

Why do we need a time-series database?

What is time-series data?

Time series data, also referred to as time-stamped data, is a sequence of data points indexed in time order. Time-stamped is data collected at different points in time. These data points typically consist of successive measurements made from the same source over a time interval and are used to track change over time.

Time series data is a collection of observations obtained through repeated measurements over time. 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. Such data has numerous applications across various industries. Let's put this in context through some examples.

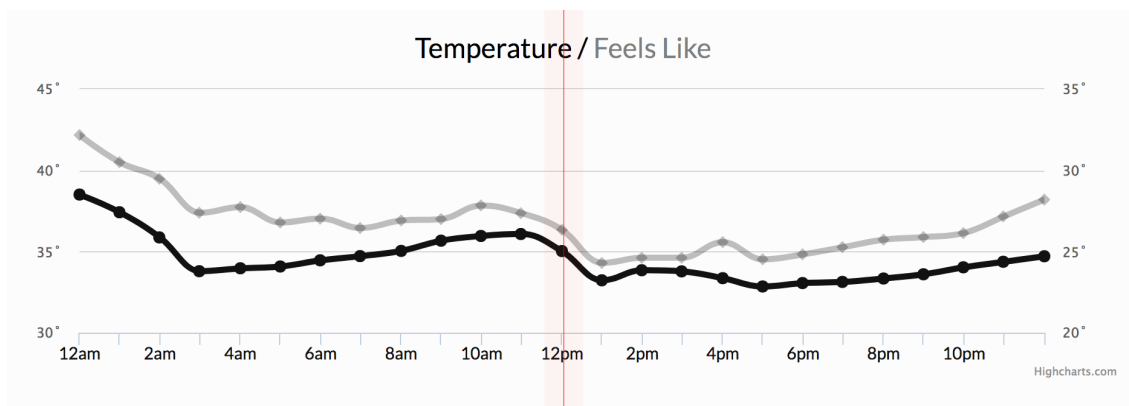
Examples of time series analysis:

- Electrical activity in the brain
- Rainfall measurements
- Stock prices
- Number of sunspots
- Annual retail sales
- Monthly subscribers
- Heartbeats per minute

Time series examples

Weather records, economic indicators and patient health evolution metrics — all are time series data. Time series data could also be server metrics, application performance monitoring, network data, sensor data, events, clicks and many other types of analytics data.

Notice how time — depicted at the bottom of the below chart — is the axis.



Example 1: Weather conditions

In the next chart below, note time as the axis over which stock price changes are measured. In investing, a time series tracks the movement of data points, such as a security's price over a specified period of time with data points recorded at regular intervals. This can be tracked over the short term (such as a security's price on the hour over the course of a business day) or the long term (such as a security's price at close on the last day of every month over the course of five years).



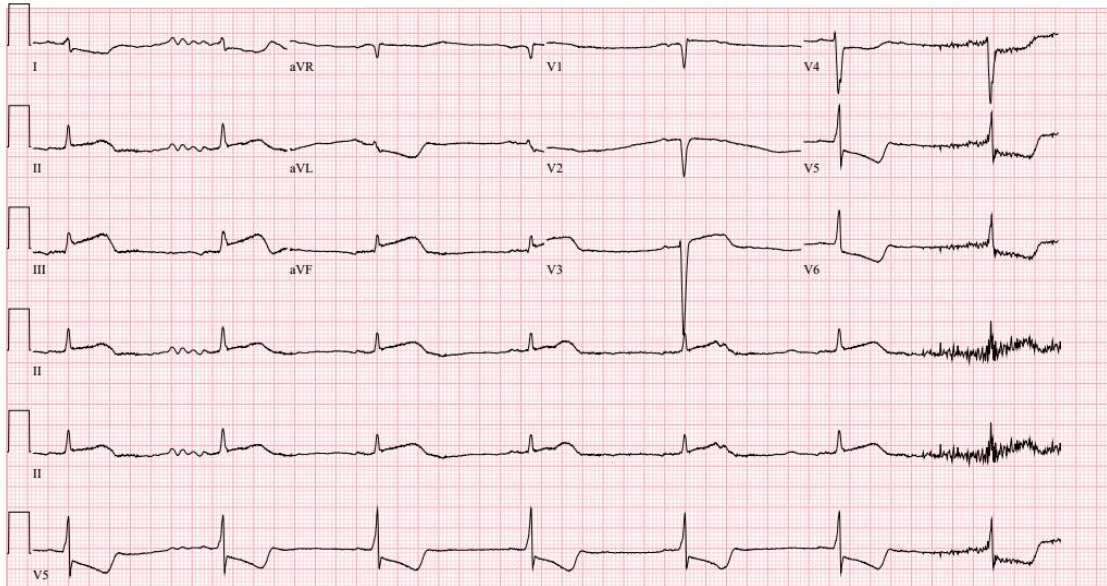
Example 2: Stock exchange

The cluster monitoring example below, depicting disk ops write and usage data, would be familiar to Network Operation Center teams. Remember that monitoring data is time series data.



Example 3: Cluster monitoring

Another familiar example of time series data is patient health monitoring, such as in an electrocardiogram (ECG), which monitors the heart's activity to show whether it is working normally.



Example 4: Health monitoring

What can we do with time-series data?

Time series data is gathered, stored, visualized and analyzed for various purposes across various domains:

1. In data mining, pattern recognition and machine learning, time series analysis is used for clustering, classification, query by content, anomaly detection and forecasting.

2. In signal processing, control engineering and communication engineering, time series data is used for signal detection and estimation.
3. In statistics, econometrics, quantitative finance, seismology, meteorology, and geophysics the time series analysis is used for forecasting.

Time series data can be visualized in different types of charts to facilitate insight extraction, trend analysis, and anomaly detection. Time series visualization and dashboarding tools include the InfluxDB UI and Grafana.

The term 'time series patterns' describes long-term changes in the series. Whether measured as a trend, seasonal, or cyclic pattern, the correlation can be calculated in a number of ways (linear, exponential, etc.), and the direction may change at any given time.

Time series data is used in time series analysis (historical or real-time) and time series forecasting to detect and predict patterns — essentially looking at change over time. Following is a brief overview of each.

Time series analysis methods

Time series analysis is a method of analyzing a series of data points collected over a period of time. In time series analysis, data points are recorded at regular intervals over a set period of time, rather than intermittently or at random.

Time series analysis is the use of statistical methods to analyze time series data and extract meaningful statistics and characteristics about the data. TSA helps identify trends, cycles, and seasonal variances to aid in the forecasting of a future event. Factors relevant to TSA include stationarity, seasonality and autocorrelation.

Time series analysis can be useful to see how a given variable changes over time (while time itself, in time series data, is often the independent variable). Time series analysis can also be used to examine how the changes associated with the chosen data point compare to shifts in other variables over the same time period.

Learn more about time series analysis methods, including spectral analysis, wavelet analysis, autocorrelation, and cross-correlation.

Time series forecasting methods

Time series forecasting uses information regarding historical values and associated patterns to predict future activity.

Time series forecasting methods include:

- Trend analysis
- Cyclical fluctuation analysis
- Seasonal pattern analysis

As with all forecasting methods, success is not guaranteed. Machine learning is often used for this purpose. So are its classical predecessors: Error, Trend, Seasonality Forecast (ETS), Autoregressive Integrated Moving Average (ARIMA) and Holt-Winters.

To 'see things' ahead of time, time series modeling (a forecasting method based on time series data) involves working on time-based data (years, days, hours, minutes) to derive hidden insights that inform decision-making. Time series models are very useful models when you have serially correlated data. Most businesses work on time series data to analyze sales projections for the next year, website traffic, competitive positioning and much more.

Learn more about time series forecasting methods, including decompositional models, smoothing-based models, and models including seasonality.

What is a time series database?

A time series database (TSDB) is a database optimized for time-stamped or time series data. Time series data 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 analytics data.

A time series database is built specifically for handling metrics and events or measurements that are time-stamped. A TSDB is optimized for measuring change over time. Properties that make time series data very different than other data workloads are data lifecycle management, summarization, and large range scans of many records.

Why is a time series database important now?

Time series databases are not new, but the first-generation time series databases were primarily focused on looking at financial data, the volatility of stock trading, and systems built to solve trading. But financial data is hardly the only application of time series data anymore — in fact, it's only one among numerous applications across various industries. The fundamental conditions of computing have changed dramatically over the last decade. Everything has become compartmentalized. Monolithic mainframes have vanished, replaced by serverless servers, microservers, and containers.

Today, everything that can be a component is a component. In addition, 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, human bodies. Everything has, or will have, a sensor. So now, everything inside and outside the company is emitting a relentless stream of metrics and events or time series data.

This means that the underlying platforms need to evolve to support these new workloads — more data points, more data sources, more monitoring, more controls. What we're witnessing, and what the times demand, is a paradigmatic shift in how we approach our data infrastructure and how we approach building, monitoring, controlling, and managing systems. What we need is a performant, scalable, purpose-built time series database.

This is why developers are increasingly adopting time-series databases and using them for a variety of use cases:

- Monitoring software systems: Virtual machines, containers, services, applications
- Monitoring physical systems: Equipment, machinery, connected devices, the environment, our homes, our bodies
- Asset tracking applications: Vehicles, trucks, physical containers, pallets
- Financial trading systems: Classic securities, newer cryptocurrencies
- Eventing applications: Tracking user/customer interaction data
- Business intelligence tools: Tracking key metrics and the overall health of the business
- (and more)

Conclusion

Time Series Databases are not new, but the first-generation Time Series Databases were primarily focused on looking at financial data, the volatility of stock trading, and systems built to solve trading. Today, everything that can be a component is a component. In addition, 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, human bodies. Everything has, or will have, a sensor. So now, everything inside and outside the company is emitting a relentless stream of metrics and events or time series data. This means that the underlying platforms need to evolve to support these new workloads—more data points, more data sources, more monitoring, more controls.

References

- Time-Series: https://en.wikipedia.org/wiki/Time_series
- What is time series data? <https://www.influxdata.com/what-is-time-series-data/>