DIANOMIC

White Paper

What and why - Data management

“Simplifying and uniting the data management infrastructure[using an intelligent edge data pipeline] will give a **single point of access** to check **operational upkeep and compliance”**

Daniel Riddoch
Date of release 24/01/2025

What do we mean?

There are many different ways in which we can gather insights about a machine or process. This can include sensors on the machine, feedback from the machine’s controllers, and even indirect measurements, like atmospheric conditions or measurements detailing the quality of the inputs or outputs of the machine.

The variety of data sources, and thus the variety and volume of data, can create several challenges, including commonality requirements, and working out how best to use the data. A simpler and more obvious problem perhaps, is that of data management. It is critical that the valuable data that has been collected is used most efficiently, and that requires an effective system to store, access, and transfer data at minimal cost.

How does the problem arise?

In many cases, machines will have different operating modes, and different indications of a problem, so it is important that data from different places is collected and used together. For example, vibration can be used to detect imbalances in a flow of liquid through a pipe, but a build up of heat in the liquid may also be dangerous, and could affect the vibration patterns and what should be inferred from them.

In the above example, it would be beneficial to build a complete view of the machine state, including vibration, operating mode, and temperatures. This would allow an ML/AI model to build up a full picture of the machine state and could result in better alerts, fewer false positives, earlier warnings and even predictions and prescriptions.

DIANOMIC

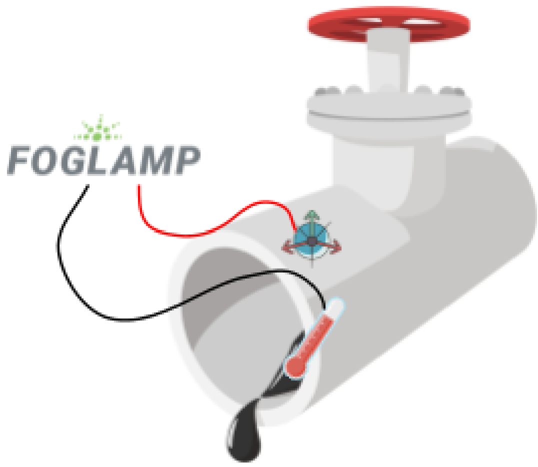

Figure 1: Sketch showing how to interconnected sensors can both provide useful information on the same issue

Several problems arise immediately from this. Perhaps the most obvious is that this kind of modelling will require vast quantities of data, collecting data from three sources instead of one will significantly increase the data volumes.

Perhaps a less obvious problem is sampling rate. It is extremely unlikely that all the sensors will sample at the same rate. Whilst this can be forced, it will often be a waste of resources and put excessive load on control systems, as operating mode does not need to be sampled at the same rate as vibration for example. Similarly, once the data has been collected, the times will need to be aligned so a consistent picture can be presented.

Finally, consider that data from different sources will be in different formats. Furthermore, it may require the purchase of additional infrastructure to connect them all to a single point of calculation. This will then be an expensive and slow transfer of data, to an expensive storage system which will be necessary to maintain the data for model training.

How can we solve the problem?

One possible solution to the last problem is data compression. In an intelligent data pipeline, lossless compression of the data can take place, which will reduce the cost of transfer and storage, without compromising the fidelity of the data.

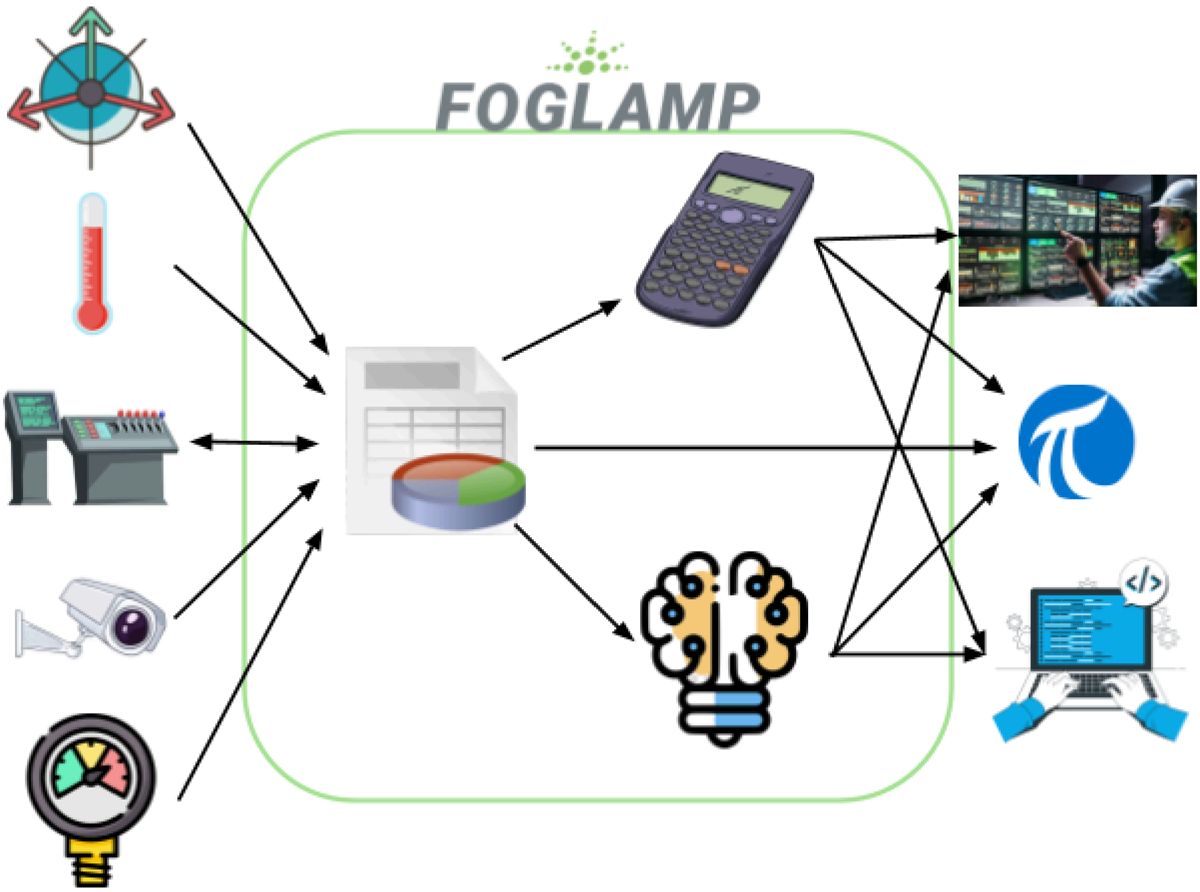

Figure 2: A sketch showing how a data pipeline like FogLAMP can ingest data from multiple sources, combine it, perform inference and send it on to various destinations

However, given that we now have an intelligent pipeline in place, why stop at compression? The pipeline can be configured to align the data and interpolate or downsample in such a way that a complete snapshot of the machine's health is always available at any time timestamp. As such, we deal with the sampling rate problems within the pipeline, streamlining data processing.

An intelligent pipeline will also be able to deal with data from different sources, and manage the data types. If the pipeline is capable of reading the input data(if not then a plug-in can be developed to facilitate this) then it will be able to convert the data into a format that can be aggregated. This format can be chosen to suit the end user.

Intelligent pipelines also have the ability to perform inference. It may be that a certain set of statistics are used for calculations, rather than raw data; these can be calculated within the pipeline, reducing the storage and transfer problems. Furthermore, logic and modelling capabilities exist within these pipelines, up to the level of ML/AI models being deployed at the edge. This completely eliminates the storage and transfer problems and all the data management challenges can be dealt with immediately and fast, fresh insights can be sent using an intelligent data pipeline.

What other benefits arise?

By reducing or eliminating the data management requirements, significant savings can be made in storage costs, and the security that may be associated with that can also be saved. Any restrictions placed on data collection by previous infrastructure will also no longer apply. For example, if previously data was only being collected at discrete intervals, perhaps because of buffering in past connections or slow conversion between types, it may now be able to be collected continuously.

Simplifying and uniting the data management infrastructure will also give a single point of access to check operational upkeep and compliance. It will unify the maintenance of the infrastructure and simplify ongoing processes.

Example

A large multinational energy company uses these techniques to monitor remote management and regulation stations on its pipeline network. For reasons of safety and security, exact details of this use case cannot be provided. Every effort is made to make clear how the use case works and the advantages offered without compromising security.

Remote metering stations are needed on long pipelines to ensure that the fluid flowing through the pipe, whether it be oil, LNG, or even water, is behaving as expected and is not being subjected to extreme conditions.

It would be very easy for a cavitation problem caused by a damaged pump or faulty filter to propagate and for the resultant vibration to cause failures. Similarly, if the pipeline is exposed, low temperatures could alter the viscosity of the fluid, which may affect its usability, or require additional efforts to insulate the pipeline. Furthermore, blockages and physical damage to the pipe may result in excessive pressure which, left untreated, could pose serious safety risks. All of these come with significant costs, safety risks, and regulatory compliance issues.

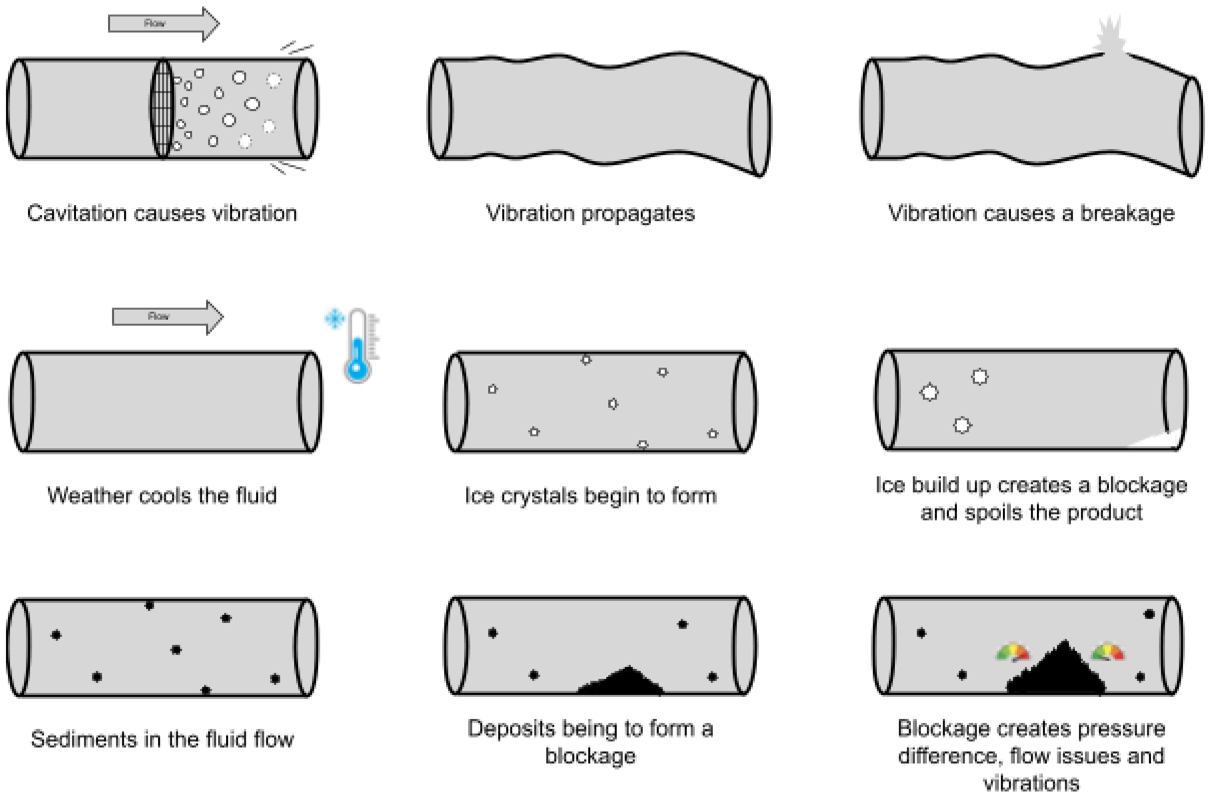

Figure 3: A sketch showing how cavitation, temperature, and blockages can cause different issues in a pipeline

It is therefore essential that these pipelines are monitored, and several different parameters should be monitored simultaneously. However, for reasons of safety and planning, the pipelines tend to run in remote areas. As such, direct monitoring is impossible, meaning a remote system is required. This leads to further challenges, as such remote locations rarely have stable power or networking, or ideal conditions for sensitive equipment.

Previously, all the sensor and monitoring systems required were installed and either connected to their proprietary gateways, supplied by the sensor manufacturers; or connected to a central gateway using bespoke infrastructure. The central gateway would then be connected directly to the network. This means that in the event of network failure, no data can be collected. Furthermore, no indication about data quality or usefulness is provided and all the raw data is sent over the network.

So, this company joined forces with Dianomic to utilise their FogLAMP suite to replace their infrastructure. A FogLAMP instance can be run on the existing gateway infrastructure, and its open source nature and library of plug-ins means it can connect to all the sensors required at the station. By installing an intelligent data pipeline, it is possible to allow for some buffering, meaning that the network connectivity requirements are not as stringent as they previously were.

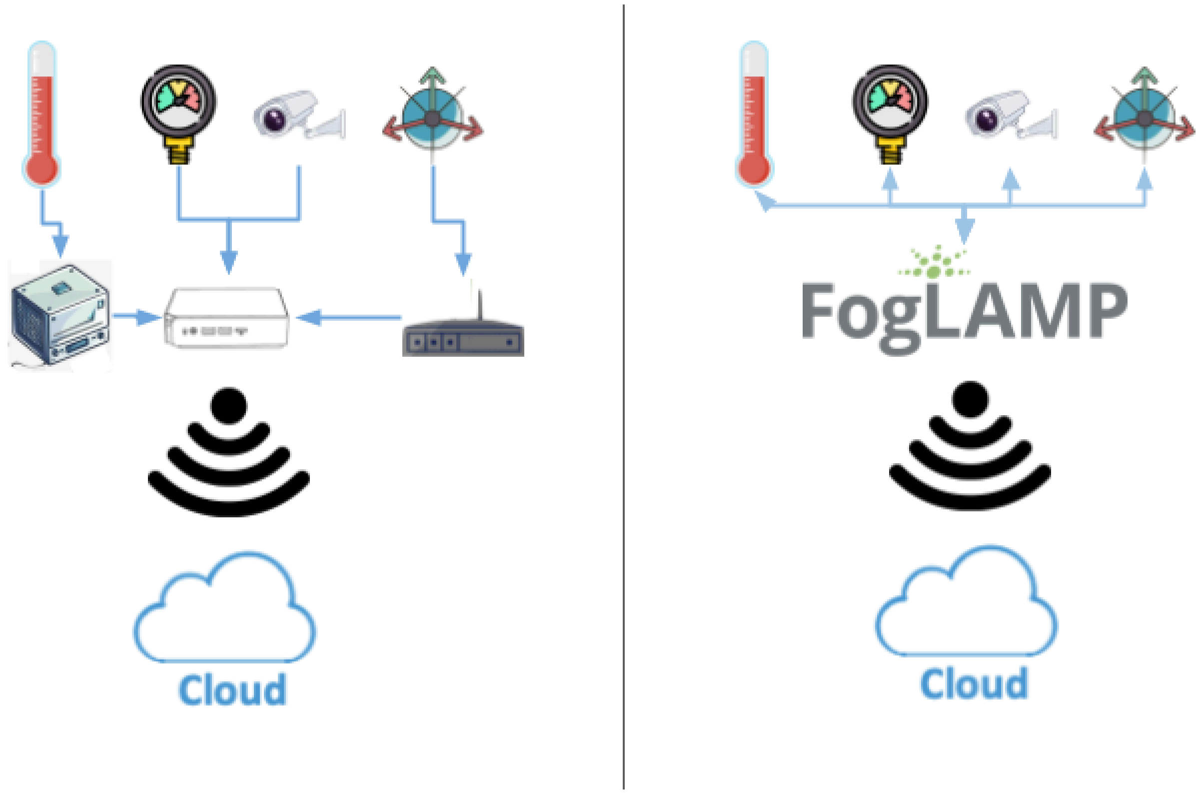

Figure 4: A sketch showing the difference between traditional architecture and a FogLAMP data pipelines structure

FogLAMP solution, data compression, scalable and reproducible, set points
The use of an intelligent data pipeline for this purpose also means that this solution is highly scalable and can be easily reproduced and deployed at multiple sites. There is also some increased functionality which can be used by data pipelines.

The data for transfer can be buffered and compressed to reduce network dependence. Similarly, inference can be performed, as simple as conditional forwarding to decide which data is necessary for condition monitoring, or even ML/AI calculations to gain deeper insights into the station performance. Finally, FogLAMP allows for set point control, which means that, with correct infrastructure, the station can be remotely controlled, perhaps even automatically, giving rapid responsiveness to changing situations.