# 《The Design and Application of Time-series Database》

# Course Report



**Subject name：Why do we need a time-series database?**

Name of the project leader：王天扬Wang Tianyang

Student ID: 2020141461066

Teacher Advisor： Tao Jianhui

Grade Assessment：

Comments：

2022 年 7月 7日

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

Software Engineer Major

**Student** 王天扬Wang Tianyang **Teacher Advisor** Tao Jianhui

In the International Curriculum Week of this semester, I participated in the course on the Design and Application of the Time Series Database. As the name suggests, the time series database is mainly used to process time labels (change in time order, namely time serialized data, data with time labels is also called time series data).

Time-series data is mainly collected and generated by real-time monitoring, inspection and analysis equipment in electric power industry, chemical industry and other types. The typical characteristics of these industrial data is: fast generation frequency, that is, each monitoring point can produce multiple data within one second; seriously depends on the collection time, that is, each piece of data requires the corresponding unique time), measuring points more information (conventional real-time monitoring system has thousands of monitoring points, monitoring points produce data every second, generating dozens of GB of data every day.

Time series database with time as the main query latitude, usually continuous multiple sequence data drawn into lines, making time-based multi-latitude report, used to reveal the trend behind the data, rules, abnormal, real-time online prediction and early warning, timing data is common in IT infrastructure, operational monitoring system and the Internet of things.With the advent of the Internet of Things era, the data volume of time-series data is a blowout explosion, and the optimal storage for this data segmentation is becoming more and more important.Initially, universal storage systems were used to store timing data, such as MySQL.The first generation of timing platforms, such as KDB +, RRDtool, Graphite and so on, they were launched 20 years ago, mainly used to store and analyze the timing data of data centers, as well as high-frequency financial data, stock volatility, etc.According to the information published by database trend tracking and industry analysis websites, sequential database is the fastest growing part of the database market share.The reason is clear that computer virtual worlds, such as databases, networks, containers, systems, applications, and physical worlds, such as home equipment, urban infrastructure, factory machines, power facilities, etc., are creating huge amounts of timing data.Now more enterprises will use temporal storage and data analysis to obtain prediction-time-to-time decision-making capabilities, so as to provide customers with a better use experience.This means that the underlying data platform needs to develop to meet the challenges of new workloads, as well as more data points, data sources, monitoring dimensions, control strategies, and more accurate real-time response, with higher requirements for the next-generation timing center.

We know that the traditional database usually record the current value of the data, time series database records all the historical data, when processing the current timing data and to constantly receive new timing data, at the same time the timing data query is always based on time query conditions, and focus on solving the problem of the following massive data scenario:

1. Designed for timing storage and high performance reading and writing: computer virtual world systems and applications, as well as the physical world of IoT devices are creating huge amounts of timing data, millions of data throughput per second is very common, and the data also need to be received in a blocking way and can be compressed to save limited storage resources.How to support the writing of ten million level / second data.How to support the aggregation and query of ten million level / second data.
2. Designed for real-time operation: prediction ability and real-time decision-making ability. After receiving the data, you can output the latest data analysis results in real time and perform predefinition operations.
3. Designed for high availability: modern software systems need to be available in 24 hours , in addition to the basic clustering capabilities, but also need to automatically expand and shrink according to the requirements, and support flexible availability.
4. Cost-sensitive: massive data storage brings about a cost problem. How to store these data at a lower cost is a key problem to be solved by sequential databases.

Therefore, on many problems, we really need the existence of the time series database.Several of the above problems are seen everywhere in our daily life, such as autonomous driving, smart home, transportation networks and open data for some enterprises, which have a lot in common.Similarly, these applications also predict that our world is changing at a curved speed, and that we capture and parse the data more quickly and faster than ever before.They rely on a form of data that measures how things change over time, where time is not just a metric, but a main coordinate axis of coordinates, namely a time series database.According to the previous introduction, you will find that the timing data is more convenient and fast, and has a lot of powerful functions, and the time series data is so powerful, because every change of the system is recorded as a new line, which can go to measure the change: analyze the changes in the past, monitor the change now, and predict how the future will change.Therefore, we define time-series data: data that uniformly represents a system, process, or behavior over time.This is not just an academic difference: with the definition of "change" around us, " we can begin to find out the time-series datasets that we should collect but do not collect at the moment.

In today's era, the industrial Internet integrates advanced technologies such as big data, the Internet of Things and cloud computing, making it possible to generate massive data storage and analysis generated by industrial equipment such as sensors and PLC. Time sequence database is an important tool for industrial data storage and analysis.Generally speaking, more than 70% of industrial data is time sequence data. Each collection point of industrial data can produce multiple pieces of data per second, and each data should have a time attribute. Moreover, the time-driven time sequence data processing system has no peak and trough of service, which requires high reliability and processing speed of the system.Therefore, a very obvious characteristic of industrial data is related to time. The storage pressure brought by this "data expansion" has become one of the pain points of industrial Internet and digital development.At present, people can deal with the rapid growth of timing data by either using open source free time series database, or using foreign commercial time series database, or relational database, but there are various problems.The time series database with independent intellectual property rights has the advantages of high reliability, high performance, scalable and exploitable, and can seamlessly and organically integrate with other components of the industrial Internet platform, and effectively promote the development of the industrial Internet ecology. Therefore, the time series database is an indispensable part of the industrial Internet platform.

Time series database has many excellent and powerful functions, which provides high concurrency, smooth, continuous, real-time writing of data, and supports efficient writing and query.There is also UTSDB-InfluxDB time series database with high-performance data storage, TSM engine supports efficient data writing and data compression; reduces storage space by 80%, accelerate data writing, reduce storage space cost, and provides professional and comprehensive timing data calculation function, support data multidimensional query analysis (Downsampling，interpolation, multiple aggregation calculation); continuous query automatically calculates aggregated data to improve the efficiency of frequent queries; retention policy effectively and automatically expire obsolete data.This powerful application will be charged based on the actual usage, effectively saving storage costs.It is not only powerful, but also easy to use, and suitable for a variety of enterprises and software today, which has been closely related to our life. Without the time series database, the production monitoring of many industrial platforms will become complex and difficult, and even the Internet will be greatly affected. In the traditional manufacturing industry, the production safety monitoring needs to realize equipment self-diagnosis, self-adjustment and self-adaptation through intelligent product intelligence, so as to reduce equipment operation and maintenance costs and improve equipment utilization rate.UTSDB-InfluxDB can help manufacturers create new business models, transform products into services, data into value, provide a variety of query methods, can view data trends and average values from the overall situation, from the multidimensional data analysis, cost budget, resource planning, etc. In addition, UTSDB-InfluxDB can also combine with UCloud-IoT platform, through industrial equipment real-time data collection points, cloud platform time sequence database storage, massive data analysis and processing, to realize the prediction and maintenance of manufacturing production equipment, to improve the production system, improve production efficiency and quality, and promote the implementation of intelligent manufacturing. Thus it can be seen that this time series database is indispensable in our daily life.

Time series database is popular not only because of his excellent performance and strong performance, but also because it will upgrade in the future, become more excellent and powerful, people use multi-level data compression technology and distributed technology, and has a rich suite, make the database become reliable and have high performance, and can constantly develop and expand. Now realize customer demand is the main development and important goal, and with the existence of the time series database, through the real-time data collection and application, can do support enterprise building plant monitoring system, manufacturing execution system, energy management system, etc., and will be widely used in power, petrochemical, metallurgy, coal mine, equipment, transportation, environment, etc.,in the final，the time series database will become the core of the big data of the wisdom factory and the wisdom city.