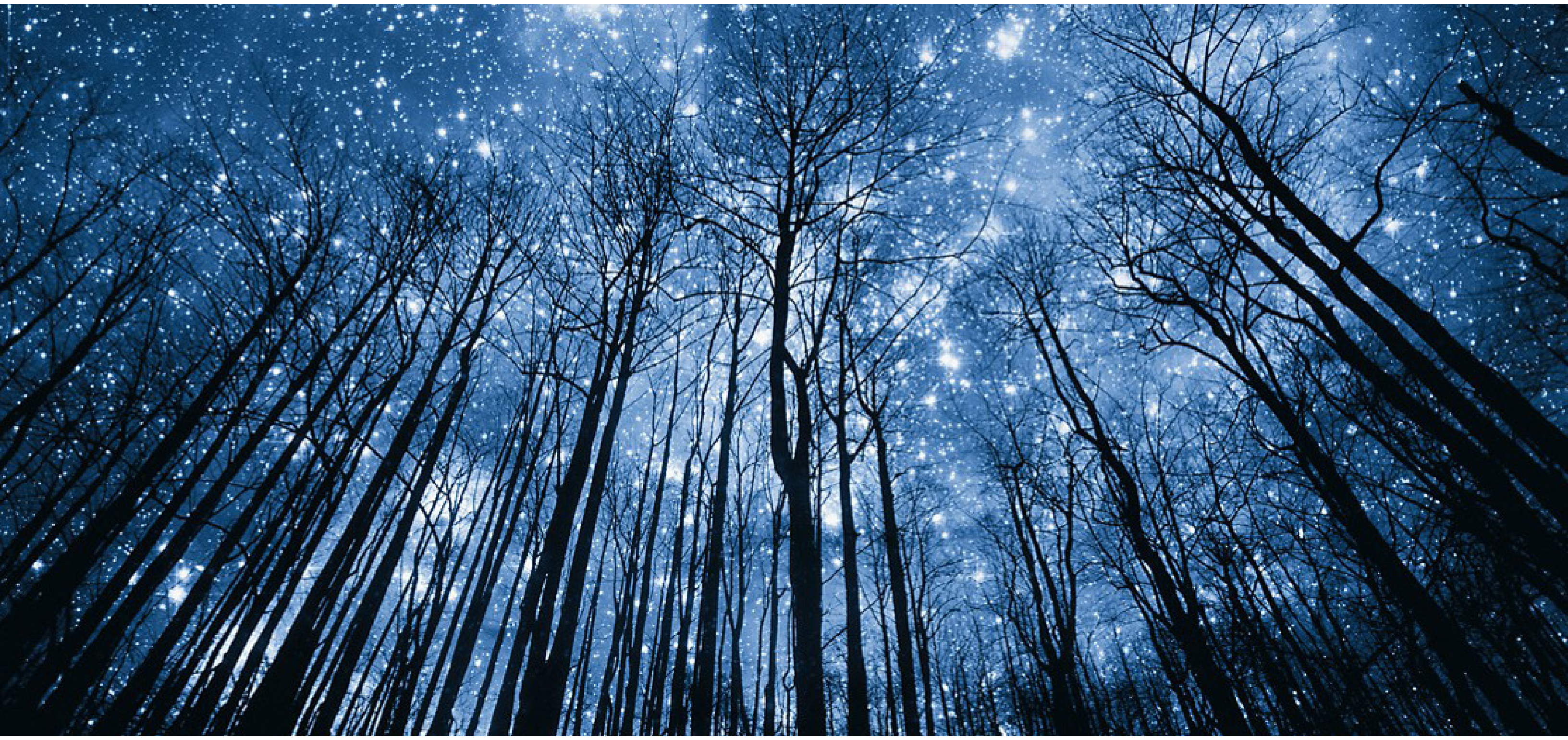
Dell Technologies

IS DATA TREATMENT ESSENTIAL AT EDGE GATEWAY?



Marcio Monteiro
Software System Principal Engineer
Dell Technologies
Marcio_monteiro@dell.com

Dell Technologies Proven Professional

The Dell Technologies Proven Professional Certification program validates a wide range of skills and competencies across multiple technologies and products.

From Associate, entry-level courses to Expert-level, experience-based exams, all professionals in or looking to begin a career in IT benefit from industry-leading training and certification paths from one of the world's most trusted technology partners.

Proven Professional certifications include:

- Cloud
- Converged/Hyperconverged Infrastructure
- Data Protection
- Data Science
- Networking
- Security
- Servers
- Storage
- Enterprise Architect

Courses are offered to meet different learning styles and schedules, including self-paced On Demand, remote-based Virtual Instructor-Led and in-person Classrooms.

Whether you are an experienced IT professional or just getting started, Dell Technologies Proven Professional certifications are designed to clearly signal proficiency to colleagues and employers.

Table of Contents

1. INTRODUCTION

The discussion developed in this article aims to debate a counterpoint to the existing information technology (IT) perspective about managing data at edge gateways. The article's focus is to highlight operational technology (OT) ideas. It can sound awkward to open space to experts coming from areas far from IT, however, OT is a real world where every single piece of data is collected. In place of talking about megabytes (MB) of data, OT sees flow, pressure, production and so on. In other words, what the virtual world treats is quite real and the best way to interpret these elements is OT.

This article's goal is to begin a discussion of how data handling can be easier with the OT approach to well-known IT problems.

2. WHAT IS EDGE?

If you search a dictionary, you could find edge defined as “the outer or furthest point of something” (Cambridge Dictionary, n.d.). With similar meaning, information technology considers edge as a device or server in a distant location.

The edge is a tenured space where physical and digital environments are connected. At this location, depending on which architecture is adopted, the software is deployed to gather data and send it to higher levels such as a local data centre or the cloud. For example, it is commonplace to find industrial gateways running on plant floors close to machines. These gateways are the link between cloud or centralized servers allowing data ingestion.

There is a wide range of market solutions with differing combinations of hardware and software to ensure all data available on the factory floor can be used. Concerns are to keep these remote devices connected using the most efficient and least intrusive technologies. In case of failure, local databases, for example, can work as a buffer to permit hundreds or thousands of machine data points or “tags” to be collected so they are not lost.

But what is the reason for collecting raw data at the data centre? There are huge improvements possible when industrial data is gathered, analysed, and used to reduce loss and make the process more reliable with better productivity. As much as we can get data and understand where the pain points are, more is possible to increase enterprise profits.

Nowadays, we are in a rush to get all data available for further processing. This data is generated in such volumes and at such high speeds that it is difficult to capture it all accurately. “Gartner expects more than 15 billion IoT devices will connect to the enterprise infrastructure by 2029“ (Katie, 2021). The exponential growth of data is revisiting our well-known architectures of centralised processing and distributed acquisition.

As a problem can leverage new opportunities, innovative ideas to overcome this exponential growth are required. The basic idea is to connect device clients to a nearby edge module for more responsive processing and smoother operations. In other words, distributed processing brings applications and data storage closer to data sources, allowing low latency and computing data efficiently to reduce network traffic (Gold & Shaw, 2021).

The technology is going beyond because thinking about spread processing led us to new ideas about pre-processing and decentralization of computing infrastructure. This movement is pushing computing resources from central data centres to close to data production. This improvement has several advantages like easing disaster recovery, increasing security, and reducing maintenance costs. Therefore, the edge concept is also motivating the reduction of data which needs to be processed in a centralized data centre (Red Hat, 2021).

It is interesting as one idea inspires another. Edge was a necessity to establish a link. Pre-processing at the edge was introduced as a solution created to answer the high volume of data created by the Internet of Things (IoT). But the exponential growth of IoT devices causes problems to arise when the number of devices transmitting data at the same time grows. So, besides pre-processing, it should be required for complete servers with databases and applications running at the edge, especially for supporting real-time data which cannot suffer latency. An edge gateway, for example, can process data from an edge device, and then send only the relevant data back through the cloud, reducing infrastructure needs (Gill, 2020).

3. WHAT IS A GATEWAY?

“A computer that sits between different networks or applications, the gateway converts information, data, or other communications from one protocol or format to another.” (Gartner, n.d.). The accuracy of this definition could be asserted for today, but not looking a few years ahead. It is the result of the gateway changes to fulfil all edge requirements. Nowadays, an edge gateway looks beyond a server with applications, database, firewall - it- it considers all needs of anyone running information through a data centre (Braunisch, Schlesinger, & Lehmann, 2021).

Even though we experience deep transformation with the gateway role, the main purpose is to allow the flow of the data from one device to another device through their network. In terms of edge, it is a bridge between sensors, devices, and data centres. It is necessary that the gateway complexity was not only increased by the movement of decentralizing processing. On the industrial side, its evolution is boosted to follow effects of Industry 4.0, for example. In this process, computers communicate to make decisions without human involvement. Industrial equipment is becoming smarter, exchanging data at a worldwide level.

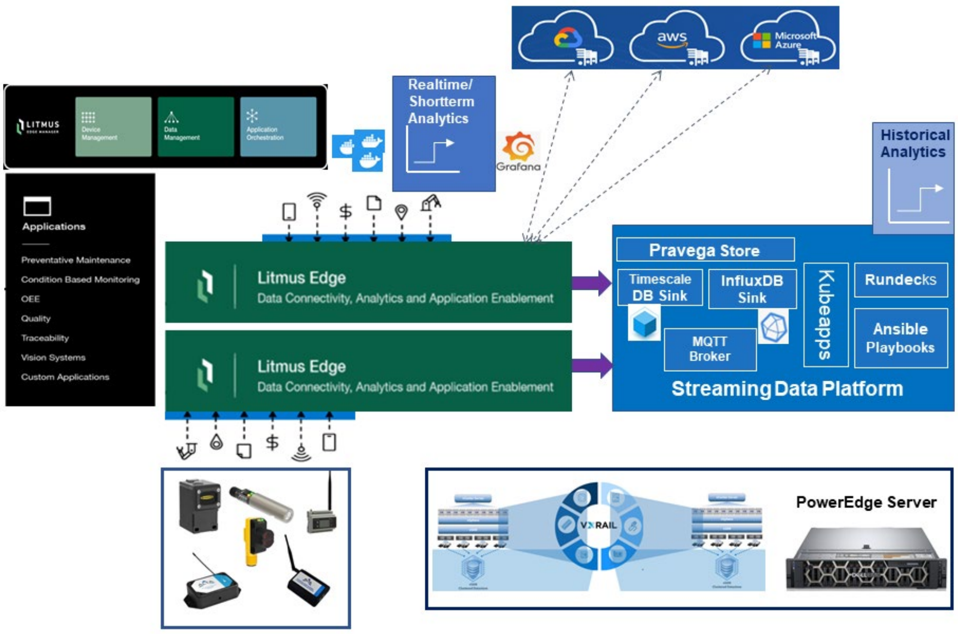


Fig 1. DTVD for Manufacturing Edge with Litmus - architecture (Dell Technologies, 2021)

IoT multiplication was another process to disrupt a well-established idea about the gateway of passive devices. The figure above shows Dell’s architecture running industrial sensors with Litmus. This solution requires thousands of tags ingested per second, with multiple links available directly to the cloud or local datacentre. This architecture is only possible for a gateway with features closer to a server than a passive gateway.

Pre-processing data locally at the edge before sending it to the cloud is a real necessity. This requirement aggregates data as a way of reducing the volume of data that must be forwarded to the cloud. It reduces response times and network transmission costs.

It is important to understand the updated gateway concept requires not only dealing with outbound traffic streams used for sending IoT data to the cloud or data centre. As responsible for part of processing, inbound traffic must receive management tasks such as updating device firmware, data visualization and basic data analytics, or deploying new applications (Open Automation Software, n.d.).

So, the gateway should contain complex features without putting aside the challenge of integrating a multitude of devices and protocols, many with varying power and connectivity requirements, while also including legacy technologies.

4. DATA TREATMENT

Machine data was treated only at the edge to be conditioned to be human-readable. For example, converting flow to tags every second or less creates new challenges to edge teams.

In smart manufacturing scenarios, most of the acquired data are time-series generated by large-scale sensor networks that are monitoring the continuous functioning of the manufacturing processes or equipment to be analysed. As a result, the volume of acquired data (i.e., time-series data from sensors collected 24 hours a day, 7 days a week) is rapidly increasing (Risse, 2018, Zhou, 2019). Such is the growth pace that in 2015, the manufacturing industry was already generating more than 1000 exabytes (EB) of data annually and it is expected to increase by 20 times in the next 10 years (Yin and Kaynak, 2015). Therefore, problems are arising related to the considerable costs associated with the resources needed to store them (Villalobos et al, 2020, Villalobos et al, 2019).

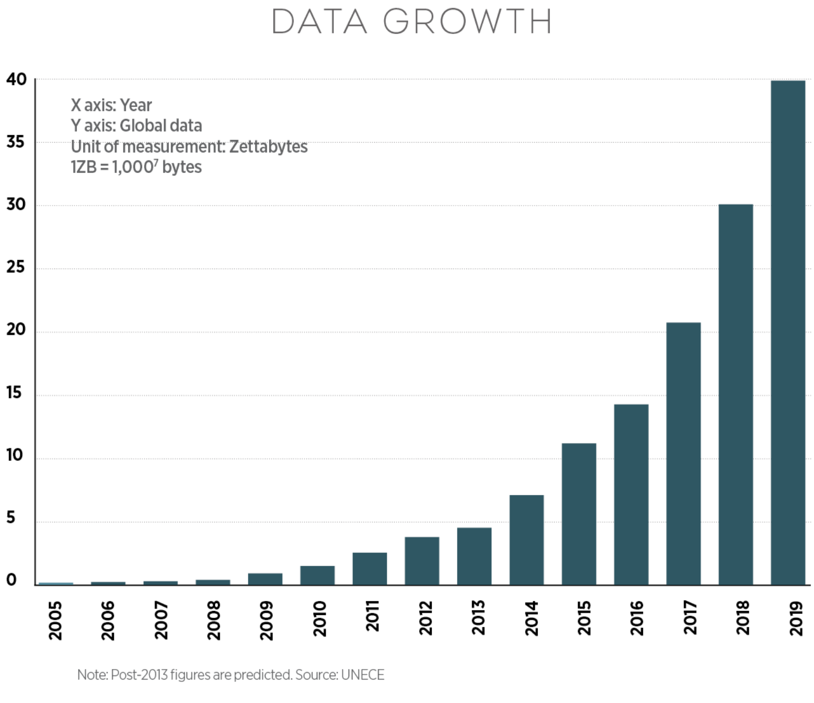


Fig 2. Data Growth (Yadav, 2018)

The industrial sector has been leading this intensive use of data. In a general way, time-series data are ingested on a large scale by sensors running 24 hours a day, 365 days a year. Dealing with a massive amount of time-series data creates high storage cost problems. In this scenario, lossless compression methods arise as an efficient solution. If compression solves the storage cost, all the processes, which includes decompression for data access, need a more powerful processor to be completely finished (Gómez-Brandón, R.Paramá, Villalobos, Illarramendi, & R.Brisaboaa, 2021).

There was no innovation in the data compression idea. It was used to compact messages with Morse code and nowadays, in social media like Twitter and WhatsApp, to send notifications with acronyms or contractions. The core idea is to transmit all the information with fewer characters using encoding. Inside the compression techniques, lossless and lossy compression can be used depending on the accuracy aimed and the end-use (Eswaran & Gastpar, 2009).

For a better understanding of how compression can deal with data, another principal element to be considered is redundancy. In a voice recording, for example, it is possible to remove high and low-frequency ranges without any loss to the result because human ears can only perceive 20 to 20,000 Hz. On another hand, removing infrared images cannot be acceptable for systems that identify small cracks in equipment. Redundancy assessment depends on the real data necessity. This evaluation of the real data necessity brings again the relevant OT role.

Working with redundant data makes it possible to do lossless compression. In this process, there is no data lost, and all the information stored reflects the original data. On another side, if redundancy is small or the storage requirements are tighter, it is possible to use lossy compression. When rebuilding a time series, for example, from any reduced representations, some data is lost. These techniques are primarily aimed at lowering the dimensionality of time series data and facilitating data analysis. In the last type, data loss does not mean incurring quality or accuracy problems if the scheme is thought of in the view of the end-user supported, again, by OT experts.

Losing data is not a unique concern of compression applications. As compression is used to improve storage and transmission data, another element is critical under edge gateway: used resources to run code and decode the data. One frequent problem is the necessity of decompressing the entire dataset to access the data. They frequently do not support random access. Compression applications accessing data without requiring decompressing all the data are more desirable because they allow data to be stored in a compressed format.

For a better understanding of compression techniques, consider the following points: (Lockerman & Kulkarni, 22):

- Delta compression stores only the difference to a reference object. It works better for redundant information like the status of a machine which is constant for hours, for example. It is especially useful for the timestamp because it is not necessary to keep all the Unix Timestamp but only the difference in seconds to the reference.
- Delta-of-delta encoding is more used in time-series datasets, which happen in regular intervals. In this way, it is not necessary to store time.
- Simple-8b reduces the number of stored digits. Integers are represented by the minimal length necessary. So, it treats it in different ways - floats and integers improve storage.
- Run-length encoding is a mix of all the above. It uses the idea of simple-8b, but it improves storage by evaluating data that does not change very often. It is used as a building block in many other algorithms.
- XOR-based compression is used to float. They are more difficult to compress because often have several “0s” to represent the number. As decimal is based on “10” and floats are binary based, the representation of decimals in binary causes excessive use of data. The first data point is stored with no compression and the subsequent points are represented using their “XOR’ed” values. It can reduce the data storage in 50%.
- Dictionary compression is an agnostic type. It works by listing the possible values that can appear, and then just stores an index into a dictionary. This dictionary has unique values. It can be used regardless of the data type, but it works fine for a limited set of values that repeat frequently.

Although new edge gateways are built with powerful processors and memory, these resources are aimed to support several new tasks imposed on them in recent years. If compression/decompression applications are requiring many resources, this solution can create more problems than it solves. Therefore, it is relevant to consider that if compression is necessary, how is it best applied to all datasets treated by the edge gateway.

New ideas are put forth all the time, like Direct Access Compression of time series (DACTS), which are remarkably interesting to be introduced in industrial time series (Gómez-Brandón, R.Paramá, Villalobos, Illarramendi, & R.Brisaboaa, 2021). For example, with equipment status, which is represented as a Boolean expression, it kept the same for value several hours. Instead of registering the same value for each second, frequent sequences of original values are represented with shorter storage.

5. KPI DATA TREATMENT AT EDGE GATEWAY

After discussing edge gateway, we can surmise that data treatment is the unique solution to handle a great volume of data. In handling the limited resources of edge gateway only from the IT perspective, data compression is a reasonable solution which is better than increasing memory, storage, and cost of a gateway. But it is necessary to think about the issue from another perspective. For example, if all raw data could be treated at the edge to find key performance indicators (KPI) before moving to long term retention storage, what improvements would exist as part of the final solution?

If we could partially move the data treatment from the data centre to the edge gateway to find KPIs, would it improve the use of its resources? Several gains assigned to KPI data treatment at the edge gateway are discussed more deeply in the following paragraphs:

- Issues and improvement opportunities
- Edge data treatment as fundamental to reducing data upload to higher levels
- Standardization of data treatments
- Moving the process evaluation and KPI elaboration closer to the real world

Issues and improvement opportunities

Every solution which involves IoT data should be built as a plug-and-play model. For example, the final user should not need deep knowledge of IT to deploy and configure a solution. It would not sound strange because we have hundreds of independent software vendors (ISV) with multiple solutions available on the market. For this reason, ISVs are developing applications more friendly for smooth customer usage.

It is not unusual for sellers to indicate that is possible to run incredible quick implantation for hundreds or thousands of tags. It is a commercial truth, but it is not true in the technical/real world. The best operational knowledge rests with the plant manager and objectively with the plant owner. The maintenance engineer or the daily operator have a deeper understanding which cannot be overlooked. They know where the issues are and how improvements can be made, but there is also a vast gap between the plant floor and the highest expectations. So, the first gain with using data treatment at the edge is aggregating this knowledge with KPIs.

KPI is obtained by correlation among several variables. It would be most beneficial to build these indicators and create value to the final solution, while still counting on operations experience. Instead of gathering all available data, these people must help to select and build high-value indicators based on only the data that is most useful. Not all data is useful.

Edge data treatment fundamental to reducing data upload to higher levels

It sounds perfectly reasonable to use compression data to gather as much information as possible to be stored and analysed. There are already several algorithms that make it feasible. However, from an OT perspective, it is unnecessary to collect all the data available on a programmable logical computer (PLC), as one example. There are several tags critical to PLC work and not all data is relevant to evaluating the process performance.

When carefully selecting which tags are important to an overall assessment, the consequence is a reduction of the amount of data in the workload. This is a natural result of an efficient data treatment at the gateway level.

For example, it is common to record a motor status - in our case, a Boolean which indicates “0” stopped and indicates “1” running. Compression technologies to reduce or avoid registering each second of the unchanged status of the same motor are quite efficient. However, it is more efficient to know if this motor status is necessary for the end customer. It is simpler and more effective, instead of doing the already mentioned data treatment at the gateway, to select which variables represent real value.

As an example, we would aim to understand the status of a pulsed solenoid valve at intervals that are meaningful and useful based on the data the valve produces. The natural consequence is the radical reduction of data. All the necessary infrastructure will be improved by defining the best KPIs.

Standardization data treatments

Standardization comes from the understanding process. Returning to the prior motor status example, instead of giving complex data compression treatment, it is possible to record how many status changes happened over an hour and which percentages were due to the running situation. If this idea can be applied to every single motor, it establishes standardization, which increases the value of data.

Moving the process evaluation and KPI elaboration closer to the real world

The last assertion is the main goal of any IoT project: looking through the data to see with fidelity the real unseen and unheard world. It is easy when millions are invested in hardware, software, and subject matter experts. The main problem is the distance between what is essential or not in the real world. It is necessary to look beyond the technology to gain a deeper understanding of what the customer’s needs really are.

6. DEVELOPING DATA TREATMENT AND KPI AT EDGE GATEWAY

The idea of using KPI is not new in the industrial environment. All equipment makers look to understand what the best setpoint is to reach maximum performance. But the greater problem is that every single piece of equipment is tested in a controlled situation. In a real factory, equipment is put together to operate in a unique mode and the isolated maximum performance is hardly reached.

At this point, you may ask what the relationship is between KPI and data treatment that is discussed in this article. It has a strong relationship. The idea is simple - Instead of pushing all data off to a data centre to be analysed and treated, the gateway can do this at the edge with all the benefits already discussed.

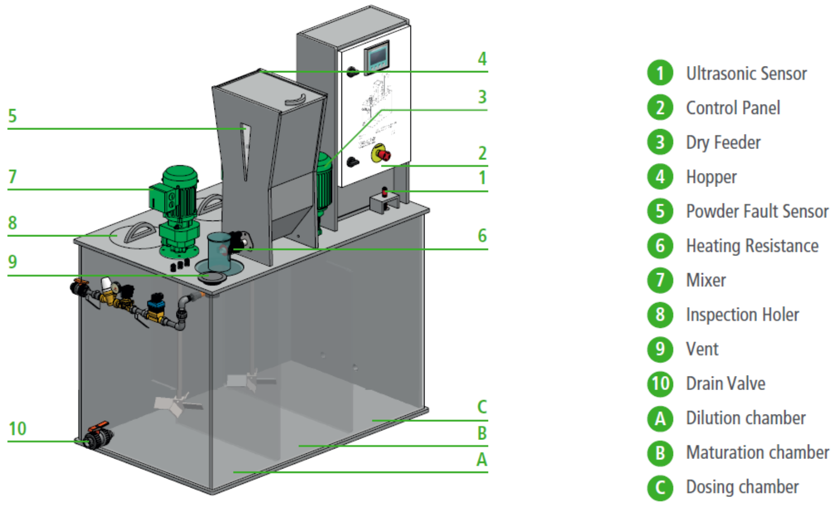


Fig 3. Polymer Preparation System (Kozegho Dosing Systems)

Let us look at an example of the equipment above, which is larger and used in dairy and wastewater treatment. Its goals are to dilute powder polyelectrolyte with water in continuous mode. It has 3 chambers. The first one (dilution chamber A) is where the dry feeder motor pushes the powder after a heating resistance is used to melt it. A mixer helps to dilute the powder. By the overflow, the mix of water and powder reaches the second chamber (maturation chamber B). By another overflow, the mix reaches the last chamber (dosing chamber C). The mixture goes out and it is used in the process.

Even though there are several motors (4 in total) and instruments (valve solenoid, level switch, level transmitter and flowmeter), this set has an operation point of maximum performance based on the inlet flow. All other variables are relevant for local control, not for performance evaluation.

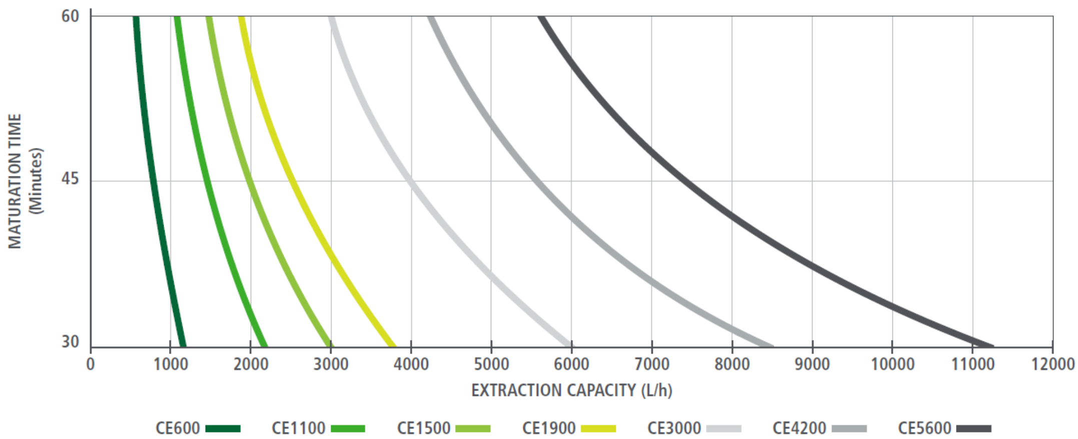


Fig 4. Extraction Capacity as a Function of Aging Time (Kozegho Dosing Systems)

In another example, the equipment CE1900 has the maximum performance at 2,500 l/h (considering a maturation time of 45 minutes). So, any other flow above or below this value will result in a performance lower than the maximum.

All this consideration of this example was done to consolidate that the most important variable to evaluate the equipment is flow. If the flow has this critical role, why should the edge gateway ingest data of each motor or level transmitter? There is no valuable use to register the current, voltage, and frequency of all the motors to evaluate the performance.

It is obvious that if any motor has an issue that results in a fault, the process is stopped. But it is an operational and local problem. So, the local failure should be monitored by the local operator to be fixed as soon as possible. This problem should be identified if our assessment includes availability, not performance. So, it is critical to understand what the necessity of the end customer is.

It is possible to go beyond this if we make a KPI evaluation at the edge gateway. With a better understanding of the equipment, it is possible to describe a parabola curve that establishes a correlation between flow and performance considering the real equipment limits (from 30% to 150% of the nominal flow). This parabola can be described as an hour-performance curve because the time between the beginning of the process and the way outflow takes roughly one hour.

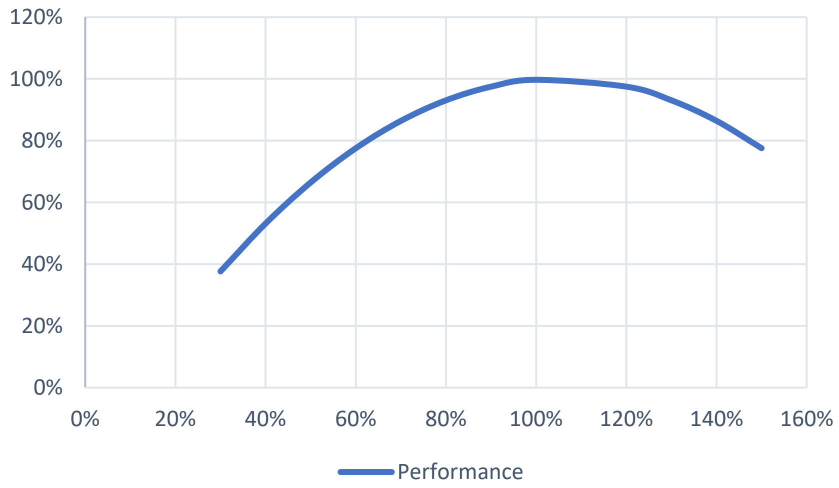


Fig 5. Performance (%) in the function of flow (average flow/maximum efficiency flow)

Using the above curve as a reference for expressing the performance calculation, it is possible to build a direct relationship between flow and performance. So instead of ingesting dozens of variables each second to be stored and processed only at the data centre, the edge gateway can do this evaluation only based on one variable: flow.

7. CONCLUSION

Nowadays, it is common to buy a mobile phone with a camera of 12 megapixels. A simple photo of your child may generate a file with several MBs. It is a nightmare to store all photos and videos of an event. For solving these simple questions, it is necessary to use compression technologies or buy a huge external hard disk or cloud storage. Do we need all this quality in megapixels to record these life moments? The same approach can be done for data at an edge gateway.

This article does not provide any newer technology or a magical solution for zettabytes (ZB) of data expected for IoT projects. Instead, there was discussion of how digital use must be compatible with and meet the customer's real necessities.

All the indicated arguments can be appreciated, but they are often set aside because such questions demand extra time investment to discuss them with plant floor operators and users. We are inclined to believe that the time spent to implement plant floor gateways is lost because no data is gathered until a gateway is integrated with existing control systems. Solutions based on wireless technologies are the best option because there is no need to wait for physical connectivity into the infrastructure. We are challenged by the time required to deploy and implement quick solutions to be available at the next board of directors meeting.

It is not possible to create a pattern for all edge gateways because the peculiarities of the process and the necessities of the end user must be considered. Think about common operational systems: Windows and Linux. They are different, but both have their place based on their features. So, let us divide the IT into only two spaces. Windows is for direct interfacing with the common user and Linux is aimed at server tasks. Sorry, but even though the IT necessity is more complex, and MAC OS has a high value to media developers, Chrome OS to hardware with limited resources and several others, besides Windows variations like Windows Server and so on.

Project standardization is not the solution because it leads to gathering thousands of tags in an abbreviated time. It should produce meaningful data to steer business decisions. How can we choose the best indicators to make choices if there is no straightforward evidence of what they are? We are overwhelmed by data with little value. The problem is not about budget, but about taking OT advice to follow the best road. Regardless of the size of the environment, or how little time is available, we need to focus on quality improvements that the data may drive in future production.

8. BIBLIOGRAPHY

Braunisch, N., Schlesinger, S., & Lehmann, R. (2021, 11 30). *Function call gateway for operational aspects in Industrie 4.0*. Retrieved from https://ieeexplore.ieee.org/document/9613179/authors#authors

Cambridge Dictionary. (n.d.). Retrieved from Meaning of edge in English: https://dictionary.cambridge.org/dictionary/english/edge

Dell Technologies. (2021). *Dell Technologies Validated Design for Manufacturing Edge with Litmus - Design Guide*. Retrieved from https://infohub.delltechnologies.com/l/dell-technologies-validated-design-for-manufacturing-edge-with-litmus-design-guide/architectural-overview-10

Eswaran, K., & Gastpar, M. (2009). Distributed Source Coding. In K. Eswaran, & M. Gastpar, *Distributed Source Coding.* Academic Press.

Gartner. (n.d.). *Gartner Glossary*. Retrieved from https://www.gartner.com/en/information-technology/glossary/gateway

Gill, B. (2020, 11 03). Retrieved from 2021 Strategic Roadmap for Edge Computing: https://www.gartner.com/en/documents/3992656/2021-strategic-roadmap-for-edge-computing

Gold, J., & Shaw, K. (2021, 06 29). Retrieved from Networkworld: https://www.networkworld.com/article/3224893/what-is-edge-computing-and-how-it-s-changing-the-network.html

Gómez-Brandón, A., R.Paramá, J., Villalobos, K., Illarramendi, A., & R.Brisaboaa, N. (2021). Lossless compression of industrial time series with direct access. *Science Direct*.

Katie, C. (2021, 02 11). Retrieved from https://www.gartner.com/smarterwithgartner/gartner-predicts-the-future-of-cloud-and-edge-infrastructure

Kozegho Dosing Systems. (n.d.). Data Sheet - PolySys CS-150608-3.

Lockerman, J., & Kulkarni, A. (22, 04 2020). *Time-series compression algorithms, explained*. Retrieved from TimeScale: https://www.timescale.com/blog/time-series-compression-algorithms-explained/

Nahar, N., & Haque, M. J. (2017). Data Compression Methods and Analysis. *Communications on Applied Electronics (CAE)*.

Open Automation Software. (n.d.). *What is an IoT Gateway?* Retrieved from https://openautomationsoftware.com/open-automation-systems-blog/what-is-an-iot-gateway/

Red Hat. (2021, 05 14). *What is edge architecture?* Retrieved from https://www.redhat.com/en/topics/edge-computing/what-is-edge-architecture?sc_cid=7013a000002pu3pAAA&gclid=CjwKCAiAlrSPBhBaEiwAuLSDUJsZOUOqe2PutyMwipZnSKnaFbqH9uS6PAtpDQOX34QHMBmgQNpm8xoCmLAQAvD_BwE&gclsrc=aw.ds

Sayood, K. (2002). Encyclopedia of Information Systems. In K. Sayood, *Encyclopedia of Information Systems.* Academic Press.

Yadav, H. N. (2018, 06 26). *Why become a Data Scientist?* Retrieved from https://medium.com/@UdacityINDIA/why-become-a-data-scientist-2deff48cbf30