A DATA ACQUISITION FRAMEWORK FOR BUILDING ENERGY MANAGEMENT

Konstantinos Touloumis, Panagiotis Kapsalis, Elissaios Sarmas, Stathis Stamatopoulos, Vangelis Marinakis and Haris Doukas
Decision Support Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, 9 Iroon Polytechniou Str., 15773 Athens, Greece

ABSTRACT

Collecting data for providing energy analytics services has closely been dependent on Building Management Systems (BMSs). A BMS connected to a set of sensors and other electronic devices aims at observing a collection of rooms in terms of temperature, humidity, and other energy related metrics with the goal of maximining user comfort while preserving energy consumption. The tasks of collecting, preprocessing, storing and querying data on a BMS faces the same challenges with those of querying big data in terms of scalability, data heterogeneity, and integration. This paper presents a framework that addresses the problem of collecting sensor data in one enriched data warehouse which follows a common ontology model. Data from various data sources is homogenized by appropriate data preprocessing and feature extraction techniques. The framework allows timeseries-oriented querying with the target outcome of providing stakeholders high detailed analytics services for decision making on energy consumption optimization, building renovation and financing.

KEYWORDS

Building Energy, Analytics, Big Data Pipeline, Energy Data Modelling, Building Data Processing

1. INTRODUCTION

Managing data on the energy sector is of significant importance since the building sector accounts for almost one fourth of the energy consumed worldwide (International Energy Outlook 2016), for electric lighting (Yun et al., 2012) and appliances (Ridi et al., 2014). The problem becomes even greater when it comes down to CO_2 emissions as the building sector is responsible for almost 40% of the global CO_2 emissions (Hu, 2022). The rapid rise in the energy demand of the building sector comes along with the exponential growth of networking capabilities including the Internet of Things (IoT) (Marinakis & Doukas, 2018), which has led to an outburst of data collected by sensors in large scale networks. Thus, efficient management on the building sector is of crucial importance for preserving energy consumption and further reducing environmental harm. Big data querying frameworks try to leverage the problem of collecting, processing, and providing data analytics services. Many approaches have integrated big data analytics methods for efficient data management on the building sector and will be analyzed thoroughly in the next sections.

The problem of handling data from buildings addresses a number of challenges as mentioned in (Pau et al., 2022). Firstly, the problem of automated data integration, as data is collected from various data sources and includes both real time and historical data. Secondly, the problem of data scaling, as sensors produce densely populated data on a high rate leading to large quantity of data within a small-time frame. The data heterogeneity is another major concern, as data is stored in different formats between sensors which makes data preprocessing a complex task, so a common ontology model is needed to project data from various data sources onto a common dataspace. Finally advanced monitoring and querying both real time and historical data along with aggregations is a complex but essential task for supporting high level user driven services (Marinakis et al., 2020).

The purpose of this paper is to present a framework for managing big data from the building sector. The proposed framework manages to efficiently deal with the aforementioned challenges. The steps of the framework are the following: Collecting data, to support efficient big data management and scalability the solution is built upon a data warehouse which receives real time streams of data from various sensors.

Preprocessing data, with appropriate feature extraction, normalization, data harmonization techniques and ontology modelling. A common ontology model will be used to address the problem of data heterogeneity which not only reuses existing ontologies models like BRICK (Balaji et al., 2016) and SAREF (Daniele et al., 2015) but is also FIWARE compliant (Rodriguez et al., 2018) to ensure data sovereignty and protection. Storing the processed data in a timeseries database that will allow high level time series querying thus allowing users to have a deeper understanding on how the energy consumption varies on different indoor and outdoor parameters.

2. BACKGROUND

As mentioned in (Minoli et al., 2017) a traditional BMS comprises of the following layers: The field layer, which consists of all the sensors installed on the building premises. The automation layer, which applies strategies derived from a set of rules. The management layer, that manages the functionality of a BMS. BMSs are often limited in storage capacity and act isolated from the entire network. In the same research a system for collecting building data is structured that uses the notion of virtual objects that can be mapped to a real sensor through an ID created by a specialized collector. Tokens are used to protect the data flow and a MySQL database is used to augment metadata with a virtual object and transmit it through Kafka messaging. However, the research later mentions that at the storage level the NoSQL databases are compromising between consistency, availability and partition tolerance known as the CAP theorem (Davoudian et al., 2019). Also, Kafka messaging is not suitable for computing time-based aggregations on streams (le Noac'h et al., 2017).

Similarly, another big data system for analytics on the building sector is analyzed in (Pau et al., 2022) that firstly preprocesses data to remove incomplete or inaccurate data, and then uses Kafka for transmitting data because of its high throughput and latency but ignores its inefficiency for time-based aggregation on time-series streams. Data is then harmonized and projected onto a common data model. The enriched data is stored in a Mongo DB that acts as the enriched data warehouse. On top of the warehouse the Presto (Sethi et al., 2019) technology is used, a distributed query engine that provides high level of abstraction for querying different data sources. However, it should be mentioned that Presto is not tailor made for querying time series data in contrast to other databases like InfluxDB (Nasar & Kausar, 2019) which can query large scale time series data without extract, transform, load (ETL).

Efficient management and monitoring of energy consumption data can benefit stakeholders in many ways. Analytics services can be built upon the upper layer of the proposed framework that will assist decision making for the following purposes: Monitoring the energy consumption, aimed at improving the building energy performance. Predicting, by using appropriate models, the energy demand towards minimizing energy consumption across different weather conditions (Karakolis et al., 2022). Designing and enhancing building infrastructure, aimed at assisting the design and development of building infrastructure, retrofitting and refurbishment actions. Designing climate resilient buildings that will ensure people comfort and well-being while preserving energy consumption (Yang et al., 2022). Creating green policies towards building sustainability, aimed at assisting policy making towards sustainable action plans, energy performance certificates, and assessing the impact of EU policies for buildings (El-Diraby et al., 2017). Decision making for energy efficient buildings, aimed at assisting in making accurate predictions on energy consumption, towards contributing to Energy performance contract conditions, EU financing institutions, and centralizing building stock data (Marinakis, 2020).

3. ARCHITECTURE

The proposed framework collects both building sensor and electricity consumption data and is depicted in Figure 1. The framework follows a modular architecture with every layer strongly interoperating with its previous and next one.

BMS: The Building Management System is an embedded system that includes sensors, meters and actuators for measuring energy related data in the building rooms. Such data includes the humidity and the temperature of the monitored building, as well as electricity consumption data from lighting, A/C meters and ventilation. All the collected information is reported in CSV files every 5 minutes. The CSV files are transformed to ZIP files and are sent to the FTP Staging Area.

Staging Area: The Staging Area uses the FTP (File Transfer Protocol) to send the collected data to the central data warehouse. The staging area authentication information and credentials should be declared in the BMS to transfer the measured data. The Staging Area is used as the intermediate layer of the framework where raw data is persisted and can further be analyzed by the next modules.

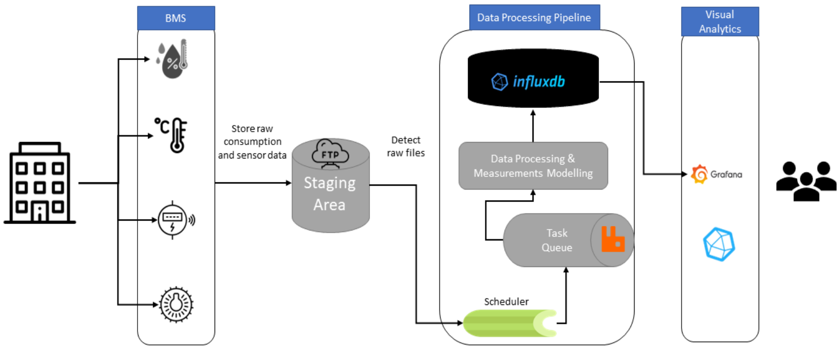

Figure 1. Building data processing pipeline

Data Processing Pipeline: The Data Processor is the module responsible for detecting the incoming raw information from the FTP Staging Area, scheduling and prioritizing data pipeline workers, pre-processing and cleaning raw data, modelling the processed information and finally storing it to an InfluxDB instance. More specifically each sub-module functionality is described below:

- **Data Processing Pipeline Scheduler**: It is a software component implemented with Python 3.7. It uses the Celery Task Scheduling Framework that is triggered periodically every 1 hour and collects all new raw information from the Staging Area. By leveraging the functionalities of the Celery framework, the tasks and the computational load are distributed to machine threads. The information is then sent to the Task Queue to be processed from the Data Processing and Measurements Modelling sub-modules.
- **Task Queue**: It is a RabbitMQ instance, connected to the Framework Scheduler that manages the collected data for processing and modelling.
- **Data Processing & Measurements Modelling**: This module receives events to be processed and modelled from the Task Queue. It is implemented with Pandas DataFrame and InfluxDB Python drivers. It conducts the pre-processing steps (removing duplicates, handling null values, detecting outliers, transforming dates) to improve the data quality. The collected data is divided in smaller data collections according to the measured information.
- **InfluxDB Storage**: It is the timeseries database that structures the stored information in measurements and tags. In the current design the metrics measured are the humidity, temperature and electricity consumption from lighting, ventilation, and air-conditioning. The tags are the device names that the measured information originates from.

Visual Analytics: The Grafana visualization engine will be used for visualizing analytics. Grafana is a well-known opensource interactive visualization web platform that allows users to create customized dashboards, charts and graphs. Grafana is connected to the InfluxDB storage and will allow users to visualize analytics that will let them further explore and infer logical relations between data coming from various data sources.

In order to define a data model to project various data sources onto a common dataspace, the ontology development 101 (Pullmann et al., 2017) is being followed which defines the steps needed for that purpose. The FIWARE smart data model offers data models for modeling many domains including the building and energy domains, since it defines the properties and the facets of the slots, but it lacks hierarchy between smart cities domain (Kapsalis et al., 2022). The BRICK ontology can be used for representing relationships between buildings and sensors, but it lacks properties and facets. SAREF defines many recurring objects from different ontologies in the building domain which can provide features and assist less experienced modelers.

However, control strategies over spaces are not included in all SAREF extensions. The EPC4EU models the energy performance certificate datasets in different geographical scales, however, the EPC4CEU doesn't store information about sensors. Finally, the created ontology is translated to a data model by following the ontology development 101 method. The model is depicted in Figure 2.

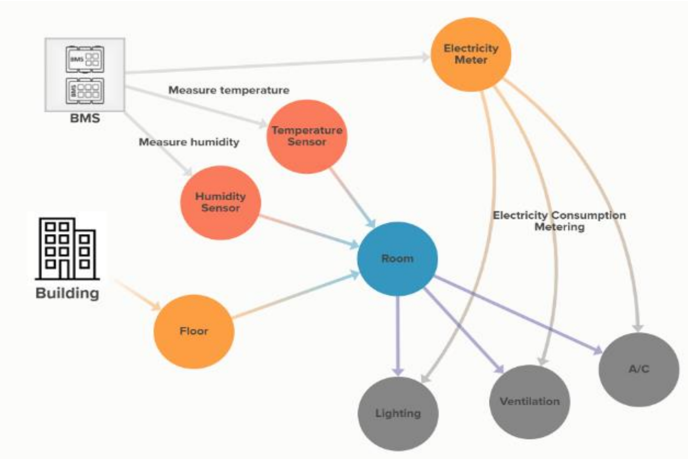

Figure 2. Constructed data model

The constructed data model is aligned with the information stored in the Influx timeseries database. More specifically, the timeseries information includes the measured values of electricity consumption from lighting, ventilation, and air-conditioning functionalities of the room. A logical entity of the constructed ontology model is the "Building", that has floors and rooms. The main entity of the conceptual Building data model is the "Room", facilitating lighting, air conditioning and ventilation. The logical model "BMS" represents the temperature, the humidity of the room and its electricity consumption from lighting, ventilation, and air-conditioning. In our experiment we have established a first level of tags, the room IDs, and a second level of tags, the source types that produce the recorded information (lights, ventilation system, air-conditioning).

4. CONCLUSION & FUTURE OUTLOOK

The framework presented in this research collects, analyzes, and homogenizes data coming from various energy sources to monitor the energy consumption of buildings. The power market to stay competitive needs to predict the electrical power consumption in both short and long term. Load forecasting aims at predicting the demand of energy consumption towards efficient power distribution construction and planning (Fekri et al., 2021). The proposed framework can be used in forecasting the energy consumption of buildings. By adding weather data such as the wind speed, solar radiation and by applying Deep Learning (DL) techniques, models can be trained to predict the energy consumption on different weather conditions and further assist building administrators in efficient energy planning. Deep learning techniques can handle big data by capturing the inherent non-linear features through automatic feature extraction methods. This work will be expanded by adding an offline training stage where data captured by the framework will be used to train robust predictive models, such as RNNs, LSTMs, for both short- and long-term energy consumption.

ACKNOWLEDGEMENT

This work has been co-funded by the European Union's Horizon Europe research and innovation programme under the DigiBUILD project, Grant Agreement No 101069658.

REFERENCES

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., Berges, M., Culler, D., Gupta, R., Kjærgaard, M. B., Srivastava, M., & Whitehouse, K. (2016). Brick. *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*, 41–50. https://doi.org/10.1145/2993422.2993577

Daniele, L., den Hartog, F., & Roes, J. (2015). *Created in Close Interaction with the Industry: The Smart Appliances Reference (SAREF) Ontology* (pp. 100–112). https://doi.org/10.1007/978-3-319-21545-7_9

Davoudian, A., Chen, L., & Liu, M. (2019). A Survey on NoSQL Stores. *ACM Computing Surveys*, *51*(2), 1–43. https://doi.org/10.1145/3158661

El-Diraby, T., Krijnen, T., & Papagelis, M. (2017). BIM-based collaborative design and socio-technical analytics of green buildings. *Automation in Construction*, *82*, 59–74. https://doi.org/10.1016/j.autcon.2017.06.004

Karakolis E., Pelekis S., Mouzakitis S., Markaki O., Papapostolou K., Korbakis G., & Psarras J., (2022). Artificial Intelligence for Next Generation Energy Services across Europe–The I-NERGY Project. *IADIS International Conference E-Society 2022 (e-Society 2022)*, 61–68.

Fekri, M. N., Patel, H., Grolinger, K., & Sharma, V. (2021). Deep learning for load forecasting with smart meter data: Online Adaptive Recurrent Neural Network. *Applied Energy*, *282*, 116177. https://doi.org/10.1016/j.apenergy.2020.116177

Hu, M. (2022). Embodied Carbon Emissions of the Residential Building Stock in the United States and the Effectiveness of Mitigation Strategies. *Climate*, *10*(10), 135. https://doi.org/10.3390/cli10100135 International *Energy Outlook 2016*. (n.d.).

Kapsalis, P., Kormpakis, G., Alexakis, K., Karakolis, E., Mouzakitis, S., & Askounis, D. (2022). A Reasoning Engine Architecture for Building Energy Metadata Management. *2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA)*, 1–7. https://doi.org/10.1109/IISA56318.2022.9904419

le Noac'h, P., Costan, A., & Bouge, L. (2017). A performance evaluation of Apache Kafka in support of big data streaming applications. *2017 IEEE International Conference on Big Data (Big Data)*, 4803–4806. https://doi.org/10.1109/BigData.2017.8258548

Marinakis, V. (2020). Big Data for Energy Management and Energy-Efficient Buildings. *Energies*, *13*(7), 1555. https://doi.org/10.3390/en13071555

Marinakis, V., & Doukas, H. (2018). An Advanced IoT-based System for Intelligent Energy Management in Buildings. *Sensors*, *18*(2), 610. https://doi.org/10.3390/s18020610

Minoli, D., Sohraby, K., & Occhiogrosso, B. (2017). IoT Considerations, Requirements, and Architectures for Smart Buildings—Energy Optimization and Next-Generation Building Management Systems. *IEEE Internet of Things Journal*, *4*(1), 269–283. https://doi.org/10.1109/JIOT.2017.2647881

Nasar, M., & Kausar, M. A. (2019). Suitability Of Influxdb Database For Iot Applications. *International Journal of Innovative Technology and Exploring Engineering*, *8*(10), 1850–1857. https://doi.org/10.35940/ijitee.J9225.0881019

Pau, M., Kapsalis, P., Pan, Z., Korbakis, G., Pellegrino, D., & Monti, A. (2022). MATRYCS—A Big Data Architecture for Advanced Services in the Building Domain. *Energies*, *15*(7), 2568. https://doi.org/10.3390/en15072568

Pullmann, J., Petersen, N., Mader, C., Lohmann, S., & Kemeny, Z. (2017). Ontology-based information modelling in the industrial data space. *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1–8. https://doi.org/10.1109/ETFA.2017.8247688

Ridi, A., Gisler, C., & Hennebert, J. (2014). A Survey on Intrusive Load Monitoring for Appliance Recognition. *2014 22nd International Conference on Pattern Recognition*, 3702–3707. https://doi.org/10.1109/ICPR.2014.636

Rodriguez, M. A., Cuenca, L., & Ortiz, A. (2018). FIWARE open source standard platform in smart farming-a review. *Working Conference on Virtual Enterprises*, 581–589.

Sethi, R., Traverso, M., Sundstrom, D., Phillips, D., Xie, W., Sun, Y., Yegitbasi, N., Jin, H., Hwang, E., Shingte, N., & Berner, C. (2019). Presto: SQL on Everything. *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, 1802–1813. https://doi.org/10.1109/ICDE.2019.00196

Yang, B., Lv, Z., & Wang, F. (2022). Digital Twins for Intelligent Green Buildings. *Buildings*, *12*(6), 856. https://doi.org/10.3390/buildings12060856

Yun, G. Y., Kim, H., & Kim, J. T. (2012). Effects of occupancy and lighting use patterns on lighting energy consumption. *Energy and Buildings*, *46*, 152–158. https://doi.org/10.1016/j.enbuild.2011.10.034