Whitepaper

Position Your Business for Big Data Success:

5 Steps to Create Sustained Value to the Bottom Line

NVP

NewVantage Partners

Understanding the Problem

Businesses are facing an environment where more and more data is collected or available, and Big Data is offering an opportunity to exploit this information in a timely and effective way. Some organizations are leveraging the technology to address new data sources, but most are still struggling to manage the data they already know well – or should know well – and find themselves dealing with new as well as current data when they look to leverage Big Data capabilities.

Data management practices often vary within large organizations. They have issues understanding what data they have, where it came from, how it got there and clarity of ownership and oversight, basically traditional data management issues. As a result, some organizations are looking to Big Data to solve their data management woes, seeing the technology as a way to <u>not</u> document data, <u>not</u> understand sources or strive to establish a common set of definitions for shared data.

Unmanaged, Big Data is like a fully loaded freight train. It's rich with opportunity when it arrives, cargo ready to be unloaded, organized and delivered to a business that knows how to use it profitably. But, out of control, it's a potentially devastating runaway that can create chaos for a business.

The truth is that organizations viewing Big Data as a panacea for all of their data challenges are doomed to make their data ecosystem more complicated, expensive and slow to react to business needs. That view almost guarantees that the value business can achieve with Big Data will be fleeting – if ever realized – as other data problems emerge and impact business activities.

The reason is simple. Infusing vast quantities of unmanaged, unclear data into your ecosystem, while okay for a fast track analysis or specific project, will become extremely complicated to manage as more initiatives and insights are desired. Imagine just five projects or programs duplicating just five terabytes of data every night. Even if it's doable, why take that step, given the impact on the infrastructure and even end-users?

It is critical that organizations seeking to leverage the large scale capabilities of Big Data and its support

for unstructured data must develop reference architectures over time and infuse formal processes, data management and oversight into their ecosystem. Otherwise, they will quickly find themselves worse off than when they started.

No Short Cuts

For example, at one large investment firm, a senior executive was convinced that Big Data was going to save him from spending more time or energy on Master Data solutions and programs to drive better governance and data management. He believed that the ability to quickly aggregate data, analyze it and derive insights would be more than enough. Big Data would figure out what data they had and what was best for them to use. With more education and a proof of concept under his belt, he realized this was not the case. In fact, he had neglected to think through the operational side of the equation and hadn't fully considered how critical it was to understand the context of the data being used.

True, Big Data can accelerate your data aggregation and time to answer. In fact, these are key benefits, but you need to understand what you're aggregating and find clarity in the data to have confidence that it is accurate and well positioned to support your insights and decisions.

Mature Data Management Enables a Faster, Repeatable Path to Realize Business Benefits

Another firm with more mature data management capabilities took a different approach. They allowed a sandbox environment to be created, but they were careful to track their data sources, and map business rules and how they developed their solution. They still iterated, they still followed a lot of dead-end hypotheses as they analyzed the data, but when they locked down the real value, they had a simple way to capture what was done. Now, even as they see a simple path to operationalize the work, they can turn it over to their development and infrastructure teams and look at their data management practices to find ways to fast-track this handoff in the future. This ensures that as the solution goes into production, the analysts stay apprised of any changes required to make it happen. They're also aware that this analysis includes a lot of unstructured data that was not

well documented before, and that they still need to work on improving their capabilities in this space.

Recognizing the risks

Of course, there are many organizations that understand the need to introduce Big Data and develop improved data management capabilities.   One chief enterprise information architect at a large insurer even voiced his concern that if all of these Big Data initiatives took off, current limited data management disciplines being applied, they would actually be worse off than they are now: data everywhere, not well understood, and always driven by one-off projects. He saw that it was critical to incrementally shore up data management capabilities or suffer the consequences.

In fact, the risks are pretty clear:

- Getting the answers you want vs. the ones you need – the right ones
- False sense of security and a poor understanding of your data
- Complex, expensive ecosystem
- Application of the wrong technology for the opportunity
- Data decisions continue to be made each and every time a project comes
- Limited or no ability to start effectively managing and exploiting unstructured information

Successful firms are taking a two-pronged approach to Big Data. They are (1) actively learning when and how to leverage the technology, and (2) working to enrich and enhance their data management, capabilities (metadata, governance, content management, reference models and active data management functions). For example, one large health insurance firm NewVantage Partners works with has actively defined a new information reference architecture and is beginning work to implement the changes as a separate, yet parallel effort to their other data initiatives, including Big Data. This seems to be an emerging trend in many insurance firms as they kick off similar efforts – perhaps a consequence of their tendency to manage risk as a part of normal business operations.

Positioning Yourself for Success

Moving forward, firms will need to actively enhance their data management capabilities. This doesn't mean they should enact onerous processes and oversight mechanisms that slow things down.   Smart organizations work to develop nimble processes and capabilities that align with their data activities' lifecycle: lighter at initiation and tighter, more robust as operational needs, quality or precision requirements increase.

NewVantage Partners recommends businesses focus on five key areas:

1) Establish clear roles and responsibilities
Just like all other data in your environment, you need a clear sense of who the key stakeholders/decision makers are. Roles may change as the data moves through your ecosystem and during its lifecycle, but they should be well understood nonetheless. As you embark on your Big Data initiatives, identify these stakeholders as soon as you can, and be prepared to refine and iterate as you go.

2) Strengthen your data governance & data management capabilities
- Ensure your processes are enhanced to support the needs of Big Data users and Big Data technology. Processes can be flexible and should take into consideration context, discovery vs. operational vs. transactional needs, with varying degrees of rigor and oversight to match.

- Ensure your reference information architecture is updated to include Big Data. This will inform future projects of how to best leverage Big Data technology and incorporate appropriate information management capabilities.

- Ensure your metadata management capabilities are enhanced to include/correlate all of the basic metadata components and over time, taxonomies, ontologies and business rules.

- Once you promote your solutions to production you will want them consistently used, so it's critical to have well-defined and functioning oversight and architecture functions. Ensure your governance processes include roles for IT controls to aid business stakeholders, and guide projects

NVP
NewVantage Partners

An Evolving Landscape

The "NEW" Reference Information Architecture
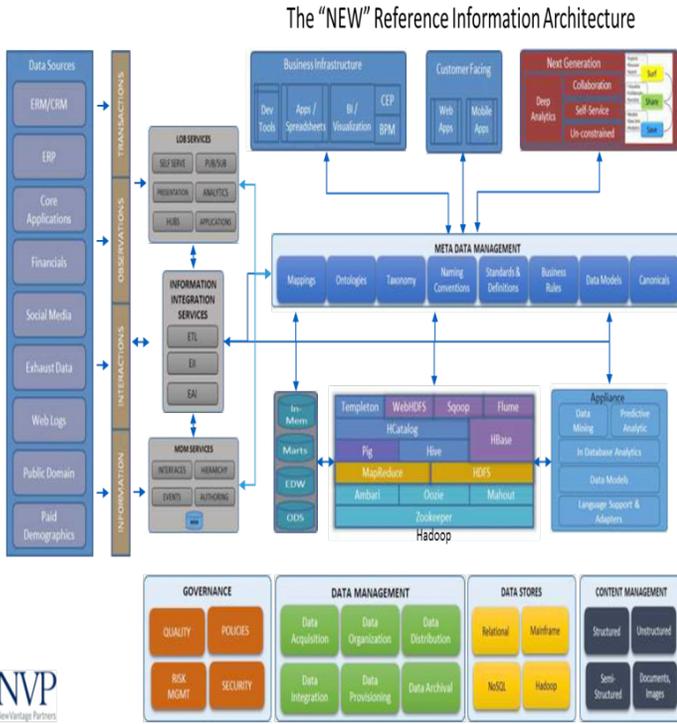
Copyright: NewVantage Partners 2013

to best use data. They should also ensure your security and legal teams. It's our experience that lightweight processes leveraging existing oversight/stage gate mechanisms work best, so long as they are applied and focused on quickly moving programs through the process and not serving as walls that prevent progress.

3) Clearly understand the purpose and required level of precision for data in the environment and adjust your expectations and processes accordingly

Whether it is a POC or a project that has moved into your mainstream business processes, be very clear on what you intend to do with the data, and what level of quality and precision you desire. This will enable the project to seek the right data sources and stakeholders to better assess value and impact, letting you decide how best to manage this data. Higher quality and precision requires stronger data management and oversight.

As your organization matures with Big Data, consider establishing a quality or precision classification approach that will allow the data users to understand what they're using and adjust their expectations accordingly. For example, you could use a white, blue gold designation to indicate raw data and cleansed data, organized and validated to support targeted analysis and usage. Some groups refine this even further into categories from 1-5 where 1 is raw data and 5 is well-understood, cleansed and organized data.

4) Incorporate unstructured content management into your DM capabilities

Unstructured data has always been a part of business operations, but now that we have better technologies to explore, analyze and infuse this content into business processes and insight work, it's critical that we finally take the step to formally include its management. Most firms struggle with this step.

Basic unstructured data in the form of comments or freeform entries is at least part of a database and should already be under some data management. But the ability to mine this information has been difficult.

Digital data outside of traditionally structured databases and business process flows is often not in scope for many governance groups and data management implementation, except when it may be seen as a technology issue. Generally, other than compliance with retention and security policies, it's not really managed today. As you move forward with Big Data implementations, you will find that this data quickly comes into scope with output impacting your BI solutions or even your operational activities. Consider actively working to include this data in the scope of your data management functions, adding clear business ownership, documenting what data information is available, usage, gold sources and more.

Don't take the 'easy route' and simply rely on Big Data technology to be your only formal data management process for unstructured data. Over time, as more and more unstructured data is added, being clear on what is good, what is bad, where it came from, and whether it is being used consistently becomes increasingly important; even its lifecycle stage is critical to leveraging this data.

NVP
NewVantage Partners

To maintain this clarity, you can use Big Data and other tools to understand what you have and identify what is valuable, and what needs to be managed. This is critical. Most unstructured data entering your Big Data Ecosystem already has some oversight, but usually as a blob and in its unstructured form. As you 'mine' this data in your business processes, it becomes more precise and valuable. It may also take on additional characteristics to comply with security, privacy or legal and regulatory elements. Ultimately, these data nuggets can become new data elements or added to existing data, BUT you must have the metadata describing it and the ability to manage it for the most effective use possible.
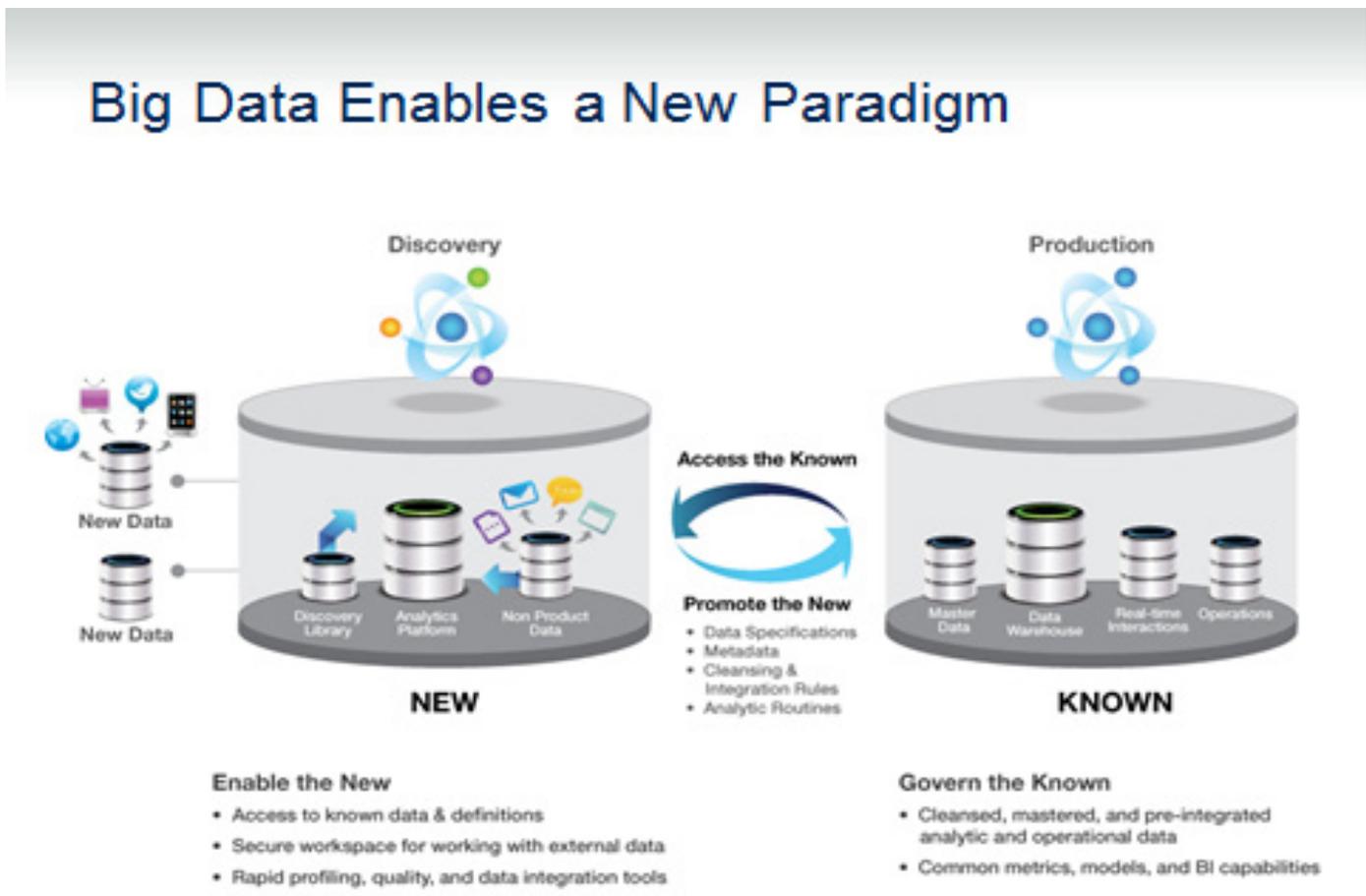
5) Formalize the hand-off from discovery and pilots to production environments
If you're doing one-time analysis or a full-fledged throw-away pilot this doesn't apply to you. But for most organizations, the initial Big Data work quickly evolves and

you find a need to leverage the highly valuable information you've mined in a sustainable way. This means taking the work done in your sandbox environments and formalizing it to run in your production environment.

Data coming from Big Data Analytics is often unfamiliar to IT organizations as you get started, and they may seek to duplicate what was done in the context of all other steps they take. Some of this activity is required to create a repeatable, production quality solution. But actions like trying to recast the database as a relational model, or assuming what they received must be rethought and remodeled – almost starting from scratch and using the work from the sandbox as just a data point in requirements – are just a waste of time and money.

Most organizations find the answer lies between taking the sandbox-promoted solution at face value and completely redoing the work. Remember, this was a data savvy

Big Data Enables a New Paradigm

Discovery

Production

New Data

New Data

Discovery Library | Analytics Platform | Non Product Data

NEW

Access the Known

Promote the New
- Data Specifications
- Metadata
- Cleansing & Integration Rules
- Analytic Routines

Master Data | Data Warehouse | Real-time Interactions | Operations

KNOWN

Enable the New
- Access to known data & definitions
- Secure workspace for working with external data
- Rapid profiling, quality, and data integration tools

Govern the Known
- Cleansed, mastered, and pre-integrated analytic and operational data
- Common metrics, models, and BI capabilities

NVP
NewVantage Partners

exercise; done well, the sandbox solution will likely contain a lot of critical, available information. It may need to be better coded, sources could be needed to better reflect as built source systems, or you may need to rework it to fit into your production Big Data environment. But don't start from scratch.

For instance, one firm spent months in their Big Data sandbox environment testing their hypothesis and leveraging Big Data technology to ultimately identify and develop a number of high impact solutions, providing real insights into customer activity that they previously could not access. Even in its temporary 'sandbox state,' this analytics work was perceived as having a huge impact on the business. With a very deep understanding of the business data, these business analysts had even created a data model and had documented the ETL they created with key assumptions. But, this was not a production-ready solution. It wasn't meant to be. It was meant to prove the value of their hypothesis, with the expectation that it would be promoted to production when the business value was proven. They knew their ETL code wasn't bullet proof and they suspected that with other development work about to go into production that some of their sources would need to change. But they had the key elements in place to fast-track the move to production. Interestingly, the IT team started the way they always do. They wanted to go back to requirements-gathering and a manual source to target-mapping of the ETL and resulting datasets in the Big Data environment. They kicked off a new requirements process and also assumed the data model was unusable. The truth was while they were not perfect, these were not Visio diagrams or Excel workbooks. They were solutions leveraging real databases and ETL tools. IT could fast-track most of this process by leveraging all of the information in the system already. With a little prompting, they concluded that they could use database models as a starting point and enhance to meet production standards. The same was true for the ETLs. So in the end, they did jump-start the process

To succeed, it's important to formalize this process for a smoother hand-off. IT and analysts in the sandbox must establish some common rules of the road, including agreeing on naming conventions, tools to use and even how to share key learnings. This is not to say the sandbox needs the entire rigor of a formal IT development

effort; that would contradict business wants from this technology. The goal is to find a balance that allows for fast iteration and discovery as well as the ability to migrate to production in a way that meets quality and production standards.

For analysts, this means not only sharing their analysis, but also sharing new approaches, new capabilities and perhaps tools with IT so they don't have to rediscover these answers.

In addition, analytics teams leveraging Big Data technology with ongoing sandbox environments will need to know the results when their solution is moved to a production/operational mode. They'll need to know if better sources are found and if any precision/quality improvements are made. For IT, this means providing feedback and lessons learned from their migration work, including sharing information about better/gold sources and ways to improve ETL for hand-off or even ways to use Big Data technology more effectively. This feedback loop will enable analysts in the sandbox to produce better results and avoid making mistakes that IT will have to address during the migration to production.

Basically, it's important to understand that work in the sandbox doesn't stop. Data analysts and scientists will continue to apply hypotheses and iterate from existing and new ideas to find data of value to the business. The team handing off Big Data solutions likely has used more automation and tools to prove business value and establish its need for a production solution than may be typical in your organization. Teams that assume they need to throw everything away and start over will find a business user base that is uncooperative. Think of the work done by your Big Data analysts in their sandboxes as an iteration in an agile development process. This will enable you to deliver more quickly and, in some cases, even automate the move to production.

Bottom-line

Over the years, NewVantage Partners has seen a few very disruptive technologies appear that drive near term benefits and change the way we approach things: the introduction of mid-range devices, PCs, mobile devices,

NVP
NewVantage Partners

and SaaS, to name a few. Organizations began adopting these technologies to help them speed up business processes and decision-making or improve productivity. They were all brought in as one-off solutions, eventually gaining such popularity and critical mass that their impacts demanded additional structure and rigor to ensure costs could be managed and operational integrity not compromised. One potential pitfall of this approach is often the very real impression that when IT gets involved and adds a production-level perspective to the technology, things slow down. Big Data changes that.

Like disruptive technologies in the past, Big Data offers great business opportunity. It also offers us a way to change our approach and create sustainable, rapid business value while incorporating these capabilities into our production environment and processes. In fact, because of the vast amount of information you can leverage in short periods of time, not taking advantage of Big Data exposes you to the risk of making really bad business decisions really quickly based on poor quality data. You also risk incurring higher costs with suspect value because of diverging solutions as multiple areas look at different approaches and gathering data inconsistently from external and internal sources.

Moving ahead with Big Data initiatives for long term success requires that businesses:

- **Establish good data management practices:** This includes data governance and incorporating Big Data into your processes, scope and capabilities.

- **Formalize promotion of Big Data solutions into production:** Develop processes to promote test and learn/sandbox Big Data solutions into production that build on the work from the sandbox, not completely replace it.

- **Learn from one another**: Facilitate business and IT learning about new Big Data capabilities and collaborate on what and when to introduce these capabilities. Big Data is still maturing and establishing inflexible processes or technology standards will stifle innovation, so make a conscious decision now to avoid those restrictions.

- **Understand your data**: Establish a data feedback loop to better understand data in production. Leveraging

your production work and data management capabilities, you will continually refine your understanding of existing data as well as the new data you are gathering and creating while using this technology. It's critical to share this understanding with key stakeholders. Transparency in data understanding will speed up analysis and realize business benefits faster.

- **Re-assess and improve**: This is a maturing technology space and so are your capabilities and needs. Regularly reassess your processes, technology, platform and architecture to ensure your success in this evolving landscape.

This should not be a heavy organizational burden, and it shouldn't slow down your discovery process or your ability to quickly deliver value. Done correctly, you will enable future projects to proceed more rapidly, support fast-track time to answer needs, gain confidence in your decisions, and be well-positioned to convert the coming onslaught of even more internal and external data into bottom line value for your business.

About NewVantage Partners

NewVantage Partners is a boutique management consulting practice established in 2001 and comprised of former C-Level business and technology executives, and senior subject experts.

Our work comprises up-front planning – current state assessment, future state design, business case, execution roadmap, as well as the development of business and technical requirements, business capabilities, and business architecture. We are frequently engaged to provide a critical link between the business and technology organizations of our clients.

NewVantage fosters a commitment to executive thought-leadership through a series of small group executive dinners, and through our executive advisory board comprised of current and former Fortune 1000 business and technology executives and well-known industry thought leaders.

For more information, contact Randy Bean at rbean@newvantage.com.

NVP
NewVantage Partners