software AG | aws

IoT that scales:
build enterprise-grade solutions with Software AG and AWS

Table of contents

Introduction

Manufacturers today appreciate the value of **smart, connected products**. From builders of heavy machinery to makers of precision instruments, the Internet of Things (**IoT**) helps enterprises better understand how their products are used, **predict when they need maintenance**, and develop new connected features that strengthen relationships with customers. IoT can also drive new business models. The global market for **Equipment-as-a-Service** (EaaS), where customers pay based on the use of a product rather than buying it outright, is projected to hit $131 billion by 2025.[1] That's up from $22 billion in 2019—made possible by the ability of IoT systems to monitor and track device usage and performance.

Yet achieving IoT that scales can be difficult. A 2020 survey by Beecham Research, based on interviews with over 400 IoT adopters, found that only 42% of **IoT projects** were either fully or mostly successful. 60% of all enterprises surveyed reported trouble scaling. Others faced challenges like a lack of clear business objectives, insufficient network connectivity, internal resistance to change, or a failure to engage outside experts.

Enterprises that have begun to explore the possibilities of IoT can likely relate. Perhaps they've gotten promising results by tinkering with early prototypes and are now looking to scale. Or they've attempted an in-house build solution, and realized it's time to seek help from a solutions provider. Whether organizations want to build entirely new IoT capabilities or extend the current ecosystem, the solution needs to be easy to implement and maintain—and, above all, cost-effective.

Software AG and Amazon Web Services (AWS): a joint solution

Software AG's Cumulocity IoT, named a *Leader* in Gartner's 2022 Magic Quadrant™ for Global Industrial IoT platforms, is an ideal place to begin developing a long-term IoT vision. As a full-scale enterprise IoT SaaS platform, its **buy and build approach** enables users to deploy their projects quickly, by leveraging strong out-of-the-box functions and at the same time developing proprietary capabilities for market differentiation.

Cumulocity IoT is most commonly deployed in the cloud through Amazon Web Services (AWS). Not only is AWS the world's most comprehensive and broadly adopted cloud provider: it also offers top-of-the-line integrated services supporting full-scale **IoT solutions**. Topics covered include analytics, **Machine Learning**, Business Intelligence, and **digital twins**. There is a great deal of synergy between Cumulocity IoT and AWS IoT offerings as many capabilities are complimentary. Real-world implementations have shown that when an organization combines technologies from both vendors, it has access to all capabilities required for cutting-edge, robust, and future-proof IoT solutions.

More than any IoT solution on the market, Cumulocity IoT's key proposition is that it removes the burden of building the nuts & bolts of IoT, so enterprises can focus on building specific capabilities. The highly intuitive user interface enables frontline engineers and non-developers to build powerful applications and easily manage their business-specific data processes, with little or no coding required. Its core **device management solutions** enable enterprises to rapidly connect existing assets and make it easy to manage millions of devices, with support throughout their entire lifecycle. Its dashboarding, alerting engines, and **streaming analytics**, are intuitive and user-friendly. To harness up-to-the-millisecond data, most platforms require technical staff to write sophisticated queries; Cumulocity IoT demands nothing more than clicking a few boxes together.

The platform is also ideal for enterprises that require IoT to operate on-premises: **Cumulocity IoT Edge** is a single-node variant of the IoT cloud SaaS platform, which can be deployed on local servers or in data centers.

1.   IoT Analytics, Equipment as a Service Market Report 2020–2025

AWS offers a broad set of complementary global cloud-based products. These include computing, storage, databases, analytics, networking, mobile, developer tools, management tools, security, and enterprise applications; available on-demand, in seconds, scalable, with pay-as-you-go pricing. AWS also offers more than 200 services, including data warehousing, deployment, directories, and content delivery. New services can be provisioned quickly, without the upfront fixed expense. This allows companies of all sizes to respond quickly to changing business requirements.

On top of this, AWS has a wide array of services that support **IoT programs** from the edge to the cloud. For example, AWS offers cutting-edge visualizations: its highly immersive Digital Twin maker gives users the ability to create 3D representations of the real world, and its integration with the 3D data platform Matterport allows for the scanning of places like factory floors to generate photorealistic maneuverable virtual spaces.

This white paper examines how **Cumulocity IoT** and AWS jointly cover a wide range of IoT capabilities: **device connectivity and management**, data persistence and query, digital twins, Machine Learning, business intelligence, and more. It shows how a solution that integrates components from both vendors scales well and avoids data and vendor lock-in. It also illustrates that this strategy is cost-effective: Working with two vendors as part of a buy and build approach does not only deliver the most robust set of IoT capabilities; it can also help cut costs. One example is **IoT on the Edge**, which reduces IoT resource consumption and therefore improves cost optimization. Another example are the low-code features that provide the option to tackle formerly outsourced tasks in-house. In sum, this strategy is the ticket to both faster time-to-market for new solutions and a lower TCO compared to custom built solutions.

IoT platform basics and device management

It's relatively easy for most enterprises to build a simple IoT pilot solution based on common software development frameworks. It's far more difficult to set up an industry-strength IoT program that grows with the demands of business—while avoiding the software maintenance trap. As always, decisions made early are critical for long-term success. Two of the most important are avoiding the urge to build an **IoT platform** in house; and choosing an appropriate tool for building IoT web applications.

Device management

The Internet of Things has two fundamental characteristics. The first is visible in its name already: Devices are being connected using *Internet* technology. This promises easy and cost-effective communication based on widely accessible hardware and software. While using basic internet technology can appear effortless, it's often overlooked that it uses many different protocols to connect systems and transfer data. The Internet protocol (IP) is a network-layer protocol with a multitude of protocols on top: On the transport layer, there is TCP but also UDP. The application layer includes HTTP, which is most used in the internet known to consumers, as well as MQTT, which is dominant in the world of IoT. Here already, a build solution capable of supporting all of these often becomes far more complex than initially expected.

The second fundamental characteristic of IoT is that IoT systems typically include many, many devices. IoT is a fundamental driver of Big Data. Big Data is not just a *little bit more* data: The data volume and the characteristics are so different that Big Data requires completely new strategies and technologies to deal with it. For the same reason, connecting to and managing IoT devices requires approaches that differ from those used to connect more traditional IT systems.

A mature IoT platform provides capabilities for efficiently and effectively dealing with both of these characteristics. **The primary mistake companies make when they embark on an IoT journey is that they try to build an IoT platform themselves**. In virtually all circumstances, this approach is not sustainable. Even efficient development teams spend too much of their time building non-differentiating capabilities instead of focusing on solutions that actually address customers' needs.

Cumulocity IoT allows enterprises to easily connect to all of their IoT devices, whatever the protocol, and however large the volume of data. It efficiently monitors and controls up to millions of devices—while providing comprehensive lifecycle management, including firmware, software, configuration, and security-related tasks. Its ready-to-use capabilities eliminate the headache of a build solution, while cutting costs, reducing the development timeline, and freeing up resources to focus on what delivers the most business value.

Building an IoT web application

Once an enterprise achieves a solid device management foundation, the next step is to begin building IoT solutions, typically in the form of web applications. Here again, the key question is which technology to rely on. While every IoT solution is different, most require a comparable set of "mid-level" capabilities. Examples include visualizing device data in intuitive ways; configuring (as opposed to hard-coding) aggregations and combinations of this data; alerting; and providing different views for different end-user roles with corresponding user management and security.

Most companies look beyond general-purpose programming to expedite these tasks. But there is a confusing variety of tools to choose from. **A second major mistake companies make is choosing a tool that's not designed specifically for IoT applications**. Many turn to modern business intelligence platforms, which have some overlap with typical IoT applications but are less suitable to handling faster, more granular IoT data. Others utilize Grafana, which was conceived as a system monitoring tool—as illustrated by the fact that it was intended to replace Kibana in the well-known "ELK" technology stack. Again, system monitoring has some similar objectives to IoT but is not identical to it. Nevertheless, enterprises will often choose such tools for IoT projects, simply because they are used elsewhere already, and/or they are considered "free." This tends to result in bad buy vs build decisions by ignoring the costs of developers' time.

Putting measurements in context: From asset hierarchies to digital twins

The measurements taken by **IoT devices** need structure and business context in order to provide their full value. For example, industrial robots typically include several IoT devices that measure different data points: positions, pressures, temperatures, etc. The robot itself is called an asset; in this context, that means a physical object that comprises several devices for collecting data.

Assets are often structured in hierarchies. The most two obvious ones are spatial topologies and so-called partonomies, structures that explain the asset and its parts. A device is installed in a room, which is located on a floor, which is located in a building. Users can assign devices to multiple asset hierarchies from different domain models at the same time. Asset hierarchies may also describe logical, non-physical structures, like security classifications or technical standards.

Digital twins and their many possibilities

Today, many enterprises have moved beyond this simple concept and embraced digital twins: comprehensive software representations of assets and related processes that are used to understand, predict, and optimize performance. A digital twin "emerges" if one or more systems collect information about a physical asset, from "slow" master data to very "fast" IoT device data, and all the data is made accessible at a single entry point. These entry points often provide an immersive user interface, all the way up to virtual representations of entire production lines. A digital twin done right is at the center of data collection and federation for the asset.

Both Software AG and AWS offer sophisticated measurement-in-context capabilities. Software AG's product, the Cumulocity Digital Twin Manager, stands out for its code-free implementation and the appeal of its dashboarding to manufacturers of industrial equipment and machinery. AWS' IoT TwinMaker provides an immersive 3D experience and seamless integration of data of all velocities using a graph database. For customers interested in a joint approach, the two solutions are easy to integrate: AWS IoT TwinMaker consumes Cumulocity asset hierarchies at design time, enabling users to define immersive, data-driven visualizations of the production line. Subsequently, TwinMaker scenes can be seamlessly integrated into a Cumulocity dashboard and consume Cumulocity IoT device measurements at run time.

Cumulocity Digital Twin Manager

The Cumulocity Digital Twin Manager allows users to define asset models and use them to build asset hierarchies. It comes with numerous user-friendly features for efficient, code-free implementation. Cumulocity IoT's dashboards can efficiently visualize assets and the measurements they send based on an extendable set of widgets—which can be made available to both equipment makers and their customers.

For example:

- Dashboards can be configured to show the IoT measurements of assets like industrial robots. This can easily be instantiated for a particular robot, switched between multiple instances, and even handle robot variants with a different set of IoT sensors by the same dashboard.

- Asset properties describing ownership and organization can be used to configure different dashboards of the same machine. In **Equipment-as-a Service** business models, one may be used by the equipment provider and another by the customer.

- Rules and triggers can also be associated with asset types. It's possible to configure Cumulocity IoT, for example, so that all rooms are kept at a certain temperature; or that all irregularities detected in a production line trigger an alert to the manager responsible.

AWS IoT TwinMaker

The AWS IoT TwinMaker is designed for a more immersive experience. Its strength is on the visual side, with sophisticated visualization of a physical, three-dimensional world.

TwinMaker enables users to compose 3D representations of the real world, known as *scenes*. These may either be created from scratch or by converting them from 3D/CAD or BIM systems, based on the glTF format. To obtain a digital representation of a physical space more quickly, TwinMaker is closely integrated with Matterport™. This product allows users to scan physical environments, like a factory floor, and directly generate photorealistic, maneuverable virtual spaces. It is especially useful for onboarding new operators and maintenance teams, rapidly orienting them to the environment and key machinery and processes.

These TwinMaker *scenes* can then be brought to life with data delivered through TwinMaker *connectors*, such as the Cumulocity connector for IoT measurements. A scene for a wind turbine, for example, might display metrics like wind speed while driving animations: a speed indicator near the rotor of a virtual turbine might actually turn at a speed based on the live measurement. Users can define rules to drive a variety of visualizations, adding to a more intuitive experience. For example, a rule may recolor the turbine's rotor in red if it is at risk of turning too fast or present a warning icon to focus an operator's attention on the most important elements of a scene at that moment.

On top of "live" IoT measurements, comprehensive digital twins include longer-living data of lower *velocity*, like maintenance records and even handbooks. Whereas the Cumulocity asset hierarchies are replicated as a hierarchy of entities in TwinMaker's knowledge graph, they can be enriched with additional non-asset entities from other data sources. This may include work orders held in enterprise ERP systems, process definitions, production schedules, or documentation repositories. Non-hierarchical associations between these rich but unconnected data sources can be added to the knowledge graph. For example, an association may relate a production line asset with all its historic work orders, fault history, engineering documentation and maintenance instructions along with upcoming sales order demand. TwinMaker supports graph-style queries which allow users to traverse these associations to create a holistic data view of the impact of production events or planned maintenance. This helps customer supervisors and planners prioritize remediation or regular servicing efforts.

AWS Glue and Amazon Athena are great companions for cataloging and easy querying (with SQL) data from structured data sources—especially data lakes—where log files, ERP exports, CMDB records, or other CSV-type files are collected over extended periods of time. TwinMaker has a built-in Athena connector which allows it to query this **data lake** data at runtime. By design, TwinMaker does not replicate data but always connects to the original systems of record at runtime. This ensures the data is as current as possible and avoids creating extra data silos which have to be kept in sync.
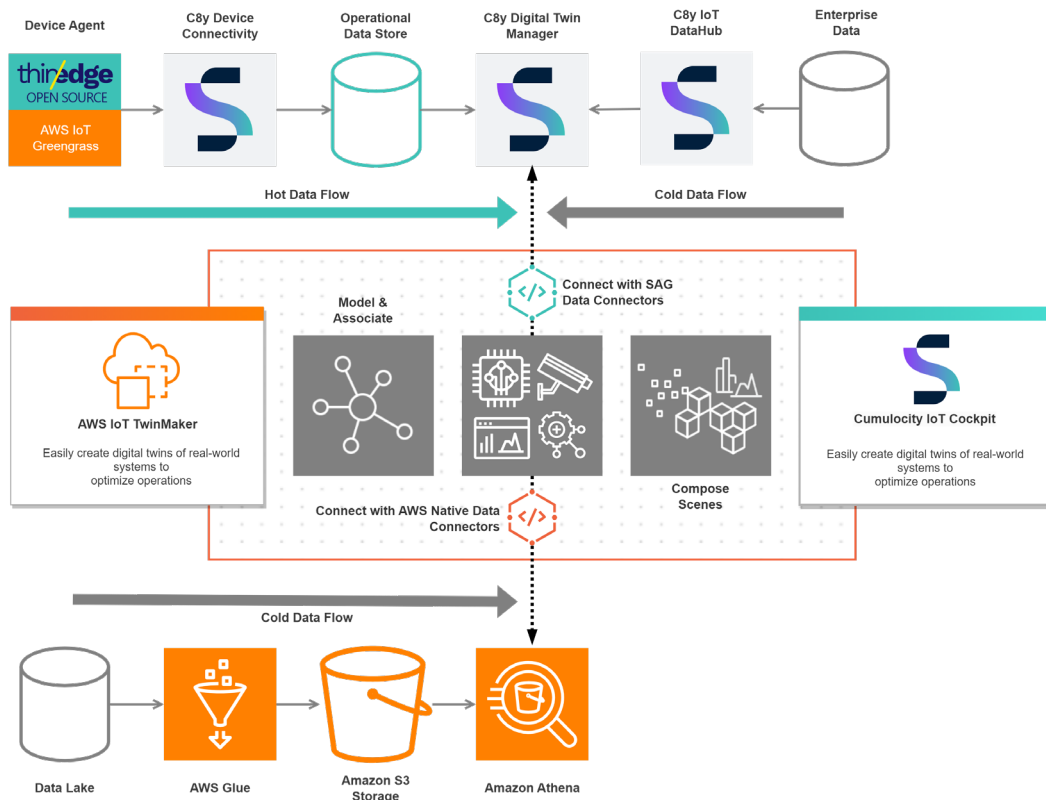

Figure 1. AWS IoT TwinMaker

Digital Twins: key takeaways

Cumulocity's Digital Twin Manager focuses on the code-free implementation of **IoT use cases** based on asset hierarchies, which can steer dashboards, rules, alarms, and streaming analytics. Its ease of use, and flexible dashboarding, makes it the ideal choice for all customers with frequently changing asset types, or for cases where digital services need customization when offered to end customers.

The AWS IoT TwinMaker powers innovative use cases that require free mobility in virtual three-dimensional *scenes*. Whenever the focus is on spatial information and the ability to navigate through 3D worlds, or enrichment with non-asset data, this is the ideal product.

Using the two tools in combination provides a more complete Digital Twin solution that will enable more representative models that provide a deeper understanding of a manufacturer's assets.

Analyzing *data at rest* and *data in motion*

Analytics and business intelligence are traditionally associated with working on stored data in a database, a data lake, or simply a file, as many Machine Learning solutions do. Such data sources are called *data at rest*.

In IoT, short-lived data from devices is continuously streaming in, and storing it *before* analysis is often highly inefficient if not impossible. One option is to use incoming data directly in computing memory (RAM). A second option is to work with streaming services like AWS Kinesis, which is designed to handle video and other types of fast streaming data. While these services also store data for a configurable, limited amount of time (a function known as *stream retention*), they prioritize low latency access over storage. These data sources are referred to as *soft real time* or (somewhat simplified) *data in motion*.

Monitoring and advanced analytics on data in motion

There are three core approaches to getting value from data in motion. In simple terms, **monitoring** is the observation of *known* system properties against a *known* target state. Consider an example from the pharmaceutical industry: If it is known that the temperature (property) of a chemical reactor must stay below 600 °C (its target state), the temperature can be measured and compared to the target state, with alerts set up to warn if anything is off. This measurement and comparison together are a form of monitoring.

Monitoring is the most basic form of **analytics**. However, advanced analytics reaches beyond this, by looking for unknown states and facts. In particular, *streaming analytics* is all about correlating data for new insights; more specifically, observing *known* properties against an *unknown* target state. During a regular production run of a batch of medicine, the temperature of the chemical reactor may follow a characteristic, but unknown curve. Streaming analytics can compare these curves over various production runs and detect outliers even if the curve is not known upfront. In this case, a single property is being correlated over various time windows. Another option is to correlate multiple properties over the same time window, for example, to find cause-and-effect relationships.

While **Machine Learning (ML)** is a highly diverse field the algorithms solving "classification" problems work in the exact opposite way: through the observation of systems with *unknown* properties against a *known* target state. A neural network trained on samples with given input and target states can later be used to consume inputs and predict outputs. For example, a network trained on pictures of both intact and defective parts can classify a new picture of a part as intact or defective. It does this *without* the user specifying what in the picture to look for, what the property is that characterizes a defective part.

While this approach can deliver major benefits for industrial IoT operations, it demands a few key prerequisites. First, data from IoT sources must be made available for both training and usage ("inference") of ML models. Second, the predictions from model usage must often be combined with data from other sources and presented in an integrated user interface. Finally, users must operate models at the best location for their intended purpose.

Streaming Analytics with Cumulocity IoT

One common challenge is knowing which approach is most effective, and efficient, for a given task at hand. While there is a broad range of problems ideally suited for ML, there are many situations where streaming analytics is the better (and more cost effective) option. Software AG understands the differences between all three approaches and can help an enterprise choose which one makes most sense.

Cumulocity IoT Streaming Analytics is an ideal solution for real-time analytics on data in motion. This tool, part of the wider Cumulocity IoT platform, focuses on performance; in fact, it was originally designed for **real-time analytics** of stock trades. It can ingest in-memory data as well as Kinesis streams. What sets it apart is that it enables users to configure solutions by assembling analytic blocks with a visual *analytics builder*. In many ways, it resembles a rule engine that is specialized for IoT: Actions are performed if complex conditions are satisfied. The analytics builder requires no coding and empowers domain experts like factory-floor engineers. This is unique for a streaming analytics engine of this kind.
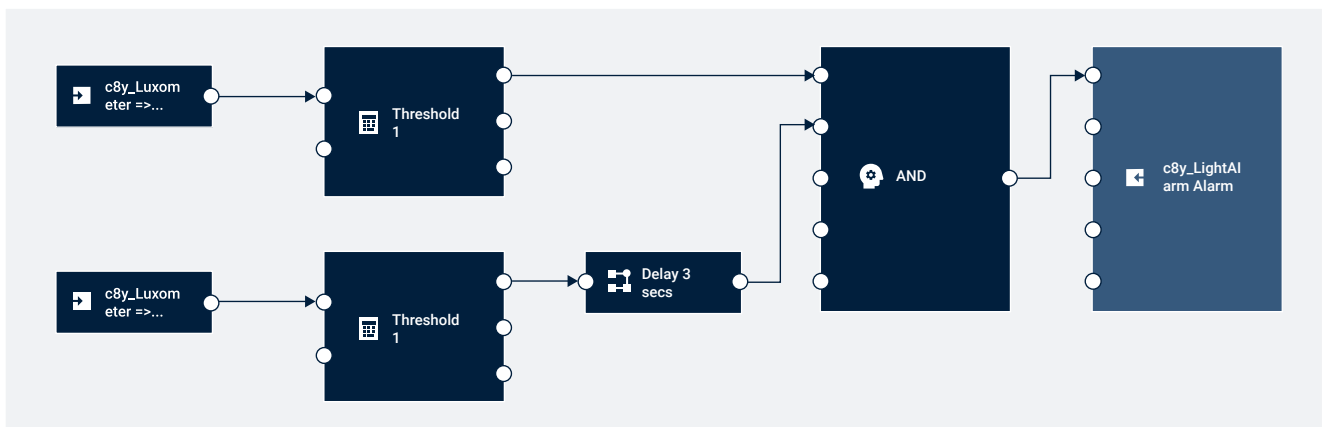

Figure 2. Drag-and-drop composition of streaming analytics solutions with the Analytics Builder

The analytics builder goes hand in hand with a coding environment that empowers developers to build even the most complex and in-depth streaming analytics use cases. Different data streams can be combined into complex "virtual" streams. Stream patterns can be detected and compared with historic patterns, which enables the detection of anomalies without having to know boundaries/thresholds upfront (like in the simple example above). Many believe that this capability can be provided by Machine Learning only. In this case, though, streaming analytics is an equally attractive or even superior option.

Analytics on data at rest

Monitoring, alerting, and streaming analytics work at the speed of business and are mostly used to support the *operational* use of IoT data. But the power of IoT goes beyond that. If an enterprise needs to leverage data analytics for strategic business insight, and IoT data is not included, then it is missing out on crucial information: for example, how customers are using connected products in the field. Fortunately, Cumulocity IoT and AWS technology empower teams to include IoT data in analytics with ease.

Cumulocity IoT comes with its own data store which is optimized for IoT, called an *operational data store*. Data analytics, on the other hand, is typically based on a data lake, with Amazon S3 as the ideal choice. In this context, the data lake is called the *analytical data store*. The operational data store is connected to the analytical data store by the Cumulocity IoT component *Data Hub*. In the analytical data store, the IoT data is combined with and put in context of other data from different sources, including non-IoT data related to the business.

Once the data reaches the analytical data store, it is fully accessible for further processing. There is no data lock-in or forced use of proprietary technology. Data Hub provides a productized, configurable integration with pre-existing data technology. This also means that existing analytics environments can easily be extended with IoT data. This is an example of the open and connected architecture of Cumulocity IoT.

There are two ways to read data from the data lake. The first is to again use Data Hub. It provides the capability to submit SQL queries to the data lake and comes with JDBC, ODBC, and ArrowFlight interfaces. Any analytics or business intelligence product that supports one of these can access the data directly from Data Hub. This option is particularly attractive if:

- A quick solution is required. Data Hub is directly embedded in Cumulocity IoT.
- The analytics solution is exclusively based on IoT data.
- The analytics solution can work with the IoT data without major preprocessing.

The second option is to index and query the data from Amazon S3 with Amazon Athena, an interactive query service, and use it for business intelligence, Machine Learning, and ML development with Amazon QuickSight. Similar to Data Hub, Amazon Athena supports using standard SQL for queries. Users can point to data in Amazon S3, define the schema, and start querying. For many use cases, this is a more efficient approach than applying the Extract – Transform – Load (ETL) paradigm which prepares the data for analysis before it is loaded into S3. This is often the case for analytic use cases based on IoT data. In essence, Amazon Athena lowers the hurdle for the analysis of large-scale datasets that include IoT data.

For more advanced analytic use cases, it may become a necessity to generate unified metadata for the data in S3. This includes so-called brownfield scenarios where S3 is already used to store analytics data from all kinds of sources. That's where AWS Glue, which is integrated out-of-the-box with Athena, becomes particularly helpful. As a serverless service, AWS Glue makes it easier to discover, prepare, move, and integrate data from multiple sources for analytics. The AWS Glue Data Catalog enables the creation of a unified metadata repository from multiple sources. It crawls data sources to discover schemas, populates a data catalog with new, modified table and partition definitions, and maintains schema versioning.

If the data needs to be processed further to make it "analytics ready," by transforming it or combining data from various sources, for example, AWS Glue can also be used to prepare the data for downstream processing.
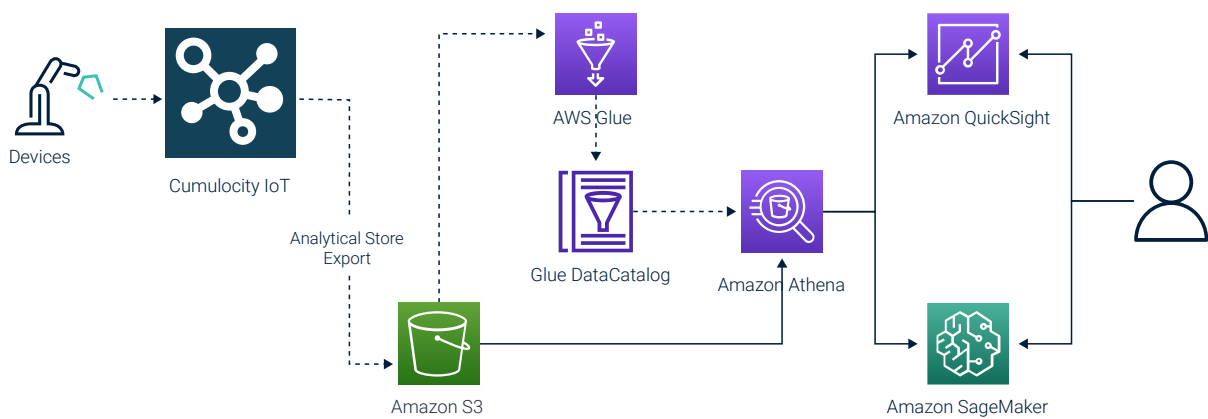

Figure 3. Analytics solutions using Cumulocity IoT and AWS product suite

Using AWS Athena and optionally AWS Glue is the preferred option if:

- A Business Intelligence and/or Machine Learning initiative is already underway (a "brownfield" situation) which should be supplemented by data from IoT.
- The IoT data needs to be catalogued, often as part of an existing initiative.
- The analytics solution is based on heterogeneous data from various sources, which needs preprocessing to make it "analytics ready".
- The data is used by other processes "downstream" of the IoT platform beyond analytics, for example, Machine Learning with Amazon SageMaker.
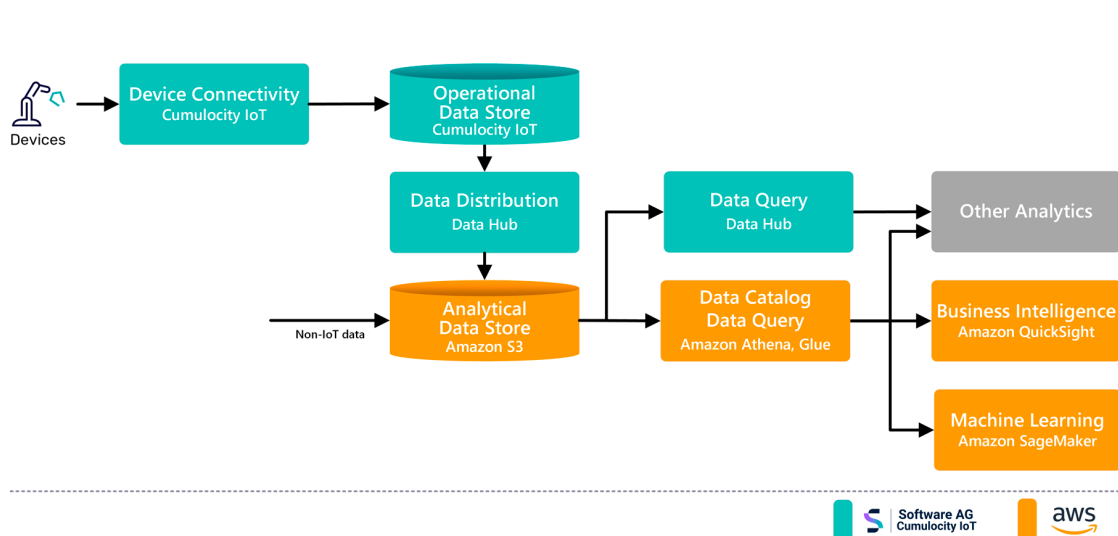- Amazon QuickSight is used as the business intelligence platform.

Figure 4. Data flow from device to analytics solution

Machine Learning in the cloud with Amazon SageMaker and Cumulocity IoT

As previously discussed, the Cumulocity Data Hub transfers IoT data to an analytical data store, such as an Amazon S3 data lake, where it can then be used for Machine Learning. Users can create datasets for both model training and usage, simplified by the provided SQL query capabilities. As the data store will typically contain data from a variety of sources, users can feed models with both IoT and non-IoT data, such as figures from production planning.

Amazon SageMaker is an ideal companion for Cumulocity IoT in this scenario. SageMaker is a fully managed Machine Learning (ML) service. With SageMaker, data scientists and developers can quickly build and train ML models, and then deploy them into a production-ready hosted environment. It also exposes ML model predictions via Amazon API Gateway. Industrial enterprises can benefit from the entirety of these features, starting with IoT data feeds from the data lake, and closing the loop with seamless integration of the results in the Cumulocity IoT cockpit.

Amazon SageMaker can be used with Cumulocity IoT as follows:

- **Data preparation**: Amazon SageMaker Data Wrangler imports, prepares, transforms, visualizes, and analyzes data. It is integrated into ML workflows to simplify and streamline data pre-processing and feature engineering using little to no coding. From Amazon SageMaker Studio, Apache Spark, Hive, or Presto on Amazon EMR or AWS Glue can be used to scale data preparation.

- **Building and training Machine Learning models**: One of the best ways to use Amazon SageMaker is to train and deploy ML models using SageMaker notebook instances. The SageMaker notebook instances help create the environment by initiating Jupyter servers on Amazon Elastic Compute Cloud (Amazon EC2) and providing preconfigured kernels with the following packages: the Amazon SageMaker Python SDK, AWS SDK for Python (Boto3), AWS Command Line Interface (AWS CLI), Conda, Pandas, deep learning framework libraries, and other libraries for data science and Machine Learning. The training stage of the full ML lifecycle spans from accessing the training dataset to generating a final model and selecting the best-performing model for deployment. SageMaker **provides** dozens of built-in training algorithms and hundreds of pre-trained models. SageMaker also provides no-code or low-code ML solutions such as **SageMaker Canvas**, **SageMaker JumpStart within SageMaker Studio**, or **SageMaker Autopilot** to train a Machine Learning model.

- **Deploying Machine Learning models**: After the Machine Learning model is trained, it can be integrated with Cumulocity's Cockpit using Amazon API gateway to get predictions.

- **Monitoring Machine Learning models**: Amazon SageMaker Model Monitor monitors the quality of Amazon SageMaker ML models in production. It allows users to set up continuous monitoring with a real-time endpoint, such as a batch transform job that runs on a regular schedule, or conduct on-schedule monitoring for batch transform jobs that run asynchronously. Alerts set up in Model Monitor notify users in case of deviations in the model quality. Model Monitor provides the following types of monitoring:

  - monitors drift in data quality
  - monitor drift in model quality metrics, including accuracy
  - monitors bias in the model's predictions
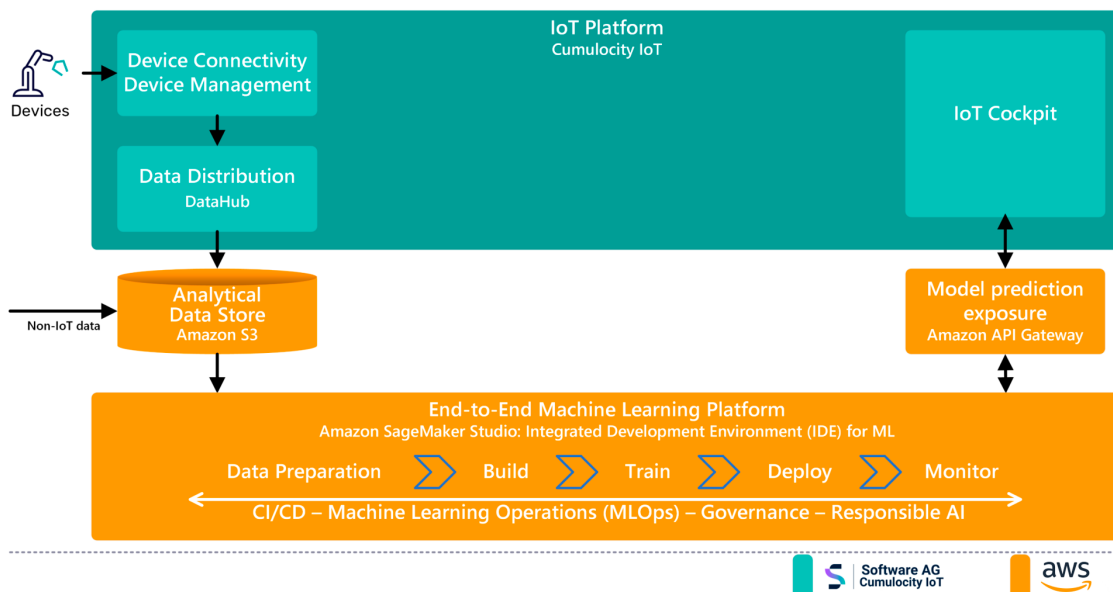 - monitors drift in feature attribution

Figure 5. Integrating Cumulocity IoT with Amazon SageMaker

The open architecture protects earlier investments in Machine Learning ("bring your own model"). Existing data lakes and ML models now also have IoT data at their disposal.

Machine Learning on fast data

IoT data is unlike any other data used for Machine Learning. It is *fast* data (of high velocity) for which the traditional ML approach of running a model against a dataset is sometimes insufficient. An example is quality control with an ML model consuming a video stream.

One way to overcome this is through **Edge computing**, which by definition means bringing computing resources closer to the place where they are needed. For ML with IoT, this means bringing the model closer to the sources of the data. There are three key drivers to do so: Performance, data security, and cost.

Cumulocity IoT is not limited to being available as Software as a Service; there is also a variant of the platform available for installation on the Edge. Every Cumulocity variant, including Cumulocity IoT Edge, has the capability to operate containerized workloads directly inside of it, a feature called *microservice runtime*. In the context of ML, microservice runtime can act as an operating environment for ML models. In this scenario, the data that the model consumes can be streamed directly to the model and never leaves the edge, with obvious advantages regarding performance, security, and cost.

ML models need an inference engine to run on. Some Cumulocity IoT users prefer to deploy microservices with both the ML model and their own inference engine. Others provide a trained model only, which complies with the widely adopted ONNX standard. For these, Cumulocity IoT provides guidance for using the open-source ONNX Runtime.

Finally, some models are operating even closer to the source of the data: Directly on the device, or on a very "thin", resource-constrained environment in its immediate vicinity. Solutions for these devices are discussed in the chapter below that covers Edge and device side computing.
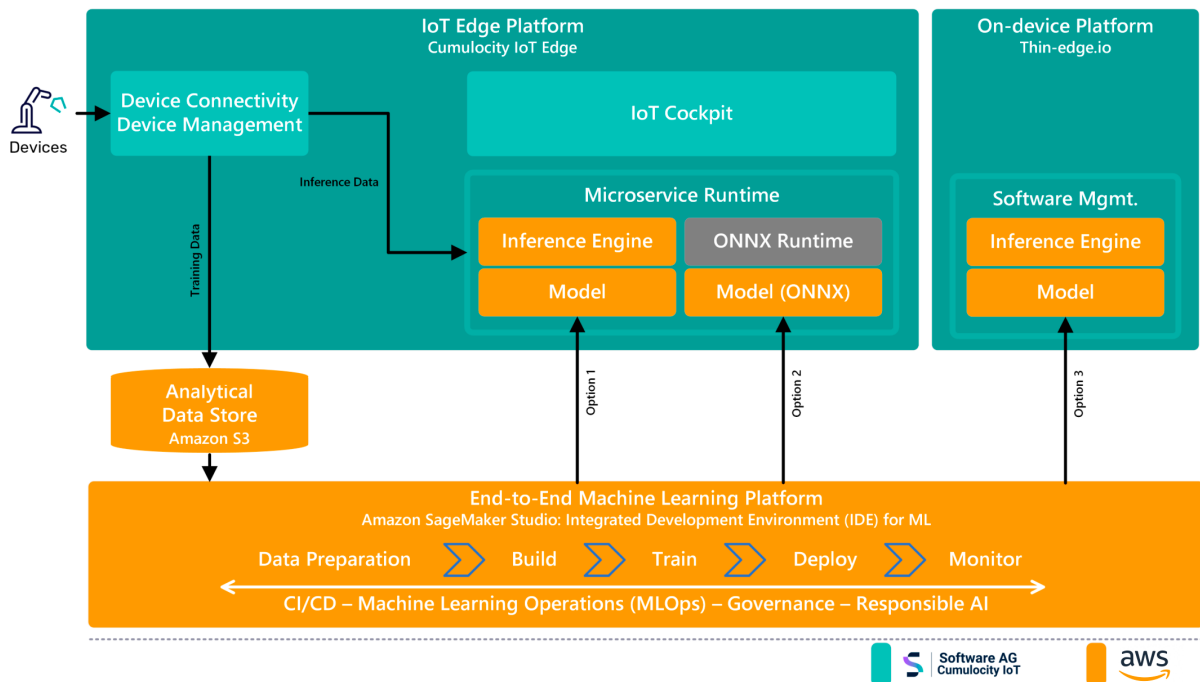

Figure 6. Deploying ML models on the Edge and onto devices

Machine Learning for Industrial Services: Amazon Lookout for Equipment

In addition to SageMaker, AWS provides a number of cloud-based ML Services that are geared towards industrial use cases. **Amazon Lookout for Equipment** is an ML industrial equipment monitoring service that detects abnormal equipment behavior so users can act and avoid unplanned downtime. Through proactive monitoring of an asset's condition, maintenance personnel can be alerted before issues occur. In addition to reducing costs, this leads to an increase in Overall Equipment Effectiveness (**OEE**).

Amazon Lookout for Equipment analyzes the data from sensors, such as pressure, flow rate, RPMs, temperature, and power, to automatically train a specific ML model based on the customer's data and equipment, with no ML expertise required. It then uses the unique, trained ML model to analyze incoming sensor data in near real-time and accurately identify early warning signs that could lead to machine failures. This enables enterprises to detect equipment abnormalities with speed and precision, quickly diagnose any issues, take action to reduce expensive downtime, and reduce the prevalence of false alerts.

Lookout for Equipment is used as follows:

1. In Cumulocity DataHub, set up a pipeline to export a dataset in a CSV format.

2. Ingest the dataset in Amazon Lookout for Equipment.

3. Train a model in Amazon Lookout for Equipment using the dataset.

4. Evaluate the Machine Learning model and **implement versioning** if the operation modes of the asset change over time.

5. Setup a pipeline to push live data from Cumulocity DataHub in CSV format to the input Amazon S3 location where the ML model is reading the data.

6. Create an inference scheduler in Amazon Lookout for Equipment.

7. Set up a pipeline to export the inference results from Amazon Lookout for Equipment and display the results in Cumulocity Cockpit using an **API Gateway**.
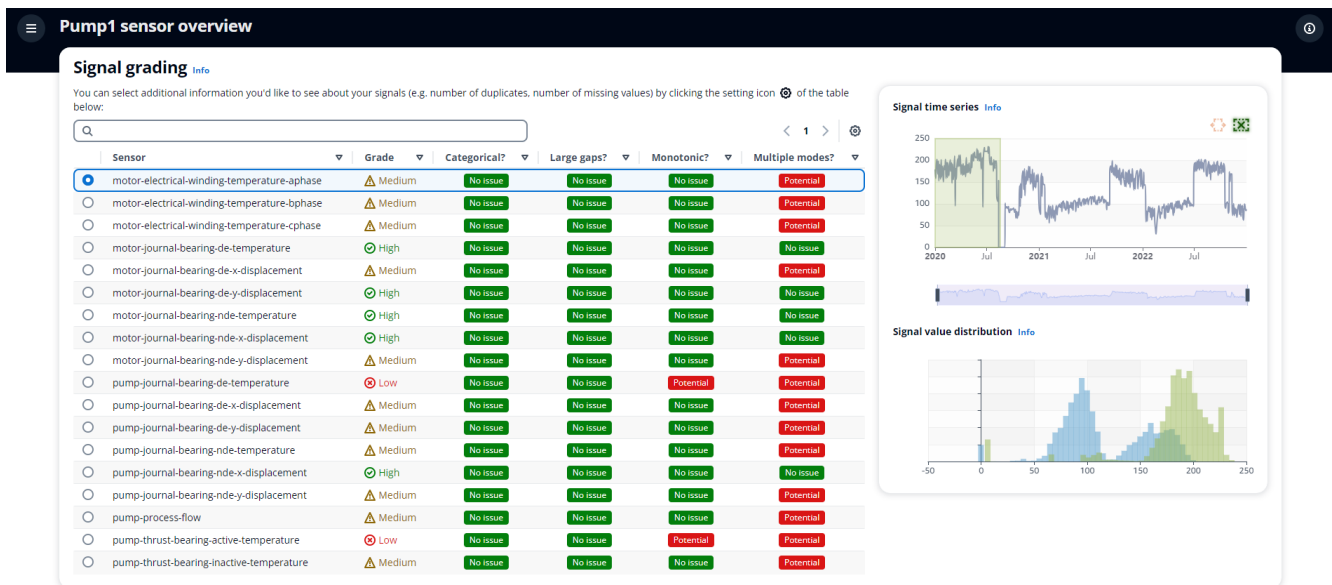

Figure 7. Signal grading with lookout for equipment, showing potential failures

Ingesting industrial media for Machine Learning

Amazon Kinesis Video Streams is a fully managed AWS service that can be used to stream live video from devices to the AWS cloud, build applications for real-time video processing, and perform batch video analytics. With Kinesis, users can ingest video and non-video data from a variety of sources, including smartphones, security cameras, web cameras, drones, thermal imagery, and audio. The video stream from Kinesis can be routed to AWS Machine Learning services like Amazon Rekognition, Amazon SageMaker or Amazon Lookout for Vision for object detection or product inspection.

On-device and Edge computation for IoT

The term *IoT device* covers a broad range of products. The simplest ones record a single measurement, such as temperature, or receive simple instructions like a binary command (for example, "on" and "off" for an IoT lightbulb). Other devices are more complex and have significant processing power that allows on-device computing, including the use of Machine Learning. (This chapter uses the term *on-device computing*, but all instances can be applied to IoT Edge computing as well.)

The fundamental idea of Edge computing is to move processing closer to the data sources, and IoT is a major use case for that approach. If a camera cannot be used to run ML models for image recognition, an option is to use an Edge computing device close to it. The term "close" has various dimensions here: it not only refers to physical proximity but also networking, security zones, etc. A major example of Edge computing is multiaccess edge computing (MEC) used in telecommunications networks.

The most important requirements for on-device computing, from basic to more advanced, are as follows:

- Network connectivity and messaging
- Device agent-supporting Cloud-based device monitoring and alert handling
- Firmware management
- General purpose processing, including software management (based on Snap, Docker, etc.)
- Machine Learning inference and model management (as a part of MLOps)

Software AG and AWS offer a complementary set of tools for on-device computing. Both tool sets have important elements in common: they are free and open-source; they support any Linux environment; and they support any programming language.

Thin-edge.io is an open-source framework to develop lightweight, smart, and secure connected devices. Software AG is a co-founder and a major contributor. Thin-edge.io supports all the functional requirements listed above. It's also cloud agnostic and known for its hardware efficiency: it is designed from the ground up for resource-constrained environments, can operate on only 16 MB of RAM, and uses Eclipse Mosquitto as a lightweight, open-source MQTT broker for communication. It's also reliable and resilient: It can deal with network outages and process restarts. The reliability mindset of its creators is also illustrated by the fact that it is written in Rust, a programming language well-suited for stable and reliable solutions.

The central benefit of thin-edge.io is its ability to reduce the development effort to build smart IoT agents without compromising on quality and feature completeness. It comes with a rich domain model for both telemetry and operations, providing efficiency and standardization. It easily combines ready-to-use software components with application-specific extensions. Its users can build an on-device agent in minutes.

AWS Greengrass is an open source IoT edge runtime and cloud service that emphasizes the software building process. On the AWS Cloud, the *AWS IoT Greengrass Cloud service* provides capabilities to build, deploy, and manage device software in any language. On the client side, the *AWS IoT Greengrass client software* supports this with:

- Messaging
- On-device processing, including the ability to execute AWS Lambda functions
- Accelerated development through pre-built components available for the devices
- On-device Machine Learning (inference)

In combination, thin-edge.io and AWS Greengrass provide capabilities for the broadest set of use cases. If AWS Cloud customers need to build a solution operating on devices from the ground up, Greengrass is likely a better bet. This includes Machine Learning that is developed in the AWS Cloud. On the other hand, the strengths of thinedge.io come into play in two types of scenarios:

- The device (possibly very resource-constrained) needs efficient device management and monitoring throughout its life cycle, based on predefined, standardized functions and domain models.
- Customers have already developed solutions in any programming language and on any cloud platform, and they need to be deployed and operated on the device.

The Greengrass and thinedge.io toolsets are not mutually exclusive. Consider a video camera managed in the cloud with Cumulocity IoT. It connects to an on-device agent built with thin-edge.io which provides connectivity, monitoring, and firmware management. The camera provides on-device Machine Learning capabilities. Amazon SageMaker provides the ML model to anonymously detect people in the room, and AWS Greengrass is used for developing supporting functions that operate on the device as AWS Lambda functions. The only information that the camera provides is the presence of people, with major advantages in terms of bandwidth and data privacy, among others.

Deployment and infrastructure

Cumulocity IoT is available as Software as a Service (SaaS) hosted by Software AG Cloud Services, with AWS as a primary provider of the underlying infrastructure services (IaaS). Additional deployment options include deployment into dedicated customer environments on AWS via VPC, and deployment of the full Cumulocity IoT product at the Edge (as described in the chapter on Machine Learning).

Why IaaS on AWS for IoT?

There are several reasons why customers choose AWS IaaS for IoT. As detailed below, AWS' primary advantages include its global infrastructure, network availability and performance, and scalability with demand to billions of requests per second.

AWS global infrastructure

AWS is the world's most comprehensive and broadly adopted cloud, offering more than 200 fully featured services from data centers around the world. Its extensive **global cloud infrastructure** has millions of customers—including the fastest-growing startups, largest enterprises, and leading government agencies. The AWS Region and Availability Zone model, which is designed to give customers added protection from system failures, has been recognized by Gartner as the recommended approach for running enterprise applications that require high availability.
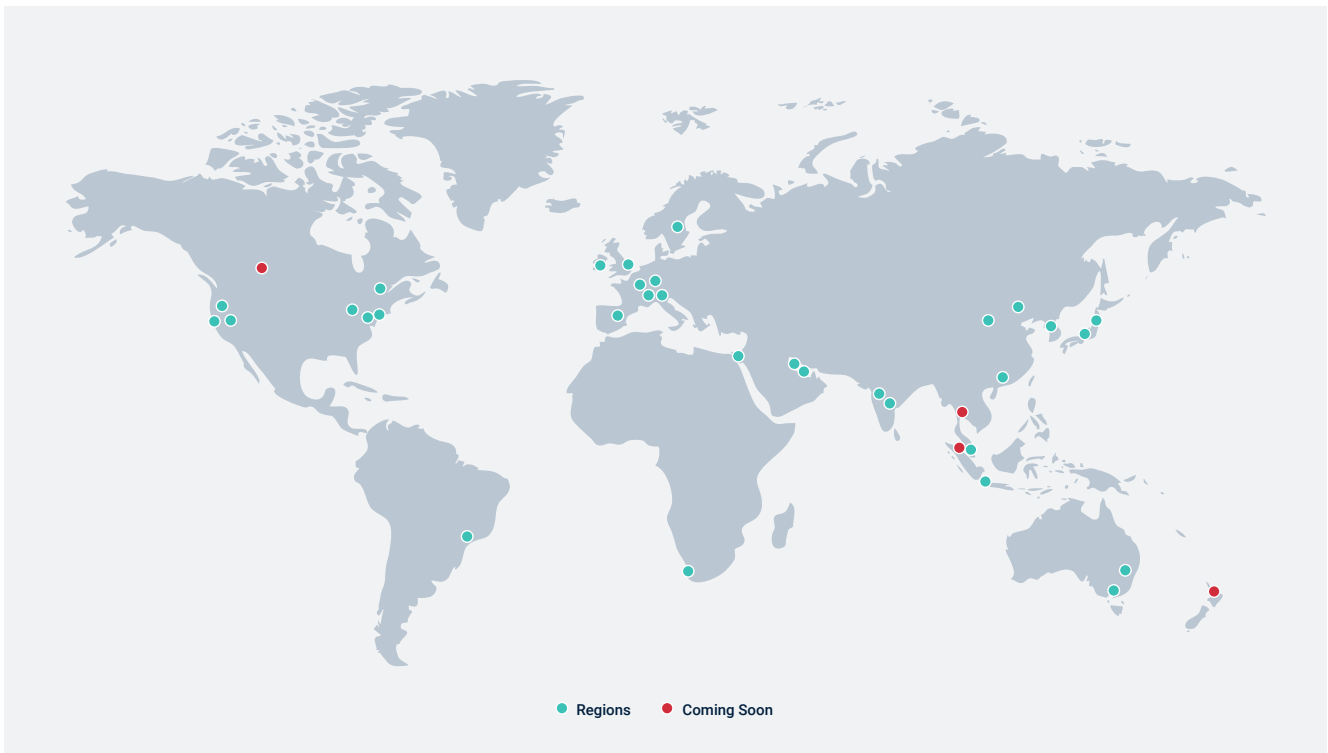
Figure 8. AWS global infrastructure

AWS network and performance

AWS delivers the highest network availability of any cloud provider. Each region is fully isolated and comprised of multiple Availability Zones, which are fully isolated partitions of infrastructure. The AWS Global Infrastructure is built for performance. AWS Regions offer low latency, low packet loss, and high overall network quality. This is achieved with a fully redundant 100 GbE fiber network backbone, often providing many terabits of capacity between regions.

Cumulocity IoT SaaS region availability

Cumulocity IoT customers can choose between the following public cloud instances that are operated in corresponding AWS regions:

- **Europe** cumulocity.com
- **USA** us.cumulocity.com
- **Japan** jp.cumulocity.com
- **Australia** apj.cumulocity.com

Customers with Cumulocity IoT operating in a dedicated cloud instance can choose any AWS region that offers at least three Availability Zones, as long as legal limitations like export control are observed.

Compliance & shared responsibility

AWS infrastructure allows customers to inherit the most comprehensive compliance controls. AWS supports 143 security standards and compliance certifications, including PCI-DSS, HIPAA/HITECH, FedRAMP, GDPR, FIPS 140-2, and NIST 800-171.

Security and Compliance is a shared responsibility between AWS and the customer. This shared model can help relieve the customer's operational burden as AWS operates, manages and controls the components from the host operating system and virtualization layer down to the physical security of the facilities in which the service operates. The customer is responsible for the management of the guest operating system (including updates and security patches), as well as associated application software and the configuration of the AWS-provided security group firewall.
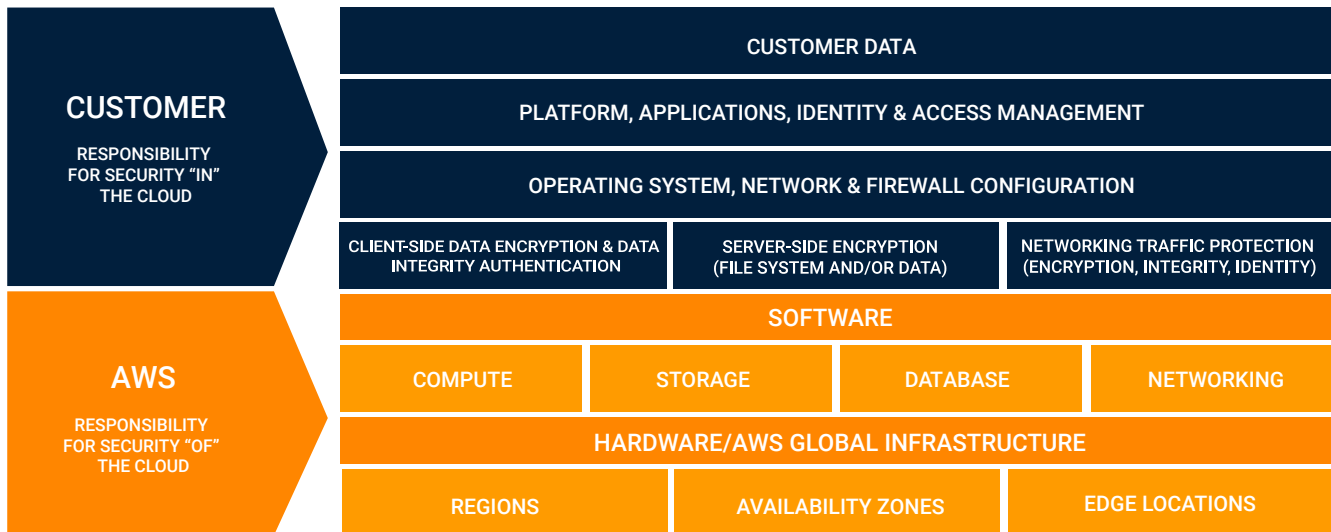

Figure 9. Compliance responsibility

Deployment options

Cumulocity IoT on AWS Marketplace
Cumulocity IoT SaaS is available through the **AWS Marketplace**, along with two distinct service packages from Software AG Professional Services to assist in developing digital solutions faster.

AWS VPC deployment
Some Cumulocity IoT customers prefer a dedicated system operating in their own AWS environment, which can be achieved through a Virtual Private Cloud (VPC) in any region of choice. With Amazon Virtual Private Cloud (Amazon VPC), AWS resources can be launched in a logically isolated virtual network that you've defined. This virtual network closely resembles a traditional network that can operate in a company's own data center, with the benefits of using the scalable infrastructure of AWS.

The following diagram shows an example of a VPC. The VPC has one subnet in each of the Availability Zones in the Region, EC2 instances in each subnet, and an internet gateway to allow communication between the VPC resources and the internet.
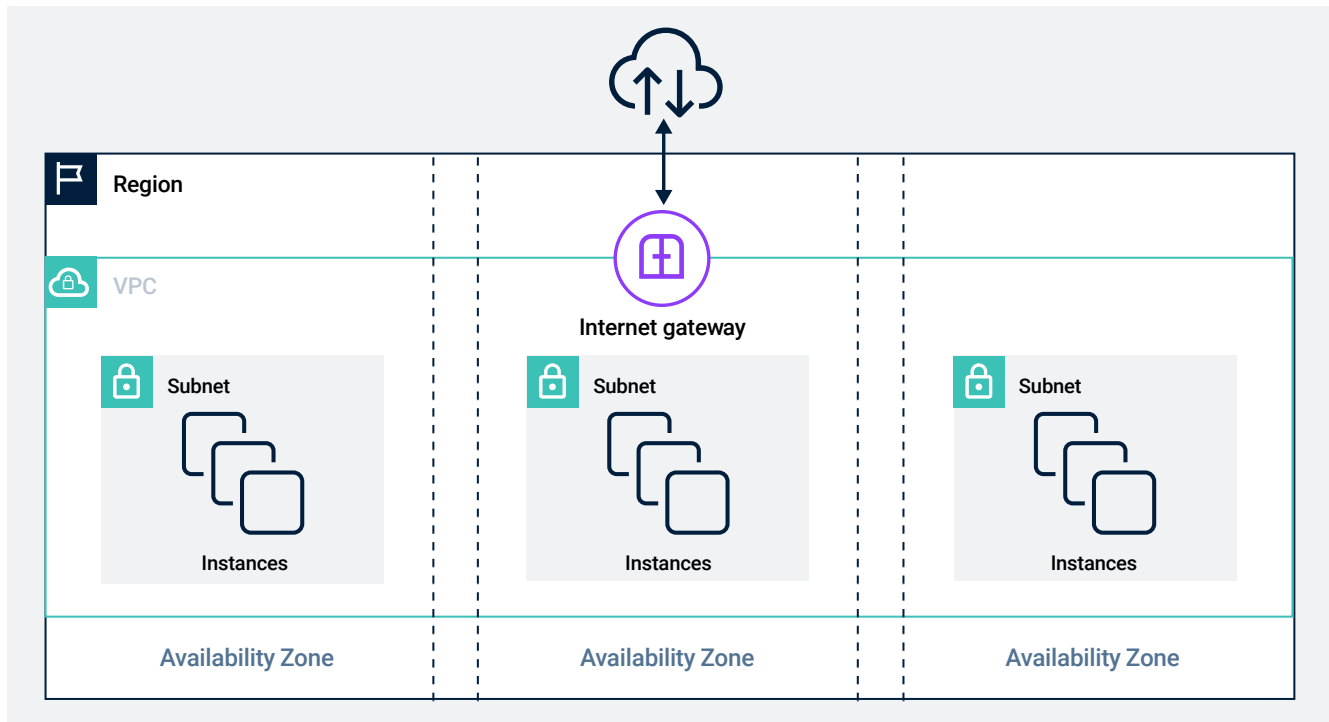
Figure 10. AWS VPC Deployment

The deployment of Cumulocity IoT into a customer owned AWS environment and VPC enables the integration and customization with over 200 fully featured services via API. One of these is AWS Lambda, a serverless, event-driven compute service that allows users to run code for virtually any type of application or backend service, without provisioning or managing servers. Code is uploaded as a ZIP file or container image; Lambda then automatically allocates compute execution power and runs based on the incoming request or event, for any scale of traffic.

The full picture

A joint capability architecture

Figure 11 shows a capability architecture that displays the key capabilities of Cumulocity IoT and AWS that are the subject of this paper. The architecture does not display all IoT-related capabilities that are available; for example, Cumulocity IoT's rule engine (*smart rules*) is missing. Instead, the architecture seeks to describe the relationship and interaction between Cumulocity IoT and AWS on a high level. The arrows indicate the main direction of data flow.
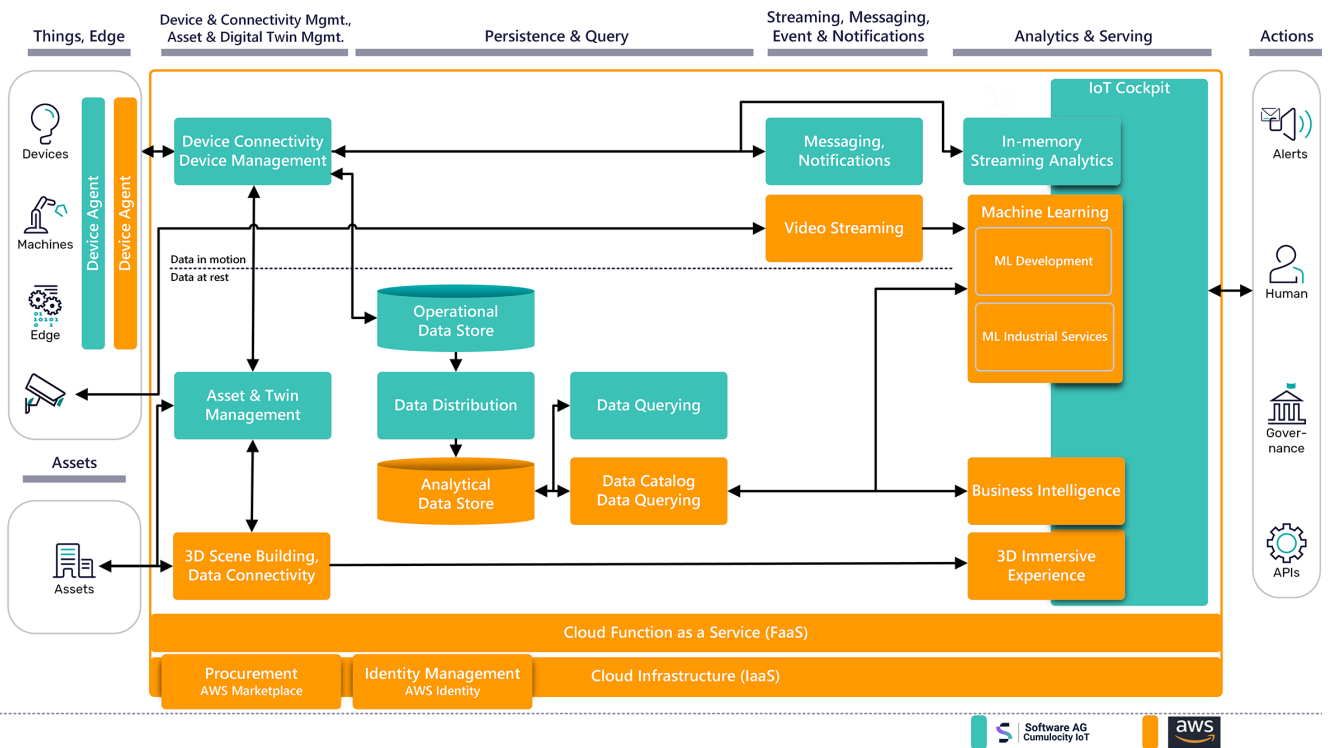

Figure 11. Joint capability architecture showing interaction and data flow between Cumulocity IoT and AWS

A joint product architecture

Figure 12 is a product architecture that displays the key components of Cumulocity IoT and AWS services in scope for this paper. Again, the objective is to indicate the general integration points between Cumulocity IoT components and the AWS services that realize the use cases described. The arrows indicate the main direction of data flow and the purpose or use case of the integration.

Figure 12. Joint product architecture showing integration points and data flow between Cumulocity IoT and AWS

Cumulocity IoT and AWS: A real-world implementation

Flexco is a leading provider of belt conveyor solutions for both light-duty and heavy-duty uses. It has customers in a variety of industries, including agriculture, mining, food processing, and logistics. Belt cleaners are an essential component as they improve belt health and overall system reliability. **Flexco Elevate**® is a remote monitoring solution for belt cleaners and belt health that is built on Cumulocity IoT and AWS. It leverages autonomous "i3" devices connected to the assets to provide sensors and secure communication.

The solution leverages the following capabilities of a Cumulocity IoT/AWS-based solution:

- The solution operates on AWS infrastructure

- The IoT data is stored in an AWS data lake

- Cumulocity IoT Device Management allows users to onboard new devices in less than 5 minutes

- Dashboards are provided for various stakeholders, delivered through a browser and the Flexco Elevate Mobile App

- The dashboards visualize a digital twin of the conveyor cleaner with live IoT data, along with maintenance records and master data, including location.

- Inbuilt analytics uses both Amazon QuickSight and Machine Learning for **predictive maintenance**

- The open solution architecture supports Flexco customers who want to use IoT data in different contexts, for example, for Digital Twins of an entire mine.

By deploying Flexco Elevate, Flexco customers have introduced live remote condition monitoring and predictive maintenance, which have enabled them to cut back on operating costs while boosting production at customer sites around the world. Within a few months of use, Flexco Elevate put one mining operation on a path to saving more than $50,000 dollars in annual labor; another firm reduced machine downtime that led to an increase in annual output of $1.1 million.

Flexco and Flexco Elevate are trademarks of Flexible Steel Lacing Company

---

Conclusion: A joint approach to enterprise grade IoT

As this paper has demonstrated, Cumulocity IoT and AWS in combination form an integrated, best-of-breed architecture that provides customers with world-class IoT capabilities. The solutions support IoT initiatives on all levels of maturity: from prototypes to millions of devices; from using simple IoT sensors to on-device Machine Learning; from the initial rollout to well-oiled DevOps and MLOps environments, which ensure that the cost of Day-2 operations stays in check.

In some areas, the required IoT capabilities are covered by one of the two partners. For example, Cumulocity IoT provides out-of-the-box device management, while AWS provides infrastructure and Machine Learning. In a few areas, both partners contribute, but with different specialties: AWS' capabilities focus on developing differentiating solutions, while those of Cumulocity IoT bring a higher level of standardization and citizen developer friendly features. Finally, AWS adds some capabilities, like Business Intelligence, that are not exclusively IoT, but benefit from interaction and data exchange with it.

Cloud-based IoT solutions are best for the majority of IoT initiatives, as they provide all the advantages of cloud computing, including cost efficiency, scalability, and flexibility. But Software AG and AWS know that IoT is distributed and also lives on the Edge and on the device itself—often in secured environments. For these scenarios, matching products and deployment options are available as well. Most importantly, wherever processing takes place, the software works with open standards. There is no vendor lock-in, as the IoT data is available on your data lake for further use. The device-side software is even open source—providing code-level visibility and the benefits of its developer ecosystem.

Finally, Cumulocity IoT and AWS are market leaders in their respective segments. They have the experience to support the entire enterprise in building a bespoke solution based on a specific architecture.

Ready to embrace the IoT era with confidence? We strongly believe this joint approach is your best bet to achieving IoT that scales—and a business that thrives on connectivity.

Speak with an IoT solution specialist

Contact us

Reading this document on paper?

Scan the QR code to get the digital version and access links referenced throughout.

ABOUT SOFTWARE AG

Software AG helps you create effortlessly connected experiences for your customers, employees and partners with an enterprise-grade iPaaS that integrates anything, anywhere, any way you want. By bringing application, data, API and B2B integration together in the same generative AI-enabled platform, you can run a high-performing enterprise and constantly improve it based on data. Get end-to-end visibility and governance across geographies, IT environments, and complex business ecosystems, with hybrid multi-cloud connectivity, and enterprise-grade security relied on by the most powerful banks, governments, and corporations in the world. Trusted by the world's best brands for more than 50 years, our technology and team of integration enthusiasts will make sure that integration is a driver of innovation for your enterprise. Ready to make integration really work for you? Now you can. Finally.

Learn more at **www.SoftwareAG.com**. Follow us on **LinkedIn** and **X**.