big **DATA**

GETTING THE RIGHT PEOPLE

on the Big Data Bus

a SAS Best Practices white paper

by
Tamara **DULL**
& Anne **BUFF**

sas best practices
THOUGHT PROVOKING BUSINESS

"Good-to-great leaders understand three simple truths:

If you begin with 'who,' you can more easily adapt to a fast-changing world.

If you have the right people on your bus, you don't need to worry about motivating them.

If you have the wrong people on the bus, nothing else matters."

– Jim Collins, author of *Good To Great*

table of **CONTENTS**

Embracing big data is an ongoing progression of change; a journey that requires a rethink in cultural norms.

EXECUTIVE SUMMARY

When considering the challenges businesses face in today's evolving digital world, business leaders expect that big data will drive big change. And it likely will.

But as with most technical revolutions, the golden trifecta of business transformation – people, process and technology – is critical. The challenge is that the three pieces of the beloved transformation trinity aren't equal when it comes to big data. This is because embracing big data to become a data-driven company is not a single transformation initiative. It is an ongoing progression of change; a journey that requires a rethink in cultural norms.

While process and technology certainly play a role in this evolution, the shift in culture is solely dependent on the people. From the structure of the organization and defined roles and responsibilities to the teams that will convene and dissolve as initiatives come and go, the *Who* of big data is what matters most – and that's what this white paper is all about.

We address the *Who* of big data from two perspectives: the *Why* and the *How*. It is important to address both of these perspectives when determining the right people to have on board for successful big data initiatives:

- The *Why* perspective looks at your organization and answers these questions:
  - *Why* is your business interested in big data and why now?
  - *Who* will make your big data vision and mission successful?
- The *How* perspective looks at the tactics of building a big data team with these questions in mind:
  - *How* will you achieve your initiatives?
  - *Who* will make your projects successful?

So as you begin your journey, get the right people on your big data bus – consider it an insurance policy for the very success of your big data program.

INTRODUCTION

Big Data So Far: What's Working, What's Not

Big data is set to have big impact for businesses. In Accenture's recent survey, *Big Success With Big Data*, 85 percent of the respondents said big data would dramatically change how they do business, with the greatest impact being customer relationships (37 percent) and product development (26 percent). As emphatic as we are about big data driving big change, we must caution that not all change will be fast or necessarily positive – at least not at first.

Industry research firms suggest that while big data initiatives are gaining steam across industries, they are not without challenges. Capgemini Consulting reports in its *Big Data Survey* that only 27 percent of the executive respondents classified their big data initiatives as successful. Put these success rates in context, however, as the same survey revealed only 13 percent of the responding companies had big data projects in full-scale production. If planned projects are not in production, there may be unexpected obstacles, and of course, executives are not going to report them as successful (yet).

In Accenture's survey, researchers only included companies that had completed at least one big data project. Of those surveyed, 92 percent reported they were satisfied with their business outcomes, and 94 percent said the big data implementation met their needs. This suggests that when companies can overcome the obstacles in completing a big data project, then the prospects of extending business capabilities and uncovering new opportunities through big data are favorable. The question then becomes, what are the obstacles businesses face when implementing big data initiatives?

Technology and budget are certainly limitations reported across almost all research surveys on big data implementations. However, without fail, the biggest and most commonly reported obstacle is the organization itself. CA Technologies found 92 percent of the respondents in its study, *The State of Big Data Infrastructure*, agreed there were major obstacles to implementing their big data initiatives, with eight of 12 obstacles being people- or culture-related.

A flexible framework for tackling big data is imperative. Having the right people, internal and external, and using them effectively on the various paths on your big data journey is the key to overcoming obstacles and realizing success.

Having the right people and using them effectively on the various paths on your big data journey is the key to overcoming obstacles and realizing success.

There are two categories of questions to answer when determining the right people for successful big data initiatives.

The Two Perspectives of *Who*

There are two categories of questions to answer when determining the right people for successful big data initiatives. For the effective and efficient use of any data throughout the enterprise, these questions are imperative, but since big data seems to have the attention of executives these days, we will couch them in terms of big data programs and initiatives. At the risk of sounding like a bad imitation of an Abbott and Costello routine, the answers to the questions of *Why* and *How* inform the answers to the two questions of *Who.*

PERSPECTIVE 1. THE WHY

Answers to the questions of *Why* help to determine the people of your business that will bring your big data mission and vision to life and execute consistently regardless of the task:

Why is the emergence of big data promising?
Why is now the time to invest in our big data assets?
Why does big data matter?

The *Why* perspective shapes organizational design, roles and responsibilities, and the culture in which they are rooted. It informs the fundamental framework of people you want on your bus no matter where it is going. Carefully outlining the perspective of *Why* will ensure your people can adapt to the ebbs and flows in the world of big data without straying from the core business values and principles.

PERSPECTIVE 2. THE HOW

The answers to the questions of *How* determine who will ensure your specific big data initiatives are completed successfully:

How will you achieve your initiative?
How will you define the scope of a specific initiative?
How will you execute?

The *How* perspective guides which people are assigned to individual projects, the timelines for when they are engaged, and the plans on how they will execute. This defines the cross-functional teams that will convene and dissolve with initiatives over time. Sticking with the bus analogy, the *How* determines the driver and passenger seat assignments depending on the destination of the current initiative.

UNDERSTANDING

the big data organization

> "People are not your most important asset. The right people are."
>
> – Jim Collins

The people of the business are what make or break it. The right people make it right, and the wrong – well, you get it. But what makes a person right or wrong? Interestingly, it is not individual characteristics that make people right or not. While there are certainly some characteristics better than others, the right fit is what really matters. This means the people have to be right for the company and vice versa.

You must first define the *Why* before you can identify who the right *Who* is. The *Why* is the purpose, the reason for "doing" big data. It encompasses the shared values and vision of the company and is the foundation for all big data efforts. The right *Who* will share the company's beliefs and values and will do everything they can to support and promote the *Why* of the business. However, finding the right *Who* and the right fit is no easy task.

What Really Matters?

When defining the *Why* perspective of *Who,* it is easy to get wrapped up in titles and organizational charts. However, titles and reporting structures are not the key aspects in defining a data-driven company. While they do have importance, role definitions, job descriptions and performance measures carry much greater weight in determining the long-term effectiveness of organizational changes.

TITLES

Titles define more about a company than they do about the job they represent. A simple search through a series of job boards will return a different description for every job title posted. No two companies have the same description or requirements for the same title. Some are very detailed, while others are suspiciously vague.

Titles superficially convey the company's values and vision and shed light on the company culture. They do not determine the expectations of the position; instead, they reflect hierarchy and reporting structures. The title *lead data scientist*, for example, suggests the company has a team with organized leadership focused on using data to advance the business. The title, however, does not indicate the skill or function.

The *why* is the reason for "doing" big data. It encompasses the shared values and vision of the company and is the foundation for all big data efforts.

Consider titles a form of external messaging about your business, but do not expect titles to define roles and responsibilities.

Titles have also become recruitment tools to attract specific candidates, but ironically research suggests titles are not effective for employee satisfaction and retention.[1] And too much play with titles often leads to title inflation. Just like you can't judge a book by its cover, you can't judge a job by its title – but you may be able to judge the company. When designing the framework for your big data organization, consider titles a form of external messaging about your business, but do not expect titles to define roles and responsibilities.

ROLES AND RESPONSIBILITIES

Roles and responsibilities are defined through job descriptions and organizational design. A job description is an overview of the position's purpose, expectations of what is to be performed/accomplished and requisite qualifications. While they should not be long or detailed, job descriptions should be specific such that all employees and potential candidates clearly understand the requirements and expectations of the position.

The organizational design is a combination of defined structure and dynamics that defines levels of responsibility and decision making, as well as patterns of communication and collaboration. Organizational design has a significant impact on the shape of the company culture.

METRICS

While job descriptions and organizational design define expectations of company roles and how they should interact, they do not provide a mechanism for ensuring expectations are met. Tactical execution and associated measures of success are functions of performance management. Effective performance management ensures individual and team contributions support and further the overarching business goals. Key performance indicators (KPIs) are quantifiable metrics used to measure the degree to which goals and objectives are met. These can and should be defined at both the individual contributor and department levels, where individual KPIs support and align with the departmental KPIs.

Defining the *Who* for your business from the *Why* perspective is not simply determining a checklist of new roles needed or a realignment of your current hierarchical structure. It requires intense comprehensive planning with consideration to titles, roles, job descriptions, organizational design and performance measures. Remember, these are the people who will be on your big data bus regardless of the journey. You want to make sure they are the right people operating in a framework that ensures success.

[1] Grasz, Jennifer. "CareerBuilder Survey Reveals Most Wanted Office Perks and ..." CareerBuilder.com. January 24, 2013. Accessed September 17, 2015.

Leadership Positions

Leadership is a critical success factor for the *Why* of big data in any business as leaders must establish, communicate and reiterate the company strategy within the big data context. The recent spate of new chief officer titles suggests companies are using these roles as a first step for important initiatives, but not necessarily with consideration to the other leadership needs of the business. When companies establish these roles prematurely, not only are the individuals not successful, but the company also suffers significant setbacks as it recovers from the failed attempt.

The definition of a CXO title, where X is almost anything these days, is based on the three elements in the title itself: authority, focus and role. Chief (C) represents the fact that this position is the designated highest level of authority for X, the defined primary or fundamental focus. Officer roles (O) historically were leadership roles appointed by the board, but they have now become more commonplace titles that represent what is important to the business. Because of this, it can no longer be assumed a CXO reports to the board or even sits near the top of the company.

It is still expected, however, that individuals holding officer titles are influential leaders regardless of where they sit. The function of any CXO position should be to align the fundamental focus of responsibility (X) with the company mission and strategy through leadership. CXOs should have a keen understanding of the opportunities endemic to their focus areas and possess the skills and authority to support business strategy within those areas.

There are two key factors that determine the success and fate of these newly defined chief roles. Of course, the leadership capability of the individual role will matter, but assuming the right individual is placed in the job, authority and organizational maturity will be the determinants of success. If given a chief title, the role must have complete authority for its focus area, with the right to make decisions autonomously to be effective. Without appropriate authority, the chief role has no value and should not be used.

Careful consideration must be given to the designated primary area of focus for the position. The business must be ready for this to be its own separate and distinct focus area companywide. This means there will no longer be pockets of X throughout the business; defining X as a primary focus area means all pieces and parts previously spread throughout the company will now fall under the purview and direction of the new chief officer. This is a major shift for which many are not prepared; thus, the position is created prematurely and subsequently fails. It is not about whether the chief is ready to lead; it is about whether the business is ready to be led.

It is not about whether the chief is ready to lead; it is about whether the business is ready to be led.

It is imperative the entire business clearly understands and accepts the definition and purpose for each leadership role.

As you can see in the following examples, there are many places where defined leadership roles could overlap. It is imperative the entire business clearly understands and accepts the definition and purpose for each leadership role. Not all roles will exist in every company. Nor should they. Chief roles should only be created for business areas ready to be seen as distinct organizational units by the entire company.

CHIEF MARKETING OFFICER

1. Leader with the highest level of authority regarding *marketing practices* companywide.
    a. The area of marketing practices may include media, public profile and brand management; customer intelligence and experience; market segmentation; and distribution channel management.
    b. The definition and all included aspects of marketing practices are clearly defined and accepted by all lines of business.
2. Provides direction and clarifies boundaries for marketing practices throughout the enterprise and will ensure key business initiatives are supported.
3. Responsible for aligning marketing practices with the company mission and strategy.

CHIEF DATA OFFICER

1. Leader with the highest level of authority regarding *data practices* companywide.
    a. The area of data practices may include data quality, data management, data governance and/or data architecture.
    b. The definition and all included aspects of data practices are clearly defined and accepted by all lines of business.
2. Provides direction and clarifies boundaries for data practices throughout the enterprise and will ensure key business initiatives are supported.
3. Responsible for aligning data practices with the company mission and strategy.

CHIEF DIGITAL OFFICER

1. Leader with the highest level of authority to lead *digital transformation* companywide.
    a. Digital transformation may combine any number of aspects, including marketing, technology, mobile applications, social networking and/or product development.
    b. All aspects of the business's digital transformation strategy are clearly defined and accepted by all lines of business.
2. Provides direction and clarifies boundaries for digital transformation and will ensure key business initiatives are supported.
3. Responsible for aligning digital transformation activity with the company mission and strategy.

CHIEF INFORMATION OFFICER

1. Organizational leader with the highest level of authority for enterprisewide *information technology.*
    a. Information technology typically includes the oversight of all technology and systems used in business operations.
    b. The scope of information technology management is clearly defined and accepted by all lines of business.
2. Provides direction and clarifies boundaries for business technologies and will ensure key business initiatives are supported.
3. Responsible for aligning information technology with the company mission and strategy.

Individuals whose characteristics align with the core business values and mission will adapt with the changes the business makes.

CHIEF ANALYTICS OFFICER

1. Organizational leader with the highest level of authority for *analytics and business intelligence.*
    a. Analytics and business intelligence can include reporting, data analysis, data science and/or data mining. The focus may be on business process, supporting technology or a hybrid of both.
    b. The definition and scope of analytics and business intelligence is clearly defined and accepted by all lines of business.
2. Provides direction and clarifies boundaries forr the use of analytics and development of business intelligence and ensures key business initiatives are supported.
3. Responsible for aligning analytics and business intelligence with the company mission and strategy.

Individual Contributors

As with C-level leaders, individual contributors involved with big data can ultimately make or break the effort, even impacting the health of the business. While they do not serve in leadership roles or have direct reports, they provide substantial input in business direction and have sincere interest in the success of the company.

Filling your bus with solid individual contributors based on key traits and capabilities rather than specific skill sets for defined positions will prove to be far more successful in the long term. Individuals whose characteristics align with the core business values and mission will adapt with the changes the business makes, whereas the need for unique skill sets will likely not transfer over time.

Getting the right people on your bus is not easy. Once they are on, it is the business's responsibility to keep them there. Employee retention of valued contributors is a job all its own. It is important these individuals understand the *Why*, the big data mission, as they get on board; and it is equally important, if not more, they understand the reasons for shifts and changes over time.

Aligning their actions and values with the company's overall mission and goals is important to these contributors. Clear, consistent communication shared regularly on behalf of the business about its direction will help ensure this alignment. Without consistent communication, the value alignment will gradually wane over time, decrease employee satisfaction, and, in turn, increase turnover.

The business is also responsible for ongoing employee development. Many individual contributors are not interested in management opportunities; they thrive instead on being influential in the roles in which they serve. Providing training opportunities and feedback for employee growth demonstrates that the company values their efforts and allows them to expand their organizational influence. While moving up the ladder does not appeal to all, career progression is still very important. Employees are more likely to stay with companies where they are valued, successful and progressing. It is the business's responsibility to ensure this happens.

Organizational Design

Organizational design is a combination of structure and dynamics. It is the culture and framework that supports the core functions of the business over the long term. The formal structure of a company is typically detailed through organizational charts that delineate reporting relationships, lines of empowerment and authority. The shape or style of the defined hierarchy (or lack thereof) is the structure that informs the coordination and supervision of operational procedures.

Organizational dynamics are the generally accepted behavioral norms for interaction and engagement throughout the business based on a shared system of values, beliefs and assumptions that evolve over time. The culture of the company is the environment created by the company's behavioral habits.

Some of the most common questions asked when tackling initial big data projects are about organizational structure:

***What** organizational changes should we make?*
***Should** we bring on a chief data officer? And if so, **when**?*
***How** many data scientists should we expect to hire?*

These questions should not be asked just because big data initiatives are on the horizon. Projects are sets of objectives and tasks that have defined start and end dates. The organizational structure should not be designed to address specific project requirements or even large groups of projects, because at some point they will reach their end. The expectation (or at least hope) is that the company's existence will not reach an end date; therefore the structure should be built to endure and weather changes over time.

Organizational restructuring may be necessary when the company adopts new business models, business processes change or new functions – for instance, a digital office – are introduced. The excitement, possibility and opportunity of big data might be the catalyst for a change in strategic direction. If this is the case, then the consideration of restructuring is appropriate.

The organizational structure should not be designed to address specific project requirements because at some point they will reach their end.

When determining the administrative style for a specific focus area, there are a variety of factors to consider.

Effective organizational design provides three key things: A functional operating model that supports the company's strategic priorities; a framework for organizing and managing physical and human resources; and reinforcement of the company's mission, vision and values. Just remember, one size does not fit all. The organization should be designed around the long-term needs of the business.

Organizational Charts

The likelihood of big data itself becoming an individual line of business in your organizational chart is very low. In fact, it probably won't happen. Big data is after all, still data. But what it does mean is enterprisewide operational data functions and practices could find their way onto the organizational chart, if they haven't already. The question is, where do all of the data roles and responsibilities fit? The answer is not so straightforward.

An organizational chart is a graphical representation of the reporting structure for the business. It defines distinct departments and the grouped functional operations for which each department is responsible. It illustrates rank and chain of command and gives a quick visual representation of the number of positions in each department.

With careful interpretation, an organizational chart can also shed light on the level of significance given to a specific focus area, where and how decisions are made, and where specific challenges might arise. No two organizational charts are identical, and there is no set template for success. However, there are a few common designs that demonstrate different styles of administration. Each style has its own benefits and challenges.

When determining the administrative style for a specific focus area (such as data) and where positions should sit within the hierarchy, there are a variety of factors to consider, including culture, company size, authority, budget, executive leadership and strategic direction. With these considerations in mind, you can determine which style is right for your company – distributed, integrated or centralized.

STYLE 1: DISTRIBUTED

A distributed administrative style does not lack authority or oversight. In some cases, this style may have the most management or administration. However, because the administrative roles and responsibilities are spread throughout the company, each position has a limited scope of responsibility and authority.

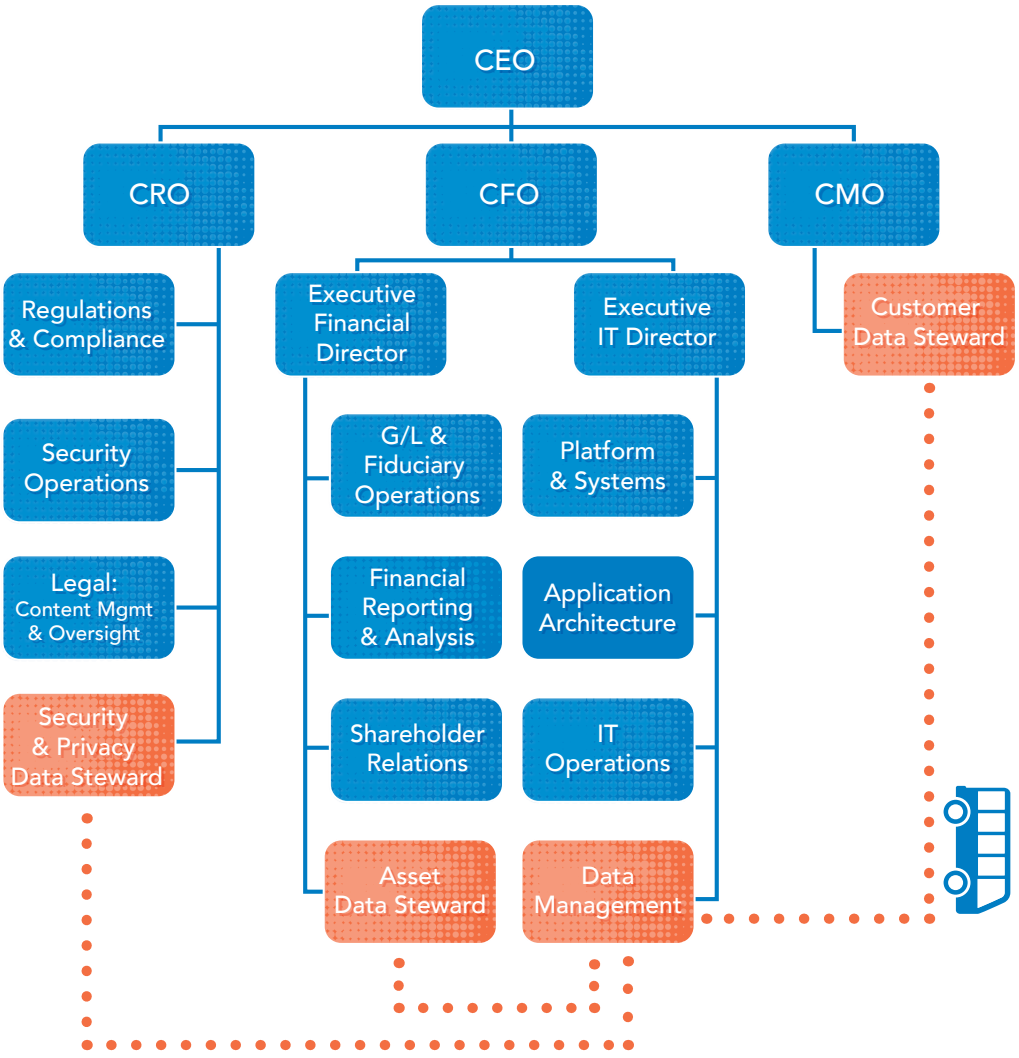

figure 1: distributed organizational chart

When considering data as a focus area with a distributed administrative style, each line of business reporting into a C-level executive has a position of data authority. These positions are responsible for the oversight, management and promotion of data within their reporting structure. They have the authority to act and manage people and resources for all things data-related within the scope of their department. Their role is to promote data-driven innovation and increase the value of data within their lines of business.

For consistency, these positions maintain a working relationship with a data management function that serves the rest of the company. The enterprise-level responsibilities of this data management function focus on the practices and tools enabling standardization and reuse. This function is also accountable for the reduction of data costs and effective resource use.

While responsible for data management across the enterprise, this role sits low in the organizational structure in the distributed model and is a peer with other data decision-making bodies spread throughout, loosely tying data roles and responsibilities together.

The distributed administrative style allows for data decision autonomy across business units and ensures data decisions are most effective for the specific department in which they are made while maintaining consistency in cross-functional operations. The distributed administrative style does not demonstrate an executive focus on data, nor does it foster executive engagement. But the repeated presence of a data authority in each business unit establishes the importance of data throughout the business and embeds the value at the operational levels. Be careful not to let the structure be an impediment to delivery.

While these lower-level data authority roles are designated as decision makers, the multiple levels of oversight can reduce the strength of power and influence they have in the business. This also typically suggests these roles have limited budgetary control, if any at all. Because budget and authority are necessary for effective management, it is imperative that the funding and procurement processes promote and support these positions to ensure they are sanctioned and credible.

A centralized leadership hub demonstrates commitment to the focus area by providing a clear vision and a framework of support.

STYLE 2: INTEGRATED

When the company's strategic direction is driven by a specific area of focus such as data, it is advantageous to have the core decision makers and influencers for the enterprise practices in the same reporting line. The streamlined oversight, targeted initiatives and explicit responsibility of a centralized leadership hub demonstrate commitment to the focus area by providing a clear vision and a framework of support.

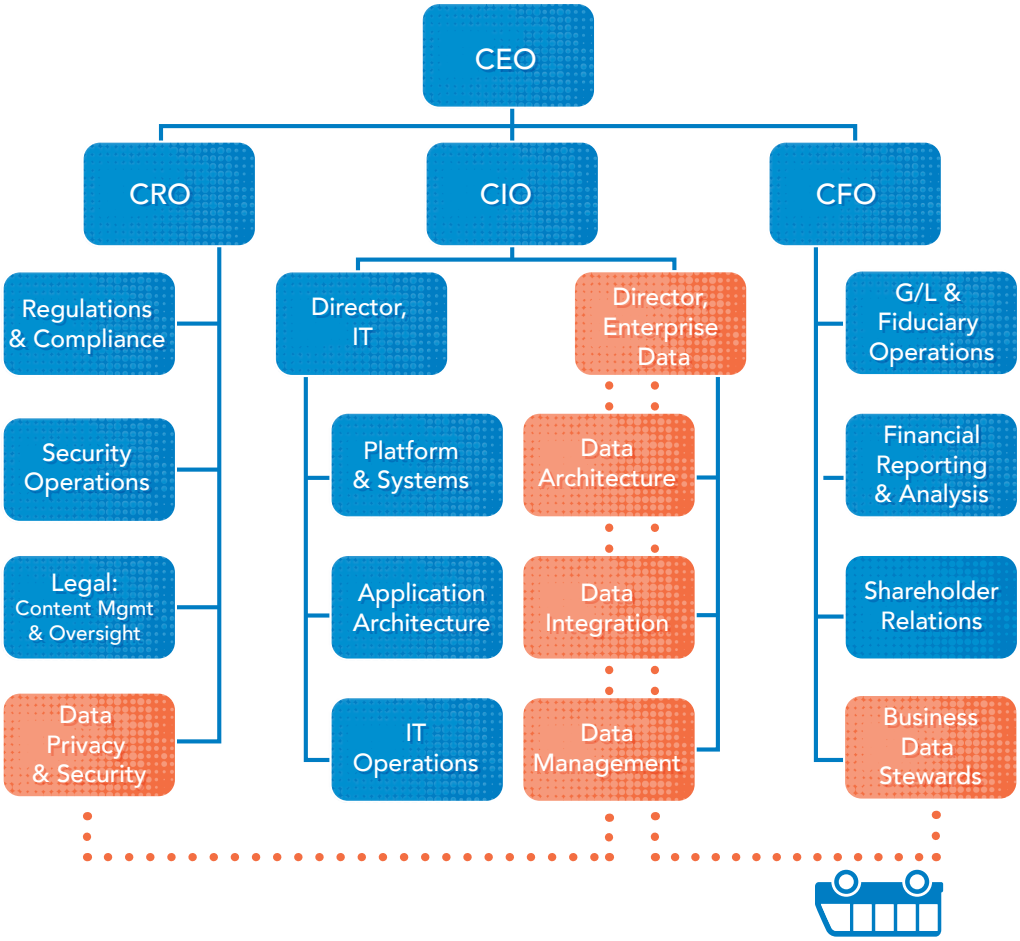

figure 2: integrated organizational chart

Each line of business has a position responsible for the execution of the company-wide data strategy. These vertical data authorities serve as extensions of the enterprise data team that's collectively responsible for the company's data strategy, including design and architecture, process and policy, and administration and enforcement. The vertical authorities help to ensure data is shared and managed proactively.

The leader of the enterprise data team sits in the upper layers of the hierarchy, typically reporting to a C-level executive, and serves as the consistent voice of data, responsible for effectively communicating data's role in all business initiatives.

This administrative model is necessary when the business is not ready (for whatever reason) for full executive engagement in enterprise data practices. The board may not authorize adding an additional C-level position; the culture may operate more effectively with less hierarchy; or the company may need the spread of support because of geographical distribution.

There could be a myriad of reasons why the integrated model is the most appropriate structure for the data organization. The centralized oversight with autonomous execution provides flexibility in administration when business departments have varying levels of maturity in their data practices. While the focused direction and strategy drive the core business initiatives, this model requires strong, consistent communication across all lines of business and with the centralized team to ensure the data responsibilities spread companywide are uniformly upheld.

STYLE 3: CENTRALIZED

When the executive view shifts from considering data as a secondary asset and business enabler to a primary asset that generates value, the company is ready for a data executive with a seat at the boardroom table.

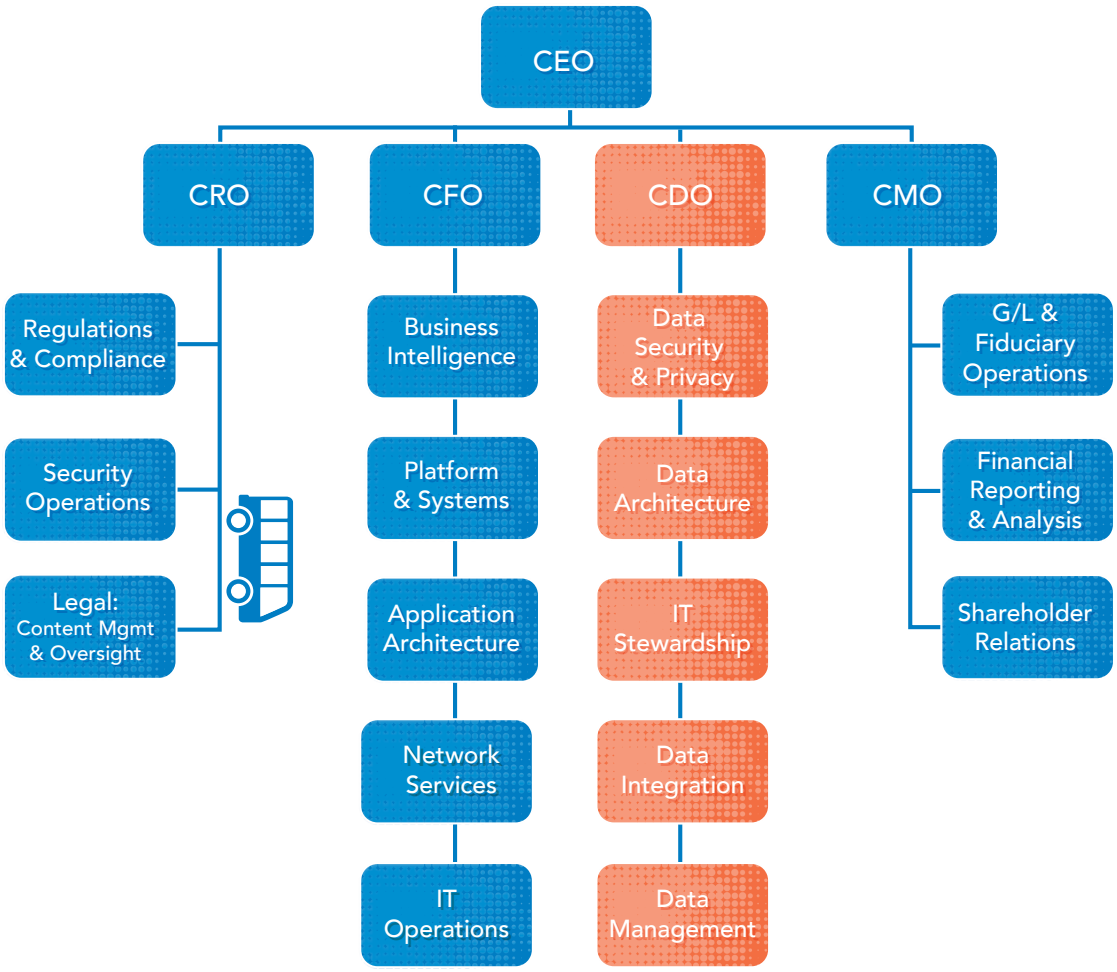

figure 3:centralized organizational chart

The organizational structure shifts, with data represented by its own line of business, as leaders recognize how data plays a key role in competition, operational excellence and innovation. Instead of asking how data can support the business, executives will ask:

What is our data capable of?
Where can our data take us next?
How can our data enable competitive advantage?

Other executives won't be concerned about getting replaced or losing responsibilities because they recognize a chief data officer (CDO) is bringing a new set of capabilities to the table that will help advance the business.

With a centralized style of data administration, all data practices are moved under the umbrella of a dedicated data division responsible and accountable for the strategy and execution of all enterprise data activities and functions. This model requires a paradigm shift and commitment from the entire company. While data practices may exist throughout the business, they will conform to, align with and are ultimately controlled by the strategy and policies managed solely by the centralized enterprise data organization.

The centralized administrative model provides constant executive engagement in the role data plays for the entire company. There is an understood importance of data inherent to the culture, and the role of the centralized data division is not only sanctioned but also highly valued.

Company Culture and Dynamics

A company's habits evolve very slowly over time as new behaviors emerge, but the habits only shift after these behaviors are repeated. Establishing repeatable behaviors is what makes change management so difficult.

As you look to start up your big data bus, identify the cultural behaviors and business practices you will need to be successful. Perform a thorough, brutally honest survey of your current culture. If there are significant gaps in the identified behaviors, build in time and tolerance to shape the culture that is most conducive to supporting your big data efforts.

While there is no right or wrong company culture, having a clear definition of what the business believes, the general rules of engagement and the commonly accepted practices will help to inform the ideal framework for the *Why* of your business. You will not only know who needs to be on your bus, but more importantly, *why* they have a seat.

The centralized administrative model provides constant executive engagement in the role data plays for the entire company.

Your big data team should include both internal and external data practitioners.

BUILDING

the big data team

In the previous section, we took a deep dive into the *Why* of the data organization and addressed these two key questions:

***Why** is your business interested in big data and why now?*
***Who** will make your big data vision and mission successful?*

In this section, we're going to shift gears and get into the tactical *How* – the building of the big data team. In so doing, we'll tackle these important questions:

***How** will you achieve your initiative?*
***Who** will make your projects successful?*

We'll first take a look at *Who* makes up the big data team, then *How* to put the team together, along with a case study.

4 Characteristics of a Big Data Team

When building your big data team, it's important to keep these four team characteristics in mind:

1. It doesn't have to be an inside job.
2. It will cross the functional divide.
3. It's not static.
4. The team sponsor will change.

Let's look at each characteristic in more detail.

1: IT DOESN'T HAVE TO BE AN INSIDE JOB

Your big data team should include both internal and external data practitioners.

In the CA Technologies study cited earlier, we learn two of the top three major investments companies need to make to ensure a successful big data initiative are in training existing resources (57 percent) and hiring new resources with the required skills (47 percent). The third investment is in infrastructure, due to the lack of resources available to manage the growing infrastructure complexity.

Most companies turn to outsourcing to offload some of this struggle with their talent shortages. In fact, Accenture found 95 percent of its survey respondents reported using one or more types of external help with their big data initiatives. And of the companies studied by CA Technologies, 40 percent invested in cloud/hosted infrastructure services, and 22 percent hired big data consulting firms.

The good news is that when it comes to staffing your big data team, you don't have to do it all yourself – nor should you. Why? Because big data technologies and capabilities are still maturing and evolving at a rapid pace. It's next to impossible these days to stay on top of everything going on in the big data space. Thus, it's more cost-effective and efficient to hire external contractors and service providers whose primary mission is to know big data and its technologies inside and out (so you don't have to). When building your big data team, select a healthy mix of team players from inside and outside your organization, namely:

- **Employees.** This includes current employees, as well as employees you're looking to hire.
- **Contractors/freelancers**. A recent report from the Freelancers Union found 53 million people – or 34 percent – in the US workforce are freelancers. This is roughly one in three Americans. They're predicting that by 2020, up to half of the workforce will be freelancers or contract workers.
- **Service providers.** This could include big data solution providers (like Cloudera and Hortonworks), cloud providers or any other provider offering X as a service (XaaS).

2: IT WILL CROSS THE FUNCTIONAL DIVIDE

Your big data team should also have a healthy mix of business and technical practitioners.

It's easy to think big data is an IT initiative, but it's not – and many businesses are learning this the hard way. You don't have to look hard to find stories about companies that have jumped on the big data bandwagon, with IT declaring success on one hand, and the business still scratching its head on the other. There's no question that implementing big data technologies is fraught with technical challenges, but without the business fully engaged and driving the requirements, your big data initiative is doomed to fail.

It's more cost-effective and efficient to hire external contractors and service providers whose primary mission is to know big data.

A well-balanced big data team will include business and technical practitioners from these six functional areas:

figure 4: the big data team dynamic

Let's explore each one.

Platform

The Platform functional area involves configuring and installing the hardware and software infrastructure that will support big data.

This may include the acquisition and setup of servers, the repurposing of existing hardware, installation of key software, configuration of storage environments, the administration of software versions and access, and maintenance responsibilities. Consider this the set of skills required to set up, maintain and enhance the big data operational environment.

SAMPLE ROLES

PLATFORM

- BI infrastructure architect
- Data warehouse appliance specialist
- Database administrator
- Database architect
- Database manager
- Hadoop architect
- Systems architect
- Security and archival specialist
- Storage engineer

Development

The Development functional area encompasses making the big data environment work.

Development activities can include – but aren't limited to – loading data into Hadoop, customizing open source software projects, writing new code, designing archival strategies within and outside the big data environment, customizing analytics software solutions, performance tuning, and working with the Platform team to help acquire additional functionality.

SAMPLE ROLES

- BI solutions developer
- Big data loading specialist
- Database developer
- ETL architect
- ETL developer
- Hadoop developer
- Hadoop engineer
- Java developer
- OLAP developer
- Predictive analytics developer
- Visualization tool developers

DEVELOPMENT

Data Specialists

The rise of the Data Specialist has occurred as Platform and Development teams confront the truth that corporate data is more heterogeneous than ever.

The data originates from sources both inside and outside the company's firewall, accompanied by a dizzying checklist of formats and business rules. Rather than straying from their core skill sets to try and understand this data diversity, Platform and Development teams are increasingly turning to Data Specialists, many of whom have worked on the business side.

For the past decade, more of your company's employees have spent time finding, correcting, loading, explaining, defining and sharing data. Many of these people are business users (see next section), and then there are those who have assumed or created new roles to support the evolving need to manage data or data-enabling technologies.

As data moves across organizations and business processes, diverse skill sets are necessary to not only ensure the data reaches its destination – be it user or machine – but that it is meaningful and consumable. Questions about data definitions, lineage, formats and usage have increased in lockstep with the business's demand for new information.

A quick note about the data scientist: Stop looking for unicorns. Because you're more likely to discover a mythical creature before you find a "data scientist" – a quant who can find that insightful needle in your data haystack while delivering a charismatic TED talk.

Remember: You're not looking for *the* data scientist. You're looking for Data Specialists – a collection of talented individuals who can collect data, analyze it, interpret the results and recommend actions. Call them data scientists if you want. These people should be able to support a variety of big data initiatives. One size does not fit all.

SAMPLE ROLES

SPECIALISTS

- Analytics practice leader
- BI analyst
- BI solution architect
- Big data researcher
- Big data solution architect
- Big data visualization specialist
- Business data analyst
- Data administrator
- Data analyst
- Data architect
- Data hygienist
- Data miner/statistician
- Data modeler
- Data scientist
- Information architect
- Machine learning expert
- Metadata administrator

Business Stakeholders and Users

This functional area encompasses businesspeople who must find, access, decipher, use, share and deploy data as part of their jobs.

Like Data Specialists, they have learned data the hard way, that is, by using it. And as business professionals, they are often consulted to help approve big data decisions, validate business rules, and identify peers who can affirm evolving big data and usage requirements and policies.

For instance, risk managers might be rewarded for reducing fraud, thus giving them a unique perspective on the type of data, its authoritative sources and the data volumes that best fit their usage needs. Indeed, the sample list is a small and often functionally specific list that is merely representative of some of the more common business stakeholder roles and by no means exhaustive.

SAMPLE ROLES

- Actuary or underwriter
- Branch manager
- Call center representative
- Data analyst (financial, marketing or risk)
- Data explorer/discovery analyst
- Marketing product manager
- Merchandiser
- Project manager
- Strategy or acquisitions analyst

STAKEHOLDERS

Governance and Policy

The Governance and Policy functional area involves senior leaders who advise other functions on how data should or shouldn't be used, according to larger corporate cultural issues, regulations and fiduciary rules.

This might involve a loose set of executive advisers, each of whom may be asked to "sign off" on a data decision before the data is provisioned to businesspeople. Or it might involve a more formal data governance council, comprising executives who may hand down sharing and security decisions about key data – for instance, usage policies for PCI (personal credit information) data. Over time, this type of data decision making might evolve into a formal process for project prioritization, hiring talent and/or breaking a tie.

SAMPLE ROLES

- Business data steward
- Chief data officer
- Customer data steward
- Data quality analyst
- Data quality manager
- Data security manager
- Risk manager
- Source system data steward

GOVERNANCE

Executives

The Executives functional area is made up of decision makers who rely on data for strategic and operational decisions, but might or might not be active users.

At a high level, executive managers might use the product of predictive models to report financial forecasts, and rely on accurate and up-to-date information in this high-visibility role. But there are far more business practitioners who rely on good, meaningful data to do their daily work, often not even aware that their dashboard reports, performance scorecards and financial measures were all freshly generated in a Hadoop environment right before they received their reports.

These individuals might not know how important they are to the big data team. But without them, big data might never have existed in the first place.

SAMPLE ROLES

EXECUTIVES

- Chief digital officer
- Chief executive officer
- Chief financial officer
- Chief information security officer
- Chief marketing officer
- Chief privacy officer
- Chief risk officer

3: IT'S NOT STATIC

Your big data team will continue to change due to a rapidly evolving big data industry.

Executives often wonder aloud whether they need to construct a brand-new, separate big data team, independent of existing BI or analytics teams. The answer will often hinge on the effectiveness and reputation of the latter team. However in most instances, the inviting existing teams to board the big data bus might be the answer.

In the previous section, we listed several sample functional roles. You may have recognized that many of these roles have been around for decades; in fact, many probably already exist in your company. But big data has also introduced some new roles to our traditional table, including:

- Big data loading specialist
- Big data researcher

- Big data solution architect
- Big data visualization specialist
- Chief data officer
- Chief digital officer
- Data explorer/discovery analyst
- Data scientist
- Hadoop architect
- Hadoop developer
- Hadoop engineer
- Predictive analytics developer

In addition to these new functional roles, big data is also requiring new and/or enhanced technical and soft skills. For example:

- **TECHNICAL SKILLS**
  - Data science, including mathematics, statistical analysis, statistical programming, analytics modeling techniques, knowledge of data subject matter and the ability to experiment with data without fear.
  - Data design for handling larger volumes of data.
  - New software frameworks, such as Hadoop, NoSQL and HBase.
  - Analytics programming languages.
- **SOFT SKILLS**
  - Ability to understand business terminology and processes across the company.
  - Understand corporate strategy and the accompanying KPIs.
  - Know which business questions are necessary to enable this strategy.
  - Ability to measure and communicate results.

These roles and skills lists are not exhaustive, but they do illustrate how big data is changing company dynamics. Big data technologies are evolving rapidly, and still a long way from maturity. Until these technologies start to stabilize, we can expect the roles and skills needed to support and effectively use them will continue to evolve.

Until technologies start to stabilize, we can expect the roles and skills needed to support and effectively use them will continue to evolve.

Your business may identify a core set of static roles that gets included in every big data initiative, such as a data scientist or Hadoop architect, but be open to the possibility that this rapidly evolving big data industry may warrant modifications to this core team periodically.

4: THE TEAM SPONSOR WILL CHANGE

Every big data strategic initiative needs to have one or more executive sponsors. The scope of your initiative will inform who the best executive sponsor(s) is for the big data team.

An executive sponsor serves as the corporate figurehead for the initiative. This sponsor will be responsible for: making the pitch, communicating the value and managing expectations with executive management; managing the decision-making and tie-breaking processes for the duration of the initiative; and potentially funding the initiative from the sponsor's own business line.

If your company is on (or has started down) the big data path, it's highly likely you have multiple strategic initiatives you'd like to support. Don't assume the CIO, CMO or CDO, if you have one, should be the executive sponsor for every big data initiative. And don't assume a single initiative requires only one sponsor; an initiative could have multiple sponsors. What you can assume, however, is the sponsor will most likely change with each initiative, depending on the deliverable.

Every big data strategic initiative needs to have one or more executive sponsors.

And don't assume a single big data initiative will involve only one sponsor. After all, business models evolve, business processes become more refined, and – in the brave new world of open source - technology projects are unveiled almost daily. Thus the an initial sponsor of a big data effort might cede to another as the big data initiative evolves, capabilities mature and data is shared more broadly. Different sponsors for different big data initiatives has become the norm at most early-adopter companies.

The table below helps illustrate this point. Several popular big data initiatives are listed. For each initiative, one or more candidate sponsors are indicated:

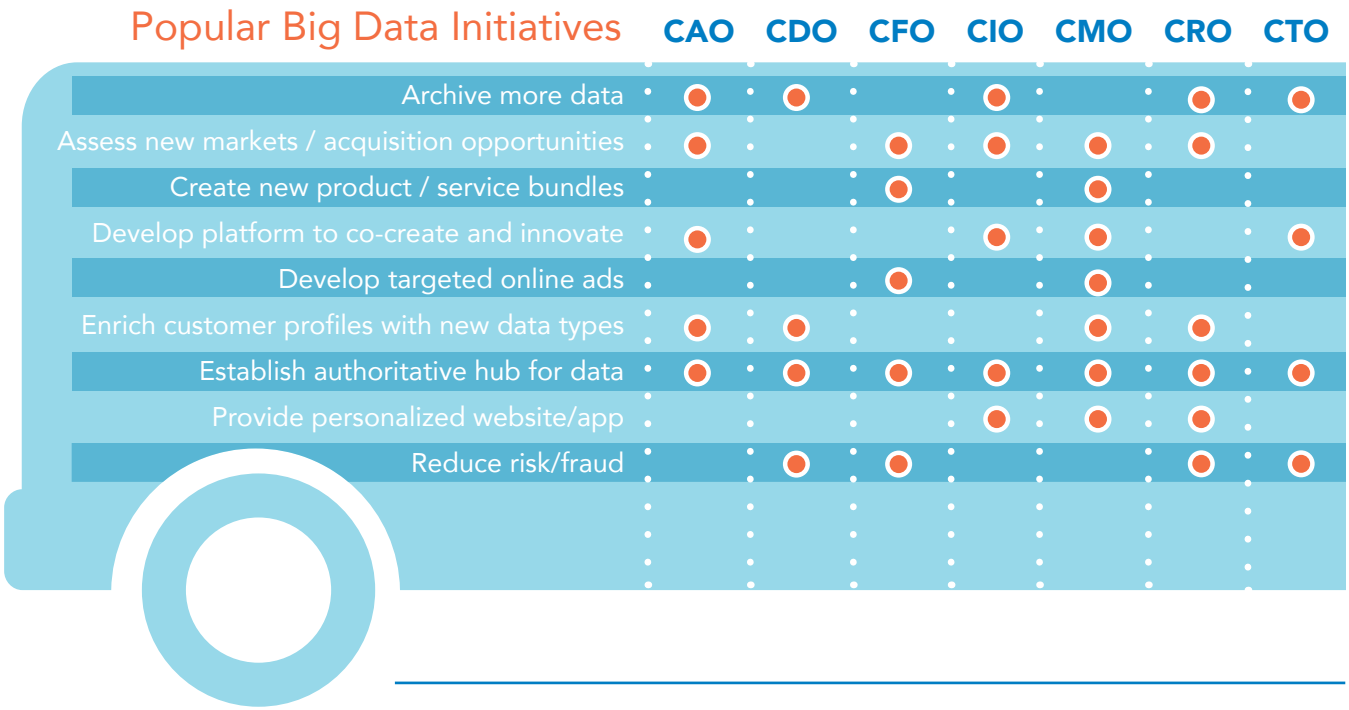

figure 5: big data initiative team sponsors

If we use “develop platform to co-create and innovate” as an example, we have four possible sponsors marked: CAO, CIO, CMO and CTO. Let’s assume the CTO is responsible for all corporate platforms, and innovation falls under the CIO’s mandate. In this case, the CTO and CIO can agree to be co-sponsors, or they can decide on one sponsor for the initiative. There’s no right or wrong choice here; it’s just a matter of what the business needs and can support at the time of the initiative.

The Team Composition at a Glance

It should be clear at this point that the task of pulling together a big data team is not a simple, once-and-done task. This team is dynamic, and it needs to be built one initiative at a time. The initiative’s scope and objectives will inform who should be on the big data team each and every time.

This diagram recaps what we've discussed so far:

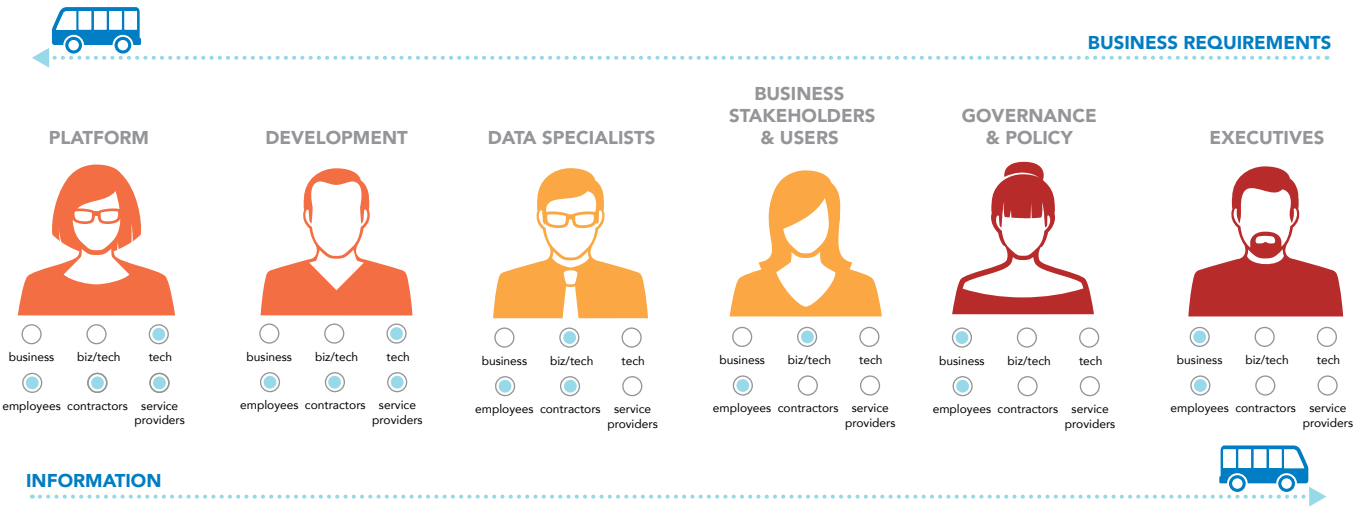

figure 6: the big data team composition

A few high-level observations to note about these six functional areas:

- If you look at the functional areas on a sliding scale from technical to business, Platform is the most technologically focused, Development is next, and then it flows through to the Executives, which is the most business-focused.
- Business requirements typically begin with Executives and flow through each functional area to the Development and Platform areas. In return, the Platform and Development areas deliver the data that gets translated into useful information for the business.
- The technical areas, Platform and Development, use resources from all three sources – employees, contractors/freelancers and service providers – whereby the remaining four functional areas (Data Specialists, Business Stakeholders and Users, Governance and Policy, and Executives) should only be staffed by employees. This is because they know best when it comes to your business, the culture and its data.

Now that we know *Who* needs to be on the big data bus, let's discuss *How* to put the team into action.

The Big Data Road Map

In this section, we'll take a look at where building a big data team fits into the overall big data road map, and then look at a case study that ties it all together.

Here's a suggested nine-step road map for companies just getting started with big data, or for those that have got off to a rocky start and are looking for some direction. Note where "build the team" fits in.

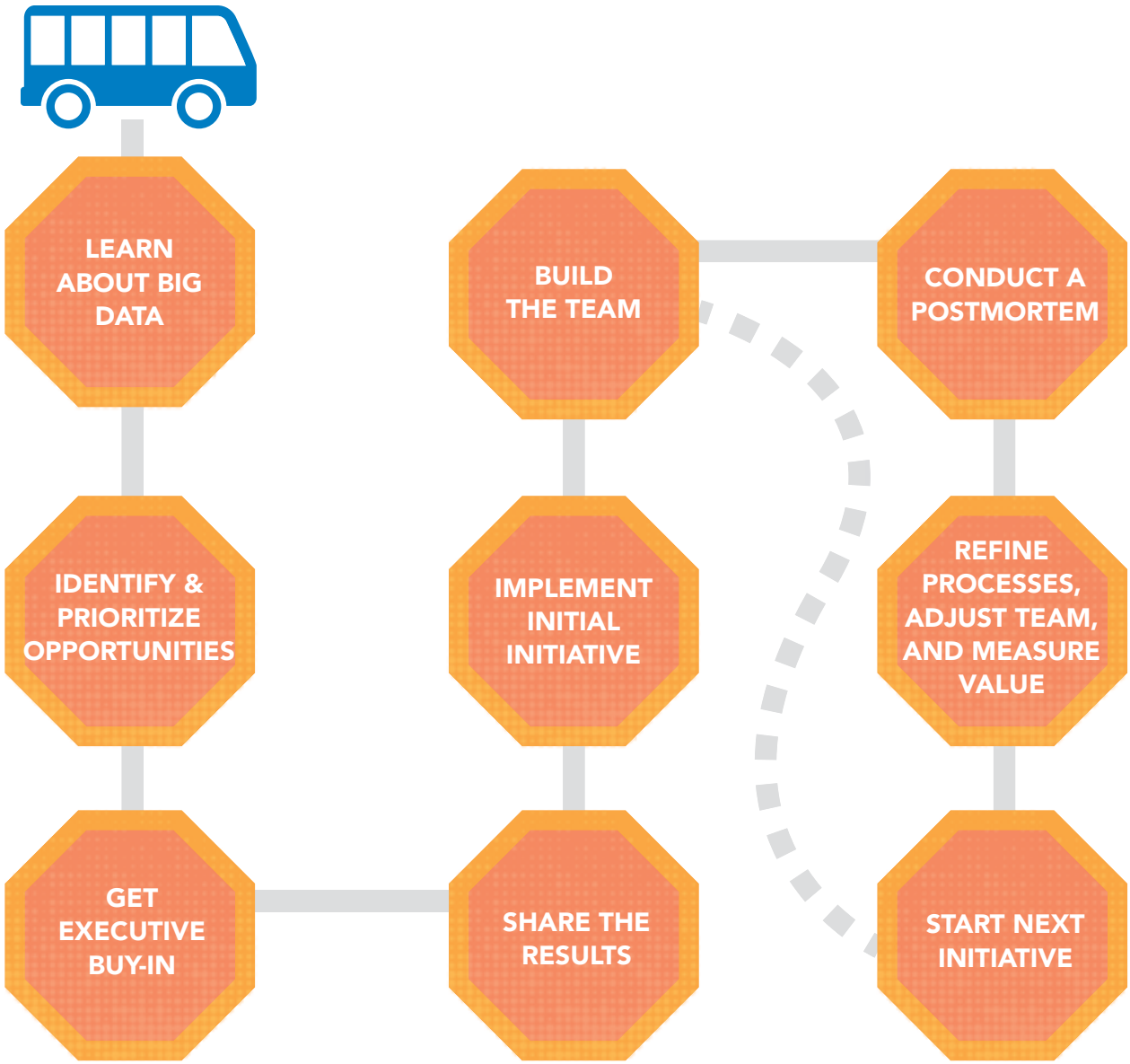

figure 7: the big data road map

Learn about Big Data

Do your homework. Read what others are doing with big data. Attend conferences. Talk to others who've both succeeded and failed in their big data initiatives. Research what your competition is doing with big data. Above all, ensure that key players across big data initiatives share a common vocabulary.

Identify and Prioritize Opportunities

As your understanding of big data develops, you begin to see big data opportunities everywhere – inside and outside of your organization. Focus your attention on those opportunities that will provide value to your business. Then systematically prioritize these opportunities, taking into consideration business value, size of effort and technical difficulty.

Get Executive Buy-in

Many big data strategic initiatives fail because there's no executive buy-in. Despite the outcome of a "skunkworks" initiative, if the initiative doesn't align with executive priorities, its survival may be compromised. Understand how business executives are measured. Is it on revenue generation? Strategy support? Cost savings? Return on investment? The best big data teams are those who stay mindful of competing priorities, and have a prioritized list of big data opportunities at the ready to share with potential sponsors at the opportune time.

If your big data initiative doesn't align with executive priorities, its survival may be compromised.

Build the Team

Once executives have approved an initial effort, it's time to build a multidisciplinary team. This team may include the following: an executive sponsor, business subject matter experts, IT practitioners and external resources, as needed. The size and scope of the initiative will help inform how many should participate.

Implement Initial Initiative

Make sure the selected effort is well-scoped. Start small. You're looking for a quick win to prove that big data is good for the business. And if the initiative is going to fail, you want to fail fast so that you can move on.

Share the Results

Enlist interested parties in reporting the results of your effort. This is true whether the initiative has succeeded or failed. The accountability means credibility, and you will most likely garner more support along the way.

Conduct a Postmortem

At the end of the initiative, gather the team to conduct a postmortem. Identify what worked well, what didn't work and what you've learned. Identify opportunities of improvement for future initiatives. If you need one, use an outside facilitator familiar with big data, its platforms and terminology, and its business applications.

Refine Processes, Adjust Team and Measure Value

Based on feedback from the postmortem and management team, refine your business and technical processes, then adjust the big data team accordingly for the next initiative. Depending on the type of initiative, start measuring the anticipated value.

Start Next Initiative

Now with one big data initiative under your belt, start the next initiative with your newly selected big data team, and repeat the cycle of sharing results, refining processes and measuring value.

Building your first big data team will take the most time, consideration and negotiation prowess. Note that the last four stages on the road map are cyclical. Each cycle represents a different initiative and a new team to make it happen. This road map further demonstrates the dynamic nature of the big data team and how each initiative informs who will be on it.

Building your first big data team will take the most time, consideration and negotiation prowess.

Case Study: Thompson's Market Inc.

To illustrate key concepts presented in this white paper, let's take a brief look at how one company, Thompson's Market, used these principles to build its big data team.

This case study is a spinoff from *The Big Data E-Book: A Tale of Customer Loyalty and Going SoLoMo*. This e-book tells the story of a struggling marketing executive at Thompson's Market, one of California's favorite (fictional) grocery chains. The new CEO wants to up Thompson's game and better meet its customers where they're at. She knows it's going to take big data – of the social, location and mobile kind – to get Thompson's moving in the right direction.

As Thompson's began its big data journey, it identified dozens of strategic initiatives to pursue. The management team then decided to prioritize these initiatives using a scoring system. The team scored each initiative on:

- **Business value.** How much value does this initiative bring to the business?
- **Size/scope.** What is the size/scope of the initiative?
- **Technical difficulty.** What is the technical difficulty of implementing this initiative?

The ideal initiative was one which had high business value, was smaller in scope and wasn't too technically challenging. The team wanted quick wins to show it was on the right track with big data.

The ideal initiative was one which had high business value, was smaller in scope and wasn't too technically challenging.

Thompson's prioritization exercise revealed which initiatives to pursue first. The management team decided to tackle the first eight initiatives over the course of two years. They learned early on that a big data team is a lot like a sports team. It has several players, but not everyone will be in the game at the same time. And just as team members can participate in one or more initiatives, they can also serve in multiple roles within the same initiative. The "Surprise Checkout Discounts" initiative helps demonstrate this concept.

The idea behind this initiative was that all shoppers should receive discounts without having to carry around coupons or being members of the loyalty card program. When shoppers checked out, they would automatically receive a discount or get one of the items for free. The discount would be anywhere from $1 off to 20 percent off.

To accomplish the Surprise Checkout Discount initiative, the Thompson's strategic team broke the initiative down into four objectives that had specific measurable outcomes and progressively worked toward successfully completing the overall initiative. The team identified six key functional roles that had to be filled by ten team members

across the initiative. Some of the functional roles, such as Data Specialists, were more prominent throughout all of the objectives, whereas other roles, such as Development, did not require as much involvement across the objectives. This helped the team understand the balance of roles required, as well as who fulfilled the requirements in each objective.

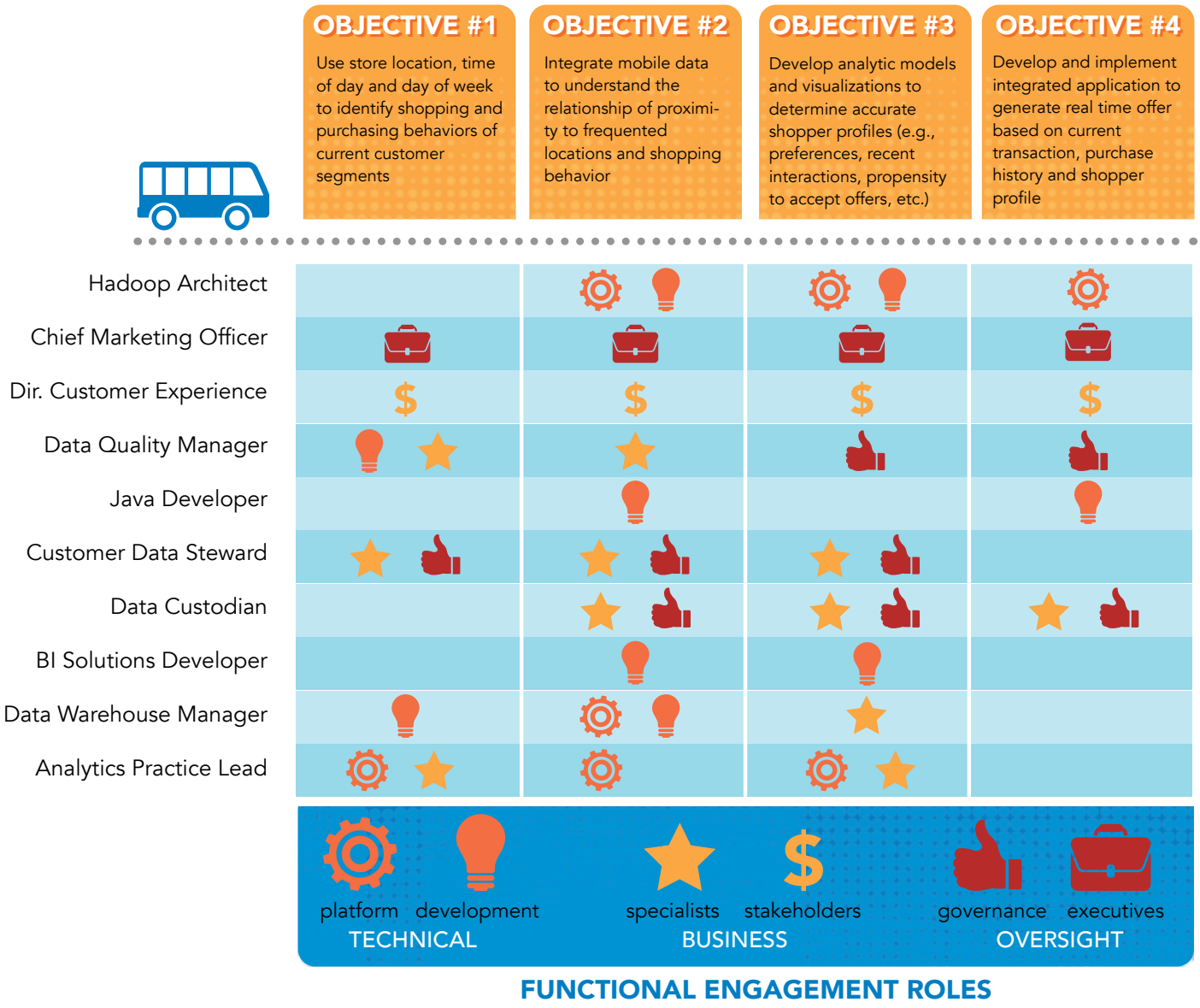

figure 8: the big data team matrix

Let's look at the Data Quality Manager (DQM), for example. This is a position selected from the organizational hierarchy because of its required data quality expertise, specifically for the data sources that will be contributing to the first objective in the Surprise Checkout Discount initiative. This position also sits on the data governance council, and therefore brings a great deal of knowledge and expertise to the team about data guiding principles, policies and procedures.

The functional roles assigned to the DQM for the first objective are development and data specialist. The first objective is SQL query intensive and has extensive data quality requirements. The DQM has the knowledge and skills that best meet the needs for this objective. The development requirements of Objective 2 overlap with the skills of others on the team so the DQM is not assigned the development role for this objective. However, the DQM will continue to serve as a data specialist to address the data requirements of Objective 2. And because of the DQM's role on the data governance council, the functional role of data governance has been assigned for Objectives 3 and 4.

Big data teams allow companies to maximize skills and capabilities across the business without requiring organizational restructuring.

CONCLUSION

In *Good to Great*, Jim Collins writes, "Great vision with mediocre people still produces mediocre results." With big data and its supporting technologies blasting the door of opportunity wide open, we are further empowered to reconsider the possibilities and realities of becoming a data-driven business. But to make our dreams and vision a reality, we will need great people – in fact, a very large bus of them. Because without them, nothing else matters.

With data volumes increasing exponentially, we literally can't afford to sit on the sidelines to wait and see how it all plays out anymore. We need to start building our big data teams now: from inside and outside the organization; from executives to platform administrators; and from business practitioners to data scientists to Hadoop developers.

Getting the right people on your big data bus is not a once-and-done academic exercise. Look at it as a disciplined process of keeping this team active and fresh for each and every big data initiative your company pursues. Not only will this be good for your employees, but also healthy for your business's bottom line.

So the question is now yours: Who's getting on *your* big data bus?

Getting the right people on your big data bus is a disciplined process of keeping this team dynamic and fresh for each and every big data initiative.

about the authors

ANNE BUFF is a Business Solutions Manager and Thought Leader for SAS Best Practices, a thought leadership organization at SAS institute. As a speaker and author she specializes in the topics of data governance, MDM, data integration and data monetization. Her energy, excitement and enthusiasm for all things data will grab you in ways you never expected, leaving you refreshed and ready to take on your toughest of data issues.

TAMARA DULL is the Director of Emerging Technologies for SAS Best Practices, a thought leadership organization at SAS Institute. Through engaging publications, rich media and industry engagements, she delivers a pragmatic perspective on topics including big data, Hadoop, open source and the Internet of Things.

Tamara regularly writes and blogs for SAS, Smart Data Collective and *The Cyberista Says* at tamaradull.com. Two popular publications include *The Big Data E-book: A Tale of Customer Loyalty and Going SoLoMo* and *A Non-Geek's Big Data Playbook: Hadoop and the Enterprise Data Warehouse*.

sas best practices
THOUGHT PROVOKING BUSINESS
SAS Institute Inc.
100 SAS Campus Drive
Cary, NC 27513-2414
USA
Phone: 919-677-8000
Fax: 919-677-4444