HORTONWORKS BIG DATA MATURITY MODEL

The strategic path to accelerating business transformations

Contents

Contact

USA
5470 Great America Parkway
Santa Clara, CA 95054
(855) 846-7866

UK
30 New Broad Street
London, EC2M 1HT
+44 (0)207 194 7710

On-line assessment available at:
http://hortonworks.com/big-data-scorecard/

About the author
Vishal Dhanuka is Managing Director and Head of Customer Innovation and Strategy at Hortonworks, and is responsible for developing the Hortonworks value-based go-to-market strategy across industries, along with helping strategic customers refine their future vision and define the business value through Big Data driven business transformation.
Email: vdhanuka@hortonworks.com

Executive Summary

Data is the new kingmaker. The explosion of data—in volume, variety and velocity—is catalyzing new business models and reshaping industries. Business leaders can no longer amble their way forward in the age of Big Data; the challenges are too great to address without a multi-year strategic roadmap, and the business potential too vast to simply dismiss.

Large volumes of data may have little or no value as individual or small groups of records, but when unified in an enterprise Data Lake, these same massive volumes can provide a global and historical perspective which may then be inspected for patterns and ultimately used to create advanced analytic applications that drive new business models.

Every industry has massive opportunities in this space. McKinsey estimates that healthcare alone stands to reap a $300 billion annual savings from Big Data innovation. The projected annual value to Europe's public sector administration is estimated at $275 billion. Retailers see up to a 60 percent increase in operating margins through Big Data programs.

But the journey to business transformation is complex. It begins with understanding the opportunities unique to your business, and understanding how the maturity of your organization enables or inhibits your ability to strategically pursue Big Data programs aligned to your business goals.

Our maturity model reflects Hortonworks' consulting experience with hundreds of companies, each entering the Big Data space with different capabilities and objectives. Hortonworks (NASDAQ: HDP) was founded in 2011 by 24 engineers from the original Yahoo! Apache Hadoop development and operations teams. At Hortonworks, we have helped more than 500 of the world's leading companies transform their businesses with Open Enterprise Apache Hadoop. We have more Big Data experience than any other company.

The unique value in our approach to assessing Big Data maturity is that we help business leaders understand and evaluate the essential focus areas for accelerating business transformations— across four stages of maturity. We couple the maturity assessment with an opportunity to plot a strategic path forward. At the top of our maturity model are businesses that are transforming through Big Data. These leaders embrace data as currency, and are creating new value streams through predictive analytics.

The second tier of our model focuses on companies that are optimizing operational performance across three dimensions: the customer lifecycle, product lifecycle and facility lifecycle. At the lower rungs of maturity, we see companies exploring or merely becoming aware of their Big Data opportunity for the first time. The exploring companies are seeing consensus emerge on the potential of Big Data and beginning to fund limited Apache Hadoop experiments with localized intent. Many newcomers to the Big Data space are simply aware of the Big Data opportunity, but have not taken meaningful steps to integrate data into business strategies or processes beyond historical analysis.

BIG DATA INSIGHTS

NEW BIG DATA TO SAVE HEALTHCARE INDUSTRY
$300B Annually

Source: McKinsey Global Institute (MGI), Big Data: The Next Frontier for Innovation, Competition and Productivity, June 2011

MACHINE-GENERATED DATA PROJECTED INCREASE BY 2020
15x

Source: IDC, IDC Maturity Landscape: Big Data and Analytics—A Guide to Unlocking Information Assets, 201Competition and Productivity, June 2011

ORGANIZATIONS PLANNING TO IMPLEMENT APACHE HADOOP BY 2017
44%

Source: Gartner Research Circle Survey, Q1, 2015

PERCENT OF COMPANIES INVESTING OVER 20% OF TECH BUDGET IN BIG DATA
73%

Source: Accenture Big Success with Big Data Survey, April 2014

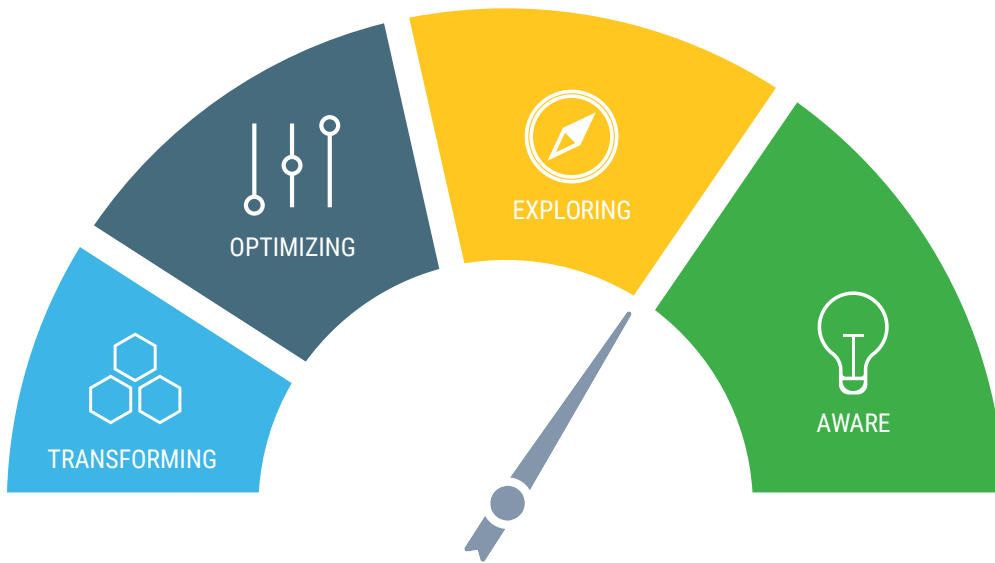## THE STEPS TO ACCELERATING BUSINESS TRANSFORMATIONS

Figure 1. Hortonworks Big Data Maturity Model

The Hortonworks Big Data Maturity Model assesses your organization's Big Data capabilities across five domains, with four focus areas inside each maturity level:

1. Data Sponsorship: Vision & Strategy, Funding, Advocacy, Business Case

2. Data and Analytics Practices: Data collection, Storage, Processing, Analysis

3. Technology and Infrastructure: Hosting, Functionality, Tools, Integrations

4. Organization and Skills: Analytic & Development Skills, Talent Strategy, Leadership and Collaboration Models

5. Process Management: Planning & Budget, Operations, Security & Governance, Program Measurement, Investment Focus

Your company's proficiency level within each of these five capability domains indicates your overall Big Data maturity. Using this information, you can compare your company to others in your industry, as well as plan roadmaps and investments for advancing the objectives and opportunities unique to your business.

Hortonworks helps business leaders understand and evaluate the essential focus areas for accelerating business transformations across four stages of maturity: Aware, Exploring, Optimizing and Transforming.

Sizing Your Business Opportunity

Business data is growing at a staggering rate, with 40ZB of data anticipated by 2020 (1 zettabyte is equal to 1 billon terabytes). Approximately 85 percent of this growth is coming from new data sources like web traffic, social media, sensors and geo-location devices. In fact, according to IDC, machine-generated data alone is projected to increase 15x by 2020.

Much of this incoming data has little or no structure, or structure that changes too frequently for reliable schema creation at the time of ingest. Since legacy IT infrastructures are incapable of handling unstructured data, many companies just decide to not capture and store it—thereby sacrificing potentially valuable business insights in the process.

Each industry has unique opportunities in Big Data, but companies tend to capture value from three common patterns of usage: discovering new patterns among previously unassociated datasets ; creating a single view of key business entities such as customers and machines; and developing predictive statistical models that enable key operational advantages.

IDC estimates that the Big Data market will experience a growth rate six times higher than the rest of the global IT market.

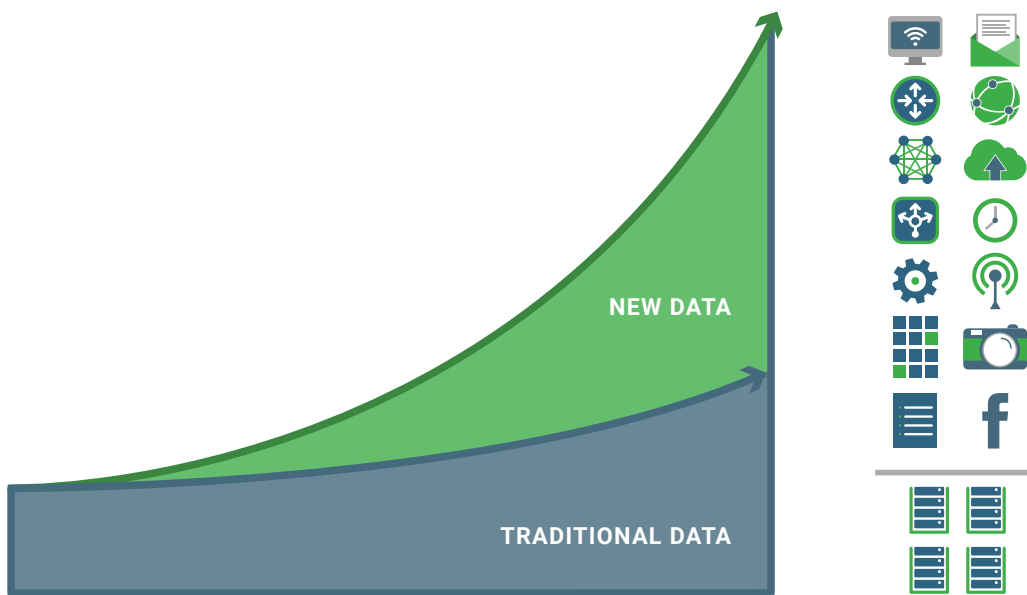*Source: IDC, "Using Big Data + Analytics to Drive Business Transformation"*

Figure 2. Numerous new data types are driving an explosion in the volume, variety and velocity of data; creating net-new value for companies that know how to capture, store and analyze it
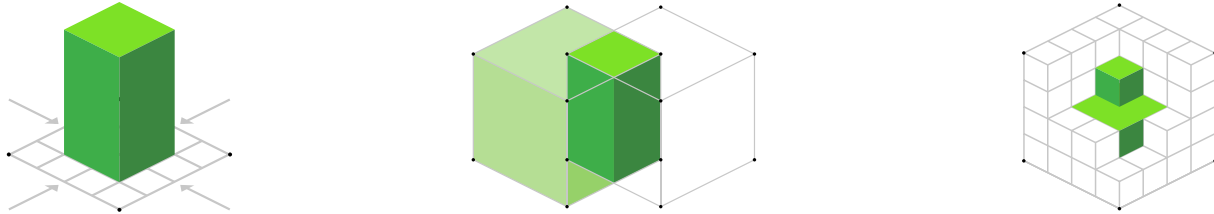
Data Security

Figure 3. Three common patterns of big data usage

DATA DISCOVERY

Business analysts use HDP to explore new data types and large data sets that were previously too big to capture, store and process. They unlock insights from data such as clickstream, geo-location, sensor, server log, social, text and video data.

This new data enriches existing analytic applications with new perspectives and insights. New data also drives the creation of innovative new apps that provide new value to the business.

SINGLE VIEW OF ENTITY

Today customers generate more data than ever before as they research, purchase and use a company's products or services. By bringing together all information about a customer including new data types such as clickstream, geo-location and social, organizations can better understand, anticipate and meet customer needs.

HDP powers a single view of each customer, allowing organizations to provide targeted, personalized customer experiences. This single view allows a company to reduce attrition, improve cross-sell and bolster its word-of-mouth reputation.

PREDICTIVE ANALYTICS

Past is prelude. Historical data provides signals that indicate what may happen in the future. By understanding signals coming from machines and sensors, server logs and other new data sources, organizations can predict future events and become more proactive.

Apache™ Hadoop® captures, stores and processes the large volumes of data streaming from connected devices and sensors that measure your business. A variety of data science and iterative machine-learning techniques can make confident real-time recommendations that reduce costs, improve safety, and inform investments.

These three common patterns of applications of advanced analytics across industries have produced several recurring business strategies. In his November 2014 Harvard Business Review study, John Hagel refers to "enterprise process economics" in order to describe how Big Data optimizes operational performance in three core dimensions: customer lifecycle, product lifecycle and facility lifecycle. Some of the common strategies we see focus on enhancing the customer experience, enabling cross-sell and up-sell opportunities, driving product targeting, scheduling preventative maintenance and optimizing resource and supply chains.

Maintaining a single view of customers enables new customer experience strategies, like coordinating customer-facing activities across departments. This requires consolidation of existing and new sources of data—often unstructured data—and simultaneous processing and delivery of data across multiple departments using different processing and analytics engines. This capability has enabled retailers, for example, to deliver a consistent experience in store, online and in marketing offers based on a customer's preferences and prior purchases. In telecom, correlating handset performance data and online account activity enables customer service to reach out proactively to solve customer problems, thereby decreasing churn.

This single-view capability can also support cross-sell and up-sell strategies. Here, data siloed by business units and departments is combined with new sources of poorly structured data and then run through multiple analytical and predictive techniques to identify new sales opportunities. For example, a bank can identify customers likely to become mortgage customers based on the number of other bank products they own and recent account activity. Likewise, online retailers can recommend other products based on a customer's previous purchases, browsing history and similarity to other visitor types.

Building a complete view across customers also enables product targeting to address large opportunities, shifting customer tastes and emerging needs—both at the aggregate and individual level. These targeting strategies combine customer and market data with new data sources, and then apply multiple advanced analytical techniques. For example, auto manufacturers combine customer purchase histories with in-car sensor data to create offerings many customers are likely to buy. Similarly, insurance companies look at lifetime buying patterns and customer claims to create bundled product offerings and price them effectively.

Preventative maintenance strategies typically involve analyzing sensor data to predict when a piece of equipment will fail, degrade in performance or act in a way requiring corrective action. For example, mobile providers may analyze sensor data in a cell tower to predict failure and proactively repair the equipment, thus avoiding costly replacement. In the airline industry, companies can analyze sensor data on planes to predict equipment failure and make repairs before schedules are impacted, and to ensure the safety of passengers and crew. And food service companies can analyze refrigeration units to repair and avoid failures that might lead to significant food spoilage and loss.

Analyzing data to more effectively predict resource requirements is a key strategy in resource and supply chain optimization. These strategies often require correlating multiple data types and combining multiple processing techniques. For example, utility companies will monitor shifting electricity usage patterns to optimize use of generating assets. Building management companies can analyze building data—people movement, elevators, lights and systems—to optimize energy use. And sales and marketing organizations may analyze market demand to optimize how field resources are deployed.

Companies acting on journey insights have seen a 15-20% reduction in cost to serve, a 10-20% boost in crossselling, and a drop of 10-25 basis points in churn.

Source: The Economist, "Do You Have the Right Stuff? Three Steps for Turning Big Data Chaos Into Customer Experience Gold ", Jan 2014

The intensity of Big Data varies across sectors but has reached critical mass in every sector.

Source: McKinsey Global Institute (MGI), Big Data: The Next Frontier for Innovation, Competition and Productivity, June 2011

Predictive analytics has never been more relevant, and easier, than it is now. Big data, gobs of compute power, and modern tools are making predictive models more efficient, accurate, and accessible to enterprises.
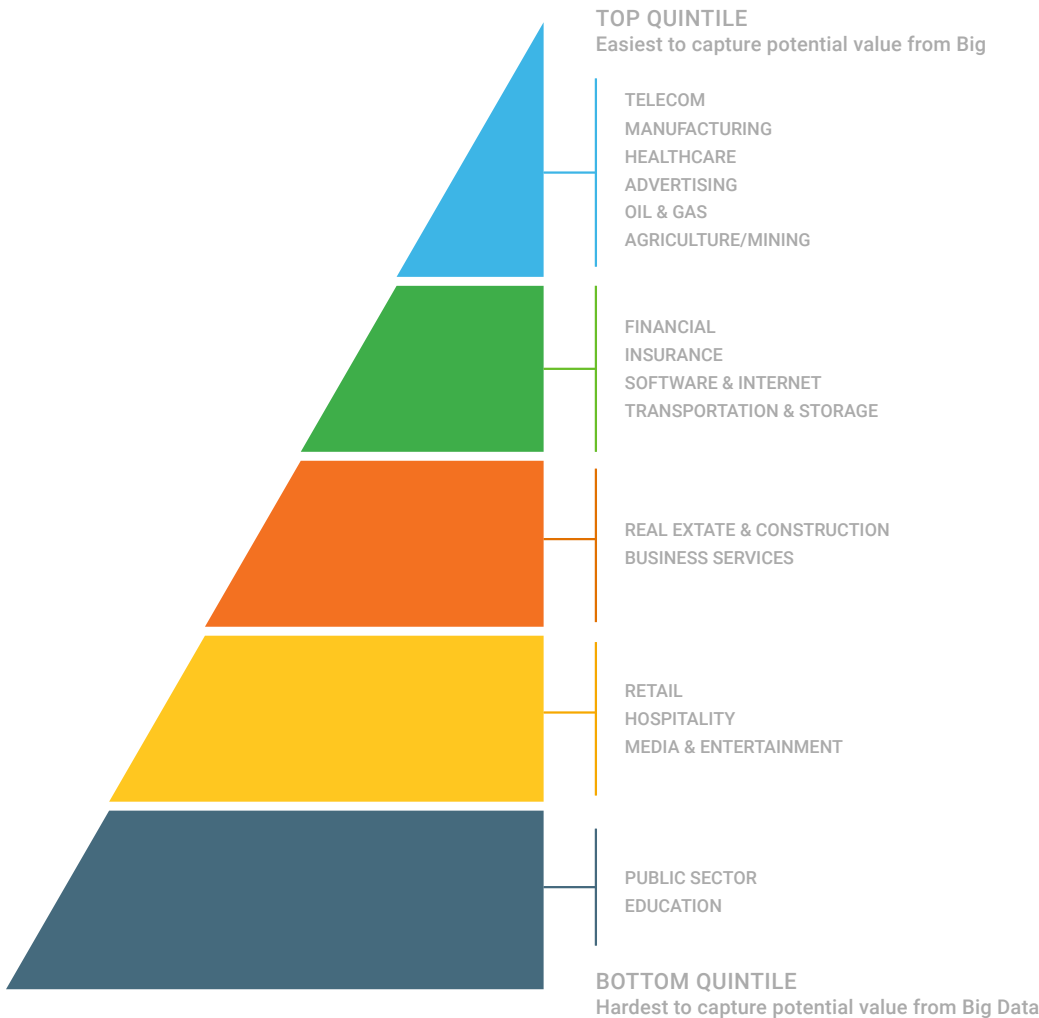
Forrester: The Forrester Wave™: Big Data Predictive Analytics Solutions, Q2 2015

Predictive analytics can provide important insight to shape a transaction before it occurs. This strategy allows a business to predict likely outcomes based on various "what if" scenarios, enabling it to better influence the outcome. Timing is critical to effectively executing this strategy, often requiring the synthesis of long-term patterns in disparate data with real-time data capture and analysis. For example, retailers may analyze publicly available data and social media content to determine a customer's propensity to buy. Automotive companies analyze sensor and geo-location data from connected cars to offer products and services in context. And banks can now identify credit card fraud as it's occurring.

A recent McKinsey study found that some industries have an easier time extracting value from Big Data than others, depending on four critical factors: talent, IT intensity, data-driven mindset and data availability. Hortonworks validated the McKinsey assessment, and came to similar conclusions—putting talent at the top of key barriers (see Section 5). The Hortonworks maturity framework highlights four focus areas within five broad capability domains, which can help you identify specific investment areas that make the most sense for your business.

WHICH INDUSTRIES CAPTURE THE MOST VALUE FROM BIG DATA?

TOP QUINTILE
Easiest to capture potential value from Big

TELECOM
MANUFACTURING
HEALTHCARE
ADVERTISING
OIL & GAS
AGRICULTURE/MINING

FINANCIAL
INSURANCE
SOFTWARE & INTERNET
TRANSPORTATION & STORAGE

REAL EXTATE & CONSTRUCTION
BUSINESS SERVICES

RETAIL
HOSPITALITY
MEDIA & ENTERTAINMENT

PUBLIC SECTOR
EDUCATION

BOTTOM QUINTILE
Hardest to capture potential value from Big Data

Some industries have an easier time extracting value from Big Data than others, depending on four critical factors: talent, IT intensity, data-driven mindset and data availability.

Figure 4. Industries are discovering unique strategies for leveraging big data and facing different levels of difficulty capturing value. Source: Mckinsey Global Institute, "Big Data: The Next Frontier for Innovation, Competition and Productivity", June 2011

Assessing Your Big Data Maturity

Hortonworks' experience working with hundreds of the world's largest brands gives us a unique perspective on how to measure your Big Data maturity. Using that perspective, we can work with you to build a roadmap aligned to your company's business goals. To help advance your organization's Big Data journey, we begin by assessing your current state along five key capability domains: Sponsorship; Data and Analytics; Technology and Infrastructure; Organization and Skills; and Process Management. Within each of these capability domains, we identify four key focus areas that indicate maturity, and then assess each area according to maturity level (i.e. Aware, Exploring, Optimizing, Transforming). Although the purpose of the exercise is to evaluate your company's maturity level in these areas, we believe it's far more important to understand how to capitalize on your existing capabilities, and to invest in those focus areas in the way that best maximizes your progress toward defined business objectives.
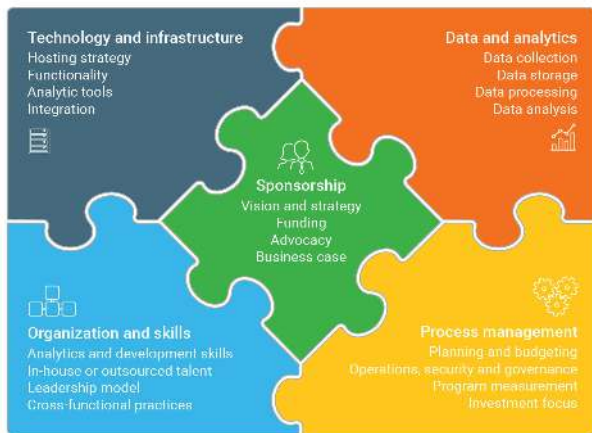
Companies need to understand where they stand in terms of big data maturity so that they can progress and identify the required initiatives

Figure 5. Capability domains and focus areas of Hortonworks' Big Data Maturity Model
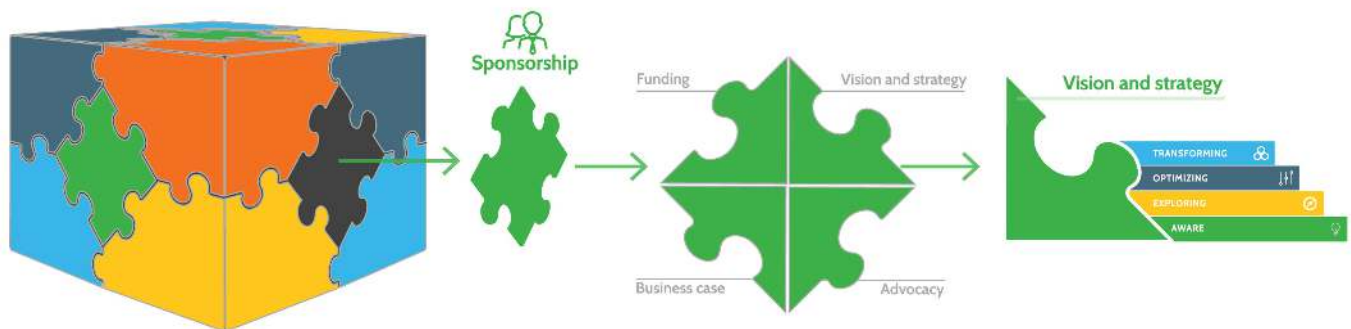
We look at five different capability domains holistically across the organization

...we then examine four key focus areas per capability domain

...finally, we measure the maturity stages for each focus area to understand the overall maturity of the organization across all capabilities

Figure 6. A prescriptive approach for assessing Big Data maturity based on four key focus areas per capability domain

CAPABILITY DOMAIN 1: SPONSORSHIP

Sponsorship refers broadly to four focus areas that determine the quality and organizational impact of your company's Big Data vision. These focus areas ask you to think about how strategically your vision is crafted and how deeply your organization has mobilized and validated the vision.

Vision and Strategy
The path to business transformation through Big Data is dependent upon a unified vision and sound strategy that emanate from the highest levels of the enterprise. At lower levels of maturity, a limited vision may exist solely among forward-thinking data architects focused on solving isolated business problems. Senior-level interest may emerge first among technology leadership, but the organization won't truly be on the road to transformation until the C-suite aligns on a vision for Big Data and creates an enterprise-wide roadmap to get there.

Funding
At lower stages of maturity, Big Data program funding is unbudgeted—because Big Data programs don't formally exist. Instead, IT budgets will typically fund one-off projects directed at specific business problems. More mature organizations fund Big Data programs at regular intervals, either by executive or business-level line items, while the most mature organizations fund centralized programs annually based on planning processes tied to the Big Data roadmap.

Advocacy
Many organizations underestimate the importance of changing culture as part of the technology and process evolution tied to Big Data transformation. In the early days of a program, individual champions inside IT will often lead the call to arms. But critical mass can't build until senior leadership across the organization consistently advocates for a culture of data-driven planning, testing and experimentation. One of the hallmarks of an Enterprise that is accelerating business transformations using data as an enabler is that new ideas for leveraging data emerge from the bottom-up as frequently as from the top-down.

Business Case
Early investments in Big Data capabilities are often justified on the basis of optimizing existing data warehouse investments. Those cases are certainly valid when you look at the 100x savings that come from offloading unused data from the EDW or expanding storage capacity to Apache Hadoop clusters. But the business case eventually evolves beyond data storage and processing efficiencies to reflect wins in projects with advanced analytics applications. Ultimately, the business case will reflect new revenue streams and business models directly attributable to Big Data insights.

The four focus areas within the sponsorship domain require you to think about how strategic your vision is crafted and how deeply your organization has mobilized and is validating the vision.

CAPABILITY DOMAIN 2: DATA AND ANALYTICS PRACTICES

The focus areas inside our data and analytics capability domain address the data pipeline—from data collection, to storage, to processing and analysis. This domain asks you consider key milestones such as the preservation and ingest of all company data, the creation of a shared data service—called a Data Lake—and the use of automation to enable speed and access to quality analytics.

The data and analytics practices domain addresses the data pipeline—from collection, to storage, to processing and analytics.

Data Collection
Data collection is at the front of the data pipeline—accordingly, how an organization targets and collects data is a primary indicator of its maturity. Low-maturity organizations typically invest a lot of manual effort in collecting primarily structured data required for specific analysis related to evaluating business performance. Newer, unstructured data sources are often ignored, and data that doesn't fit because of its size or format are routinely discarded. Mature organizations view data as currency, and never discard it. Increasing levels of automated collection streamlines ingest processes, and the most mature organizations actively seek out new data sources.

Data Storage
The limited storage capacities and prohibitive costs of legacy EDW infrastructures often lead less mature organizations to suffer from what Gartner calls "dark data"—data that is unused or lost because of a lack of ownership, visibility or access. Maturing organizations realize the potential future value of storing all data, even if that value cannot be determined in the near-term. A unified enterprise data architecture signals the beginning of the optimization phase and the pending arrival of the coveted enterprise Data Lake—a shared data service open to users across the business.

Data Processing
Most organizations are already routinely processing structured data, although manual processes can often make this a tedious and expensive endeavor to scale. As enterprise metadata standards emerge, organizations will more consistently organize their data for discovery and reuse. Aligning legacy, department-level metadata schemas to a unified enterprise architecture can be a challenge, but it's a necessary requirement of the enterprise Data Lake, and the foundation of advanced enterprise analytics. A fully operational data processing engine will work across multiple data types and sources, including real-time data.

Data Analysis
Lower maturity businesses are using data in an ad hoc manner, reporting on basic business Key Performance Indicators (KPIs) usually related to performance management. The arrival of new kinds of data eventually gives way to deeper analytic exercises based on new "what if" scenarios, but the time-to-value for such exercises is still lengthy. The analytics practice matures around predictive modeling and real-time analysis. The most mature analytics organizations provide consistently high quality, availability and value across multiple business areas using automated processes based on optimized business rules.

CAPABILITY DOMAIN 3: TECHNOLOGY AND INFRASTRUCTURE

Our technology and infrastructure focus areas look at the quality and scalability of your Apache Hadoop clusters, and the adoption of hybrid hosting scenarios that maximize data access across the organization. This domain also addresses your level of integration between analytic tools across the business, again targeting shared access to key capabilities.

Hosting Strategy

Concerns for security and management prevent many companies from exploring hybrid cloud/ on-premises hosting solutions. But as demand grows for enterprise-wide applications and data, organizations design their hosting environments to optimize availability and speed. Best-of-breed hybrid cloud solutions deliver high availability, security and dependability across unified public/ private cloud and on-premises architectures.

Functionality

An organization's ability to work with its data is a key indicator of maturity. The primary roadblock to Big Data maturity is a legacy EDW storage-only infrastructure that cannot support multiple workloads or emerging data types. A proof-of-concept Apache Hadoop cluster and/or NoSQL project often ignites the possibilities of moving beyond a single-workload scenario to batch, interactive and real-time data processing that serves advanced business analytics, custom applications or packaged applications driven by multi-structured, multisourced data.

Analytic tools

Tooling typically follows the same journey as the Data Lake. In the beginning, departmentlevel tooling supports local requirements for routine, canned and ad hoc reports on business performance. The investments in tooling are project-specific. As analytic tooling matures, multiple fit-for-purpose technologies will converge in an enterprise analytics platform that sits atop the Data Lake. The presence of enterprise analytics is the precursor to adoption. The companies that are transforming their business models have created a culture where centralized analytics resources are integral to every critical business decision.

Integration

The least mature organizations spend a disproportionate amount of time maintaining and tuning their technology and infrastructure to meet the ever-changing requirements of the business. As the enterprise Big Data strategy and roadmap emerge, integration begins to unify deployed technologies around a common architecture. The data itself also becomes more integrated over time. The most mature organizations have created a single view of key business entities, maintained and delivered to the business by a centralized data service or center of excellence.

The technology and infrastructure domain looks at the quality and scalability of your Apache Hadoop clusters and investigates your level of integration between analytic tools across the organization.
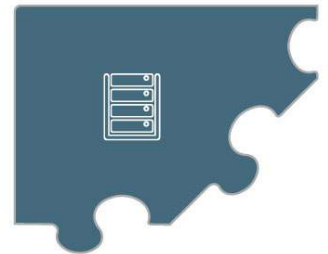
CAPABILITY DOMAIN 4: ORGANIZATION AND SKILLS

The keystone of modern Big Data strategy, your company's approach to identifying the right organizational structure, talent focus and staffing model will largely predict your ability to achieve transformational results. The focus areas in this domain address your company's approach to outsourcing, management, program leadership and cross-functional collaboration.

Analytics and Development Skills
Similar to the way a Big Data vision begins to form in organizations, internal skills tend to cluster initially at the department or line-of-business levels in technical roles. Lone data analysts may emerge with compelling insights that spark new conversations, and pockets of expertise in data storage optimization may start conversations about Big Data architecture like Apache Hadoop. Once the basic Apache Hadoop infrastructure is in place, the demand for skills shifts to data mining and data discovery, combined with data application. The most mature organizations have implemented formal enterprise-wide training programs and support systems to deliver a functional level of analytic expertise across business roles.

In-house or Outsourced Talent
The availability of talent is among the top barriers to increasing Big Data maturity. While many enterprises have in-house expertise in traditional EDW practices, and some have added specialists in collecting and storing new data types, fewer can find and hire the right people to strategically staff the right capabilities for the enterprise roadmap. Outsourcing becomes a key strategy for delivering successful proof of concepts, and even more mature organizations depend heavily on outsourced talent to augment central Big Data resources.

Leadership Model
The journey to Big Data transformation requires a centralized pool of resources: centralized data, centralized analytic tooling, centralized support and most importantly, centralized leadership. Because most Big Data programs begin with localized POCs, the creation of a central center of excellence can often draw from existing leadership. A cross-functional management structure will drive the cultural and procedural changes necessary to transform the way an enterprise thinks and behaves.

Cross-functional Practices
One of the most important ways of growing localized data successes into enterprise-wide capabilities is to encourage cross-functional collaboration. In the early stages of Big Data maturity, cross-functional collaboration will focus on optimizing data collection, standardizing metadata and creating access to the Data Lake. But as the organization matures, collaboration will expand beyond Apache Hadoop and EDW teams to include cross-functional business teams optimizing their own Big Data programs. A Big Data steering committee helps the largest organizations ensure alignment between the business and the center of excellence.

The organization and skills domain looks at your company's approach to outsourcing, management, program leadership and cross-functional collaboration.
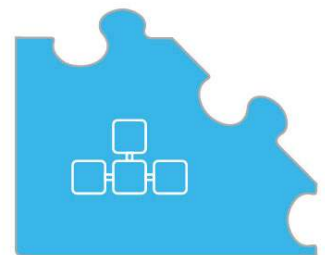
CAPABILITY DOMAIN 5: PROCESS MANAGEMENT

Our final capability domain focuses on how your company manages and evaluates your Big Data program. From program planning to data governance and security, this capability domain sets standards for the way your organization's investment in Big Data is protected and optimized over time.

Planning and Budget
At the lowest levels of maturity, planning and budgeting for Big Data programs happens in an ad hoc manner—i.e., as ideas come up, usually from technical staff dealing with a storage crisis or a forward-thinking business analyst with their eye on a proof of concept. Proof of concept success typically creates planning cycles at the department or business-unit level, with little or no cross-functional coordination. As POC successes take root, multiple Big Data initiatives will emerge from the bottom-up and top-down, until annual budget planning at the executive and departmental level makes funding a shared priority tied to the enterprise roadmap.

Operations, Security and Governance
Data security tends to be one of the first capabilities an enterprise invests in, continuing as a key focus area into the second stage of maturity. Data operations and governance tends to emerge as a priority in stage three. Collaboration between IT and business groups signals the movement toward a centralized policy. Unified data collection, processing and monitoring policies are essential for the integrity and growth of the Data Lake. At the highest level of maturity, the organization will strictly comply with enterprise-grade security, disaster recovery and backup across the public and private cloud infrastructure.

Program Measurement
The notion of measuring the measurement program is a bit of a meta-game, but without a formal evaluation process dedicated to optimizing efforts in data-related programs, it is possible that under-performing data strategies will flounder unchecked. A culture of "test and learn" is key to creating an environment where failure is embraced as a pathway to discovering success. The goal of "failing fast" is integral to the ability to succeed in the long run. Advanced analytics is a process of creating hypotheses and testing them. The losses become lessons learned and the wins become profitable new business rules. Routine cost/ benefit analyses driven by formal processes to evaluate the ROI of program investments is critical to fast-tracking business transformation.

Investment Focus
Investment should be tied closely to measurement. In less mature organizations, investments in Apache Hadoop and other Big Data infrastructures are easy to justify because of the immense data storage and EDW efficiencies that deliver almost immediate ROI. But as the roadmap shifts to analytic applications and data mining, mature organizations expand their investment beyond optimizing operations and services, to invest in the development and refinement of new business models.

The process management domain analyzes how your company manages and evaluates your Big Data program.

"What matters is to reshape the internal decision-making culture so that executives base their judgments on data rather than hunches. Research already indicates that companies that have managed this are more likely to be productive and profitable than the competition."

Source: strategy&, Big Data Maturity—An Action Plan for Policymakers and Executives, 2014

Comparing to Industry Benchmarks

In 2015, the modern enterprise is still clearly finding its legs when it comes to Big Data. Across our five key capability domains for Big Data maturity, only the technology-focused domains consistently score higher than the bottom rung of the ladder. IT appears to be leading the Big Data revolution, and it figures because IT maintains the beachhead where the tidal wave of new business data is coming ashore. But IT alone cannot forge the path to business transformation.

IT departments make a compelling business case for modernizing the data center with Apache Hadoop. With the Enterprise Data Warehouse under strain, IT has used Apache Hadoop to offload low-value ETL processes and reduce incremental spend while dramatically increasing storage capacity. But the economics of data storage and processing pale in comparison to the transformative power of advanced analytic applications built on top of a Big Data architecture like Apache Hadoop. Forrester stated in 2014, "Apache Hadoop's momentum is unstoppable… a must-have data platform for large enterprises, forming the cornerstone of any flexible future data management platform."

A 2015 study by Gartner found that nearly half of organizations plan to implement Big Data architecture (Apache Hadoop) by 2017, with 37 percent of companies surveyed already deployed (26 percent) or deploying by 2016 (11 percent). Compare those strong technology adoption figures to more vital strategic capabilities: A recent IDC survey of over 4,000 organizations of all sizes found that only 25 percent had a business analytics strategy, and 50 percent either didn't know how or found it too difficult to measure return-on-business analytics project investments. Despite difficultly creating measurement strategies for Big Data investments, the investments are clearly flowing. A study from Accenture showed that 73 percent of companies are investing more than 20 percent of their technology budgets in Big Data initiatives. It also showed 22 percent of companies are spending more than 30 percent of their technology budgets on Big Data.

Our two-year outlook on industry-wide Big Data maturity is bright, but it follows the current pattern: technology will continue to lead among Big Data capability domains, constrained by a lack of strategic vision and organizational alignment. The presence of strategy and vision is directly correlated to the presence of consistent sponsorship, advocacy and funding. An explosion of software and hardware solutions targeting Big Data is driving growth across industries, and the emergence of common platform standards through projects such as the Open Data Platform is hastening the market's ability to develop new analytic applications that leverage Big Data. IDC predicts that the Big Data technology and services market will grow at a 26.4 percent rate annually, to $41.5 billion in 2018—about six times the growth rate of the overall information technology market.

Forrester stated in 2014, "Apache Hadoop's momentum is unstoppable…a must-have data platform for large enterprises, forming the cornerstone of any flexible future data management platform."

Source: The Forrester WaveTM: Big Data Apache Hadoop Solutions, Q1 2014

Early adopters of Big Data technology have paved the way for fast followers, and a mass-market awakening is now underway, with businesses of all sizes launching proof of concept projects. Technology, infrastructure and data analysis capabilities will continue to pull the market forward into 2017, but these tech-centric capabilities won't drive business transformation without similar advances in three key capability domains: sponsorship, skills and process management. Rising awareness of the Big Data opportunity among CEOs and other non-technical C-suite executives is igniting vision and strategy planning, but a lack of talent and the ability to change traditional organizational processes will continue to limit the potential of organizations to reach a transformative maturity level by 2017.

Apache Hadoop is the rising star of the business technology agenda for a simple reason—it disrupts the economics of data, analytics, and someday soon, all enterprise applications.

Source: The Forrester WaveTM: Big Data Apache Hadoop Solutions, Q1 2014

Stages of Maturity			
1. Aware	2. Exploring	3. Optimizing	4. Transforming

Capability Domains: Sponsorship, Data and analytics, Technology and infrastructure, Organization and skills, Process managment

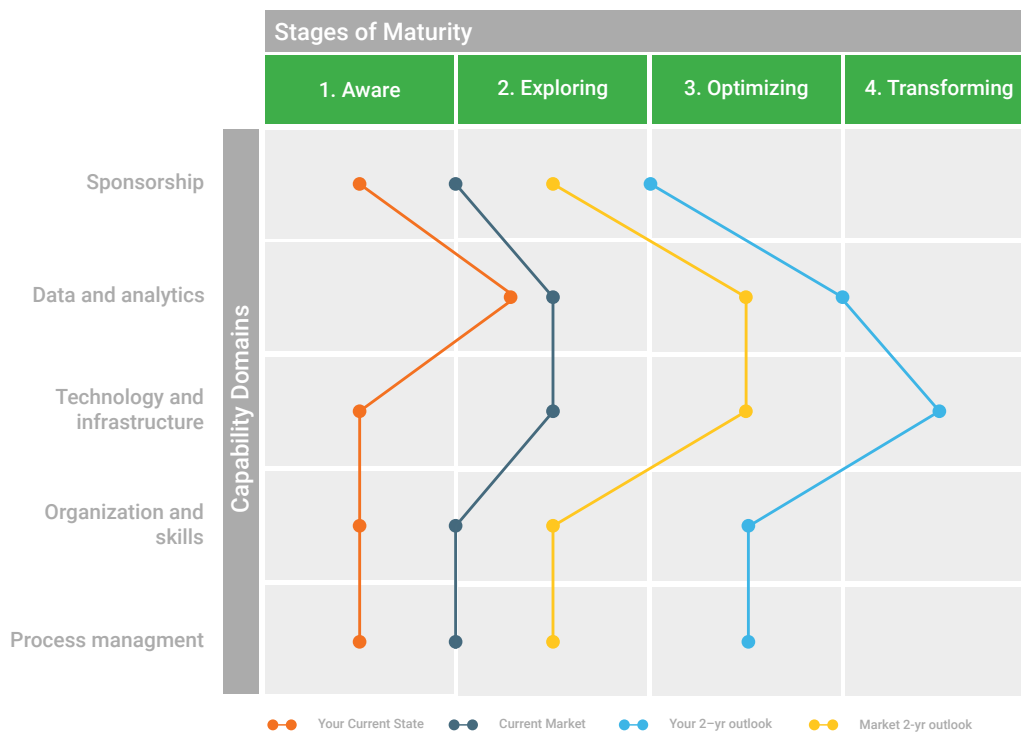Legend: Your Current State · Current Market · Your 2–yr outlook · Market 2-yr outlook

Figure 7. (For illustrative purposes only) Sample analysis that highlights a company's current and future state big data maturity vis-a-vis that of the average industry's current and future state maturity.

Plotting Your Strategic Path Forward

Given the general lack of maturity in this bourgeoning space, your near-term focus will likely address up-leveling your vision and business case, addressing your skills gaps, and defining your process management. If your organization's Big Data maturity is like most (see Figure 6), you're staring at a graph that puts you in the lowest stage of maturity across 60 percent of key focus areas. Knowing where to focus specific investment is critical, but nearly every organization faces four fundamental blockers:

1. Executive leadership

2. Organizational model

3. Human resources

4. Strategic planning

The broad strokes of your Big Data roadmap might align to the particular challenges and opportunities of your industry sector. Generally speaking, manufacturers need to invest more in IT assets. Real estate, wholesale trade and management sectors face critical challenges in data availability. Government and education face unique vision and leadership challenges. And every industry faces severe challenges around skills—especially retail, construction, hospitality and entertainment. Conversely, inherent advantages exist for some industries. Utilities, telecommunications, manufacturing and health care tend to have the most readily available data. Transportation, management and warehousing tend to be better invested in IT infrastructure. Finance, manufacturing and professional services tend to worry less about access to talent.

But the finer details of your Big Data strategy should address the gaps between where your business should go with Big Data and the capabilities you need to get there—developing an executive-level vision based on understanding your Big Data opportunities, and targeting the focus areas most relevant to executing your roadmap. Many roadmaps begin by aligning the right people and skills with an enabling organizational structure. Hortonworks launched a Joint Innovation program in 2015 to address the widespread need for these strategic fundamentals.

EXECUTIVE LEADERSHIP IS A MUST

Several industry studies point to the increasingly vital role played by CEOs and IT Chiefs in moving Big Data projects forward. A 2014 study by A.T. Kearney found that analytics success is tied directly to executive leadership. Among organizations with the highest levels of Big Data maturity, 59 percent vested responsibility for Big Data in a CXO, compared to 18 percent who put the company's Big Data leadership with the head of IT.

IDC's 2014 Enterprise Big Data study revealed that CEOs are the driving force behind Big Data strategies (47 percent of respondents said the CEO is sponsoring their project), while IT Chiefs are most likely to lead analysis initiatives. It's no surprise CEOs are throwing their weight behind Big Data, considering that only 26 percent in the IDC survey pointed to ROI as an obstacle for their Big Data projects. Nearly half cited budget as a challenge, while the other leading obstacles were the availability of skilled employees (41 percent) and development time (39 percent).

When it comes to implementing a Big Data solution, the Accenture study identified cross functional integration as the number one challenge. The study showed that 36 percent of respondents said system barriers between departments prevent collection and correlation of data. This is where Apache Hadoop's shared data service—Data Lake—can be particularly valuable.

Big Data strategy should address the gaps between where your business should go with Big Data and the capabilities you need to get there—developing an executive-level vision based on understanding your Big Data opportunities, and targeting the focus areas most relevant to executing your roadmap.

Source: A.T. Kearney, Big Data and the Creative Destruction of Today's Business Models, 2013

OPTIMIZE YOUR ORGANIZATIONAL MODEL

How should companies organize to capture the benefit of Big Data and move swiftly to higher maturity stages? According to A.T. Kearney, three models are currently emerging to deliver critical analytics support across the modern enterprise: a) decentralized, b) embedded and c) standalone.

In the decentralized model, analytics resources sit inside each business unit or functional group. From our perspective, this model tends to be most common in lower stages of maturity, providing a quick-start capability to early-adopter business groups. Close collaboration between analysts and the business result in faster time-to-value, but the model falls short when the company's Big Data vision eventually centralizes, creating the need for integration between data and tooling across the organization. But the drawbacks of this model are obvious: analytics programs are inherently local in focus, lacking a broader strategic roadmap—a condition that is only reinforced by the fact that analytics lack a voice in upper management. The lack of integration inherent in this model also breeds duplication of effort and an inability to scale insights.

In the embedded model, analytics resources are situated between the business groups and business processes. As a single analytics organization, this model creates new efficiencies. Processes and methodologies for data capture and processing are uniformly applied, creating conditions necessary for a Data Lake. Shared services help scale the impact of functional expertise, and a company-wide Big Data strategy can be evangelized and supported consistently. The difficulties of this model stem from a less transparent allocation of resources, which is connected to the fact that the analytics group has essentially grown into a crossfunctional unit from inside a leading business group. And without direct reporting to the C-suite, the embedded group will struggle creating cross-functional alignment on strategy and processes. We tend to see the embedded model more commonly for companies that are along the optimization stages of maturity.

In the standalone organizational model, the analytics group sits between executive functions and business groups. This model is typical in more mature companies with a strong executive vision and mandate for cross-functional integration. This resource-intensive analytics organization may take the form of a center of excellence, and may be supported by a crossfunctional advisory group that ensures the representation of business interests in the administration of company-wide data policies. The standalone model creates faster adherence to centralized policies and processes, but sometimes results in a perception that the analytics group lacks the business-level insight or expertise more common in a decentralized model. The allocation of Big Data resources is most often contested in this model as well.

The optimal organization model for most enterprise-level Big Data programs is a hybrid decentralized/standalone structure where there is a combination of decentralized services for business intelligence and standalone shared services for analytics. With business intelligence embedded within the business units, the entire company develops a view of performance insights with a culture of hypothesis and data-driven decision making as opposed to the "gut feeling" approach. The key value of the centralized standalone analytics function is that it helps the leadership pursue a predictive, rather than reactive, view of the opportunities. In addition, as an executive function, the standalone analytics organization will have the budget and authority it needs to effect change, and the participation it requires from across the organization to advance an enterprise-wide roadmap.

Many times, the most effective business model for big data analytics is a combination of decentralized services for business intelligence and standalone shared services for analytics.

Source: A.T. Kearney, Big Data and the Creative Destruction of Today's Business Models, 2013

HAVE A PLAN FOR HUMAN RESOURCES

How to structure your Big Data organization may seem a simple challenge when compared to the difficulty you will face staffing it. Companies at lower levels of big data maturity will find it beneficial to leverage the rapidly expanding data and analytics ecosystem. The options have never been better for turnkey solutions in foundational technologies like Apache Hadoop, operational and analytic infrastructure, data visualization and BI tools, databases and pre-packaged vertical applications. Deciding which resources to outsource and which to bring in-house is a critical decision, and your approach will evolve throughout your journey.

Similar to the A.T. Kearney study, we at Hortonworks have characterized eight key roles for Big Data maturity. Across these eight roles are ten overlapping skillsets. Core to every enterprise that is accelerating business transformation using data as a key enabler are data scientists and database administrators, followed in importance by software engineers and business analysts. When it comes to skills, many organizations are surprised to learn that creativity is one of most important qualities to screen for across roles; creativity is especially important among data scientists and analysts. Together, these core team members help the organization work through a constant stream of "what if" scenarios—helping business stakeholders identify opportunities, and testing hypotheses around data. Creativity is also essential in communicating data insights. Hence, the rise in popularity of data visualization.

Your decision to hire or outsource Big Data talent is likely to be influenced by your geographical location. Europe and North America both have a wide range of required talent, but lack the depth of Asia. North American companies in particular face a serious drought of key talent over the next several years. In the United States, Big Data skills are geographically unequal in their distribution. A recent Hortonworks study found that 51 percent of U.S. Apache Hadoop engineering jobs (including data scientists, developers and administrators) exist on the West Coast, with 24 percent of jobs located in the South, 18 percent in the Northeast, and only eight percent in the Central U.S. Demand for Apache Hadoop engineering skills has increased by more than 1500 percent between 2011 and 2015. McKinsey predicts a 50- to 60-percent gap in deep analytic talent by 2018, exacerbating a problem that already forces many organizations to scramble for talent and rely primary on outsourcing for key Big Data capabilities.

Deciding which resources to outsource and which to bring in-house is a critical decision, and your approach will evolve throughout your journey.
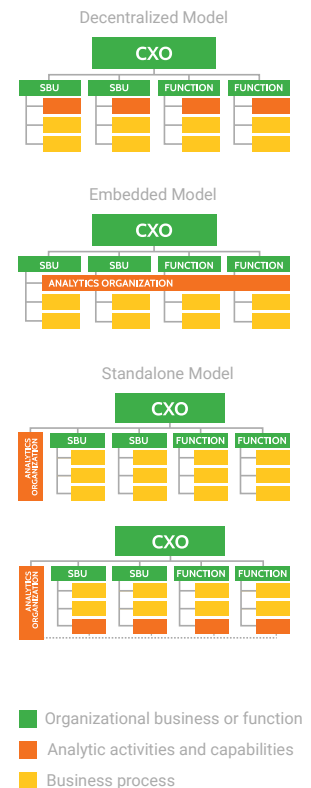

Figure 8: Four organizational models for Big Data analytics Source: A.T. Kearney, Big Data and the Creative Destruction of Today's Business Models, January 2013
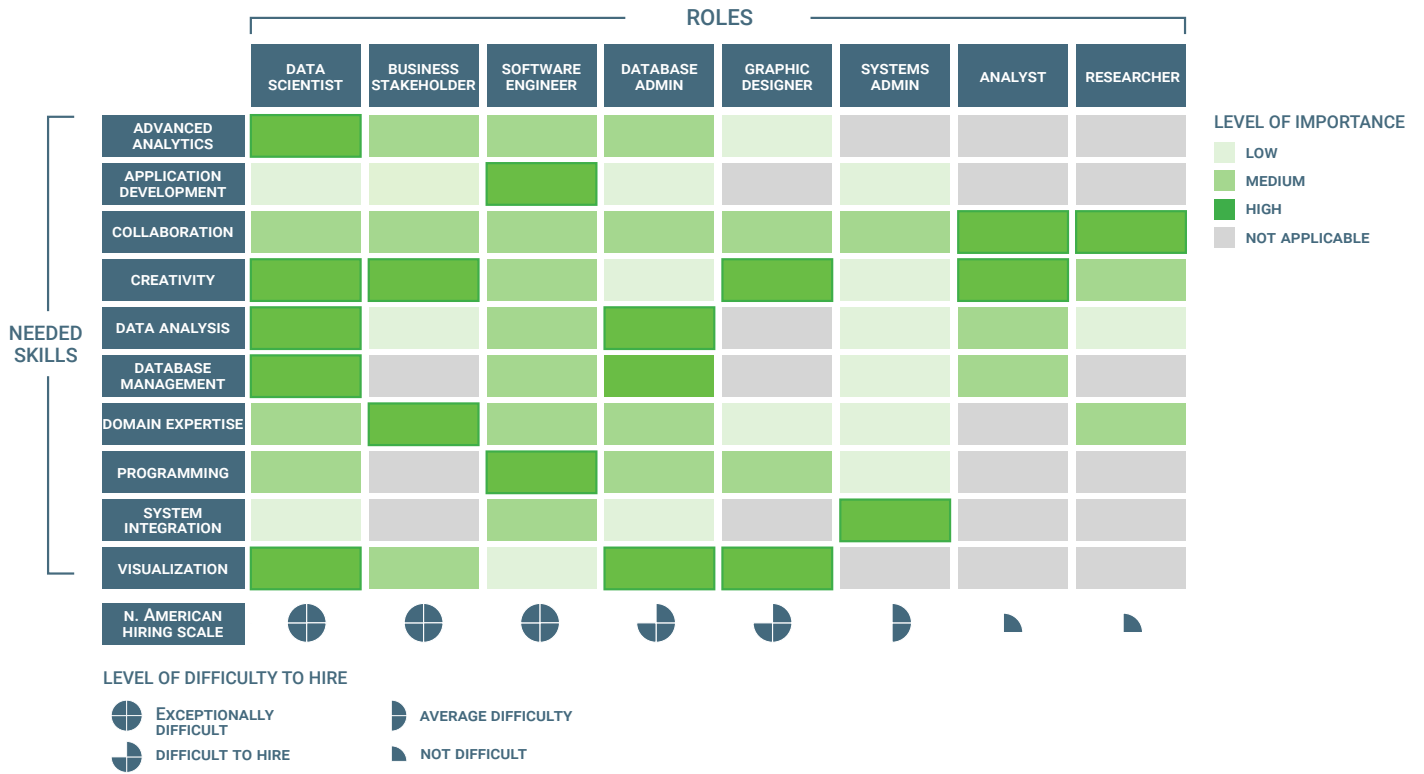
BIG DATA SKILLS CHESSBOARD

Figure 9: The skills required to form a strong data-first organization

"DIFFICULTY TO HIRE" RATINGS BY ROLE FOR SILICON VALLEY
(Based on 1-100, 100 being VERY difficult)

99	Data Scientist	80	Graphic Designer
95	Business Stakeholder	60	Systems Administrator
90	Software Engineer	50	Analyst
80	Database Administrator	50	Researcher

Source: Hortonworks Analysis

INVOLVE EXPERTS IN STRATEGIC PLANNING

The shrinking availability of key talent is likely to force a greater reliance on outsourced Big Data resources in the coming years. Deciding where and how to focus hiring efforts is among several strategic roadmap decisions that can be supported by experts. Assessing your organization's Big Data maturity and having a basic understanding of the opportunities in your industry are precursors to participating in a planning program like the Hortonworks Joint Innovation Program.

The Hortonworks Joint Innovation Program is designed to jumpstart the strategic planning process for your company's Big Data roadmap, resulting in a company vision, strategy and execution plan. The Joint Innovation Program is a co-innovation process that helps executive teams navigate their path to higher stages of Big Data maturity. The program helps you:

- Improve the success rate of innovation by applying Design Thinking methodology in

- Innovation Workshops

- Identify transformational opportunities for your business

- Define use cases and proof of concept plans

- Identify value opportunities, barriers and risks

- Create a value-based roadmap with defined milestones

- Define an operating plan aligned to your scope

Hortonworks' experience with the world's largest enterprises creates a valuable base of knowledge that can be leveraged to benefit businesses in multiple industries. The Joint Innovation Program is a series of collaborative workshops focused on value discovery, roadmap development, and executive planning for governance and operations. The business case that emerges from the program helps organizations align on the right reasons for investing in Big Data, the projected benefits and costs and the path to value realization.

Hortonworks is committed to working closely with enterprise customers to facilitate the migration to higher levels of maturity around Big Data and Apache Hadoop solution deployment.

Figure 10. From vision to value, the Hortonworks Joint Innovation Program offers access to the world's most experienced Big Data practitioners

OPEN LEADERSHIP

Hortonworks has a singular focus and commitment to drive innovation in the open exclusively via the Apache Software Foundation process.

Hortonworks is responsible for the majority of core code base advances to deliver Apache Hadoop as an enterprise data platform.

ECOSYSTEM ENDORSEMENT

Hortonworks is focused on the deep integration of Apache Hadoop with existing data center technologies and team capabilities.

Hortonworks has secured strategic relationships with trusted data center partners including Microsoft, SAP, Teradata, Rackspace and many more.

ENTERPRISE RIGOR

Hortonworks has a world-class enterprise support and services organization with vast experience in the largest Apache Hadoop deployments.

Hortonworks engineers and certifies Apache Hadoop with the enterprise in mind, all tested with real-world rigor in the world's largest Apache Hadoop clusters.

For an independent analysis of the Hortonworks Data Platform and its leadership among Apache Hadoop vendors, you can download the Forrester WaveTM: Big Data Apache Hadoop Solutions, Q1 2014 report from Forrester Research.

About Hortonworks

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

Contact

For further information visit
visit www.hortonworks.com.

+1 408 675-0983
+1 855 8-HORTON
INTL: +44 (0) 20 3826 1405

A16_078

References

Angel, Nasry. "Are You Using Data To Win? Discover Your Data Management Maturity." Forrester. Forrester Research, Inc, 26 Feb. 2015. Web. 6 Aug. 2015.

El-Darwiche, Bahjat, Volkmar Koch, David Meer, and Walid Thome. "Big Data Maturity: An Action Plan for Policymakers and Executives." Strategy&. PwC Network, 2014. Web. 6 Aug. 2015.

Evelson, Boris and Martha Bennet. "Establish Ongoing Processes To Assess Your Business Intelligence Maturity." Forrester. Forrester Research, Inc, 15 Dec. 2015. Web. 6 Aug. 2015.

Goetz, Michele. "Do Data Quality The Big Data Way." Forrester. Forrester Research, Inc, 23 Apr. 2015. Web. 6 Aug. 2015.

Hagen, Christian, Hugo Evans, Dan Wall, Khalid Khan, Bharath Thota, and Anand Seshadri. "IT's Challenge: Bringing Structure to the Unstructured World of Big Data." A.T. Kearney. A.T.Kearney, Inc, 2014. Web. 6 Aug. 2015.

Hagen, Christian, Dan Wall, Khalid Khan, Marco Ciobo, Ajay Yadav,a nd Jason Miller. "Big Data and the Creative Destruction of Today's Business Models." A.T. Kearney. A.T.Kearney, Inc, 2013. Web. 6 Aug. 2015.

Halper, Fern, and Krish Kirshnan. "TDWI Big Data Maturity Model Guide: Interpreting Your Assessment Score." TDWI Research. 1105 Media Inc, 2014. Web. 6 Aug. 2015.

Hopkins, Brian, and Noel Yuhanna. "Critical Big Data Insight From Strata 2015." Forrester. Forrester Research, Inc, 29 May 2015. Web. 6 Aug. 2015.

Hopkins, Brian, and Ted Schadler. "Digital Insights Are The New Currency Of Business." Forrester. Forrester Research, Inc, 27 Apr. 2015. Web. 6 Aug. 2015.

Hopkins, Brian. "Brief: An Elastic Analytics Platform: SQL-For-Apache Hadoop Foreshadows A Next-Generation Platform For Elastic Analytics On Big Data." Forrester. Forrester Research, Inc, 24 July 2015. Web. 6 Aug. 2015.

Hopkins, Brian. "Brief: Turning Big Data Into Business Insights, 2015." Forrester. Forrester Research, Inc, 20 Feb. 2015. Web. 6 Aug. 2015

Kisker, Holger. "TechRadar: Business Intelligence, Q1 2015." Forrester. Forrester Research, Inc, 23 Feb. 2015. Web. 6 Aug. 2015.

Manyika, James, Michael Chui, Brad Brown, Jaques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. "Big Data: The next Frontier for Innovation, Competition, and Productivity." McKinsey Global Institute. McKinsey & Company, June 2011. Web. 6 Aug. 2015.

McCormick, James. "Measure Your Digital Intelligence Maturity." Forrester. Forrester Research, Inc, 14 Oct. 2014. Web. 6 Aug. 2015.

Radcliffe, John "Leverage a Big Data Maturity Model to Build Your Big Data Roadmap. " Radcliffe Advisory. Radcliffe Advisory Services Ltd, 7 Jan. 2014. Web. 6 Aug. 2015.

"Rise of the Data-First Enterprise: A Hortonworks White Paper." Hortonworks. Dec 2014. 6 Aug 2015.

"Using Big Data + Analytics To Drive Business Transformation." IDC. IDC Corporate USA, Spring 2015. Web. 6 Aug. 2015.

Vesset, Dan, Greg Girard, Cynthia Burghard, Debra Osswald, Michael Versace, Adelaide O'Brien, Jill Feblowitz, and Simon Ellis. "IDC MaturityScape: Big Data and Analytics — A Guide to Unlocking Information Assets." IDC. IDC Corporate USA, 2013. Web. 6 Aug. 2015.