

Optimización del Almacenamiento de Datos en la Gestión Energética de Edificios Inteligentes

Samuel Otero Paz^{1,2}, Jose A. Taboada González¹, Jose R.R. Viqueira¹, Juan E. Arias Rodríguez²

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. El objetivo principal del proyecto LIFE-OPERE es mejorar la gestión energética en grandes instalaciones de la Universidad de Santiago de Compostela (USC). Se almacenará y analizará la información de sensorización de los edificios para determinar el impacto de las nuevas medidas adoptadas. Para ello, se ha diseñado e implementado un sistema de adquisición, almacenamiento y publicación de datos. Este artículo se centra en la optimización del modelo de datos para hacer frente a los problemas y retos que han surgido durante el proyecto. Además, se valora también la necesidad de implantar una infraestructura big data.

Keywords: sensor, automatización, big data, base de datos, edificio inteligente

1 Introducción

Los edificios representan el 40% del consumo de energía final de la UE [1], por lo que resultan claves de reducir las emisiones de gases de efecto invernadero a través de la mejora de la eficiencia energética.

La Universidad de Santiago de Compostela (USC) participa como socio principal en el proyecto LIFE OPERE (2013-2016), cuyo objetivo principal consiste en desarrollar medidas orientadas a mejorar la gestión energética de las instalaciones de gran consumo que la universidad posee, así como demostrar su eficacia mediante el registro y análisis de toda la información necesaria. Para responder a las necesidades del proyecto, se diseñó e implementó un sistema de adquisición, almacenamiento y publicación de datos de la sensorización de los edificios.

En este contexto, se diseñó un modelo de datos basado en el estándar Observations and Measurements (O&M) [2] del OGC bajo el precepto de poder incorporar la mayor cantidad de fuentes y tipos de información (dataloggers, datos meteorológicos, etc.), partiendo de otro más genérico [3]. Este artículo se centra en la implementación

de dicho modelo en estructuras de datos relacionales, concretamente sobre bases de datos Postgres, y su capacidad para cumplir con los tiempos de respuesta requeridos, tanto para el almacenamiento en tiempo real de los datos de sensores como para la consulta de los mismos, bien para su monitorización (consultas de tipo OLTP) o su análisis (consultas de tipo OLAP).

2 Descripción del sistema

El sistema, implementado en Java, está caracterizado por una arquitectura en 3 capas, tal y como muestra la siguiente figura.

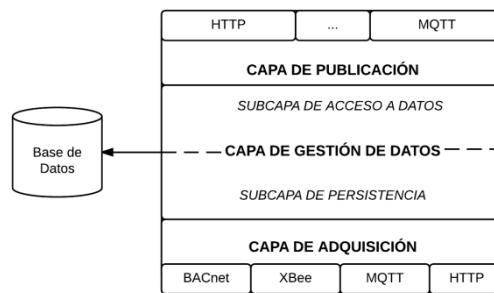


Fig. 1. Arquitectura del sistema

La capa de adquisición de datos introduce la flexibilidad necesaria para comunicarse con sensores que empleen diferentes protocolos de comunicación, aunque en el marco del proyecto OPERE únicamente se está empleando BACnet [4]. La capa de gestión de datos aborda la persistencia de datos aportando independencia sobre la tecnología de almacenamiento a usar. Finalmente, la capa de publicación aísla a las aplicaciones cliente de las capas inferiores ofreciendo una API de servicios, permitiendo conexiones tipo cliente-servidor (servicio web REST sobre HTTP) y o tipo publicador-suscriptor (MQTT). La flexibilidad de esta arquitectura permite que una instancia de la aplicación pueda enviar datos a otra instancia, pudiendo establecerse un árbol jerárquico con nodos hardware (sensores) o software (aplicaciones) en los que cada nodo envía datos al nivel superior.

El modelo de datos empleado en el sistema define una serie de sensores (device) que realizan distintos tipos de observación (boolean, measure o category) de una propiedad (property).

3 Optimización del Almacenamiento de Datos

Para evaluar el rendimiento del sistema de almacenamiento, se diseñó un banco de pruebas consistente en 70 millones de registros con datos de 398 sensores (4GB en disco) y un conjunto de 10 consultas tipo de interés para el proyecto. Los datos procedían de 1.114 señales de 4 edificios, el 25% muestreadas cada 10 segundos (datos

síncronos o disparados por tiempo) y el 75% registradas ante cambios de valor (datos asíncronos o disparados por eventos).

Se probaron 3 versiones diferentes de estructuras de datos relacionales derivadas del modelo de datos mencionado. El entorno de pruebas consistió en un equipo con procesador Intel Core 2 Duo a 2GHz, 2GB de RAM, 4 MB de caché L2, disco duro SATA de 7.200rpm de 500 GB y sistema operativo Windows 7 Professional de 32 bits, ejecutando un servidor Postgres 9.3 con una caché (shared_buffers) de 128MB. La caché de la BBDD se renovó en cada ejecución y el tiempo de respuesta de cada consulta se estimó como la media de 5 ejecuciones.

<i>Device</i>				
Identifier {text}	Description {text}			
<i>Property</i>				
Identifier {text}	Description {text}	Type {text}		
<i>Measure</i>				
Device_id {text}	Property_id {text}	Uom {text}		
<i>Booleanobservation</i>				
Device_id {text}	Property_id {text}	Uom {text}	Measuretime {timestamp}	Value {boolean}
<i>Measureobservation</i>				
Device_id {text}	Property_id {text}	Uom {text}	Measuretime {timestamp}	Value {real}
<i>Categoryobservation</i>				
Device_id {text}	Property_id {text}	Uom {text}	Measuretime {timestamp}	Value {text}

Fig. 2. Estructuras de datos de la versión 1

La figura 2 representa la estructura de datos de la primera versión, que emplea identificadores de tipo texto para sensores y propiedades. Esta versión demostró ser poco eficiente tanto en cuanto a espacio de almacenamiento como a tiempo de respuesta de las consultas. Los registros ocupaban una media de 72 bytes, siendo su tamaño variable en función de la longitud del identificador del sensor y la propiedad, y los tiempos de respuesta se situaron en torno a los 3 minutos independientemente del tipo de consulta.

En la segunda versión se emplearon identificadores numéricos de longitud fija. En este caso los registros se caracterizaban por un tamaño fijo de 19 bytes y los tiempos de respuesta fueron mucho menores y dependientes de la complejidad de la consulta. La figura 3 muestra la comparativa de los tiempos de respuesta entre ambas versiones. Cabe destacar que la consulta más habitual (#6) en OPERE se resolvía en menos de 10 segundos.

Conforme el volumen de datos almacenados crecía, la consulta de datos síncronos seguía respondiendo en el mismo orden de magnitud. Sin embargo, la obtención de datos asíncronos implicaba un tiempo de búsqueda excesivo, ya que éstos se encuentran separados entre sí por una gran cantidad de registros procedentes de señales síncronas. Para solventar este problema, la tercera versión del modelo separa en tablas diferentes los datos asíncronos de los síncronos, y añade una nueva tabla con el último dato de todas las señales. Esto permite que la tabla sobre la que se consultan las seña-

les asíncronas sea mucho más pequeña y que la consulta del último dato de cualquier señal presente un tiempo de respuesta del orden de milisegundos.

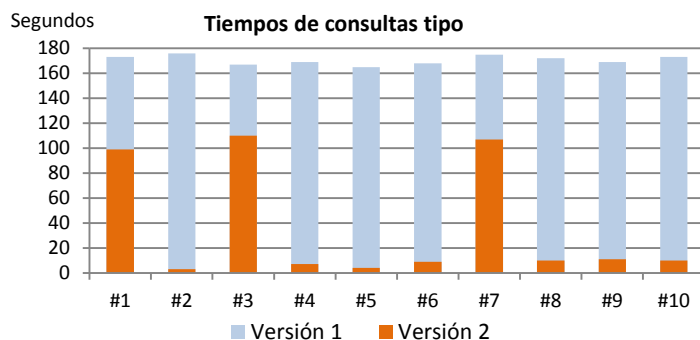


Fig. 3. Comparación de los tiempos de respuesta entre las versiones 1 y 2

4 Conclusiones y trabajos futuros

El uso de base de datos relacionales con estructuras de datos optimizadas demostró ser una solución válida en términos de rendimiento para el proyecto OPERE, en el que se están almacenando 1.114 señales procedentes de 4 edificios durante 18 meses y a una frecuencia de 10 segundos, que es mucho mayor que la necesaria para la mera gestión y control de las infraestructuras. Si bien en este contexto no es precisa una migración hacia entornos big data, sí se considera necesaria una evolución hacia los mismos en otros escenarios factibles. Actualmente la USC dispone de 47 edificios inteligentes controlados por más de 10.000 señales. Por lo tanto, aplicaciones que precisen el almacenamiento de toda esta información sí requerirían de la fiabilidad y escalabilidad que estos sistemas proporcionan, a pesar su mayor complejidad frente a las soluciones relacionales clásicas.

Agradecimientos. Este trabajo ha sido parcialmente financiado por la CE bajo el amparo del Programa LIFE+ Política y Gobernanza Medioambiental (referencia LIFE12-ENV-ES-001173).

Referencias

1. Directiva Europea 2010/31/UE relativa a la eficiencia energética en edificios. Diario Oficial de la Unión Europea del 18/06/2010, p. 13 2010
2. Cox, S. Geographic Information - Observations and Measurements. OGC Abstract Specification Topic 20. Open Geospatial Consortium (OGC). 2013
3. David Martínez et al.: Common Data Model in AmI Environments. En: 8th International Conference on Ubiquitous Computing and Ambient Intelligence. Personalisation and User Adapted Services, pp 212-215. (2014)
4. ANSI/ASHRAE Standard 135-2012: BACnet®--A Data Communication Protocol for Building Automation and Control Networks. ASHRAE, Atlanta, USA. 2012.