A scalable database model for multiparametric time series: a volcano observatory case study

Placido Montalto, Marco Aliotta, Carmelo Cassisi, Michele Prestifilippo, and Andrea Cannata
INGV, Osservatorio Etneo, Italy

The variables collected by a sensor network constitute a heterogeneous data source that needs to be properly organized in order to be used in research and geophysical monitoring. With the time series term we refer to a set of observations of a given phenomenon acquired sequentially in time. When the time intervals are equally spaced one speaks of period or sampling frequency.

Our work describes in detail a possible methodology for storage and management of time series using a specific data structure. We designed a framework, hereinafter called TSDSystem (Time Series Database System), in order to acquire time series from different data sources and standardize them within a relational database. The operation of standardization provides the ability to perform operations, such as query and visualization, of many measures synchronizing them using a common time scale.

The proposed architecture follows a multiple layer paradigm (Loaders layer, Database layer and Business Logic layer). Each layer is specialized in performing particular operations for the reorganization and archiving of data from different sources such as ASCII, Excel, ODBC (Open DataBase Connectivity), file accessible from the Internet (web pages, XML).

In particular, the loader layer performs a security check of the working status of each running software through an heartbeat system, in order to automate the discovery of acquisition issues and other warning conditions. Although our system has to manage huge amounts of data, performance is guaranteed by using a smart partitioning table strategy, that keeps balanced the percentage of data stored in each database table. TSDSystem also contains modules for the visualization of acquired data, that provide the possibility to query different time series on a specified time range, or follow the realtime signal acquisition, according to a data access policy from the users.