21, rue d'Artois, F-75008 PARIS
http : //www.cigre.org

CIGRE US National Committee
2012 Grid of the Future Symposium

Grid Operational Data Management

J. TAFT[1]
Cisco Systems, Inc

P. DE MARTINI
Newport Consulting

USA

SUMMARY

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, intelligent electronic devices, variable and distributed energy resources, and the enabling of market participation by millions of customers. This transition from vertically oriented electric system to a hybrid, more decentralized 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.

In this new context, operational data will need to support a wide range of uses including planning, enginering, workflow, asset management, system optimization, real-time operations and controls, electricity markets, and new services for end customers and other market participants. This places a new emphasis on the need for updated data management architectures that evolve 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.

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. 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 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.

[1] jetaft@cisco.com

KEYWORDS

Data, Analytics, Intelligence, Data Classes, Controls, Latency

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 variable and distributed energy resources (including renewable resources and load management), as well as the enabling of market participation by millions of customers. This transition from a centrally managed and deterministic system to a more decentralized and stochastic grid creates the need for a convergence of data, controls, and transactions to enable reliable and secure grid operations. [1] In this new context, data includes a wide range of types to enable use by transmission and distribution system operations, electricity markets, and new services for end customers and other market participants. Without a clear understanding of the potential analytics and business use from the development of a data management strategy and analytic architecture at the outset, there is a risk of creating stranded costs from having to rework data stores, insufficiently designing telecommunications infrastructure and possibly buying the wrong data management solutions.[2] This paper is focused on data management strategies that support event-driven real time operation and accommodate multiple uses for the same data at differing latencies to realize both operational and business value.

Data Classes

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. Data arising from grid devices and systems may be grouped into five classes. Each has its unique characteristics and business value. An understanding of these classes is important in the development of a data management strategy for electric operations. Table 1 below describes these five key data classes.

Table 1. Data Classes

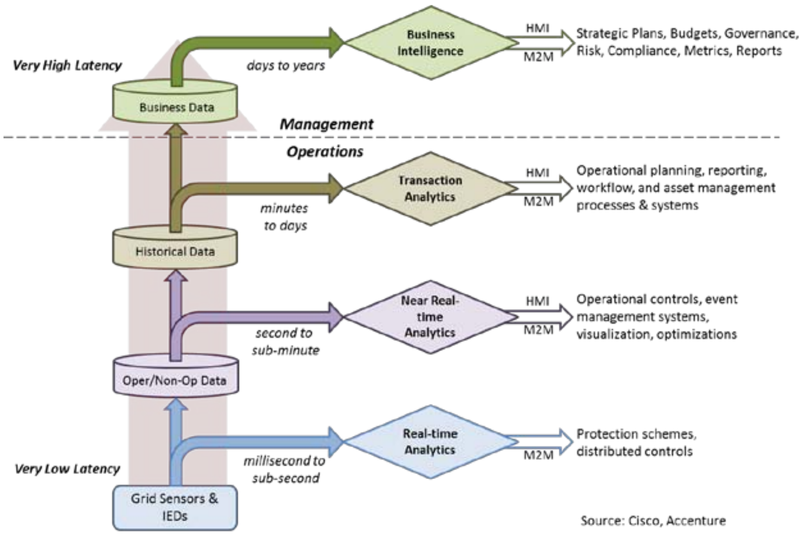
Data Class	Description
Telemetry	Measurements made repetitively on power grid variables and equipment operating parameters
Oscillography	Sample data from voltage and current waveforms
Usage Data	Typically meter data captured by time-integrating demand measurements combined with voltage to calculate real power
Asynchronous Event Messages	Typically event messages generated by a grid device in response to some physical event or asynchronous business process
Meta-data	Data that is necessary to interpret other grid data or to manage grid devices and systems or grid data

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.

Data 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 of the industry focus has been directed at customer energy consumption data 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 1 below illustrates the issue of latency.

Figure 1. Operational Data Latency Hierarchy



A similar view of this concept has been provided by von Meier. [3] 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. 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. 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.

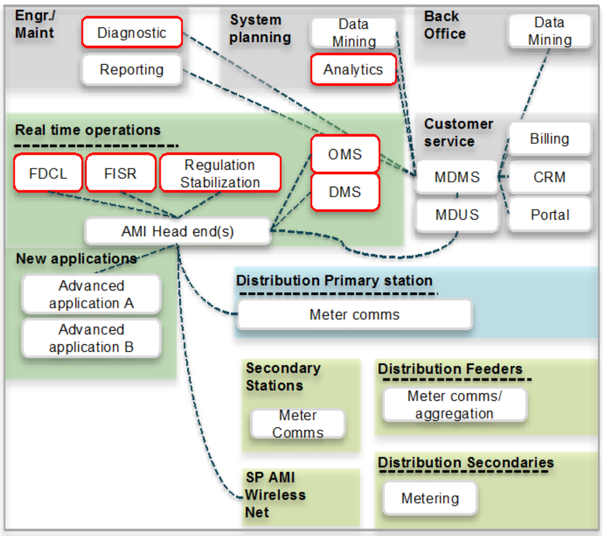
Table 2. Data Lifespan Classes

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.

Meter Data Example

In reality, any particular data element may have multiple uses, destinations and related latency requirements and lifespans as described above. Figure 2 below illustrates the various potential uses of smart meter data in an operational context. Each line in the figure represents a data exchange that has a specific data classification, latency requirement, lifespan and other attributes. Nearly all grid operational systems have similar multiple use requirements that neccesitate a holistic approach to ensure that an effective data management strategy can be developed.

Figure 2. Meter Data in Operational Context

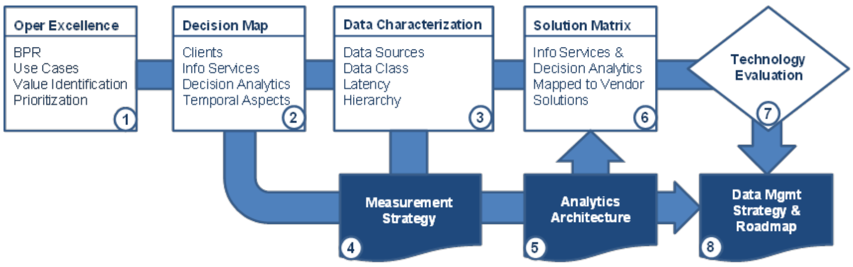


Additionally, decentralization of the grid creates new uses and temporal requirements for operational data management and system integration. For example, it may be necessary to persist data in a distributed manner as well as continuing to manage and potentially govern it from a central location. Also, this may drive the need to combine various data storage types into a hierarchical multi-store scheme that, in turn, 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 has been required for certain sources to ensure data quality; in addition, correlation has been 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.

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. Such a framework (see Figure 3) 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. Another foundational element is the identification of internal and external data clients with respective information services, analytics and timing requirements for each identified decision process. Data characterization, as described earlier, is required to align the data client decision processes with subsequent architectural and technology decisions. The result provides the input for developing both a measurement strategy and solution matrix.

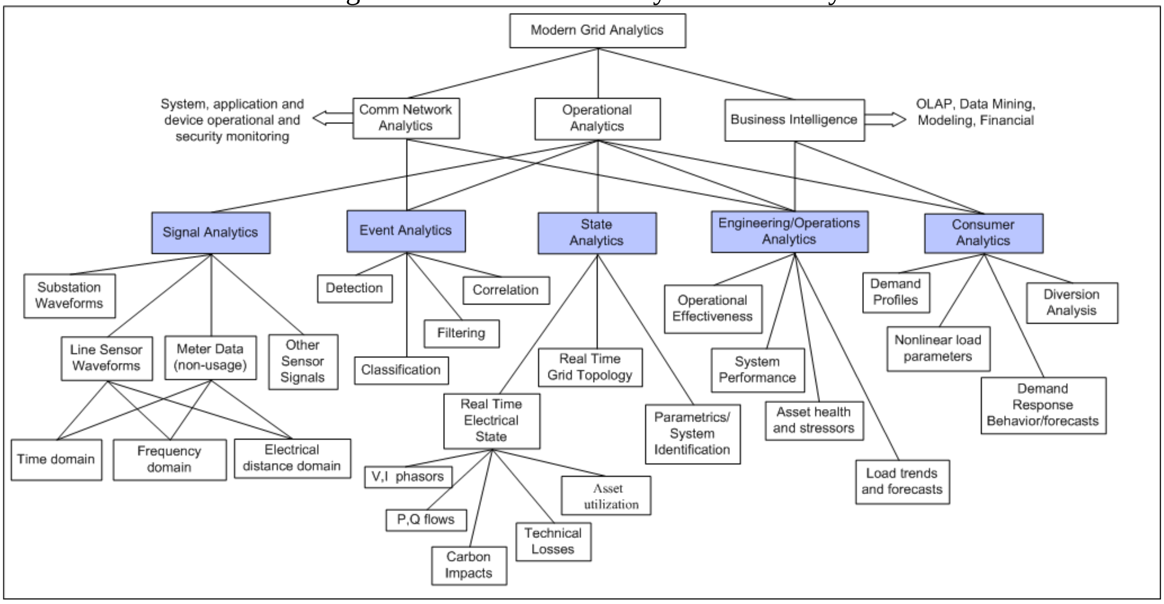
Figure 3. Data Management Strategy Development



A measurement strategy includes the development of an observability [4] strategy and sensor/measurement system architecture that are necessary to achieve observability of the grid for support of the protection/control system as well as other applications such as outage and fault management, asset utilization optimization, and asset lifecycle management. The observability strategy and measurement system architecture are both driven by requirements derived from a protection and control system architecture, and the related 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. The measurement strategy, in turn, enables 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 as well as grid and communication network structure. 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 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 4 below.

By combining this taxonomy with the measurement strategy and requirements for protection and control, as well as 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 driven processing with centralized management.

Figure 4. Modern Grid Analytics Taxonomy



Conclusion

Electric distribution operations are changing from mostly relatively high-latency, batch-oriented processes to low-latency, streaming and asynchronous event message-driven real-time operations. With multiple classes of data, and the recognition that the data classes, or 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 than in the past. 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 in both 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 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 as well as industry stakeholders are well informed of the data-related needs, value proposition, architectural considerations, and commercial options that are properly aligned for 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.

BIBLIOGRAPHY

[1] P. De Martini, L. von Prellwitz. "Gridonomics™", Cisco, 2011

[2] J. Taft, P. De Martini, L. von Prellwitz. "Utility Data Management & Intelligence, A Strategic Framework for Capturing Value from Data", Cisco, 2012

[3] Alexandra von Meier, Integration of Renewable Generation in California, California Institute of Energy and Environment, http://ucciee.org/downloads/EPQU_von_Meier_2011.pdf

[4] G. Franklin, J. D. Powell, and A. Emani-Naeini, Feedback Control of Dynamic Systems, 6th Edition, Pearson Higher Education, Upper Saddle River, NJ, 2010.