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

CIGRE US National Committee
2013 Grid of the Future Symposium

Advanced Utility Analytics with Object Oriented Database Technology

PTM	JJS	BT
Paul Myrda	John Simmins	Bert Taube
EPRI	EPRI	GE
USA	USA	USA

SUMMARY

In general, severe and wide-area electric power system outages tend to have two recurring causes. These are lack of situational awareness and breakdown or faulty behaviour of protection and control systems due to multiple power system elements failing which pushes the grid beyond its design criteria. Although power is usually restored to most customers within hours, some localized areas may not have power for a longer period.

Recent reports about detrimental wide area outage events have more specifically stated that causes were due, in part, to "lack of awareness of deteriorating conditions." They pointed out that the technology now exists to establish real-time transmission monitoring systems based on a measurement network composed of synchrophasors. Those reports recommended to establish real-time measurement systems and develop computer-based operational and management tools.

Object oriented database technology is a way to model and track large sets of objects that have a natural relationship. This relationship can be modelled in the database providing insights into relationships on the grid that may not be easily obtained using other technologies. This paper provides broad implications for situational awareness using object oriented database technology. It reports on the broader implication of the adoption of object oriented database technology on the situational awareness of the Smart Grid.

KEYWORDS

Scalable situational awareness, Big Data, advanced analytics, NoSQL data management, smart grid big data demands, real time analytics

INTRODUCTION – Why the Purpose of Big Data is really Situational Awareness

The concept of Big Data has been around for more than a decade, and its potential to transform the effectiveness, efficiency, and profitability of virtually any enterprise has been well documented. Yet, despite the concept of Big Data being well-defined, and the general enormity of its opportunity well-understood, the means to effectively leverage Big Data and realize its promised benefits still eludes many.

Big Data's remaining challenge to realizing these benefits comes in two parts. The first is to understand that the true purpose of leveraging Big Data is to take action - to make more accurate decisions, more quickly. We call this part situational awareness. Regardless of

Taube@ge.com

industry or environment, situational awareness means having an understanding of what you need to know, have control of, and conduct analysis for in real-time to identify anomalies in normal patterns or behaviors that can affect the outcome of a business or process. If you have these things, making the right decision in the right amount of time in any context becomes much easier.

Although the term of situational awareness itself is fairly recent, the concept has roots in the history of military theory - it is recognizable in Sun Tzu's *The Art of War*, for instance. Today, the concept of situational awareness has been expanded to a variety of areas such as air traffic control, nuclear power plant operation, vehicle operation, and anesthesiology. Defining the parameters of situational awareness for any industry is not simple, and thus surmounting Big Data's other remaining challenge of creating new approaches to data management and analysis to support these needs is also no small feat. Today, new data is created at a hugely exponential rate, and therefore any data management and analysis system that is built to provide situational awareness today must also be able to do so tomorrow. Thus, the imperative for any enterprise is to create systems that manage Big Data and provide *scalable* situational awareness.

The utility industry is in particular need of scalable situational awareness so that it can realize benefits for a wide range of important functions that are critical for enabling Smart Grid paradigms. Scalable situational awareness for utilities means knowing where power is needed, and where it can be taken from, to keep the grid stable. When power flow is not well understood, the resulting consequences can quite literally leave utilities and their customers in the dark.

Utilities can learn much about how to achieve scalable situational awareness from other industries, most notably building management and telecommunications, which have learned to deal well with Big Data's complexity and scale.

This paper will describe the diversity of utility data, the criteria which define a big energy data problem, and how utility business value can be built through increased situational awareness. The various problems that arise from storing the large number of data types in smart grid systems using traditional software technologies (data historians and relational databases) will be summarized. Overcoming these problems can be accomplished by applying NoSQL-based data management and analytics solutions, which can be seamlessly integrated with object-oriented languages. As a result, NoSQL solutions enable grid reliability and situational awareness, and help utilities achieve other critical business goals.

The Smart Grid Network's Unique Big Data Demands

In 2001, Doug Laney, now an analyst with Gartner, created what has become the widely-accepted three-dimensional definition of Big Data, also known as the Three V's of Big Data: volume, velocity, and variety.

Laney's definition, though, was created for the paradigm of a typical business, such as manufacturing, where profitability is often achieved by the minimization of fixed assets, where work in progress is measured in days, weeks, or months, the value chain for manufacturing almost always crosses company boundaries and real-time data collection and analysis are often not critical to ensure the profitability of the organization. In the utility industry, there are vertically-integrated and deregulated variants that have to act exactly the same.

In this environment, the acquisition of real-time data can both be costly and can seriously impact the bottom line. This adds more dimensions to the utility industry's Big Data needs as enterprises must not only deal with data's volume, velocity, and variety challenges, but also with two new V's: validity and veracity.

Validity - Information in the utility environment often has a "shelf life" and therefore may only be needed for storage and evaluation for a fixed period of time. The questions of when to archive or dispose of data become relevant given the cost of storing large volumes of data.

Veracity – This fifth variable recognizes that data is not perfect, and that achieving "perfect" data carries a cost. Utilities must consider two questions: 1) how good must the data be to achieve the necessary level of accurate analysis, and 2) at what point does the cost of correcting the data exceed the benefit of obtaining it?

The utility industry's time scales vary over 15 orders of magnitude due to the unique diversity of sensors and critical business processes. When trying to create scalable situational awareness, this impacts all five V's of the industry's Big Data pressures.

Big Energy Data Features	**Energy Data Types**	**Energy Data Sample Rates**
Data Volume (e.g. TBytes per Day)	**Telemetric Data** (e.g. in SCADA Systems; normally in Data Historian)	**μs-Range** (e.g. High Frequency Switching Devices)
Data Velocity (e.g. 300,000 Data Objects/sec)	**Oscillographic Data** (e.g. in Power Quality Monitor; normally in Data Historian)	**ms-Range** (e.g. PMU Devices)
Data Variety (e.g. Large Variety of Data Object Types/Classes)	**Usage Data** (e.g. in Meter Data Management System; normally in RDBMS)	**sec-Range** (e.g. DER Output Variations)
Data Validity (e.g. Large Variety in Data Object Shelf Life)	**Asynchronous Event Messages** (e.g. in Distribution Management System; normally in RDBMS)	**min-Range** (e.g. Service Restoration)
Data Veracity (e.g. Large Variety of Data Objects with different Data Quality)	**Meta Data** (e.g. in Geospatial Information System; normally in RDBMS)	**hour-Range** (e.g. Demand Response)
		Day-Range (e.g. Day-ahead Scheduling)
		Year-Range (e.g. Life of IT Asset)
		Decade-Range (e.g. Life of OT Asset)

Analyzing huge volumes of data that span multiple orders of time-scale magnitude falls short of traditional data-management technologies' abilities. Traditional methods of data management, such as relational databases (RDB) or time-serialized databases, may not have the capability to capture the causal effects of years or decades of events that may occur in a millisecond or microsecond range, and therefore cannot meet the real-time smart grids' scalable situational awareness needs. Additionally, such an array of devices and processes create an especially-wide variety of data types and formats that must be considered when making any decision, and thus for enabling scalable situational awareness.

Utilities' Current Data Analytics and Management Methods and the Need for Change

Utilities have primarily used the same two types of databases to manage and analyze their data for the last 30 years, despite neither of them ever being ideal for these critical tasks.

The relational database (RDB) rose to prominent use by utilities during the seventies when storage media was very expensive. The main advantage that the RDB was that storage costs could be minimized, but at the expense of needing to write a lot of proprietary code to describe the relationships between data (commonly known as a JOIN).

Data historian technologies, sometimes referred to as streaming data stores, rose to similar prominence for storing utilities' time-serialized data. These, essentially, were another form of the RDB, in this case optimized for storing data with a time-stamp, which emphasized reduction in storage costs at the price of not being able to easily correlate it with important variables other than just time.

The shortcomings of these technologies have become especially apparent as the need to conduct analysis across multiple data types, formats, and domains has become more important for the smart grid. The failure of traditional databases to scale well in the face of rising data volumes, complexity, and speed has been well proven, with alternative technologies often outperforming them by more than ten-fold in benchmarking tests.

What is needed are data management technologies that are optimized for analysis rather than constraints like speed and storage space. Ideally, these technologies would also be built much like the grid itself, with classes of assets that have natural, pre-defined relationships between them. These capabilities are readily found in proven object-oriented databases (ODB) and emerging NoSQL technologies, and which have been deployed across multiple industries with similar data challenges with great effect.

The ODB has been used for years in telecommunications, transportation, building management, and many other industries to track and analyze large numbers of data types and classes. Unlike relational or serialized databases, ODB's offer seamless integration with object-oriented languages. This means that the objects' application descriptions translate directly to the database objects themselves, making analysis much easier and faster, and thus supporting the goals of scalable situational awareness.

Additionally, query is used for optimization based on use cases and the business application of the data, not as the sole means of accessing and manipulating the underlying data. There is no application code needed to manage the connectivity between objects or how they are mapped to the underlying database storage. ODBs use and store object identity directly, bypassing the need for the CPU and memory-intensive, set-based join operations used by RDBs.

Requirements of Data Analytics Systems to Create Scalable Situational Awareness for Smart Grids

The underlying data management and analytics solutions required to provide scalable situational awareness for smart grids must have five key characteristics: Flexibility, Interoperability through Connectivity, a Control Network, it must use Open, Standards-Based Data Management Technologies, and it must support Scalable Data Analysis.

Flexibility - Unlike many industries, power delivery is notoriously variable, with daily, weekly, and annual variations due to variability in customer load, generation dispatch, delivery system outages, and other reasons. This variability has challenged the industry to discern patterns that can be used to identify abnormal conditions and anomalies that spur critical decisions-making processes.

Linked NoSQL technologies can deal with data that looks at voltage and current rate data just as easily as any other type of data from any other industries. By embedding a variety of different data object models to capture the different energy data types, as well as its corresponding sample rates, object-oriented programming allows for an integrated data management and analytics concept. It creates the necessary flexibility to deal with the challenging characteristics of Big energy Data in real-time. Critical values of a fully-integrated object-oriented data management and analytics solution include (1) fast and reliable data retrieval, (2) suitable data formats for data analysis, (3) one object-oriented programming language (for DDL and DML), (4) connectivity between objects without application code, (5) direct use and storage of object identities, and (6) combined advanced and traditional data management. This provides the situational awareness that is needed for utilities: understanding the immediate value of making a decision to solve an abnormality in normal data patterns within a relevant time frame.

Control Network - Not only is collecting all of the data that smart grid sensors and devices produce a challenge, but all of these devices must be fully communicative, interconnected, and, controllable. The decisions made based on having full situational awareness must be rapidly translated in to the functioning grid, which, like enabling interoperability, requires a single, cohesive control system enabled heavily through network connectivity.

Open, Standards-Based Data Management Systems – A network as complex, variable, and fast-moving as the smart grid requires millions of devices, sensors, and machines. It is impossible to expect that any one data management technology vendors' systems will be used across every grid application and scenario. But more to the point, smart grids will be integral to the everyday life of billions of people, so as new technologies are developed and adopted over time the smart grid must be able to adjust and change the data management systems to meet new requirements. To enable this, utilities must leverage open system architectures across five specific areas to permit ease of adoption and avoid costly vendor lock-in:

- *Network Infrastructure:* Includes protocol, routers, media type, IT connectivity, etc.
- *Control Devices:* Heavily-utilized devices that produce, consume, and manipulate data, as well as control and monitor the energy grid network.
- *Network Management and Diagnostic Tools:* Enable configuration, commission, and maintenance for the system.
- *Human-Machine Interface (HMI):* Includes the visualization tools through which users and managers obtain a view into the system, including both PC software and instrumentation panels.
- *Enterprise/IT Level Interface:* Connects the control network into the data network. No gateways other than open systems standards-based routers and IT-based data exchange mechanisms are used.

A critical sixth factor is the object-oriented data management system itself, which must also be considered part of this open standards-based architecture. The ODB represents the configuration database for the complete network of the grid, storing the configuration profile data of every device participating in the open, fully interoperable and integrated control network, and enabling effective communication and control between them all.

Scalable Data Analysis – Utilities will face immense data volume increases over the next several years, making the job of ensuring the validity and veracity of data analysis ever harder. Open architectures and data management technologies will play a pivotal role in enabling data analysis that scales to these new volume demands. These systems must not only be capable of dynamically scaling to account for and manage increased data complexity, but also sheer volume as new types of devices are deployed on the grid network.

Linked NoSQL's abilities to scale for both extreme data volumes and analytics complexity have been proven in large network environment deployments like telecommunications and building energy management. These industries are directly relevant to the smart grid because of their scale and complexity challenges, as well as the nature of the data itself. This technology has been deployed by dozens of telecommunications network operators and equipment vendors, and in more than 150,000 buildings around the world.

Conclusions

The potential power of scalable situational awareness through object-oriented data management for the smart grid is very substantial. Utilities are faced with the simultaneously large challenges and opportunities of Big Data, which make achieving scalable situational awareness harder, but also more important and rewarding. By turning virtually every piece of utilities' infrastructure in to a sensor, and making them fully interoperable and controllable with object-oriented data management technologies, utilities can prevent outages, mitigate potential threats to the network, and realize a range of other important business benefits. The energy industry should learn from other industries that have already conquered the challenges of Big Data, and apply the lessons learned to turn the current grid in to a smart grid.

BIBLIOGRAPHY

[1] Doug Laney, 3-D Data Management: Controlling Data Volume, Velocity and Variety, February 2001

[2] Codd, E.F. (1970). "A Relational Model of Data for Large Shared Data Banks". In: Communications of the ACM 13 (6): 377-387

[3] Jeffery Taft, Paul de Martini, Leonardo von Prellwitz. "Utility Data Management and Intelligence". Cisco White Paper, May 2012.

[4] Don v. Dollen and Bert Taube. "Advanced Computational Techniques for Situational Awareness and Analytics in Power Grids." (in cooperation with J. Simmins, P. Myrda, G. Gray, S. Sternfeld, V. Bagga). Redwood City: Presentation at Smart Grid Update's Conference "Data Management and Analytics for Utilities", June 2012.

[5] Bert Taube. "How Object-Oriented Data Management enables Smart Energy Control in Buildings." Santa Clara: Presentation at ConnectivityWeek, May 2012.