Data Storage Optimization for Energy Management in Intelligent Buildings

Samuel Otero Paz[1], Jose A. Taboada González[1], Jose R.R. Viqueira[1], Juan E. Arias Rodríguez[2]

[1] COGRADE – Centro de Investigación en Tecnoloxías da Información,
[2] Systems Laboratory – Instituto de Investigacións Tecnolóxicas,
Universidade de Santiago de Compostela,
Santiago de Compostela, Spain
{samuel.otero, joseangel.taboada, jrr.viqueira, juan.arias}@usc.es

Abstract. The main objective of the LIFE-OPERE project is the improvement of energy management in the facilities of the University of Santiago de Compostela (USC). Data analytics are applied to sensor data to check the impact of new policies in the energetic behavior of the facilities. To achieve this, system for the acquisition, management and publication of building sensor data was designed and implemented. The contribution of this paper is focused on the optimization of the data model that enables overcoming the main problems and facing the main challenges that arose during the project. Besides, the applicability of big data technologies is also studied.

Keywords: sensor, automation, database, intelligent building, big data.

1 Introduction

Nowadays, the energy consumption in buildings reaches the 40% of the whole energy consumption in the EU [1]. Therefore, buildings are key to achieve the reduction of strong greenhouse gases, mainly by the improving of energetic efficiency.

The LIFE OPERE (2013 – 2016) project, whose main partner is the University of Santiago de Compostela (USC), aims at improving the energy management on an intelligent building with large consumption. The solutions proposed in the project are compared with other intelligent buildings of the university to prove its efficient. To fulfill the building sensor data requirements of the project, a system for data acquisition, management and publication was designed and implemented. An important part of the system is a data model that enables the representation of sensor data from a wide variety of heterogeneous sources (dataloggers, meteorological agencies, etc.).

This paper describes the process of optimization of the data storage structures of the aforementioned system that enabled facing the main challenges that arose during its construction. In particular, such an optimization enabled reducing the response time to the queries required by the monitoring and analytical processes. Real time On Line Transaction Processing (OLTP) functionality over recent data is required by monitoring processes, whereas On Line Analytical Processing (OLAP) functionality is required by analytical process over historical data. Finally, the need of technologies form the area of big data was also analyzed.

2 System Design

The three-layer architecture of the system is depicted in Fig. 1.

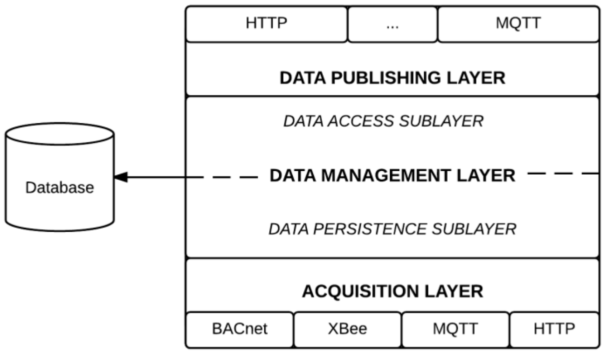

Fig. 1. System Architecture

The acquisition layer enables the uniform access to the data produced by the building sensors. The adaption to the heterogeneity of the different sensor and communication technologies is achieved by specific wrappers. Currently, however, only devices connected through the BACnet [2] protocol are incorporated in the OPERE Project. The data management layer is divided into two sub-layers, namely a data persistence sub-layer that records the data in appropriate data structures and a data access sub-layer. Finally, the data publishing layer enables the communication with the external clients, either through client/server or publish/subscribe interfaces. In particular, an HTTP REST web service (client/server) and a MQTT messaging service (publish/subscribe) are currently implemented. It is noticed that the architecture design is generic and enables its deployment in cascade by enabling data publishing services to be sensor data sources of a higher level.

The data model used by the data persistence sub-layer is based on the Observations and Measurements (O&M) OGC standard [3] and it enables the representation of *sensors* (devices) that may *measure* one or various *properties*, which in turn may be measured by one or various sensors. For each combination of sensor and property (measures relationship) various *observations* may be recorded at distinct time instants. Various types of such observation are supported, including Booleans, Measures (reals with an associated unit of measure) and categories.

3 Data Storage Optimization

To test the performance of the system a benchmark was established composed of a collection of sensor data and a set of relevant queries to be executed, representative of the requirements of the OPERE project. The test data was obtained from 1114 signals from 4 different buildings, 25% of them sampled every 10 seconds (time-triggered samplings) and 75% of them recorded only when their value changed (event-triggered samplings). The database contained 70 million records from 398 sensors (4GB of disk space). The benchmark included also 10 representative SQL queries.

Three different variations of the relational data structures derived from the aforementioned data model were tested in the following hardware configuration. Intel Core 2 Duo, 2GHz processor, 2GB RAM, 4 MB L2 cache, hard disk SATA 7200 rpm, Windows 7 Proffesional 32 bits, and PostgreSQL 9.3 relational DBMS, with a cache (shared_buffers) of 128 MB. The DBMS cache was renewed on every execution and the response time for each query was obtained from the average of 5 executions.

Device

Identifier {text}	Description {text}

Property

Identifier {text}	Description {text}	Type {text}

Measure

Device_id {text}	Property_id {text}	Uom {text}

Booleanobservation

Device_id {text}	Property_id {text}	Uom {text}	Measuretime {timestamp}	Value {boolean}

Measureobservation

Device_id {text}	Property_id {text}	Uom {text}	Measuretime {timestamp}	Value {real}

Categoryobservation

Device_id {text}	Property_id {text}	Uom {text}	Measuretime {timestamp}	Value {text}

Fig. 2. Data structures version 1

Fig 2 represents the first version of the data structures used for the data storage. Note that text typed identifiers were used both for devices and properties. Such version showed to be not efficient at all, both in storage size and query response time. All records were of variable size (72 bytes in average) and query response times were around 3 minutes, with small variation from one query to another.

In a second version of the data structures, text identifiers of devices and properties were replaced by numeric ones of a fixed size. Now, most structures had a fixed record size (19 bytes for measure observations) and query response times were much lower in most queries. The comparison of response times for the 10 queries is shown in Fig 3. Note that query (#6), which is the most common one in OPERE, is now executed in less than 10 seconds with such a small variation on the data structures.

The response time of queries on time-triggered samplings was growing in a reasonable proportion with the size of the data, however, the response time for queries on event-triggered samplings was too slow, due to the fact that consecutive event-triggered samplings were separated in the structures by too many intermediate time-triggered ones. To solve this, a third version of the data structures was designed and implemented. Now, time-triggered and event-triggered samplings were separated in different data structures and last observations of each combination of device and property were also recorded in a separate structure. Now queries on even-triggered samplings are really fast and accessing the last observations, which is very common in monitoring, is a matter of few milliseconds.

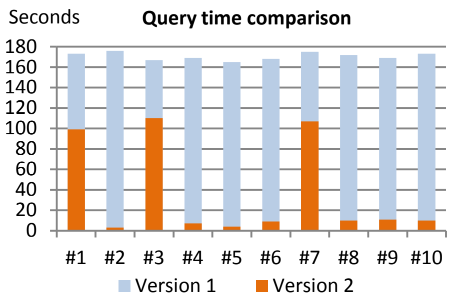

Fig. 3. Comparison between first and second version of the data structures

4 Conclusions

The use of optimized data structures in a relational database showed to be an efficient solution for the OPERE project, which is recording 1114 signals from 4 buildings during around 18 months and with a maximum frequency of 10 seconds, which is much higher than the one required for common management and control of infrastructures. Although such application domains do not require solutions from the area of big data, other possible scenarios will clearly demand them. Currently in the USC, 47 intelligent buildings are controlled with more than 10000 signals. Thus, applications that require the recording of all the data being acquired from all the buildings, would demand reliability and scalability requirements that would justify a migration to big data technologies, despite of the cost associated to their increased complexity compared to classical relational solutions.

Acknowledgement. This work has been partially supported by EC under the LIFE12-ENV-ES-001173 project application from Life+ Environment Policy and Governance Programme.

References

1. Directive 2010/31/UE related to Energy Performance of Buildings. Official Journal of the European Union on 18/06/2010 , p. 13 2010.
2. ANSI/ASHRAE Standard 135-2012: BACnet®--A Data Communication Protocol for Building Automation and Control Networks. ASHRAE, Atlanta, USA. 2012.
3. Cox, S. Geographic Information - Observations and Measurements. OGC Abstract Specification Topic 20. Open Geospatial Consortium (OGC). 2013.