

Report on the Second International Workshop on Energy Data Management (EnDM 2013)

Torben Bach Pedersen
Center for Data-intensive Systems (Daisy)
Aalborg University
9220 Aalborg, Denmark
tbp@cs.aau.dk

Wolfgang Lehner
Database Technology Group
Technische Universität Dresden
01062 Dresden, Germany
{wolfgang.lehner}@tu-dresden.de

1. INTRODUCTION

The energy sector is in transition—being forced to re-think the current practice and apply data-management based IT solutions to provide a scalable and sustainable supply and distribution of energy. Novel challenges range from renewable energy production over energy distribution and monitoring to controlling and moving energy consumption. Huge amounts of “Big Energy Data,” i.e., data from smart meters, new renewable energy sources (RES—such as wind, solar, hydro, thermal, etc), novel distributions mechanisms (Smart Grid), and novel types of consumers and devices, e.g., electric cars, are being collected and must be managed and analyzed to yield their potential.

Energy is at the top of the worldwide political agenda. For example, The European Union has stated the “20-20-20 goals” (20% renewable energy, 20% better energy efficiency, and 20% CO₂ reduction by 2020). Even more ambitious goals are set for 2030 and 2050. This situation is reflected in research funding schemes such as the EU Horizon 2020 Framework program as well as national programs. Increasingly, such programs include joint calls involving both energy and IT partners. Data management is at the heart of this development, as witnessed by the following story headlines from key players: “The Smart Grid Data Deluge” (O’Reilly Radar); “Big data for the Smart Grid” (theenergycollective); “The Coming Smart Grid Data Surge” (Smart-GridNews.com).

Thus, data management within the energy domain becomes increasingly important. The International Workshop on Energy Data Management (EnDM) focuses on conceptual and system architecture issues related to the management of very large-scale data sets specifically in the context of the energy domain. The overall goal of the EmDM workshop is a) to bridge the gap between domain experts and data management scientists and b) to create awareness of this emerging and very challenging application area. For the workshop’s research program, the organizers especially try to attract contributions that push the envelope towards novel schemes for large-scale data processing with special focus on energy data management.

The Second International Workshop on Energy Data Management (EnDM’13)¹ was held in conjunction with

¹<http://endm2013.endm.org>

EDBT 2013 in Genova, Italy, on March 22, 2013. This half-day event brought together researchers and engineers from academia and industry to discuss and exchange ideas related to energy data management and related topics. The workshop featured one industrial keynote, five research papers, and finished off with a panel/roundtable discussion. The accepted papers spanned a number of exciting topics within energy data management, including (in no particular order) representation of smart meter data, ontologies for emissions trading, and forecasting of renewable energy production. Two papers concerned the important topic of capturing and managing flexible energy demands, specifically the visualization of flexible energy demand objects and the extraction of consumption flexibilities from consumer consumption time series. The workshop proceedings have been published in a joint volume of all EDBT/ICDT 2013 workshops [1].

2. INDUSTRIAL KEYNOTE

The keynote was given by Data Warehouse Architect Jens Otto Sørensen from the Danish Transmission System Operator (TSO), Energinet.dk, and was entitled “The Danish DataHub Solution.” The keynote first described the un-bundling and liberalization which has taken place in the Danish electricity market over the past decade. This process has led to a number of new problems, including the lack of separation between grid companies and electricity suppliers, competitive market barriers, (too) varying quality of readings and master data, no overview of errors and delays in transactions and data exchange, and the inability of lack of (corporate) customers to get sufficient overview of their electricity consumption. These problems led to new market regulations in order to solve them. However, in order to implement these regulations there was a need for a common place to exchange detailed energy data among all the market players. This so-called “DataHub” has now been implemented and entered into production². The talk explained the benefits of the solutions, including a generally improved data management, technocratic un-bundling, lowering market entry barriers, improved efficiency, facilitating new products and services, and better market integration. The talk also outlines the desirable properties for such a solution, includ-

²DataHub web page

ing (market) transparency, seamlessness, real-time operation, decoupling of (business) processes, full “transaction time flexibility” (rolling back committed transactions, processing transactions in the past and the future), and low technical entry barriers. As of March 2013, the system processes around 2 million inbound and 2 million outbound transactions per day, contained in around 800,000 messages, the largest of which was 44MB(!). The keynote summarized the lessons learned, which included the never-ending importance of data quality, the non-trivial involvement of IT vendors, and the significant technical and organizational challenges of communicating with 130+ organizations using 25+ IT systems. However, an even bigger challenge was to get everyone involved to understand the new business requirements (new timelines and regulations). Finally, the keynote discussed the taken design decisions and concluded that would have been different if only technical, and not also political, considerations had to be taken. For example, the political requirements meant that EDIFACT (in addition to XML) formats were allowed and the master data ownership was distributed to 78 local grid companies rather than a single central authority.

3. RESEARCH PAPERS

The first paper “Symbolic Representation of Smart Meter Data” by Tri Kurniawan Wijaya, Julien Eberle, and Karl Aberer focused on the topic of smart meter data analytics, which allows utility companies to analyze of smart meter data in real-time to understand customer behavior. However, the data volumes are very large, leading to performance problems, and detailed meter data is furthermore a potential privacy breach. Thus, the paper instead proposes to generalize the detailed readings into symbolic units that reduces both the volume and privacy risks significantly, while still allowing interesting analyses, e.g., data mining, to be performed on the symbolic data. A number of experiments on real-world data showed the feasibility of this very interesting proposal.

The next paper “Visualizing Complex Energy Planning Objects With Inherent Flexibilities” by Laurynas Siksnys and Dalia Kaulakiene focused on visualization of smart grid data. Specifically, it considered visualizing objects capturing the inherent flexibilities in (intended) electricity consumption and production, so-called “flex-offers.” The paper first presented the planning and control activities involved in balancing demand and supply, which are made harder by increasing rates of (non-schedulable) renewable energy, faced by current energy companies. The paper then presented its OLAP-inspired approach to navigate and explore flex-offers, including several specific visualizations and a histogram-based technique for the visualization. The paper finished by outlining the research challenges ahead in visualizing energy flexibilities.

The paper by Umberto Ciorba, Antonio De Nicola, Stefano La Malfa, Tiziano Pignatelli, Vittorio Rosato, and Maria Luisa Villani called “Towards Ontological Foundations of Knowledge related to the Emissions Trad-

ing System” discussed the European Union’s Emissions Trading System (EST). It first analyzed some of the EST-related challenges that can be handled by ICT systems. A significant challenge in this area is the need for a precise understanding of the area, in the form of a common and formalized model, i.e., an ontology of the area, for which the paper presented the first step. The paper then discussed the ontological foundations for the development of ontologies related to the EST, a first example and a vision for practical implementation, and the associated challenges encountered with their development.

The fourth paper “Optimized Renewable Energy Forecasting in Local Distribution Networks” by Robert Ulbricht, Ulrike Fischer, Wolfgang Lehner, and Hilko Donker considered the role of forecasting in integrating renewable energy sources (RES) into local energy distribution networks. Since RES are not controllable, it is essential to be able to accurately forecast the supply delivered by RES. However, a number of challenges exist, including the wide variety of RES installations, and the non-availability of fine-grained metering data. The paper presents a generalized optimization approach for determining the best forecasting strategy for a given scenario, including the choice of forecasting model, forecasting granularity (single RES installation or aggregated view), and model parameters. The approach is tested on real-world data and directions for future research are given.

The final paper by Dalia Kaulakiene, Laurynas Siksnys and Yoann Pitarch was called “Towards the Automated Extraction of Flexibilities from Electricity Time Series.” Like the second paper, it also concerned the topic of flexibilities in energy consumption and demand, but from a different perspective. Specifically, the paper considered how to derive the available flexibility in the energy consumption of a given customer based only on metering data. The paper presented a number of approaches, ranging from basic to advanced, and requiring various amounts of background knowledge, e.g., knowledge of appliances or usage frequencies. Some of the approaches have been implemented in a software tool used in the simulation trials of the EU project MIRABEL. The paper rounded off by presenting a number of directions for future research.

4. ROUNDTABLE/PANEL

The workshop finished off with a panel/roundtable discussion on Research Challenges for Energy Data Management. The workshop organizers first suggested some important topics. First, there is currently a lack of common definitions of data and information concepts within the area, e.g., community-wide agreed-upon standard ontologies specifying common concepts. Second, there is a lack of standardization of the units of the technical architecture within smart grid systems, e.g., which types of layers exist, and what the nodes at each layer does. Further challenges include optimized forecasting and prediction techniques, seamless integration of past, present and future data, and developing scalable and robust data management techniques tailor-made for en-

ergy data management systems. In this context, the domain of energy data management is a driving force to build robust solutions combining data-intensive applications (classical analytical workloads on large datasets) and compute-intensive applications (simulations, numerical optimizations etc).

In more general terms, energy data management systems are a prime example of massively distributed systems managing large amounts of data in real-time while operating vital societal infrastructures. Thus, techniques developed within energy data management will have further applications in other demanding application domains. This impact will embrace a variety of different areas in database and information systems research. For example, on the one side, domain-specific modeling techniques can be adapted to suit other application areas as well. On the other side, optimizations at the system architecture layer are required to deal with massive amounts of time series data and allow flexible aggregation and sampling techniques. Since time series are relevant for many other domains as well, the technological impact sparked by energy data management will help to push the envelope of sophisticated data management techniques in general. In the long term, we also consider the domain of energy data management as one of the most prominent use-case of cyber-physical systems (CPS, see cyberphysicalsystems.org/) to seamlessly combine activities within the real and virtual world by an omnipresent monitoring and activity triggering mediation layer.

The roundtable discussion added further perspectives. It was mentioned that electricity consumers will change their behavior if the incentives are right, e.g., in a case from Florida, consumers changed their use significantly in return for less blackouts. In general, financial incentives is not enough, one must also look into “earthsaver points” and friendly competition with peers and neighbors. Another upcoming issue is charging electric vehicles (EVs), which can in some areas at some times exceed the available capacity. Thus, intelligent approaches for handling such flexible demand are needed. Here, a lot can be gained from analyzing and understanding user behavior, all the way down to the person and device levels. However, there is a lack of good datasets for this, also due to privacy concerns. However, open datasets would be a significant asset in this area.

5. DISCUSSION AND OUTLOOK

Summing up, if we first look at the topics of the presented papers, we note that they span a wide range of topics ranging from smart meter data representation and use, ontologies for emissions trading, forecasting

of renewable energy production, and managing flexible energy demand. Compared to the first workshop, the important issue of privacy of energy data was now addressed. The papers are generally the result of interdisciplinary collaborations, including contributions from several areas within computer science.

Next, when looking at the topics which occurred in the Call for Papers, but not within the accepted (or submitted) papers, we see that more systems-oriented topics such as data processing architectures, partitioning, caching, and replication schemes, query languages and query processing, robustness aspects are not covered. We believe this is not because the topics are not important, but rather due to the fact that energy data management is still new, and thus most systems are still in the development phase. While most papers are based on small case studies, only the keynote described large industrial case studies of already running systems. We again attribute this to the fact that smart grids are still in development.

Summing up, we conclude that there is a lot of interesting work going on in the area of energy data management, with many remaining challenges to be met. This supports the need for venues that focus on this issue. The EnDM workshop series will continue at EDBT 2014 in Athens where the 3rd International Workshop on Energy Data Management will be held on March 28, 2014³. For the 3rd edition of the workshop, it is the intention to organize a special issue of a journal for extended versions of the best papers.

6. ACKNOWLEDGEMENTS

The EnDM Chairs would like to thank all the authors of submitted papers for their interest in the workshop and the high quality of the submitted papers. We would also like to thank the distinguished PC members for their careful and dedicated work, both during the reviewing and the discussion phases. Finally, we would like to thank the Organizing Committee of EBDT/ICDT 2013, especially the General Chair Giovanna Guerrini, the Workshop Chair Kai-Uwe Sattler, and the Proceedings Chair Anastasios Gounaris. for their support in organizing the EnDM 2013 workshop.

7. REFERENCES

- [1] G. Guerrini, editor. *Joint 2013 EDBT/ICDT Conferences, EDBT/ICDT '13, Genoa, Italy, March 22, 2013, Workshop Proceedings*. ACM, 2013.

³<http://endm2014.endm.org>