Why We Need a Time-series Database

1. The Age of Big Data

With the development of the Internet and information industry, especially the Internet of Things (IOT) in recent years, "big data", which has existed for a long time in fields such as physics, biology and environmental ecology and industries such as military, finance and communication, has become more and more well-known and attracted widespread attention in the world.

* 1. Background and Impact

From 2012 on, the term "big data" has been mentioned more and more, which is usually used to describe and define the massive data generated in the era of information explosion, and to name the technological development and innovation related to it.

Data is rapidly expanding and becoming larger, which determines the future development of enterprises. Although many enterprises may not be aware of the potential problems caused by the explosive growth of data, as time goes by, people are more and more aware of the importance of data to enterprises. As a New York Times column argued in February 2012, the era of "big data" has arrived, in which decisions in business, economics and other fields will increasingly be based on data and analysis rather than experience and intuition.

"It's a revolution," said Gary King, a professor of sociology at Harvard University. "The sheer volume of data has led to a quantification process in every field, whether it's academia, business or government.

Now the society is a high-speed development of society, science and technology developed, information flow, and the communication between people more and more close, life is becoming more and more convenient, the application of large data more and reveal his advantage, it occupied area is becoming more and more big, e-commerce, O2O, logistics distribution, etc., all kinds of using big data field is assisting in the development of the enterprise constantly develop new business, Innovate operation mode. With the concept of big data, the judgment of consumer behavior, prediction of product sales, accurate marketing scope and replenishment of inventory have been comprehensively improved and optimized.

"Big data" in the Internet industry refers to the phenomenon that Internet companies generate and accumulate data on users' online behavior in their daily operations. The scale of the data is so vast that it cannot be measured in terms of G's or T's. Every day, more than 500 million images are uploaded around the world, and 20 hours of video are shared every minute. Even all the information people create every day that is all the communications, including voice calls, emails and messages, all the pictures, videos and music uploaded doesn't match the amount of digital information people create about themselves every day.

* 1. The Value of Big Data

As the large amount of data volumes and data types, especially with the wide application of the Internet of Things, information perception is everywhere and there is a large amount of information, the value density becomes relatively low. But in fact, it is very important and valuable. How to complete the value "purification" of data more quickly through powerful machine algorithms is an urgent problem to be solved in the era of big data.  
  
There is so much data that we can be no longer keen on accuracy, while previously there was very little data to analyze, so we had to quantify our records as accurately as possible, and as scale increased, the obsession with accuracy diminished; With big data, we no longer need to get to the bottom of a phenomenon, as long as we master the general direction of development, appropriately ignoring the accuracy at the micro level, we will have a better insight at the macro level.  
  
"What is the most expensive thing in the 21st century?" "Talent", I think so. But now, the era of big data has also brought with it the value of data that keeps doubling. That is, I think, the data is also the most expensive thing in the 21st century.   
  
The growth of data has never stopped, or even exploded, thanks to the rapid expansion of network bandwidth and the huge amount of data brought by various wearable devices.Within a minute, more than 100,000 new tweets were posted on Weibo and Twitter. The social network Facebook has more than 6 million views...  
  
What do these huge numbers mean? It means that a whole new way of getting rich may be on the horizon, worth as much as oil or gold.In fact, while we're still using twitter and other social media platforms as a way to express our feelings and opinions, Wall Street's money gurus are tapping into the Internet's "data wealth," using it to predict market trends, and to great effect.

1. The Disadvantages of Regular Database, that is, The Advantages of Time-Series Database

For the big data, actually we can use a non time-series database,such as relational database, and some people do. But with the data volume becoming huge, data ingestion rate becomes a challenge, query latency becomes a challenge, storage becomes a challenge, special data analysis is required, down sampling is required, interpolation is required, time weighted average is required and so on.

For the storage and processing of time-series big data, relational database is often used for processing, but because of the inherent disadvantage of relational database, it can not be efficient storage and data query. Time-series big data solutions can efficiently store and rapidly process massive time-series big data by using special storage methods, which is an important technology to solve massive data processing. This technology adopts special data storage mode, which greatly improves the processing capacity of time related data. Compared with relational database, its storage space is halved, and the query speed is greatly improved. Time series functions have superior query performance over relational databases.

There are huge differences between time-seriesl business and ordinary business in many aspects, which can be summarized as follows:

1. Consistently generating huge amounts of data without peaks and troughs. To take a simple example, a sentry monitoring system that monitors 100 metrics per second on 1W servers would generate 100 WATTS of TPS per second. For example, if 100 million people are wearing a popular sports wristband and each wristband only collects 3 metrcis (heartbeat, pulse and step count) per second, it will also generate 300W TPS per second.
2. Data is inserted, and there is no update or deletion. Based on the fact that data generated by time-series services is rarely updated or deleted, there can be considerable simplification in the design of time-series database architectures.
3. More attention will be paid to recent data, and streaming processing will be paid more attention in the future. Time-old data are rarely accessed or even discarded. It's easy to understand, sentry systems we tend to care most about the last hour, most about the last three days, very little about the last three days. With the advent of streaming computing, time-series data will inevitably pay more attention to the value of real-time data in the future development, and this part of data is undoubtedly the most valuable. It is a very common and important scenario that the data can be generated and the alarm can be generated according to certain rules. The more timely the alarm is, the better for the business.
4. Labels of multiple dimensions exist in data, which often requires multi-dimensional joint query and statistical query. Another very important function of time series data is multi-dimensional aggregated statistical query. For example, the business needs to calculate the click-through rate and total revenue of advertisements published by Google in USA in the last hour, which is a typical multi-dimensional query.

And based on the basic characteristics of time-series business, TSDB pay attention to the following technical points in particular: high throughput write capability, storage tiering /TTL, higher Compression rate, multi-dimensional query capability, high efficiency polymerization capacity, real-time anomaly detection and future prediction.

So to sum up, why do most respondents use time series databases instead of regular databases? Why is TSDB the fastest growing database today?

There are two reasons:  
(1) Scale. Time series data is accumulated very quickly. (For example, a connected car can collect 25 gigabytes of data per hour.) Conventional databases are not designed to handle data of this size, and relational databases are very poor at handling large data sets; The NoSQ database L handles scale data well, but not as well as a database fine-tuned for time series data. In contrast, time series databases (which can be based on relational or NoSQL databases) treat time as a first-class citizen, processing such large data volumes with increased efficiency and performance gains, including: Higher Ingest Rates, faster large-scale queries (although some support more queries than others), and better data compression.

(2) Availability. TSDB also typically includes some common capabilities and operations for analyzing time series data: data retention policies, continuous queries, flexible time aggregation, and so on. Even if you don't care about scale right now (for example, if you're just starting to collect data), these features can still provide a better user experience and make your life easier.

That’s all why we need a time-series Database.