Utility Data Management & Intelligence

A Strategic Framework for Capturing Value from Data

Connected Energy Networks Business Unit

Jeffrey Taft, PhD
Paul De Martini, Newport Consulting Group
Leonardo von Prellwitz

May 2012

Introduction

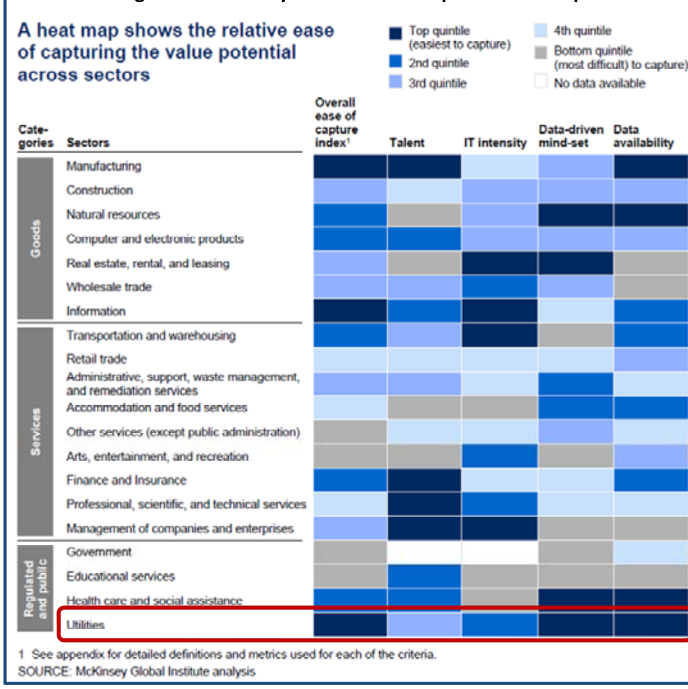
21st Century electric networks are rapidly evolving on multiple dimensions, including the development of energy information and operational platforms, in response to the adoption of a wide array of sensors, the penetration of significant distributed energy resources (including renewable resources and load management), and the enabling of market participation by millions of customers. This transition from vertically oriented value chain to a hybrid, more horizontal industry structure creates the need for the convergence of data, controls, and transactions into a unified energy platform enabling reliable and secure market and grid operations. The resulting platform is an emerging Enernet.[1]

In this new context, data includes a wide range of types to enable both use by electric systems for planning, operations, asset management, and electricity markets applications, and new services for end customers and other market participants. When combined with an exponential growth in volume and diversity of data sources and in variety of uses and related latency requirements, developing an effective data management strategy presents a very large challenge.

Although this growth has direct implications on computing applications, analytics and computing infrastructure investment, it's important to consider all of the data management steps: **collect, store, organize, analyze** and **share**. It's also important to note the role communications infrastructure plays in this process, as it's often a limiting factor in many smart grid systems and existing utility operational and enterprise networks.

Cisco has estimated the value of grid modernization investments in the US at about $210 billion through its Gridonomics™ analysis.[2] To realize this value, a 21st Century unified energy platform must be based on solid architecture and informed technology selections to fully harness the convergence of data, controls and transactions. Moreover, if this convergence is accomplished, the value number for grid modernization may be even higher. This is because the Gridonomics™ value estimate does not take into account the effects of managing the data in the most effective manner, and applying and leveraging advanced analytics to the data. That is, the $210 billion value does not fully account for the potential to extract greater network operational and business benefits by following the methods proposed.

Figure 1. McKinsey Ease of Value Capture Heat Map



While the subject of data management and analytics has tended to be treated in the

[1] Bob Metcalfe, GigaOM Interview, October 2, 2007
[2] Paul De Martini and Leonardo von Prellwitz, Cisco Systems, "Gridonomics", September, 2011

industry in rather confusing and hyperbolic language, the approach described in this paper is intended to provide simplifying frameworks that build upon the existing end-to-end process orientation and data centric operations of utilities. As such, it is interesting to consider the heat map from McKinsey[3] in Figure 1 above illustrating industry segments' orientation toward value realization from data. McKinsey's "*relative ease*" index rating suggests that utilities should be able to capture the value inherent in the data if the approach described in this paper is followed.

Identifying and Classifying Data Characteristics

The current "Big Data" discussion often does not differentiate between the several types of data found in utility operations or the temporal aspects related to the data types. It is important to distinguish the different types of data, including, among others, energy characteristics, operational state for energy production/storage/use, economic utility values, building/plant, process/device performance characteristics, market participant/customer data, geospatial information, electric network contextual information, and temporal/service attributes. To manage data effectively, it is essential to understand the differences related to each data class, their potential applications, and their respective latency considerations. Framing the data characteristics correctly allows proper treatment and identification of effective management solutions. Much of the industry discussion today on data management solutions seems to ignore this initial step in understanding the nature of the architectural and engineering problems to solve, causing potential challenges when integrating into the unified energy platform. Further, without a clear understanding of the potential analytics and business use from development of a data and analytic architecture at the outset, there is a risk of creating stranded costs from having to rework data stores and possibly buying the wrong data management solutions.

Data Class

Data arising from smart grid devices and systems may be grouped into five classes. Each has its own key characteristics and business value; an understanding of these classes is important in the development of networking solutions for electric utilities. Table 1 below describes these five key data classes.

Table 2. Data Classes

Data Class	Description	Key Characteristics
Telemetry	Measurements made repetitively on power grid variables and equipment operating parameters; some of this data is used by SCADA systems	Constant volume flow rates when the data collection technique is polling; standard SCADA polling cycles are about 4 seconds, but the trend is to go faster; telemetry can involve a very large number of sensing points. Telemetry data usually comes in small packets (perhaps 1500 bytes or so).
Oscillography	Sample data from voltage and current waveforms	Typically available in bursts or as files stored in the grid device, captured due to a triggering event; transferred on demand for use in various kinds of analyses. For some kinds of sensing systems, waveform data is acquired continuously and is consumed at or near the sensing point to generate characterization values that may be used locally or reported out (e.g., converting

[3] McKinsey & Company, "Big data: The next frontier for innovation, competition, and productivity", May 2011

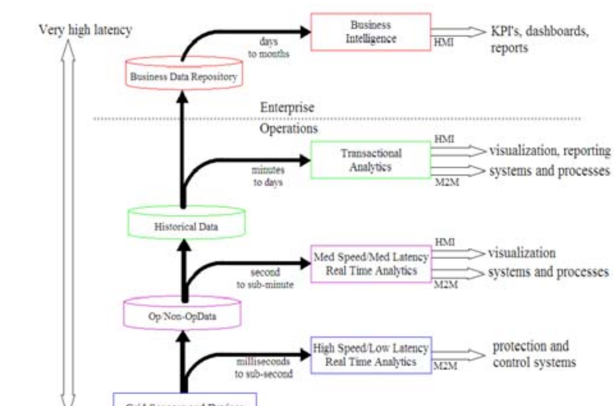
		waveform samples to RMS voltage or current values periodically); waveform sampling may be at very high rates from some devices such as power quality monitors.
Usage Data	Typically meter data, although metering can occur in many forms beside residential usage meters; typically captured by time-integrating demand measurements combined with voltage to calculate real power	May be acquired on time periods ranging from seconds to 30 days or more; residential metering may store data taken as often as 15 minutes, to be reported out of the meter one to three times per day
Asynchronous Event Messages	May be generated by any grid device that has embedded processing capability; typically event messages generated in response to some physical event; this category also includes commands generated by grid control systems and communicated to grid devices; may also be a response to an asynchronous business process, e.g., a meter ping or meter voltage read	For this class, burst behavior is a key factor; depending on the nature of the devices, the communication network may be required to handle peak bursts that are up to three orders of magnitude larger than base rates for the same devices; also, since many grid devices will typically react to the same physical event, bursting can easily become flooding as well.
Meta-data	Data that is necessary to interpret other grid data or to manage grid devices and systems or grid data	Meta-data includes power grid connectivity, network and device management data, point lists, sensor calibration data, and a rather wide variety of special information, including element names, which may have high multiplicity

Note that the business value of each class is not necessarily equal to that of other classes. It is important that each utility understand this concept and define the business value of each data class, perhaps to the point of subdividing the classes as appropriate for the specific utility's drivers and constraints, so that proper data management solutions may be derived that reflect the utility's business requirements. Also, data often is used by multiple departments within a utility and may have quite different perspectives on the classifications above. It is critical that a holistic approach is utilized along with an effective governance process to reconcile and differences. The governance process used for enterprise business process management should be utilized as the potential prioritization and ownership issues with data are part of this domain.

Latency

Identifying the temporal aspects of the underlying business processes and control systems is a critical consideration to develop effective data management strategies and architectures. A lot of grid data has multiple uses; in fact, it's an element of synergy that has significant impact on smart grid economics and system design (networking, data architecture, and analytics) to ensure that data is used to support as many outcomes as possible. Latency in this context can be defined as both the time interval between the time data is requested by the system and the time the data is provided by a source and/or the time that elapses between an event and the response to it. This is why it is important to understand how data is consumed in a variety of ways and places in a power grid and utility operations. While much industry focus has been directed at customer energy consumption data

Figure 2. Grid Data Latency Hierarchy

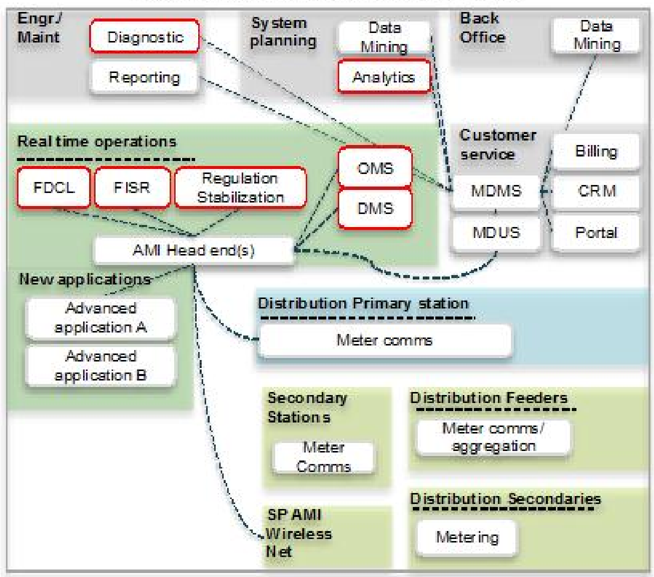


Source: J. Taft, Accenture

generated from smart metering systems, it is also important to understand the implications of the growth in grid sensor and control data streams. This is because much of this sensing and control data does not enter the data center and some does not even enter the control/operations center, as it must be consumed while streaming in grid devices and systems. Consequently it is important to classify data according to the latency requirements of the devices, systems, or applications that use it and define appropriate persistence, or actually, lack of such.

Figure 2 above illustrates the issue of latency. Latency hierarchy is a key concept in the design of both data management and analytics applications for physical networks with control systems or other real time applications.

Figure 3. Latency in Operational Context



What the latency hierarchy chart does not illustrate is that a given data element may in fact have multiple latency requirements and uses, meaning that any particular datum may have multiple destinations. Figure 3 to the right illustrates how data from a smart meter system may be used to support multiple operational process and analysis. Each use has unique performance and latency requirements.

This is why latency considerations must be included in the design of an energy platform. Otherwise, significant, and potentially fatal architectural issues will arise. These include; inability for applications or data stores to scale, inability to access data on a timely basis to meet business and operational needs, and/or creation of choke points on underlying telecommunications and computing infrastructure. Latency is probably the most overlooked and least understood aspect of utility data management today.

The latency hierarchy issue is also directly connected to the issue of lifespan classes, meaning that depending on how the data is to be used, there are various classes of storage that may have to be applied. This typically results in hierarchical data storage architecture, with different types of storage being applied at different points in the grid corresponding to the data sources and sinks, coupled with latency requirements. Table 2 below lists some types of data lifespan classes that are relevant to smart grid devices and systems.

Table 2. Data Lifespan Classes

Data Lifespan Class	Characteristics
Transit	Data exists for only the time necessary to travel from source to sink and be used; it persists only momentarily in the network and the data sink and is then discarded; examples are an event message used by protection relays, and sensor data used in closed loop controls; persistence time may be microseconds
Burst/Flow	Data that is produced in bursts or is processed in bursts may exist temporarily in FIFO queues or circular buffers until it is consumed or overwritten; examples include telemetry data and asynchronous event messages (assuming they are not logged) – often the storage for these data are

	incorporated directly into applications, e.g., CEP engine event buffers
Operational	Data that may be used from moment to moment but is continually updated with refreshed values so that old values are overwritten since only present (fresh) values are needed; example: grid (power) state data such as SCADA data that may be updated every few seconds
Transactional	Data that exists for an extended but not indefinite time; typically used in transaction processing and business intelligence applications; storage may be in databases incorporated into applications or in data warehouses, datamarts or business data repositories
Archival	Data that must be saved for very long (even indefinite) time periods; includes meter usage data (e.g. seven years), PMU data at ISO/RTO's (several years); log files. Note that some data may be retained in multiple copies; for example, ISO's must retain PMU data in quadruplicate.

Data Storage, Complex Processing, and Analytics

To manage Big Data it is essential to apply technology solutions appropriate to the data class and intended business processes to achieve the expected results. New data sources and flattening of business processes across a utility have created confusion as to the right technical solutions. In many instances, it is not a question of one technical approach versus another but rather what is the best combination for the business need. The following discussion provides a foundation for consideration of the use of data stores, complex processing and analytics for the 21st Century unified energy platform.

Data Stores

The foregoing suggests that data storage is not a simple matter of a storage area network at the utility enterprise data center. Many types of storage and database technologies are useful in the smart utility context. However, Hadoop, Oracle, SAP and Teradata all impact when and how analysis takes place and how to link together so the entire value chain can be viewed. Table 3 below summarizes principal types. Some types are specialized for specific purposes; others like standard SQL databases are used for more general utility applications.

Table 3. Data Store Types

Store Type	Comments
Operational Data Stores	Used to hold state data which is continually refreshed, such as power and device state data, real time grid topology.
Time Series Stores	Used to hold telemetry that will be processed in various ways over various time scales, but specifically including very long times.
FIFO Queues and Circular Buffers	Very short term storage for data being consumed quickly by applications; often implemented in the application itself as memory resident small volume buffers
Meter Usage Data Repositories	Large scale repositories for meter data; these often hold the data of record for billing; generally associated with meter data management systems, although some independent MUDR's have been implemented.
Relational Databases	Widely used in a variety of operational and enterprise contexts; built using either standard relational database technologies or memory-resident versions for faster response, especially in business intelligence and decision support applications. Utilities may have many such databases that have grown organically over many years of operation.
Warehouses and Datamarts	Used for storage of very large data sets for business intelligence, data mining, and the like; generally relational, but newer approaches are emerging.
True Distributed Databases	Databases in which various data elements exist in non-duplicated form on various physical stores, non-duplication being key to scalability; useful for operational data/grid state in distributed intelligence environments.
Waveform Repositories	Used to hold waveform files (oscillography); the waveform files may be treated as BLOB's; repositories can be special purpose or a general content management tool.

GIS as a Data Store	Geographic Information Systems are often the system of record for as-built physical network topology (occasionally it may be the Outage Management System that performs this function for Distribution); some smart grid applications need access to the as-built topology meta-data, so it can be necessary to use the GIS as a database, although most are not built for real time or near real time query support. Consequently, as-built topology may be staged to a datamart for near real time access, with periodic updates from the GIS to the datamart.
Federated Databases	This is not a database type so much as a middleware for databases; federation can tie together heterogeneous databases so that querying systems do not need the details of the multiple underlying databases; this technology, along with CIM-structured relational databases has been used to integrate multiple operational, transactional, and time-series databases in smart grid data management solutions
No-SQL/NoReL databases	Developing in response to "big data" requirements, these databases avoid the use of relational structure (hence the names "No-SQL" and "NoRel"), these databases are intended to scale to petabytes and beyond. These are beginning to see some use for business intelligence applications but have not penetrated utilities much as yet.
Content Manager Stores	Databases designed specifically for content management, so that files of various kinds can be stored, access-controlled, version-controlled, etc. Useful for BLOB-like objects, hence the mention above for waveform repositories, but also useful for engineering drawings, video, manuals, and grid device settings/configurations

As an example of why it is important to understand the relationships among the data classes, persistence models, and data store types, consider the present interest in the Hadoop "big data" storage model. Hadoop was originally designed by Google to analyze very large data volumes with a mix of complex and structured data that don't fit nicely into relational data bases. As such, the Hadoop model can be very good for enterprise-level business data repositories. However, for operational data it has several drawbacks[4]:

- The centralized data store model cannot satisfy the needs of low latency multi-objective/multi-controller (MO/MC) systems where analytics must often be consumed close to the point of data generation.
- The Hadoop Distributed File System (HDFS) coherency model does not work for dynamic operational state information and bursty event message data flows that are huge components of the big data challenge of smart physical systems. "HDFS applications need a write-once-read-many access model for files. A file once created, written, and closed need not be changed."
- The HDFS data access model is not suitable for highly interactive real-time system operations – "HDFS is designed more for batch processing rather than interactive use by users. The emphasis is on high throughput of data access rather than low latency of data access."
- The destination for much of the data in a smart grid/smart physical system environment is NOT an enterprise data center.

[4] The discussion and quotations regarding the use of Hadoop is based on material from the Hadoop website: (http://hadoop.apache.org/)

As another example of how data characteristics influence data and analytics architecture, consider the processing of data that is logically treated as a batch (such as meter usage data) as compared to data that is available as a stream, such as Phasor Measurement Unit (PMU) data. In the batch case, there is a data collection process that aggregates and accumulates data into a large store and then various data processing steps are applied to the entire set of data. In the meter case, this includes VEE and the processing of billing determinants into actual billing statements. For PMU data on the other hand, the situation more nearly resembles live streaming video. In this case, it is grid power state measurement data that is streaming, but the real-time use of such data eliminates the possibility of accumulating it into a large store to be processed offline, when the goal is to provide on-going operator decision support for grid operations, for example. In such case, the data is processed live, so that real-time data stores such as First In First Out (FIFO) queues are the first destination of the data, which is then piped into streaming analytics. As a secondary issue, the data may then be sent to a time series repository, so that offline analyses can in fact also be performed later. Thus a two-stage data management and analytics architecture is needed. Streaming data is increasingly common in smart grid implementations, as is bursty asynchronous event message data, leading to the need for a newer model for processing such data types, namely Complex Event Processing, which has lately also become known by the name "streaming database."

Complex Event Processing and Streaming Databases

Considering data class characteristics, latency hierarchy, and storage classes, we begin to see that the architecture of the modern utility data management system is increasingly distributed in nature and complex in form. Complex Event Processing (CEP) is a technology that has found wide use in industries as varied as financial systems, homeland security, and sensor data processing. In each of these cases, the common element is that data from edge devices must be processed "on the fly", whether it comes in streams or asynchronous bursts. The CEP technology is capable of applying complex queries to multiple data streams simultaneously to detect specified conditions ("events"), thus triggering appropriate actions in real time.

This distributed evolution poses new system integration issues as well as data management issues; even though it may be necessary to persist data in a distributed manner, the utility will certainly want to manage and potentially govern it from a central location. In addition, the need to combine various data storage types into a hierarchical multi-store scheme suggests the need for the use of data federation techniques in order to integrate the various stores into a unified data management solution. Finally, there is a need to integrate data quality management at the various levels of this data management hierarchy. While the tools for doing this at the enterprise level are well established, the same is not true for the lower latency aspects of grid data management. For example, several utilities have found that some level of filtering was also required for certain sources to ensure data quality correlation used to eliminate potential false positives. Utilities may wish to look at complex event processing (which has many other uses in an advanced grid environment) as also being useful for monitoring data quality in a streaming fashion.

The data quality issue bears further consideration. Tools and techniques for data quality monitoring and assurance at the enterprise data center level are fairly well established. However, these tools and methods presume that data will be resident for a period of time and that as behavior patterns gradually emerge, it will be possible to specify a set of rules that can then be applied against the data base going forward to detect data quality issues. Clearly this model is not applicable to real time data flows and streams in a distributed smart grid environment.

CEP technology is well established and can be implemented in centralized or distributed form. For the centralized form, the implementation resides in a data center and can be scaled to handle millions of events per second. For the distributed form, small footprint, embeddable versions of the CEP engine and rule base exist and can be implemented in distributed, cooperative fashion at the control center, at primary substations, and in distribution feeder devices. In addition to monitoring data flows for data quality issues, CEP can also perform the functions of message burst/flood management and filtering, and can be used to process sensor data to extract core information to be acted upon locally or to be forwarded to higher levels of the utility decision control hierarchy. Distributed CEP can be implemented such that individual CEP elements cooperate to generate complete event processing results.

CEP can support a range of relevant utility business functions. These include meter data management, fault detection, outage management, SCADA support, and remote device/system monitoring. In line with the previous discussion about the value of various grid data classes, CEP is a flexible tool that when included in an overall data management strategy and architecture, can significantly augment the flexibility needed to implement modern utility grid data management solutions.

Analytics

Utility operational decision making is evolving to address growth in intermittent resources and responsive load as well as technology advances to address grid reliability objectives. The time periods for making grid operational decisions are declining and the situational analysis is increasingly becoming automated as opposed to human centric. These changes are being facilitated by an increase in grid sensors providing greater fidelity of operational data and operational systems migrating from deterministic approaches to more stochastic methods.

Depending on the regulatory construct, reductions in operating expense may create headroom within an existing utility's "flat-lined" revenue requirements for additional capital spend and associated earnings growth. The opportunities are significant. As such, operational excellence programs have become a business imperative for utilities worldwide. For this reason, business processes across a unified electric platform will be an increasingly important area of industry differentiation in terms of both financial performance as well as customer experience. Also, customer service expectations are following the broader consumer technology evolution – substantially changing every 18-24 months. This is why the leading utilities are following many examples of successful Fortune 500 firms over the past decade by embracing and competing on analytics.

The objective for analytics is to automate high-volume decisions on a unified energy platform across a utility with precision in a consistent, scalable, fast and economical manner that allows a high degree of adaptability. The current range of commercial data management solutions is rather broad, including many enterprise data analytic tools and specialized energy and water analytics. Applying the right solution to a particular business process can be challenging given the variety of data characteristics and related business process within and across four primary operating units; transmission and distribution (T&D), customer service (Cust Svc), energy procurement (Energy Trading) and information technology (IT). In simple terms, these processes involve asset management, operations, customer engagement, and trading and risk management.

Figure 4 below illustrates a data management solution-business process matrix that may be developed to identify and clarify the various commercial solutions available to support several utility business unit processes.

Figure 4. Data Management-Business Process Matrix

	T&D		Cust Svc		Energy Trading		IT	
	Opers	Asset	Mktg	Opers	TRM	Opers	Opers	Asset
Data Stores								
Complex Event Processing								
Analytics								
Tools								
Data Integration (inc ESB)								
Algorithms								
Presentment								

The current analytic solutions market is comprised of commercial products in four general segments: tools, data integration, algorithms and presentment.

- "Tools" are products that provide a set of analytical tools that are configurable to create algorithms and decision models specific to a customers' need. These tools require professional services and customer domain expertise to implement and maintain them successfully.

- "Data Integration" solutions combine utility proprietary data with commercially available data sources and with pre-designed and/or configurable algorithms to yield unique information. Data integration, for example, can enable utilities to develop their own primary research data on customer behavior eliminating the dependence on costly and less effective secondary research.

- "Algorithms" are products that have been developed to address specific business areas to at least 80% of the functional requirement. While these solutions require external professional services and domain expertise for the remaining configuration, it is much less than that required for tool-based applications.

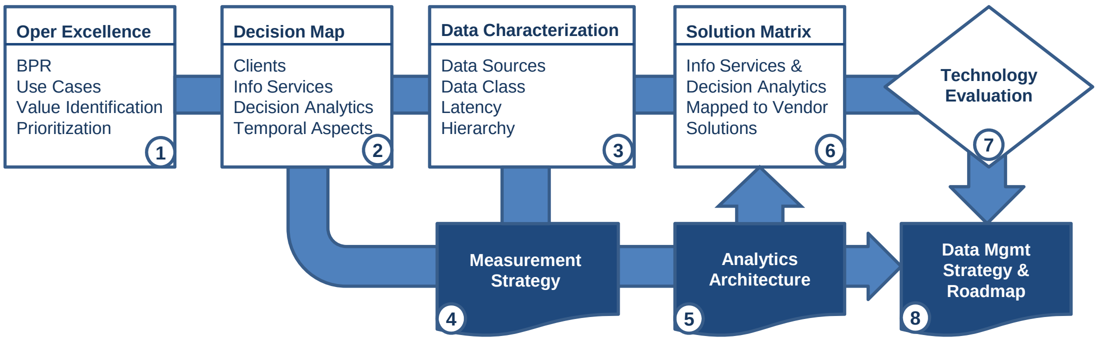
- "Presentment" solutions usually provide geospatial context, in addition to sophisticated graphical human interface. These applications can leverage proprietary asset locational information and real-time operational data in addition to mash-ups of commercially available geospatial information. Like tools, presentment products require system integration services and domain expertise to yield the best results.

Many commercial products combine functionality from two or more segments and can be leveraged across several business processes. As described in this paper, it is important to map business needs with the core capabilities of analytics solutions when assessing applicability as part of a data management strategy. This will allow better identification of value creation synergies and avoid overlapping functionality that can result in complicated system integration, higher costs and over paying.

Data Management Strategy

Effective use and management of data requires a holistic technology management framework to align business needs with technology decision processes to achieve the desired results. The figure below illustrates such a multi-step framework.

Figure 5. Data Management Framework



This framework incorporates both existing and future **business needs** (e.g., use cases, processes, values, priorities) to set a foundation for assessing appropriate architectures and data management solutions as an initial step. This is followed by development of a **decision map** of internal and external clients with respective information services and decision analytics. This mapping includes identification of the timing requirements for each identified decision process. **Data characterization**, as described earlier, is required to align the decision processes identified in step two with subsequent architectural and technology decisions. The result provides the input for developing both a **measurement strategy** and **solution matrix**.

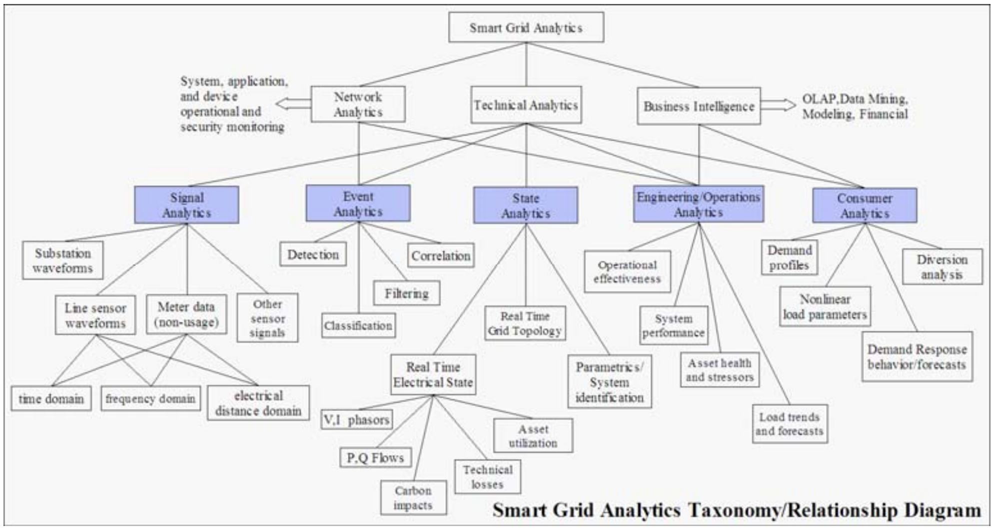
The measurement strategy includes the development of an observability strategy and sensor/measurement system architecture. These are both driven by requirements derived from the protection and control system architecture, and the applications architecture. The measurement strategy takes into account the real time data needs, as well as constraints on cost of implementation, and therefore takes into account the structure of the grid in question. By applying the control theory concept of observability[5], it is possible to minimize the cost of the measurement subsystem while achieving the necessary observability of the grid for support of the protection/control system and other applications such as outage and fault management, asset utilization optimization, and asset lifecycle management.

The decision map and measurement strategy enable development of an **analytics architecture**. The analytics architecture derives from the set of business process requirements for the advanced grid, and takes into account measurement strategy, grid and communication network structure. In the Cisco GridBlocks™ Reference Architecture, the communications network is also a platform for distributed intelligence, a prime element of which is distributed analytics. The reason for this is that in a distributed environment, data may have to be processed via analytics to extract actionable information with very low

[5] Observability is a measure for the effectiveness of a system's sensor data to determine the behavior of the entire system.

latency that precludes central processing at a control or data center. To facilitate the development of the analytics architecture, it is helpful to develop a taxonomy of analytics classes such as Figure 6 below[6].

Figure 6. Analytics Taxonomy



By combining this taxonomy with the measurement strategy and requirements for protection and control and other applications, the utility can develop a full analytics architecture that indicates not only what types of analytics are needed, but also where in the grid and in the network these analytics should reside. The resulting architecture supports development of the solution matrix, **technology evaluation,** and selection of the final **data management strategy and roadmap**. Some of the key implications for data architecture include the need to provide multi-level persistence modes, coupled with analytics matched to latency requirements, and the need to transition from fully centralized, batch-oriented processing to distributed, event drive processing with centralized management.

Conclusion

In response to the adoption of intermittent renewable resources, distributed energy resources, and the enabling of millions of customers to participate in electricity markets, 21st Century electric networks are evolving on multiple dimensions including the development of unified energy platforms. Concurrently, operational excellence programs have become a business imperative for utilities worldwide striving to meet earnings objectives. Management of data and use of business analytics offer the potential to address several critical needs that arise from these two trends. It is essential to consider the inter-relationships between business processes, business applications, data management and telecom and computing infrastructure.

Specifically, the utility information universe is changing from mostly relatively high-latency, batch-oriented processing to low-latency, streaming and asynchronous event message-driven real-time operations, not only on the grid operations front, but also on the customer experience and energy markets/transactive load fronts. With multiple classes of data, and the recognition that the data classes or

[6] Note: "Network Analytics" in Figure 6 refers to telecommunication networks and "Technical Analytics" refers to electric operational analytics.

even specific subsets of the data classes have differing economic values, the problem of managing utility data, processing the data and consuming it have become both larger, more complex, and far more crucial to the utility that in the past. Utilities are making the transition from being in the energy delivery business to being in the energy delivery and energy information management business. Data volumes are rapidly increasing and new uses for grid-derived data are being developed as new market models for energy delivery evolve.

This places a new emphasis on the need for updated data management architectures that move from batch to event-driven real time operation, that accommodate multiple uses for the same data at differing latencies, and that recognize that the value of data is in part based on the various ways it can be used. These ways often depend on the application of multiple analytics to the same data sets to extract differing information or differing views of the information inherent in the data. Ultimately, utilities must view their grid data as significant assets which need new data management strategies, roadmaps and architectures to preserve, extract, and realize the full value of grid modernization. The asset value holds whether a small or large utility, but the level of complexity increases significantly with the scale and scope of a utility's operations and supporting systems.

Value realization requires thoughtful planning, design, technology selection and implementation of data management strategies. Failure to comprehensively address these considerations in a worst case scenario, may lead to potentially tens of millions of dollars in stranded IT assets. It is clear that a structured approach can successfully harness these technologies to realize the opportunities. The data classification and management framework, data-related technology management, and data taxonomy process proposed in this paper will ensure that electric utility and industry stakeholders are well informed of the data-related needs, value proposition, architectural considerations, and commercial options that are properly aligned for success. In many cases utilities have already started on a path to manage data from their new systems, like smart metering, and create operational information. However, the recommendation is to apply these methods as part of the next architecture review or data management/analytics project pre-engineering effort to ensure future success. A utility's success in the 21st Century depends on the development and execution of a successful data management and intelligence strategy that effectively converges operational processes and the technology stack from application through telecommunications infrastructure.

Acknowledgements

The authors, Jeff Taft, PhD, distinguished engineer and chief architect, Cisco Converged Energy Networks, Paul De Martini, managing director, Newport Consulting, and Leo von Prellwitz, senior energy strategy analyst, Cisco Converged Energy Networks acknowledge the following individuals who contributed to the development of this paper:

- Tom Walker, director, data management & analytics, Southern California Edison
- Erin Gault, principal, Neptune Consulting
- Marianne Dickerson, sr director, t&d technology strategic planning & architecture, PG&E
- John Simmons, sr project manager, Intelligrid, EPRI
- Richard Hammond, vice president, Gridiant Corp.
- Jonathan Poe, director, utilities, Cisco
- Kristan Gainey, sr manager, marketing communications, Cisco

For more information please contact Jeff Taft at jetaft@cisco.com or Leo von Prellwitz at lvonprel@cisco.com .