

Time Series DB

Design a Time Series Database

Examples of time series values are data points spit out by monitoring and logging services. Periodic values, either single or key-value pairs. Usual queries are aggregate queries e.g. How many times in last 5 minutes did this request take more than 1s?

In theory, any database can be used to store time-series values. But one needs to think hard about how to design that schema. There are some OSS that are tuned just for time-series data e.g. OpenTSDB.

Reading pointers:

1. List of possible requirements of a TSDB

- <https://softwareengineering.stackexchange.com/questions/269590/how-to-efficiently-store-big-time-series-data>
- <https://www.xaprb.com/blog/2014/06/08/time-series-database-requirements/>

2. Optimizing a Relational Database for Time Series Data

- <https://www.youtube.com/watch?v=X4TfveHzBwM> (Short, 5 minute video, but lays the foundation well)

3. Using a columnar database as TSDB

- <https://medium.com/@hellomichibye/column-oriented-database-draft-part-2-21199a2de18a> (A little cursory, but sets the foundation well)

4. Cassandra for Time-series Data

- <https://academy.datastax.com/resources/getting-started-time-series-data-modeling>
- <http://www.datastax.com/dev/blog/advanced-time-series-with-cassandra>

5. Mongo Schema for a TSDB

- <https://www.mongodb.com/blog/post/schema-design-for-time-series-data-in-mongodb>

6. An analysis by an engineer who works in a company that uses a TSDB as their core product:

- <http://jmoiron.net/blog/thoughts-on-timeseries-databases>