Proceedings of the ASME 2012 6th International Conference on Energy Sustainability
ES2012
July 23-26, 2012, San Diego, CA, USA

ES2012-91198

DATA MANAGEMENT FOR A LARGE-SCALE SMART GRID DEMONSTRATION PROJECT IN AUSTIN, TEXAS

Kazunori Nagasawa
Department of Mechanical Engineering
The University of Texas at Austin
Austin, Texas, USA
kazunori.nagasawa@austin.utexas.edu

Charles R. Upshaw
Department of Mechanical Engineering
The University of Texas at Austin
Austin, Texas, USA
crupshaw@utexas.edu

Joshua D. Rhodes
Department of Civil, Architectural, and Environmental Engineering
The University of Texas at Austin
Austin, Texas, USA
joshdr@utexas.edu

Chris L. Holcomb
Pecan Street Inc.
Austin, Texas, USA
cholcomb@pecanstreet.org

David A. Walling
Texas Advanced Computing Center
Austin, Texas, USA
walling@tacc.utexas.edu

Dr. Michael E. Webber
Department of Mechanical Engineering
The University of Texas at Austin
Austin, Texas, USA
webber@mail.utexas.edu

ABSTRACT

This paper presents a data management scheme for the Pecan Street smart grid demonstration project in Austin, Texas. In this project, highly granular data with 15-second resolution on resource generation and consumption, including total consumption of electricity, water, and natural gas and solar generation, are collected for more than 100 homes. Furthermore, this testbed, see Figure 1, of homes represents the nation's highest density of rooftop solar PV and electric vehicles, and includes a substantial subset of homes that are highly instrumented with meters on up to 6 sub-circuits in addition to the whole-home meter. Consequently, this demonstration project generates a one-of-a-kind dataset with excellent temporal and geographic fidelity.

One consequence of this extensive dataset is that there are hundreds of parallel data streams that need to be remotely (wirelessly) collected, filtered, processed, managed, stored and analyzed to be useful for researchers. Cumulatively, they represent 100s of gigabytes of data after just a few months of collection, which represents a formidable barrier to conducting research.

In partnership with the Texas Advanced Computing Center (TACC), which is an NSF-sponsored cluster of supercomputers at UT-Austin, a data collection and management scheme has been developed. For storing the data, we have built a single column oriented database that so far has shown tremendous performance benefits. This paper shows the data schema, an example of MySQL query, and a developed program for rapid and auto-

Figure 1: Mueller district of the project testbed in Austin, Texas includes approximately 100 homes that are highly instrumented [1].

mated data extraction, analysis and display. We expect that the findings of this work will be beneficial to researchers interested in grid-scale data management.

INTRODUCTION

Measuring home energy use data is significant for reducing energy consumption and impact on the environment. Started in the 1972, the Twin River project was one of the first quantitative studies on home energy consumption and conservation [2]. In this project, environmental and building envelope data, electricity and natural gas consumption data, and behavioral data are captured for determining energy usage profiles. As the result of this five-year field study of residential energy use, it was found that daily feedback lowered energy use about 10 to 15% and user behavior had the largest impact on consumption [2–4]. At that time, there was no Internet in residential homes so the data were collected over telephone lines and recorded on a magnetic tape recorder and a teletypewriter.

Currently, given the development of information and communication technology, data are easily collected over ubiquitous integrated networks such as the Internet. An early example of such data collection was the Laredo project - one of the first smart grid demonstration projects [5]. In this project, a hybrid fiber-coaxial network was built for on-demand automated meter reading. The system also enabled each meter to act as a load research meter by transmitting that information over the network to database. However, collection and control of sub-circuits proved difficult and deploying a two-way broadband network was deemed too high of a cost [5].

In the United States, approximately 14 smart grid demonstration projects have been conducted [6]. These projects are attractive for the energy industry due to their potential for increased reliability, availability, and efficiency as compared to the current electrical grid. The Pecan Street smart grid demonstration project in Austin, Texas is one-of-a-kind project due to its customer-focused approach. Pecan Street Inc. is a public-private partnership, whose members include: The University of Texas at Austin, the City of Austin, Austin Energy, the Greater Austin Chamber of Commerce, the Austin Technology Incubator, and the Environmental Defense Fund [1]. This project seeks to show the utility of the smart grid to the homeowner due to the highly visible rollout of integrated end-user systems such as home energy management systems (HEMS) that can manage the energy usage of numerous consumer systems including appliances, consumer electronics, and electric vehicles. The project will deploy HEMS from up to five companies that will be able to test their systems on a real grid with real customers in real homes. Many of these homes will also have electric vehicles, home energy storage equipment, and rooftop solar PV systems. To our knowledge, this testbed of homes represents the nation's highest density of rooftop solar PV and electric vehicles, and includes a substantial subset of homes that are highly instrumented with meters on up to 6 sub-circuits in addition to the whole-home meter. As the number of connected devices and homes increases, data management will be a formidable barrier to conducting research. However, given the novelty of smart grids, little or no standards exist to manage large volumes of data for these types experiments. Therefore we are developing our own unique approach and are making it public with the hope that it will be useful to others as they confront similar data challenges.

DATABASE

In the Pecan Street smart grid demonstration project, there are three distinct data acquisition systems installed. While there is some overlap of installed systems, not every home has all systems.

Incenergy/Sequentric systems monitor current for both whole home main phases and 6 sub-circuits. Two encoders are deployed at each home that each measure 4 circuits. The encoders sample the current on each circuit at 1 kHz. The current is averaged every 2 seconds and sent to the gateway which is polled every 15 seconds. The data undergoes additional processing once it is polled to reflect an estimated average Watts for each polled 15-second interval. Itron water and gas systems simultaneously collect cumulative gallons of water and cubic feet of natural gas, respectively. These meters communicate with encoder receiver transmitter (ERT) enabled Incenergy/Sequentric gateways and are recorded at the same 15-second interval as the electric data.

eGauge systems, which are at this time only installed on homes that have solar PV installations, sample frequency, voltage, and current waveforms for monitored circuits with a sampling frequency rate greater than 2 kHz. The system calculates the root-mean-square (RMS) of current, RMS of voltage, cal-

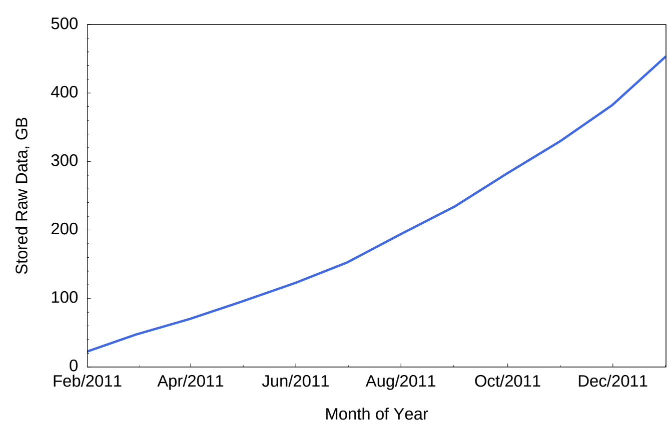


Figure 2: Stored raw data in the database in a year.

culates average real power, apparent power, and other measurements while considering phase differences between currents and voltages. These data are then saved in 1-minute interval.

The electric meter from Austin Energy (the local electric utility) system collects the real power retrieved from the grid, sent to the grid, and generated during each 15-minute interval.

In addition to the granular energy and water data stored on the database, home energy audit, tax appraisal, weather, solar, Electric Reliability Council of Texas (ERCOT) electricity price, and home energy survey data are also stored in the database.

TACC, which is supported by the National Science Foundation, the University of Texas at Austin, the UT System, and grants from other federal agencies feeds all the collected data into a single column oriented database. We are using the InfiniDB running in MySQL [7]. The filesystem for the database is a RAID0 setup with two 240 GB SSDs. MySQL is an open source relational database management system (RDMS) that delivers a fast, multi-threaded, multi-user, and robust SQL database server [8]. Data feeds from the various HEMS are gathered in a variety of structured formats, primarily some form of CSV or XML file.

Figure 2 shows stored raw data in the database from February 1, 2011 to January 31, 2012. It is shown that the steady increase is a result of more and more homes and HEMS coming online as the project moves forward, and the stored raw data reached approximately 454 GB or 2.7 billion raws after one year of study.

DATA ACCESS

Figure 3 shows a data access sequence expressed by the unified modeling language (UML). TACC provides instructions to users on proper access to the database, including user name, password, hostname, and port number. The connection information allows users to connect from any MySQL capable client to the database. Thus, users may have access to run queries against the tables by using GUI clients such as MySQL Workbench and SQLyog. These clients offer the ability to extract subsets of the data into CSV, HTML, and XML for further analysis.

The following is a sample query of the database:

```
SELECT DataID, DATE_FORMAT(SUBTIME(timestamp,
'05:00:00'), '%Y-%m-%d %H:%i:00') AS
correct_time, consumption, generation, grid
FROM pecanstreet_dev.egaugenew
WHERE timestamp BETWEEN '2011-08-29
05:00:00' AND '2011-08-30 04:59:59' AND
DataID = 8084
ORDER BY timestamp ASC
LIMIT 86400;
```

This sample query extracts data id (unique participant identifier), the CDT corrected time (the original timestamp is in UTC), whole home consumption, solar generation, and demand from the grid for the selected home on August 29, 2011.

Figure 4 shows a part of the result of the sample query on MySQL Workbench. Users could export the result as various file extensions for further analysis. However, for executing very large queries, such as data for all the homes, MySQL Workbench's built-in export capabilities don't suffice, but MySQL Connectors enable users to connect and execute MySQL statements from another languages or environments, including Java, Perl, Python, PHP, Ruby, C, and C++ [8]. As a result, we have developed a Java program that can connect to the database, execute queries and generate graphs via integration with Gnuplot, which is a graphing utility for Linux, Windows, OSX, and many other platforms [9].

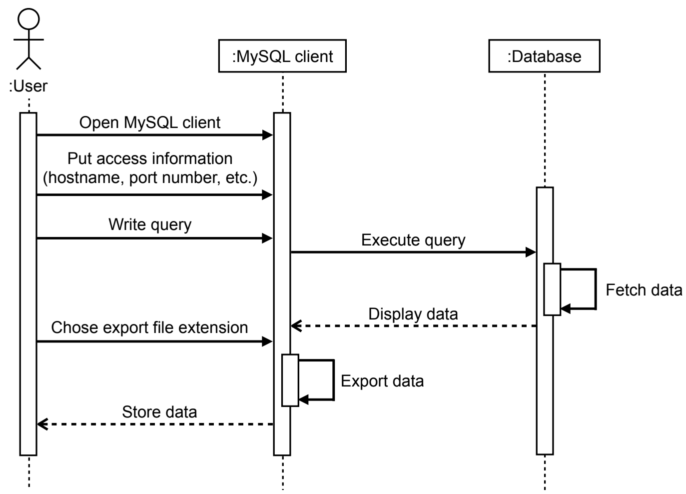


Figure 3: A data access sequence diagram through using MySQL client.

DataID	correct_time	consumption	generation	grid
8084	2011-08-29 00:00:00	5.779633333	0.000000000	5.779633333
8084	2011-08-29 00:01:00	5.746133333	0.000000000	5.746133333
8084	2011-08-29 00:02:00	5.791033333	0.000000000	5.791033333
8084	2011-08-29 00:03:00	5.578850000	0.000000000	5.578850000
8084	2011-08-29 00:04:00	5.357733333	0.000000000	5.357733333
8084	2011-08-29 00:05:00	5.332050000	0.000000000	5.332050000
8084	2011-08-29 00:06:00	5.308083333	0.000000000	5.308083333
8084	2011-08-29 00:07:00	5.294916667	0.000000000	5.294916667
8084	2011-08-29 00:08:00	5.322566667	0.000000000	5.322566667
8084	2011-08-29 00:09:00	5.348550000	0.000000000	5.348550000
8084	2011-08-29 00:10:00	5.343900000	0.000000000	5.343900000
8084	2011-08-29 00:11:00	5.352250000	0.000000000	5.352250000
8084	2011-08-29 00:12:00	5.331050000	0.000000000	5.331050000
8084	2011-08-29 00:13:00	3.580366667	0.000000000	3.580366667
8084	2011-08-29 00:14:00	1.401383333	0.000000000	1.401383333
8084	2011-08-29 00:15:00	1.620150000	0.000000000	1.620150000

Figure 4: A result of the sample query.

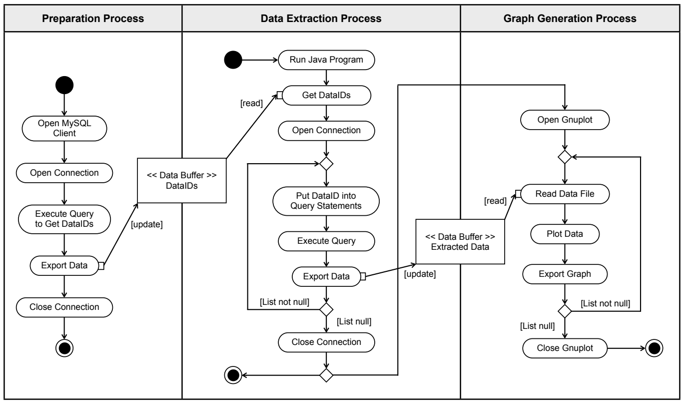


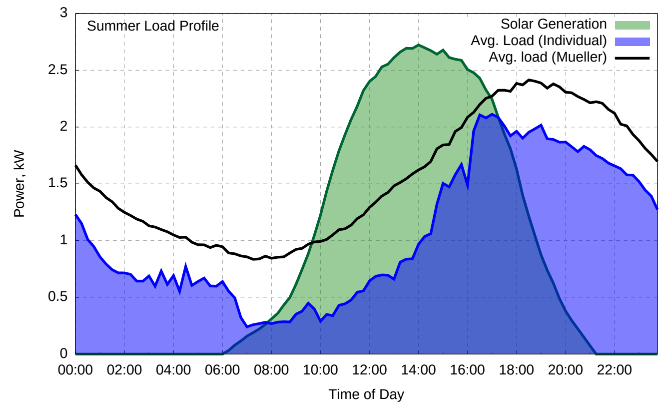
Figure 5: An activity diagram of automatic graph generation.

Figure 5 illustrates an activity diagram of automatic graph generation expressed by the UML. This activity diagram consists of three processes: preparation, data extraction, and graph generation. In the preparation process, the user will export the DataIDs using the MySQL client. In the data extraction process, the Java program initially reads the DataIDs and stores them in an array list for a looping condition. Each DataID is read line by line and input into a query statement routine until the list of the DataIDs becomes null. At the end of every routine, the program exports the extracted data file. After finishing the routine, the program opens Gnuplot. This graph generation process also runs the same routine and generates graphs.

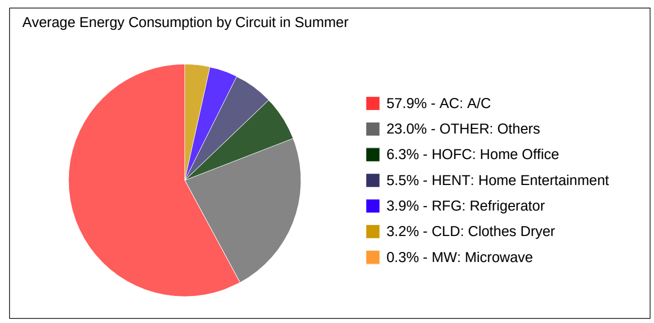
INPLEMENTATION

After one year of data acquisition and analysis, we are attempting to deliver an annual report to individual participants. There are 100 homes for the baseline study, which carried out from February 2011 to February 2012. Each report consists of 8 unique graphs, such as seasonal load profiles, to convey to residents how much energy they have consumed over the entire time period. This resulted in about 800 graphs that needed to be generated, hence our automated process.

Figure 6 shows examples of the graphs. Figure 6 (a) illustrates the summer load profile of a home. The green line represents averaged solar generation, the blue line represents individual load profile, and the black line represents averaged load profile in the Mueller district (all homes in the study). Figure 6 (b) illustrates an example of the average energy consumption broken down by circuit during Summer for a home. The developed program was robust and capable of generating these graphs automatically.



(a)



(b)

Figure 6: An example of summer load profile of a home (a) and an example of average energy consumption broken down by circuit during Summer for a home (b).

FUTURE WORK

This work has been performed over the last year as away to solve the enormous data challenge that face large-scale experiments of this kind. Future work in this area will quantify the accuracy of data between Incenergy/Sequentric, eGauge, and Austin Energy systems, including data drop-outs, repeated samples, and instrumented errors. Once the data has been validated and the management and query systems have been finalized, these systems will be used to conduct analysis of energy use in the built environment. Additionally, energy usage feedback to residents is effective for reducing energy consumption [10], so the non-trivial task of data visualization will be an important role for analyzing the effect.

CONCLUSIONS

This work presented a data management scheme that was developed for the Pecan Street smart grid demonstration project in Austin, Texas. This schema demonstrated how to collect, filter, process, manage, store and analyze large volumes of data to be useful for research analysis. Although the schema was tailored for this project, it is applicable to any other smart grid implementation.

ACKNOWLEDGMENT

Kazunori Nagasawa's portion of this work was funded by 2011/2012 Nihon University Overseas Scholarship Student System in Japan. Other funding comes from the DoE and the Doris Duke Foundation via Pecan Street Inc.

REFERENCES

[1] Pecan Street Inc., 2010 [Online]. "2010 pecan street report of recommendations". http://www.pecanstreet.org/wordpress/wp-content/uploads/2011/08/Pecan_Final_Report_March_2010.pdf, March.

[2] H., R., and Socolow, 1978. "The twin rivers program on energy conservation in housing: Highlights and conclusions". *Energy and Buildings,* ***1***(3), pp. 207 – 242.

[3] Harrje, D. T., and Grot, R. A., 1978. "Instrumentation for monitoring energy usage in buildings at twin rivers". *Energy and Buildings,* ***1***(3), pp. 293 – 299.

[4] Seligman, C., Darley, J. M., and Becker, L. J., 1978. "Behavioral approaches to residential energy conservation". *Energy and Buildings,* ***1***(3), pp. 325 – 337.

[5] Babin, M. L., 1998 [Online]. "The laredo pilot program". oriontechnologyinc.com/swf/thelaredoreport.pdf, March.

[6] SmartGrid.gov, [Online]. "Project information and locations". http://www.smartgrid.gov/recovery_act/project_information.

[7] Calpont, [Online]. "Infinidb". http://www.calpont.com/products/infinidb.

[8] Oracle Corporation, [Online]. "Mysql 5.6 reference manual". http://dev.mysql.com/doc/refman/5.6/en/index.html.

[9] Williams, T., and Kelle, C., 2011 [Online]. "gnuplot 4.4 an interactive plotting program". http://www.gnuplot.info/docs_4.4/gnuplot.pdf, November.

[10] Ehrhardt-Martinez, K., Donnelly, K. A., and Laitner, J. A. S., 2010. "Advanced metering initiatives and residential feedback programs: A meta-review for household electricity-saving opportunities, report number e105". *American Council for an Energy-Efficient Economy*, June.