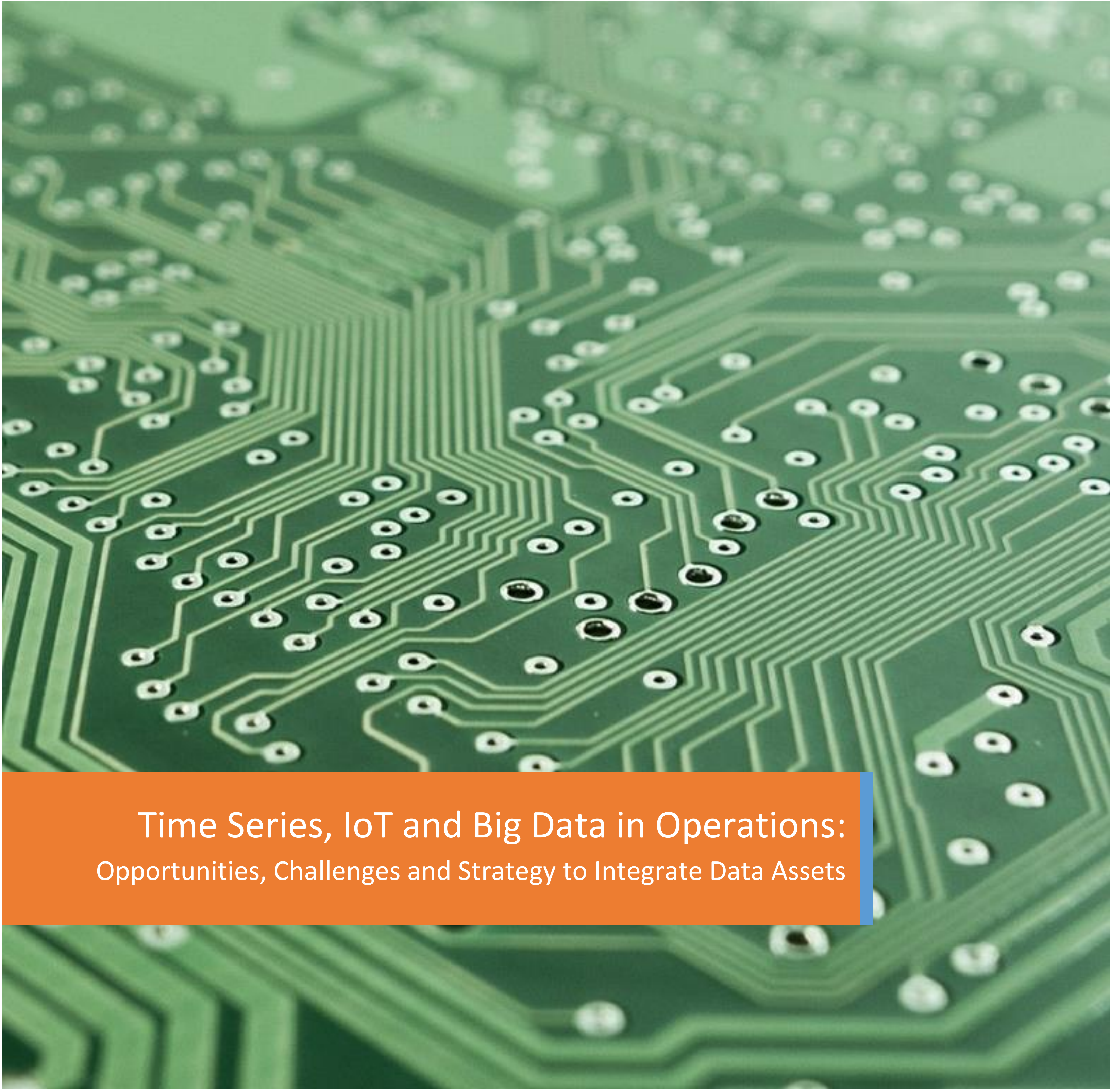


Time Series, IoT and Big Data in Operations:

Opportunities, Challenges and Strategy to Integrate Data Assets

REVONOS
REVOLUTION IN UNDERSTANDING

Executive Overview

As companies have streamlined and modernized, massive amounts of data are being created by equipment and processes. The Internet of Things (IoT) movement will add to this data boon as up to 6.4 billion devices are expected to be connected and in use by 2016 (a 30% increase from 2015) and 21 billion by 2020. It is this data asset that companies will turn to in order to gain the next step change in their operations.

The technological barriers to integrating Information Technology (IT) with Operational Technology (OT) are falling. Often companies that are beginning to recognize and embrace the opportunities to combine IT and OT implement new technology components prematurely without a clear understanding of all the requirements, how they work together, and proper support from disparate business units and executives.

- An oil and gas company invested in cloud server architecture, Hadoop cluster, and visualization tools to gain insight into the massive operational data stores available within the company, but failed to capitalize on the investment by focusing on low value-add network monitoring with input only from the IT department.
- A $13B market cap corporation invested in several overlapping new technologies to optimize field logistics, but abandoned the project after not getting proper input from OT groups, and failing to evaluate existing in-house technologies to complement the new technologies or supplement the new technologies with their capabilities.
- A utility company hired a team of personnel to learn data science, and invested in in-memory analytics modeling tools in addition to a business intelligence reporting engine to streamline operations. The data science team is currently idle and waiting for input from operations to develop strategy and value cases to address. Additionally, while the data science corporate roadmap is robust, the ability to monetize and provide actionable information to front-line workers was not included in the roadmap.

To realize the most value from company data assets, a holistic strategy and understanding of technology options and business drivers is needed. The need to combine these systems is a given, but the roadmap to successful IT-OT convergence is complex, and the path must incorporate cultural change, process knowledge, systems knowledge and business knowledge.

This paper addresses the opportunities and challenges that companies are presented with when moving towards an IT/OT convergence, and propose a strategy to integrate data systems within your company.

Opportunities

The potential economic impact of the Internet of Things (IoT) is staggering. By some estimates, IoT could have an impact of $4 trillion to $11 trillion per year, with some of the biggest impact in operations management and predictive maintenance. Engineers, analysts, decision makers and others can access IoT data for day to day tasks as well as root cause analysis, reporting, process engineering and operating excursion identification and resolution. Integrating other data sources like operations application data, spreadsheets, public records and unstructured data can reveal new insights in the hands of users across the company.

As storage costs decrease, I/O rates increase, visualization technology improves, and organizations empower employees lower in the organization to make right-time decisions, the uses for IoT data increases. Also, industries have seen the development of tools that could enable the convergence of IT and OT technologies. The [synergy of IT/OT technology would] allow greater numbers of technologies, devices, and systems to connect…to improve efficiency, provide access to data normally held in silos, and enhance productivity. The potential in IoT technology by select industries is outlined below in Table 1: IoT Maturity and Potential.

Table 1: IoT Maturity and Potential

Industry	IoT Maturity	2015 Investment	Approximate Potential Compound Annualize Growth Rate (through 2018)
Retail	Medium-High	$196M	20%
Logistics	Medium-High	$192M	20-25%
Wholesale	Medium	$50M	10-15%
Discrete Manufacturing	Low-Medium	$462M	15-20%
Process Manufacturing (E&P, etc)	Low-Medium	$104M	15%
Healthcare	Low	$167M	25%
Utilities	Low	$114M	15-20%

Challenges

Operations rely on several purpose-built applications to run operations, all specializing in their type of data, or process. The purpose-built model leads to siloes of data, and segmented organizations. Historically, the usage of purpose-built data beyond the specific group or discipline was limited for several reasons – data security considerations, data standards, data quality or integrity, storage or replication restrictions, perceived uses for the data and issues with cross-integration of specialized data.

- Data Standards - SCADA systems have been the main provider of high resolution time series data, also now referred to as part of the Industrial Internet of Things (IIoT) spectrum. With the IoT movement, massive amounts of data are being created and will need to be democratized across operations. Unlike SCADA, there is still some uncertainty of standards with new IoT device data, which will pose a challenge for companies attempting to make the data available. The data standard challenge will need to be addressed through investment in application-level normalization or increased data governance and stewardship efforts.
- Siloes of data and expertise - Purpose-built applications causing data separation across operations are fairly common, even with similarly structured data. These applications are integral to operations, but have different customers (internal and external), requirements, guidelines, rules, level of detail, etc. Different requirements and customers lead to silos of data and knowledge and a lack of overall traceability and accountability regarding who has access to both critical and noncritical assets. Attempting to integrate these systems without proper support, governance and buy-in from operations, IT and the executive team will lead to delays and possible failure for IT-OT convergence programs.
- Data quality and integrity - Time, personnel, and changing business conditions have created data within systems that is very disconnected from other systems. Data quality, integrity and standardization are poor which makes efforts to democratize a cross-platform dataset very difficult. Attempts have been made to interface disparate data to a data warehouse and then build models or commonality in the warehouse. Alternately, much effort is expended to interface key items between systems to build commonality. Both paths can be valuable, but can often be derailed because a holistic strategy and stewardship program is not in place.
- Storage and Replication – Effective storage, timely availability and access data is an additional challenge for many industries. It is estimated that 90% of the data collected is never utilized, and 60% of that data loses its value moments after it is collected. There is a wealth of information available in real-time time series data that can be enhanced with real-time analytics. A post-facto approach to storage and analysis of time series data causes companies to miss opportunities to take action, automatically implement transactions, and make decisions based on information available in live data streams.

An oil and gas corporation made an effort to utilize raw time series data from a SCADA system within a decision management engine to optimize field dispatch operations. The company also had a purpose-built time series analytic framework that specialized in capturing, cleansing and contextualizing raw time series data (SCADA). Instead of utilizing the built-in functionality of the existing application to create real-time calculations and analysis, the company tried to develop the functionality in-house in their big data platform against the SCADA data. The added development costs were a main reason the project was not successful and abandoned.

Strategy for Creating Integrated Data Systems

Architecture Description

The architecture is composed of three parts: purpose-built applications, a Big Data platform and the analytics delivery tools. For purposes of this article we have broken out operations from financial purpose-built applications. The different components of the architecture can be found in Table 2: Components of the Big Data Architecture.

Table 2: Components of the Big Data Architecture

Type	Data	Functions	Examples
Purpose-Built Operations Applications	Real-Time Events Ops Structure Transactions Rules Process	Specialization Operations Transactions Specific User Base	Various operational systems
Purpose-Built ERP Financial Applications & CRM	Financial Driven Processes Corporate Structure Transactions Corporate Interests	Accounting Budgeting Asset Information Financial Reporting	Oracle EBS, SAP, JD Edwards, Salesforce
Big Data Platform	All of the Data + Non-Corporate (3rd Party) Relational Time Series Unstructured	"Data Lake" Large Data Sets Holistic Analysis Advanced Analytics Storage and Processing	Hadoop (Hortonworks, Cloudera), Apache Spark™, AWS, Other
End User Analytics/Reporting	Transient Data Store	Visualization Data Tables Graphs Discovery & Investigation Drill Down Ad Hoc	Tibco Spotfire®, Oracle Business Intelligence Enterprise Edition, Microsoft Excel, Microsoft SQL Server Reporting Services, SAP HANA

Big Data Ecosystem

Big Data platforms have multiple, varied tools to handle different methods of storage and manipulation of data. Each implementation can have different components, but each component generally falls under one of the roles in Table 3: Roles in Big Data Ecosystems. These tools work together to provide the necessary components to run an effective data platform. The tools are not specific to any industry or application. They are built for performing a specific purpose applicable across industry, function and for the most part, data type.

Storage role - responsible for representing the data persisted on disk in a format useful for programs.

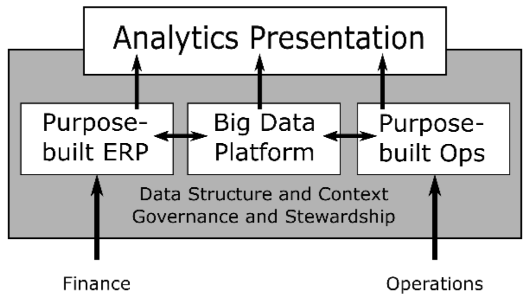
Processing role - takes data from the storage role or another source, performs some manipulation of the data and passes the results to a service in the storage role or outside the system.

Interface role - help to persist and process data from virtually any source. Tools in the interfaces role simplify interaction with the processing or storage tools.

Security and Administrative roles – control access, and provide overall administration of the platform. These roles are necessary for operations support in an enterprise system environment.

Table 3: Roles in Big Data Ecosystems

Role	Purpose	Examples
Storage	Handle data persistence	Apache HDFS™, Apache HBase™, Oracle Berkeley DB, MongoDB, Amazon Simple Storage Service (S3)
Processing	Perform actions on stored data	MapReduce, Apache TEZ™, Apache Spark™
Interfaces	Interact with data storage and processing	Apache Hive™, Apache Pig™, Apache HCatalog™, Amazon Kinesis Firehose
Security	Protect data access through authentication, authorization and encryption	Apache Knox Gateway, Apache Ranger, AWS IAM
Administration	Tools to ease administration and human interfacing	Apache Hue™, AWS CloudFormation



Key Components of a Successful Integrated Data Strategy

Technology is a major component of a successful IT-OT convergence. Additionally, culture, governance and accountability are key items to consider and incorporate. In order to gain value from your data assets, a company should incorporate the items found in Table 4: Components of an Integrated Platform. For a successful implementation, interface across the components below is necessary.

Table 4: Components of an Integrated Platform

Component	Roles Required	Purpose
Governance	Operational Experts, OT, IT	Maintain and update data standards
Community of Interest (COI)	Program Lead, Multi-Discipline Representation, IT	Align with company goals and objectives, discuss and prioritize company projects and initiatives
Purpose-built Applications	Specialized discipline personnel	Maintain source-of-record data and interface with Governance and COI personnel
Big Data Platform	Data Scientists, Data Context Experts, IT, OT	Steward data components within the platform, administer security and interfaces, collaborate with COI group to prioritize and iterate for value realization cases
Analytics Presentation	All	Present actionable information, decision management, data democratization

Examples

Use Case 1 – Liquid Dispatch Optimization

In a low commodity environment, it is essential to leverage time series data assets to describe and infer conditions in the field. Time series alone is effective in describing historical and current conditions. Leveraging the time series data by adding context about the data being reported, and additionally presenting holistic optimization models is a desired outcome for combining time series, big data tools and analytics presentation. In the following example, an oil and gas producer was able to take two readings from tank equipment in the field, enhance the data with real-time calculation, and subsequently solve a prioritization problem across an entire basin utilizing big data toolsets. Specifically, the operator was able to create an optimal dispatch decision with resource constraints in regard to hauler trucks available, market nominations and pricing, multiple oil sales near pickup capacity, while limiting downtime due to capacity limitations at multiple sites.

Purpose-Built Application Data – Asset information about the tank (capacity, volume per inch, location, hierarchical, etc.) was integrated to the time series data platform. The purpose-built applications utilized were production accounting, wellwork, and GIS applications.

Time Series Data – For the time series data component, the company utilized time series data in a read-only repository interfaced from a SCADA server. The readings of interest for this example were fluid level and water interface level in approximately 6,000 oil production tanks. The reading units were in feet with decimals. Utilizing real-time calculations in the time series application, the fluid and water readings were transformed to total inches. The “inches” data stream was then used to transform inches into inventory volume in real-time. Additionally, real-time calculations were designed to determine a rate of change in inventory volume over time.

Big Data – Rates of inventory change, capacity in the tanks, available capacity at the site, and locational information from time series and purpose-built applications were interfaced to the Big Data platform for all tanks in the basin. With this data, a time to capacity and a time to shut-in estimate was calculated individually for all tanks. Based on hauler capacity, potential lost production due to shut-in, mileage and several other prioritization factors, a dispatch recommendation was presented to office personnel to approve via custom analytic review tools. In this case, processing speed for model calculation results was the primary driver for utilizing the big data platform.

The initial implementation of this example created several efficiencies. The first to be realized was adoption by personnel in the company. This combination of electronic assets had an immediate effect on time management, both for field and office personnel (100+ hours per day across field personnel). Additionally, gains were seen in safety and environmental arenas, as well as overall optimization of limited resources for the oil company and hauling companies.

Additional use of this model was explored in “predicting” sales into the near-term, which also has a direct benefit on financial reporting (estimated improvement of 2% per day), marketing nomination, and other key operational tasks.

Use Case 2 – Smart Grid

For the most part, IoT data comes in the form of time series. In order to get the most value from the data, processing, analyzing, visualizing and acting on this data needs to occur in near real-time. For utilities, usage and load on their grid change quickly during the day. This has an effect on efficiency of the grid, power portfolio pricing and overall health and reliability of power delivery to consumers. In some areas, entire transmission systems have become overused and fragile due to population growth and aging equipment. At the same time, consumers have added more electronic devices to the home, such as computers, high-definition TV's, microwave ovens, wireless telephones, and even electronic controls on refrigerators, ovens, and dishwashers. These new appliances are more susceptible to voltage variation. To address these issues without investing in expensive overhauls of their system, utilities are implementing Smart Grid technology to increase efficiency and reliability of their grids which has the added benefit of decreasing overall energy usage. Smart Metering or Smart Grid technology incorporates grid management, resource planning, customer operations and compliance.

Purpose-Built Application Data – Master relationship data, including location, asset type, customer identification and geospatial information was used to maintain a single view of the grid. Weather and demographic information was also utilized. The purpose-built applications or data used were gathered from contract management, accounting, GIS, grid monitoring applications and 3rd party subscriptions and data feeds.

Time Series Data – For the time series data component, the company collected, stored and analyzed in real-time over 50,000 data points per second. The readings of interest for this example were mainly usage and load statistics from homes, businesses, equipment, renewable sources, storage locations and digital substations located on the grid. Current, historical and real-time calculated readings were stored and interfaced to a big data platform.

Big Data – The historical usage and load data was processed to offer predictive and prescriptive recommendations and action. The predictive models were combined with customer information to identify potential preferred demand/response offerings. Prescriptive models managed the power portfolio mix to optimize power sourcing, market pricing and downtime and maintenance planning. In this case, the geographical dispersion of assets, varying asset type, multiple inputs and complex optimization models were the basis for incorporating the smart grid model to the big data platform. Several in-house and commercial analytic visualization tools and applications were used for the different users or customers based on their specific needs.

Implementation of smart grid modelling resulted in efficiencies in grid management and customer experience. The utility experienced an 50% decrease in short-term peak loads, and an 15% decrease in overall peak loads. Additionally, the customers were able to realize an 10% decrease in electricity costs.

Additional benefits could be found in reduction of greenhouse gases, more customer control of their power needs and overall decreased demand on the infrastructure. Predictive maintenance for aging assets could also be a value driver for utilities either by utilizing purpose built analytics in the time series application, or holistic modelling of the entire grid in the big data platform. Ideally, the big data platform could be used to train models that could be applied in the time series application in real-time.

Conclusion

Big Data is integral to IT/OT convergence and complements time series and other purpose-built applications, particularly as the IoT movement gains momentum and streaming data becomes ubiquitous across devices of all types. Companies that leverage the very best of the purpose-built applications while enabling democratization of structured and stewarded data for new insights and value discovery in the Big Data platform can enable a step change in performance and provide a strategic advantage in their market.

For companies to successfully navigate the complex roadmap of an IT/OT convergence and become data-driven with existing data streams and new IoT devices coming on-line, organizations must employ the following best practices.

- Evaluate the state of existing technology, particularly purpose-built applications - Existing systems should incorporate best practices and some unique functions, calculations, or data structuring applicable to a specific discipline that can easily interface to a big data platform. Industrial companies need to have applications that excel in capturing and federating time series, relational, and unstructured data.
- Capital or Resource Investment and Prioritization of key systems and applications - Applications that can capture and provide contextualized, high resolution and standardized data, or consume actionable data streams from advanced analytic platforms (or both) should be prioritized in planning. These systems are key components of a successful Big Data integration initiative.
- Development of strong data governance, stewardship and management programs or oversite – Data Governance and Management should be cross-discipline, and review and enhance data standards within the company based on operational expertise from operations, IT and OT perspectives. Applications that can easily capture and disseminate asset context should be utilized where possible.
- Development of a Community of Interest - Any projects that could monetize the underlying value of Data Assets across a corporation should align with corporate and operational goals and objectives, and be driven by a community that is accountable for results. The Community of Interest (COI) is also integral to developing a data-driven culture. Interface between front-line workers and the COI is integral to identification of value, communication of standards and maximization of the value of data assets.
- Democratization of the data and feedback into Decision Management – Data is only useful if it is discoverable, actionable, and can be put in the hands of decision makers. Additionally, actions taken as a result of data or calculations needs to be tracked, evaluated and utilized to make future decisions better. Investing in tools that can enhance the value of the data, either with in-application real-time calculations, contextualization, visualization or decision management is integral to a successful IT/OT convergence.

Each organization should evaluate the benefits, determine their level of data and application maturity and review their holistic strategy around IT/OT convergence. These five points summarize a recommended approach to combining operational, time-series and IIoT data with business-oriented data that exists in traditional IT platforms. With the proper approach and planning, many common pitfalls can be avoided on the road to becoming a data-driven organization.

References

Gartner - Gartner Says 6.4 Billion Connected "Things" Will Be in Use in 2016, Up 30 Percent From 2015

http://www.gartner.com/newsroom/id/3165317

McKinsey Global Institute - Unlocking the potential of the Internet of Things

http://www.mckinsey.com/business-functions/business-technology/our-insights/the-internet-of-things-the-value-of-digitizing-the-physical-world

National Institute of Standards and Technology (NIST) – Definition of Time Series

http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc41.htm

Digitalist Magazine - How Quickly Are Industries Adopting the Internet of Things?

http://www.digitalistmag.com/iot/2016/03/02/how-quickly-are-industries-adopting-the-internet-of-things-2-04044088

NIST Cybersecurity Practice Guide, Special Publication 1800-2: "Identity and Access Management for Electric Utilities"

https://nccoe.nist.gov/projects/use_cases/idam

siliconAngle - Digging up dark data

http://siliconangle.com/blog/2015/10/30/ibm-is-at-the-forefront-of-insight-economy-ibminsight/

insideBIGDATA - Deriving Value from Data Before It Goes Dark

http://insidebigdata.com/2015/10/12/deriving-value-from-data-before-it-goes-dark/

Gartner IT Glossary – Big Data

http://www.gartner.com/it-glossary/big-data/

NEMA - What is Smart Grid and why is it important?

https://www.nema.org/Policy/Energy/Smartgrid/Pages/What-Is-Smart-Grid.aspx

LinkedIn SlideShare - Big Data Analytics in Energy & Utilities

http://www.slideshare.net/AndersQuitzauIbm/big-data-analyticsin-energy-utilities