Operationalizing Smart Meter Data with "Medium Data" Tools

Dennis J. Nelson, Andrew J. Berrisford, Scott Albrechtsen, and Cathy Rupp
BC Hydro - Power Smart

ABSTRACT

While many utilities will have to wait years to have operational reports from their Smart Meter investments provided with "Big Data" solutions; by using an approach we call "Medium Data" solutions, BC Hydro's system planning and load management teams can get real knowledge of where the peaks come from much sooner. From a data sourcing and handling perspective, this is an interim stage between traditional Load Research's "Small Data" sampling and modeling, and the future "Big Data" state, where one would simply aggregate the hourly loads for all meters.

The "Medium Data" approach works with a daily extract of a "super sample's" hourly data for 60,000+ meters which includes all large commercial accounts. That information is integrated with the daily register reads for the whole population of ~1.8M meters. An existing tool built in Statistical Analysis System (SAS) software referred to as the Distribution Load Shape Estimation (DLSE) model can provide excellent estimates of hourly demand (billed sales) by feeder, substation, and region; for segments such as building types, rate classes, and communities. This information moves to an SQL database with a simple viewer, each day.

The DLSE data sets also serve as an efficient repository for historical aggregated hourly, daily and monthly consumption data, enabling simpler decisions on smart meter data retention strategies.

Background

This paper begins by reviewing some of the history of interval data modeling in BC Hydro and the evolution of the Load Research function, and then describes how the utility will make use of smart meter data, while awaiting the "Big Data" solutions. A summary of benefits from the "Medium Data" and examples of outputs are presented at the end.

Since 1991, BC Hydro's Load Research (now Load Analysis) team has collected load research data (15-minute interval kWh) from interval data recorders at a representative sample of ~1,500 customer locations, many by phone line and the rest manually downloaded by Load Analysis staff or contractors. The Load Research "Post-stratified Ratio Expansion" process uses the sample 15-minute data along with population billing data (from the SAP/CCS billing system) and distribution topology (the relationship between meters, transformers, phase and feeders) data to produce hourly load shapes by Rate Class or Site Code (building types at account level). The Rate Class load shapes are used for estimating energy (kWh) and peak demand (MW) figures for the Regulatory Fully Allocated Cost of Service (FACOS) studies, Rate Design, and Load Forecasting's demand forecasts. The Site Code hourly load shapes (8,760 hours per year) support Distribution Planning's work with the DLSE tool developed by Load Analysis staff, with direction from Dr. Roger Wright and representatives of BC Hydro's Distribution Planning, Load

Forecast and Rates teams. The load shapes are made available to users via DLSE web application, described in more detail later.

Load Analysis also has maintained the company's monthly billing history for 20 years, accessed by over 500 BC Hydro users via the AHR application (developed in 2004), and all the data associated with the statistical analysis and reporting of energy information required by clients throughout the company. Along with these primary data sources, the team collects and integrates data from several other sources, including the Outage Management System and the SCADA System (PI). Figure 1 illustrates the range of internal and external clients served by the team. This has led to the evolution of a unique utility analytic team familiar with data from many systems used in the company, which has been taking on more of the role of a Business Intelligence team for the past few years. The previous Load Analysis computing requirements might be termed "Small Data", with datasets in the Gigabyte range, using a total of 6 TB of SAS working/storage space.

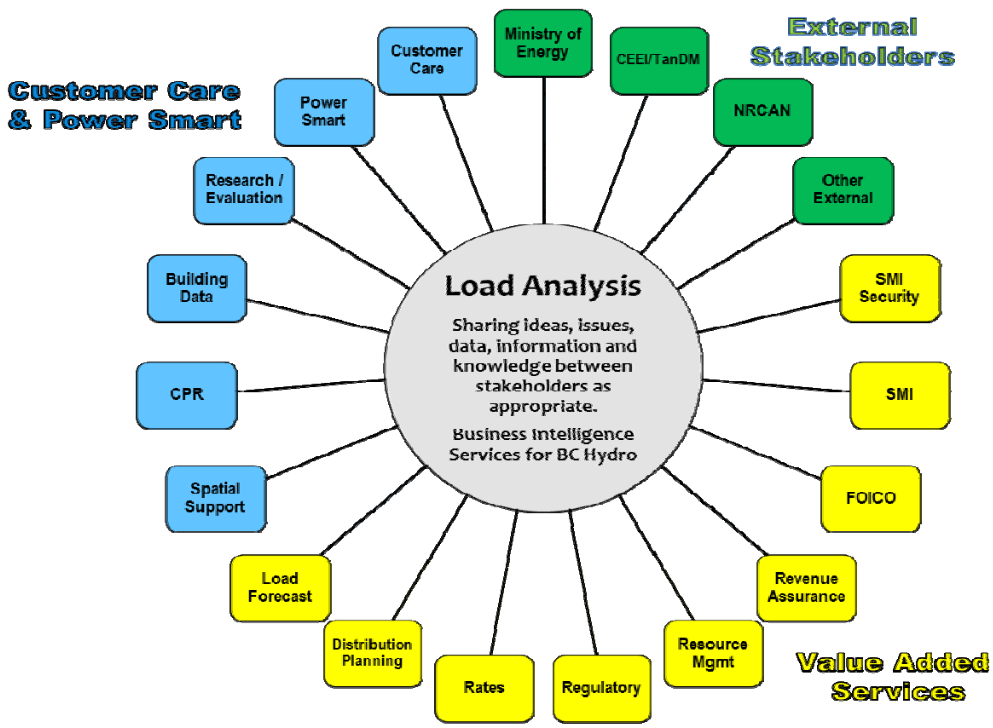


Figure 1. BC Hydro load analysis group stakeholders.

Evolution of "Load Research"

The Load Analysis Group has been intimately involved with the Smart Metering & Infrastructure (SMI) Project, from its conception several years ago based on our proof-of-concept work in 2005 and through several "pilots", then to the definition (2007), procurement (2008 and again in 2011), and implementation (2012) of the project presently nearing completion in 2014. The SMI project has major implications for many areas of the organization, including Load Analysis, which will have to adapt to the new SMI environment, with new data sources, new tools, and more clients with new analytics requirements. In conjunction with BC Hydro's IT staff, the SMI team, and the Revenue Assurance group, several Load Analysis staff have been working on the "Big Data" efforts, developing the Energy Analytic Solution (EAS) project within the new "Massively Parallel" Greenplum computing environment. The EAS will form the core of a new corporate area of interest, called Enterprise Analytics, with the goal of providing analytics solutions built from all the major data systems, to all analytics users in the company. The advent of SMI and EAS will require the evolution of the Load Analysis group to a more comprehensive Business Intelligence role over the next two to three years. This evolution process and the associated data approaches are shown in Figure 2.

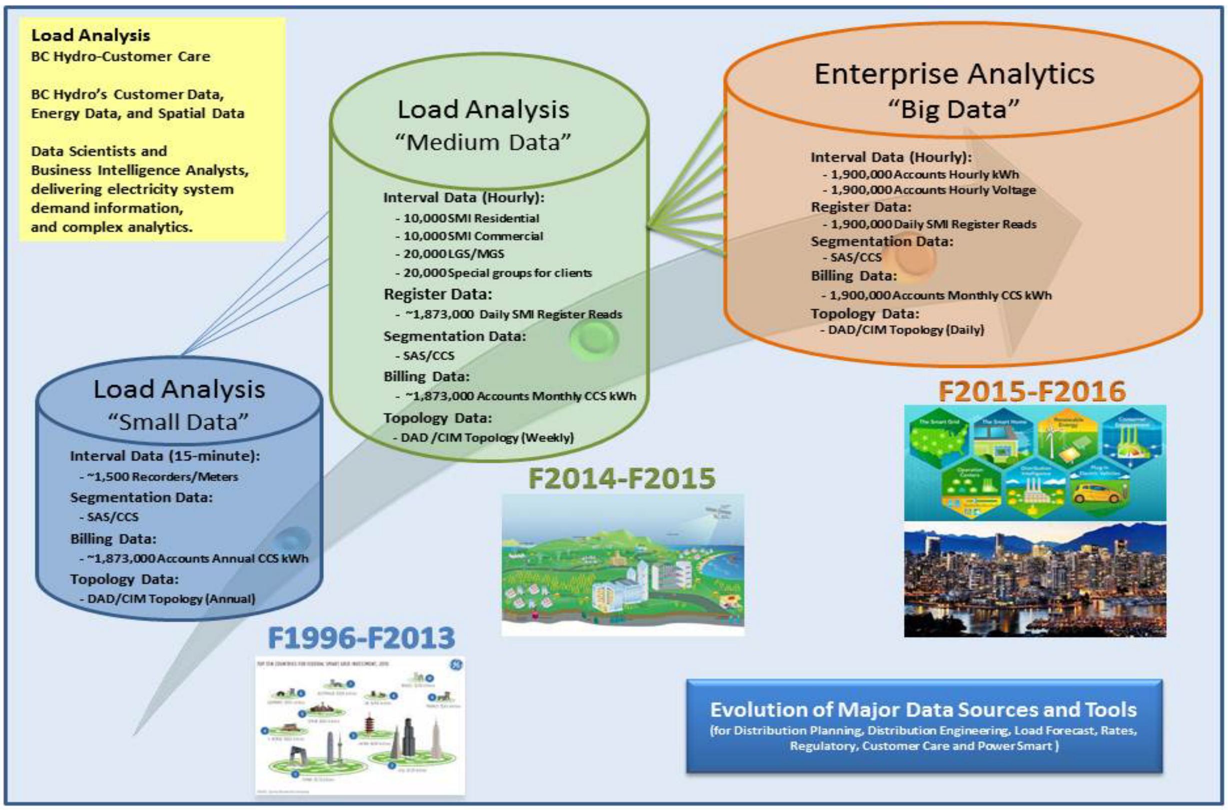


Figure 2. The road to "big data" analytics.

The first stage of the transformation is completed. We have retired the fleet of ~1,500 recorders and specialty meters, as SMI meters were installed at those locations. Interval data from SMI meters now provides the team with a larger sample of hourly (hourly is the finest granularity we are allowed for residential accounts by the Provincial Privacy Commissioner) interval data; ~40,000 meters for FACOS and DLSE and another ~20,000 meters to support our work for Revenue Assurance, Power Smart Evaluation and Distribution Planning.

This second (interim) stage will increase the size of accessible interval metered data by a factor of about 40. By late 2014 or 2015 we expect that all the SMI interval data for ~1.8 million meters will be easily available for analytics, by specific teams within BC Hydro, including the Load Analysis Group.

The third stage of the transformation will then take place. Most of the remaining sample-based analysis may be replaced by summing actual metered loads, and the "Big Data" Enterprise Analytics environment will make energy, peak demand, and load shape information readily available to the whole company, with a little help from the transformed Load Analysis Business Intelligence team.

Towards a "Medium Data" World

As described in Figure 2, the Load Analysis team will in the short-term, focus on "Medium Data", and relevant tools, to provide value added information from SMI, until the "Big Data" environment is fully developed and available. At that point the data flows to the tools described below will change, and some new solutions will be developed.

The "Medium Data" Flow

Every eight hours the smart meter data is uploaded via the Automated Data Collection System (ADCS) to the Itron Meter Data Management System (MDMS) (see Figure 3).

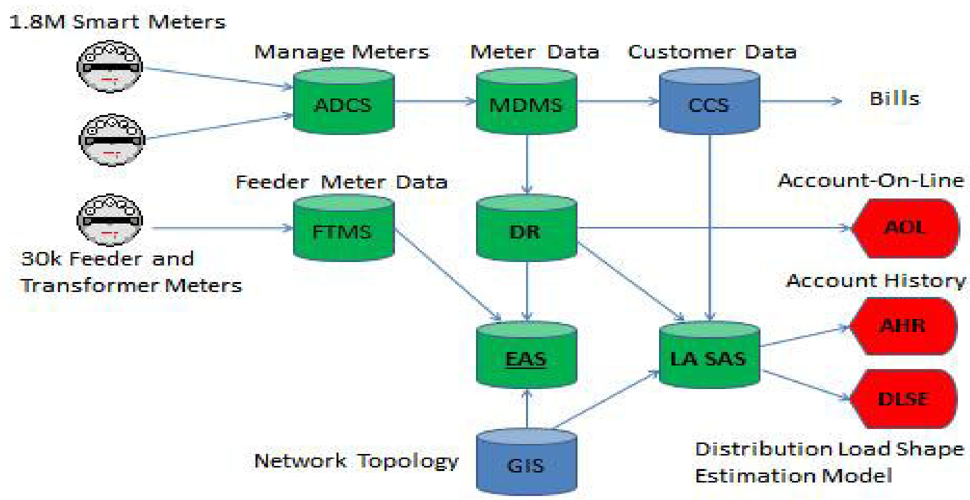


Figure 3. Data movement in the SMI environment.

The midnight uploads contain the final register reads for the day. The MDMS does data validation and estimation, and uploads SMI data (hourly KWh, Vh, kvarh, and register reads) to the "interim" Data Repository (DR) each morning, mostly to service the customer portal views of hourly usage, accessed through Account On Line (AOL). Customers can see their interval consumption, and daily consumption the day after, through the portal. All of the SMI consumption data also flows from the MDMS to EAS, daily. The Data Repository stores all the SMI energy and voltage data but from a production perspective, can only really handle the customer portal work. Now, daily updated files formed in the Data Repository can be extracted each night, by Load Analysis, and moved into our new Enterprise Analytics SAS environment to replace and enhance the data from the retired Load Research recorders. In the flow diagram below, we include the Feeder and Transformer Metering System (FTMS), although it is not included in our "Medium Data" tools.

Daily Register Reads

To operationalize the available, rich SMI interval data in the near-term, Load Analysis is accumulating a new and distinct stream of data. The daily register reads are to be updated each night. (The whole of the available history of SMI daily register reads has been acquired and will be retained by Load Analysis.) This provides daily billing/usage data for ~1.8+ million meters. This can be used for:

- Load Forecast/Finance accrual systems replacement.
- Quick calendarization of billing data, with all the weather effects captured properly.
- Sales tracking by segment, by day.
- Improving a vast array of forecast and analytic efforts.
- Fine-tuning the account level heating and A/C codes.
- Enabling flexibility in a data retention strategy.

Hourly Interval Data

A "super sample" is being collected and will be updated nightly. Hourly interval data from all over 35kW demand-metered accounts (~20,000), as well as the official Load Research sample (for FACOS) of ~10,000 residential and ~10,000 commercial meters, along with another ~20,000 specially selected meters (for Distribution Planning, Revenue Assurance, Power Smart Evaluation and others) is transferred to the SAS server each night. This can be used for:

- Estimating the hourly loads (i.e. spreading the daily energy use) by feeder and substation, by site code and/or rate for DLSE and other planning work. This analysis will now use this week's sales data, not last year's.
- Conducting analysis on end-use consumption, supporting end-use metering and Conditional Demand Analysis efforts.
- Generating new knowledge about where our system and regional peaks really come from.
- The building of tools for residential customer support (end-use analytics, and advice)
- And for all of the earlier Load Research functions; such as supporting the Load Forecast, the FACOS, Rates Analysis, Power Smart Planning or Evaluation work, and Key Accounts queries.

The "super sample" includes all over 35kW demand-metered interval data to meet its accuracy requirements. To properly support Distribution Planning and Distribution Engineering work, a staff portal housing all of the demand-metered accounts interval data is being built, with the data to be secured in an SQL database and with a staff viewer built in .NET (see Figures below).

Access to All SMI Interval Data in Greenplum

Greenplum (EAS) stores more data in a “Big Data” device, with “Big Data” tools, initially developed to support theft detection analytics. Eventually this data will be available for non-theft related analytics. “Big Data” is an evolving term, used mostly by consultants, to describe the integration of huge data sets, measured in Terabytes, for a variety of business needs. The SMI hourly interval data for 1.8+ million meters belongs in big data hardware. Standard servers and software cannot effectively run analysis on datasets that are so large.

In late 2014 we expect that access to the EAS data will be operationalized, for non-theft analytic purposes, and with that final transition, our team will gain access to the interval data for the whole population of SMI meters, as opposed to the ~60,000 meters in the “super sample”. This will allow the development of more sophisticated models to support energy reporting for planning, forecasting, risk management, and other applications throughout the business. For the DLSE tool, the EAS energy balance database will hopefully replace the modeling work currently done for the residential and small commercial hourly patterns, within the day.

DLSE Methodology

The DLSE tool essentially transforms subsets of disaggregated, incomplete and confidential “Big Data” sources into complete and usable “Medium Data” data marts available across the organization. The methodology is a hybrid of Load Research techniques that have been used within utilities for decades, significantly enhanced by the use of very large sample sizes for interval data and the 1.8m smart meter daily register reads. The tool has evolved from a Load Research product that only modeled hourly data (using annual energy by account, annual topology, and annual load research hourly load profiles by site code) to that of a Data Mart based on exponentially more real (smart metered) data. The methodology emphasizes using the best and most accurate metered data available for aggregation, while using many different data sources and techniques so the end product is as complete possible.

The processes of the DLSE methodology, in the early phase, are shown in Figure 4 and numbered items within the figure are described below.

(1.1) Compile Monthly kWh Population Data: Multiple data sources throughout BC Hydro are joined together using shared keys to create a data source containing all BCH customers’ bills, account information, facility/building classifications (site codes) and connectivity information. This data is the basis for Load Research ratio expansion that transforms monthly consumption to hourly load shapes.

(1.2) Load Research Expansion Process: Hourly interval data for a large statistical sample of ≈10,000 Residential meters and ≈10,000 Commercial meters is joined to the monthly kWh data for the BC Hydro customer population to do combined ratio expansion. The very large size of these samples allows for further disaggregation of building/heating segments into 50+ site codes across 4 billing regions while maintaining high statistical accuracy of each profile. As a result, over 200 distinct hourly load profiles are created for modeling. These Load Research shapes are the foundation to appropriately weight on an hourly basis:

- the kWh consumption of the monthly bills of customers without a smart meter and
- the daily kWh consumption from SMI daily register reads, for residential and small commercial.

(2.1) Compile Hourly Loads from Daily Register Reads: The daily energy of the available SMI daily registers is weighted for each particular day to create hourly loads per customer. The resulting data set per month is ≈1.26 Billion records (≈1.75 Million meters * 24 hours * 30 days).

(2.2) Compile Hourly Estimates for All BC Hydro Meters: The monthly kWh consumption of all BC Hydro smart meters is weighted for each hour of the month. The meter's monthly consumption is applied one of the 200 Load Research load profiles created for the month based on the meter's account's site code and geographic location. The resulting data set per month is ≈1.30 Billion records (≈1.80 Million customers * 24 hours *30 days).

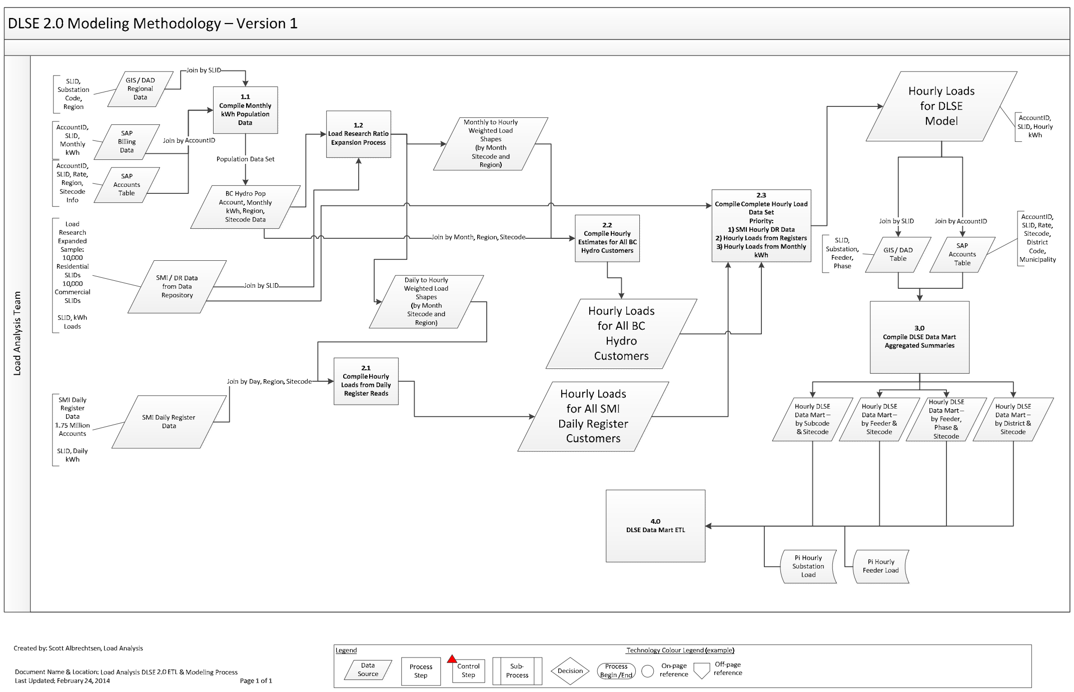


Figure 4. DLSE modelling methodology.

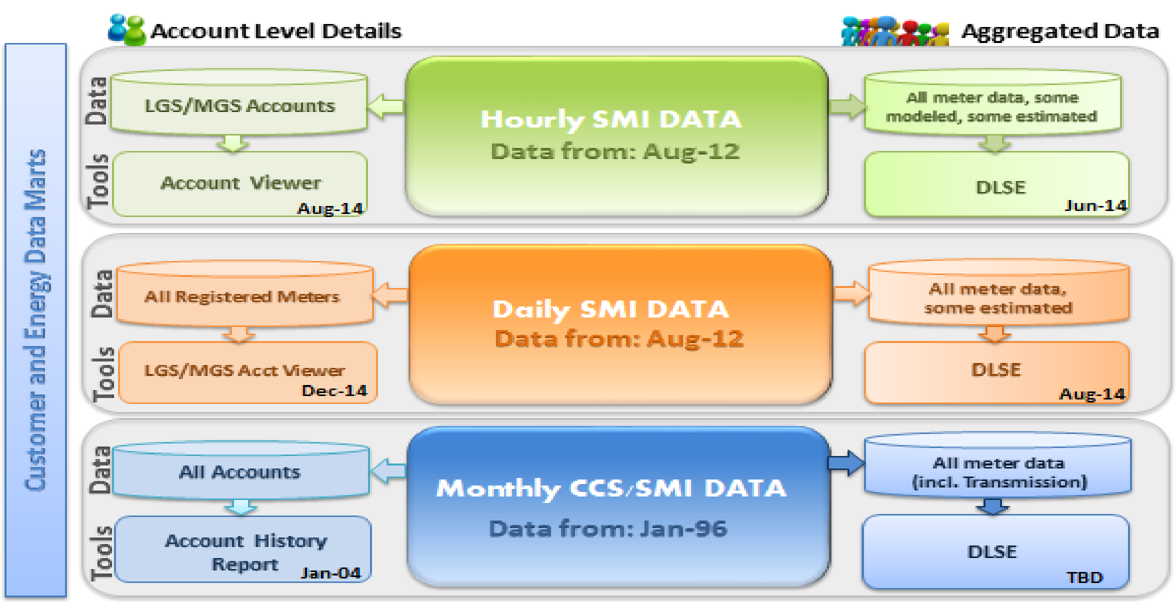
(2.3) Compile Complete Hourly Load Data Set: The hourly data created from the expansion of monthly kWh bills, register data for ≈1.8 Million meters, and the actual interval SMI data available, is joined together for a 100% complete data set of metered BC Hydro customers. This data is the "Medium Data" underlying the DLSE Data Marts. Priority is examined on an hour-by-hour basis where real SMI interval data is used first, and then data from hourly-expanded data from registers second, then finally data from estimated hourly loads via Load Research methods.

(3.0) Compile DLSE Data Mart Aggregated Summaries: The DLSE tool is a series of hourly Data Marts at multiple levels of aggregation. The appropriate GIS topology data as well as appropriate district and municipality information is joined to the "Medium Data" in 2.3. The data is then summarized per hour. The hourly loads can be shown at many levels of aggregation:

- Distribution Feeder and Phase Level by building type (site code)
- Distribution Feeder Level by building type (site code)
- Distribution Substation Level by building type (site code)
- Forecast Region by building type (site code)
- Billing Region by building type (site code)
- Transmission Region by building type (site code)
- District Code by building type (site code)
- BC Municipality by building type (site code)
- BC Hydro Distribution System Level by building type (site code), and rate code

(4.0) DLSE Data Mart ETL: Once the DLSE data marts are generated, there is an ETL process to load the SQL database of the DLSE web tool.

Customer and Energy Data Marts



Notes:
1. Load Analysis uses Statistical Analysis System (SAS) in the background for all of its analytic work. Display tools for shared data marts are built in .NET and the data is stored in SQL Server.
2. Residential Daily Register reads and Super Sample Small Commercial and Residential interval data are restricted to LA and RA staff.

Figure 5. Customer and energy data marts.

Figure 5 presents BC Hydro's customer and energy data marts, and their historical or expected in-service dates. It is clear that the data from the smart meters will be the source for most of the data marts made available to utility staff. The monthly data comes from two sources, the CCS billing system, and the SMI daily register reads for the last day of the calendar month. The calendarized data view in the AHR will be greatly enhanced by having actual data from month-end, rather than waiting three months for the meter reading cycles to complete, and calendarizing the billing data. Load Analysis uses SAS software in the background for all of its analytic work. Display tools such as account viewer and AHR for shared data marts are built in .NET and the data is stored in MS SQL Server. For privacy reason, Residential Daily Register reads and Super Sample interval data are restricted to LA and RA staff only. In Figure 5, for each data mart we show the starting date for the data, and the estimated or actual implementation

Privacy & Business Confidentiality

In general, the Data Marts produce summarized metered consumption information to a high enough level that privacy and business confidential information cannot be discerned. As a result, information from these Data Marts can be distributed freely throughout the organization without concern. The data included in the DLSE has all been managed to ensure that no cells of <20 residential accounts by dwelling type, or <3 commercial businesses by site code, will be found. A special privacy and business confidentiality process lies between the data aggregation and the viewing tools. As a result this DLSE data has no security issues, and only if and when it is combined with the very large transmission voltage customer data would we need significant security.

The access levels are defined in security profiles. A screen explaining the reason for the query will be built to maintain the data security, similar to our billing data viewer tool (AHR).

Current Situation

Contributions of Smart Meter Data via "Medium Data" Approach

With the implementation of this "Medium Data" approach, BC Hydro is already seeing benefits in several areas of our business. The following summary of current and expected benefits expands upon the bullet points made above.

Load Forecast/Finance accrual systems replacement. With daily register reads for all of the registered smart meter accounts, and hourly/daily estimates for the rest of the meters, the Finance team can begin to plan for the end of their accruals system. The official data for daily register reads will probably need to flow directly from EAS to Financial systems, but all the R&D can now be done.

Quick calendarization of billing data, with all the weather effects captured properly. Most of BC Hydro's energy planning is conducted using sales data which is calendarized (bi-monthly read billing data is re-allocated based on the number days in each month) to fit in the months it was used. This was important as most meters were read on a bi-monthly schedule. Now the daily register reads can be used, showing actual electricity consumption by month, with all the weather effects captured properly.

Sales tracking by segment, by day. At certain times of the year it is important to understand which segments of domestic sales are meeting the forecast, and where sales are falling short. We can now see sales by hour for the ~25,000 largest distribution voltage accounts, and with daily register reads, by day for virtually all accounts, and segments.

Fine-tuning the account level heating and A/C codes. With daily sales data for all accounts we can better define electrically heated and non-electrically heated accounts, and calculate their temperature response functions. More important, we will be able to identify the "hybrid" homes, those who put a significant electric space-heating load on the system on only the coldest days, but who could not be detected with bi-monthly billing data.

Enabling flexibility in a data retention strategy. BC Hydro is required to retain data used for billing, for seven years. The aggregation and storage of hourly interval data after adjustments are complete (90 days); by site code by feeder, and by rate and community ensures that a concise history of hourly consumption is available, by segments of interest, stored efficiently but at a very granular level, with no privacy or business confidentiality risks. The retention strategy for smart meter raw data can be based purely on billing rationale, as planning needs have been dealt with in Customer Care's Customer and Energy Data Marts.

Improving a vast array of forecast and analytic efforts. We can see accuracy and timing improvements in most standard reports, but what is most exciting is the fact that we will discover new uses for SMI data (daily register reads, actual large customer hourly sales, dynamically shaped aggregated small commercial and residential sales), providing answers to business questions that have not been asked, and could not be answered previously.

Estimating the hourly loads (i.e. spreading the daily energy use) by feeder and substation, by site code and/or rate for DLSE and other planning work. This analysis will now use this week's sales data, not last year's. This data will be used to do field planning for feeder balancing and storm response. Beneath the general feeder level, are the additional phase and the phase "section" levels for engineers to utilize. At the conservation planning level, the impending Conservation Potential Review (CPR) for Power Smart will benefit by having the utility's actual hourly load shapes, by 50+ building types (site codes plus facility sizes), to use for the base year data. This will improve the capacity planning, and better inform the energy planning aspects.

Conducting analysis on end-use consumption, supporting end-use metering and Conditional Demand Analysis (CDA) efforts. The super sample includes a statistical sample of 10,000 residential households, and Load Analysis maintains interval level data for ~5,000 2010 Residential End-use Survey (REUS) participants, and the ~8,000 from 2012. This provides an amazing depth of data for CDA, and for other purposes.

Generating new knowledge about where our system and regional peaks really come from. When the Load Forecast team can look at their peak day load by hour, by business/building segments, with any regional summaries they wish, uncertainties about what caused system peaks will dissipate. Power Smart's capacity planning can now benefit from knowing the hourly load shapes for each of 50 building types (site codes). This enables planners to target by building type, and by region. BC Hydro planners have been very pleased with the knowledge gained by looking the recent peak week, by dwelling type (Figure 7).

The building of tools for residential customer support (end-use analytics, and customer advice). With the residential survey samples noted above, BC Hydro or contractors working with our data should be able to build and test homeowner support information and software. Dis-aggregation of residential hourly loads comes in two flavours; utility-based where the customer is not involved, and customer-based, where the customer can answer 10-20 questions about their household, and improve the dis-aggregation results, and the advice provided by the utility. Where the customer answers questions, to gain better information about their use of our product, privacy is not an issue. Whether we can send billing alerts without the customer opting-in is yet to be determined, but we will prompt the Customer Service Reps, with possible reasons (heaters left on, failing refrigeration unit), when we suspect high bill queries might be expected.

For all of the earlier load research functions. Supporting the load forecast, the FACOS, rates analysis, Power Smart planning or evaluation work, and Key Accounts queries, the new SMI data, via the "Medium Data" tools, will improve data quality, and provide much faster response to business questions. Comparisons of the super sample load research analytics (sample = ~40,000 locations) to the previous sample of 1,500 locations will be made before the next regulatory hearing.

Sample Outputs from the DLSE Tool

Figure 6 shows the peak day (Monday) for Fiscal 2014. We can also see the contributions of each of the three sectors to the distribution peak, for the day before and the day after the peak day. With a total system peak of 10,023 MW, we can see that one needs to add ~2,000 MW for the transmission voltage load to customers with their own substations (for clients who need that data, there is a short delay), and about ~1,000 MW of system losses.

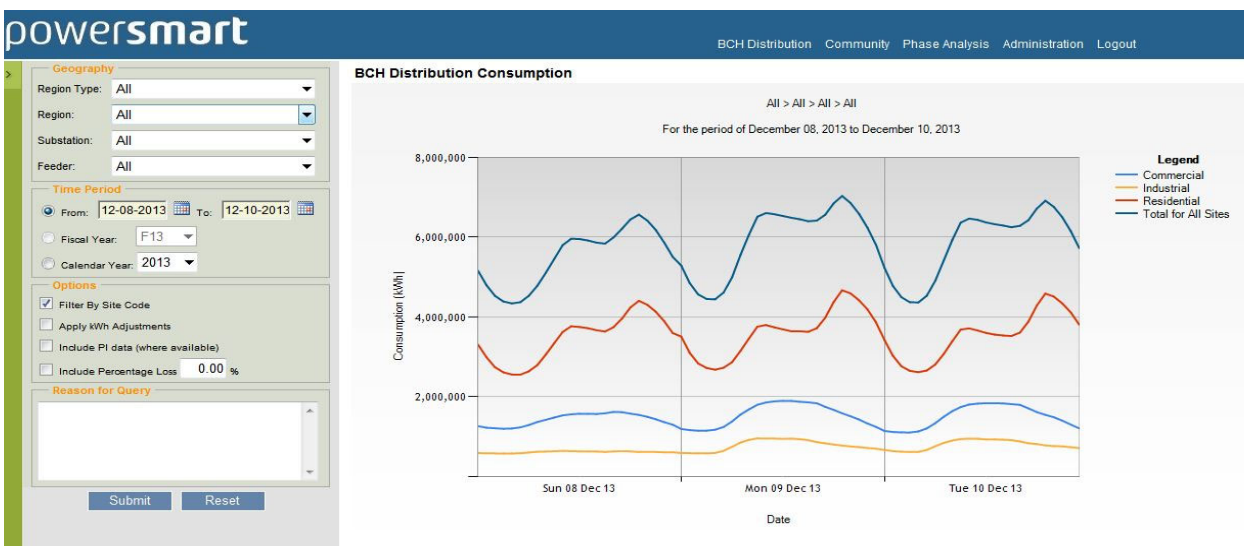


Figure 6. Peak days by sector.

Figure 7 below, shows regional residential site codes' load profiles. Each of the four regions can be viewed separately, but for inter-regional comparisons, one would need to do Excel extracts. Phase 2 of the DLSE tool development will likely add improved comparison features.

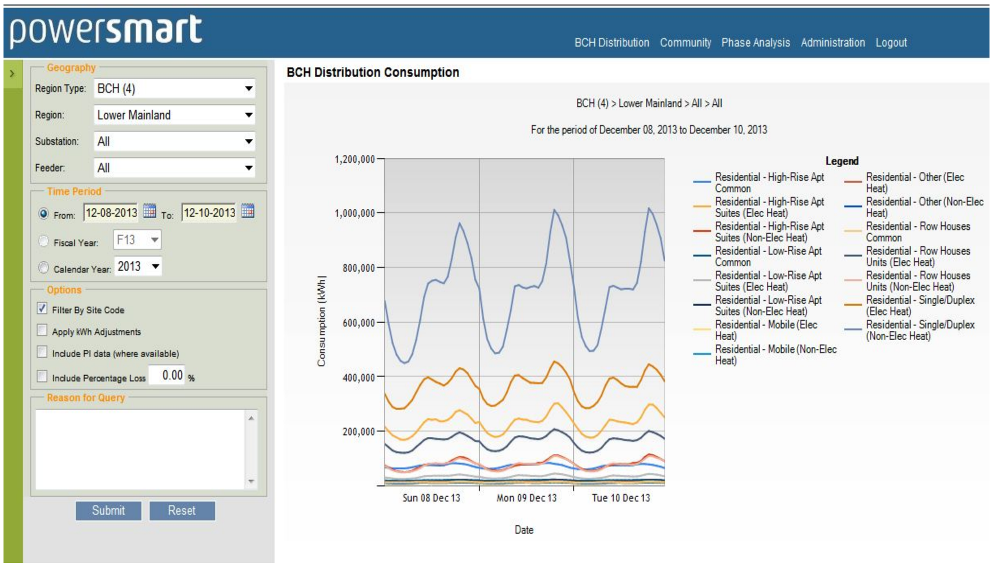


Figure 7. Peak day regional residential hourly use by site codes.

Closing Remarks

While smart meter data may initially have difficulties being worked with in a "Big Data" solution, one can perhaps more quickly solve the information problems with "Medium Data". With daily register reads for all meters, and good samples of hourly interval data, load research techniques can bridge the gap.

Even with "Big Data" solutions the utility will need to maintain historical data by segment, in the various time intervals. Using aggregation data marts as efficient historical repositories enables clear decisions to be made regarding retention of smart meter data.

The data marts BC Hydro is building in this "Medium Data" phase, will still be used in the "Big Data" phase, just the source and detail level of the incoming data will change.

Business groups always look to use the best source for their data. In BC Hydro, a "temporary" data repository, built to support the customer portal displaying the interval data, was the best source in 2013. It will also be the best source through 2014, and if it still operates in 2015 decisions need to be made as to whether to continue with the same source or to use the "Big Data" Energy Analytic Solution as the source for the Data Marts. The business units should be data source agnostic, but select their data sources based on: sources of record, data availability, reliability and security.