



Doing more with your IoT data

Turning historical data into
valuable business insight

White paper



Doing more with your IoT data

Each connected sensor in your IoT solution can create thousands—if not millions—of data points every hour. A treasure trove of value exists in that raw data! But to leverage that value, your business needs to store data over time and process, organize and analyze your data properly. That's where Cumulocity IoT DataHub comes in. Using Cumulocity IoT DataHub, you can analyze historical IoT data in a cost-efficient data lake of your choice to make business decisions that have impact.



The power to be gained from long-term IoT data

Most IoT projects begin by focusing on the sensors to be used for collecting data. Next comes connecting and managing those sensors, then setting alarms and triggering actions based on incoming sensor data. You might also run ad-hoc analytics for real-time data analysis and visualization. Using “if this, then that” scenarios, you can trigger actions based on outcomes. For example: If the pressure of a paint nozzle on a painting robot goes below X mBar for more than Y milliseconds, then the system triggers the appropriate remedial action.

Analyzing and acting on real-time IoT data is incredibly powerful. You can prevent major issues by quickly detecting faults. You can do this using Software AG's Cumulocity IoT platform.

Now let's go one step further: Deriving real business insights and analyzing the entirety of information from all your different devices, as well as setting up responsive actions and combining that data with other data held in business systems. You can do all this with Cumulocity IoT as well.

By adding Cumulocity IoT DataHub, you can do analysis on longer-term data—beyond the standard six weeks of data stored in Cumulocity IoT. Access to this historical data expands your understanding of the “why” behind events. For companies using TrendMiner, they will be able to access both the historical data stored in Cumulocity IoT DataHub as well as the live data stored in Cumulocity IoT.

Here's what we mean. Let's say a machine's temperature rises above what's considered the norm. Rather than just reacting to live data, it'd be helpful to know:

- Why is the equipment hot?
- What's been the outcome in previous similar circumstances?
- Is this a common occurrence?
- Are there outside influences?
- What might you change in the setup and systems?
- Can you predict and prevent this issue in the future?

The more data you have, the more reliable your answers will be. To find these answers, you'd need to connect your usual business intelligence (BI), machine learning and other business systems to this longer-term IoT data without adversely affecting ongoing operational tasks. Cumulocity IoT DataHub is ideal for this scenario.

Analyzing IoT data from the past enables you to understand trends, patterns, issues and compare with other data and information collected across the business. Cumulocity IoT DataHub extracts, condenses, optimizes and transfers that data in the right format to your choice of data lakes for economical and scalable long-term storage. Then you can run number-crunching analytical tasks in the data lake, where it's the most cost efficient to do so.

Cumulocity IoT DataHub bridges the gap between your IoT data and the larger analytics ecosystem. You gain more value from your IoT data, because you know how to make your business more reliable and better equipped to improve the customer experience. You move from "IoT lite" to IoT exceptional.

Doing more with Cumulocity IoT DataHub

Data storage in Cumulocity IoT is designed for low-latency operational tasks and quick turnaround—instant alarms and instant reactions—for example, sending an alarm if the oil level of a machine is below a threshold. That's on purpose to help businesses improve agility and react quickly to changing real-time data. An operational data store runs different operational tasks with respect to a recent time period.

By comparison, DataHub is optimized specifically to store long-term data for historical analytics. Cumulocity IoT DataHub migrates the data from within your operational store in Cumulocity IoT, organizes it, condenses it and puts it into a data lake where it can be kept efficiently, at less cost, for the longer term and in a way that it can be analyzed or queried by your BI and data science tools.

Most BI and analytical applications use the de-facto industry-standard SQL to query data sources, such as data warehouses and relational databases. Cumulocity IoT DataHub ensures efficient execution of SQL queries of data stored in data lakes. This opens up your analytics for wider analysis on all data, so you can learn from the past and make better decisions for the future.

How businesses use historical IoT data

Keeping customers informed

An engineering company that develops and tests systems like combustion engines, electric drives, batteries and fuel cells wants to provide its customers better insight. Using Cumulocity IoT DataHub, the company combines real-time IoT data with historical IoT data. Cumulocity IoT DataHub takes the data from Cumulocity IoT's operational store, puts it in a data lake and gives the company's data scientists access to the historical data sets through one common layer in Cumulocity IoT DataHub. Based on the analysis, the company recommends how its customers can improve operational performance and save money. There would be no way to access this data other than with Cumulocity IoT DataHub.

For its marine customers, the company can recommend routes that have less turbulence, faster travel and use less fuel. The company feeds live data on its machine into Cumulocity IoT's operational store and combines this data with the historical information gathered through Cumulocity IoT DataHub. Live information is displayed en route.





Automotive production line planning

Another company builds machines to fill the fluids in cars on the production line. The stand-alone **Cumulocity IoT edge** solution runs on an industrial PC in the factory—in the hands of production planning staff with no link to the cloud needed. The company hooks up the machines to all parts of the car that need brake fluid, oil and washer fluid, runs pressure tests, and then uses the data to plan for future car production. Insight into the average time it takes to fill a similar capacity model helps in planning production time and fluid requirements. Having fast access to reliable time and materials estimates eases the planning process.

What you can do better with Cumulocity IoT DataHub

With Cumulocity IoT DataHub, you can build data analytics solutions that help you understand:

- Why things happen
- Whether a situation is a recurring issue
- The previous impact

As a result, you can predict a situation and build a process to handle it—even respond automatically to prevent potential problems.

The **IoT analytics** software provides:

- Incremental offloading of IoT data to low-cost yet high-performance data lakes
- Scalable SQL querying on long-term IoT data in data lakes
- Standard interfaces to BI and data science tools

The benefits are:

Cost efficiency

Cumulocity IoT DataHub transfers your IoT data into a data lake, which is 1,000 to 3,000 times cheaper than storage in an operational store.

Ease of use

No in-depth data engineering skills required. In a few mouse clicks, you can configure a data pipeline that automatically and continuously copies your IoT data into the data lake and runs analytical queries. There are no cumbersome mappings or transformations to define for converting raw data in the operational store into structured data in the data lake.

High performance

Clever storage layout, combined with scalable query capabilities, enable you to query huge data volumes in seconds.

Wider business insights

Cumulocity IoT expands who can see value from IoT data and offers the insights that can be used across the business. Data scientists can perform complex analytical tasks on the data and integrate it with other business data to enhance business decisions.

How Cumulocity IoT DataHub works

Offloading data

Cumulocity IoT DataHub moves your IoT data from an operational store in Cumulocity IoT to a data lake to turn raw IoT data into a structured, condensed format needed for efficient SQL querying, the basis of reporting and intelligence. This “offloading” process allows you to build a low-cost and long-term archive of device data.

The operational store of Cumulocity IoT oversees the document-based storage of events, alarms, inventory and measurements data. To keep the store performative, data is stored for a limited time in Cumulocity IoT. Cumulocity IoT DataHub migrates all data into a data lake of your choice, such as Amazon® S3 (and API-compatible products such as Min.io®) or Microsoft® Azure® Data Lake Storage Gen2 (Azure Storage). Those data lakes are purpose-built for storing huge amounts of data in a scalable and inexpensive manner.

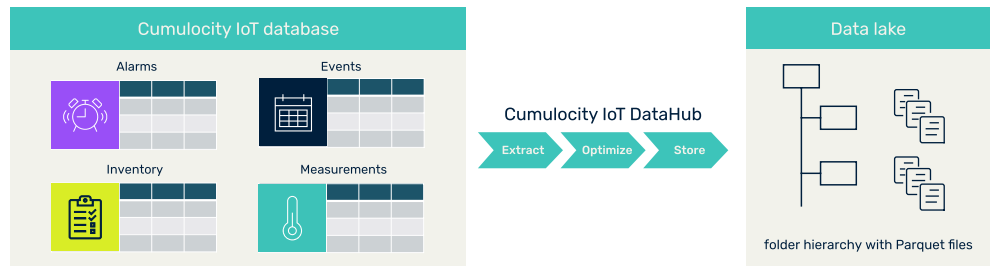
Getting device data into the data lake requires you to define an offloading pipeline. Such a pipeline defines which data from the operational store you want to offload and where the results are to be stored in the data lake. Once configured, the pipeline runs at a fixed cadence and offloads not-yet offloaded data into the data lake (i.e., only delta changes need to be transferred over the network).

With Cumulocity IoT DataHub, you needn't deal with cumbersome definitions of how the original document-based data is to be transformed. Using detailed knowledge of Cumulocity IoT's internal data model, Cumulocity IoT DataHub auto-transforms the data into a tabular format, which is optimal for machine learning, BI and most custom application use cases.

Provided you want to pre-process or clean the data before it is offloaded, you can define additional result columns or add filter predicates for a finer granular selection of the data to be offloaded. Once you have defined the pipeline, you simply need to activate it. The initial run of a pipeline will transfer all data from the operational store into the data lake. From then on, Cumulocity IoT DataHub will periodically check for new data not yet offloaded and trigger the pipeline so that these increments are copied into the data lake. Cumulocity IoT DataHub takes care that neither duplication occurs, nor data gets lost.

What does the data in the data lake look like? The transformed, tabular data is stored in the Apache Parquet™ format, which gives you an analysis-friendly and storage-efficient columnar representation of your data. Apache Parquet is the de-facto standard data format in “big data” tools, giving you the freedom to process the data with tools such as Apache Spark™ in addition to using Cumulocity IoT DataHub. Taking common analysis patterns into account, Cumulocity IoT DataHub arranges the Parquet files in a temporal folder hierarchy. Additional housekeeping mechanisms in the background regularly compact smaller Parquet files, which boosts overall query performance (think of defragmenting a hard disk back in the days).





How Cumulocity IoT DataHub works

Once offloading is complete, you can analyze and gain insights from your data at interactive speed, using your favorite BI and data science tools. You can then gain business insights by extracting exactly what you need and integrate the information with that from other business systems.

Combining insight from business systems with IoT data

Cumulocity IoT DataHub enables you to connect your BI querying and reporting tools to your IoT data, so you can extract all sorts of powerful business insights from the data. It offers SQL as the query interface, which is the lingua franca of data processing and analytics. Dremio™ is the internal engine which executes the SQL queries. Due to its highly scalable nature, Dremio can easily cope with many analytical queries.

With Cumulocity IoT DataHub, you can quickly connect the tool or application of your choice, including:

- BI tools using JDBC or ODBC
- Data science applications using Python® scripts, which connect via ODBC
- Custom applications using JDBC for the Java® ecosystem, ODBC for .NET, Python, etc. and REST for (Cumulocity IoT) web applications

Training machine learning models

Nowadays, machine learning is a popular choice for gaining deeper knowledge into business and production processes. The more data you have, the more reliable the insight from your machine learning models will be. Cumulocity IoT DataHub prepares the ground for training complex machine learning models by making the entirety of your IoT data available in a well-structured and analysis-friendly format. Simply connect your favorite data science tool through ODBC, JDBC or REST, and start processing your data.

You can, for example, train a model on the failure states of a valve in order to learn which factors indicate that the valve will soon fail. Then use these insights, combined with your current live data in Cumulocity IoT, to proactively change a valve before it breaks. This is the power of combining live data with historical data.

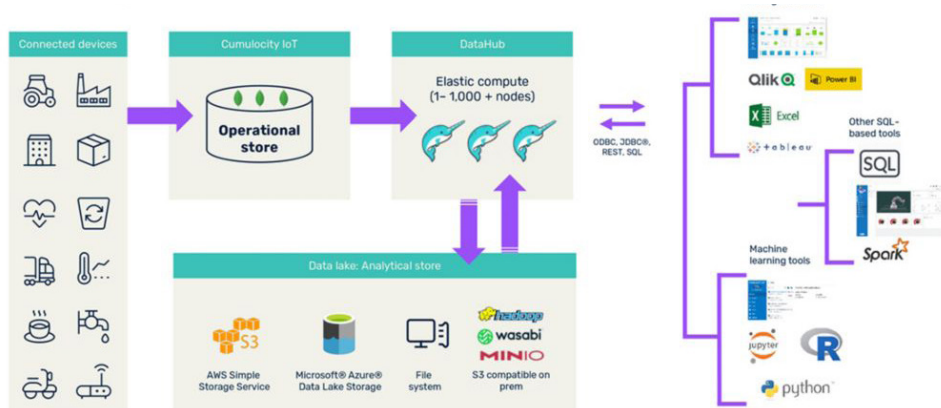
Live data helps you respond immediately to current circumstances. Historical data helps you gain insight from long-term data to see trends and patterns and make decisions accordingly. By using historical data to train machine learning models, then combining that with the live data, you can use knowledge of the past to know what's best for today and the future.

Cumulocity IoT DataHub architecture

Cumulocity IoT DataHub is designed to:

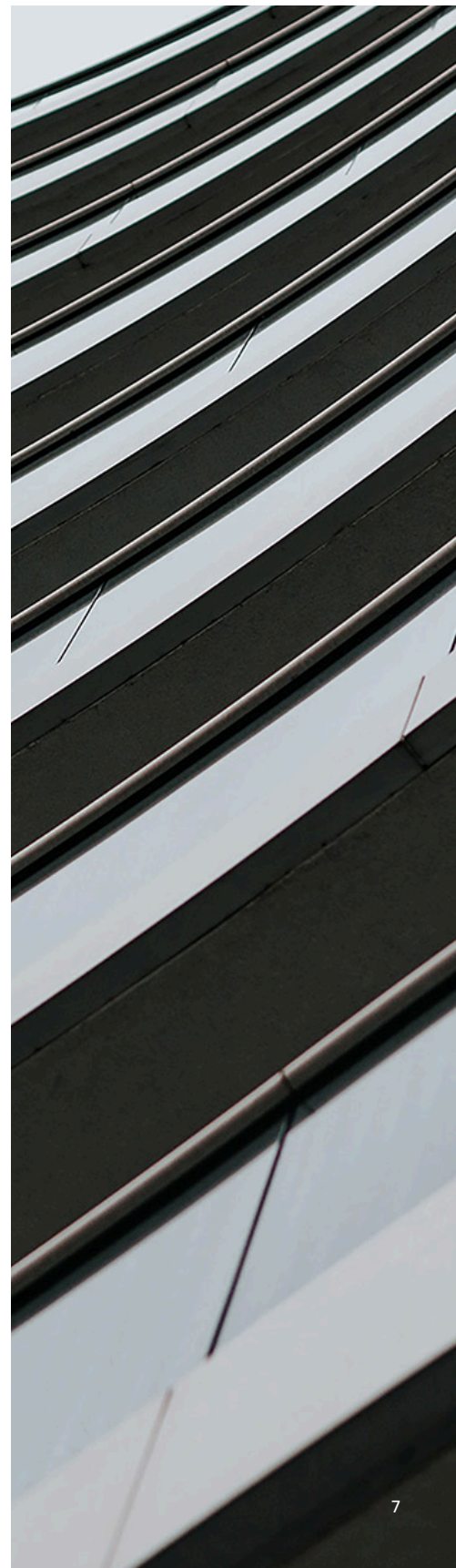
- Automatically move data from the operational data store into a data lake
- Store flattened data in an analysis-friendly layout
- Execute complex analytical queries at high speed
- Easily scale with the amount of IoT data being processed—one pillar of that architecture is the separation of storage and compute capabilities

Cloud data lakes allow you to easily scale your data storage needs with the onslaught of data emitted by your IoT sensors. Cumulocity IoT DataHub ensures the data is well structured within a temporal hierarchy and complemented by internal housekeeping mechanisms ensuring compact file representations. Cumulocity IoT DataHub uses Dremio to move data into the data lake. Dremio is also in charge of executing queries on that data lake. Dremio delivers leading-edge query performance by using innovative technologies like Apache Arrow™, reflections and Columnar Cloud Cache. Scaling Dremio nodes allows you to process ever-increasing amounts of IoT data in seconds.



Cumulocity IoT DataHub takes data periodically from the operational data store, either on-premises or at the edge, and transforms it into a compact format that's highly efficient for analytical queries and places it in an analytical store in the data lake. Analyze that data using your favorite BI tool.

Cumulocity IoT DataHub is designed as a cloud-native application, with all its components running as microservices/containers in Kubernetes® clusters in private or public clouds. Need local processing? On a shop floor, for example, IoT devices are often connected to local computers instead of remote cloud platforms and do local processing instead of moving all data to the cloud. Cumulocity IoT DataHub serves those use cases by providing an edge edition. As a storage layer, Cumulocity IoT DataHub Edge uses the local storage of the edge device. Other than that, Cumulocity IoT DataHub Edge offers the same capabilities as the cloud edition, excluding horizontal scalability.





Take the next step

To get the most value from IoT, your organization needs to move beyond just connecting and managing devices (what some call “IoT lite”) to doing more with your IoT data. That means integrating your IoT data with enterprise systems and analyzing that data to glean strategic insights that shape profitable business decisions. Instant actions based on live data is powerful. Even more powerful is analyzing historical data on its own and combined with live data. Cumulocity IoT DataHub enables you to extract all that data, organize and structure it and store it in the right format so it can be queried by reporting and BI tools and analyzed for business insight and decision making.

Use Cumulocity IoT DataHub to do more with your IoT data, from seeing trends in your equipment to delivering value-added insights to serve your customers. Learn from the past to know what’s best for today and the future.



Take the next step

To find out more, talk to your Software AG representative and visit www.softwareag.com/iot.

ABOUT SOFTWARE AG

Software AG began its journey in 1969, the year that technology helped put a man on the moon and the software industry was born. Today our infrastructure software makes a world of living connections possible. Every day, millions of lives around the world are connected by our technologies. A fluid flow of data fuels hybrid integration and the Industrial Internet of Things. By connecting applications on the ground and in cloud, businesses, governments and humanity can instantly see opportunities, make decisions and act immediately. Software AG connects the world to keep it living and thriving. For more information, visit www.softwareag.com.

© 2021 Software AG. All rights reserved. Software AG and all Software AG products are either trademarks or registered trademarks of Software AG. Other product and company names mentioned herein may be the trademarks of their respective owners.

wp_cumulocity-iot_datahub_en