Grid Analytics: How Much Data Do You Really Need?

Stanley E. McHann, Jr.
Member, IEEE
4400 Buffalo Gap Rd. Suite 5150
Abilene, TX 79606
stan.mchann@milsoft.com

Abstract -- Electric utilities are experiencing a technology transformation, which includes the deployment of distributed intelligent devices, two-way communications systems, and information technologies to enable greater monitoring and control of their distribution systems. These technologies will increase the volume of data flowing into a utility's Information Systems, which in turn will need to be stored and managed. The amount of data that can be generated from AMI/AMR, SCADA, and other intelligent devices can be significant. One strategy, which seems to be a popular option, is to collect all of the data possible and figure out what to do with the data at a later date. As the amount of data increases that needs to be stored, there is a corresponding increase in Information Technology infrastructure, skill sets, and cost. If collecting all data were possible, it is doubtful the data would be stored or structured in a way that would be useable in the future.

An alternative to capturing and storing all data possible is to use Information Engineering methodologies to focus on what data is needed for a given task or application. Information Engineering is defined as a set if interrelated disciplines that are needed to build a computerized enterprise based on information systems. The focus of Information Engineering is the data that is stored and maintained by computers and the information that is distilled from this data. [1] It is foundational to understand that data and information are terms used interchangeably, but are distinctly different terms with different meanings. A kWh is a piece of data that is useful information when added to other data such as meter number, account name, or service location. The amount and kind of data required, as well as the necessary time frames, (i.e., historical, real-time, predicted) depend upon the applications, which may include:

1. Geospatial Information System
2. Meter/billing information
3. Asset management
4. Work and workforce management,
5. Network modeling and analysis for planning (e.g., voltage drop, power flow, short circuit, arc flash, contingency studies, reliability metrics, loss analysis, protective device coordination)
6. Operations (outage management, fault location, Volt/VAR control, power quality, demand management, distributed generation and storage)
7. Real-time, active grid management, grid analytics.

This paper will discuss how to apply Information Engineering principles to turn data into useful information for a utility as an alternative to the "Big Data" approach to capturing and storing data. Using kWh, and Voltage as examples, we will outline the Information Engineering process to turn these data elements into useful information. Once you have the information engineered, the next step is to use data management methodologies to manage the data that is being gathered. Data Management is a detailed topic on its own and will not be covered on this paper. What data has been captured needs to be managed across the enterprise.

***Index Terms*—Advanced Metering Infrastructure (AMI), Data Analytics, Smart Grid, Smart Meters, SCADA (System Control and Data Acquisition), Distribution Analysis, Distribution Automation, Distribution State Estimation, Load Flow Analysis, Real Time Distribution Feeder Analysis, Active Grid Management, Big Data**

I. INTRODUCTION

There is a lot of discussion in the utility industry about "Big Data." The information technology definition of big data is it is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. [2] There are some examples of Big Data applications that are worth mentioning to give context to how this technology is applied.

Science is on one extreme end of the Big Data spectrum with the Large Hadron Collider (LHC) near Geneva Switzerland that conducts experiments which represent approximately 150 million sensors that deliver data at 40 million times per second which equals 600 million collisions per second. Filters are applied to the data and about 99.999% of these data streams from the sensors are not recorded, but this still results in 100 collisions of interest per second. [4][5][6] Only 0.001% of the sensor data stream is actually used by the four experiments presently ongoing at LHC. This represents 25 petabytes of data annually, as of 2012, before data replication. If all sensor data was recorded it would represent over 150 million petabytes annually, or almost 500 exabytes per day. This does not include replication. To appreciate the magnitude of 500 exabytes a day, it is equal to 500 quintillion ($5x10^{20}$) bytes per day. It is estimated that this data storage requirement would be over 200 times greater than all other sources combined in the world.

A more moderate example of large data sets is the Sloan Digital Sky Survey (SDSS), which started collecting astronomical data in 2000. In the first few weeks of operation, SDSS collected more astronomical data than all the data collected in the history of astronomy to date. SDSS collects data at approximately 200GB per night and has amassed 140 terabytes of information. The Large Synoptic Survey Telescope (LSST) is anticipated to come online in 2016 replacing SDSS. It is anticipated that the LSST system will accumulate 140 terabytes of data every 5 days. [7]

In the commercial sector Wal-Mart handles an estimated 1 million customer transactions per hour, which is then imported into databases that are believed to contain more than 2.5 petabytes of data. 2.5 petabytes of information is about 167 times the information contained in all the books in the US Library of Congress. [3] This is a much more manageable amount of data than what the scientific community captures, but the amount of data is still significant.

At the foundation of the data explosion in electric utilities is the meter read. Most customers had their meter read 12 times per year or once a month. Early AMR (Advanced Meter Read) systems were designed to bring a kWh read once a month for the purpose of providing the utility customer a bill. Today, 15-minute read intervals are what is expected of an AMI/AMR (Advanced Meter Infrastructure/Advanced Meter Read) which is 96 meter reads per day per meter or approximately 2880 reads per month (30 day month) per meter. This equals an increase of 287,900% in just kWh data.

For this example a dial on a meter is equal to 1 byte of information and a meter has 5 dials. Every meter read is equal to 5 bytes. Also in our example, the grid has 50,000 meters that are installed. The number of meters is not uncommon in a utility COOP. So 50,000 meters multiplied by 5 bytes per meter will equal 250,000 bytes of information every time we accomplish a meter read or kWh.

Time Interval	Number of Meters	Data	Amount of Data
1 Day	50K	5 Bytes	24MB
1 Month	50K	5 Bytes	720MB
1 Year	50K	5 Bytes	8.76GB

The data in the chart above does not include demographic information such as the customer name, address, phone number, meter number, and service location, which would also be in the database. The demographic information may vary up and down depending on the number of customers at any given time, but the meter data grows substantially when the utility choses to receive meter reads every 15 minutes for all meters. The numbers in the chart also do not take into account data replication or mirroring in Storage Area Networks (SANs), backups, or data being stored in other applications. For example, the Customer Information System will have demographic and meter data along with the Outage Management System, which further increases the storage required for the same data.

Austin Energy in Austin, Texas has implemented 500,00 smart meters. Austin Energy's smart meters send data every 15-minutes, which requires 200TB of storage with disaster recovery redundancy factored in. To manage the data takes more disk space than just storing the information. If Austin energy moves from 15-minute to 5-minute interval data their data storage needs will grow to 800TB.

Austin Energy is certainly not a rural power system, but it is an example of how much data that smart meters can generate that has to be managed and hopefully used effectively. There are certainly two extremes in collecting smart meter data. One extreme is to gather everything and sort out later how it might be used, or use Information Engineering concepts to make decisions on what data is important.

A. Information Engineering – A Foundational Concept

Information Engineering is used to help with the problem of data designed by applications for use by applications. The Customer Information System (CIS) may have data in it's relational database that is easily consumable by the CIS application, but not easily consumable by the Outage Management System without some level of systems integration to map the data between the systems. Data is easily siloed within applications and in some cases, siloed in ways that even limit the data being used by another part of the same application. Add in another level of complexity, the departmental management of a utility, and the data within an application can be further siloed due to organizational boundaries.

MultiSpeak and CIM (Common Information Model) are two methods by which utility applications can share data. This helps solve the problem of sharing data between applications, but does not solve turning the data into information. Hopefully, the applications are doing a good job of this, but in many cases, the data in different application databases need to be tied together to give the utility useful information about load, line losses, voltage drop, outages, and other essential operational information.

ANSI C12.19 defines end device data tables that are used in AMR meters. These data definitions are a good start for sharing data between AMR/AMI vendors. Many utilities have more than one AMI/AMR vendor and having this data standard helps the utility use the meter data across the utility to analyze problems and issues without the further complexity of having to translate or convert data definitions between two different metering systems.

Information Engineering is taking the data the utility may have and building information from this data information that enhances or improves the utilities operations. GIS, CIS, OMS, Engineering Analysis (EA), Asset Management Systems (AM), all have data in them that can be consumed by the respective application. By implementing a MultiSpeak or CIM interface, data can be shared among many of these applications and turned into useful information. For example, the circuit model in EA is essential for OMS to understand how the gird is put together so it can accurately show outages. Also, the CIS sends OMS it's consumer information so that OMS knows who is out of power, not just what is out of power. The OMS system is a consumer of data sets from different applications and turns that data into useful information to help a utility manage outages.

Many utilities rely on their software vendors to do the data integration needed to generate the information the utility needs to manage it's operations. Information Engineering methods can be implemented by a utility to further enhance the value of their data to lower operational costs and improve customer service. An Information Engineering process will help a utility identify all data sources, timeliness of the data, and what data is really needed to solve the utilities engineering and operations needs. Of course there is the get all data and store it for use sometime option.

B. Networks – Power Line Carrier, RF Mesh, IP

Utility networks have reached out from inside the Information Technology or Data Processing department to engineering, mapping, operations, substations, and sensors and meters. Through the implementation of different communications technologies, more data can make it back to the IT department and be used effectively by the different departments in the utility. The meter may have every C 12.19 register full of data, but depending on the network fabric, Power Line Carrier (PLC) or Radio Frequency (RF) Mesh, it may not be possible to get the data back to the head end system. These networks do not have unlimited bandwidth nor a long transmit time when the power goes out. In a RF Mesh network, sending a power out notification is a last gasp measure that is over once the capacitor is drained that feeds the transmitter. Once the Information Engineering is completed, take a look at the meter network and make sure it is capable of delivering the data that may be required.

At customer site had just been upgraded to a new RF Mesh network fabric with new meters. They sustained a major outage involving about 1800 meters. The customer had a SCADA system, an OMS system, a EA and GIS system, on a common data bus with the new head end system for the meters. Of 1800 meters, approximately 1450 meters reported an outage. Why? The RF Mesh could not transport all messages to the head end system. However, the OMS system and an accurate circuit model from the EA system, had data integration to the SCADA system, and it also had integration to the CIS system. The OMS system was able to identify the fault, realize that it did not have all the messages from the AMI/AMR head end system, but predicted the outage correctly so the customer could dispatch crews right away. Through the Information Engineering the customers' software vendor had accomplished, the utility had excellent situational awareness in the storm even though not all the messages could make it through the RF Mesh network.

C. kWh

Getting kWh readings to bill customers over an AMI/AMR system is commonplace today. Some customers get this information once a day and others get the kWh data every 15 minutes. Creating a bill every month without having the need to have meter readers was how AMI/AMR systems found their way into most utilities. Load Allocation is another analysis that we can use kWh for. Using REA (Rural Electric Association) Bulletin 45-2, we can setup a Load Allocation study using the kWh data we have in the CIS system. The chart below shows the applications and data that will be needed to set up the Load Allocation study.

Application	Data	Information Delivered
CIS or AMI/AMR	kWh	kWH by service location
WindMil ™	Circuit Model	Load Allocation

D. Voltage

Voltage is available as a data field in most AMI/AMR meters. Most customers configure the AMI/AMR networks to bring back billing information first, outage information second, and then anything else after these two priorities are met. It is very important to understand that if a voltage reading is sent from the meter to the head end system, another piece of information is not being transmitted from the meter to the meter head end system. These trade offs should be an Information Engineering exercise to determine which service locations need to send voltage data back, a "bell weather" meter, that gives the utility a good view of voltage from the substations to the end of the feeders. A trade off analysis needs to be made as in most cases, the voltage data can come back, but there may not be a kWh reading for that day or hour depending on packet setup.

As a general rule, voltage is not needed from every meter, every 15-minutes, to create a good voltage drop study. Do the Information and Electrical Engineering analysis to decide which meters need to be sending the information.

E. Summary

Many of the concepts in this paper are a series of papers or books in themselves. Information Engineering is a good example of this. There have been many books written on the topic. Today, because of the large data sets that can be created, Systems Analysts that performed the Information Engineering function are now being called Data Scientists.

Before making an investment in the tools to handle "Big Data" or even try to bring back significant amounts of data form an AMI/AMR or SCADA system, take the time to evaluate what data your utility really needs, where it will be stored and managed, and how it will benefit the organization.

So how much data do you really need to do an effective job at Grid Analytics? That will vary from utility to utility, but I would suggest that if you look at the data you already have, you would have a very capable analytics system with a modest investment and utilizing proven methodologies such as Information Engineering.

F. Author information

Stanley E. McHann, Jr. is Chief Technology Officer at Milsoft Utility Solutions, Inc. He has served in engineering, and senior technical staff at Dell Computers, Tandem Computers, 3Com, Cadence Design Systems, and Landis + Gyr prior to joining Milsoft.

ACKNOWLEDGMENT

The author gratefully acknowledges the contributions of Matt Goodman and the staff at Itron for their technical support on meters and networks for this paper.

REFERENCES

Periodicals:

[1] Brumfiel, Geoff (19 January 2011). "High-energy physics: Down the petabyte highway". *Nature* 469: pp. 282–83. doi:10.1038/469282a.
[2] "Data, data everywhere". *The Economist*. 25 February 2010. Retrieved 31 January 2012.
[3] "Data, data everywhere". *The Economist*. 25 February 2010. Retrieved 31 January 2012.

Books:

[4] Martin, James, *Strategic Information Planning Methodologies, 2nd ed, Prentice-Hall, 1989, p. 24*. ISBN 0-13-850538-1.
[5] White, Tom (10 May 2012). *Hadoop: The Definitive Guide*. O'Reilly Media. p. 3. ISBN 978-1-4493-3877-0.

Other References:

[6] *CERN-Brochure-2010-006-Eng. LHC Brochure, English version.*. CERN. Retrieved 20 January 2013.
[7] *CERN-Brochure-2008-001-Eng. LHC Guide, English version.*. CERN. Retrieved 20 January 2013.