Why do we need a time-series database?

Sequential data is a series of data generated over time, which is simply time-stamped data. A Time Series Database (TSDB) is a Database optimized for ingesting, processing, and storing timestamp data. Such data may include metrics from servers and applications, readings from sensors in the Internet of Things, user interactions on websites or applications, or transaction activity in financial markets.

The main data attributes of time series data are as follows:

Each data point contains timestamps for indexing, aggregation, and sampling. The data can also be multidimensional and correlated;If the write level is high and the read level is low, the high frequency write level at the second, millisecond or even nanosecond level must be supported. Queries are usually multi-dimensional aggregated queries, which require high latency. Summary views of data (for example, down-sampled or aggregated views, trend-lines) may provide more insight than individual data points. For example, given network unreliability or abnormal sensor readings, we might set alerts when a certain average exceeds a threshold over a period of time, rather than at a single data point; Analyzing data usually requires accessing it over a period of time (for example, give me click through data from the past week);

While other databases can handle time series data to some extent at smaller data sizes, TSDB can handle data ingestion, compression, and aggregation over time more effectively. Take the Internet of Vehicles scenario as an example, 20000 cars, each car 60 indicators, assuming that the collection is once per second, then 20000 \* 60 = 1200000 index value will be reported per second, namely 120W data index value per second, each index value is 16 bytes (assuming only including 8-byte timestamp and 8-byte floating point number), This generates about 64 gigabytes of data per hour. In fact, each index value comes with additional data such as labels, which actually requires more storage space.

In short, the sequential database is a database specially designed to store and process time series data, supporting efficient read and write of sequential data, highly compressed storage, interpolation and aggregation.

But why do we need to use time-series database? What are the benefits of a time-series database?

The first is the increase in data writing efficiency, which can be proved by the example of Ali Cloud. Ali Cloud released the timing database TSDB, which improved the data writing efficiency by a hundredfold. Ali Cloud time series database TSDB. Users can complete data query and analysis without developing codes, which can help enterprises mine the value of time series data from any dimension.

Under normal circumstances, there are many monitoring points of timing business data, the concurrency exceeds one million, and the real-time requirement is high, so the difficulty of multi-dimensional attribute combination analysis is greatly increased. If the traditional relational database is used to store the monitoring indicator data, the writing and analysis efficiency is low and the storage cost is high. When using open source Open TSDB to build time series cluster, not only the operation and maintenance cost is high, but also the stability is difficult to guarantee.

TSDB optimizes the storage structure of time series data and reduces the size of single record data through bulk memory compression. Write efficiency is more than 100 times higher than relational databases and storage costs are reduced by 90%. At the same time, TSDB has time sequence insight capability and can realize interaction. Visual data analysis helps enterprises grasp the data change process in real time, discover data anomalies, and improve production efficiency.

According to the actual pressure measurement, TSDB read efficiency is one order of magnitude higher than open source Open TSDB and Influx DB. In addition, TSDB also provides professional integrated time series data computing capabilities, supporting down-sampling, data interpolation and space. Converged computing can meet a variety of complex business data query scenarios and can aggregate and analyze millions of data points in seconds.

Ai Leqiang, senior product manager of Ali Cloud Time series database, said, "The time series database is responsible for the storage and analysis services of the most valuable data assets of the Internet of Things. In the future, it will certainly play a huge role in smart cities, smart transportation, smart hotels and smart agriculture. Infrastructure for the future of everything.

Taking the Vehicle Internet as an example, TSDB time insight can quickly and accurately obtain each vehicle's mileage, driving speed, power supply, engine speed and other indicators in real time, and track the vehicle's operating conditions and operating trends within the time zone.

Of course, there are many other advantages of timing database, here is another example of shenzhen Institute of Digital Research - Polar Star timing database .

Shenzhen Institute of Digital Research - Star time series database is a domestic database with completely independent intellectual property rights. The product has a number of time sequence data processing patent technologies, and has passed the test and verification of China Electric Power Research Institute, Guangdong Electric Power Research Institute, China Saibao Laboratory and other authoritative institutions, with excellent performance, leading in China. In addition, SHENZHEN Institute of Digital Research - Gestar timing database has three unique advantages:

First, strong performance

Based on microkernel technology, a single server node can meet the capacity of more than 20 million label points. The second generation of hybrid network processing engine, read and write performance is more than 6 million records per second; Data redundancy and database cluster technology ensure the reliability and expansibility of products.

Second, cloud side full coverage

There are different application modes for different scenarios of shenzhen Institute of Digital Research - Polar Star timing database. At present, there are mainly three types: cloud application, suite application and embedded application. Thus, the full coverage of multiple products on the cloud side is realized, which can effectively reduce the development cost and complexity of enterprises and achieve seamless data synchronization between different databases.

Third, Wide application scenarios

Shenzhen Institute of Digital Research - Star time series database has a wide application space in three scenarios of enterprise production, big data center and Internet of Things. Enterprise production scenario: it has broad application prospects in production process data management, factory-level monitoring system, intelligent real-time information interactive platform and other fields in various industries; Data center scenario: As the core product of enterprise big data center, Jstar timing database is used in centralized EMS/SCADA, DMS, metering automation, online monitoring and other automation scenarios. Internet of Things monitoring scenario: It can be widely used in Internet of Things (IoT) equipment monitoring system, enterprise energy management system (EMS), production safety monitoring system, equipment detection system and other scenarios.

Number of shenzhen institute - pole star performed very well in the domestic powerful performance of temporal database, by the market, at present has been southern power grid, the three gorges power plant the world top 500 enterprises such as the recognition and trust, the product can be applied to electric power, petrochemical, transportation and logistics industry, environment protection and so on more than 10 a sequence of data fusion, storage, analysis and release.

So what are the applications of the time-series database?

Internet of Things timing data is a digital record of physical quantities of industrial equipment, which is widely used in rail transit, energy management and control, intelligent manufacturing and other fields.

Rail transit

In the field of rail transit, by installing sensors on Bridges and collecting and analyzing sensor data, the state of Bridges, such as deflection, strain, vibration and bearing displacement, can be detected, so that targeted health detection and reinforcement can be carried out to avoid bridge collapse and ensure the safety of people's travel. IoTDB already has user cases in this area.

Energy control

Some open source users of IoTDB monitor energy consumption, alarm management and forecast optimization of chemical plants, realize energy saving, emission reduction and efficiency enhancement, and respond to the policy of "carbon peak and carbon neutrality".

Energy consumption monitoring: by monitoring the real-time water flow of the factory, we can find problems such as running, leaking, dripping, leakage and hydraulic balance in the water supply system. If the water consumption continues to increase after the factory shuts down, we can conduct timely investigation.

Alarm management: according to the value of the measurement point for abnormal judgment, timely sound, light, telephone alarm, reduce the impact of energy system failure.

Prediction optimization: establish energy prediction model, optimize production schedule, achieve the goal of reducing pollution.

Intelligent manufacturing

Taking the application of IoTDB in a cigarette factory as an example, the cigarette factory can improve production efficiency and upgrade the manufacturing industry by monitoring the equipment status and controlling the statistical process.

Equipment operation monitoring: collect equipment key data, check equipment production process key data (number of mouthsticks, cigarette paper, water paper) at a certain time.

Silk quality evaluation: judge the quality of tobacco according to the real-time monitoring data (whole rate of tobacco, broken rate of tobacco, etc.).

Statistical process control (SPC) : according to the leaf moisture content, discharge temperature and other process quality data, distinguish the random fluctuation, abnormal fluctuation and development trend of product quality, according to the trend to determine whether to add or reduce materials, to ensure the stability of the production process.

The above applications are all equipped with a large number of sensors on a variety of equipment, and how to add, each industrial enterprise needs to determine the business characteristics. While designing data collection, we also need to design how to store the data. Generally, the more data collected and the finer the granularity, the better. A high-end device, such as an airplane, has 80,000 sensors; An excavator will be equipped with four or five hundred sensors, but the number of excavators is large, a user has 20,000 excavators, there are tens of millions of time series data; 20,000 wind turbines, generating 120 petabytes of data per year. The scale of sequential data in the Internet of Things is enormous.

By the end of this article, you should have a better understanding of the time-series database, its benefits, and how it can be used.  Better understand why we use the time-series database.