**The introduction about TDengine**

In the era of big data, massive amounts of real-time data are sent to the cloud. All the data generated by machines, devices, sensors and trading systems are temporal and often contain location information.  In addition, due to the large number of data records, real-time data writing becomes a bottleneck and query analysis is extremely slow, which has become a new technical challenge, so TDengine came into being in this situation.

TDengine is the creation of a Taosi team led by Tao Jianhui, founder of Beijing Taosi Data Technology Co., LTD. In 2017, Taosi started to develop TDengine;  In 2018, TDengine launched its first product;  In 2019, Taosi data officially announced that TDengine kernel and community edition will be 100% open source;  In 2020, Taosi data received A pre-A round investment of nearly ten million DOLLARS from GGV and other companies to develop TDengine, an open source big data platform for the Internet of Things.

TDengine is a high-performance, distributed, SQL-enabled sequential database with its core code, including clustering, all open source.  TDengine does not rely on any open source or third party software, with completely independent intellectual property rights, with high performance, high reliability, scalability, zero management, easy to learn and other features, providing caching, data subscription, streaming computing and other functions, minimize the complexity of research and development and operation.

TDengine syntax uses a subset of standard SQL syntax, the native interface is C/C++ interface, while supporting ODBC,JDBC,Python,Go, C# and other development interfaces.  Clients need native dynamic link libraries to interact with servers. Dynamic link libraries currently support only Windows and Linux operating systems.

TDengine has all the features needed to process iot data, including :1.  Sql-like query language to insert or query data;  2. Support C/C++,JAVA(JDBC),Python,Go,RESTFUL,and Node.js and other development interfaces;  3. Various Ad Hoc query analysis can be done by TDengine Shell or Python/R/Matlab;  4. Support streaming computing based on sliding window through continuous query;  5. Super table is introduced to make data aggregation between devices simple and flexible through labels.  6. Embedded message queue, reference can subscribe to the latest data;  7. Built-in cache mechanism, the latest status or records of each device can be quickly obtained;  8. There is no distinction between historical data and real-time data, which is transparent and identical to the application;  9. The installation package is only 1.5m, and it only takes a few seconds from download to successful operation.  At the same time, after studying the characteristics of the Internet of Things,TDengine makes two technological innovations :1. The data model of "one table for one device" can greatly improve the efficiency of data insertion and query for a single device.2.  Each table is labeled with static label, and the static label data is stored separately from the collected dynamic data to solve the problem of multi-table aggregation query.

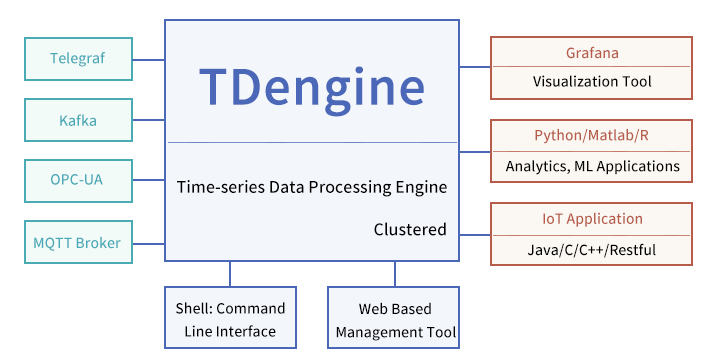
TDengine makes full use of the characteristics of sequential data, such as structuralization, no transaction, little deletion or update, more write and less read, and designs a new storage engine and calculation engine for sequential data. Therefore, compared with other sequential databases, TDengine has the following advantages :1.  More than 10 times performance improvement: defined innovative data storage structure, single core can process at least 20,000 requests per second, insert millions of data points, read more than 10 million data points, more than 10 times faster than the existing universal database;  2. The cost of hardware or cloud service is reduced to 1/5: Due to super performance, the computing resources are less than 1/5 of the general big data scheme, and the storage occupies less than 1/10 of the general database through column storage and advanced compression algorithm;  3. Full-stack real-time data processing engine (All in One): it integrates database, message queue, cache, streaming computing and other functions, eliminating the need to integrate HDFS and other software, greatly reducing the complexity cost of application development and maintenance, and seamlessly connecting with third-party tools;  Strong analytical skills;  4. High availability and horizontal scaling: With distributed architecture and consistent algorithms,TDengine ensures high availability and horizontal scaling and supports mission-critical applications through multiple replication and clustering features;  5. Zero operation and maintenance cost, zero learning cost: easy and quick cluster installation, no need to separate databases and tables, real-time backup;  6. Core open source: In addition to some auxiliary functions,TDengine's core open source, enterprises will no longer be bound by the database, which makes the ecology more powerful, product more stable, developer community more active.  Interactive Console: Through the command line Console, you can perform AD hoc queries, database operations, management, and cluster maintenance by executing SQL statements without programming.

And TDengine has two other highlights: data subscription and asynchronous insertion.  Data subscription is a subscription feature provided by TDengine.  An application can subscribe to the contents of a table or supertable in the database and be notified immediately of changes in the table's records.  The same table can be subscribed by multiple applications.  Asynchronous insert: To avoid the performance degradation caused by network latency and to improve data insert speed,TDengine also provides a set of apis for applications to insert data asynchronously.  With asynchronous apis, performance can be dramatically improved.

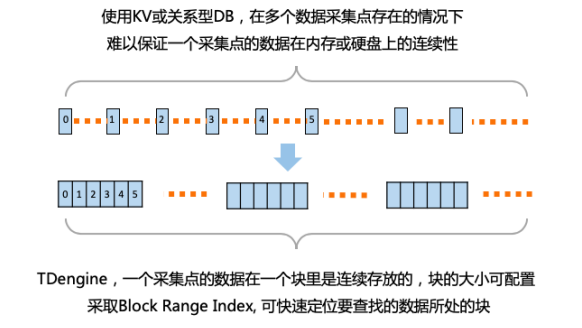
Through the functions of TDengine, we can analyze some of its features :1.  It is specially designed for Internet of Things data, and realizes the function that each collection point corresponds to one table by taking advantage of the timing characteristics of Internet of Things data.  But not for general iot data;  Supports high availability, and divides each physical node into multiple virtual data nodes and virtual management nodes.  In terms of storage structure, each collection point is used to create an independent table for storage.  6. The installation package is very small and easy to install and use.

After understanding the characteristics and functions of TDengine, we can know that TDengine is a basic software with a wide range of applications. In principle, all the places where machines, equipment and sensors are used to collect data can be used.  Some typical scenarios are as follows: public security, electric power industry, communication industry, financial industry, travel tools, transportation industry, oil industry, Internet, logistics industry, environmental monitoring, Internet of things, military industry, manufacturing industry.  In different scenes, use their own characteristics and functions to complete the corresponding functions.  However, because TDengine makes full use of the characteristics of timing big data, it cannot be used to process general data such as web crawler, Microblog, wechat, e-commerce,ERP,CRM and so on.

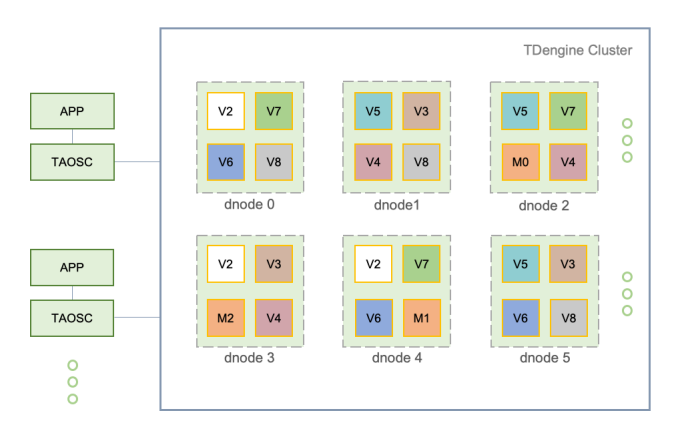
TDengine comes in three versions: Community, Enterprise, and Cloud Services.  Community edition is the open source version, under the AGPL license, is a small to medium scale Internet of Things data platform;  Enterprise edition is a carrier level distributed edition, it has ultra-high reliability, super horizontal scalability, and to deal with the challenges of big data, in addition to all the features of the community edition, there are additional features;  Cloud Service edition is the enterprise edition of TDengine running on AWS and Ali Cloud, with the characteristics of elastic scaling and zero management, through professional technical service team, to provide carrier-level Internet of Things big data platform services.  Each of the three versions has its own benefits, targeted at different individuals or groups.

With the basics of TDengine behind us, let's take a look at TDengine in the technology ecosystem.  TDengine plays the role of "hub" in the whole sequential big data platform. Opc-ua,MQTT,Telegraf, and also Kafka are located in various data collection and message queues, and they continuously write data to TDengine, Grafana,Pytho  N /Matlab/R,IOT Application and other visualization,BI tools, configuration software, applications to carry out the corresponding processing.  These relationships are shown below: 

Then, let's take a quick look at TDengine's two constructs: storage and distributed.  TDengine storage structure: In memory, In order to reduce memory overhead, effective processing time out-of-order problem,TDengine uses line storage, SkipList index, and memory according to First In First  In order to make full use of the characteristics of sequential data,TDengine adopts column storage for persistent storage, and the data of each table is stored consecutively to maximize the compression rate and data reading speed. In addition, each data block has predictive calculation to further improve the speed of data analysis.



TDengine distributed architecture :TDengien is based on the hardware, software system is not reliable, there will be failures, is based on the assumption that any single computer is not capable of handling large amounts of data design, so TDengine from the first day of development, is designed in accordance with the distributed high reliability architecture, is  Completely decentralized.  The overall system structure of TDengine is shown in the figure below:



There are three types of nodes in the distributed structure, namely, data node, virtual data node and virtual management node.  It handles interface interactions between applications and clusters and is embedded in JDBC, or C,Python, or Go connection libraries.  Applications interact with the entire cluster through TAOSC rather than directly connecting data nodes in the cluster.

That's the basics of Tdengine. If you're interested in other aspects of Tdengine, you can query it.