Avinash BM  
9-August-2023

Evaluation Of Timeseries DB For Live Information and Tracking Information

**Table of Contents**

[**Evaluation of TimeSeries DB for Live Information and Tracking Information** 2](#_Toc142476122)

[**Popular TimeSeries databases that you might consider evaluating for your use case includes:** 3](#_Toc142476123)

[**InfluxDB** 4](#_Toc142476124)

[**Major Features:** 4](#_Toc142476125)

[**Common Use Cases:** 5](#_Toc142476126)

[**Pros and cons of InfluxDB are:** 5](#_Toc142476127)

[**Cost / Pricing** 6](#_Toc142476128)

[**Timescale DB** 8](#_Toc142476129)

[**Major Features** 8](#_Toc142476130)

[**Common Use Cases:** 8](#_Toc142476131)

[**Pros and cons of Timeseries DB are:** 9](#_Toc142476132)

[**Cost / Pricing** 9](#_Toc142476133)

[**Azure TimeSeries Insights** 10](#_Toc142476134)

[**Major Features** 10](#_Toc142476135)

[**Common Use Cases:** 10](#_Toc142476136)

[**Pros and cons of Azure TimeSeries Insights are:** 11](#_Toc142476137)

[**Cost / Pricing** 11](#_Toc142476138)

[**Summary** 12](#_Toc142476139)

[**Conclusion** 13](#_Toc142476140)

# **Evaluation of TimeSeries DB for Live Information and Tracking Information**

**Description:**

TimeSeries database (TSDB) for managing live information and tracking data, there are several important factors to consider. A TimeSeries database is designed to handle large volumes of time-stamped data points efficiently, making it well-suited for scenarios such as real-time monitoring, tracking, and analytics.

Here's a list of key criteria to assess when choosing a TSDB for your use case:

1. **Data Model and Flexibility:**

Evaluate whether the TSDB supports the data model required for your use case (e.g., numeric values, events, geospatial coordinates, etc.).

Determine if the database allows for flexible schema design, accommodating changes and additions to your data over time.

1. **Scalability and Performance:**

Assess the TSDB's ability to handle high write and read throughput, as real-time tracking generates a constant stream of data.

Check if the TSDB supports horizontal scaling to accommodate growing data volumes without sacrificing performance.

1. **Latency and Real-Time Capabilities:**

Consider the TSDB's ability to handle real-time data ingestion and provide low-latency access to queried data.

Look for features like streaming support and push-based APIs for live data updates.

1. **Data Retention and Compression:**

Evaluate how the TSDB manages data retention policies and aging out of older data.

Check if the TSDB offers compression techniques to optimize storage and reduce costs.

1. **Querying and Analytics:**

Examine the TSDB's query capabilities for retrieving historical and real-time data.

Look for support for complex queries, aggregations, filtering, and time-based analytics.

1. **Integration and Ecosystem:**

Check if the TSDB integrates well with your existing tech stack and tools, such as visualization tools, monitoring solutions, and programming languages.

Consider whether the TSDB has connectors to popular frameworks or services.

1. **Availability and Reliability:**

Assess the TSDB's features for data replication, high availability, and disaster recovery.

Look for support for multi-region deployment and automatic failover.

1. **Security and Access Control:**

Ensure that the TSDB provides robust security features, including encryption, authentication, and authorization.

Check if the TSDB supports fine-grained access control for different users and roles.

1. **Ease of Management:**

Evaluate the ease of setting up, configuring, and managing the TSDB.

Consider whether the TSDB offers tools for monitoring and managing the database's health and performance.

# **Popular TimeSeries databases that you might consider evaluating for your use case includes:**

1. InfluxDB
2. OpenTSDB
3. TimescaleDB
4. Azure Time Series Insights (if you're considering Azure as your cloud provider).

## **InfluxDB**

A diagram of a company

Description automatically generated

InfluxDB is a time series database designed to handle high write and query loads. InfluxDB is the open-source time series database that is part of the TICK (Telegraf, InfluxDB, Chronograf, Kapacitor) stack.

### **Major Features:**

**High performance:**

InfluxDB is a high-performance data store written specifically for time series data. It allows for high throughput ingest, compression and real-time querying. InfluxDB is written entirely in Go and compiles into a single binary with no external dependencies. It provides write and query capabilities with a command-line interface, a built-in HTTP API, a set of client libraries (e.g., Go, Java, and JavaScript) and plugins for common data formats such as Telegraf, Graphite, Collectd and OpenTSDB.

**SQL-like queries:**

InfluxDB works with InfluxQL, a SQL-like query language for interacting with data. It has been lovingly crafted to feel familiar to those coming from other SQL or SQL-like environments while also providing features specific to storing and analyzing time series data. InfluxQL supports **regular expressions, arithmetic expressions, and time series-specific functions** to speed up data processing.

**Down sampling and data retention:**

InfluxDB can handle **millions of data points per second**. Working with that much data over a long period can lead to storage concerns. **InfluxDB automatically compacts data to minimize your storage space**. In addition, you can easily down sample the data; keeping high-precision raw data for a limited time and storing the lower-precision, summarized data for much longer or until the end of time. InfluxDB has two features that help to automate the down sampling and data expiration processes — **Continuous Queries and Retention Policies**.

It is especially suited for **Industrial IoT applications** that require **real-time analytics** and **monitoring of sensor data**.

### **Common Use Cases:**

1. **IoT:** InfluxDB can collect sensor and device metrics from factories, manufacturing plants, satellites, and smart devices, and provide real-time analytics and monitoring.
2. **Monitoring:** InfluxDB can act as a central platform where all metrics, events, logs, and tracing data can be integrated and centrally monitored.
3. **Data Science:** InfluxDB can be used for forecasting, creating alerts, and API development, as well as monitoring infrastructure networks and data reconciliation.

### **Pros and cons of InfluxDB are:**

**Pros:**

* Schema less ingestion, which means you don't have to define the data structure before inserting data.
* Huge community and integrations with popular tools such as **Kafka**, Grafana, AWS, etc.
* Fast, scalable, and open source.

**Cons:**

* Performance decreases on datasets with high cardinality, which means having many unique values in a field.
* Open-Source version supports a single instance, which means **no clustering or replication for high availability**.
* Multiple versions with different database engines, supported languages, and features, which can cause confusion and compatibility issues.
* New querying language (Flux) to learn, which is different from the previous InfluxQL.

**The TICK Stack**

TICK stands for Telegraf, InfluxDB, Chronograf, and Kapacitor, which are integrated in a cohesive architecture, or "stack." Together these technologies provide a platform that can capture, monitor, store, and visualize all data in a time series, allowing for informed business decisions in real-time.

1. **Telegraf is a plugin-driven server agent** for collecting and reporting metrics. Telegraf plugins source a variety of metrics directly from the systems it runs on, pulling metrics from third-party APIs or even to listen for metrics via a **StatsD and Kafka consumer service**. It also has output plugins to send metrics to a variety of other datastores, services, and message queues, including InfluxDB, Graphite, OpenTSDB, Datadog, Librato, **Kafka**, MQTT, NSQ and many others.
2. **InfluxDB is a time series database** built from the ground up to handle high write and query loads. InfluxDB is a custom high-performance datastore written specifically for time-stamped data, and especially helpful for use cases such as **DevOps monitoring, IoT monitoring, and real-time analytics**. Conserve space on your machine by configuring InfluxDB to keep data for

a defined period, and to automatically expire and delete unwanted data from the system. InfluxDB also offers a SQL-like query language for interacting with data.

1. **Chronograf** is the administrative user interface and visualization engine of the stack. It makes it easy to setup and maintain the monitoring and alerting for your infrastructure. It’s simple to use and includes templates and libraries that allow you to rapidly build dashboards with real-time visualizations of your data and to easily create alerting and automation rules.

### **Cost / Pricing**

A white table with a check mark

Description automatically generated with medium confidence

A screenshot of a white screen

Description automatically generated

A screenshot of a computer

Description automatically generated

A grid of white paper with purple ticks

Description automatically generated

InfluxDB POC

InfluxDB offers several ways to get started with a POC, including:

1. InfluxDB Cloud Serverless: For smaller workloads, multi-tenant infrastructure.
2. InfluxDB Cloud Dedicated: For scaled workloads, dedicated infrastructure.
3. InfluxDB Enterprise: For scaled workloads, on-prem.

During a POC, you can test various features of InfluxDB, such as its REST API, data replication and scaling, automated backups, disaster recovery, high availability, data encryption, and security1. You can also evaluate its performance and scalability to ensure that it meets your requirements.

## **Timescale DB**

Timescale is a PostgreSQL data platform for time-series, events, and analytics. It gives you the reliability of PostgreSQL, the time-series superpowers of Timescale DB, and the peace of mind of a fully managed service. It provides features like automatic backup and restore, high availability with replication, seamless scaling and resizing, and much more.

### **Major Features**

**Hypertables** are PostgreSQL tables that automatically partition your data by time. You interact with hypertables in the same way as regular PostgreSQL tables, but with extra features that makes managing your time-series data much easier.

In Timescale, hypertables exist alongside regular PostgreSQL tables. Use hypertables to store time-series data. This gives you improved insert and query performance, and access to useful time-series features. Use regular PostgreSQL tables for other relational data.

Timescale supports full **SQL**, so you don't need to learn a custom query language. This section contains some simple queries that you can run directly on this page.

* SELECT, WHERE, LIMIT
* ORDER BY, GROUP BY
* FIRST AND LAST VALUE

Aggregation refers to several different calculations that you can perform on your data. For example, if you have data showing temperature changes over time, you can calculate an average of those temperatures, or a count of how many readings have been taken. Average, sum, and count are all examples of simple aggregates. Time-series data usually grows very quickly. And that means that aggregating the data into useful summaries can become very slow. Continuous aggregates make aggregating data lightning fast.

There are three main types of aggregation:

1. Materialized views
2. Continuous aggregates
3. Real time aggregates

### **Common Use Cases:**

* **Monitoring computer systems:** virtual machines, servers, container metrics, CPU, free memory, net/disk IOPs, service and application metrics such as request rates, and request latency.
* **Financial trading systems:** securities, cryptocurrencies, payments, and transaction events.
* **Internet of things:** data from sensors on industrial machines and equipment, wearable devices, vehicles, physical containers, pallets, and consumer devices for smart homes.
* **Eventing applications:** user or customer interaction data such as clickstreams, pageviews, logins, and signups.
* **Business intelligence:** Tracking key metrics and the overall health of the business.
* **Environmental monitoring:** Temperature, humidity, pressure, pH, pollen count, air flow, carbon monoxide, nitrogen dioxide, or particulate matter.

### **Pros and cons of Timeseries DB are:**

**Pros:**

* **Optimized for time-stamped data:** A TSDB is built specifically for handling metrics and events or measurements that are time-stamped. It is optimized for measuring change over time.
* **Data lifecycle management:** A TSDB has properties that make time series data very different than other data workloads, such as data lifecycle management, summarization, and large range scans of many records.
* **Forecasting Trends and Patterns:** One of the most significant advantages of time series analysis is its ability to forecast future trends and patterns.

**Cons:**

* **Limited Application Scope:** One of the main drawbacks of time series analysis is its limited application scope.
* **Dealing with Outliers and Noise:** Time series analysis can be challenging when dealing with outliers and noise in the data.

### **Cost / Pricing**

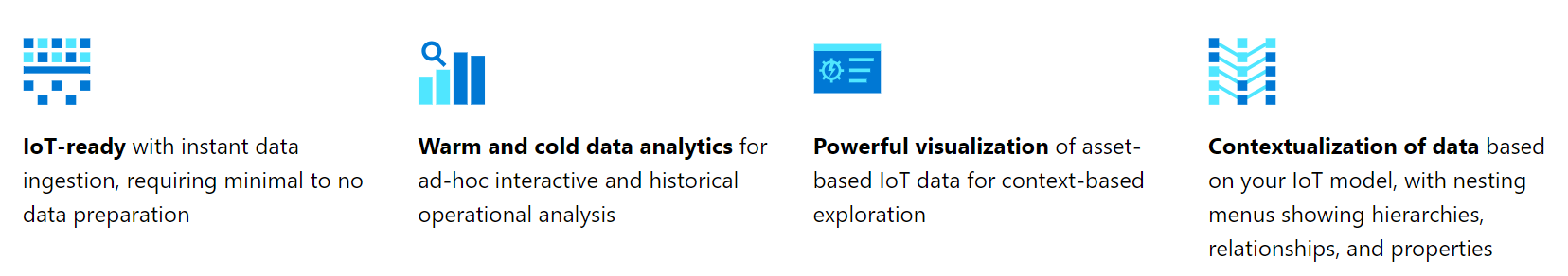
A screenshot of a computer

Description automatically generated

## **Azure TimeSeries Insights**

Improve operations and decision-making with decades of IoT data delivered with rich visualization and a turnkey experience. Use time series data insights and interactive analytics to accelerate IoT analytics data use throughout your organization.

Turn disparate data streams into insights and provide context using Time Series Model. From the beginning of your engagement to the full connection of your industrial assets, access near-real-time data streaming and historical data for a complete picture.



### **Major Features**

**Inherently scalable:** Uncover new insights as your IoT data grows. Ingest and store billions of IoT events as you scale to hundreds of thousands of assets. With our pay-as-you go pricing, you get an end-to-end industrial IoT analytics platform that matures with your business.

**Open and flexible:** Make IoT data accessible to the masses. Use the Time Series Insights explorer for rich visualizations and a turnkey experience. Link industrial IoT data with other business metrics with the native Power BI connector. Build custom dashboards with the open-source data formats, rich APIs, and Javascript SDK.

### **Common Use Cases:**

* Data exploration and visual anomaly detection

Instantly explore and analyze billions of events to spot anomalies and discover hidden trends in your data. Azure Time Series Insights Gen2 delivers near real-time performance for your IoT and DevOps analysis workloads.



* Operational analysis and process efficiency:

Monitor the health, usage, and performance of your assets to drive operational efficiency at scale. Time Series Insights gives you the capacity to manage billions of data events from diverse and unpredictable IoT devices, without compromising performance.

* Advanced analytics with Azure IoT:

Integrate Time Series Insights with advanced analytics services, such as Azure Machine Learning, Azure Databricks, Apache Spark, and others. Compare real-time data against your models to quickly discover new patterns or potential issues.

### **Pros and cons of Azure TimeSeries Insights are:**

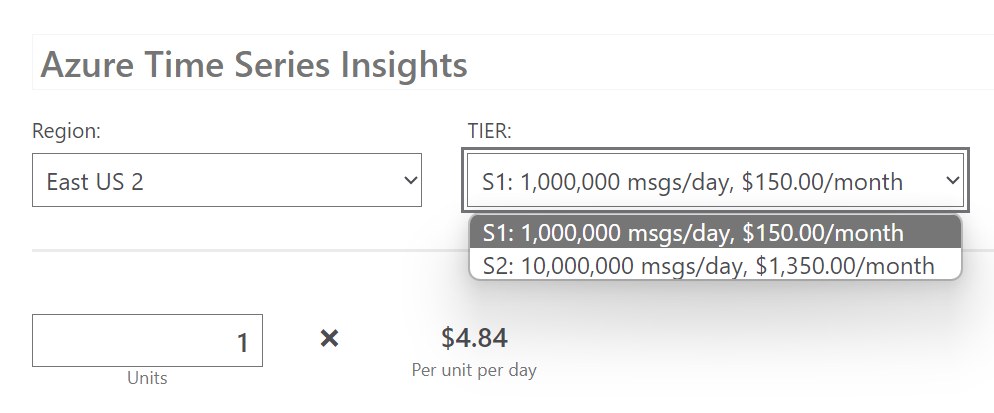
**Pro:**

* **Scalability:** Azure Time Series Insights is designed to support IoT at scale. It can ingest from 1 million to 100 million events per day, with a default retention span of 31 days. You can visualize and analyze live data streams in near real-time, alongside historical data.
* **Insights:** Azure Time Series Insights helps you gain insights into your sensor data. Use it to spot trends and anomalies, conduct root-cause analyses, and avoid costly downtime. Cross-correlation between real-time and historical data helps you find hidden trends in the data.
* **Near real-time insights:** Azure Time Series Insights can ingest millions of sensor events per day, with one-minute latency. Azure Time Series Insights helps you gain insights into your sensor data.

**Cons:**

* Training: At the beginning, you may need comprehensive training and IT experience for configuration.
* Internet dependency: If you don’t have internet access, you won’t have access to resources.
* Region Supported – Canada, United States

### **Cost / Pricing**



## **Summary**

InfluxDB, TimescaleDB, and Azure Time Series Insights are all popular time-series databases. They are designed to handle large amounts of time-series data, such as that generated by IoT devices. Each has its own strengths and weaknesses, and the best choice for you will depend on your specific needs.

1. InfluxDB is a market leader in the time-series database space and was first released in 2013. It is an open-source database that is licensed with MIT License for a single node, with paid plans available for InfluxDB Cloud and InfluxDB Enterprise that provide clustering and other production-ready features.
2. TimescaleDB is a relational database that relies on the relational data model commonly found in PostgreSQL, MySQL, SQL Server, Oracle, etc. It is purpose-built for time-series data and provides scalability, insights, and near real-time insights.
3. Azure Time Series Insights is a fully managed analytics, storage, and visualization service for managing IoT-scale time-series data in the cloud. It provides massively scalable time-series data storage and enables you to explore and analyze billions of events streaming in from all over the world in seconds.

When it comes to choosing between these three databases, there are several factors to consider. These include the data model, query language, reliability, performance, ecosystem, operational management, company and community support.

A side-by-side comparison of Azure Time Series Insights vs. InfluxDB based on preference data from user reviews shows that Azure Time Series Insights rates 4.1/5 stars with 52 reviews while InfluxDB rates 4.5/5 stars with 98 reviews.

|  |  |  |
| --- | --- | --- |
|  | **Azure Time Series Insights** | **InfluxDB** |
| **Criteria** |
| **Overview** | **G2 Rating** | **G2 Rating** |
| **Star Rating** | 4.1 | 4.5 |
| **Largest Market Segment** | Small-Business (40.9% of reviews) | Small-Business (58.9% of reviews) |
| **Entry Level Price** |  | $0.00 |
| **Total Number of Reviews** | 44 | 95 |
| **General Ratings** |  |  |
| **Meets Requirements** | 9 | 8.7 |
| **Ease of Use** | 8.3 | 8.5 |
| **Ease of Setup** | 7.7 | 8.8 |
| **Ease of Admin** | 8.5 | 8.7 |
| **Quality of Support** | 8.5 | 8.6 |
| **Has the product been a good partner in doing business?** | 8.3 | 9.1 |
| **Product Direction (% positive)** | 8.3 | 8.6 |

## **Conclusion**

InfluxDB easier to use, set up, and administer and preferred doing business.

Azure Time Series Insights meets the needs of their business better than InfluxDB.