

Intelligent OT Data Pipelines

What is a data pipeline?

Data enables business leaders to make optimal decisions that increase productivity and manage costs in their organizations. But ensuring decision makers have the data they need is a major challenge. Crucial data is scattered throughout the organization, in numerous formats, and often subject to corruption and bottlenecks. As an organization grows, these problems magnify in scale and impact.

Data “pipelines” are the heart of a robust and secure data practice. They automate the collection of data where it is generated, modify it into more usable forms, and reliably transport it to where it is needed. To accomplish this, pipelines offer numerous capabilities to buffer, analyze, transform, tag, filter and otherwise enhance the data as it travels.

Data pipelines provide an underlying foundation for faster, more accurate data-driven decisions. They enable your organization to:

- Analyze and act on data in a distributed environment.
- Enable machine learning and artificial intelligence tools and processes.
- Perform real-time data analysis.
- Respect systems of record.
- Integrate data stores, clouds, OT/IT, ISA95 systems.

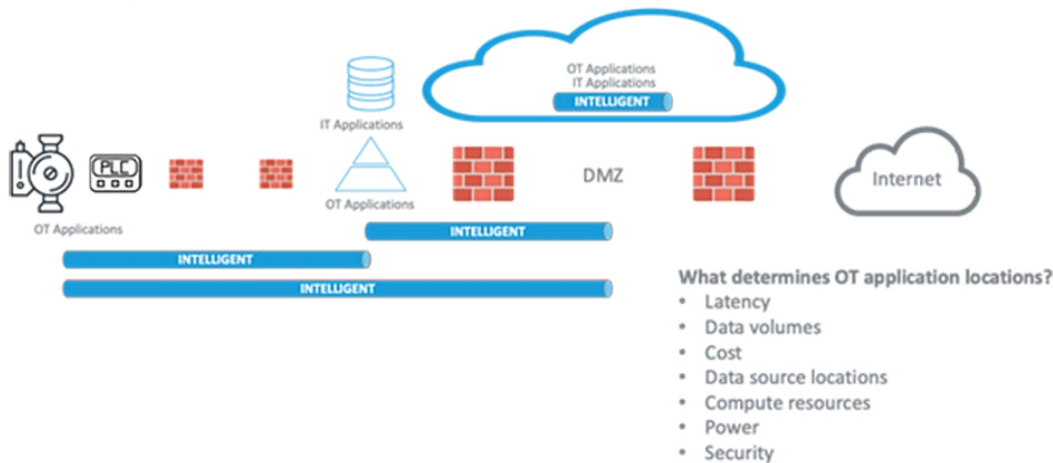
OT data pipelines are different

Operation’s technology requires scalable, secure, reliable, bidirectional, managed data pipelines from sensors and machines to on-prem operational systems and clouds. Pipelines enable OT-IT convergence that help deliver the goals of 21st century industrial transformations. However, compared with IT environments, OT environments have unique complex challenges on their data pipeline infrastructures. Among these are:

- Geographically distributed sensors, machines and processes.
- Challenging communications infrastructures.
- Highly heterogeneous data.
- Proprietary brown field systems with long life-times.
- Scalability — Enterprises may have thousands of machines and millions of sensors.
- Highly regulated and secure environments.

Machine learning places additional requirements on OT data pipelines. Machine learning inference is often required at the edge where the highest fidelity data can be used without latency, cost, compliance or bandwidth concerns. Edge ML also enables faster response times for services like automated controls. Additionally, OT data pipelines must support more distributed applications and analysis compared to cloud-centric pipelines.

OT Data Pipelines Operate Where Applications Live



OT Data Pipeline Capabilities

OT data pipelines require a unique set of capabilities to handle operational sources (machines, sensors, SCADA, DCS...), data types (time-series, event streaming, image, vibration...), and applications (OEE, MES, historians, HMIs...)

OT Data Collection and Delivery

There are over 100 proprietary OT “source” protocols and many cloud and ISA95 destination protocols. OT data pipelines architecture must allow for the use of any protocol to connect to both the sources and destinations of data. As such, OT data pipelines represent a significant improvement over 1990s-era multi-protocol gateways that provide one-to-one mappings of machine protocols to application protocols, creating silos of data.

OT Data Modeling

The machines and applications in the OT infrastructure all generate and consume data in unique formats. The 1995 multi-protocol gateway architecture requires scripting a translation for each machine-application pair, resulting in endless trouble shooting of broken integrations. Modern OT data pipelines solve this issue by coding a single plug-in for each machine and application that translates between the proprietary data format and a common JSON format. This architecture increases pipeline robustness while significantly reducing the time and effort required for implementation and support.

OT Data Filtering and Enhancement

Because of the volume and nature of OT data, OT data pipelines must support sophisticated processing at the edge of the OT infrastructure. Intelligent data reduction at the source means network, bandwidth, processing and storage savings for everything else downstream.

In addition, as machine learning and automation is pushed to the edge, all of the data transformation and enhancement required by the machine learning models must also be performed at the edge. As such, OT data pipelines require far more capable edge data processing capabilities than the more cloud-centric IT pipelines.

OT Data Pipeline Triggers

OT data pipelines must do more than just transform and move data. Machine learning inference provides insights into process and machine states or other conditions. OT data pipelines provide mechanisms to use this information to send notifications, trigger new processes (such as self-update) or even automate control functions.

OT Data Transport

OT environments can be challenging for networks. Power availability, long distances and wireless interference make reliable connectivity more difficult. OT data pipelines must support functions like buffering and data validation to reliably deliver data over intermittent or unreliable links.

OT Data Pipeline Management

Plants, mines and factories can have 1000s of machines with 100s of thousands of sensors and actuators streaming millions of tags and other data. Further, OT data pipelines often implement sophisticated data processing and applications distributed across numerous edge locations. To address these challenges, OT organizations must implement processes to build, deploy, secure and monitor OT pipelines at scale.

OT Expertise

The expertise of engineers, scientists, operators, maintenance teams and business managers are often required to develop OT-data pipelines and edge applications. Many of these employees do not know how to code. To build sophisticated OT data pipelines, OT organizations should implement no-code and low-code tools that enable the various teams to directly contribute their expertise.

Deploying OT Pipelines

OT data is often consumed by multiple applications, users and processes. For example, a machine's PLC data may be used to monitor machine health, power consumption, and output quality. Distinct OT data pipelines are required to filter and transform the PLC datastream to meet the requirements of each different application. These requirements result in rapid growth of the number and complexity of pipelines and configurations.

Deploying machine learning to the edge further complicates OT data pipeline management. Models and their pre-processing filters change frequently, requiring updates to the data pipelines that house and serve them.

OT organizations will require robust enterprise management capabilities to cost-effectively ensure this infrastructure remains reliable. Some of the capabilities that will be required are:

- Abstraction of assets, data sources, pipeline logic, destination integrations and Edge MLOps to allow operators to understand the infrastructure.
- Automation of pipeline provisioning, configuration, updates, rollback and deletion.
- Robust change control and versioning.
- High-Availability.
- Auditing.

Securing OT Pipelines

OT systems are often located in critical infrastructure, healthcare, and other environments that have strict regulatory, safety and security requirements for their data. These policies may restrict users access, premises and networks that data is allowed to transit and data flow directions. As automated control functions are integrated into OT data pipelines, security becomes even more paramount. OT data pipelines must offer a rich enough set of security policies to support these requirements at scale. The pipelines must also support the security technologies and policies of the data sources and data integrations.

Monitoring OT Pipelines

As OT data pipeline use grows, they form the underlying data feed for multiple mission critical applications and functions. To ensure reliability, OT organizations will need systems to monitor their data pipelines performance. In addition to monitoring data flow availability, these systems should also monitor the data itself for anomalies to ensure that sensors and transformations are operating properly.