**Why do we need a time-series database?**

In order to explain the question,we may start from the concept of time-series so that we can figure out why we use time-series data instead of other kinds of data.

Time-series data is a series of data points indexed/listed in time order.Time-series data is a collection of observations obtained through [repeated measurements over time](https://www.influxdata.com/blog/what-is-time-series-data-and-why-should-you-care/). Plot the points on a graph, and one of your axes would always be time.Time-series metrics refer to a piece of data that is tracked at an increment in time. For instance, a metric could refer to how much inventory was sold in a store from one day to the next.

Time-series data is everywhere, since time is a constituent of everything that is observable. As our world gets increasingly instrumented, sensors and systems are constantly emitting a relentless stream of [time series data](https://www.influxdata.com/time-series-database/). Such data has numerous applications across various industries. Let’s put this in context through some examples.Nowadays,time-series data have become part of our life.In aspects like finance,transportation,manufacturing and so on,time-series data are widely used.

Time-series data features its time.In other words,it’s time stamped and stable data which is immutable.Since time series data comes in time order, it is almost always recorded in a new entry, and as such, should be immutable and append-only (appended to the existing data). It doesn’t usually change but is rather tacked on in the order that events happen. This property distinguishes time series data from relational data which is usually mutable and is stored in relational databases that do online transaction processing, where rows in databases are updated as the transactions are run and more or less randomly; taking an order for an existing customer, for instance, updates the customer table to add items purchased and also updates the inventory table to show that they are no longer available for sale.

The fact that time series data is ordered makes it unique in the data space because it often displays serial dependence. Serial dependence occurs when the value of a datapoint at one time is statistically dependent on another datapoint in another time.

Though there are no events that exist outside of time, there are events where time isn’t relevant. Time series data isn’t simply about things that happen in chronological order — it’s about events whose value increases when you add time as an axis. Time series data sometimes exists at high levels of granularity, as frequently as microseconds or even nanoseconds. With [time series data](https://www.influxdata.com/blog/getting-started-influxdb-cloud-telegraf-on-macos/), change over time is everything.

Since time-series data is usually kept and rarely deleted or updated,real-time data computing is preferred.What’s more,query for time-series data is always in time and space range.

With so many features,we can operate on time-series data to get desired results.Briefly speaking,we can make real time alert,forecasting,identifying the problems while keeping insights of operation.

Taking real time alert as an example,when you see the world as time series as we do, alerts appear as simple series of boolean values, with true meaning that the alerting condition is met and false or the absence of value that it is not. The meaning and complexity of the underlying condition are very flexible. It could be a simple threshold that is passed, a diversion from a previous trend, a detected anomaly, or any other thing you might think of.

Treating alerts as time series means to periodically assess those conditions and, depending on the outcome, to insert boolean values in a set of time series. Those series will accrue values and give you a very good understanding of your systems.

In order to make use of time-series data,the tool for dealing with time-series data is time-series database.A time-series database captures a set of fixed values and dynamic values. For example, in a web application, a set of data points might be actions performed by a group of visitors. The dynamic values are the number of desired actions, such as eCommerce purchases, performed by the visitors over time. By analyzing purchases over time, the organization can understand the value of each group of users, and prioritize marketing activity for different [customer segments](https://formation.ai/blog/the-problems-with-customer-segmentation/).

Ideally, time series records should be written into a repository in a format that enables quick time-based writes and reads. Because the records are time-stamped, the order of data points becomes a native aspect of the data. We can then use this order to deliver the data to a stream processing engine, which can then treat the ordered data like a data stream.

Now coming back to the topic:why do we need time-series database?

·Amount

The first reason comes from the character of time-series data:amount.Obviously,time-series data is collected with time and increase with time at the same time.Once we collect data with seconds,the amount of data is predictably huge.In order to get data as accurate as impossible,the rate of collection can’t be too low.In this case,there are three challenges urgent to solve:data ingestion rate,query latency and storage. Briefly speaking,a powerful tool is needed to keep and process these data.

·Special data analysis

In addition to huge volume,special data analysis is another question faced with time-series data processing like down sampling,interpolation,time weighted average and so on.In trend research,down sampling is usually used.By sampling in proper interval,we can get the trend of data and make prediction at some extend.Even if you’re just starting to collect this type of data and scale is not a concern at the moment, these features can still provide a better user experience and make data analysis tasks easier. Having built-in functions and features to analyze trends readily available at the data-layer often leads you to discover opportunities you didn’t know existed, no matter how big or small your data set.

Time weighted averages are used in cases where a time series is not evenly sampled. Time series data points are often evenly spaced, for example every 30 seconds, or every hour. But sometimes data points are recorded irregularly, for example if a value has a large change, or changes quickly. Computing an average using data that is not evenly sampled is not always useful.

For example, if you have a lot of ice cream in freezers, you need to make sure the ice cream stays within a 0-10℉ (-20 to -12℃) temperature range. The temperature in the freezer can vary if folks are opening and closing the door, but the ice cream only has a problem if the temperature is out of range for a long time. You can set your sensors in the freezer to sample every five minutes while the temperature is in range, and every 30 seconds while the temperature is out of range. If the results are generally stable, but with some quick moving transients, an average of all the data points weights the transient values too highly. A time weighted average weights each value by the duration over which it occurred based on the points around it, producing much more accurate results.

·Internet of Things

IoT technologies generate and use massive quantities of time-series data. For example, mobile devices, eCommerce applications, automobiles, and systems for inventory management, all time-stamp data according to events. Quick ingestion of time-series data is critical to ensure IoT devices and metrics can continuously capture data and store it for analysis.

·Monitoring Analysis

Time series data is often used to monitor computer system metrics. This process works by reading data from computer systems, of users who have agreed to let others monitor the computer system. Typically, metrics include process count and memory utilization, which help check the utilization of computer resources and assess if resources should be reallocated.

·Key Performance Indicators (KPIs)

KPIs are time-oriented and sampled repeatedly and fit easily into time series data. Some examples of these KPIs may include profit, revenue, cost, conversion rate, number of transactions and average order value. Once this information is collected and stored, it can be used to create dashboards.

·Anomaly Detection

Anomaly detection helps detect unexpected deviations in time series data. Time series data captures a value whenever a system change occurs. Organizations can use these values to measure change, discover ways in which changes occured in the past, monitor what is currently happening, and leverage this accumulated data to predict future events.

 Virtualization is a main factor in achieving anomaly detection. A time series plot, for example, provides the visualization people often need to spot outliers. Automated anomaly detection is another way, one that often expedites the process, offering insights in real-time. This can allow you to quickly correlate outliers.

In brief,time-series data has become part of our life and benefit ourselves in various aspects.Under such circumstances,time-series databases will predictably be the next widely welcomed and used kind of database.The best way to follow the step of era is to understand learn time-series data and make use of time-series databases.