四川大学
SICHUAN UNIVERSITY

题　　目 Why do we need a time-series database?

学生姓名 张皓为

课　　程 The Design and Application of Time-series Database

学　　号 2020141461093

专　　业 软件工程

指导教师 Jeff Tao

二 〇 二 二 年 七 月 七 日

Abstract

With the advent of the new era, a variety of new technologies have developed vigorously.Under the development trend of 5G and the Internet of Things(LOT), the era of the Internet of everything has arrived. The high-speed communication brought by 5G technology brings more efficient operation of the data transmission speed of the Internet of Things. An important technology in the development of the Internet of Things is time-series database. A large amount of data generated by iot devices needs to be processed and stored through effective timing database, so the application of time-series database plays a very important role in our life.

Definition

Time-series data refers to the data of a series of recorded values arranged according to the size of the timestamp as time-series data. In daily life, temporal sequence data is ubiquitous. For example, the records of user visits to websites and the system log data of application systems are constantly changing according to time. Therefore, a series of data composed of user information and other attributes determined according to time is a group of temporal sequence data. In recent years, the application of time-series data more widely, including the field of Internet of things, economic and financial field, environmental monitoring, medical, industrial manufacturing, agricultural production field, system monitoring and so on various aspects of hardware and software, the extensive use of time-series data, the trend of the research object, regularity, abnormal sex, and with a wave of artificial intelligence in 5G. As can be seen from the above example of vehicle positioning, there is a big difference between time-series data and relational data:

1. Traditional relational databases focus on adding, deleting, modifying and checking and transaction functions, while sequential databases write massive data and read and query mostly data within a period of time. Traditional relational database are adopted B Tree,which reduces the intermediate process that goes through when positioning records, thereby speeding up access, is a random read/write mode, will find on consume more time, for more than 90% of the scene is written to the time-series database efficiency is too low, so the mainstream of time-series database is used the LSM Tree to replace B Tree, its core idea is to give up part of the reading ability for writing. Of course, in addition to the LSM Tree as a storage mechanism, there are many other attempts and improvements to the time-series database, and they are constantly evolving. In addition to storing queries, time-series database often need real-time analysis and calculation operations. For most Internet big data applications, offline analysis is more common. Even if real-time analysis exists, the requirements for real-time analysis are not high. For example, user portrait can be carried out after accumulating certain user behavior data. However, for Internet of Things applications, the real-time calculation of data is often very high, because real-time alarm is needed according to the calculation results to avoid the occurrence of accidents.

2. Time-series data does not care about the relationship. In car positioning, we do not need to know other attributes of the owner of the car, such as age and occupation, so there is no association with the table of the owner of the car. It has high cohesion and low coupling properties.

3. Time-series data are rarely updated, and the generation of measured values at a certain moment will not change, so there is almost no need to update time-series data. For relational data, existing data are often updated, such as students' personal information, including age, height and other attributes.

4. The data volume of time-series data continues to increase linearly. New data will be generated every certain period of time and massive data will continue to be generated, so the data volume is huge. However, the growth of relational data usually does not continue to grow over time, for example, the amount of student data in a school is relatively stable over a period of time. Take smart electricity meters as an example. A smart electricity meter collects data every 15 minutes and automatically generates 96 records every day. There are nearly 500 million smart electricity meters in China and nearly 50 billion records are generated every day. An internet-connected car collects data every 10 to 15 seconds and sends it to the cloud. A car can easily generate 1,000 records a day. If all 200 million cars in China were connected to the Internet, 200 billion records would be generated every day. Within five years, iot devices will generate more than 90% of the world's data.

Therefore, the sequential database has many advantages over the traditional relational database, and will be full of new vitality in the new era.

Application scenarios:

1. Intelligent emergency command and integrated communication scheduling for smart cities and energy industry

Smart city, smart factory, intelligent emergency command and integrated communication command and scheduling scheme is a visual command and scheduling scheme that integrates monitoring, command, scheduling, conference, communication and other functions by adopting digital BIM+GIS+NBIOT+AI+5G+ algorithm technology. In the emergency warning, reporting, response, command and other aspects of the realization of timely and effective visual command, to meet the emergency scene real-time image transmission and video consultation requirements for rapid response.

2. Emergency command and integrated communication command and scheduling scheme

There are many application scenarios in daily inspection, hidden trouble reporting, 3D map and fusion scheduling of various parks. Equipment management operation status, HSE risk level, process flow, process control operation parameters and other maintenance conditions of all kinds of business site and management real-time data and information

intuitive display, timely find problems, analyze causes, put forward rectification suggestions, and implement.

3. Intelligent inspection and security application scenarios

Comprehensive monitoring of combustible gas, smoke and electrical fire; Full link linkage of fire sensing, video monitoring and fire water; The whole process of fire, alarm, evacuation and fire fighting is covered.

4. Intelligent operation and maintenance of energy industry equipment

Iot platform can also be applied to unified management, operation and maintenance of massive equipment terminals, online monitoring and diagnosis of equipment status, and timely failure warning. It can also display operation and maintenance data through multidimensional charts.

Future

In the future, with the increasing amount of data, we need sequential database to process our daily data.

1, non-relational databases will overtake relational databases, and NoSQL will overtake SQL. Since most of the newly generated data will be non-relational data from the perspectives of the way of social construction, the language expression system delivered, the way of video collection and the massive generation of sequential data of industrial production IoT, etc. And IDC predicts that by 2023, the world will generate more non-relational data than relational data.

2, the data will be more on the edge; Especially at the edge of consumption. In the future, smart wearable devices, smart hardware, smart devices in factories, Iot devices, automotive batteries and sensors will be the main sources of future data. These data will be collected at the edge, transmitted to the central, integration, storage, computing, monitoring in the central, which will be the future development of a large space, is also the focus of the sequence database for the scenario.

3, here's a saying that 'Al for DB,DB for Al'. DB for Al means that about 60%-70% of the scenarios supported by the database will be ai-related in the future, and the database needs to better support Al's training, reasoning, rapid data iteration and response.

In the future, time-series database will also develop the following five advantages:

1. Both insert and query performance is much higher.

2. Because of its superior performance, it requires less than 1/5 of the computing resources of other software.

3. Use column storage and adopt different compression algorithms for different data types, requiring less than 1/10 of the storage resources of other software.

4. No separate database, separate table, no real-time data and historical data, zero management cost.

5. Using standard SQL syntax, the application can insert or query data through standard JDBC, ODBC interface, learning cost is almost zero.

The purpose of the Internet of Everything is to be more intelligent. In the future, we can give memory to some devices that have no memory ability through super-fusion of time-series database, and derive more possibilities for intelligence, leaving minimalism to users. In conclusion, those are why we need times-series database.