed from the manufacturer's perspective, the problem takes on an additional complication. The parallel themes of "sweating an existing asset" and "if it ain't broke, don't fix it" are potent inhibitors to change and innovation. Therefore, a new approach is required. This dichotomy highlights where IoT and, more specifically, intelligent edge IoT can play a pivotal role in delivering an essential oversight capability that complements legacy operational technologies (OT).

The Third Way

Now, more than merely delivering protocol translation for downstream OT sensors and actuators, edge IoT provides a critical enabling capability: application distribution, orchestration, and management. With this, manufacturers and operators can complement existing systems - those too expensive to rip-and-replace or too brittle to re-engineer - with overlay solutions that add a layer of insight and operational integrity. Comprehensive edge-based applications enablement empowers feature-rich and innovative applications. The follow-on effects deliver solution diversity, negate the reliance on cloud connectivity and processing, enhance resilience and operational robustness, reduce latency, and lay the groundwork for localized machine learning and artificial intelligence advances. Intelligent edge IoT has the potential to change everything.

Seamlessly distributed localized versions of cloud-native applications can be applied to multiple real-world use-cases. Crucially, for the industrial, manufacturing, and commercial sectors, this includes many of the most common and business-critical:

- Telematics (both original and aftermarket)
- Asset tracking
- Condition-based monitoring
- Usage-based insurance
- Inventory management
- Environmental monitoring and safety compliance

Taken as a whole, these use-cases represent the lion's share of current and projected future use of IoT within these sectors. Interestingly, there is increasing cross-over and interplay between IoT use-cases, and we see this playing-out with examples such as condition-based monitoring and usage-based insurance.

When it comes to maintaining major plant and equipment, the traditional approach of pre-defined scheduled maintenance has proven to be unnecessarily intrusive and costly. Condition-based monitoring - together with the associated reporting - takes a more nuanced and individualized approach, ensuring that proactive maintenance is delivered only as, where, and when needed. IoT sensors collect all necessary data points, with these analyzed against aggregated baselines and detailed failure models. These are the finely-tuned actuarial tables of the maintenance world. Of course, the sweet spot is to truck-roll while efficiency levels remain acceptable but as close as possible to projected sub-optimal performance or failure.

Usage-based insurance functions much like its private auto insurance equivalent: provided operators run their machinery per the manufacturer's guidelines, using only trained personnel, do not exceed the defined duty cycle, and service it correctly, insurance premiums will be lower. However, this model only works when IoT sensors are in place to collect the necessary data that validates compliance.

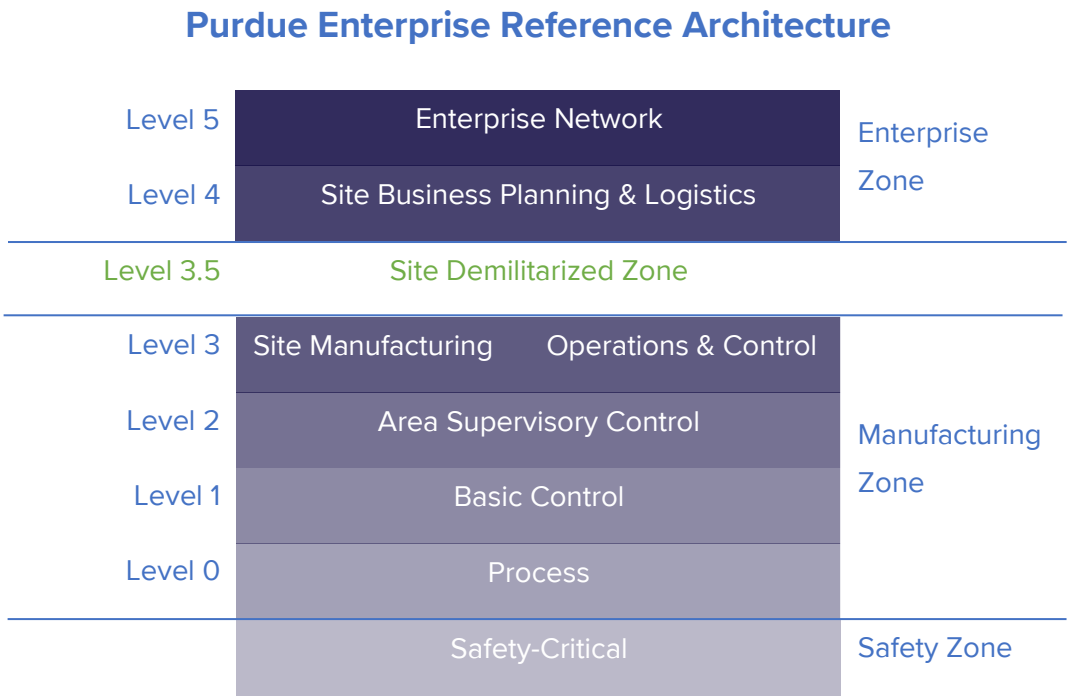
Both these use-case innovations also highlight an important transition: the need to move the application environment from closed and proprietary to open and collaborative. The third-party maintenance provider needs real-time data to precisely time their truck-roll, while the insurance underwriter needs similar direct, unfiltered access to usage data. Relying on the manufacturer or operator to accurately collect and promptly share actionable data is problematic from a timing and transparency perspective. In many situations, this approach will not deliver the necessary speed or agility. Hence, the requirement to create an application ecosystem that works for both in-house and third-party applications.

However, establishing an open application ecosystem is not a trivial matter. There are significant considerations to address; integrating with the entrenched enterprise architecture and defining an appropriate security framework top the list.

Integrating IoT with CIM

In the context of computer-integrated manufacturing, a model known as the Purdue Enterprise Reference Architecture (PERA) provides a structured methodology for delineating roles and responsibilities. Although initially developed before the widespread availability of cloud connectivity, PERA has proven itself highly adaptable to evolution, the Internet revolution, and continues to play a role, influencing, at the very least, major industrial and manufacturing organizations' enterprise architectures.

As we experience increasing integration of emerging IoT solutions with traditional Operational Technology (OT), the PERA model provides a well-understood means of positioning and contextualizing these new components. For example, merely referring to a device as an "IoT Edge Gateway" does little to advance our knowledge of its architectural significance. However, adding the relevant PERA definition provides much-needed context. For example, "Level 2" tells us that the component provides functions related to Control Systems - including the supervision, monitoring, and control of physical processes - and "Level 3" defines Manufacturing Operations Systems that deliver management of production workflows, perform batch management, record data, and manage operations and plant performance.



Indeed, the traditional PERA model has evolved to recognize a ubiquitous Internet. Additionally, the reality of Internet access has necessitated the introduction of a "Level 3.5", which is used to denote the De-Militarized Zone (DMZ), a security demarcation point between the on-site and outside worlds, defining those systems and components that are and are not directly accessible from the Internet. Cybersecurity is an obvious, non-negotiable must; however, the Level 3.5 DMZ does create challenges in introducing third-party applications and data sharing. Here, an intelligent edge IoT solution delivers unique capabilities that bridge the seemingly conflicting demands of simultaneously securing and enabling access to lower-level system elements, particularly applications.

Several standardization frameworks have emerged that seek to build on the PERA model and provide specific guidance to secure the systems and connectivity deployed in industrial networks. One example is the International Electrotechnical Commission's IEC 62443 specification, a multi-faceted approach to industrial networks and communications' security posture.

The standard describes industrial cybersecurity's technical and process-related aspects and defines specific roles for manufacturers, operators, and integrators (the providers of integration and maintenance services). Each role defines and implements risk-based assessments designed to prevent and manage security risks for industrial automation and control systems (IACS).

IEC 62443 takes a wide-ranging approach to industrial cybersecurity, including so-called maturity levels and security levels. Product development maturity levels (1 through 4) range from a basic "initial" through "managed", "defined", and - ultimately - up to "improving." Products are said to be "improving" when there are demonstrable levels of continuous improvement in terms of the process metrics that deliver effectiveness and performance. In terms of security, the various levels of 0 through 4 indicate the preparedness and resistance to progressively more advanced cyber-attack classes, considering the need to protect the specific system component and the attacker's sophistication, domain-specific knowledge, and motivation. Level 0 defines that no protection is required or provided, all the way through to Level 4, which protects against highly motivated attackers with significant levels of domain knowledge, using sophisticated means, and leveraging extensive resources.

Processes, systems, and products used in industrial automation environments can be certified to comply with IEC 62443 or part thereof. Organizations may wish to obtain certification merely as part of an internal best-practice program, or certification may be necessary to comply with a customer's or partner's contractual requirement.

Crucially, for those organizations seeking formal accreditation, comprehensive offerings such as Pelion Device Management Edge creates a path to compliance for the relevant sections, such as Part 4-2 which refers to "Technical Security Requirements for IACS Components." Additionally, to be deemed compliant, individual system components must also satisfy the Common Component Security Constraints (CCSC).

Unifying the IoT Edge

Digitally ambitious organizations - especially those in the industrial, manufacturing, and commercial sectors - need an on-site edge computing solution that enables various use-case applications, interfaces with multiple systems, monitors for suspicious or anomalous events, and maintains operating software across a range of downstream system components. This suite of requirements makes a unified solution essential; operators expect more of the gateway than the traditional OT/IoT device management role. It is now also the focal point for significant functionality. Any serious offering needs to combine extensive downstream device life cycle management, comprehensive gateway system management, and - increasingly - the all-important secure applications enablement capability.

Device Management

Full-featured device management warrants a detailed paper in its own right; however, in the context of an edge gateway solution, suffice to say that there is a need to support both legacy OT sensors and actuators - typically through protocol translation services - and their genuine IP-based IoT siblings. In terms of protocol translation, this needs to extend above-and-beyond basic one-to-one protocol mapping to capabilities such as device discovery and state awareness; indeed, an effective solution requires a complete protocol translation engine that is robust, flexible, and extensible. Necessary device management functionality also includes seamless onboarding, secure activation, an agile first-to-claim ownership process, extensive resource management, proactive health monitoring, diagnostics, and - of course - ongoing software bug fixing and security patching.

Gateway Systems Management

In terms of the gateway systems management function, here, we are concerned mainly with managing and maintaining the edge gateway itself. The requirements include monitoring performance, alerts, resource utilization, logs, diagnostics, and status and availability. An important characteristic, given the edge gateway's network placement, is to provide a secure method for remote access. Typically, edge gateways are implemented at Level 2 in the PERA model - behind and protected by the perimeter firewall at Level 3.5 - and connectivity is needed to facilitate administration of gateway software, features, and network and communication configurations. APIs that automate remote management capabilities will streamline the process across a fleet of gateways and significantly ease the burden.

Again, it's vital to securely maintain the edge gateway's complete software stack, including bootloader, kernel, operating system packages and not merely the user space.

Applications Ecosystem Management

Proceeding to the main event - applications enablement at the edge - and, as previously discussed, there appears to be an inexorable move towards increasing device data processing on-site, at the IoT edge.

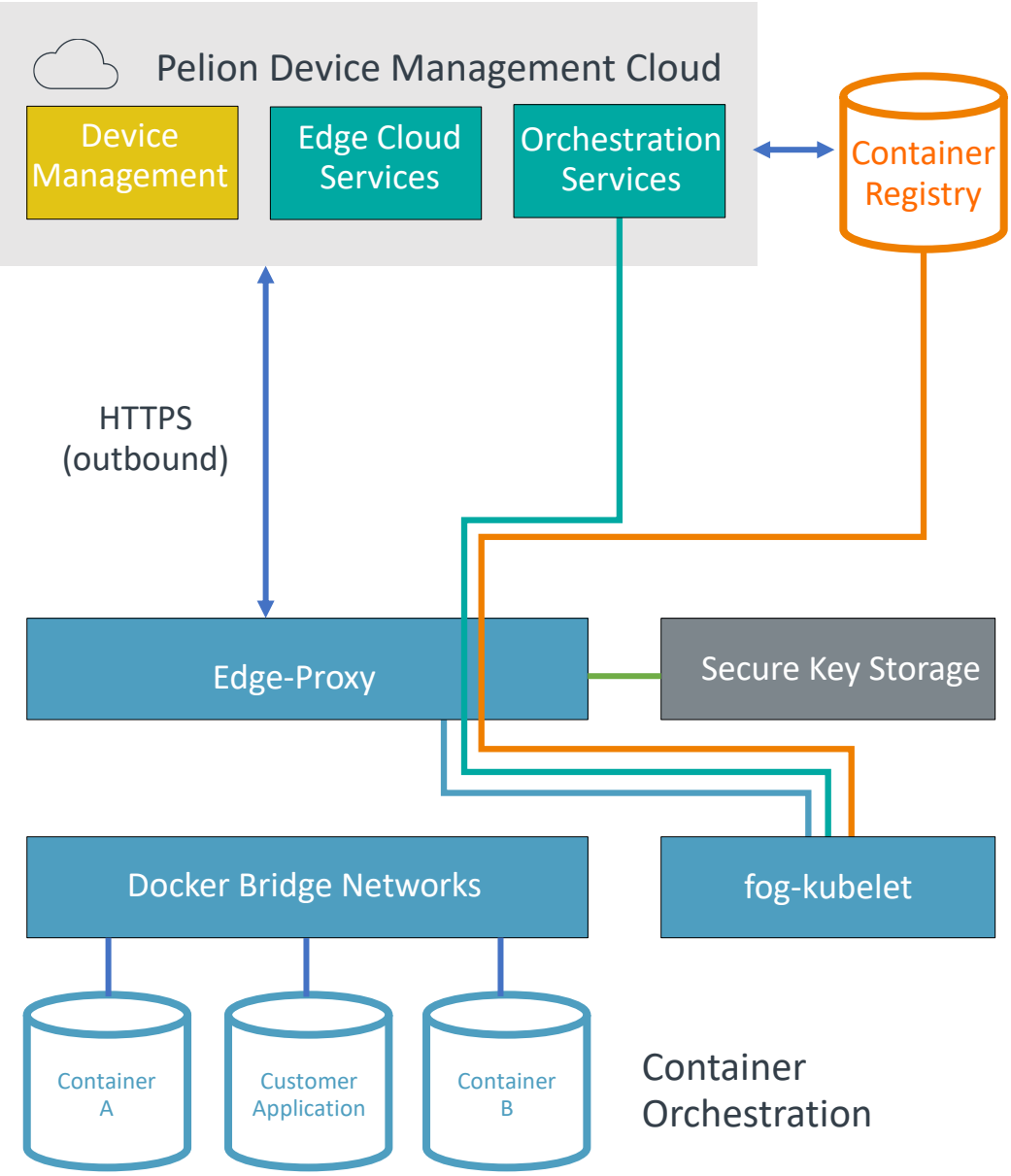
The operational drivers for this trend are many and growing:

- **Keeping relevant data on-site** and sharing it between local applications enables and accelerates critical operational decision-making.
- **Locally redacting data improves the efficiency** of DataOps and reduces the overall data transport and storage burdens.
- **Keeping both the data and its processing local** facilitates the application of machine learning, or even artificial intelligence.
- **Reducing the need to transport all traffic to the cloud** has a positive impact on bandwidth requirements, latency, and large-scale cloud-based server demands.
- **Removing the total reliance on cloud-based processing** allows local application autonomy, mitigating loss of availability or production downtime caused by unreliable or limited cloud connectivity.
- **And finally, opening up the applications ecosystem** to trusted partners and other third-parties sets the scene for true innovation.

At this point, it's worthwhile pointing out that it's the establishment of a framework supporting cloud-native applications that make the deployment of an on-site applications ecosystem feasible. Being cloud-native ensures that applications retain most if not all of their original characteristics and development equity. This commonality is vital when seeking to leverage existing developer skillsets and accelerate deployment and implementation. Notably, developers can usually implement cloud-native applications without the need for extensive changes to core code or APIs, and leveraging a common base also maximizes the return on existing integration tools and techniques. In short, existing DevOps investments can be fully leveraged and exploited.

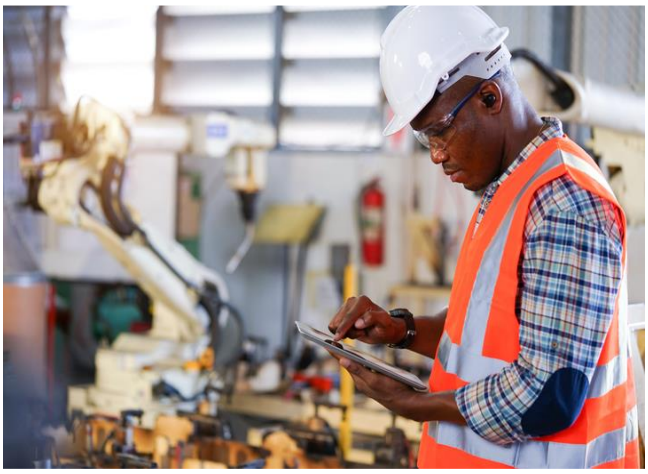
Next, a container-application registry function plays an essential role in managing the process of authorizing, distributing, and maintaining applications. The registry policies those applications that individual gateway systems are permitted to install and operate, and interaction with the appropriate key storage component maintains software integrity.

Essentially, the registry is a distribution platform for containerized applications running on gateway devices, holding diverse images as tagged versions, and acting concurrently as both storage and content delivery system. The registry establishes a single place from which owners/operators manage images, allowing them to deploy containers and securely manage their applications reliably. Security of the registry itself is critical, and solutions should implement a unified system for authentication and access control.



Working in unison, the store of cloud-native applications and the container-application registry deliver the core service: the ability to host a series of cloud-based applications securely and readily deploy selective applications to individual gateways. Simple enough to say, but of course, these things are never that easy, and there are many functionality requirements and implementation considerations.

The Small Print



Linux containers have become the standard mechanism to package and distribute portable applications, and they are a natural fit for gateway-based devices. For server-side cluster management and application deployment frameworks, Kubernetes, the open-source, container-orchestration system, is a popular option. Pelion's view is that a Kubernetes-based container environment delivers the optimum strategy for IoT edge computing. Kubernetes is known for its flexibility; as such, it is ideal for integrating applications, legacy and emerging.

As useful as Kubernetes is, however, there's a vital need to complement its generic application deployment, scaling, and management automation functionality with a robust and extensible security framework that's fit-for-purpose in the industrial IoT context. Considerations include operating systems, logging, registration, digital certificates, and integrity checks. Security, as always, is king.

Starting with the basics, the edge gateway offering should utilize a hardened operating system (OS), reducing the surface of vulnerability and the risk of a successful cyberattack. Hardening will include removing unnecessary software, usernames, logins, and the disabling of unnecessary services or processes. The principle is that a single- or limited-function device is more secure and less exposed to attack than one supporting a broad range of functionality.

There is a slight dichotomy here - after all, we are implementing containerization to explicitly support multiple applications - and so a fine line needs to be tread. Too strict, we may limit effectiveness or too lax, and we may be exposed, inviting an attack.

Creating and maintain a system for device identity trust is vital; for example, leveraging the Bootstrap private key and certificate is a reliable method for ensuring device identity. PARSEC is open-source software that provides a standard API to hardware security and cryptographic services in a platform-agnostic way. Integrating PARSEC security software with the operating system's crypto library provides hardware-backed protection. Similarly, attestation and integrity checking is important during system boot. Consolidated and accessible via the solution's portal, these checks provide operators with a means of establishing and verifying a trusted environment.

Secure boot loading is also key to ensuring a trusted system. By default, it's possible to update a boot loader by rewriting the relevant storage partition with the updated image. This approach, however, is vulnerable to corruption and does not provide a rollback facility for proper robust updates. Separating boot loaders, creating "normal" and "secure" modes, allows us to distinguish between those components of the operating system that can run as standard - for example, the Linux kernel - and those that need to run in a secure space.

Logs provide valuable oversight of gateway and applications behavior and status. However, trust is vital; hence, solutions should enact a tamper-evident system. Ideally, this will include some form of integrity verification based upon unique crypto keys that are factory provisioned and activated during the deployment phase.

Network segmentation for individual containerized applications prevents cross-application contamination, protecting and isolating the platform and neighboring containers from a compromised application. It is also essential to partition between Internet-facing containers and device-facing processes. Additionally, a suitable access control mechanism would control API access to protocol translators.

As part of full life cycle integrity verification, the gateway management solution should include checking the verification and revocation status of digital certificates. Should certificates no longer be valid, the system would re-bootstrap and, upon completion of this process, invoke updated credentials (assuming that these are now available).

And finally, communications between the on-site edge gateway solution and the backend platform and registry would be encrypted at all times.

Summary

To date, the cloud and centralized data centers have been the epicenter of digital transformation. However, as more distributed devices connect, and IoT permeates every aspect of life, the edge will light up with intelligence and become increasingly autonomous. This transition means that data processing and insights extraction will progressively migrate from the cloud to the IoT-enabled edge, driving convergence around a unified, intelligent edge gateway solution. While centralized applications will continue, local applications will frequently dominate, running on edge systems that concurrently support legacy OT and new IoT devices.

How Pelion can help

Pelion Device Management Edge offers a unique solution for edge applications enablement and management. It is based on open-source components, leverages open APIs, and integrates well with the existing application management ecosystem and cloud-based DevOps toolchains. Security for IoT devices is critical, from hardware to connectivity and into the cloud; end-to-end security is enabled from gateway provisioning to secrets management to application-level security. Pelion's Edge offering provides a wide range of features that ensure device-to-data security, regardless of the industry, market, or application use-case, facilitating the large-scale deployment of robust IoT solutions.

Crucially, Pelion Device Management Edge establishes a unified control plane for orchestrating hybrid cloud-edge applications. Pelion Edge makes remotely deploying and managing microservices a genuine reality. Leveraging many advancements in cloud application management technologies such as rolling updates, containerization, rollbacks, health monitoring, and checkpointing, Pelion Edge opens the door to new possibilities. Owners or operators can establish an application marketplace ecosystem of their own; once approved and registered, cloud-native applications - both in-house and third-party - to be selectively deployed. The ability to seamlessly support application distribution and management is vital as organizations seek competitive advantage in an increasingly challenging economic environment, and Pelion Edge makes this happen.

Edge applications complement cloud-based analytics with real-world data acquisition and inferencing. Edge gateways consolidate and unify a broad range of IoT sensors and actuators, translate legacy protocols, and deliver secure, scalable, and reliable application distribution. Supporting scale-out deployments with a cloud-based service empowers the orchestration of legacy software and innovative cloud-native applications on edge gateway systems. Extending decision-making to the edge will improve responsiveness, operational efficiency, and increase security for IoT deployments. Specifically designed to address the requirements of demanding, digitally transformed organizations, the Pelion Device Management Edge offering is particularly appropriate for those operating in the industrial, manufacturing, and commercial sectors.

To find out about how Pelion can accelerate your business transformation,
visit our website or **request a consultation**.