



AN INFLUXDATA TECHNICAL PAPER

Why Time Series Matters For Metrics, Real-Time Analytics And Sensor Data

Paul Dix
CTO and Founder,
InfluxData



OCTOBER 2023 (REVISION 6)

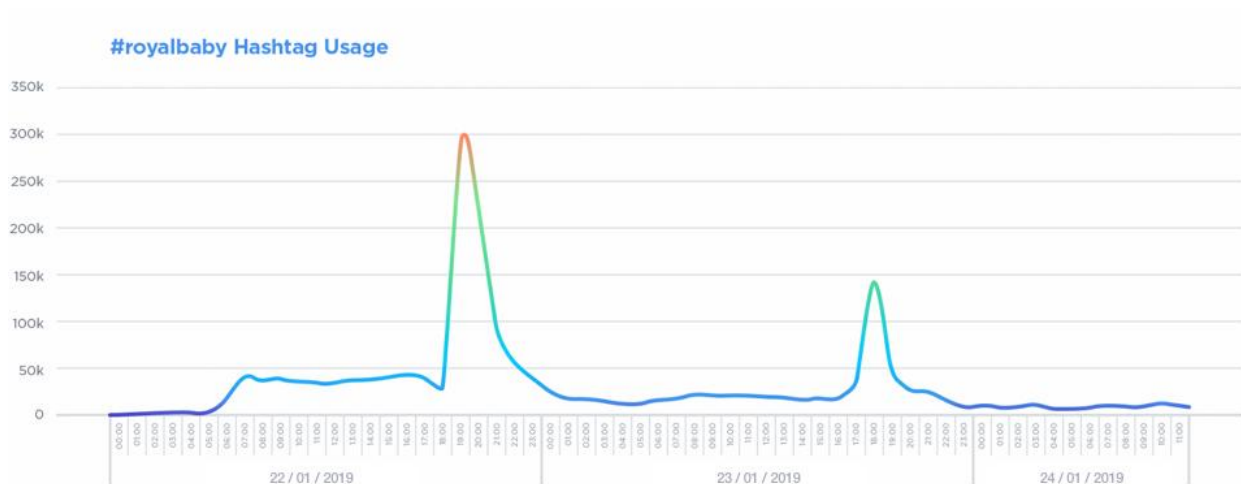
Introduction

Time series data has historically been associated with applications in finance. However, as developers and businesses move to instrument more of their servers, applications, network infrastructure and the physical world, time series is becoming the de facto standard for how to think about storing, retrieving and mining this data for real-time and historical insight. This paper will:

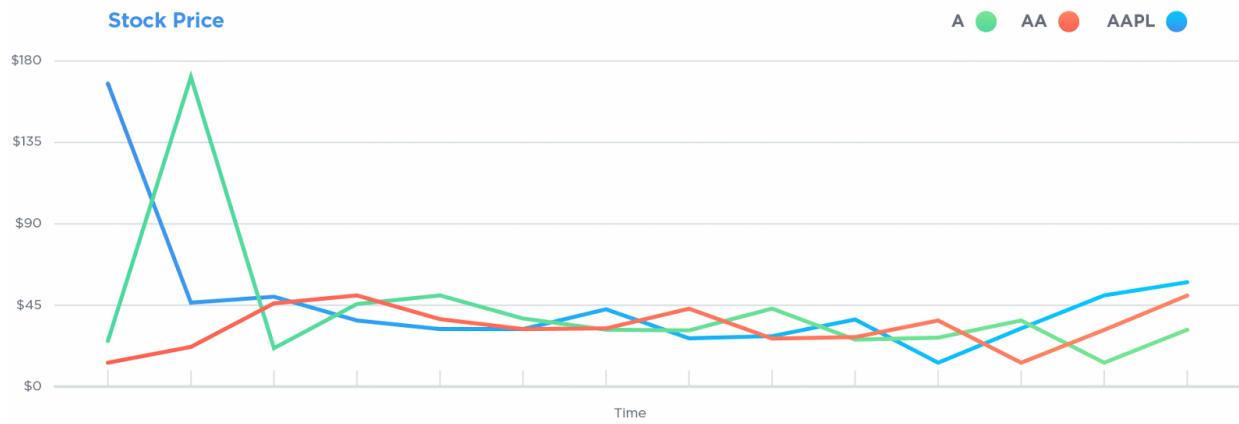
- Define [what time series data is](#) (and what it isn't)
- Explain how the time series data domain differs from more traditional data workloads like OLTP or full-text search
- Examine what makes the InfluxData platform different from other proposed solutions

What is time series data?

Time series are simply measurements or events that are tracked, monitored, downsampled and aggregated over time. This could be server metrics, application performance monitoring, network data, sensor data, events, clicks, trades in a market and many other types of analytical data. The key difference that separates time series data from regular data is that you're always asking questions about it over time. A simple way to determine if the dataset you are working with is time series or not, is to look and see if one of your axes is time. Below are a few examples of time series data plotted on graphs:

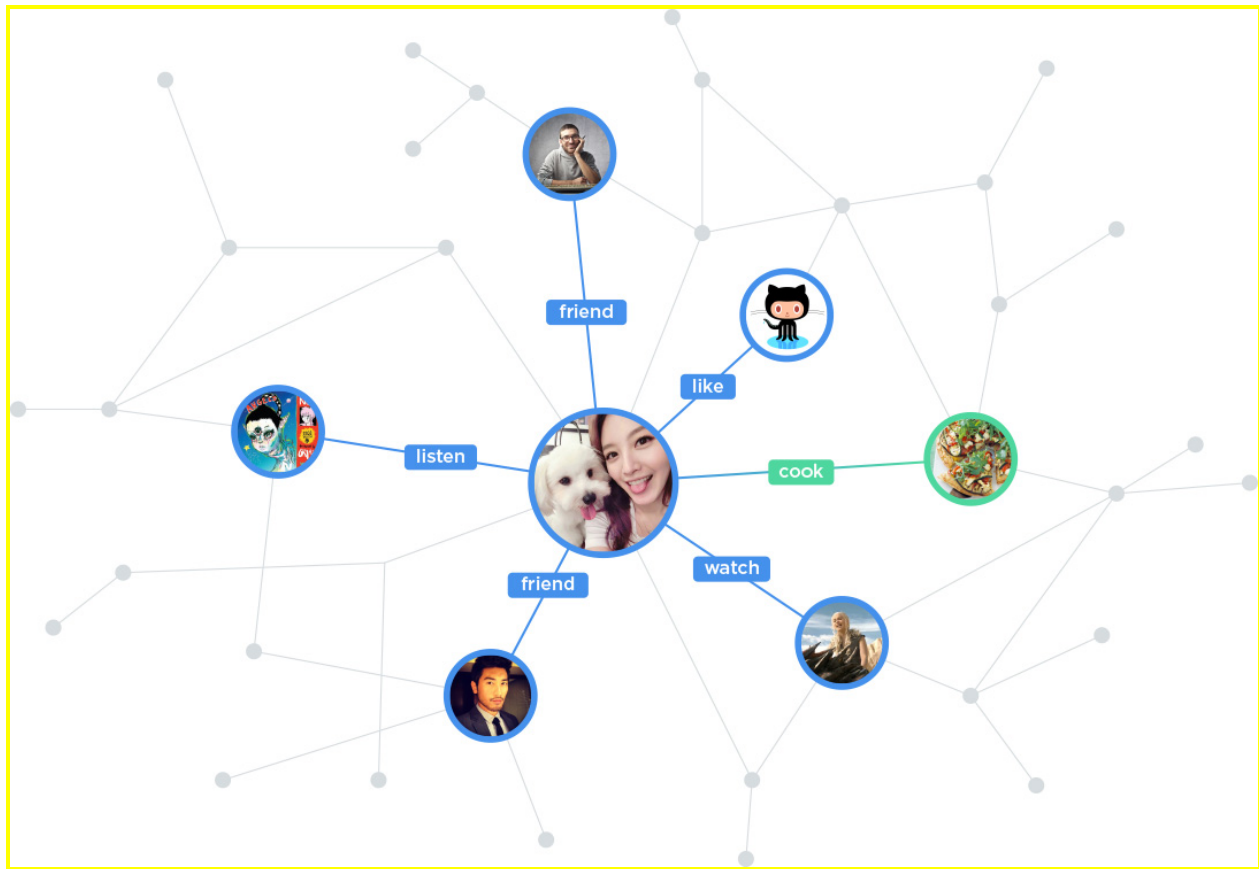


Time Series Example: Hashtag Frequency Over Time

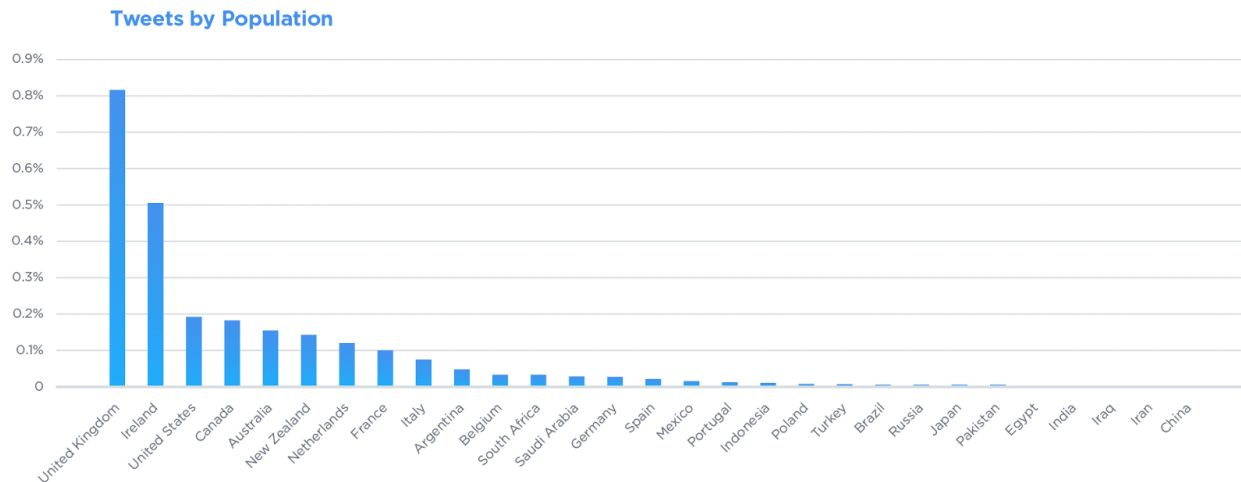


Time series example: stock ticker prices over time

Now, a few examples that are NOT time series data, plotted on graphs:



NOT a time series example: graph of relationships



NOT a time series example: tweets by population

Time series data comes in two forms: regular and irregular. Regular time series consist of measurements gathered from software or hardware sensors at regular intervals of time (every 10 seconds, for example) and are often referred to as metrics. Irregular time series are events driven either by users or other external events. Summarizations of irregular time series become regular themselves. For example, summarizing the average response time for requests in an application over one minute intervals or showing the average trade price of Apple stock every 10 minutes over the course of a day.

InfluxDB is optimized for both use cases. Many services and time series databases support only the regular time series metrics use case. InfluxDB lets users collect from multiple and diverse sources, store, query, process and visualize raw high-precision data in addition to the aggregated and downsampled data. This makes InfluxDB a viable choice for applications that require storing raw event data.

The InfluxDB platform organizes time series in a structured format. At the top level is a measurement name, followed by a set of key/value pairs called tags that describe the metadata, followed by key/value pairs of the actual values called fields. Field values in InfluxDB can be boolean, int64, float64 or strings. Finally, there is a timestamp for the set of values. All data is queried by the measurement, tags, and field along with the time range.

This structure makes it easy for developers to build tools around the APIs that InfluxData provides. Unlike relational or document databases, InfluxDB organizes time series data into a set structure. This structure is also what sets InfluxDB apart from other solutions. The richness of the time series data structures that can

be represented open InfluxDB up to more time series and metrics use cases, while also widening the performance lead that InfluxDB has over other solutions. With the right schema and setup, a single InfluxDB server can handle over [4,000,000 writes per second](#), something the competition is unable to match.

Time series data use cases

In recent years, time series has become a common use case across many industries and a database category of its own. Metrics, events, and other time-based data are being generated at an exponential rate, as there is a growing requirement for analyzing today's complex environments. The InfluxData platform provides a comprehensive set of tools and services to collect and accumulate metrics and events data, analyze the data, and act on the data via powerful visualizations and integrations with specialized tools..

Whether the data comes from humans, [sensors](#) , or [devices](#), InfluxData empowers developers to build next-generation monitoring, analytics, and IoT applications faster, easier, and to scale delivering real business value quickly.

InfluxData has customers and users that span three primary use cases: [DevOps monitoring](#), [real-time analytics](#), and [IoT monitoring](#). Anyone who has sensors, servers, VMs, containers, applications, users, or events to track could benefit from using InfluxData.

DevOps monitoring

For our users in the DevOps monitoring and metrics space, most are in medium to large organizations. Some of these users are building custom monitoring solutions from scratch, deploying the InfluxDB platform to first track their servers, VMs, containers, data stores, and network hardware and later as a generalized metrics platform for application developers within the organization.

Others are using InfluxData to supplement commercial APM products to instrument aspects of the InfluxDB platform for which no probes or agents exist, or to stitch together metrics from multiple monitoring solutions. And in both cases, users are not only gathering metrics to determine a baseline about the health of their systems, they are also using InfluxDB as the place to accumulate their log events. Having access to log data is secondary and an important contextual source to help further triage and resolve issues.

Real-time analytics

We see organizations of all sizes working with real-time analytics. Some are building applications that will face their users with InfluxDB as the underlying database, others are instrumenting business, social, or development metrics in real-time for internal consumption. We often see users start with the DevOps

custom monitoring and metrics use case, who then move into real-time analytics once the platform is deployed. InfluxDB eventually becomes the central store for all time series, sensor, observability, and analytics data.

Internet of Things (IoT)

We have also found that there are a number of IoT use cases. We've seen users in industrial settings like factories, oil and gas as well as renewable energy plants, agriculture, smart homes, roads, and infrastructure. There are also users in consumer-grade IoT-like wearables, consumer devices, and trackers.

Solutions built on the InfluxData platform

For [organizations](#) around the world, in nearly every industry, InfluxDB has become the system of insight for unified metrics and events – enabling the most demanding SLAs and providing a foundation for solutions such as [Application Performance Monitoring \(APM\)](#), [Industrial IoT](#), [Kubernetes Monitoring](#), [Network Performance Monitoring](#), and [Stream Processing](#). Here's a brief overview of each.

InfluxDB for application performance monitoring

In a digital economy where complexity is a given and a responsive application is a requirement, visibility into your entire application has become a necessity for enterprises. APM can be performed using InfluxData's platform [InfluxDB](#). InfluxDB is a purpose-built time series database. It is a central platform where all metrics, events, logs, and tracing data can be integrated and centrally monitored.

InfluxDB for Industrial IoT monitoring

The industrial world has a long history of modernizing processes in order to keep production running efficiently and safely while minimizing downtime. Yet many are locked in legacy [data historian](#) solutions that are costly and lack the methods needed to provide innovation and interoperability. In contrast, [InfluxDB](#) provides diverse system architecture options not available from a single software vendor. It provides the freedom to integrate with other solutions and allows you to adapt the code to fit your ever-changing system requirements. This is why many industrial enterprises around the world are choosing InfluxDB for [IIoT monitoring](#).

InfluxDB for Kubernetes monitoring

Kubernetes orchestration provides built-in fault tolerance, automating scaling and maintenance for a desired cluster state. However, visibility must come with the necessary granularity and information for fast identification of the source of trouble. Monitoring and accountability are what makes automation reliable. InfluxDB helps to identify and resolve problems before they affect critical processes, and most importantly,

offers ways to implement [Kubernetes monitoring](#) that accommodate developers' need for instrumentation without overloading IT operations.

InfluxDB for network performance monitoring

When network speed degrades or connectivity fails, the data flow sustaining applications and IT operations will struggle or halt along with the network. Networks — the lifeline of IT infrastructure — are dynamic environments. They require monitoring to deliver consistent, predictable network performance. Dealing with so much monitoring data, it can be easy to be consumed by it. However, there is a way to effectively manage your IT infrastructure, by centralizing, analyzing, and automating it. InfluxData enables you to do this with its network monitoring tools. Its collection agent Telegraf, with 300+ plugins, supports protocols such as ICMP/Ping, SNMP, NetFlow, SFlow, and Syslog. InfluxDB, for its part, contains a powerful query engine for processing multiple data sources in real-time. To gain the necessary visibility in the status, performance, and responsiveness of all network devices in their enterprise, cloud, or hybrid application environments, enterprises are deploying the InfluxData platform for network performance monitoring.

InfluxDB for stream processing

[Stream processing](#) is the processing of data in motion. It unifies applications and analytics by processing data as it arrives, in real-time, and detects conditions within a short period of time from when data is received. The key strength of stream processing is that it can provide insights faster, often within milliseconds to seconds. Stream processing naturally fits with time series data, as most continuous data series are time series data. And time series data needs a purpose-built database to ingest, store, and process it. This is exactly what InfluxDB is. And this is why, given its high-write throughput and the scalability it allows, InfluxDB suits stream processing.

The time series workload

Time series data has a few properties that make it very different from other data workloads. Data lifecycle management, summarization, and large range scans of many records are what separate time series from other database use cases.

With time series, it's common to request a summary from a larger period of time. This requires going over a range of data points to perform some computation like a percentile to get a summary of the underlying series to the user. This kind of workload is very difficult to optimize for a distributed key-value store or relational database. InfluxDB is optimized for exactly this use case, giving millisecond-level query times over months of data.

With time series, it's common to keep high-precision data around for a short period of time. This data is

aggregated and downsampled into longer-term trend data. This means that for every data point that goes into the database, it will have to be deleted after its period of time is up.

This kind of data lifecycle management is difficult for application developers to implement on top of the typical database. They must devise schemes for cheaply evicting large sets of data and constantly downsampling that data at scale.

InfluxDB is designed as a time series database with solutions built-in for summarization and data lifecycle management at a large scale. These come out of the box with no application-level code required from the developer. To learn more, see the [retention policies documentation](#).

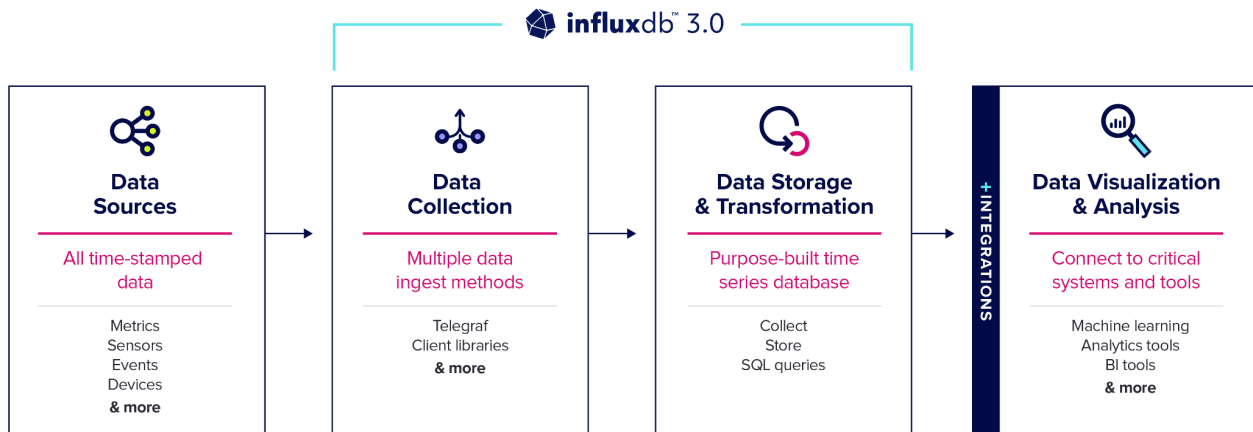
What makes InfluxData different?

We often get asked questions about what makes InfluxData different from other technology solutions. Generally, these can be organized into three categories: applications, databases, and stream processing systems. In the next few sections we'll look at various solutions within these categories. So, let's start with a look at the InfluxDB platform and InfluxDB's data model, which is a key differentiator.

The InfluxData Ecosystem

The InfluxData open source time series platform consists of two main components: InfluxDB and Telegraf. These tools work together to make it easy to collect, store, and analyze your time series data. The most important part of the platform is that InfluxDB makes it easy to integrate with external tools and services for things like data visualization, forecasting, automation, and anything else you'd want to do with your time series data. Let's take a look at what Telegraf and InfluxDB can do and some of the tools they integrate with.

InfluxDB Platform



Telegraf

Telegraf is the data collection component in the InfluxData ecosystem, designed to gather metrics from a variety of sources and write them into InfluxDB. It's a plugin-driven server agent that can collect data from a wide array of sources and communication protocols.

One of the key advantages of Telegraf is its extensibility; its plugin system supports inputs, outputs, processors, and aggregators, allowing you to tailor Telegraf to meet the specific needs of your infrastructure with over 300 plugins currently available. This seamless integration between Telegraf and InfluxDB ensures that you can collect just about any metric you need, transform it as necessary, and store it efficiently and effectively in InfluxDB. Telegraf can also be used to tie together different storage systems for situations where you have different types of data that aren't a good fit for a time series database.

InfluxDB

InfluxDB serves as the core database responsible for storing and analyzing time series data in the influxData ecosystem. It is designed to handle high-velocity and high-volume data streams with low latency, making it especially well-suited for applications that require real-time analytics. InfluxDB is built on open standards like Parquet for data storage, Apache Arrow for in-memory data representation, and FlightSQL for data transfer.

Because of these [design decisions](#), one of the key strengths of InfluxDB is its ability to integrate with best-in-class tools and services. InfluxDB supports both SQL and InfluxQL query languages, making it easier for users to get started. InfluxDB also offers native HTTP APIs that allow for seamless integration with various external tools for data visualization, monitoring, alerting, and machine learning, among other tasks. For example, it can integrate with [Grafana](#) for advanced data visualization, [Pandas](#) for data analysis,

and any other tools that support SQL or standards like [JDBC](#). This makes InfluxDB not just a database but a vital part of a larger, flexible ecosystem aimed at solving complex problems related to time series data.

The InfluxDB data model

The InfluxDB data model is quite different from other time series solutions like Graphite, Prometheus, or OpenTSDB. InfluxDB has a line protocol for sending time series data which takes the following form:

```
<measurement>,<tag set> <field set> <timestamp>
```

The measurement name is a string, the tag set is a collection of key/value pairs where all values are strings, and the field set is a collection of key/value pairs where the values can be int64, float64, bool, or string. The measurement name and tag sets are kept in an inverted index, which make lookups for specific series very fast.

For example, if we have CPU metrics:

```
cpu,host=serverA,region=uswest idle=23,user=42,system=12 1549063516
```

Timestamps in InfluxDB can be by second, millisecond, microsecond, or nanosecond precision. The micro and nanosecond scales make InfluxDB a good choice for use cases in finance and scientific computing where other solutions would be excluded. Compression is variable depending on the level of precision the user needs.

On disk, the data is organized in a columnar format where contiguous blocks of time are set for the measurement, tagset, fieldset. There is no limit to the number of tags and fields that can be used. Other time series solutions don't support multiple fields, which can make their network protocols bloated when transmitting data with shared tag sets. Most other time series solutions only support float64 values, which means the user is unable to encode additional metadata along with the time series data.

InfluxDB vs. APM and logging platforms

We often get asked how InfluxData is different from applications like Datadog, SumoLogic, Splunk, New Relic, and other metrics and monitoring systems. First and foremost, InfluxData is purpose-built as a time series data platform. It exists so that developers can build their applications on top of the platform.

The above-mentioned applications have their user interfaces and business logic built-in. They're meant to

be off-the-shelf solutions that give developers whatever they need for the given problem. Most developers wouldn't build their custom applications on top of these bundled solutions.

InfluxData is a platform for developers to build upon. It's meant to be customized for the unique business logic of each organization it's deployed by. This makes it an ideal choice for larger organizations that are looking to develop solutions designed specifically for their needs or for application developers creating solutions for customer-facing products. InfluxDB makes it easy to integrate best-in-class tools together so you can build the ideal solution for your use case, rather than making tradeoffs by choosing an all in one platform that is weaker in certain areas.

InfluxDB vs. other databases

InfluxDB is often compared to other databases. However, when doing a comparison, the entirety of the InfluxDB platform should be considered. There are multiple types of databases that get brought up for comparison. Mostly these are distributed databases like Cassandra or more time series-focused databases like Prometheus or TimescaleDB..

When comparing InfluxData with Cassandra or HBase, there are some stark differences. First, those databases require a significant investment in developer time and code to recreate the functionality provided out of the box by InfluxDB. Specifically, developers will need to write code to shard the data across the cluster, aggregate, and downsample functionality, data eviction and lifecycle management, and summarization. Finally, they'll have to create an API to write and query their new service.

When the rest of the InfluxDB platform is brought into the picture, developers using more general-purpose databases have even more ground to make up. They'll need to write tools for data collection, introduce a real-time processing system, and write code for monitoring and alerting. While some of these tasks are handled with other time series databases, there are a few key differences between the other solutions and InfluxDB. First, other time series solutions like Graphite or OpenTSDB are designed with only regular time series data in mind and can't store raw high-precision data and downsample it on the fly.

When compared the services designed for time series data that are available from major cloud providers like AWS, Google Cloud, or Microsoft Azure, developers choose InfluxData for the following reasons:

- **Open source** — InfluxData is, first and foremost, an open source company. It is committed to sharing ideas and information openly, collaborating on solutions, and providing full transparency to drive innovation. An energized group of developers help continuously improve InfluxData's products, helping to make them more reliable, secure, and awesome. The power of the open source community to drive innovation is unsurpassed by any proprietary software solution.

- **Hybrid cloud and on-premises support** — Distributing assets on one or across multiple cloud-hosting environments is often the best choice for companies. Key among many advantages is avoiding vendor lock-in, which can limit the ability to customize systems and negotiate better rates, ultimately making it very difficult — and expensive — to change cloud providers to meet ever-evolving business and technical needs. Every business has specific functionalities, and a hybrid cloud system offers the flexibility to choose services that best fit their needs, whether to support GDPR regulatory requirements or teams that are spread across multiple providers. From an operations perspective, a multi-cloud system increases efficiency and provides another layer of security to ensure that there is no downtime.
- **Time series focused** – InfluxData as a company is focused entirely on making the best database and ecosystem for working with time series data, while for the major cloud providers, their product offerings are just one of many in their portfolio. Often these cloud-provided services are simply made to be “good enough” for existing customers, rather than best in class in terms of performance and efficiency.

InfluxData *is* time series data

Fueled by the massive growth of connected devices (i.e., IoT) and rapidly increasing instrumentation requirements for next-generation software, time series technology has become more popular. Since launching [InfluxDB](#), an open source time series platform in 2013, we have seen millions of downloads, built an expanding list of enterprise customers, and fostered a growing community that is always finding new ways to deploy and build on our platform.

InfluxData also offers the following commercial editions of InfluxDB:

- **InfluxDB Cloud Serverless** – [InfluxDB Cloud Serverless](#) is a fully managed, elastic, multi-tenant solution with pay-as-you-go pricing that makes it quick and easy to get started with InfluxDB.
- **InfluxDB Cloud Dedicated** – [InfluxDB Cloud Dedicated](#) is a fully managed, single-tenant solution that can be optimized for your workload and provides isolation for security guarantees.
- **InfluxDB Clustered** – [InfluxDB Clustered](#) is a self-managed, on-premises solution that allows you to run an InfluxDB cluster on your own infrastructure.

InfluxData has a narrow focus for what has rapidly become a horizontal use case. Our narrow focus means that the entire stack can have optimizations for performance and developer productivity that other general-purpose solutions can't match, such as high-compression, super-fast engines, and powerful query language designed to best work with flow-based models. At the same time, InfluxData provides a platform that is broadly customizable, making it a perfect choice for developers who want greater control over their tooling than what out-of-the-box applications and solutions provide.

With time series as the key ingredient for custom DevOps monitoring and metrics, real-time Analytics and sensor data, and the Internet of Things, InfluxData continues to be at the forefront of what is sure to be the next wave of data platforms after NoSQL.

According to DB-Engines, over the last 24 months, time series has been the fastest growing database category. InfluxData's InfluxDB is the overwhelming leader among [time series database management systems](#), according to [DB-Engines' results](#).

About InfluxData

InfluxData is the creator of InfluxDB, the open source time series database. Our technology is purpose-built to handle the massive volumes of time-stamped data produced by IoT devices, applications, networks, containers, and computers. We are on a mission to help developers and organizations, such as Cisco, IBM, PayPal, and Tesla, store and analyze real-time data, empowering them to build transformative monitoring, analytics, and IoT applications quicker and to scale. InfluxData is headquartered in San Francisco, with a workforce distributed throughout the U.S. and across Europe.

[Learn more.](#)

InfluxDB documentation, downloads & guides

[Download InfluxDB](#)

[Get documentation](#)

[Additional resources](#)

[Join the InfluxDB community](#)

[Get InfluxDB Cloud](#)



Try InfluxDB

Contact Sales

Contact us for a personalized demo
influxdata.com/contact-sales

