OnPoint Consulting

The Big Data Lifecycle

Sean Kramer, Business Experience Designer

The rapid digitalization of information is forcing the hand of companies to leverage 'big data' to solve increasingly complex organizational and business challenges of our always-on digital world, or otherwise risk being overtaken by more data-driven competitors. Yet the allure of big data is often met by uncertainty about how to make sense of it all. Organizations interested in advancing their economic interests, improving the efficiency of services, and even predicting future outcomes with remarkable accuracy, should focus on the "Big Three of the Big Data Lifecycle": data preparation, data visualization, and data security.

Why It Matters

The right data-driven strategy can make sense of our interconnected world by enabling near real-time decision-making based on sound information analysis rather than guesswork. 'Big data' in its raw state is simply an abundant resource that is available to inform, rather than explain. The key is unlocking the data and presenting it in a consumable fashion for rapid, predictive, and integrated insights. **Effective data utilization first requires curating, visualizing, and securing big data to realize its potential**. Whether the organizational goal is web application or IT infrastructure optimization; revenue assurance; traffic optimization; healthcare behavior analysis; or even protecting the warfighter, big data can positively steer an organization's decisions and actions.

Data preparation

Organizations should first assess their data preparation services to turn disorganized and unprocessed bits and pieces of data into structured dynamic content to more quickly influence actionable behavior.

Figure 1. Data preparation helps turn raw data into meaningful data.

Data preparation organizes unstructured, raw information for later processing and deeper analysis. Refining the data preparation process is especially important given that big data talent has not kept pace with market demand, resulting in a widening gap between organizations and the individuals who can unlock valuable business information -- typically hard-to-find data scientists[1]. Hadoop and SQL database developers help businesses manage their data, but non-coders need tools that provide accessible paths to new insights. **By better structuring how data is collected and prepared, organizations will more quickly connect their people to data that matters.**

[1] http://techcrunch.com/2015/12/31/how-to-stem-the-global-shortage-of-data-scientists/

Data Visualization

Improving big data visualization is a better, more innovative and cheaper strategy to express the value that's been curated from collected data. Improving big data preparation and visualization would enable queries (without the need for a SQL specialist or enterprise database developer) to turn potentially overlooked data points into dynamic content and imagery.

Removing technical barriers to interact with an organization's databases in a more visual manner can bring to the forefront revelations or patterns that enable better decision-making. Emphasizing more visually focused data-driven content will help make once elusive data points more recognizable to more personnel. The power of data visualization is often directly related to their simplicity and many visualization tools already exist. Examples include the topological subway map, word clouds to highlight a document's prominent topics or themes, scatter plots that depict correlation, a Sankey diagram to see if one data set flows into another, or a circular tree diagram to show major groupings, sub-groupings, and their interconnections.

As businesses continue to rely on data collection as a means to business value, data-driven visuals will help an organization recognize trends and opportunities to seize market advantage.

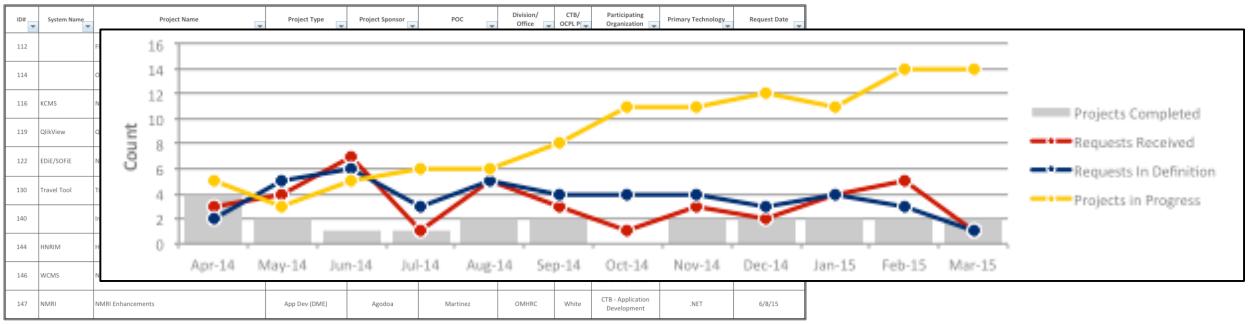

Figure 2. An example line and bar graph to visualize the underlying textual data.

Data Security

Many organizations still rely on legacy systems with incompatible formats and outdated standards vulnerable to intrusion. Worse yet, these stove-piped systems create stand-alone sets of individual data points that can appear meaningless when not aggregated and analyzed against all available information.

This lack of synchronization in both analytics and secure systems architecture results in diminished situational understanding and less confidence in strategic decision-making, hindering long-term success. Legacy systems are not only vulnerable from a security standpoint, but also prevent the integration of new capabilities that produce more sophisticated analytics. With the evolving cyber-threat landscape, data's transition to the cloud, and increasing government compliance objectives, organizations should implement scalable security architecture to store and access their data with confidence.

The Take-Away

Data preparation software eases the burden of sourcing, shaping, and cleansing messy data sets to accelerate its usefulness for analytics and decision-making. Emphasizing visualization of data will empower users to more easily spot, process, and develop insights into their meaning in a more efficient and effective manner. And securing data with redundant prevention, encryption, and detection services, as well as clear governance models (particularly as more organizations outsource their data to the cloud), will protect big data throughout its lifecycle.

The OnPoint Approach

OnPoint understands the enormity of this challenge. Faced with significant data volume, search limitations, and growing storage costs of legacy systems, the Department of Energy needed a new way to improve performance, deliver results and manage risk. Faced with data sets so large and complex that it was difficult to process using traditional data management approaches, OnPoint developed a Hadoop-based solution that reduced expenses by $1.5M per year, while providing substantial security intelligence and analytics capabilities with an easy to expand and cost-effective architecture. OnPoint turned data overload into actionable information and strategic advantage, transforming how a government agency manages and interprets their data.

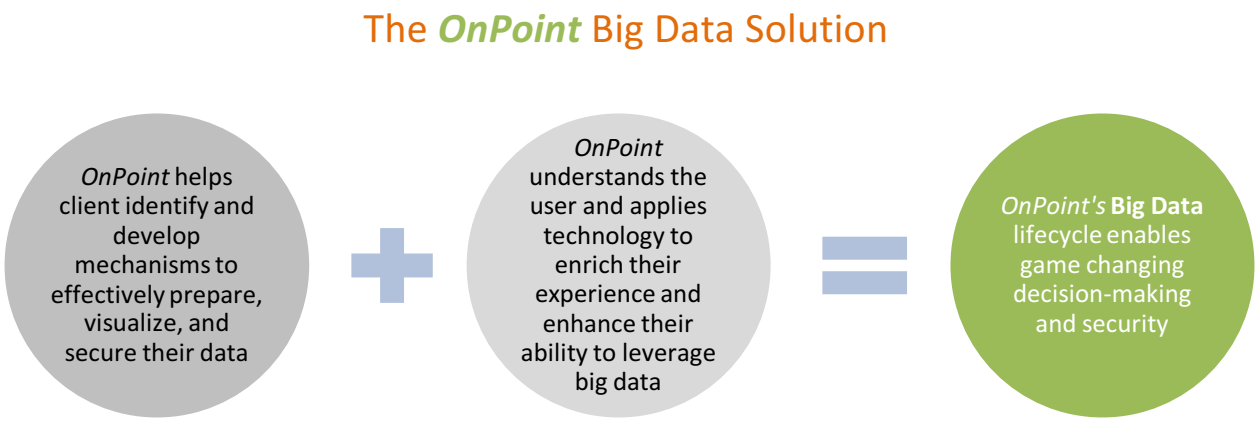

Figure 3. OnPoint Big Data Process

About OnPoint Consulting

OnPoint has proven experience in big data systems analysis, service oriented architectures, enterprise integration, and data visualization. OnPoint's team of information architecture, user experience, and design professional ensures solutions consume and present the data in a manner that enables users to tackle the most pressing and complex problems their organization faces. As the National Security Arm of Sapient (Adobe's 2015 Digital Marketing Partner of the Year), OnPoint has access to innovative commercial capabilities to support the government's mission in a time of digital transformation. As organizations face increasing pressure to harness big data, OnPoint's experience and processes are well positioned to help leverage solutions to derive exacting insight and secure a competitive advantage.