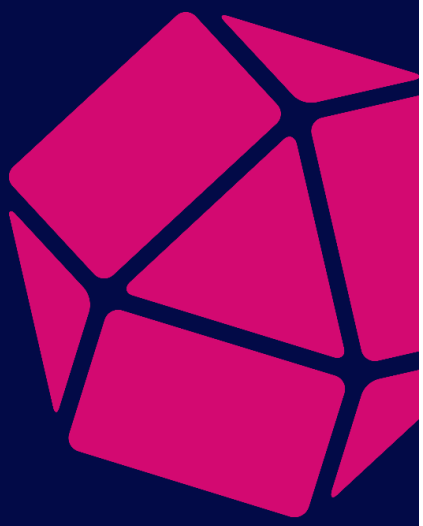
influxdb™

AN INFLUXDATA TECHNICAL PAPER

Architecting for IoT: The Need for an IoT Data Platform



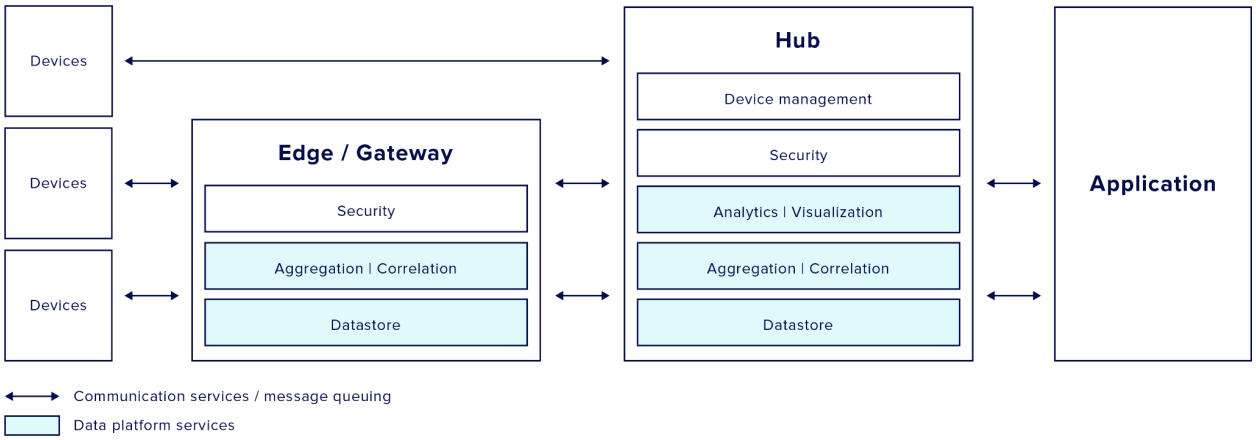
DECEMBER 2023 | REV 1

Architecting for IoT - choosing the right IoT data platform

We are witnessing the instrumentation of every available surface in the material world—streets, cars, factories, power grids, ice caps, satellites, clothing, phones, microwaves, milk containers, planets, and even human bodies. These devices generate more data than we have ever experienced. This data is streaming in real-time and will force companies to determine which IoT platform architecture will be resilient, scalable, and extensible enough to handle these new workloads.

IoT projects gather data from sensors or devices to gain real-time insights, accelerate decision-making, perform automated tasks, and create value by enabling organizations to become data-driven. So, the ability to ingest large amounts of data without interruption is a key differentiator in IoT data platforms. Pushing data services further to the edge is the current direction of IoT data platforms—it will continue to be a defense against rising network costs and as edge environments become more robust.

Distributed IoT platform



Although the list below is not exhaustive, we see three key architectural components in a typical IoT architecture:

Devices: Sometimes called “Things” or “endpoints,” these can be software or hardware devices that generate data from the object they are monitoring and/or controlling. Some examples are temperature sensors, voltage sensors, humidity sensors, and machine rotation or vibration sensors. Sensors can be co-located, as in the case of a wind turbine or car, while others form part of another device (like the Smart Watch with biometric data). Some even have control features like industrial valves.

Edge/gateways: Gateways are usually deployed closer to the edge (some might be part of a sensor collection and deployed with the sensor) and allow for the collection of data, the selective transfer of data to the Hub, and provide some command-and-control functionality back to an automated device. This area is further expanding into what is called edge computing. With edge computing, users can make critical decisions at (or near) the data source instead of sending it back to the Hub.

Hub: The Hub is conceptually the IoT solution's central processing area. Although hubs deploy in a distributed fashion, conceptually, they process data streams from the gateways (or directly from the sensors), to store, monitor, and analyze data, and provide the basis for new applications to enable competitive differentiation.

To support this architecture, an IoT platform needs to provide a variety of services:

Communication services provide device connectivity, message/event queuing, transportation services across Wi-Fi, cellular or other wireless protocols, as well as fixed connections.

Security services provide verification, encryption, and authentication services to guarantee that devices (and software on those devices) are secure and tamper-proof. Additional security that protects communication services between devices is typically necessary as well.

Device management services support device provisioning and lifecycle management. This includes operations like firmware updates, diagnostics, and monitoring.

Data services support the collection, aggregation, storage, visualization, and analysis of sensor data. The next section of this paper explores this set of services in more depth.

IoT data characteristics

Data from devices and sensors, collectively known as IoT data, has unique characteristics that data services must address:

IoT data is time series data: IoT is synonymous with time-stamped data (or time series data) because the purpose of any sensor is to measure change over time. Systems evaluate and analyze data based on timeframes and ranges, allowing them to act on specific insights. Time is not an afterthought for IoT data—it is a constituent of data. Therefore, a data platform for IoT needs to be optimized for time series data.
IoT data is streaming data: IoT is synonymous with streaming data. Myriad sensors create data and each sensor emits a continual stream. A data platform for IoT needs to be optimized for streaming data and have

an abundance of actionable analytics to find the signal amid the noise.

IoT data is real-time: Sensors actively capture and emit data in real-time. A data platform for IoT that can handle real-time ingestion and real-time streaming analytics equals higher business value. The system can then ingest, analyze, interpret, and act upon data in real-time.

IoT data platform services

Modern IoT workloads and IoT data characteristics mean more data points, data sources, monitoring, and controls. They demand a shift in how we approach building systems that support this type of data. A modern IoT data platform provides a comprehensive set of tools and services optimized for IoT data, including:

Data aggregation services are lightweight and deployable in a distributed manner to collect, normalize, correlate, and aggregate metrics and events from sensors and other data sources across a range of IoT data protocols. Because of the wide array of hardware and communication protocols used in IoT applications, a tool that simplifies integrating and collecting data is essential to success.

Data storage services must support high write volumes as well as the efficient storage and real-time retrieval of large sets of time series data. In addition, the database should provide time series functions that support queries and analytics without needing developers to reinvent the wheel.

Streaming analytics and visualization services provide visualization and dashboarding services, real-time pattern detection services, and notification, control, and action services to automate the entire system.

InfluxData - the modern IoT data platform

InfluxData delivers a modern open source IoT data platform built from the ground up to support organizations building solutions to take advantage of IoT data. Specifically, InfluxData provides the following services:

Data storage services: At the heart of the InfluxData offering is InfluxDB, an open source time series database that supports high write loads, storage for large data sets, and conserves space through high-efficiency compression, downsampling, retention policies for automatically expiring and deleting unwanted data, as well as backup and restore features. InfluxDB also makes it easy to analyze data by providing easy-to-use, powerful data query languages with built-in time-series-specific functions via SQL and InfluxQL.

Data Aggregation Services: InfluxData provides a comprehensive set of tools and services to get metrics and event data from sensors, devices, systems, and machines. InfluxData's collection services are built from the open source Telegraf project. With over 300 plugins and counting, Telegraf enables the collection of IoT data across multiple protocols popular with IoT devices.

Telegraf deploys as a lightweight binary directly on devices or as part of an edge deployment. Telegraf can filter and process data to reduce bandwidth consumption or clean data before sending it to storage.

Streaming Analytics Services: InfluxDB integrates easily with your streaming analytics tool of choice. You can query data stored in InfluxDB directly via API or use client libraries for the most popular programming languages.

Visualization Services: InfluxDB has direct integrations for many visualization tools, making it simple to connect and visualize your data using best-in-class solutions. These include Grafana, Apache Superset, and Tableau.

In short, InfluxData provides a comprehensive set of open source IoT data services you can deploy from the hub to the gateway and the edge.

When you take a step back and consider all the different devices sending data to an IoT platform, you realize that it's all about data. IoT devices generate large amounts of data with nanosecond precision from diverse locations, with the consequent necessity for quick data aggregation and an increased need to index, store, process, and react effectively.

In the next section, we provide additional insight and some lessons from our clients who have adopted an IoT data platform.

Learning from our clients: choosing the right IoT data platform

At InfluxData, we strive to provide our customers with a clear roadmap for selecting the right IoT data platform. By avoiding the potential risks and costs associated with unsuitable databases, we can help them to make the most of their journey and learn from the experiences of others.

Lesson 1: Don't start with the wrong data platform architecture

The following list includes some common database technologies that clients try to use with IoT data:

General-purpose relational databases: MySQL is a popular open source relational database management system. But database technologies are written to solve a particular use case. Relational technologies such as MySQL and Postgres keep references to data (e.g., Customers have Orders, Orders have Order-lines) but struggle at scale with IoT data—streaming, time-stamped, and real-time. Relational databases provide poor compression, time-performance, query, and scalability for IoT data.

NoSQL databases: HBase and Cassandra are columnar or key-value databases designed for arbitrary amounts of "big data" but not the characteristics of IoT data. For instance, neither database supports time as a key constituent. They don't compress time-stamped data for better resource utilization and can't handle the millions of writes per second that streaming IoT data requires. They have no built-in functions to perform time-dependent queries and cannot automatically downsample the time precision of old data to free up resources. All these capabilities could be added with some engineering, but why waste precious resources when you can use a platform that already has them built-in?

Elasticsearch: Elastisearch is an open source, distributed, full-text search engine for enterprise workloads with a datastore suitable for time series data. However, this flexibility comes at a price: to correctly utilize the primitives Elasticsearch provides, you must model any particular use case. Moreover, optimal usage requires knowledge of how the internal mechanisms work and has a much steeper learning curve. InfluxDB is purpose-built for time series and is, therefore, easy to set up and use for an IoT use case. Elasticsearch typically sees worse results for IoT data in performance factors like write throughput and data compression. Adopt a Time Series Database: Trying any of these databases reveals their limitations when handling time series data. A time series database (TSDB) is ideal because IoT data is time series data. A modern TSDB is built specifically for handling time-stamped metrics, events, and measurements and optimized for measuring change over time. Properties that make time series data different from other data workloads include data lifecycle management, summarization, and large-range scans of many records.

Lesson 2: Don't settle for first-generation time series databases

Time series data comes in multiple forms: metrics, events, traces, and logs. Legacy time series databases such as Graphite, RRD, or OpenTSDB only support regular time series metrics. In contrast, modern time series databases, like InfluxDB, can handle the full spectrum of time series data.

Customer case studies

Here are some sample case studies from customers with successful InfluxDB implementations:

1. BBOXX – This solar energy product provider transitioned from its outsourced monitoring solution to real-time data with InfluxDB Cloud to continuously monitor their geographically dispersed

85,000 solar rooftop units.

2. Loft Orbital – Loft Orbital is the leading provider of space infrastructure as a service. Companies turn to Loft Orbital when they want to focus on their area of expertise and trust Loft Orbital to operate their satellites using its mission-agnostic, flexible operating system, and interfacing technology. This system is monitored using InfluxDB and uses Telegraf for data collection.
3. Spiio – This IoT company focused on plant analytics started with MySQL as a proof of concept, considered generic/branded IoT platforms, proceeded to try other time series databases, and then adopted InfluxDB.
4. tado° – This IoT-based home climate control company started with MySQL as a proof of concept, progressed to next-gen MySQL, then settled on InfluxData.
5. Teréga – Teréga is a gas storage and transportation company in France that manages 5,000 kilometers of pipelines. Teréga replaced their legacy data historian with InfluxDB.

About InfluxData

InfluxData is the creator of InfluxDB, the leading time series platform used to collect, store, and analyze all time series data at any scale. Developers can query and analyze their time-stamped data in real-time to discover, interpret, and share new insights to gain a competitive edge. InfluxData is a remote-first company with a globally distributed workforce. For more information, visit www.influxdata.com.

InfluxDB documentation, downloads & guide

Get InfluxDB
Try InfluxDB Cloud for Free
Get documentation
Additional tech papers
Join the InfluxDB community