Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 122 (2017) 589–594

Energy Procedia

www.elsevier.com/locate/procedia

CISBAT 2017 International Conference Future Buildings & Districts Energy Efficiency from Nano to Urban Scale, CISBAT 2017 6-8 September 2017, Lausanne, Switzerland

Smart Cities (Urban Simulation, Big Data)

Big Building Data - a Big Data Platform for Smart Buildings

Lucy Linder[a], Damien Vionnet[b], Jean-Philippe Bacher[b], Jean Hennebert[a,*]

[a] *Institute of Complex Systems, HEIA-FR, Boulevard de Perolles 80, 1700 Fribourg, Switzerland*
[b] *Institute of Energy, HEIA-FR, Boulevard de Perolles 80, 1700 Fribourg, Switzerland*

Abstract

Future buildings will more and more rely on advanced Building Management Systems (BMS) connected to a variety of sensors, actuators and dedicated networks. Their objectives are to observe the state of rooms and apply automated rules to preserve or increase comfort while economizing energy. In this work, we advocate for the inclusion of a dedicated system for sensors data storage and processing, based on Big Data technologies. This choice enables new potentials in terms of data analytics and applications development, the most obvious one being the ability to scale up seamlessly from one smart building to several, in the direction of smart areas and smart cities. We report in this paper on our system architecture and on several challenges we met in its elaboration, attempting to meet requirements of scalability, data processing, flexibility, interoperability and privacy. We also describe current and future end-user services that our platform will support, including historical data retrieval, visualisation, processing and alarms. The platform, called BBData - *Big Building Data*, is currently in production at the Smart Living Lab of Fribourg and is offered to several research teams to ease their work, to foster the sharing of historical data and to avoid that each project develops its own data gathering and processing pipeline.

Peer-review under responsibility of the scientific committee of the CISBAT 2017 International Conference – Future Buildings & Districts – Energy Efficiency from Nano to Urban Scale

Keywords: Big Data; Building Management Systems; Smart Buildings; Web of Buildings;

1. Introduction

Concerns about climate change have triggered the need for more sustainable and dynamic energy management systems. This is especially true for the building sector, which accounts for about 20% of the energy consumed worldwide [1].

The standardisation of building sensors, for example through IoT technologies, and their coupling to smarter control systems are forming the basis of Smart Buildings, providing new ways for the owners, operators and facility managers to improve both the reliability and performances of building assets. Research wise, smart control systems are intensively investigated with the objectives to save energy and at the same time increase the comfort level. Examples of smart controls are promising, reporting savings from 10% for simple controls up to 60% for the most advanced

* Corresponding author. Tel.: +41-26-429-6596
E-mail address: jean.hennebert@hefr.ch

1876-6102
Peer-review under responsibility of the scientific committee of the CISBAT 2017 International Conference – Future Buildings & Districts – Energy Efficiency from Nano to Urban Scale
10.1016/j.egypro.2017.07.354

systems. Smart controls are addressing the different sources of consumptions in buildings such as heating, cooling, ventilation, electric lighting and solar shadings [2], lightning [3], electric appliance [4], etc. We observe that such technologies will rely more and more on the gathering of large amounts of data from multiple sensors, actuators and dedicated networks.

The handling of Smart Building data is challenging in many ways. First, the building automation domain has a long history with interoperability problems due to the diversity of systems and technologies [5], leading to interoperability and integration concerns [6]. Second, the density of sensors and actuators as well as the sensing frequency tend to increase for a finer observation and control of the equipments, leading to large quantities of data to process. As an example, in Project Dasher, the database reached its limits after only three months and the collection of 2 billion rows of data [7]. Third, advanced control, monitoring and post occupancy evaluation are relying on more and more complex models using both real-time and historical data, as well as aggregations and correlations in the range of several years. While real-time and time-aggregation processing have received a renewed attention and many frameworks try to address those needs, none of them are currently fully mature [8]. Finally, while the coupling between the IoT and the Big Data is strong, there are few comprehensive approaches to support the collection of data from building sensors to their exploitation. Current research efforts are mainly focusing on the collection of data from the data producer tiers, the reception tiers or the exploitation one [9].

In the BBData project, we attempt to answer those challenges by providing a full-featured data processing platform for Big Building Data. BBData is an ingestion, processing and sharing system able to scale up to the Big Data expectations of Smart Building environments. The project is an applied research contribution of HEIA-FR in the context of the Smart Living Lab in Fribourg. Still in the earlier stages of development, it has been running continuously for several months to gather data from more than two thousands sensors located at the Halle Bleue of the BlueFactory site.

Section 2 presents a general overview of the BBData platform. Subection 2.1 shows how Big Data technologies changed the landscape of Smart Building solutions. Subsection 2.2 presents the architecture of BBData, as well as the technologies that support it. In section 3, we review some of the challenges of big data and how we met them in our solution. In section 4, we present some key features of BBData and how Building professionals could benefit from it.

2. System description

2.1. From BMS to Web of Buildings

In this work, we rely on the inclusion of a dedicated system for the storage and processing of sensors data equipping smart buildings. More specifically, the architecture of this system is remote, de-coupled and based on Big Data technologies. This choice enables new potentials in terms of data analytics and applications development, the most obvious one being the ability to scale up seamlessly from one smart building to several, in the direction of smart areas and smart cities. More than a simple evolution of Building Management Systems (BMS), such open Big Data architectures are probably an important paradigm shift from BMS to what is called *Web of Buildings* [10].

As illustrated on the left part of Figure 1, traditional Building Management Systems (BMS) are standalone softwares running on a server located in the premises of a building. Standard BMS includes three levels of functions: (1) the field layer is made of sensors and actuators connecting to the physical word; (2) the automation layer applies strategies derived from a set of rules and parameters; (3) the management layer configures and manages the other layers. BMS usually offer limited capacities in terms of sensor data storage and act as isolated systems often unreachable from the outside.

The right part of Figure 1 illustrates our BBDATA architecture with the inclusion of remote solutions for storage and data processing. More than the ability of scaling up at the level of a city, this approach brings new opportunities in terms of BMS evolution through the de-coupling of sensors, data storage and applications levels. This opening is supported by the emergence of the open and standardised principles of the Internet of Things and Web of Things [11][12]. It allows for seamless integration of new data sources, such as weather forecasts available in web services or mobile sensors. It also allows for shorter application development cycles, enabling for example mashup approaches as in modern web application development. Finally, the availability of long term historical data allows to use advanced mathematical models and machine learning technique [13].

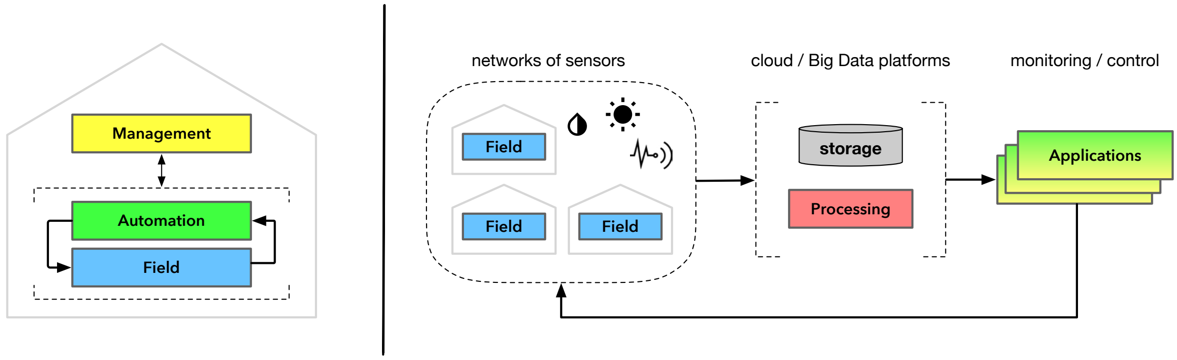

Fig. 1. On the right, a traditional BMS architecture; On the left, a *Web of Buildings* architecture.

2.2. The BBData architecture

Figure 2 shows the journey of a measure in the BBData system. In (1), different kind of sensors produce measures. Those sensors might come from different manufacturer and use various data encodings. To ensure the compatibility of the platform with any kind of equipment, BBData uses the concept of *virtual objects*. A virtual object has metadata, such as name, unit and type and can be mapped to a real sensor through an ID. This mapping is handled by *collectors* (2). A collector creates a *bbdata record* for each measure and sends it to the *input api* in a structured JSON format. In case some sensors lack an internal clock, the collector can also produce a timestamp. The *input api* (3) is a JavaEE REST service running on GlassFish[1] whose role is to validate the incoming measure and ensure its authenticity using the object's ID and a secure *token*. If the check succeeds, security informations are dropped and the resulting JSON is added to an Apache Kafka[2] message queue (4) for processing. The content of this queue is also dumped periodically to HDFS, so the raw inputs can be replayed anytime. Before processing, the measure is first "augmented" with the virtual object's metadata pulled from a MySQL[3] database (5). The result is stored in a second message queue using a compressed format (6).

In BBData, processing covers a wide area of tasks. It goes from the saving of raw values into a persistent store to the detection of anomalies or the computation of time aggregations. Each of these tasks is handled by a specific processor. Processors are independent streaming applications running in a hadoop cluster. They subscribe to the augmented Kafka topic, carry their task and save their output, if any, in a Cassandra database. This design makes it possible to add or remove processors without any impact on the system. We have currently two kind of processors, both implemented with Apache Flink[4]: the first saves the raw records to Cassandra, the second one computes live time aggregates (mean, max, last measure, standard deviation) with a granularity of fifteen minutes, one hour and one day.

Users and building automation applications can access the data and manage virtual objects through a standard REST interface called *the output api* (7) or via HTML5 web applications.

3. Big Data: challenges and pitfalls

The potential of Big Data technologies is undeniable. However it raises many challenges when it comes to design a sustainable and efficient data ecosystem. First, even if Hadoop is a de-facto standard in the Big Data landscape, it is made of many components distributed on multiple machines. The maintenance of such system is rather intensive

[1] https://glassfish.java.net/
[2] https://kafka.apache.org/
[3] https://www.mysql.com/
[4] https://kafka.apache.org/

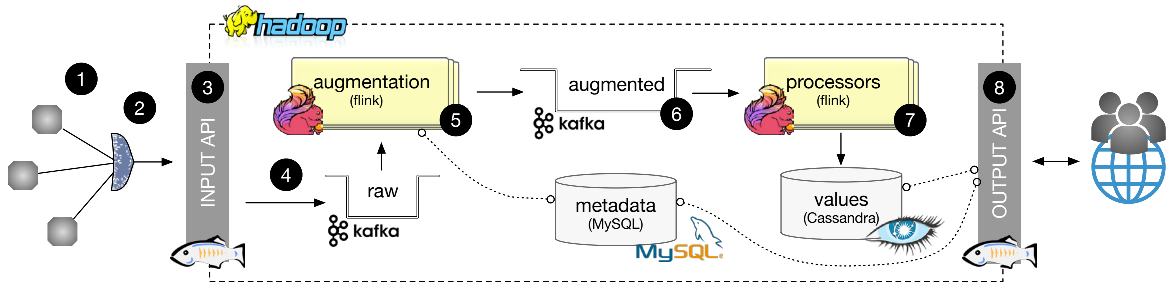

Fig. 2. Architecture of BBData: the journey of a measure

and the learning curve is steep. Second, the technology landscape in Big Data is fast evolving and improvements are constant, especially in the analytics domain. With our BBData system, we had to and we still are compromising between stabilising our environment and incorporating the latest techniques, sometimes questioning initial choices in front of emerging Big Data design patterns. We can report here on two examples we faced while designing BBData and that are specific to the storage and processing of building data.

At the storage level, NoSQL databases are compromising between consistency (C), availability (A) and partition tolerance (P), known as the CAP theorem [14]. It has been revised since to include latency in the equation [15]. Our choice of using Apache Cassandra was motivated by its good reported performance as an AP database and also by the fact we could reach good consistency with a proper configuration. But this performance comes at a cost: the supported queries are tightly dependent on the data model. Such scalable databases use the notion of *shards* that are partitions of the data on different nodes. A given data will be stored in a given shard according to a user defined strategy. In BBData, we privileged a fair distribution of the data on several shards using as distribution key the object's id and the timestamp's year and month. This ensures data are well-balanced between nodes, but also limits our queries to those where the object id and the time period are known. Moreover, since measures from different month are potentially stored on different machines, the wider the time range, the less the performance.

At the processing level, our focus on building data and energy management requested the ability to compute time-based aggregations on streams. While aggregations are relatively common needs, the main stream-processing frameworks such as Apache Spark, Flink and Kafka Streams are still looking for satisfying general solutions. In our case, we needed aggregations based on timestamps embedded in the data and also that the time advanced differently for each virtual object. This conjonction of requirements are not supported by any framework yet. As a result, we had to implement our own tool to offer on-the-fly time aggregations to our users.

4. User services

BBData can offer many tools to different types of user. It includes the building user, the building owner, the architects and the engineering consultancies. The first set of tools is related to data monitoring. At the building user point of view, these tools offer access to live monitoring, to historic data and to different type of pre-calculated values. The data monitoring is also useful to the architects and the engineering consultancies. Presently, these two actors have access to few information once the building is in operation. A data monitoring service would be a real added value for them. Such a service allows to understand the behavior of the building and the behavior of the occupants. From the observed behaviors, the engineering consultancies are able to do post-occupancy evaluation and optimization to reduce the energy consumption and/or improve the user satisfaction [16][17]. Figure 3 illustrates a potential for energy savings based on correlating door sensors and lighting consumption. For the architects, they can use this information to build best practice guides and to enrich their experience for the next constructions.

Building owners have also expressed their interest for other types of services. The first one is the "anomaly detection" [18]. By taking advantage of data history, it is possible to identify data patterns that define the standard

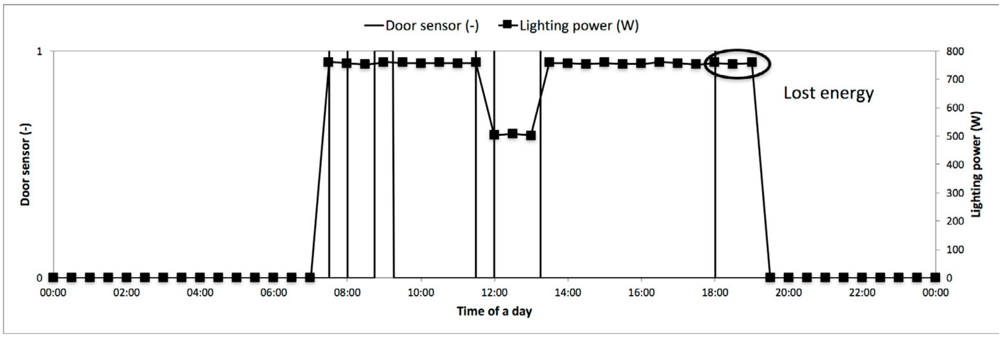

Fig. 3. Post occupancy optimization: modify lighting automation to save energy after the last person leaves the office

behavior of the building and its users. By comparing these patterns to the current state (see Figure 4), discrepancies can be detected automatically and alerts can be triggered, for example a maintenance alert. Based on the same idea, a predictive maintenance can be developed. Like patterns for standard building behaviors, it is possible to define patterns from the occurred anomalies and associate to them metadata such as the equipment or user behavior causing the anomaly and the way the issue was solved. If one of these anomalies arise one more time, the BBData system could detect it and suggest the right issue resolution. These different services are not yet provided by the BBData system but are under development.

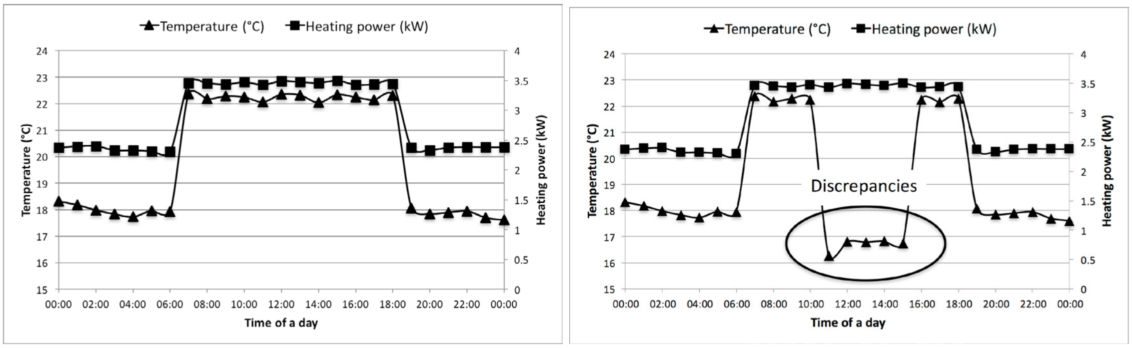

Fig. 4. Standard pattern vs. abnormal pattern - Simulated data

We are also currently developing a third set of tools useful to the building owner for planning building renovation and comparing the performances of different buildings. These tools are able to compute yearly building energy signatures and environmental CO2eq footprints. With such tools, the building owner can compare the current state versus the previous year and check if the established action plan has any effect. Prediction models will also be included based on the current state and on historical data for the rest of the year [19].

5. Conclusion and future work

In this paper, we presented the BBData platform, a distributed system for storing and processing building data. Based on Big Data technologies, the platform enables new potentials in terms of data analytics and applications development, with the ability to scale up seamlessly from one smart building to several, in the direction of smart areas and smart cities. We detailed the architecture of the system and the different technological choices we made by describing the journey of a data emitted from a sensor and going trough the modules of the system. We then reviewed some of the challenges we met setting up the platform, some of them being related to classical constraints of Big Data environments. We finally described several services, existing and under development, that leverage on BBData and that are dedicated to building users, owners, architects and managers.

In our future work, we plan to further develop end-user services such as alarms based on advanced anomaly detection and prediction. Other areas of development will go in the direction of multivariate stream processing including, for example, time aggregations on a group of objects or correlations between streams.

References

[1] Residential and commercial buildings. U.S. Energy. Information Administration, International Energy Outlook 2016; 2016.

[2] Daum, D.. On the Adaptation of Building Controls to the Envelope and the Occupants. Ph.D. thesis; ENAC; Lausanne; 2011. doi:10.5075/epfl-thesis-4935.

[3] Rea, M., of North America, I.E.S.. The IESNA Lighting Handbook: Reference & Application. IESNA LIGHTING HANDBOOK; Illuminating Engineering Society of North America; 2000. ISBN 9780879951504. URL: https://books.google.mu/books?id=0Ot4QgAACAAJ.

[4] Ridi, A., Gisler, C., Hennebert, J.. A Survey on Intrusive Load Monitoring for Appliance Recognition. In: 22nd International Conference on Pattern Recognition - ICPR. IEEE. ISBN 9781479952106; 2014, p. 3702–3707. doi:10.1109/ICPR.2014.636.

[5] Ploennigs, J.. A modular, adaptive building automation system ontology. IEEE IECON 2012;.

[6] Moreno, M.V.. Applicability of big data techniques to smart cities deployments. IEEE Transactions on Industrial Informatics 2016;.

[7] Khan, A.. Big data from the built environment. Proceedings of the 2nd international workshop on Research in the large 2011;.

[8] Maninder Pal Singh Mohammad A. Hoque, S.T.. A survey of systems for massive stream analytics. Tech. Rep.; Cornell University; 2016.

[9] Cyril Cecchinel Matthieu Jimenez, e.A.. An architecture to support the collection of big data in the internet of things. Conference: 2014 IEEE World Congress on Services 2014;.

[10] Bovet, G.. A scalable and sustainable web of buildings architecture. Ph.D. thesis; Telecom ParisTech; 2015.

[11] Bovet, G., Hennebert, J.. A web-of-things gateway for knx networks. In: Smart SysTech 2013; European Conference on Smart Objects, Systems and Technologies. 2013, p. 1–8.

[12] Bovet, G., Hennebert, J.. Offering web-of-things connectivity to building networks. In: Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication. UbiComp '13 Adjunct; New York, NY, USA: ACM; 2013, p. 1555–1564.

[13] Ridi, A., Zarkadis, N., Bovet, G., Morel, N., Hennebert, J.. Towards reliable stochastic data-driven models applied to the energy saving in buildings. In: International Conference on Cleantech for Smart Cities & Buildings from Nano to Urban Scale (CISBAT 2013). 2013, p. 501–506.

[14] Gilbert, S., Lynch, N.. Perspectives on the cap theorem. Computer 2012;45(2):30–36.

[15] Abadi, D.. Consistency tradeoffs in modern distributed database system design: Cap is only part of the story. Computer 2012;45(2):37–42.

[16] Zimmerman, A.. Post-occupancy evaluation: benefits and barriers. Building research and information 2001;.

[17] Tarrio-Saavedra, J., al., . Assessing thermal comfort and energy efficiency in buildings by statistical quality control for autocorrelated data. Applied energy 2017;.

[18] Burton, K., al., . Development and alpha testing of a cloud based automated fault detection and diagnosis tool for air handling units. Automation in Construction 2014;.

[19] Mathieu, J., al., . Quantifying changes in building electricity use, with application to demand response. IEEE Transactions on SMART GRID 2011;.