A CommodityPoint Whitepaper

Big Data – Challenges and Opportunities

Sponsored by

SUNGARD®

Prepared by
CommodityPoint, a division of UtiliPoint International, Inc.
19901 Southwest Freeway, Suite 121
Sugar Land, Texas 77479
www.Commodity-Point.com
January 2013

Introduction

Much has been written about "Big Data" in the last couple of years, but just what is it? As now commonly used, the term Big Data refers not just to the explosive growth in data that almost all organizations are experiencing, but also the emergence of data technologies that allow that data to be leveraged. Big Data is a holistic term used to describe the ability of any company, in any industry, to find advantage in the ever increasingly large amount of data that now flows continuously into those enterprises, as well as the semi-structured and unstructured data that was previously either ignored or too costly to deal with.

The problem is that as the world becomes more connected via technology, the amount of data flowing into companies is growing exponentially and identifying value in that data becomes more difficult - as the data haystack grows larger, the needle becomes more difficult to find. So Big Data is really about finding the needles – gathering, sorting and analyzing the flood of data to find the valuable information on which sound business decisions are made.

When applied to energy related businesses, Big Data implications vary by market segment – Big Data concerns for utilities are not the same as those for energy trading organizations; however the necessity for solving those problems can be as equally pressing.

In this whitepaper, we will examine:

- Components of a Big Data solution
- Big Data deployments for the emerging SmartGrid, utilities and energy retailers
- Value of Big Data to energy trading companies
- Big Data Framework for Regulatory Compliance and Trade Surveillance, including a case study on utilizing Big Data analytics in Fraud Detection in trading companies
- How do get started addressing Big Data Challenges

Big Data Requires New Technologies

Dealing with Big Data sets up to multiple petabytes in size (a single petabyte is a *quadrillion bits of data*) requires new technologies and new approaches to efficiently process large quantities of data within tolerable elapsed times. Traditional relational database technologies, like SQL, have been proven

inadequate in terms of response times when applied to very large datasets such as those found in Big Data implementations. To address this shortcoming, these Big Data implementations are leveraging new technologies, most commonly including the use of NoSQL (Not only SQL) data stores and Hadoop, a framework for processing the massive data stores that define Big Data.

NoSQL & Big Data stores first started appearing around 2009 and offered several intriguing capabilities versus relational databases, including:

- Easily distributed across multiple "commodity" servers and are horizontally scalable
- Open-source technology providing low-cost implementation with near constant improvements
- Robust and proven reliability in production with technology leaders such as Google, Amazon and Facebook ensuring high levels of availability
- Schema-free with easy replication support and relatively simple API's (application program interfaces)
- Support for structured, semi-structured and unstructured data
- Eventually consistent, meaning that overtime, data updates will eventually propagate throughout the database

Despite these advantages, NoSQL data stores are not intended to replace relational databases in true "systems of record" as they are currently not without limitations, including:

- No access control
- Inability to process complex transactions at the database level
- Relational concepts cannot be implemented (although that is what a RDBMS is for)
- No industry standardization, so each implementation is a "greenfield" effort
- A general lack of tools and frameworks that provide easy access for business users

The Hadoop distributed file system, as described on the Apache Software Foundations Hadoop website[1] is a "*…framework that allows for the distributed processing of large data sets across clusters of computers using a simple programming model. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to*

[1]http://hadoop.apache.org/

failures." Essentially, Hadoop was designed to act as the "operating system" for Big Data, providing a number of necessary capabilities that could not otherwise be provided through scaling of traditional technologies, including:

- Designed from the ground-up for batch use with high data throughput
- Ability to manage file sizes up to tens of petabytes in size
- Designed for write once, read many times access
- Provides interfaces for code to move to data, rather than other way around
- Allows the deployment and use on relatively cheap commodity machines
- Highly tolerant to failure

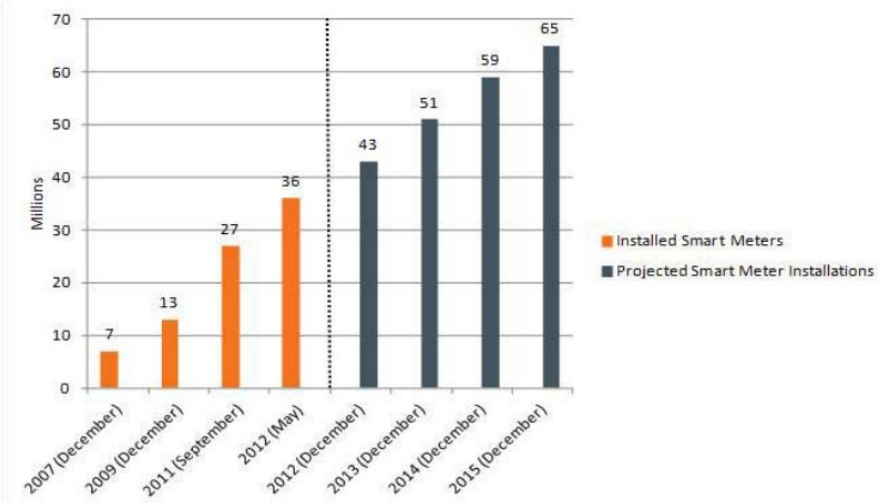
Hadoop is an open source technology, meaning that the technology is available for free and is constantly updated by a global user community, again including development resources at companies such as Yahoo and Facebook. It is also supported by various commercial providers such as Cloudera and Hortonworks, in a model very similar to that of Linux.

Big Data Promises for Utilities

Utilities have been deploying smart meters in ever growing numbers, spurred by the promise of improvements in operational efficiency, reliability and customer service. With improved demand data, utilities can better incorporate variable generation assets such as renewables, reduce the need for new or expanded base load generation such as coal fired facilities, and deploy innovative programs that allow customers to better control their electric usage while simultaneously reducing peak demand across the utilities' grid.

The Institute for Energy Efficiency (IEE) is forecasting that the US will see a 500% growth in smart meter deployments from 2010 to 2015, reaching an estimated 65 million units in service.

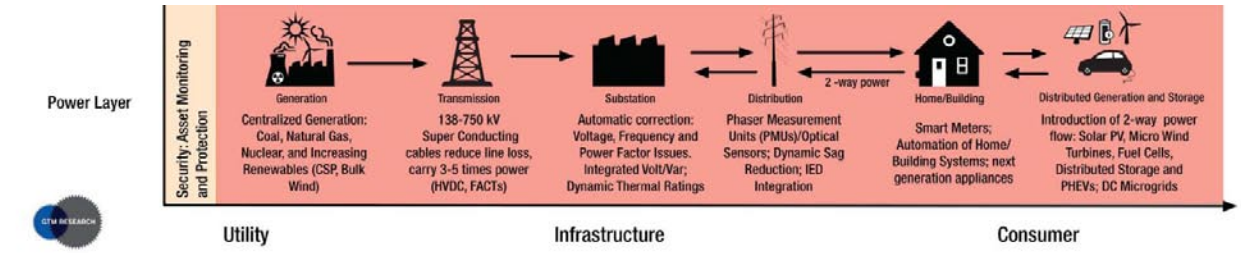
Smart Meter Deployments 2007 – 2015 (Millions)



Source: Institute for Electric Efficiency, Federal Energy Regulatory Commission[2]

These newly deployed smart meters are now starting to flood many utilities with millions of data points arriving in sub-hourly intervals. For example, for a mid-sized utility servicing a half million customers, the deployment of smart meters will result in an explosive growth of customer data, increasing the number of reads from 6 million reads per year to almost 18 billion per year, a 3000 fold increase in the amount of data that must be captured, ordered, stored, and analyzed in near real-time.

Upstream from the consumers' meters, operational data from newly deployed smart grid technologies, such as synchrophasors (which monitor power flow and quality and report data dozens of times per second), combined with sophisticated generation monitoring and control systems, and diverse and sometimes unstructured sources such as weather information, facility surveillance systems, and even social media (such as twitter) and emails should be considered as part of a consolidated data store. In doing so, utilities can gain a holistic view of their business environment and employ predictive analytics tools, enabling those companies to better anticipate outcomes, not merely react to events.



Source: GTM Research

Utilities that are able to harness and derive actionable information from this mass of highly diverse data, can realize significant benefit. For those companies with operations on the upstream side of the utility sector, such as generators and transmission operators, the advantages include:

- Better utilization/optimization of generation assets, reducing the need to invest capital in new power plants
- Improved grid balancing, reducing stresses on power transmission facilities and improving reliability
- Improved predictive equipment analytics reducing likelihood of failure of key, high value assets

On the retail side of the utilities, Big Data solutions have the ability to provide insight into the efficacy of demand response programs and provide utilities and retailers rapid feedback as to the value of innovative pricing programs based upon customer usage patterns. Other demonstrated uses include improved forecasting of bill payments/accounts receivable, enabling retail providers to better manage their collections processes by allowing them to stratify their accounts by probability of default. This allows them to prioritize their collections process and pursue payments on those accounts that are past due but still likely to pay, and not expend resources on those may merely lag past the invoice due date. In doing so, utilities can optimize their collections processes, reducing costs and realize real value from improved revenue recovery.

Big Data Advantages for Energy Trading

Wholesale market participants face similar challenges, however from an even more diverse set of data inputs than those found in the utility sector, both externally and internally, including sensor data from producing assets, real-time or end of day price data from a multitude of markets, counterparty credit data, position management information, and many others. The ability to programmatically collect and quickly analyze this data is key to realizing its benefits including:

- Reduced risks at all levels, including market risk, execution risk, and enterprise risk
- Improved decision response times enabling traders and decision makers to quickly respond to rapidly changing market conditions
- Maximized returns on costly assets via the use of optimization tools that combine operational and market data, both real-time and forward looking;

- Meeting time critical reporting and transparency requirements as mandated by regulators and financial markets.

Unfortunately, achieving these goals in a wholesale trading or marketing IT infrastructure can be difficult given the number and variety of systems generally found in these operations, including one or more core ETRM systems, risk management and analysis tools, optimization tools, market and exchange interfaces, ERP systems such as Oracle or SAP, and of course the ubiquitous spreadsheets found within every shop. Despite the potential complexities and effort of creating such a Big Data store within a trading environment, such solutions can provide a number of commercial and technical advantages, including:

- The ability to process more data, more quickly and with more accuracy, improving calculation of risk metrics, price volatilities, moving averages, and intraday price movements
- The ability to maintain all data, not just that which might appear be immediately relevant, from the ever increasing streams flowing into trading organizations, allowing companies to:
  - Better identify market opportunity by having more visibility into historical trends and previously unexplored markets
  - Discover new direct or cross correlations amongst commodities and geographies
- End of data triage – capture all the data arriving within the business as, all else being equal, more data is always better than not enough
- Capture and maintain better historical data for retrospection, including back testing of new risk models and price curve retrospection to improve accuracy of forecasts
- Decrease costs and reduce time required in acquiring data
- Achieve faster processing, yielding quicker answers to key questions such as VaR measurements, stress testing and position reporting
- Improve performance of the vital ETRM system resources by offloading data processing to big data clusters
- Gain the ability to scale and improve data acquisition/data processing times via rapid addition of relatively cheap hardware
- Improved compliance and control with improved visibility into all data and trading activities (trading surveillance), including intraday activities, not just end-of-day.

Big Data Framework for Regulatory Compliance and Trade Surveillance

Recently enacted Dodd-Frank regulations require trading shops maintain and be able to recall, within 5 days, all information related to the full life cycle of any swap or other regulated transaction, including data and records from all trading systems, documents, and logs that relate to that particular deal or set of deals. A Big Data architecture can capture and provide rapid recall and access to all artifacts related to the life cycle of the deal, including CTRM system data, ancillary system data, emails, twitter, IM conversations, voice recording transcripts, and other unstructured data sources. Utilizing a Big Data strategy, companies do not have to perform data triage, discarding much of the historical information that was previously too expensive (due to storage costs) to capture and maintain. These solutions provide a singular repository in which companies can keep all data, both structured and unstructured; and as these solutions utilize "commoditized" equipment, the cost of doing so is significantly reduced versus specialized high speed, high availability equipment that would otherwise be required. This ability to quickly recall and reconstruct the entire circumstance of a trade or trades is a prerequisite for being able to answer regulatory inquiries or internal audits.

As previously noted, Big Data approaches can add significant benefit to trading organizations in the area of trading surveillance – identifying aberrant trading behaviors in order to limit potential financial and regulatory exposures. Trading surveillance solutions, such as SunGard's Protegent Surveillance for Energy and Commodities, help firms review trading activity and identify behaviors that could indicate potential market manipulation or fraudulent behaviors. While these types of solutions do work with traditional relational databases, when utilized in conjunction with Big Data datasets, additional value can be derived, including:

- Full view of all historical trading activities, including highly granular power transactions down to the smallest time interval.
- Ability to quickly recall, review and analyze historical data from high volume algorithmic trading programs, permitting better views of trading patterns in order to uncover anomalies.
- Positions the company as a proactive, compliance-oriented organization in the eyes of regulators, rating agencies and shareholders.

Deploying a Big Data Solution

For companies wishing to take advantage of a Big Data solution, SunGard Global Services, a company experienced in Big Data deployments, recommends a structured approach that ensures program goals are achievable, and all components of the solution are scalable to meet current and future requirements.

At a high level, the company recommends the following approach:

- **Readiness Assessment** – Assesses business processes and IT operations ability to support the paradigm
- **Business Case** – Defines a strategy, assesses current initiatives, and analyzes business implications
- **Proof of concept**
    - Use of existing hardware vs. cloud services
    - Create a model for your data, understand structured vs. semi-structured vs. unstructured
    - Understand the current size & projected growth of your data, also the velocity (where does the information need to be & by when)
    - Find the biggest bang for your buck
    - Establish real time data replication to allow for parallel, intraday processing
- **Application Development** – Building applications that take advantage of big data features (scalability, etc)
- **QA & Testing** – Develop test plans, scenarios, and automated testing of big data-based applications
- **Service development** – Creating highly scalable, self-provisioned and self-healing applications
- **Metrics and performance** – Identify critical metrics that ensure optimal end-user experience

Conclusion

Given the torrents of data from the multitude of sources currently flooding utilities, trading firms and other wholesale energy market participants, managing that mass of data is extremely challenging – finding actionable market intelligence and using that data to rapidly respond to market developments is even more so. Given the impending exponential growth of data coming from the "Smart Grid",

increased regulatory oversight of energy markets, and the lost commercial opportunities buried in all that data, traditional capture, storage, retrieval and analysis techniques will prove ineffectual for many companies in this market.

Big Data solutions, already established and in-use in other industries, have the potential to address the data management and analysis issues faced by energy market participants in a rapidly evolving market and regulatory environment:

- Ability to capture, store, and retrieve all data, both structured and unstructured, eliminating data triage and the resultant "holes" in traditional data stores.
- Reduced cost of developing and maintaining a large-scale data store via the use of readily available and low cost "commodity" equipment.
- Facilitates the ability to rapidly analyze extremely large data sets, like those associated with emerging Smart Grid/Smart Meter technologies, improving performance of generation and transmission assets, demand response programs, account collection programs, and ultimately, bottom-line financial performance.
- Facilitates improved analysis of historical trend information, including both trade data and macro market information, providing improved performance via tuning of trading programs, including algorithmic programs.
- Supports and enhances regulatory compliance and trading surveillance through complete capture of, and rapid access to, all historical trading data as necessary to meet regulator inquiries

For companies seeking to deploy a Big Data solution, a structured approach is necessary to ensure all stakeholder goals and issues are addressed, the business case is properly developed and vetted, and the technology deployment is fully aligned to meet the enterprise goals both now and in the future.

About SunGard Global Services

SunGard Global Services helps financial services and energy companies solve business problems by managing their complex technology and operational needs. Focused on large scale data and process management, custom application development and systems integration; we combine business and technology consulting and managed services to deliver innovative custom solutions. We help our customers capitalize on and manage the consequences of business, technology and regulatory change. For more information, visit us at www.sungard.com/globalservices/learnmore.

About SunGard

SunGard is one of the world's leading software and technology services companies. SunGard has more than 17,000 employees and serves approximately 25,000 customers in more than 70 countries. SunGard provides software and processing solutions for financial services, education and the public sector. SunGard also provides disaster recovery services, managed IT services, information availability consulting services and business continuity management software. With annual revenue of about $4.5 billion, SunGard is the largest privately held software and services company and is ranked 480 on the Fortune 500. For more information, please visit www.sungard.com.

About CommodityPoint

Commodity*Point* is the industry leader in providing Commodity Trading & Risk Management (CTRM) research, analysis and advisory services. Our services bring insight into business issues, trends, processes and technology, to utilities, energy companies, banks, brokers, funds, investors and vendors that enhance their competitive position and support critical business decisions around the wholesale commodity trading markets. Our team provides expert analysis of market trends and, in particular, the technologies and applications supporting those that participate in regional or global commodity markets.

With offices in Europe and the US, and backed by an experienced research team, our organization provides an unparalleled view of the marketplace. **Commodity*Point*** is a division of leading energy and utilities analyst and consulting firm, UtiliPoint International, Inc.

Other Resources

The CTRM Blog – http://www.ctrmblog.com

CommodityPoint – http://www.Commodity-Point.com

TRM Products and Services Online Directory – http://www.trmdirectory.com

Other CommodityPoint Reports

CommodityPoint research reports are available at http://www.Commodity-Point.com

For Additional Information regarding this report or any CommodityPoint product or service, please contact Patrick Reames at preames@utilipoint.com or Mark Tredway at mtredway@utilipoint.com.

CommodityPoint Locations:

The Americas
19901 Southwest Freeway, Suite 121
Sugar Land, Texas 77479 USA
Phone: 281-207-5440

Europe & Asia/Pacific
International Business Center
Prikop 4
602 00 Brno
Czech Republic
Tel: +42 0 533 433 658

UtiliPoint International is a wholly owned subsidiary of Midas Medici Group Holdings, Inc. (OTCBB:MMED).