

Introduction of Machbase

Machbase is a new generation columnar time series database that ingests and stores large amounts of log or time stamped data. Machbase is designed for collecting, in real time, "machine or sensor data" from various IoT environments while simultaneously allowing the data to be queried and analyzed using standard SQL and APIs at very high speeds. Machbase supports any IoT data architecture, and works extremely well on limited edge compute devices, or gateway/fog platforms, as well as cloud/cluster implementations.

Machbase solves the problem of storing and processing massive amounts of sensor data that could not be solved by existing current big data solutions, while safeguarding the future emergence of new sensor data requirements through various capabilities and functions.

The Emergence of New Data

Index

- The Emergence of New Data
- Types of Sensor Data
- Sensor Data Structure in PLC

Through the recent development of applying big data and real-time analytics to the internet of Things (IoT) domain and with the promise of Machine Learning (ML), various and vast amounts of data are being captured and accumulated. With the future of 5G systems entering the market, machine to machine data traffic will continue to explode.

Therefore, by analyzing these data in real-time, new applications, better efficiencies, greater reliability and fewer failures will be achieved. In particular, the number of source devices, such as edge devices or sensors, have dramatically increased in the last few years, and as a result the generated data from these devices have also increased exponentially. However, traditional, and even big data processing software are not providing adequate solutions to meet the volume and performance needs. The reasons why the current, conventional solutions are not suitable for the current data processing are summarized as follows:

First, the rate of data generation is increasing exponentially.

As the number of data sources grows and the types of information that need to be processed increases, the rate at which the server can receive and store the data is entirely different from before. There is no software solution optimized for storing tens of thousands to hundreds of thousands of data per second aside from saving it as a plain text file in the file system.


Second, the demand for real - time data analysis is increasing in proportion to the rate of data generation.

Using big data technics to help decision making is very important; however, existing solutions are not technically advanced enough to index hundreds of thousands of data per second and deliver results to the user, through search and analysis, in real-time.

Finally, as previously mentioned, the characteristics of "sensor data" is completely different.

Conventional databases have architectures that are inadequate for processing sensor-type machine data, and to address this, a new technological approach is needed. For this reason, Machbase has been newly developed with the best architecture to handle new "sensor-type machine data" and is the only database solution that can store, process, and analyze real-time data.

Types of Sensor Data



The figure above shows typical sensor data divided into three types. The properties and characteristics of each types are as follows.

ID

This value represents a symbol and a number indicating the uniqueness of the device (source code) where the corresponding machine data is generated.

It is the serial number of the machine or sensor and is represented as a 32-bit or 64-bit integer.

TIME

This value represents the time when the corresponding machine data occurred. This time stamp has the tendency to increase continuously and is usually in units of a second but can also be as fine as in nanoseconds.

DATA

This value is binary data, mainly in the form of integer type, real number type, or IP address value. Typically, this domain includes numeric values such as temperature, acceleration, and brightness from a specific sensor, or fixed data of 4-byte or 16-byte similar to IPv4 or IPv6. In certain applications the value can be in binary format as from camera or audio devices.

By providing a Tag Table, Machbase offers optimized architecture that can receive hundreds of thousands of sensor data per second.

Sensor Data Structure in PLC

In actuality, sensor data used in Programmable Logic Control (PLC) is not always represented by the above three types.

In most cases, a large number of sensor data are stored in a collected form as shown below. Therefore, the following types of data can be accepted together and should be easily converted.

data from PLC										
Time	SN01M	SN02M	SN03M	SN04M	SN05M	SN06M	SN07M	SN08M	SN09M	SN10M
04:01:56.005	11.1	1	0	0	0	1	0	0	0	0
04:01:56.057	11.3	1	0	0	0	1	0	0	0	1
04:01:56.109	11.1	1	0	1	0	1	0	1	1	0
04:01:56.161	12.3	1	0	0	0	1	0	0	0	1
04:01:56.213	9.1	1	1	0	0	1	0	0	0	0
04:01:56.266	0.9	1	0	0	1	1	0	0	0	0
04:01:56.319	8.9	1	0	1	0	1	0	1	0	1
04:01:56.370	1.3	1	0	0	0	1	0	0	0	0
04:01:56.422	33.1	1	0	0	0	1	0	0	0	0
04:01:56.474	3.3	1	0	0	0	1	0	0	0	0
04:01:56.526	5.6	1	0	0	1	1	0	0	0	1
04:01:56.578	5.5	1	0	0	0	1	0	0	1	0
04:01:56.630	4.5	1	0	0	0	1	0	0	0	0
04:01:56.682	5.3	1	0	0	0	1	0	0	0	0
04:01:56.733	1.2	1	0	0	0	1	0	0	0	1
04:01:56.785	3.4	1	0	0	1	1	0	0	0	0
04:01:56.838	11.3	1	0	1	0	1	0	1	0	0
04:01:56.890	11.2	1	0	0	0	1	0	0	0	0
04:01:56.942	9.9	1	0	0	0	1	0	0	0	1
04:01:56.994	8.4	1	0	0	0	1	0	0	0	0
04:01:57.046	8.4	1	0	1	0	1	0	0	0	0
04:01:57.097	1.2	1	0	0	1	1	0	1	1	1
04:01:57.149	1.3	1	0	0	0	1	0	0	0	0
04:01:57.200	11.2	1	0	0	0	1	0	0	0	0
04:01:57.252	3.1	1	0	0	0	1	0	0	0	0

Machbase provides a Log Table, which also provides a structure to store the above PLC data.

This table also provides the ability to enter and analyze tens of thousands of data per second.