Introduction to Edge Computing in IIoT

An Industrial Internet Consortium White Paper

IIC:WHT:IN24:V1.0:PB:20180618

Edge Computing Task Group

Almost every use case and every connected device focused on by the Industrial Internet Consortium (IIC) requires some sort of compute capability at its source, at the edge. Multiple sources define *edge computing* as "cloud computing systems that perform data processing at the edge of the network, near the source of the data". While this is certainly true, it only scratches the surface of the immense power and remarkable capabilities that edge computing applications and architectures can provide to solve industrial internet users' toughest challenges. But, as is typical with any powerful technology, innovative architectures and new terminology are needed to facilitate implementation, bringing increased complexity with it.

Purpose and Audience

This white paper provides practical guidance on edge computing, architectures and the building blocks necessary for an edge computing implementation. It defines edge computing architectural functions and highlights key use case considerations

Consequently, there is a need to identify:

- where the edge is,
- its defining characteristics,
- key drivers for implementing edge computing and
- why compute capabilities should be deployed at the edge in Industrial Internet of Things (IIoT) systems.

It also informs architecture and testbed teams through:

- identifying and evaluating standards, practices and characteristics best suited for addressing edge computing holistically, and highlighting gaps where needed,
- identifying deployment models and crosscutting functions that address patterns and characteristics for edge computing deployment and
- exploring and identifying extensions to the current edge computing model that expand and enhance the functionality of edge computing devices.

Several detailed use cases are highlighted, edge computing for industrial analytics is explored and security challenges for edge computing implementations are also considered. (Detailed security information from the IIC can be found in the *Industrial Internet Security Framework*.[1])

[1] "The Industrial Internet of Things; Volume G4: Security Framework," Industrial Internet Consortium (IIC), 2016: www.iiconsortium.org/IISF

Viewpoints and Crosscutting Concerns

This whitepaper targets the technical community and focuses on the functional and implementation viewpoints from the IIC's *Industrial Internet Reference Architecture*[2] (IIRA):

The *functional viewpoint* focuses on the functional components in an IIoT system, their structure and interrelation, the interfaces and interactions between them and the relation and interactions of the system with external elements in the environment to support the usages and activities of the overall system. These concerns are of interest to system and component architects, developers and integrators (see IIRA Chapter 6).

The *implementation viewpoint* deals with the technologies needed to implement functional components (functional viewpoint), their communication schemes and their lifecycle procedures. These elements are coordinated by activities (usage viewpoint) and supportive of the system capabilities (business viewpoint). These concerns are of interest to system and component architects, developers and integrators and system operators (see IIRA Chapter 7).

It identifies the attributes of edge computing related to the Edge, Platform and Enterprise Tiers described in the IIRA as well as the crosscutting functionality such as data management, connectivity, orchestration, analytics and security.

This white paper has a technical focus to identify deployment models and edge-computing implementation patterns across the several crosscutting functions. Figure 1 shows a multi-layer edge computing architecture with crosscutting functions useful in deploying edge-computing architectures as defined by the IIRA.

We fully recognize that the software driving edge computing is an essential component of edge computing implementations and architectures, but the sheer volume of operating system and platform choices options are beyond the scope of this white paper.

[2] "The Industrial Internet of Things; Volume G1: Reference Architecture" Industrial Internet Consortium (IIC), 2018. www.iiconsortium.org/IIRA

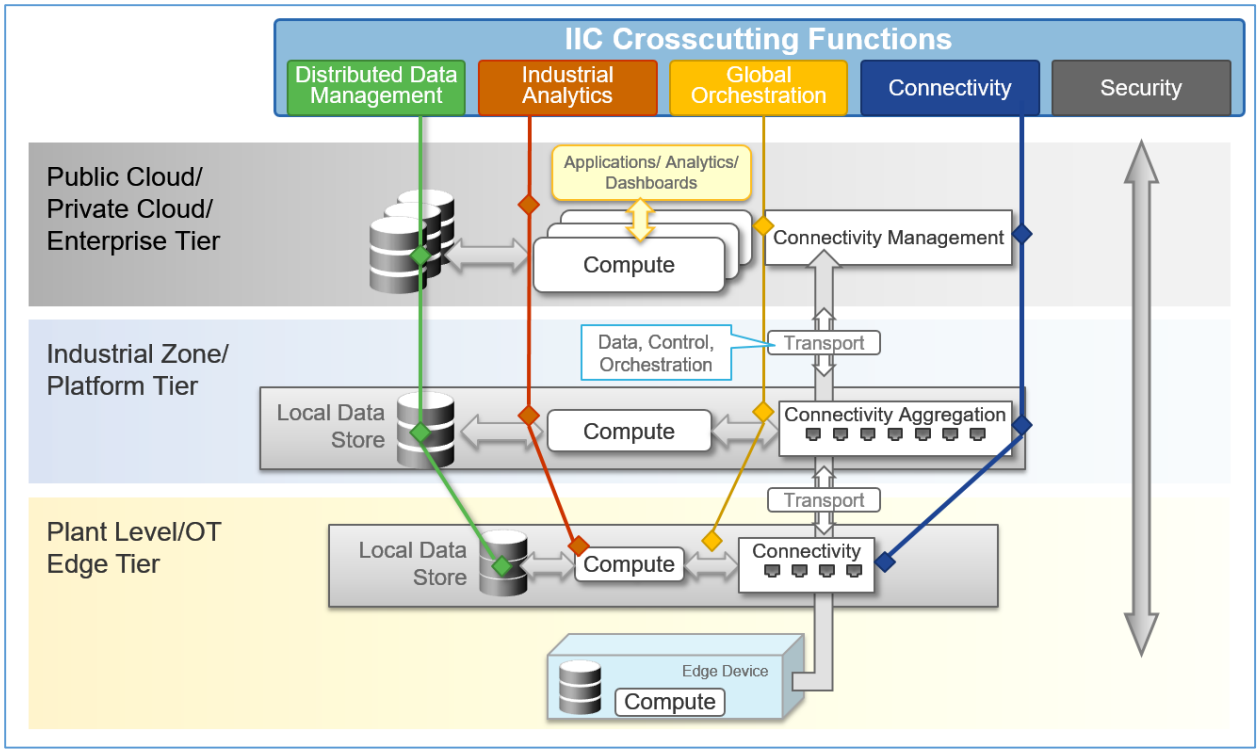

Figure 1: IIC Crosscutting Functions Across Edge Computing Architectures

Where is the Edge?

The edge is a logical layer rather than a specific physical divide, so it is open to individual opinion and interpretation of "where" the edge is. The business and usage viewpoints provide clues, while the functional and implementation viewpoints deal with the technical aspects.

From the business perspective, the location of the edge depends on the *business problem* or "key objectives" to be addressed.

> "Key objectives are quantifiable high-level technical and ultimately business outcomes expected of the resultant system..." and "Fundamental capabilities refer to high-level specifications of the essential ability of the system to complete specific major business tasks". (Ref IIRA).

There is a continuum of fundamental capabilities for an IIoT solution and "the edge" moves along this continuum based on the requirements of the problem at hand, as shown by the following examples found in typical industrial operations.

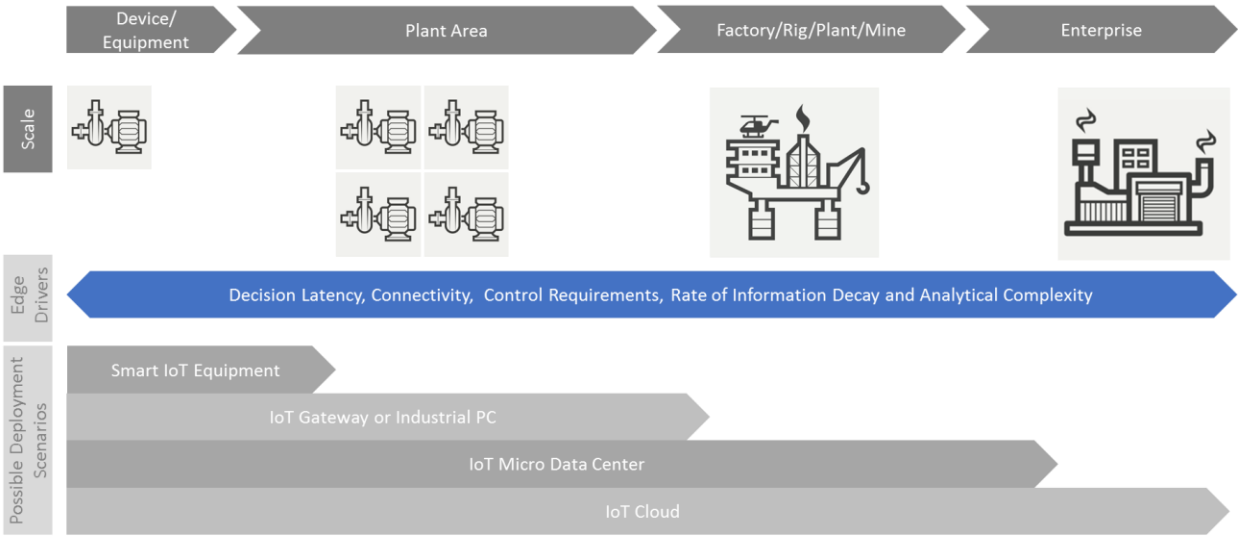

Figure 2: Edge continuum for a typical industrial environment

EXAMPLE 1: PROTECT EQUIPMENT FROM DAMAGE BY OVERHEATING

In this scenario, a "dumb" thermocouple measures temperature on a pump. A pump with edge computing capability can perform basic analytics to determine if a defined threshold is exceeded and shut the pump down in milliseconds. There is no decision latency and no need for connectivity to perform this function. Connectivity is not necessary, but it may be used for notification. The time value of the temperature information decays rapidly as delayed response can result in equipment damage. In this case the edge is at the device level as it can achieve the key objective, even if connectivity to higher-level systems and networks are interrupted.

EXAMPLE 2: MONITOR THE PERFORMANCE OF PLANT AREAS OR PRODUCTION LINES

The performance of equipment and production lines are often expressed through performance indicators like Overall Equipment Effectiveness (OEE). Near real-time analytics on multiple data points from sensors in the plant area can be processed on a local gateway and provide OEE trends and alerts to operational systems or personnel. In this case, the fundamental capability requires information from multiple equipment sources to perform simple analytics. The time value of information is high as response delays waiting for decisions from the cloud can cause significant losses. This business problem suggests that the edge is at the plant area level.

EXAMPLE 3: OPTIMIZE SUPPLY CHAIN FOR A LOCATION OR FACTORY TWICE DAILY

Optimizing supply chain processes for a local facility, factory or an oil field requires data from multiple sources at short intervals to apply optimization algorithms and analytics that will adapt supply-chain plans in business systems such as SCM or ERP. The fundamental capability requires local or factory-level connectivity with decisions made in hours. Additional information outside the perimeter of the factory may be useful, but not mandatory for effective optimization. In this instance, the edge is at the perimeter of the factory, plant or local facility.

EXAMPLE 4: PREDICT EQUIPMENT FAILURE AND SCHEDULE PROACTIVE RESPONSE

Machine learning models to predict Electric Submersible Pump (ESP) failures require data from multiple offshore platforms. The analytics models are complex and a large amount of data is needed to train and re-train the models. It also requires regular data feeds from operating ESPs to determine each unit's remaining useful life. The data from individual ESPs need to be analyzed regularly but information decay is much slower than in the other scenarios and decisions can be taken daily or weekly. Computation is typically performed at the enterprise level using a public or private cloud and is at the top end of the edge continuum.

The edge can be anywhere along the time-value graph (see Figure 3) as these examples illustrate. It is "where" data for sensors is used to achieve a specific key objective or address a specific business problem.

Why Compute at the Edge

Edge computing is a decentralized computing infrastructure in which computing resources and application services can be distributed along the communication path from the data source to the cloud. That is, computational needs can be satisfied "at the edge," where the data is collected, or where the user performs certain actions. The benefits are:

- improved performance,
- compliance, data privacy and data security concerns are satisfied and
- reduced operational cost.

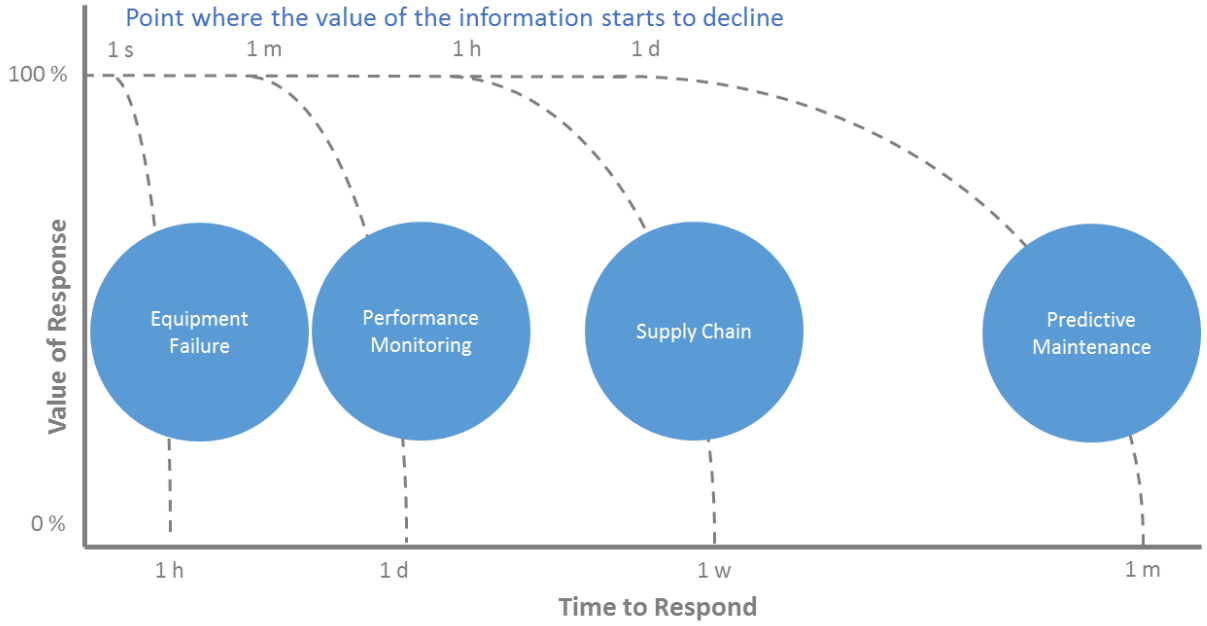

Figure 3: Time-Value curves for Edge information

We examine each in turn.

IMPROVE PERFORMANCE

The edge is not merely a way to collect data for transmission to the cloud, it also processes, analyzes and acts on the collected data at the edge within milliseconds and is therefore essential for optimizing industrial data at every aspect of an operation.

In a windfarm, for example, if wind speed or direction changes, the edge software onsite can analyze this data in real-time and adjust individual turbines to optimize overall wind farm production. Only aggregated data is sent to the cloud, reducing communication bandwidth requirements and improving data transfer time.

In addition, the turbines generate terabytes of data. Sending this data to a cloud platform to run advanced analytics may be technologically achievable, but cost prohibitive to do daily. Through edge computing, the end user can capture streaming data from a turbine and use it in real-time to prevent unplanned downtime and extend the life of the equipment while reducing the data set to a more manageable size for transmission to the cloud.

The challenge of transmitting large quantities of data in real-time cost-effectively from remote industrial sites can be mitigated by adding intelligence to devices at the edge of the network, in the plant or field. Edge computing on the device brings analytics capabilities closer to the machine and provides a less expensive option for optimizing asset performance.

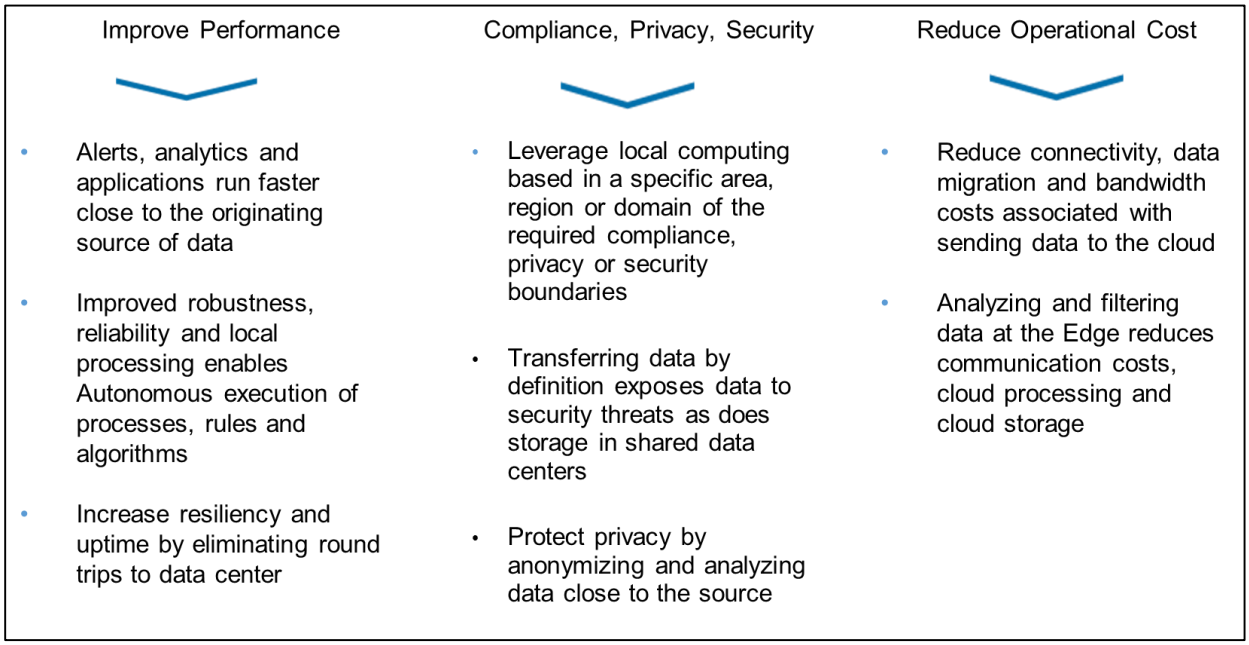

Figure 4: Benefits of Edge Computing

COMPLIANCE, DATA PRIVACY AND DATA SECURITY

Public cloud creates a long list of privacy, regulatory and compliance issues related to classified or sensitive data. Today, service providers can guarantee private access and control but at the price of being cumbersome, costly, inelastic and difficult to manage.

Edge computing allows enterprises to operate independently using a public/private cloud by using local computing based in that area, region, domain or the required local security boundaries.

REDUCE OPERATIONAL COST

Connectivity, data migration, bandwidth and latency features of cloud computing are expensive. Edge computing addresses these by reducing bandwidth requirements and latency.

If an oil and gas company drilling in Nigeria, for example, requires computing to predict oil-well production-decline rate, the alternatives are to build their own data centers (with the associated cost and scale limitations) or to use a cloud provider (where the nearest datacenter can be 5,000 miles away) with significant costs and unreliable service. With edge computing, the end user can process data in real time locally at a fraction of the cost of the public cloud, while still maintaining the flexibility that a cloud infrastructure provides.

Edge computing creates a valuable continuum from the device to the cloud to handle the massive amounts of data generated from IIoT. Processing data closer to where it is produced and at the response times required by the local applications addresses the challenges of rapidly increasing data volume. Edge computing decreases response time to events by eliminating a round trip to the cloud for analysis. It avoids costly bandwidth additions by eliminating the need to transmit gigabytes of data to the cloud. It also protects sensitive IIoT data by analyzing it locally within a private network.

Consequently, enterprises using edge computing may improve and optimize operational performance, and address compliance and security concerns while efficiently managing costs.

DIAGRAM OF EDGE COMPUTING EXAMPLES

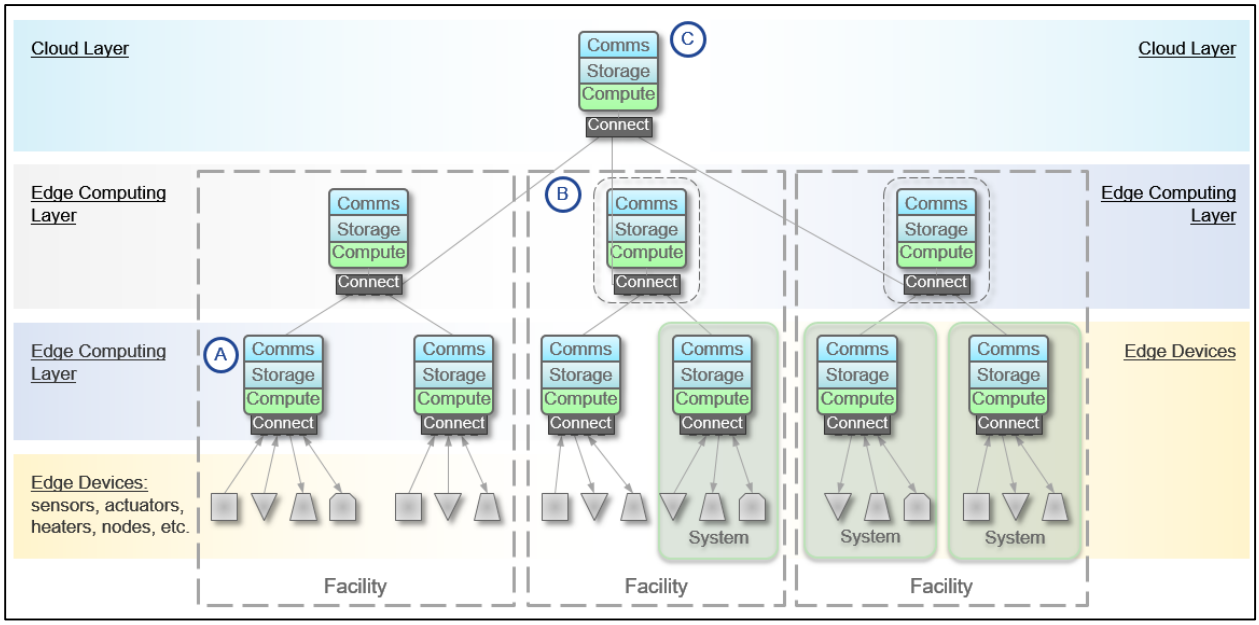

Figure 5: Edge Computing Topology

Figure 5 illustrates multiple examples of edge computing implementations based on differing business points of view. The examples progress from left to right as the edge layer becomes more complex and aggregates multiple system functions below. The computing layer moves up the architecture stack, aggregating processing capabilities, information and data from below.

The multitude of choices means there is a layered edge-cloud synergy rather edge versus cloud. Where possible, digitalization is always going to use edge and cloud synergistically in which fast and localized compute take place at the edge while global compute, model development, management and security can benefit from the "wisdom of the cloud".

Starting with an example of a simple temperature controller (see "A" on the left in Figure 5), the problem to be solved is temperature monitoring and control of a specific device or zone. In this case, the edge devices would be the thermocouple sending temperature data and the element providing the heating or cooling, and the edge computing device would be the temperature controller running the control algorithm and making the adjustments.

If the objective is to orchestrate temperature across several devices or areas, then the edge becomes the temperature controllers themselves (whether individual components or standalone systems) and the edge-computing layer becomes the system coordinating the control, typically a PLC or SCADA system ("B" in Figure 5).

If the business objective is to monitor and manage multiple geographically dispersed facilities, then the edge is each individual facility reporting its status to a compute layer in the cloud ("C" in Figure 5).

Characteristics of the IIoT Edge Computing Model

Edge computing exists vertically within the full stack from device to cloud and horizontally across IIoT subsystems. The new computing model is fully distributed and can support a wide range of interactions and communication paradigms including:

- peer-to-peer networking, for example security cameras communicating about objects within their scope,
- edge-device collaboration such as self-organizing vehicles that travel together or a community of wind turbines in remote locations,
- distributed queries across data stored in devices, in the cloud and anywhere in between,
- distributed data management, defining where and what data is to be stored, and for how long and
- data governance including quality, discovery, usability, privacy and security aspects of data.

Key Drivers: Cloud to Edge Computing

IIoT disrupts the cloud-computing model with new usage scenarios leading to these requirements:

Time sensitive: Often decisions need to be made within milliseconds while a round trip to the cloud introduces undesirable latency. Reliability and critical-path control management make it too risky to rely solely on remote logic. A good example is autonomous guided vehicles; although an anti-collision algorithm can execute in the cloud, it is best to run the algorithms at the edge.

Communication: Mobile network infrastructure tends to follow the pattern of deploying to highly populated urban areas, before trickling down to rural or remote locations. For assets that are truly remote, satellite connectivity may be the only option. This creates a paradigm where IIoT use cases for industries such as mining, oil & gas, chemicals and shipping are not well served by robust affordable communication.

Data boundary: In some applications, the data produced and consumed by devices is required by other devices only within the local area. This local data can be acquired and served with low latency by the edge to the users in the local area. Depending upon the use case, the radius of the local area can vary from a few centimeters from the device to an entire neighborhood or city. In augmented-reality scenarios, for example in smart cities, local edge infrastructures can store information about points of interest of a neighborhood. Since most of the access to the data (or consumption of the data) will be made in the same local area, there is no need to store all information in the cloud. As a truck transitions from private to public network and across sovereign boundaries both enterprise policies and local data regulation will determine what can be stored locally and what can be sent to the cloud.

Data volume: The amount of data generated by sensors can be huge. For example, hundreds of high-resolution cameras creating video streams at 30 frames per second could clog communication channels. Edge computing allows data to be processed and stored locally with only preprocessed data being transferred to the cloud.

IT/OT convergence: Historically the operational technologies (OT) that are used to manage and automate industrial equipment exist at the edge of the network while information technologies (IT) have been more centralized. Though these systems have been treated separately, there is value in having an integrated IT/OT strategy that offers:

- business data needed for interpreting or contextualizing IoT data for decision making,
- availability of both existing and new business outcomes, business models that leverage integrated data and
- standard processes to drive outcomes.

Data governance deals with quality, discovery, usability, privacy and security aspects of data. Insufficient data governance can leave a company vulnerable to major business disruptions. On the other hand, extreme data governance can stifle innovation. Edge computing helps simplify data governance by:

- reducing data clutter: high volume time-series data can be analyzed at the edge,
- refining data usability: edge computing allows data to be contextualized resulting in better usability,
- improving data privacy: security policy at the edge allows only relevant data shared with the systems up in the hierarchy and
- lowering the impact of security breach: since edge computing allows for the data storage and analysis to be federated, impact of a security breach can be contained.

Use Cases

This section describes use cases that illustrate the benefits of edge computing. Figure 6 shows logical entities residing on either the cloud or the edge, connected through WANs. When clouds were first introduced, the trend was to “shift everything into the cloud”, but, due to network latency and the cost to transmit a large amount of data, more logical tasks remained at the edge. With the improvement of the processing power and capability, the amount of tasks performed on the edge will continue to grow.

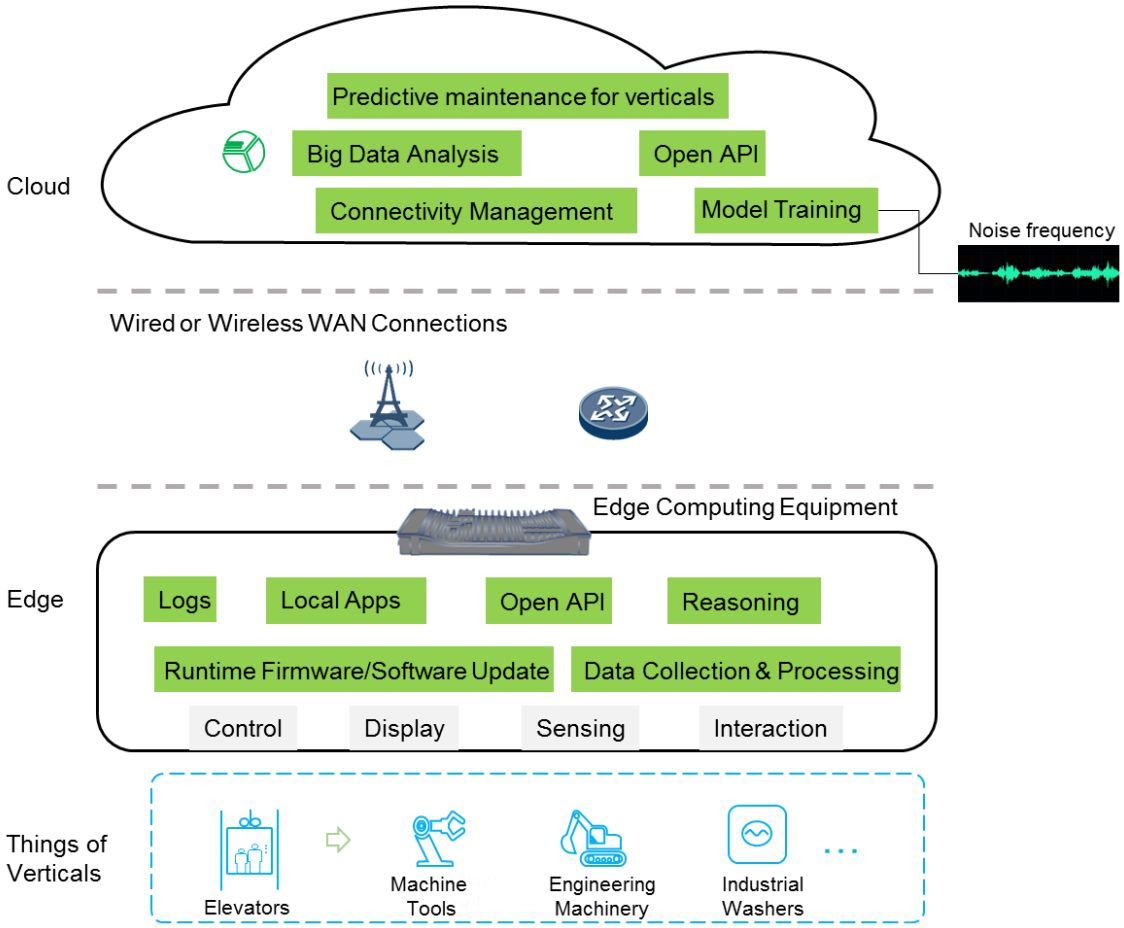

Figure 6 Logical Architecture Diagram for Edge Computing

To facilitate discussions on the boundaries and the necessary means to enable edge computing, there are “Key Requirements”, “Edge Boundary” and “Edge Devices” clauses added to each use case. “Key Requirements” are not intended to be a standard so they are not normative. The “Edge Boundary” is one view from the contributor to the uses case and is not intended to be definitive as typically, the boundary of each service or application varies as drawn by the system designers. Similarly, the term “Boundary Devices” portrayed in the use cases by the contributors is not an exhaustive list.

Through the examination of all the use cases described here, we inferred general requirements for edge computing that may be common to all use cases:

Communications: Edge devices must continue to function even though data communications may be temporarily interrupted.

Edge device capability: The edge devices need to support edge-computing capabilities: communication, local computing and local storage.

Edge device functionality: The edge devices can be customized with features to fit various vertical industries, such as compact size, low power consumption, anti-vibration, electromagnetic shielding, waterproof and dust proof.

CREW SAFETY MANAGEMENT

Objective: Use wearable multi-gas detectors to monitor employees' exposure to harmful gases during a shift. Create real-time exposure profiles using data from the sensors and adjust the work schedule or work flow to prevent health issues.

Description of the use case: Safety in a hazardous or life-threatening operating environment, such as a mine, is a top issue. Other issues are production related, such as the status of the tools used, the status of the conveyer or vehicles used to carry out the ore, and the amount of ore produced in a measured period.

Key Requirements:

1. (Environment monitoring) poisonous gas detection, ambient temperature detection and control and lighting control need to be implemented.
2. (Personal monitoring) the personal vital-sign monitoring system must be installed locally and the data sent out to a central monitoring station in the operation center.

Edge Boundary: Operation site (e.g. at the mine operation office or operation center).

Edge Devices: Personal vital sign monitor (body temperature, heart rate and blood pressure, CO_2 level), personal tracker, environment monitors (ambient temperature, CO level, hazardous gas detection, lighting), tools, trucks and unmanned vehicles, conveyer belts and weighting devices.

FLEET TRACKING AND PLATOONING

Objective: Combine real-time GPS data with vehicle usage data from sensors to monitor and optimize the location and status of the fleet.

Description of the use case: A trucking company must operate its vehicles safely and efficiently. In addition to optimal routing in the delivery routes, the mechanical status of each vehicle can contribute to timely maintenance to improve the operation efficiency and safety. The operation is originally designed for individual trucks but it can be extended to several trucks to form a "platoon" of unmanned trucks to increase operational efficiency.

Platooning is typically used for self-driving vehicles. The lead vehicle detects the lane and traffic condition to decide the optimal speed to maneuver. This information is then communicated through the following vehicles in the platoon. Each car has vehicle-to-vehicle (V2V) communication capabilities and controls for driving safety. The lead vehicle gathers the information, such as the location of the fleet and the operating condition of each vehicle and reports to the office.

Key Requirements:

1. GPS based vehicle tracking supported by a local map.
2. The vehicle status monitoring system is installed locally and the data shall be sent out to a central monitoring station in the operation center.
3. A connection to a central location providing traffic updates

Edge Boundary: The lead vehicle of the platoon.

Edge Devices: Vehicle status monitor (engine speed, temperature, break pad thickness and hydraulic subsystem, transmission subsystem, weight and shock status, tire pressure, fuel level, etc.), GPS tracker, environment monitors (ambient temperature, lane detection, vicinity vehicle detection), container status monitor (for refrigerating) and V2V driving safety devices.

PREDICTIVE MAINTENANCE–CONNECTED ELEVATORS

Objective: With edge applications installed on connected elevators, operators and technicians are able to perform predictive maintenance of the elevators based on the data they provide at the edge.

Description of the use case: Operators of connected elevators rely on edge functions to achieve predictive maintenance of their systems for the systems to become more reliable and reduce system downtime. The operational cost of these systems is greatly reduced since the efficiency of the system can be significantly improved.

A connected elevator uses many sensors for gathering data on noise, vibration, temperature, etc. The operational status of the elevator can then be derived from analyzing the sensed data. With elevators connected to edge computing devices, and the sensed data uploaded to the cloud, elevator operators can obtain the running status of all of their elevators. Elevator technicians are then able to perform predictive maintenance using edge computing data, and data in the cloud, to check and maintain those elevators selectively that are more likely to fail based upon analytics. Predictive maintenance increases the operational efficiency of the equipment while reducing the maintenance expense through targeted failure prevention and avoidance of unplanned downtime.

Key Requirements:

1. The edge devices offer containers, open APIs that allow third parties to develop applications to be installed on the edge devices.
2. To support 7 x 24 monitoring, the edge devices support runtime update of its software and firmware.

Edge Boundary: The elevator operations center or the elevator itself

Edge Devices: Infrared sensors, weight sensors, smoke detectors, vibration inductors, noise sensors, cameras and operator interfaces.

PRODUCT TRACEABILITY

Objective: Regulations in the food industry (e.g., EC 128/2002) require manufacturers to establish systems that enable traceability of food products across all stages of production, processing and distribution. While this use case focuses on the food industry, product traceability is important across multiple industries.

Description of the use case: Pieces of plastic in chocolate bars, bacterial contamination of cream cheese, falsely declared ingredients in pasta-based ready meals—a food product may be recalled for any number of reasons. Time is of the essence when it comes to product recalls. As well as damaging the manufacturer's reputation, these situations can be expensive, with costs rising as the whereabouts of the end products become less clear.

Bar codes, 2D codes, or electronic transponders are used to identify objects depending on whether they are individual items, primary and secondary packaging, pallets, trucks or containers. When it comes to deciding which technology to use, financial factors and the objects and processes involved must be taken into account. For example, a bar code can be printed onto an egg, while the cartons holding six or twelve eggs can be labeled with 2D codes and additional plain text such as the best-before date. A transponder, on the other hand, can be added to shipping cartons, pallets and other aggregated containers. The various methods of product identification described above ensure that the flow of materials across the supply chain are labeled, identified and tracked.

Industrial machinery, automated guided vehicles (AGVs) and collaborative robots or "cobots" are increasingly prevalent on the factory floor. The most targeted applications are packaging and palletizing, pick and place, machine tending and assembly and quality inspection. As issues may occur at any step in the supply chain including the quality or handling of materials, contamination introduced by people or machines or faulty processes, product quality and product traceability require orchestrating, recording and verifying the people, processes and machines involved.

Sensors with edge-computing capabilities allow these product identification methods to be checked against stored data for verification to ensure the flow of goods, people, process and machines. The right product goes into the right package and onto the shelf with critical information appearing correctly on the package and full genealogy available in the cloud.

Key Requirements:

1. Consider ambient conditions when selecting edge devices within the food industry, such as humidity, cold storage and outdoors.
2. Location and tracking of items across all stages of production, processing and distribution is important.
3. Sensors and computer vision systems identify particulates or contaminants in food.
4. Sensors and edge computing to orchestrate, record and verify the people, processes and machines involved.

Edge Boundary: machine on which the sensor is located.

Edge Devices: bar code readers for 1D or 2D codes, vision sensor for recording images, RFID tags and readers.

Edge Computing and Industrial Analytics

Analytics is broadly defined as a discipline transforming data into information and business value through systematic analysis. *Industrial analytics* is the use of analytics in IIoT systems.

Advanced analytics is at the core of this next-generation level of transformation and, when applied to machine, process and grid data, provides new insights and intelligence to optimize decision-making and enable intelligent operations leading to transformational business outcomes and social value. These new insights and intelligence can be applied across any level of any industry if the appropriate data can be collected and analytics applied correctly. Some say data is the new oil. If that's the case, then data analytics is the new engine that propels the IIoT transformation.

Analytics can be classified in a number of different ways depending on where they are performed, the window of time for a relevant and meaningful response, and what functionality they are trying to achieve. A single analytical flow could involve edge analytics for initial distillation of data and immediate actuation, analytics in the cloud comingling the latest news from the edge with historic big data stores and back to the edge for further actuation.

Technology and the Evolution of Industrial Analytics

Advances in IT and OT capabilities such as compute capacity, communication bandwidth, low latency, software capability and sensor technology have removed technological constraints and allowed analytics to be deployed through an entire IoT system. For instance, looking at the edge tier of a system, the processing capability available at the edge in conjunction with low-latency communication have enabled algorithms to be run in real time supporting models that generate insights and real-time control for the system. Similarly, looking at the cloud tier, what was once impractical, performing streaming analytics on enormous data sets, is now possible thanks to big data compute capabilities and high-bandwidth communications. These same advances have also enabled the distribution of analytics so that they need not be centralized and can be implemented across the IIoT ecosystem.

Where Should the Analytics be Performed?

Most industrial analytics deployments use a hybrid approach where analytics run at all tiers from edge to cloud, with analytics at a particular tier addressing a specific business objective.

Cost benefits stem from reducing the amount of data being sent to and stored in the cloud. Edge analytics mitigates the cost of storing and processing low-value and oft-repeated data. Analytical

models are not helped by data noise. Instead of creating an unnecessary noisy big data problem, edge analytics can distill data prior to sending it on to the cloud.

Security Considerations for Edge Computing

Security is an important consideration for edge computing. More components and communication channels create a greater potential for attack vectors. Innovations are required to monitor, manage and secure globally distributed systems and contain inevitable breaches. The IISF documents a generalized end-to-end security framework. In edge computing implementations:

- security must be built-in to each device and at every level of the architecture,
- computing and networking endpoints must be monitored and managed,
- latest patches must be applied,
- attacks must be isolated and quarantined and
- affected components must be able to be healed.

Orchestration

The centralized nature of cloud computing enables access to a scalable and elastic pool of shareable physical or virtual resources. As computing is distributed to the edge, resources can still be shareable but elasticity is challenged because:

- compute resources could be in separate islands where they cannot communicate to coordinate computation,
- locations of compute resources may be difficult or costly to access,
- compute in ruggedized enclosures might not be expandable and
- technicians to perform the work may not be easily available at the edge.

With these limitations, the approach is inverted. Understanding both the “as-built” compute target and net-available resources is critical to deploying the right software to the right location. Once deployed, tools to manage, monitor and secure the entire lifecycle are required. Software may need to be throttled or redeployed, memory usage restricted, databases and logs truncated if resource thresholds are challenged. We also need to predict usage trends to address issues before they occur.

The challenge for developers and administrators is to understand not only the physical requirements of their applications (computing inputs, outputs, connectivity, etc.), but also the security and processing requirements and how those requirements translate to different CPU and OS types. Industry standard calculations and metrics may be required.

The two main activities essential to deliver an orchestration solution are:

- infrastructure management to handle the lifecycle of devices at the edge including the commissioning and provisioning of resources and

- orchestration to manage the lifecycle of services and applications and the dependencies between them.

Various standardization efforts share this understanding of orchestration, such as the ETSI Multi-Access Edge Computing (MEC) initiative, the ETSI NFV Management and Orchestration (MANO).

Orchestration and infrastructure management at the edge poses challenges not faced in the cloud mainly due to:

Heterogeneity of devices and application domains: At the edge, there are no expectations on homogeneity regarding devices, or the hardware and software platforms. An infrastructure management system needs to be flexible enough to manage a plethora of devices to consume their resources seamlessly. To provide some level of homogeneity at the edge, both virtualization and containerization technologies may be employed. Also, devices can behave differently based on the application domain. An orchestration solution in a smart factory environment where the nodes are static and the network reliable will have a different behavior from those orchestrating a logistics or smart cities domain where vehicles are consistently mobile and subjected to variable connection quality.

Different connectivity and communication technologies: IIoT gateways must handle multiple connectivity solutions using different protocols. The orchestrator must be aware of the available solutions to guarantee communication between deployed functions and applications.

Differences in capabilities, requirements and constraints: The higher level of homogeneity and the virtually infinite availability of resources in the cloud ease the orchestration process. Conversely, a broader range of service requirements, device capabilities and constraints are observed at the edge. For example, devices at the edge have different sensors, actuators, real-time operating system or networks, some nodes can provide accelerators, others will not. The orchestrator must be aware of the capabilities found in the infrastructure. Also, the constraints on these nodes need to be known beforehand (e.g. bandwidth, battery, CPU power, memory). At orchestration time, a service must be able to describe its requirements, and the requirements will be checked against the available capabilities and constraints found in the infrastructure.

With that in mind, orchestrators can operate both vertically and horizontally. Vertical orchestrators handle services in a specific domain, while horizontal orchestrators manage services across different domains providing integration among them. An example would be a smart factory that relies on a logistics company, and each has its own orchestrator. A horizontal orchestrator composes services that span the different domains (e.g. a service that adjusts throughput of a production line based on the current location of the necessary supplies).

Orchestration is an important aspect of edge computing to provide a platform to support both IT and OT activities in IIoT. The ability to coordinate the deployment of new services and applications gives the edge the capacity to be programmable and deliver the services required

by its consumers. While trying to ensure the quality of the service required, its presence in edge solutions needs to be enforced.

Conclusion

Edge computing has been implemented in a variety of IIoT deployments; however, the need to modernize edge architectures became apparent with the emergence of cloud computing. The rapid decline in processor and memory cost enables more advanced decision-logic closer to where the data is created, at the edge. The industry has learned that a "one-size-fits-all" approach has never been adequate for IIoT. It is also true that the IIoT system designers always know where the edge boundary is, and what devices in the system can be categorized as edge devices. System designers are challenged to implement an architecture that is managed, orchestrated, trustworthy and secure. The next phase of the work will be to address these concerns in the Technical Report.

While we have tried to lay out the significance of edge computing of future IIoT systems, we know it is a never-ending task as new IIoT applications and new considerations appear every day. We intend this paper to trigger more in-depth conversations and invite your participation.

This is not the end; rather, a beginning.

Authors and Legal Notice

This document is a work product of the Industrial Internet Consortium Edge Computing Task Group, co-chaired by Mitch Tseng (Huawei), Todd Edmunds (Cisco) and Lalit Canaran (SAP SE).

Authors: The following persons have written substantial portions of material content in this document: Mitch Tseng (Huawei), Todd Edmunds (Cisco) and Lalit Canaran (SAP SE).

Contributors: The following persons have contributed valuable ideas and feedback that significantly improve the content and quality of this document: Pieter van Schalkwyk (XMPro), Cliff Whitehead (Rockwell Automation), Mike McBride (Huawei), Mingui Zhang (Huawei), Prof. Thomas Magedanz (Fraunhofer Fokus), Mathias Santos de Brito (Fraunhofer Fokus), Eddie Lee (Moxa), Michael Thomas (SAS), Jill Oertel (SICK), Manish Sharma (Ligado), Shi-Wan Lin (Thingswise), Will Sobel (VIMANA), Wael Diab (Huawei), Eric Harper (ABB), Vijay Ujjain (PricewaterhouseCoopers) and Mark Crawford (SAP SE).

Technical Editor: Stephen Mellor (IIC staff) oversaw the process of organizing the contributions of the above Authors and Contributors into an integrated document.