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

CIGRE US National Committee
2019 Grid of the Future Symposium

Zero to One: A Utility Playbook for (Finally) Delivering Sustainable Value from Your Synchrophasor Data by Enabling Easy Data Exploration and Rapid Use Case Prototyping

K. D. JONES
Dominion Energy
USA

SUMMARY

In the utility sector, synchrophasor data has long been heralded as a vehicle for transformative change across multiple business units of a utility and the industry, at large. Its unique characteristics - high resolution, time synchronized, phasor domain quantities - set it apart from virtually all other grid telemetry. Dominion Energy's belief in the potential for synchrophasor data has driven nearly a decade of investment in the deployment of phasor measurement units (PMUs) across its transmission grid. Despite the tremendous promise across the industry, the value of synchrophasor data remains trapped in a combination of inactive sensors, legacy data historians, and black-box tools. Why has the utility industry struggled to unlock this truly opportunity laden technology? Dominion has, over recent years, developed and tested hypotheses aimed at answering this exact question. The answer is simple; for Dominion and the industry to succeed, the cost of analytic experimentation must be driven down beneath a threshold that allows data to be frequently interrogated in complex and meaningful ways. To make this a reality, Dominion established a collaboration with a technology startup company to deploy a platform data management, visualization, analytics, and AI platform for high density telemetry such as synchrophasor data deployed in the cloud with the express purpose of driving the cost of analytic experimentation to its physical minimum. This paper distils Dominion's 2019 successes into a playbook for other transmission owners (or related entities) to create similar transformations within their respective organizations.

KEYWORDS
synchrophasor, data, platform, use cases, analytics, value proposition

kevin.d.jones@dominionenergy.com

1. ZERO TO ONE

Peter Thiel, American entrepreneur and venture capitalist, writes in his widely read book *Zero to One: Notes on Startups, or How to Build the Future* [1] that there are two types of transformations in the world. These are (1) the transition from *Zero to One* and (2) the transition from *One to N*. From his perspective, the former is vastly more difficult than the latter. Succinctly translated, creating something from nothing is far more difficult than building on something that already exists. He posits that all truly transformative changes are step changes, from *Zero to One*.

In the utility sector, synchrophasor data has long been heralded as a vehicle for transformative change across multiple business units and the entire industry. Its unique characteristics - high resolution, time synchronized, phasor domain quantities - set it apart from virtually all other grid telemetry. Dominion Energy's belief in the potential for synchrophasor data has driven nearly a decade of investment in the deployment of phasor measurement units (PMUs) across its transmission grid. Originally fueled by government stimulus [2], then made sustainable by forward-looking modifications to internal construction standards [3] that called for PMUs and phasor data concentrators (PDCs) to be deployed anywhere new construction was performed. Most recently, these deployments have been augmented by enabling latent PMU functionality within Dominion's entire fleet of digital fault recorders (DFRs). The combination of these efforts has yielded hundreds of PMUs or PMU-enabled devices across the Dominion transmission grid, measuring nearly 40,000 individual quantities and providing nearly full transmission system coverage via PMU measurements; in many places this coverage is redundant.

Despite the tremendous promise across the industry, the value of synchrophasor data remains trapped in a combination of inactive sensors, legacy data historians, and black-box tools [4]. When framed in this context, the utilization of synchrophasor data for business value has long been a *Zero*. While there are noteworthy exceptions to this assessment, they are often isolated to singleton applications - unable to broadly impact the industry or even the host organization itself. Why has the utility industry struggled to shift this truly opportunity laden technology from a *Zero* to a *One*?

Dominion has, over recent years, developed and tested hypotheses aimed at answering this exact question - taking focused actions to transform its synchrophasor data from *Zero to One*. At the heart of these hypotheses is the notion that the fundamental bottleneck is resources; it is simply still too costly (read *difficult*) to work with any meaningful volume of synchrophasor data. This is exacerbated by the sheer number of potential use cases for synchrophasor data [5] and the challenge of identifying what makes sense within the nuance of a particular organization. Therefore, the utility industry does not work with meaningful volumes of synchrophasor data as part of day-to-day or core business activities. Even control-room applications use miniscule amounts of synchrophasor data (i.e. the most recent minutes or hours) compared to the volumes available for those willing to store it and make it readily available for historical analysis. For Dominion and the industry to succeed, the cost of analytic experimentation must be driven down beneath a threshold that allows data to be frequently interrogated in complex and meaningful ways. Low cost experimentation opens the door to a world of possible use cases and is the key to transforming synchrophasor data from a *Zero to a One*. Once achieved, the answer to the question "just what can be done with all of this data?", the transformation from *One to N*, can be swiftly addressed.

1.1 Purpose

This paper distils Dominion's 2019 successes into a playbook for other transmission owners (or related entities) to create similar *Zero to One* transformations within their respective organizations. This work serves as evidence validating previously presented hypotheses on the best approach for extracting value from synchrophasor data *that is also applicable to the coming point-on-wave sensor deployments*. This playbook contains everything from driving philosophies and validated hypotheses to dos, don'ts and specific success metrics. It is intended to be evidence that the industry can and should modernize its posture towards not only synchrophasor data but also many other transformative mechanisms such as the cloud, working with startup companies, and challenging legacy business drivers and assumptions such as the value of experimentally derived business cases for data. This paper is the story of Dominion's *Zero to One* transformation and a glimpse into the company's subsequent move from *One to N*.

1.2 Just How Much PMU Data Does Dominion Have, Anyway?

Dominion has three classes of PMU devices in the field.

1. Hundreds of standalone PMU devices and PMU-enabled relays - These devices monitor the grid within a critical network and, therefore, can one day serve real-time operations. The total number of individual measured quantities originating from these devices is approximately 15,600, representing 65-70% coverage of the network. Through normal construction, this number will quickly grow to saturation nearing 100%. The centrally located servers that are positioned to facilitate the ingest of this data are slotted for replacement in 2020.
2. Hundreds of DFR devices with PMU functionality enabled - These devices monitor the grid from a non-critical network. The total number of individual measured quantities originating from these devices is approximately 23,800. This represents virtually 100% transmission network coverage.
3. Six (6) distribution-feeder-head PMU devices (not microPMUs) - These devices monitor the grid from a non-critical network.

At scale, the future synchrophasor footprint at Dominion's will include approximately 100% measured coverage of the transmission grid from both the critical network and non-critical networks. This means Dominion will eventually have duplicate measurement redundancy from two separate networks. At current capacity, approximately ***275 TB (e.g. 275,000,000,000,000 bytes)*** of synchrophasor data per year can be generated. This number neglects the need for redundancy or option to perform (lossless) compression. This also excludes the potential storage of derived or calculated values. Dominion estimates that the saturation point for streams from PMUs will be approximately 50,000 streams compared to the current volume of approximately 39,400. This would yield approximately ***350 TB*** of measured values per year.

Of the deployed capacity, Dominion is currently live streaming about 35% of the currently available volume of synchrophasor data with two remaining bottlenecks to be addressed near term. First, the remaining DFR PMUs will have new communication cards installed in the devices to enable stable PMU functionality. Second, the IT infrastructure used to bring back Relay-PMU data will be replaced/upgraded in 2020, thereby opening the door for the full 15,600 streams from these devices. Therefore, Dominion anticipates an aggregate stream volume of 40,000 individually measured quantities by mid 2020 or sooner.

1.3 Why Take Action Now?

A combination of factors contributed to the sense of urgency in (finally) developing a sustainability path for Dominion's synchrophasor program.

1. Performance Bottlenecks in Enterprise Data Historian - Dominion's enterprise data historian could not ingest a small fraction of Dominion synchrophasor data; basic export of data to other systems was glacially slow; visualization of high density data within the historian was virtually impossible; and no architectural elements existed to enable historical data analytics. This realization was actually the driver that triggered the re-evaluation of the state of Dominion's synchrophasor program.
2. Legacy of Investment - Dominion has invested ten years into its synchrophasor program. It has achieved great success in terms of sensor deployment and also been a champion of new technology development such as the Open Source Linear State Estimator [6][7][8]. Despite the noteworthy impact of these successes, they do not alone cement synchrophasors' value into the business model of Dominion. Given the level and duration of previous investment, it is only natural to question the ROI in the present.

3. Organizational Changes - Many of the advocates responsible for the earlier success with synchrophasor technology have moved from engineering roles into formal leadership roles, creating two forces. The first, a positive force, is higher level advocacy for transformation via synchrophasor data. The second, somewhat dangerously, is a void left at the grass roots level where strong organizational dependencies on synchrophasor data have not been established. If meaningful business processes and decisions depend on synchrophasor data, the synchrophasor program can live on sustainably without its originators.
4. Time; And the Second Law of Thermodynamics - There are certainly natural forces within the industry that will, eventually, solidify the value of synchrophasor data by creating dependencies in perpetuity. The danger in waiting for these outcomes is entropy. Between now and that time, Dominion would need to continue to invest in its synchrophasor program based solely on the promise that *one day* it will become necessary. If Dominion fails in this respect, entropy takes over and the last decade's work will decay. A value chain for synchrophasor data must be established today so that it will still be around in the future when it is critical to day-to-day operations. Otherwise, the entire program is at risk.

1.4 What Did Dominion Actually Do to Go from Zero to One?

Dominion established a collaboration with a technology startup company, PingThings (www.pingthings.io) to deploy a Platform-as-a-Service (PaaS) data management, visualization, analytics, and AI platform for high density telemetry such as synchrophasor data [9], deployed in Amazon Web Services' (AWS) GovCloud [10], with the express purpose of driving the cost of analytic experimentation to its physical minimum. Minimizing cost, in this case, includes not only actual cost but also the minimization of the time to iterate. In short, iterative experimentation must be fast and affordable. In doing so, Dominion can move faster than the industry believes possible and ingrain synchrophasor data into as many aspects of the business as possible - creating a virtuous cycle of maximally delivered ROI and sustainability of the synchrophasor program.

A collaboration with a startup company is a significant risk. Because of the importance of the work and the capabilities of the company and its technology, Dominion felt that not only was the risk warranted but that it was manageable. Dominion tempered the risk of working with a startup by (1) taking the time to educate SMEs on the core problems, (2) engaging in deep technical due diligence with the startup's architect (3) effectively lobbying for the solution, and (4) following through with laser focused execution.

Educating the SMEs - There is tremendous noise in virtually every sector surrounding data analytics and machine learning. The author, in particular, dedicated time and effort to get to the heart of the problem Dominion was trying to solve and understand which technologies are appropriate and cost effective and which are not. The physical and technological delineations amongst various offerings made clear that our chosen collaborator was the right choice. This transformation started as a grassroots initiative.

Engaging in Deep Technical Due Diligence – The author spent a significant amount of time directly engaged with the architects and engineers in the startup to execute a rigorous technical evaluation of the technologies that makeup the platform.

Socializing for Success - The author started with a problem, communicated the impact of not addressing the problem, identified the appropriate solution and, bit-by-bit, built an internal coalition amongst technical and managerial leadership.

Excellence in Project Execution - Execution of the initiative required intensive collaboration with the startup company and wrangling of internal resources that were often focused on other, seemingly more important tasks. Without the close partnership provided by a startup company, who by their very nature must be invested in Dominion's success, the *Zero to One* transformation would not have happened.

To be explicitly clear regarding the definition of the *Zero to One* transformation. Dominion classifies *Zero* as having synchrophasor data but not having the ability to cost-effectively store and iteratively interrogate large volumes of historical synchrophasor data. Dominion classifies *One* as having removed any and all impediments to fast, easy data exploration, low-cost analytic experimentation and rapid use case prototyping. Dominion can approach *N* by developing an ensemble of compelling use cases that impact the business in a multitude of ways.

2. THE APPROACH AND THE SOLUTION

Dominion's approach to transition synchrophasor data from *Zero to One* existed first as a set of philosophies that tersely put are (1) minimize the cost of data exploration and analytic experimentation of large volumes of data and (2) develop tools and systems that focus on high complexity workflows of super users as opposed to the broader employee population. Second, it consisted of the research and identification of a purpose-built technology solution as well as a collaborative partner with world-class capabilities and strong well-aligned incentives.

2.1 The Core Hypothesis - Low-Cost Data Exploration and Experimentation

The driving philosophy for the transformation of Dominion's synchrophasor program can be summarized as follows:

While many complex cultural, technological, and organizational blockers limit synchrophasor use across the industry, the fundamental constraint is the cost of working with data. Essentially, *if it is too difficult to work with synchrophasor data, nobody will work with synchrophasor data*. Therefore, success is not granted by the identification of one or more use cases but rather by the adoption of the philosophy that if the cost of working with data can be driven to a minimum, numerous use cases will emerge.

Therefore, to ensure ROI from the Dominion synchrophasor program, every bottleneck should be mitigated to the greatest degree possible, including:

1. <u>Physical Bottlenecks</u> - Maximize the volume and integrity of data captured from the field. Do not downsample or discard data. Do not move data from one system to another unnecessarily to perform analysis (*i.e. don't move the data to the computation, move the computation to the data*).
2. <u>Technological Bottlenecks</u> - Consider the technical challenges at every layer of the technology stack for ingesting, storing, compressing, querying, visualizing, performing computation on, and learning from data. Architect for the most demanding functionality and the complex and unforeseen use cases, not only the regulated use cases (*greatest common factor not lowest common denominator*).
3. <u>Organizational Bottlenecks</u> - Create healthy incentives for users to engage with the data. Socialize successes. Mitigate dependencies on inflexible groups and systems by migrating to the cloud (there are *technological* bottlenecks relieved by this as well).

Why Does Cost Effective Data Exploration and Analytic Experimentation Matter?

Put most simply, cost effective data exploration and analytic experimentation matter not because there are *small* numbers of use cases but because there are *large* numbers of potential use cases. Therefore, the problem is one of searching for the right use cases and evaluating their effectiveness in the context of a particular organization. Broadly speaking, there are three categories of sources for use cases:

1. <u>*a priori*</u> - These use cases ask the question of "What existing use cases might be useful to my business?" There are literally thousands of previously identified use cases in the literature and across the industry [5]. Which ones are most effective for Dominion?

2. _a posteriori_ **-** These use cases ask the question of "How can I address this *specific* issue or event?" With cost effective access to data, when *something* happens (like an equipment failure) or a serendipitous moment uncovers a new truth, lessons learned from data can be transformed into concrete use cases. [11]
3. _a experimentalem_ - These use cases are identified by asking the question "What does the *data* tell me that I might not think of or directly observe on my own?" These are derived purely through experimentation and are not known a priori.

Thus, experimentation and iteration are central to an effective synchrophasor data analytics value chain and driving down the cost is key to enabling it.

2.2 Focus on the Right People: The Super Users

Synchrophasor discussions often focus on serving the wrong people. The typical framing falls into one of two categories. First, the goal is to "get the data in front of operators." This is certainly a healthy goal but an incomplete proxy for the holistic value of synchrophasor data. Second is the enterprise approach to create "one place for all the data" so that it can "be made accessible across the organization" and "create enterprise-wide efficiencies". The downfall of this latter approach cuts even deeper than a focus on the control room. The fact is that only a small population of users within the organization will create use cases of value. The troubling part is these use cases are far more complex than a general-purpose user may require but conversely will be the most valuable. Therefore, when selecting or designing tools, systems, or platforms, focus on the *super users*. These are the individuals that, if properly enabled, can transform the organization with data.

Build for the Super Users, Not the General-Purpose Users

Utilities often develop lowest-common-denominator systems that serve the enterprise. In the world of data, these traditionally manifest as *historians*. The data historian is designed to *archive* the company's data but often does little more. The evolution of historians has led to *full-feature data historians,* which bolt on simple functionalities for supporting low-complexity use cases, addressing the basic needs of the many. Furthermore, the vast majority of users are not skilled in extracting value from high density telemetry, limiting the volume and value of their use cases. Rather, specialists capable of use case development must be able to get their hands on the data. Therefore, mirroring the design of systems for synchrophasor data from traditional data historians is fundamentally the wrong approach. Technologies and data system architectures should be designed to serve a small population of *super users* that have high-complexity, high-value use cases. This hypothesis fundamentally changes the requirements of a system. [12]

Choose the Users that can Contribute the Most

Dominion took the approach of rolling out a solution to enable synchrophasor analytics that focuses on the super user population. In contrast to the 1100 employees in Electric Transmission, the population of super users at Dominion just eclipses two dozen. The criterion for inclusion are simple: (1) deep subject-matter-expertise in electric power systems via graduate degrees or years of experience and (2) the proclivity to solve engineering problems by writing computer code. The super user team spans numerous groups within Electric Transmission: Operations Planning, EMS Engineering, Transmission Planning, Fault Analysis, T&D System Protection, Special Studies, and Engineering Analytics & Modeling.

Untether from the Control Room to Accelerate Business Value Extraction

While utilities often suffer tunnel vision regarding synchrophasor applications by over-emphasizing the role of the system operator, it is understandable how this can occur. One of the many outcomes from the Northeast Blackout of 2003 [13] was a recommendation to increase system monitoring of voltage and current phasor measurements to increase situational awareness [14]. What makes this recommendation special is the specificity given to synchrophasor measurements as part of the solution. If angle walkout, detectable with wide area PMU data, had been seen, [14] operators could have prevented or mitigated the impact of cascading outages and ultimately, the blackout. This catapulted a previously nascent technology into the spotlight and catalyzed investment to deploy synchrophasor technology across the grid over the next 15 years. In doing so, it also solidified synchrophasor technology in the minds of many as a tool for the control room.

The control room is undoubtedly an important audience and consumer for synchrophasor data; however, it is not the only one. Dominion has found need to challenge the view of the operator as the singular or most valuable user of synchrophasor data with a broader perspective that includes engineering and business value propositions for synchrophasors. In fact, one of the enabling factors for Dominion's recent acceleration in the use of synchrophasor data was dividing the synchrophasor program into two (somewhat) orthogonal paths. One path remains focused on control room applications while the other focuses on deriving non-operational business value from the data. Before this bifurcation, the timelines of new technology adoption in the control room artificially limited the adoption of technology by other parts of the business. Dominion could no longer consider use by the control room as the only metric of success. Thus, the importance of non-operational functions - engineering, design, and planning purposes - was elevated so that Dominion could immediately start extracting value from synchrophasor data and chart a path to program sustainability. This shift requires a fundamentally different approach to technology, architecture, and engineering workflows. In one sentence, this is a shift from just-in-time computing with real-time analytics to bulk-computation of historical analytics. The latter much more resembles the "big data" problem and data science workflows seen in many other verticals undergoing data-driven transformations.

2.3 The Case for a High Performance Sandbox Architecture

Focusing attention on super user workflows impacts technology and design decisions throughout the architecture of a data analytics pipeline; these ideas are captured by the idea of a *high performance sandbox* [12]. What exactly is meant by *high performance*? And what exactly is meant by a *sandbox*?

The performance of such a system is not simply measured by base performance metrics of the platform: computational, storage, or memory capacity. Instead, the system concept expands to include users and the critical performance metric of interest is the effectiveness of these users. Therefore, a high performance sandbox is one that utilizes *whatever resources and architecture required to make the team high performance*. This impacts design decisions at every layer of the architecture all the way from user interfaces and APIs all the way down to the underlying data structure persisted on disk.

Furthermore, *sandbox* does not suggest that the system is unimportant, insecure, or only for toy problems. Sandbox implies a high degree of flexibility, adaptability, imagination, and a focus on high-value creation potential. It is owned and managed by the business, hosted in the cloud, and facilitated by experts in cloud and devops. Fundamentally, it does not force users into black box analytics but rather exposes all functionality programmatically.

Seek Transparent, Flexible Capabilities not Black-Box, Singleton Applications

During research into the right technology and vendor platform, it was evident that many offerings attempted to provide out-of-the-box analytic tools (vertical silos) as opposed to highly flexible platforms with programmatic interface/API that were designed to be extended (horizontal layers). Because of the broad spectrum of potential use cases, Dominion sought out a solution which puts the SME in the driver's seat. Rather than letting a vendor (often poorly) define a methodology or a use case, Dominion chose to enable its subject matter experts to define the use cases through high-performance programmatic interfaces into the analytics platform.

2.4 Focus on the Right People: Choosing the Right Vendor

Organizations attempting to develop and deploy a data platform can either build one or buy one. If the organization decides to take on the project itself, it can be led by either the IT department or the group within the organization requesting the platform. If an external supplier is chosen, there is a wide range of different sized companies that can be selected ranging from the large incumbent, dominant in the industry (the 800-lb gorilla) to the small, technology-forward startup.

Don’t Roll Your Own Big Data Platform

Dominion has seen many organizations (not just utilities) attempt to construct their own systems. The problems with this approach are manifold. Making sense of the diverse collection of the ever-evolving technologies is difficult even for experts. Power systems engineers, however bright, are not equipped to develop suitable technology stacks to solve the fundamental problems of managing an analytics pipeline for high density telemetry. Furthermore, distributed computing systems are difficult to build and once built must evolve continuously over time. This is a massive undertaking, best left to people focused on the craft. Lastly, individuals with these skill sets are drawn to career paths outside of core utility/energy business by strong interests or compensation. Therefore, even if an organization should succeed, it is then beholden to the original, internal authors of the platform, who may ultimately leave the team or even the company.

Be Wary of Enterprise IT Solutions

While most utilities’ first instinct is not to “roll their own big data platform” they may often look first to their IT organization to either specify and procure from an outside vendor or build and assemble themselves. However, there are still some fundamental shortcomings to an IT designed or IT owned approach. First, as previously mentioned, be wary of the proclivity of IT to build enterprise data systems that serve the broadest user base possible at the expense of high-value use cases. Second, IT will often require business cases upfront when specifying large data systems. This mindset results in solutions without the ability to experiment cost-effectively. Third, IT is not a subject matter expert in synchrophasor data analytics, nor do they understand the burdens of working with high density telemetry. There must be a champion in the business that understands these profound differences defining the requirements and perhaps even taking ownership of the initiative. Fourth, an IT solution may compromise the flexibility of infrastructure. For example, it is important for driving down the cost of analytic experimentation to be able to easily and quickly scale up (or down) compute to address dynamic, day-to-day analytic needs, add storage as the data footprint grows, and adapt architectures and software installations to meet the ever-changing demands of the analytic team. Lastly, IT is typically a cost-center for the utility and is, therefore, not driven by many of the necessary financial incentives to match the performance of a small, nimble team of experts. Make no mistake, the role of IT in the success of the initiative is critical but should be structured as described later in this paper.

Consider the Cost and Speed of the Large Industry Incumbent

There are many scenarios where the large industry incumbent can provide the exact solution required by the business. That is one of the reasons they have become the large industry incumbent. However, when the solution requires advanced, new technology to solve the problem, there can be several impedance mismatches between the incumbent’s solution and the right solution. Broadly speaking, these mismatches are high cost, insufficient performance, long lead-times, and long upgrade cycles. While there are absolutely many merits to the contributions of these types of organizations, when it comes to working with large volumes of synchrophasor data, these four aspects each undermine a different aspect of the core hypothesis of driving down the cost and iteration time of progression through the analytic pipeline.

Consider an Alternative

Working with a startup offers a very different experience than working with a large vendor. They can be fast, responsive, largely devoid of bureaucracy, and offer state-of-the-art technology delivered by employees that would never be hired directly by a utility. Thus, startups offer a significant risk/reward option. There can be high risk given the size and potential under capitalization of the startup, but also tremendous reward. A utility’s business directly impacts a startups ability to survive and raise capital. Therefore, the success of the initiative within the utility and the success of the startup are intertwined. This creates strong incentives for all involved to achieve excellence. This is the path chosen by Dominion.

This may not be a frictionless endeavor. Everything within a utility—contracting, legal, IT, etc.—is designed to interface with (and protect itself from) large vendors where the alignment of incentives may be shallow at best. Therefore, working with a startup will be a learning experience for the organization involved. However, if risks and incentives are managed properly and there are adequate champions within the business with an appetite to create value from the solution, it can be a step change for the utility.

2.5 The Right Tool: Choosing Technology Designed to Solve this Specific Problem

Choosing the correct technology stack is of utmost importance to enable the *Zero to One* transformation and drive down the cost of data exploration and analytic experimentation. Success in this regard requires understanding the fundamental characteristics of the data problem to resolved, understanding the big world of "big data" and time-series databases, and working inside secure but flexible infrastructure. The author dedicated approximately six months starting in the fall of 2017 to understand the underlying physical limitations preventing Dominion's at-scale utilization of synchrophasor data. The following sections are high-level summaries of the lessons learned that guided the decisions made that led to the successes in 2019.

Understand the Uniqueness of the Problem

Neither data nor the mathematical transformations applied to it are created equal. Data can take on many shapes and sizes and accumulate into a variety of lengths (time), widths (number of sensors), and depths (resolutions). Analytic methodologies range from simple algebraic transformations all the way to deep learning algorithms. Not surprisingly, each industry has a unique set of requirements for the time series sensor data that captures the behavior of their systems. First, the sensor data itself is what one might consider *low complexity*. In other words, it is simply a timestamp and a floating-point value (measured quantity) of some precision. Second, sensors must sample data at frequencies fast enough to capture the underlying, complex dynamics of the physical world. Given the speed of electrical and electro-mechanical disturbances, there is an obvious need to sample at higher frequencies. Finally, interesting behaviors can be intermittent and infrequent. Thus, to learn from such system dynamics whose importance may not be initially recognized, industrials want to retain data at full temporal resolution indefinitely. This becomes absolutely non-negotiable when the organization wants to deploy machine learning and deep learning algorithms to extract additional value from these sets of dense telemetry. Such algorithms need as much data as possible both to establish baseline behaviors of the system in question and for examples of events of interest for detection, classification, categorization, and prediction. ML and DL become essential at such scales simply because the inbound data fast outstrips the ability of humans to comprehend and process. This is in contrast to (1) lower resolution telemetry (SCADA) and (2) non-telemetry time-series data such as click-stream data from the web or Twitter feed data. Systems that are optimized for these data types fail to provide the desired performance for high density telemetry of 1Hz and above (i.e. synchrophasor data).

Understand the History of the Field of Big Data and Distributed Computing

There have been three generations of cost-effective big data architectures in this millennium. To disambiguate these terms from buzzwords, *big data* is considered to mean any amount of data whose processing and/or storage exceeds the capabilities of a single computer. Given that the amount of desktop computing power changes over time, this is a dynamic threshold. The term *cost-effective* is used to denote systems that employ a mostly software-only approach to big data versus one that employs unique or specialized hardware that tends to follow an exponentially increasing cost curve. More specifically, cost effective systems are built on large arrays of commodity hardware with software designed to orchestrate large computation jobs.

The first generation of big data systems unofficially started in 2006 with the development and release of the open source distributed file system and map-reduce framework supported by Yahoo known as Hadoop [15]. It's focus was to enable the rapid processing of the type of data prevalent at the time using the analytics of the day. As large volumes of data became more prevalent, machine learning (ML) rose to prominence due to its ability to extract more value from this data. However, training ML algorithms required a data access pattern different from the map-reduce paradigm, which persisted each intermediary computation to slow permanent storage. This need sparked the rise of a second generation of big data platforms that enabled distributed, in-memory system that made iterative calculations orders of magnitude faster. The first Apache Spark paper, from UC Berkeley's famous AMPLab, was in 2010 [16]. While ML algorithms effectiveness continued, a new generation of neural networks, the so-called deep networks due to their ever-increasing number of network layers, came to prominence, again because they could create even more value from the ever-growing data sets. The computing resources required to train deep learning (DL) networks is so great that it is prompting new silicon to be created using an entirely new architecture for AI-specific processors. Today, a third generation of big data systems are arising that are architected specifically for a type of data (images, video, relational, time series, etc.), optimized in every way possible for performance. Now that nearly all things are possible in the cloud given enough virtual machines, the question is how cost effectively can a particular task be done. Therefore, because of the uniqueness of the problem of working with large volumes of synchrophasor data, Dominion needed to identify a platform technology stack from the third generation of big data platforms.

Understand the Underlying Technology: Time Series Databases

Over the last few years, time series databases have risen to prominence in Silicon Valley with a contemporaneous commitment of over $150M investment in this space. Companies like Timescale are adding time series capabilities to the open source relational database Postgres. This is a great fit for the organization that already uses Postgres and wants to add relatively low frequency time series data. InfluxData has built its own custom time series database for consumer IoT and the metrics world, where data rates seem to top out at approximately 1Hz. While these data stores make it easier to work with some time series data, they were designed for a far different velocity of time series data than what utilities now face with synchrophasor data.

There is another group of relevant database technologies that are often targeted toward industries: so-called in-memory databases. These datastores are blindingly fast because they keep all data in the computer's main memory versus persisting the data to disk. This allows very low-latency data queries and analytics if and only if the data is in memory. Thus, they perform exceptionally well supporting the use cases for which they were designed. However, these types of databases are not well suited for historical analytics of timescales beyond 24-48 hours.

Dominion's decision to utilize the PredictiveGrid was strongly influenced by the design and performance of the Berkeley Tree Database (BTrDB) (pronounced "better-D-B") [17]. This technology, originally a brilliant research project that has been engineered into a scalable, enterprise-grade component of the PredictiveGrid platform, is specifically designed and optimized for high density (10Hz to 1GHz) physics-based (low-complexity) time-series data (e.g. high-density telemetry such as synchrophasor data). The author recommends pursuit of the associated references for more information about BTrDB.

Select a *Platform*, Not a *Solution* or a *Tool*

The word *platform* has been used frequently in this paper and not *solution* or *tool* for good reasons. All too often, a vendor will provide a *tool* or a *solution* that attempts to solve a very particular, narrow problem for the utility. Unfortunately, this solution is a black box, by design, that will not interoperate with other solutions, by design. Such solutions represent a vertical integration strategy designed to promote vendor lock in and generate substantial, potentially unnecessary fees. Innovation happens best when silos are broken down into interoperable horizontal layers of capabilities. As alluded to earlier, Dominion believed that it was more valuable to enable the SMEs with high performance tooling and programmatic access to the data as opposed to a black-box, off-the-shelf, fixed, analytic solution. This is reflected in our selection of the PredictiveGrid over "competing" technologies.

Build for Today *and* Tomorrow

For a sensor data platform, the industry's general directions are obvious - more sensors will be deployed, and future sensors will generate far more data. Machine learning and deep learning will be required to maximize the value created from this data. One such example is the transition to point-on-wave (POW), a sensor that will increase the amount of data by one or more orders of magnitude. Thus, it is ideal to pick a platform that can scale to manage this amount of data and the analytic use cases over the next decade. The PredictiveGrid enables performant read, write, and compute for physics-based time series down to nanosecond resolution. This will enable Dominion to fuse together its synchrophasor data and (eventual) continuous POW for greater analytic capability in the future.

Cost effective data ingest is also still valid in the future. The total expected synchrophasor data load for Dominion is approximately 50,000 streams. Each stream contains 30 measurements per second, yielding a total of 1.5M points per second. A test instance of the PredictiveGrid platform of approximately similar hardware specifications was benchmarked to support ingest and processing of 90,000 streams, each with 120 measurements per second, nearly 11M points per second - nearly 7x the theoretical maximum synchrophasor data load of Dominion Energy. With an aggressive assumption of a future where PMUs have been completely replaced by continuous point on wave sensors. Given 20,000 point-on-wave streams provides 100% coverage (fewer physical quantities, essentially only voltage and current waveforms, would be measured and reported at this frequency). Assuming a 2KHz sampling rate, this sensing system generates 40M points per second. PingThings has demonstrated data platforms of capable of ingesting over 50M points per second.

Establish an Appropriate Role for IT

While not the driving force behind this project, IT still played a supporting role in the overall project and was brought into the initiative from the beginning. Alongside IT, Dominion engineers performed the due diligence risk assessments prior to the project's inclusion in the 2019 capital budget. This was the first risk assessment that Dominion IT had done for a PaaS in the cloud and the first time working with the vendor. The selection of AWS GovCloud as the cloud-provider was a significant component of the success of the risk assessment. The other highlights of the security requirements needed for passing the risk assessment include:

1. Secured network connection via VPN for streaming data
2. Secured VPC within AWS GovCloud
3. Encryption of data in motion via Gateway Exchange Protocol
4. Encryption of data at rest via the PredictiveGrid
5. Single-Sign On (SSO) services like SAML, OAuth, and Active Directory so Dominion controls user access

Several key lessons learned from this exercise helped to reduce much of the negative mythology surrounding the cloud. All typical security mechanisms required by the utility can be provided within the cloud (sometimes better than on-premises) if the vendor is competent and invested in working with customers with sensitive data. Upon completion, the IT Solutions Architect said that "this was one of the best risk assessments that I have ever done." This is a testament to the maturity of the cloud and the capability of the startup company - both of which tend to be plagued by rampant, negative mythology in the utility sector. IT sign-off on the approach was fundamental to its success.

With the risk assessment completed, IT had two well defined responsibilities while every other milestone was dependent on the business or the vendor, allowing for a greatly accelerated timeline. First, they were responsible for establishing the network connection from the Dominion network to the AWS GovCloud environment. Second, they were responsible for establishing the Dominion side of the single-sign-on handshake for Windows AD authentication into the PredictiveGrid platform. From this perspective and others, there are tremendous advantages to working with a PaaS solution for the sake of initial deployment timelines, future upgrades, support, and maintenance.

Always Have a Backup Plan
There were still several opportunities for significant delays to Dominion's *Zero to One* transition of synchrophasor data. The objective was to get engineers working with data in the platform as soon as possible. Dominion included two extra steps to ensure that, if critical dependencies suffered delays, we could still get engineers working with data in the platform as soon as possible. The first step was for the vendor to seed the platform with a small amount of historical data before streaming went live. This allowed engineers to begin working with data and learning the tools 2 months earlier. The second step was the establishment of a VPN connection into AWS GovCloud and the PredictiveGrid instead of waiting on a direct physical connection (the actual mechanism is an IPSec tunnel). The earliest a direct physical connection could be established was estimated to be three months after a VPN connection could be established. Also, the direct physical connection is subject to the schedules of the telecommunications provider which may change dramatically. At the time of this writing, Dominion is still working to establish a direct physical connection into AWS GovCloud. If a delay is possible, assume that it will happen. This serves as a testament to the platitude of hoping for the best and planning for the worst. With forward thinking, Dominion engineers created resilience mechanisms preventing six months of delays, drastically accelerating the creation of use cases.

2.7 Seek Out Informal Industry Advisors
Another component of Dominion's orchestration of its synchrophasor analytics initiative was an informal collection of industry advisors. These individuals are active in the synchrophasor community, are stakeholders in the proliferation of synchrophasor technology within their respective companies and have collaborative relationships with key Dominion personnel. The power of these relationships comes from their lack of formality. Each advisor relationship with Dominion is one-on-one. Communication and ideas can flow much more freely in this environment than within an industry working group. Again, the desire here is not to form yet another obligation, but rather to serve as a medium for easy communication regarding ongoing synchrophasor analytic use case development. The outcome is two-fold. First, Dominion can solicit innovative ideas from its advisors on use cases for synchrophasor data. Second, Dominion can socialize its success with particular use cases to help accelerate the overall synchrophasor adoption across the industry.

2.8 Facilitate Short (But Frequent), Targeted Workshops
No matter the technology stack, working with large sets of data through code is a skill that most power systems engineers do not possess. Furthermore, it is still atypical for a utility to have data scientists on staff. High performing individuals with these skill sets typically do not seek employment in the energy sector. Therefore, it is more feasible to teach the power systems engineers the coding and data science skills than the other way around. Therefore, appropriate training is key.

Dominion also approached training from a slightly different perspective. Rather than host lengthy training classes, Dominion has hosted a series of short workshops, consisting of a half-day structured class followed by a half-day practical exercises where the users actually write code and work with actual Dominion data - not sample data from the vendor. This latter half is an opportunity for the super users to engage one-on-one with the vendor to solve problems directly related to their workflows. These workshops are hosted quarterly as Dominion evolves the skill sets to this new paradigm. The workshop format also allows Dominion to evolve the focus of the workshop over time to meet the current needs of the users. Essentially, if there is a large, fixed volume of training material, it can either be covered all at once with the risk of forgetting most of it or, preferably, divided up into smaller sessions that occur more frequently as each session serves as spaced reinforcement of the material and a reignition of the underlying project.

Select Targeted Content and Optimize User Engagement
The first workshop was hosted at the end of May and was focused on:

1. Python development using Git/Github, a distributed version control system.
2. Literate programming and data analytics with Jupyter Notebooks [18]

3. Basic data science libraries like Pandas (Python Data Analysis Library) and NumPy (fundamental scientific computing library for Python)
4. Visual analytics using the platform's multi-resolution plotter [19]
5. Querying/writing data from/to the platform
6. Basic time series and power engineering analytics.

The second workshop is scheduled for mid-September and will be focused on the following:

1. Efficient utilization of computing resources for large data sets
2. More Pandas and NumPy
3. Long-running stream-analytics using the DISTIL Framework [20]
4. Using *conduit* (a Python library for Power Engineering Analytics on the platform)
5. Working with platform metadata

The third and final workshop for 2019 will be scheduled for early December. Furthermore, the vendor persisted the training material in each user's profile on the platform so that they can engage in self-guided learning after the workshop and users unable to attend can still get much of the learning experience. Something that Dominion has not yet done but is considering is to have internal code-a-thons or hack-a-thons targeting synchrophasor analytics.

3. ZERO TO ONE IN UNDER 4 MONTHS

One remarkable characteristic of Dominion's transformation of synchrophasor data from a Zero to a One is the speed at which the transformation took place. It is remarkable not just on utility/corporate time scales but even on objectively measured time scales. There is no shortage of stories where utilities struggle with digital transformations and big enterprise data historian rollouts. These initiatives may be drawn out for years yet still not reach their full potential. From project kickoff to live data streaming into Dominion's PredictiveGrid platform took less than four calendar months and occupied the partial focus of less than 10 individuals (including engineers, IT, and the vendor). This initiative is proof that truly impactful change is possible in utility companies when roadblocks are removed and incentives are aligned. This was made possible by an ensemble of factors:

1. Selection of the correct technology stack (data platform)
2. Collaboration with the right vendor partner
3. Minimizing organizational overhead and bureaucracy by
    1. Leveraging a PaaS solution in the cloud
    2. Championing the initiative from within the business instead of from IT

The ensemble of these decisions made the below timeline possible.

Project Timeline & Milestones

- <u>January 2019</u> - The IT capital project for the deployment of the PredictiveGrid and streaming of synchrophasor data into AWS GovCloud officially begins. Internal kickoff meetings are held by the business to establish project goals, milestones, and timelines.
- <u>End of January</u> - The contract for 2019 with PingThings is finalized and executed. Work begins for PingThings. Design discussions begin for the network connection into AWS GovCloud. Two plans emerge that play out on different timescales. A VPN for the short term and a direct physical connection for the long term.
- <u>Mid-February</u> - Dominion's instance of the PredictiveGrid platform is live in AWS GovCloud. Historical data is pre-loaded for testing. Feature testing begins by the engineering team. Design work continues for the VPN connection. Direct connection on hold until June.

- March - Small group meetings with super users from each department/group to socialize Dominion's new capabilities and its analytic roadmap. First super user Workshop scheduled for the end of May. VPN testing is successfully completed.
- Early April - GEP ingestor testing begins and finishes. Dominion successfully streams its first PMU over the VPN connection into AWS GovCloud and the PredictiveGrid.
- Mid-April - Dominion successfully streams its first ten PMUs over the VPN connection into AWS GovCloud and the PredictiveGrid.
- Early May - Dominion scales up the synchrophasor stream into PredictiveGrid to include ~25% of the total available volume of PMU streams.
- Late May - Graduate student interns arrive for the summer. Super User Workshop #1 is hosted for over 20 individuals across multiple groups within Electric Transmission. Graduate interns receive work assignments from their mentors.
- June-July - Use case prototyping and development across graduate interns and employees. See the following section for identified use cases in this time period.
- August - Graduate intern report out to Dominion leadership including 3 projects focused on synchrophasor analytics using the PredictiveGrid. This document was written.

4. OUTCOMES

In summary, Dominion was able to work with the vendor to go from contract to live data in 4 months. This is objectively impressive on any timescale. Secondly, with little training, Dominion was able to build a team of super users and graduate interns and identify numerous use cases with little guidance or direction over a 2-month period. These represent substantial shifts in capabilities that should continue to pay dividends into the future and serve as foundational lessons for future endeavors. Much like the internet or smartphones represent irreversible changes, Dominion's new capabilities have reset expectations beyond the point of no return.

4.1 Identified Use Cases in the First Two Months

A useful qualitative metric demonstrating the impact of Dominion's transformation is the volume and type of synchrophasor data use cases prototyped and evaluated for usefulness thus far. The use cases below, described by methodology and utility, makeup only the use cases explored between June 1 and July 31, 2019. This work product does not represent the full capacity of the entire super user population but only a small handful that were able to dedicate portions of their working time to this effort (recall that utilization of the platform and synchrophasor analytic prototyping was not mandated but simply enabled). Again, these are the use cases derived by the super users and graduate students without mandate or provocation. Furthermore, none of these use cases were economically feasible before deployment of the platform. Each of the use cases falls into one of three categories:

1. Ready-to-use on a daily basis
2. A new but foundational component to support future analytic workflows
3. Ready-to-scale to create a new analytic workflow

Human-Scale Data Exploration and Visualization - The platform ships with a data exploration tool for quickly and easily visualizing and exploring dense telemetry. The entire history of one or more streams can be selected and instantly viewed on a dynamic chart control that functions like "Google Maps for time series." A user can zoom and pan many orders of magnitude in resolution and time to explore streams. This enables the user to quickly identify the stream(s) valuable to their workflow and specific segments of time with the data they need. There are a plethora of minor use cases that involve casual or very specific data exploration. Previously, because the data was not easily available, Dominion did not utilize synchrophasor data where a result was needed in a timely manner. Near instantaneous synchrophasor visualization has fundamentally altered workflows too numerous to list here. Once Google Maps made using cartographic information fluid, fast, and easy, unfolding a map to determine directions seems downright archaic. Previous attempts at data exploration in other systems would result in crashing of the applications when viewing 8 - 24 hours' worth of data for a single stream. In the words of one of the frequent consumers of synchrophasor data, “the data exploration capabilities alone were worth every penny”.

Fault Timestamp Identification through Data Exploration - The Fault Analysis team can use the fast and easy data exploration tools to identify the exact timestamp of a fault on the system. This enables them to identify and query the waveform data they need much faster. This results in a smoother and faster workflow for the fault analysis engineers who are supporting restoration operations and can positively impact system reliability metrics. The impact to reliability metrics greatly outweighs the superficial time savings experienced by the engineer. Dominion is also working to integrate waveform data into the PredictiveGrid to support further fusion of synchrophasor data and waveform data.

Data Quality Assessments - The underlying time-series database provides built-in stream statistics for associative values such as *point count* within hierarchical time windows. This makes it very easy to frequently interrogate and monitor stream health. This helps Dominion to minimize the resources required to battle the entropic forces of data quality issues.

Post-Event/Post-Mortem Analysis and Reporting - Engineers have developed varieties of Jupyter notebook templates to support different event types. When an event of interest or an equipment failure happens, the engineers (leveraging the templates) curate an analysis report that can be socialized across the company. The net benefits are two-fold. First, this creates a positive psychological association between synchrophasor data and feeling *in-the-know* regarding a recent event, malfunction, or failure. Second, this increases overall organizational situational awareness. Over longer periods of time, event analysis can derive quantitative KPIs useful for tracking macro system behavior. Because of the labor associated with generating these types of reports, few were generated in the past. The platform enables rapid delivery of reports while the content still matters to stakeholders.

Symmetrical Component Analysis - As an example of a descriptive analytic, symmetrical components are an engineering metric that can tell us much about the performance, health, and design of the network. Engineers used the DISTIL framework to run a symmetrical component analysis on the archive of data. Of particular interest are the geo-spatial or temporal correlations of the negative and zero sequence magnitudes. Dominion anticipates comparisons across days/weeks/months/seasons as well as regions of the grid to inform power quality assessments, equipment health issues, equipment design issues, and overall system performance. This analytic has been developed and tested at scale and is ready to use on a daily basis. However, the resulting data could potentially be analyzed from a number of different perspectives, opening up further opportunities to transfer into other workflows.

Conduit: A Power Systems Engineering Library - It quickly became evident that engineers would benefit from a library that sits just above the database API to provide for a number of engineering analytic features such as (1) unit conversions (L-L, L-N, p.u.), (2) power calculations (P, Q, S, pf), (3) compound streams (Phasor, Phasor Pair, Phasor Group), and (4) their associated calculations such as symmetrical components. Dominion worked with the vendor to specify and develop this library with a focus on expressiveness and memory safety. This is a foundational analytic designed to increase the effectiveness of the engineer but also natively contains several analytic use cases ready for daily use.

Spatio-Temporal Data Conditioning - Engineers also spent time developing a spatio-temporal data conditioning tool for recovery missing data from streams that can be included in other analytic workflows to maximize the integrity of the input data streams. This analytic is seen as a foundational component geared towards enabling further analytic workflows.

Voltage Stability Metrics - Engineers developed metrics for bus Thevenin Equivalents to measure relative bus loadability, voltage coherency metrics for grouping similarly behaving clusters of buses, and voltage sensitivity metrics for assessing maximum controllability. The impact of such performance metrics is two-fold. Primarily, it can be used as a methodology to identify locations for voltage regulating equipment such as STATCOMs. Secondarily, it can be used as a methodology to identify potential locations for renewable generation. These prototyped performance indicators are ready to scale up to a large-scale analysis to aid in identification of investment opportunities for Dominion.

4.2 Be Open to Non-Obvious, Positive Impacts

There were many unintended, though positive, side effects of this initiative. There are specifically three worth identifying.

Business-Led IT Capital Projects - Dominion Electric Transmission had success with the model of an IT capital project that was envisioned and largely driven by the business instead of IT alone. This is a model that, when appropriate, Dominion will continue to use in the future.

Working with Startups - Dominion's work with PingThings has shifted internal perspectives on working with startup companies. While not all startup companies are equal, Dominion now understands ways to engage with startup talent while mitigating risk factors.

The Cloud can be Better than On-Premises - It turns out that synchrophasor data was an excellent vehicle for Dominion to gain experience in working with the cloud and commercial cloud providers. The reasons for this are because it (1) has not yet become a mission critical system and is therefore a great guinea pig and (2) because it is a very demanding use case due to the sheer volume of data transported and analyzed, perfect for the scalability of the cloud. This was highlighted by Dominion IT's satisfaction with how the vendor handles maintenance and upgrades of the PredictiveGrid with zero data loss juxtaposed with the sheer volume (in TB) of data that is managed. Thusly, IT is celebrating this project as a success story for the cloud at Dominion.

5. FUTURE WORK AND NEXT STEPS

As compelling as 2019 has been, far more opportunity lies ahead. Dominion has transitioned its synchrophasor data from *Zero to One*. What remains is the transformation from *One to N*. Dominion's focus going forward will occupy several key dimensions. These include:

Increasing Super User Engagement and Celebrating Successes - Dominion will work to sustain an environment for sensor analytic development free of bottlenecks and properly incentivized users so engagement happens organically. While some use cases may require mandated development, unbridled creativity often results in the most impactful results.

Aggressively Prototype and Evaluate Meaningful Use Cases - As time passes, more and more data will populate Dominion's PredictiveGrid platform. This happens through the addition of new devices to the stream and simply more time elapsing. The more data that is available, the more natural resources available for extracting business value.

Data Fusion - Dominion is also exploring the integration of other telemetry into the platform. Targets include SCADA, DFR records, and eventually, continuous point-on-wave data. The purpose is not to create "one place for all the data" but to unlock the value of these data sets in the same way that Dominion has unlocked synchrophasor data.

Focus on Descriptive Analytics First - In contemporary discussion, it's hard to talk about *data* without talking about *machine learning (ML)* and *artificial intelligence (AI)*. However, ML and AI are not magic boxes that suck in data and output insights. An organization typically evolves through a series of analytic phases. This begins with descriptive analytics that characterize a data set followed by predictive analytics that attempt to make predictions based on a data set only to be superseded by prescriptive analytics. There will undoubtedly be applications on the right-hand side of this spectrum that use synchrophasor data. However, Dominion is strongly pursuing the realm of descriptive analytics that characterize the performance, health, and as-built design of the grid. Dominion believes this to be a rich set of information with many opportunities for transformation of the business. A premature emphasis on ML and AI (often simply because of the hype) will undermine the fundamental mission of extracting value from synchrophasor data. Will there be exceptions? Absolutely. But on the whole, the industry's initial focus should be on descriptive analytics that characterize the grid.

Scaling Analytic Workflows - Dominion has positioned the correct technologies to support large-scale historical analysis of synchrophasor data from across the system. However, there are many analytic-specific challenges to scaling a prototype of a particular methodology into an analytic workflow that can be applied to the entire archive. Many of these challenges are related to the software architecture of the analytic. Dominion's strength will be in applying its subject-matter-expertise in an environment conducive to rapid prototyping. Frequently, use cases will be identified that require a greater degree of software engineering expertise. For many of these cases, Dominion will work with our vendor partner to scale up individual use cases into full-blown analytic workflows.

BIBLIOGRAPHY

[1] Thiel, P. A., & Masters, B. 1. (2014). Zero to one: notes on startups, or how to build the future. Unabridged. London, Virgin Books.

[2] Recovery Act: Smart Grid Investment Grant (SGIG) Program, https://www.energy.gov/oe/information-center/recovery-act-smart-grid-investment-grant-sgig-program

[3] Thomas, Kyle, PMU Installation and Placement, October 14, 2015, NASPI Working Group Meeting, Chicago, IL.

[4] Murphy SP and Jones KD, "Learning from Data: Fixing the Analytics Pipeline to Increase the Rate of Grid Evolution," 2017 Grid of the Future Symposium, Cleveland, OH, Oct, 2017.

[5] Jones K, The Challenges of PMU Scalability – NASPI Working Group Meeting, Oct 23-24, 2018, Philadelphia, PA

[6] Jones, Kevin D. et al. "Three-phase linear state estimation using Phasor Measurements." *2013 IEEE Power & Energy Society General Meeting* (2013): 1-5.

[7] https://github.com/kdjones/lse

[8] https://github.com/kdjones/openLSE

[9] Murphy SP, Schuman J, Jones KD, Bariya M, Andersen M, A Universal Platform for Utility Sensor Analytics and Artificial Intelligence, 2018 Grid of the Future Symposium, Dulles, VA, Oct, 2018, Paper Submitted.

[10] https://aws.amazon.com/govcloud-us/

[11] Jones, Kevin D, Turning 10 - A Decade of Synchrophasor Technology at Dominion Energy, April 16, 2019, NASPI Working Group Meeting, San Diego, CA

[12] Jones, Kevin D, The Role of a High-Performance Sandbox in Your Synchrophasor Analytics Pipeline, April 24, 2018, NASPI Working Group Meeting, Albuquerque, NM

[13] https://en.wikipedia.org/wiki/Northeast_blackout_of_2003

[14] Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations (PDF). Energy.gov – Office of Electricity Delivery & Energy Reliability (Report). U.S./Canada Power System Outage Task Force. United States Department of Energy. April 2004.

[15] https://en.wikipedia.org/wiki/Apache_Hadoop

[16] Matei Zaharia, Mosharaf Chowdhury, Michael J. Franklin, Scott Shenker, and Ion Stoica. 2010. Spark: cluster computing with working sets. In Proceedings of the 2nd USENIX conference on Hot topics in cloud computing (HotCloud'10). USENIX Association, Berkeley, CA, USA, 10-10.

[17] Andersen M and Culler D, BTrDB: Optimizing Storage System Design for Timeseries Processing, Fast '16 14th USENIX Conference on File and Storage Technologies, Feb 2016.

[18] Kluyver, Thomas, Ragan-Kelley, Benjamin, Pérez, Fernando, Granger, Brian, Bussonnier, Matthias, Frederic, Jonathan, Kelley, Kyle, Hamrick, Jessica, Grout, Jason, Corlay, Sylvain,

Ivanov, Paul, Avila, Damián, Abdalla, Safia, Willing, Carol and Jupyter development team, (2016) Jupyter Notebooks – a publishing format for reproducible computational workflows. Loizides, Fernando and Scmidt, Birgit(eds.) In Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press. pp. 87-90 . (doi:10.3233/978-1-61499-649-1-87).

[19] Kumar S, Michael P Andersen, and David E. Culler, Unifying data reduction in storage and visualization systems, SIGMOD'18, June 2018, Houston, Texas, USA

[20] M. P. Andersen, S. Kumar, C. Brooks, A. von Meier, and D. E. Culler, "DISTIL: Design and implementation of a scalable synchrophasor data processing system," in Smart Grid Communications (SmartGridComm), 2017 IEEE International Conference on, 2017, pp. 271–277.