Time Series Data

- Time series data is a sequence of values, typically consisting of successive measurements made from the same source over a time interval.
- 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
- The key difference with time series data from regular data is that mostly you ask questions about it over time.
- With time series, it’s common to request a summary from a larger period. This requires going over a range of data points to perform some computation.

Uses

• Time Series Analysis:

Explore how a given variable changes over time
E.g., presidential candidates in U.S. elections can be considered a Time Series Analysis task.

• Regression Analysis:

Examine how the changes associated with a specific variable can cause shifts in other variables over the same time period.
E.g. : Stock Market Analysis - How one stock's prices over time affect other stocks’

• Time Series Forecasting:

Use information regarding historical values and associated patterns to predict future activity.
E.g. : Weather Forecasting, Sales Forecasting

Time Series Databases

- Optimized for the collection, storage, retrieval, and processing of time series data.
- Specifically for handling metrics,events or measurements that are time-stamped.
- Optimized for measuring change over time. Allows its users to create, enumerate,update, destroy and organize various time series in a more efficient manner.
- It’s common to keep high-precision data around for a short period of time.

- For every data point that goes into the database, it will have to be deleted after its period 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 summarizing that data at scale.

Required Properties

Data Location:

If related data is not located together in the physical storage, the data queries can be really slow and even result in timeouts.

A TSDB co-locates chunks of data within the same time range on the same physical part of the database cluster and hence enables quick access for faster, more efficient analysis.

Fast and easy range queries:

The correlated data together it ensures that the range queries are fast.

Regular databases produce an index out of memory error because of the sheer volume of time series data and subsequently affect the performance of read and write operations.

High write performance:

Databases are not able to serve requests predictably and quickly during peak loads.

TSDBs should ensure high availability and high performance for both read and write operations during peak loads

Time series data is usually being recorded every second or even less than that, so write operations need to be fast.

Data compression:

As time-series data is mostly recorded per second they usually need a better data compression technique.

TSDBs should provide functionality to perform roll-ups in such scenarios for data compaction.

Scalability:

Time-series data increases very quickly and regular databases are not designed to handle this scalability.

Time series databases are designed to take care of scale by introducing functionalities that are only possible when you treat time as your first concern.

This can result in performance improvements,including: higher insertion rates, faster queries at scale, and better data compression.

Usability:

TSDBs typically include functions and operations that are common to time series data analysis.

So, this increases the usability by improving the user experience in case of dealing with time related analysis.

Popularity

TSDBs are to handle time series data, but their popularity seems to have increased with the emergence of Internet of things (IoT).

IoT is a network of physical devices/objects with connectivity which enables them to exchange and collect data.

Such technologies are generating large amount of data which is usually time-stamped, so with the increase in popularity of IoT, TSDBs popularity increased even more, because they can be used to efficiently store sensors and devices' data in this domain.

Facebook ,eBay are using TSBS instead of relational database for data monitoring purposes

Top TS Databases

Top few TSDBs and their ranking can be seen in the figure according to DB-Engines Ranking of Time Series DBMS.

DB-Engines is an independent website which ranks databases based on search engine popularity, social media mentions, number of job offers, and technical discussion frequency.

Influx DB is ranked number one in this list as of March 2023.

include secondary database models

41 systems in ranking, March 2023

Rank Mar 2023	Rank Feb 2023	Rank Mar 2022	DBMS	Database Model	Score Mar 2023	Score Feb 2023	Score Mar 2022
1.	1.	1.	IninfluxDB	Time Series, Multi-model	29.15	-0.29	-0.54
2.	2.	2.	Kdb	Time Series, Multi-model	8.03	+0.73	-0.98
3.	3.	3.	Prometheus	Time Series	7.33	+0.27	+1.00
4.	4.	4.	Graphite	Time Series	6.59	-0.16	+1.11
5.	5.	5.	TimescaleDB	Time Series, Multi-model	4.52	+0.02	+0.06
6.	6.	↑ 7.	RRDtool	Time Series	3.02	+0.02	+0.41
7.	7.	↑ 9.	DolphinDB	Time Series, Multi-model	2.69	-0.16	+1.16
8.	8.	↓ 6.	Apache Druid	Multi-model	2.63	+0.14	-0.63
9.	↑ 10.	↑ 15.	TDengine	Time Series, Multi-model	2.43	+0.02	+1.67
10.	↓ 9.	↓ 8.	OpenTSDB	Time Series	2.32	-0.12	+0.53
11.	↑ 12.	↑ 12.	QuestDB	Time Series, Multi-model	1.89	-0.06	+0.73
12.	↓ 11.	↓ 11.	GridDB	Time Series, Multi-model	1.83	-0.13	+0.49
13.	13.	↓ 10.	Fauna	Multi-model	1.77	-0.12	+0.42
14.	↑ 15.	↑ 24.	M3DB	Time Series	1.02	-0.07	+0.77
15.	↑ 17.	↑ 18.	VictoriaMetrics	Time Series	1.01	+0.02	+0.39
16.	↓ 14.	↓ 13.	Amazon Timestream	Time Series	1.00	-0.15	-0.02
17.	↓ 16.	17.	KairosDB	Time Series	0.95	-0.11	+0.30
18.	↑ 20.	↓ 16.	eXtremeDB	Multi-model	0.83	+0.03	+0.13
19.	19.	↓ 14.	CrateDB	Multi-model	0.82	-0.09	-0.15
20.	↓ 18.	↑ 21.	Apache IoTDB	Time Series	0.70	-0.29	+0.30
21.	21.	↓ 19.	Raima Database Manager	Multi-model	0.59	+0.06	+0.02
22.	↑ 25.	↑ 27.	Heroic	Time Series	0.45	+0.07	+0.24
23.	↑ 24.	23.	Axibase	Time Series	0.38	-0.03	+0.12
24.	↑ 26.	↑ 34.	Machbase	Time Series	0.33	-0.03	+0.31
25.	↓ 22.	↑ 29.	ArcadeDB	Multi-model	0.31	-0.11	+0.22