**Why do we need a time-series database?**

Pan Youhe 2020141461065

Database management system is an important field in computer science, the most common of which is relational database. In recent years, time series database has become a new hot research direction. This is closely related to the development of big data, Internet of Things and other fields. This article attempts to analyze why we need to study and apply time series databases.

**1. Application areas of time series database**

Time series database has rich application scenarios, and it has begun to play its role in various fields. With the rapid development of information technology, computers have entered all walks of life, which has created new demands for database management systems. The following examples illustrate the application areas of time series databases.

(1) IoT devices in daily life. Various smart devices have gradually entered our daily life, bringing us great convenience. For example, smart meters can automatically record readings on a regular basis and transmit the data back to the company in time to remind users to pay their bills.

(2) Automotive field. Self-driving cars are one of the hottest developments in the automotive field right now. Various sensors on driverless cars need to continuously collect vehicle status and road condition information, and perform real-time calculations on these data to ensure driving safety. This has extremely high requirements on the speed of data processing.

(3) Financial field. Stock prices change rapidly, so time series databases are also needed to store and calculate these constantly generated data. Data in the financial sector is often of high value and requires timely analysis. This places a lot of demands on time series databases.

There are many other areas where there is an urgent need for time series databases. In order to meet these needs of the industry, it is necessary to conduct in-depth research on time series databases.

**2. Characteristics of time series data**

The various fields mentioned above generate large amounts of time series data. Time series data is a large amount of data that is continuously generated in chronological order. Time series data has many typical characteristics. The following takes a smart meter that automatically records data as an example to introduce the characteristics of time series data.

(1) Time stamped data. Time series data is continuously generated in chronological order, and each piece of data will store a timestamp to record the time when the data point was generated. The data generated by the same meter is also generally automatically arranged in timestamp order.

(2) Structured data. In addition to the time stamp, each data point also records the meter's number, location, current, voltage, and other information. That is to say, the data has a relatively fixed structure.

(3) Data source is like a stream. A data source for time series data continuously produces new data at short intervals and tries to write it to the database. Time series databases need to be able to cope with the speed of inserting data like this.

(4) Data rate is pretty stable. Data sources usually collect and record data at regular intervals, so the rate of data generation is relatively stable.

(5) Immutable. A data point stores the meter readings at a time. After these data are generated and stored, the content will generally not change. That is to say, time series data generally does not need to be modified.

(6) More write than read operations. The operation of writing the readings to the database is very frequent, and the reading of the data is performed at regular intervals, which is relatively infrequent. Therefore, the write operation of the time series database requires higher efficiency than the read operation.

(7) Data is rarely deleted or updated. Time series data is generally stored for a fixed period of time or a certain amount, and then automatically deleted or overwritten by new data. There is usually no need to manually delete data, nor is it generally necessary to manually modify data that has already been generated.

(8) There is always retention policy. After the data is generated, it is generally stored for a period of time for the purpose of calculation in the following period of time.

(9) Real-time data computing is desired. After the time series data is generated, it often needs to be calculated in real time. For example, the electricity bill is calculated based on the amount of electricity used, and it is judged whether the remaining electricity bill is sufficient. Therefore, the operation efficiency of time series database is very important.

(10) Query is always in time and space range. For example, to query the daily electricity consumption of an electric meter in the past month, it is necessary to read the data generated by the electric meter in the previous month. Therefore, the time series database needs to have a class of functions that can easily search for data by time and space.

Time series data has these unique characteristics, many of which are not found in traditional data. Therefore, it is difficult for ordinary relational databases to meet the needs of users, and it is necessary for us to design a special time series database for these characteristics.

**3. Disadvantages of traditional databases**

Traditional databases mainly include relational databases, NoSQL databases, real-time databases, and general big data platforms. Each of these databases has different characteristics and applicable fields. They each have some different advantages and disadvantages when it comes to working with time series data. On the whole, none of them can fully meet the needs of time series data related fields. For example, time-series data is usually generated very frequently, and the accumulated quantity is large, and the speed of data calculation is also very high. These characteristics make it difficult to directly apply various traditional database management systems in the field of processing time series data.

The volume of time series data is huge and growing at an exponential rate. Related industries are also developing rapidly, and the need for more efficient databases is more urgent. Therefore, we are motivated to design a specialized class of time series data to meet the growing demand in this area.

**4. The advantages of time series database**

Time series database is a database management system specially designed for processing time series data. At present, the most popular time series database products mainly include TDengine, Prometheus, InfluxDB, TimeScaleDB, OpenTSDB, etc. They have different characteristics in terms of data model, query language and so on. They are all purpose-designed time series databases with many advantages over various traditional databases.

**5. Summary**

Time series data has many typical characteristics. Time series data has more and more applications in various fields. Traditional databases are difficult to meet the needs of use, while time series databases have many unique advantages. All in all, time series databases have broad prospects for development. Therefore, we want to study and apply time series databases.