Time Series Databases

Introduction

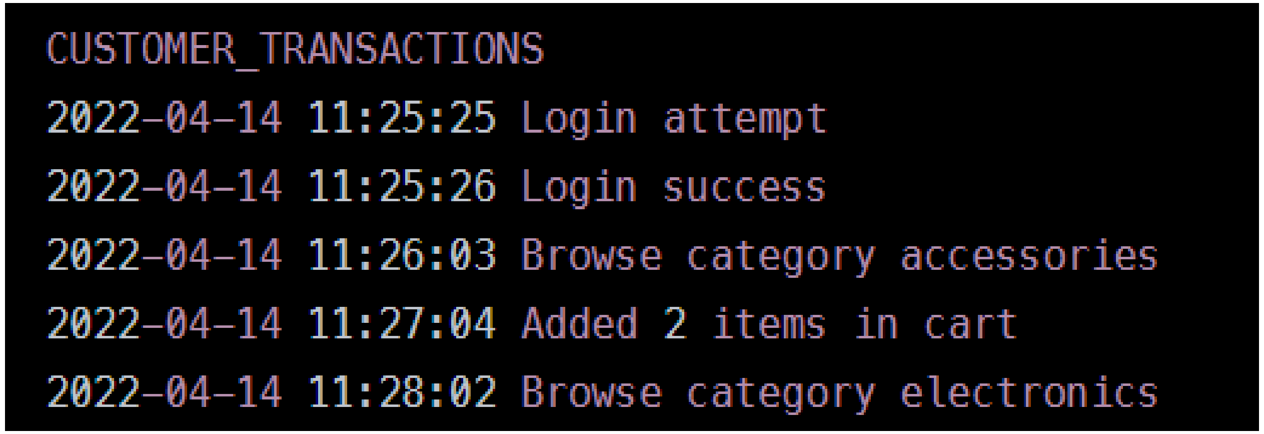
By, Dr. Minakshi More

What is Time series Data?

Time series data is the collection of data that is queried and indexed based on time period.
Relational and non-relational databases have timestamp data types to store time-related data. Time series databases are specifically designed for time series data management.

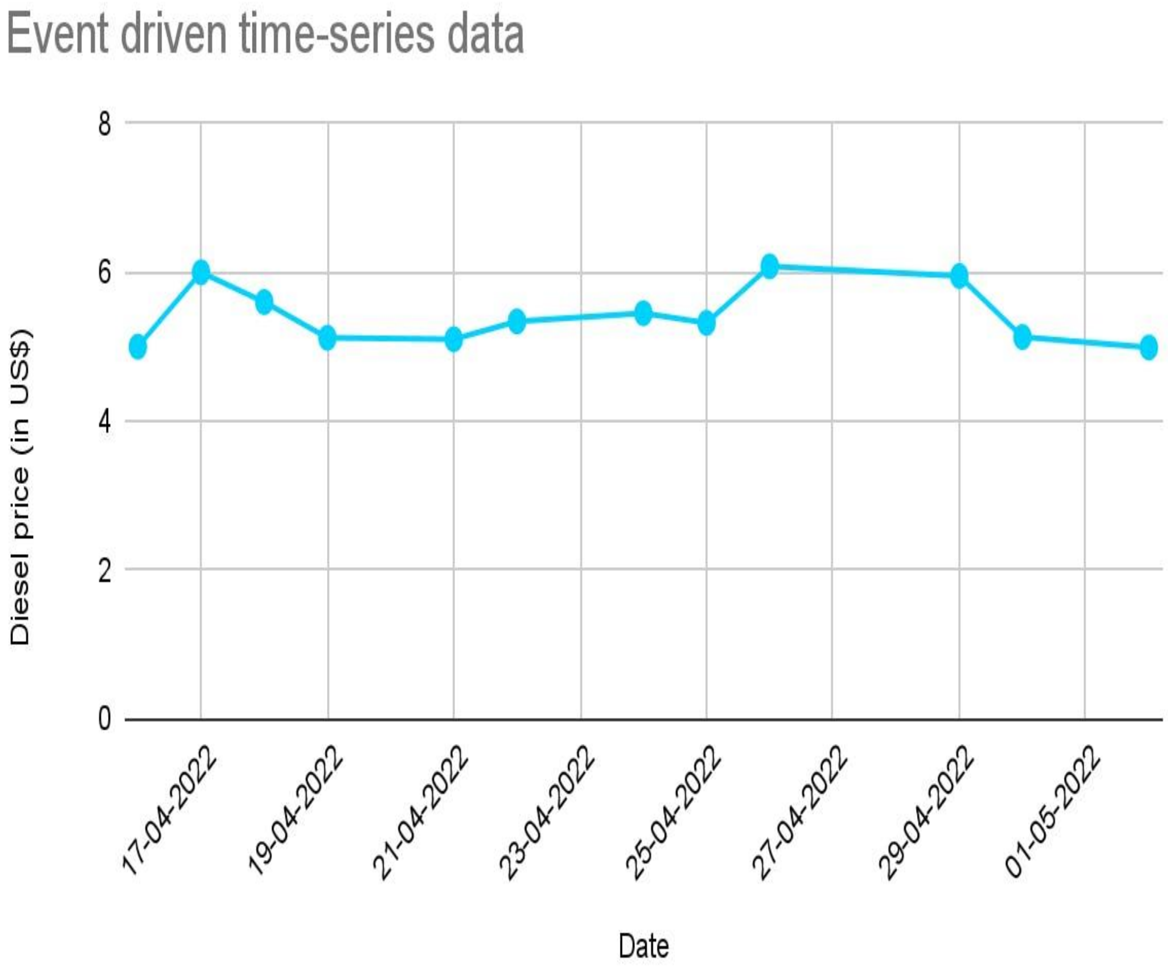
Time series data

If you think about it, all the data that we store has a timestamp attached to it. For example, log files, customer login times, sensor data from IoT devices, traffic data, weather data, and browser history all have timestamps attached.



```
CUSTOMER_TRANSACTIONS
2022-04-14 11:25:25 Login attempt
2022-04-14 11:25:26 Login success
2022-04-14 11:26:03 Browse category accessories
2022-04-14 11:27:04 Added 2 items in cart
2022-04-14 11:28:02 Browse category electronics
```

Time series data can be measured in seconds and minutes (like sensor-based devices), hourly (like phone usage), daily (petrol price), weekly (timesheets), monthly (electricity consumption), quarterly (performance reports), half-yearly (company growth), or annually (profits and revenue). Time series data can be at regular intervals or event-driven (irregular):



Regular time series data

Date	Diesel price (in US$)
16-04-2022	5
17-04-2022	6
18-04-2022	5.36
19-04-2022	5.12
20-04-2022	5.12
21-04-2022	5.1
22-04-2022	5.34
23-04-2022	5.34

Time series database

Databases that provide special features to efficiently handle (store, manipulate, and retrieve) time series data are called time series databases. Some popular time series databases are Prometheus, InfluxDB, and TimeScaleDB. Databases like MongoDB provide time series collections to handle time series data, so you can get the benefits of both a time series and a non-relational database in one.

How is data stored in a time series database?

As shown in the example, data in a time series database has a timestamp and at least one metric related to it. For example, the diesel price was $5.45 (metric) on 24-04-2022. We can add more metrics as well—for example, petrol prices, stock prices, or the number of cars visiting the state museum.

Date	Diesel price	Petrol price	Stock price	Number of cars visiting the state museum
24-04-2022	5.45	7.13	100	780

How does a time series database work?

Time series databases store data as time-value pairs for easy analysis and querying. Time series databases can efficiently handle concurrent series—i.e., multiple metrics parallel—making them well-suited for banking and financial transactions.

A time series database has three aspects: database features, time series features, and data features.

Database features

This includes the basic CRUD (Create, Read, Update, and Delete) features, as well as features like high availability, scalability, and reliability. The database should be able to handle large amounts of writes, and reads/updates should be at particular time windows.

Time-series features

The time is stored as a timestamp, which includes the time in precision of seconds and milliseconds. Date can be stored in various formats using the DateTime data type. Timestamp supports calendar and time zone adaptation. Time series databases also provide support for getting aggregations and statistics about the data based on time.

Data Features

Data is appended in the sequence of time and is stored as time, value, and events. Data can have many dimensions. The data often does not require relationships between entries of different tables and older data is removed or compressed and archived.

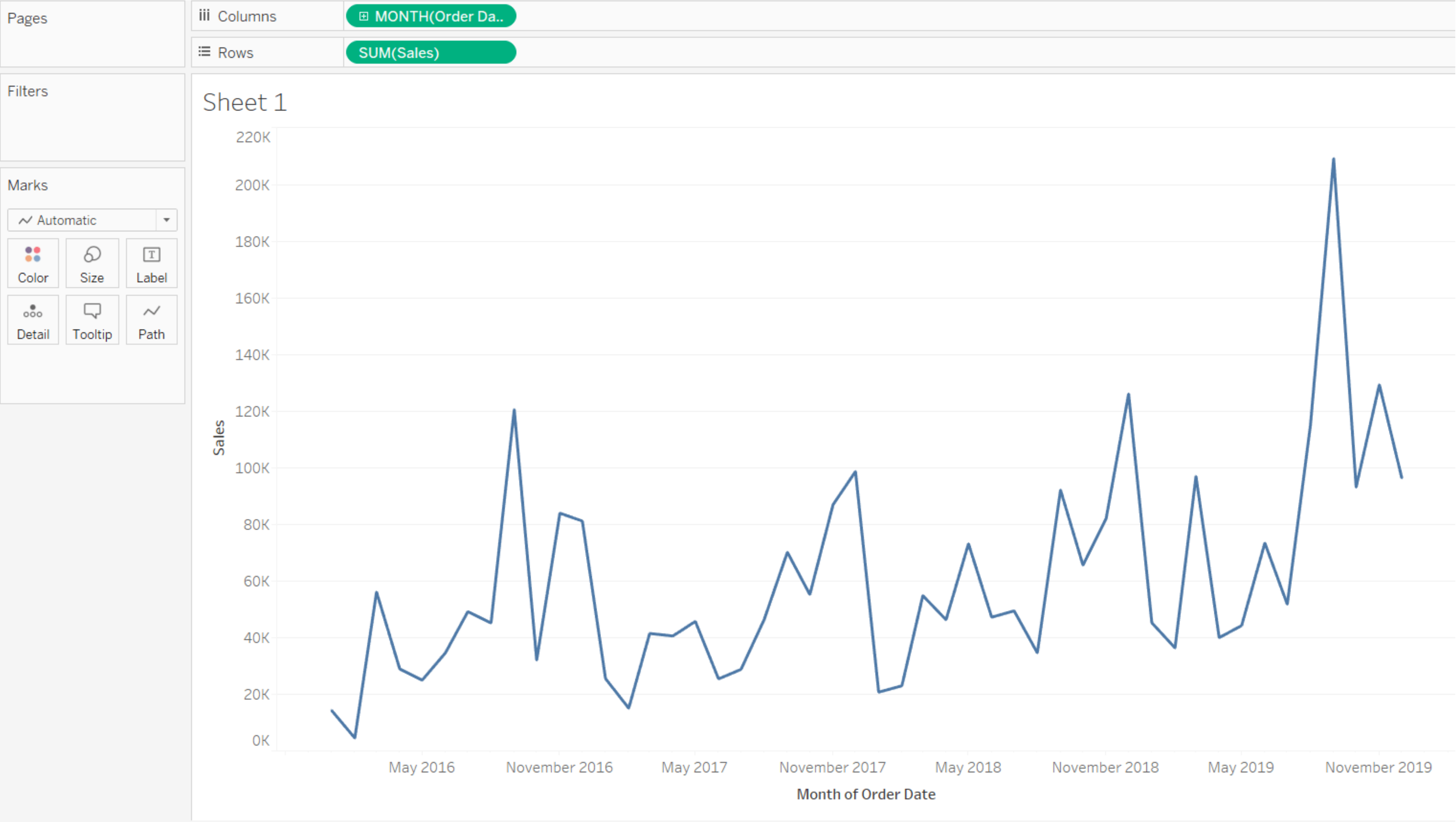
Why is a time series database important now?

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. Massive 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.

Time series analysis examples



Time series analysis is used for non-stationary data—things that are constantly fluctuating over time or are affected by time. Industries like finance, retail, and economics frequently use time series analysis because currency and sales are always changing. Stock market analysis is an excellent example of time series analysis in action, especially with automated trading algorithms. Likewise, time series analysis is ideal for forecasting weather changes, helping meteorologists predict everything from tomorrow’s weather report to future years of climate change. Examples of time series analysis in action include:

- Weather data
- Rainfall measurements
- Temperature readings
- Heart rate monitoring (EKG)
- Brain monitoring (EEG)
- Quarterly sales
- Stock prices
- Website Activity data
- Industry forecasts
- Interest rates

Time Series Databases

- QuestDB
- Timescale DB
- Influx DB
- Amazon Timestream
- Cassandra (NoSQL)

Thank You