DIANOMIC

White Paper

What and why - storage and latency

“Data pipelines shift processing capabilities to the edge, the **need for large scale data storage is removed**, as the outputs from models can be sent directly to storage”

Daniel Riddoch
Date of release 24/01/2025

What do we mean?

As we digitise industry, more and more data will be collected. Depending on how the data is to be used, it will have to be stored and transferred. Data which is to be used for compliance, safety, or record keeping must be stored indefinitely, and that which is to be used for operational purposes should be stored for a set period of time. In all cases it may be necessary to transform some or all of the data so that it may be most efficiently used and stored. The exact nature of the end use will determine exactly how this is managed, but all solutions have a few common elements.

All solutions are costly, either renting cloud space or buying and operating physical hardware costs money. All solutions introduce a potential security weak point. All solutions introduce extra steps before data insights can be acted on and introduce a slowing down of reaction time.

How does the problem arise?

Events that we are attempting to monitor, capture and react to can often happen fast, particularly in optimised manufacturing processes. In some cases, such as high speed assembly lines or heavy and fast moving machinery, this can become extremely dangerous. Consequently it is critical to be able to act based on data as quickly as possible.

A data driven approach is necessary because most machines are too complex to be understood mechanically and for some simple measurements to be enough to calibrate a mechanical model. As such, a large volume of high quality data will be required to develop and train an AI/ML model.

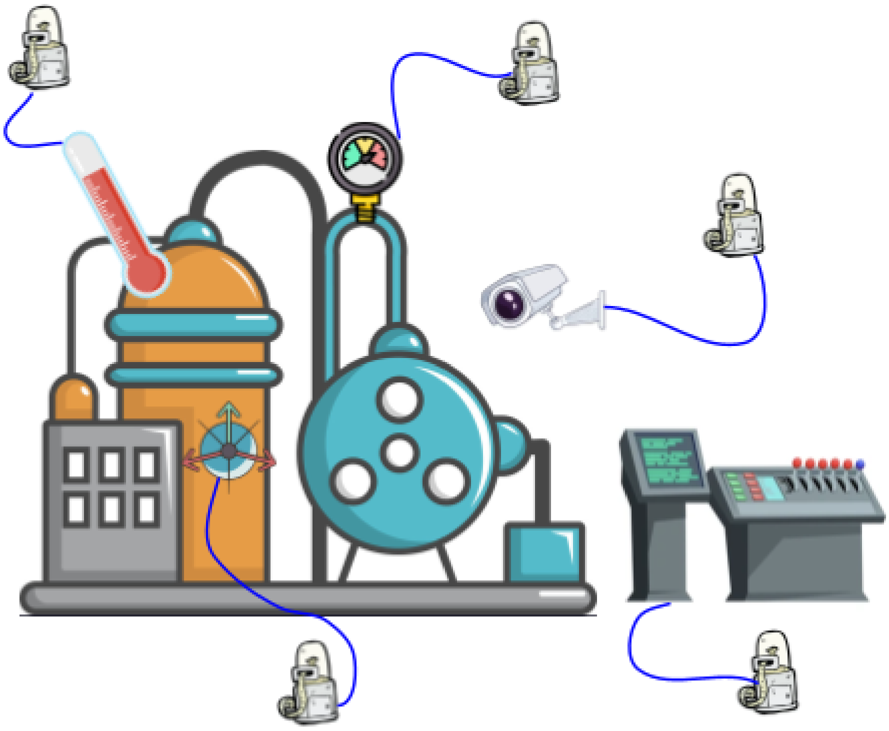

Figure 1: A sketch showing how one machine can produce huge volumes of data if properly monitored

Systems must, therefore, be developed to collect, handle, process, and apply data in very large quantities and at very high speed. This presents two problems with traditional approaches: moving such large volumes of data will create a latency problem; storing data in the quantities required will be difficult and expensive.

How do we solve the problem?

Let's re-examine what we need from the data. Often, we do not need to actually keep the data values, we only need to keep the insights the data provides. In that case, the data itself does not need to be moved far from where it is collected, we can simply move the processing closer to the collection. In other cases, not all of the data is required for modelling. Some data may be erroneous, some may be metadata which is not required once the data is normalised, and this data can be utilised or filtered at the edge, once the insights it provides are exploited.

How do we get the insights from the data, at the edge, store and apply these insights to new data, and do it all in a user-friendly manner that allows expertise to be developed and conclusions shared. The simplest way to achieve these objectives is with intelligent data pipelines.

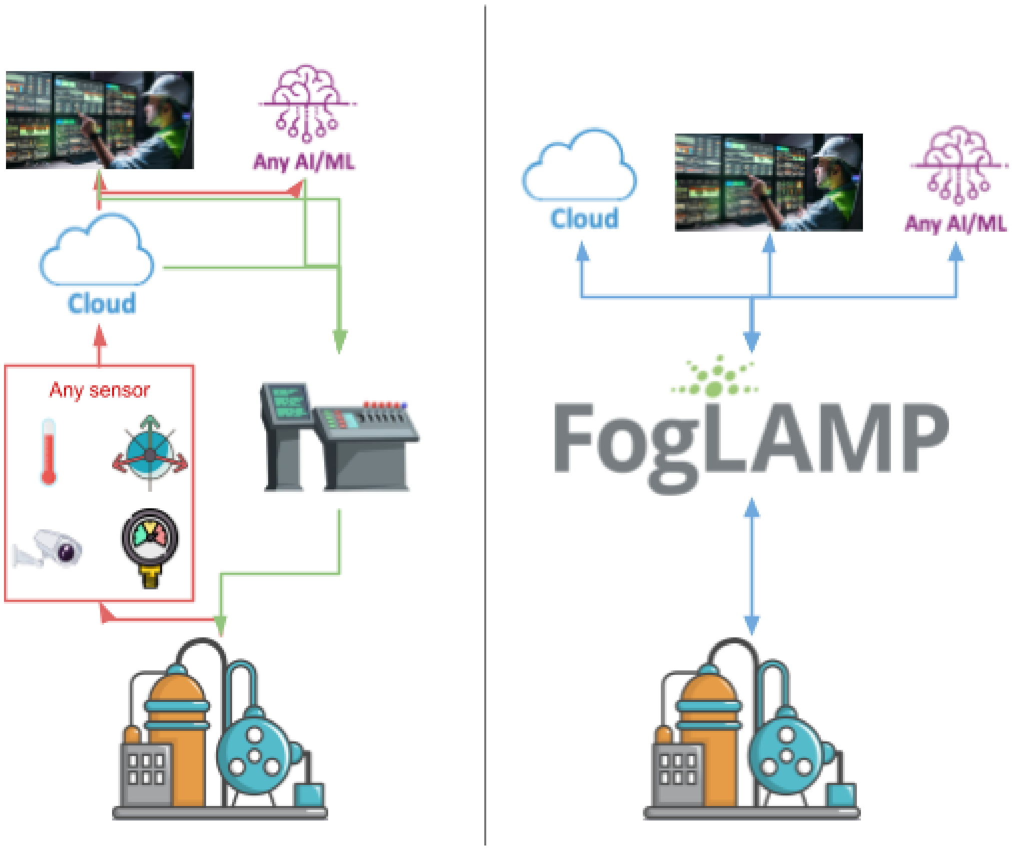

Figure 2: A sketch showing the simpler route taken by data using FogLAMP compared to traditional infrastructure

Data pipelines shift processing capabilities to the edge, the need for large scale data storage is removed, as the outputs from models can be sent directly to storage and/or display, rather than having to deal with raw data. This allows us to react faster to events, removing latency concerns, improving efficiency and safety.

What other benefits arise?

As already highlighted, all storage solutions cost money, and by drastically reducing the need for storing large volumes of data, big savings can be made without compromising the insights gained or the ability to look back at previous machine health and product quality.

By dealing with the latency issue, money will also be saved by efficiency gains. The ability to have a constant picture of the machine state is beneficial to catch quality issues early and to ensure compliance with safety regulations.

DIANOMIC

Example

Neuman Aluminum used Dianomic’s FogLAMP suite to change their data collection and storage infrastructure.
https://neumanusa.com/

Neuman Aluminum, working with the Austrian Centre for Digital Production, the Linux Foundation and Dianomic have redesigned their data collection, storage and transfer architecture to meet their specific requirements.

Neuman manufactures aluminium products for a wide range of industries, such as aviation, automotive, and defence; and across a wide variety of processes, including impact extrusion, heat treatment, and machining. In all of these, quality is of absolute importance, and this can be achieved by monitoring the machine processes and the manufactured parts.

Figure 3: Some of the components manufactured by Neuman Aluminium, including the car parts that the project deals with(Credit: Neuman Aluminium)

Neuman has multiple production facilities across Europe, and the project was focussed on facilities at Žarnovica in Slovakia. In this plant, quality control data was required by the end user of the product. In this case, the plant makes floor panels for cars, and the data is to be passed along with the product to assembly facilities.

Previously, Neuman used connected devices to send data into the Azure cloud. This data was then stored and could be accessed by the end user at any time. Unfortunately, there was no local buffer and as such, interruptions in network connectivity of any single monitoring device would lead to a stoppage in production because data could not be collected and so the part could not be sent to the end user.

The buffering capabilities of FogLAMP can be used to overcome the connectivity issue. Local data can be stored temporarily on the edge, and when network connectivity is restored, it can then be sent, meaning that production won't be interrupted by network issues.

The next step that Neuman will take is to investigate the possibilities of using ML/AI. Some possibilities include machine control, conditional forwarding, and adjusting safety devices. This has the potential to improve the efficiency of the manufacturing process, reduce the data burden and reduce or even mitigate safety risks.