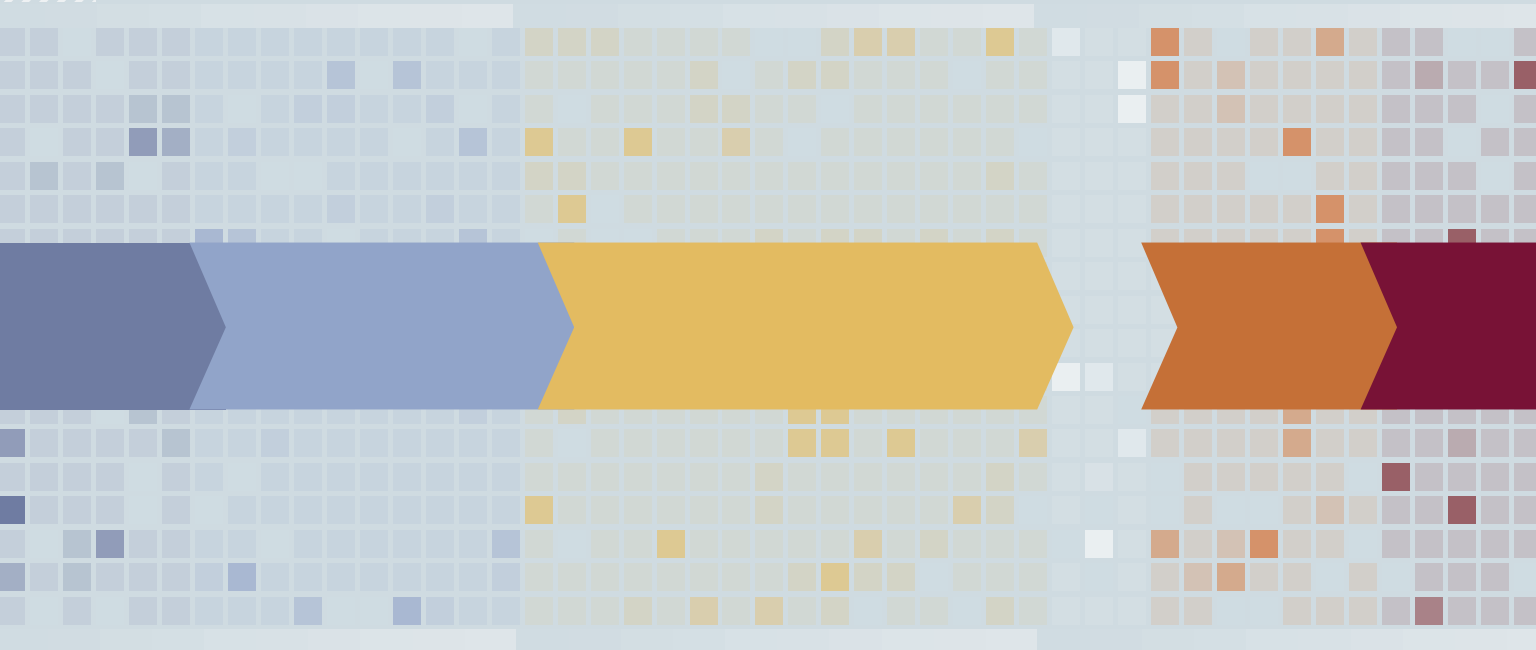
TDWI RESEARCH

TDWI BENCHMARK GUIDE 2013–2014

TDWI BIG DATA MATURITY MODEL GUIDE

Interpreting Your Assessment Score

By Fern Halper and Krish Krishnan



tdwi.org

tdwi

Sponsors

Cloudera

IBM

MarkLogic

Pentaho

TDWI BIG DATA MATURITY MODEL GUIDE
Interpreting Your Assessment Score

By Fern Halper and Krish Krishnan

Table of Contents

About the Authors

Fern Halper, Ph.D., is director of TDWI Research for advanced analytics, focusing on predictive analytics, social media analysis, text analytics, cloud computing, and other "big data" analytics approaches. She has more than 20 years of experience in data and business analysis, and has published numerous articles on data mining and information technology. Halper is co-author of "Dummies" books on cloud computing, hybrid cloud, service-oriented architecture, service management, and big data. She has been a partner at industry analyst firm Hurwitz & Associates and a lead analyst for Bell Laboratories. Her Ph.D. is from Texas A&M University. You can reach her at fhalper@tdwi.org, or follow her on Twitter: @fhalper.

Krish Krishnan is CEO for Sixth Sense Advisors, Inc. He is a recognized industry thought leader and practitioner in the area of data warehousing and big data. A sought-after speaker at various industry conferences, Krishnan's expertise spans all areas of big data, business intelligence, and data warehousing. Krishnan specializes in providing high-performance solutions for big data, artificial intelligence, and BI/DW initiatives. He has designed and implemented data warehouse solutions that range from hundreds of gigabytes to multi-petabyte data volumes. Krishnan leads the data warehouse appliance expert channel at the BeyeNETWORK. He has written two books on big data and is finishing a new book on social analytics. He has also authored 200 case studies and white papers on topics surrounding the subjects of data. You can reach him at rkrish1124@yahoo.com.

About TDWI Research

TDWI Research provides research and advice for business intelligence and data warehousing professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of business intelligence and data warehousing solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

Sponsors

Cloudera, IBM, MarkLogic, and Pentaho sponsored the research for the TDWI Big Data Maturity Model, online assessment tool, and this guide.

Foreword from the Authors

Reason for the Model

Today, many organizations are collecting increasing amounts of disparate data. In fact, they are collecting more than they can manage or analyze; yet, they realize that data and data analysis can provide an important strategic competitive advantage. According to TDWI Research, a new group of user organizations is currently commencing or expanding solutions for analytics with big data. We created the TDWI Big Data Maturity Model and assessment tool in response to requests from organizations to understand how their big data deployments compare to those of their peers in order to provide best-class insight and support. The assessment measures the maturity of a big data program in an objective way across various dimensions that are key to deriving value from big data. Based on our experience, data becomes useful when it can be analyzed, so big data analytics is a primary focus of the model.

Value of the Model

The model can help guide IT and business professionals on their big data journey. It provides a framework for companies to understand where they are, where they've been, and where they still need to go on their big data deployments. The model can also provide guidance for companies at the beginning of their big data journey by helping them understand best practices used by companies that are more mature in their deployments. Of course, most companies are at the beginning of their big data journey, and so there is a great deal of confusion in the market. Therefore, the model also helps organizations recognize that they are not alone in their big data challenges.

A great feature of the Big Data Maturity Model is the interactive benchmark assessment. After taking the benchmark survey, you will be able to quantify the maturity of your deployment in an objective way, understand your progress, and identify what it will take to get to the next level of maturity. This benchmark guide is designed to help you understand the phases of maturity in big data as well as help you interpret your benchmarking scores. We hope you find it useful.

Fern Halper, Director for Advanced Analytics, TDWI Research

Krish Krishnan, Founder, Sixth Sense Advisors

Introduction

Big data is not a single technology, but a combination of old and new technologies that help companies gain actionable insight while effectively managing data load and storage problems. Big data analytics requires the ability to collect, manage, and analyze potentially huge volumes of disparate data, at the right speed, within the right time frame, while providing the right-time analysis and activity to the end consumer. Big data includes structured databases, unstructured data from various internal and external sources, streaming data, social media, geospatial data, and so on. To successfully leverage these various kinds of data requires infrastructure, data, analytics, organizational structure, and governance processes to make big data analytics operational and actionable.

Many companies begin to utilize big data solutions because they have a specific business problem to solve that requires big data. For instance, an Internet company might monetize big data as part of its business model. This is often a new breed of company that is very fast and aggressive, where data is driving business. Others, often because of a business imperative such as losing market share, have begun to realize that data is a corporate asset they can use to become more competitive. Still others are simply outgrowing their current data infrastructures.

What's Different about Big Data Analytics?

Big data analytics is not necessarily new. Some companies have been dealing with big data for years and performing some pretty sophisticated analytics. This is true for the Internet giants as well as the early adopters in retail, financial services, and telecommunications. What's different as big data starts to enter the mainstream is that organizations are using it to both extend what they have been doing in analytics as well as use it for new forms of analysis. For instance, maybe a model for churn took 20 hours to run before a big data infrastructure was in place. Now it takes less than 10 minutes. An automobile manufacturer might utilize sensor data from its autos and combine that with other data for preventive maintenance. That's taking a new form of data (machine sensor data) and running it across new infrastructure, combining the outcomes with existing analytics, and creating more opportunities for everyone in the process.

Other analytics differences include:

- **More data and attributes.** With big data technologies, you can utilize many more attributes in an advanced model because you now have the data and the associated computing power to do so. In addition, you can address a wider set of data and potentially use that data in unforeseen ways.
- **Automatic and programmatic.** Because so much data or streaming data is involved in big data analysis, it is often important to make use of code for analysis rather than doing it manually. Stock trading is one example of this. Online advertising is a newer example. A coupon you receive on your smartphone that makes use of your GPS data and your past behavior is another.
- **Hypothesis free.** This is sometimes called data-driven analysis. With a hypothesis-driven approach, you first define the problem—for example, I believe that customers are acting this way because of reason X. In a hypothesis-free environment, you set your algorithms to work over the data and discover patterns. Of course, it makes sense to have some problems identified in the first place, but this is still considered by some to be a different approach to analysis.

Even if you have one big data implementation at work in your company—such as analyzing telecommunications network traffic—it doesn't mean that your organization is mature in its big data analytics efforts. Likewise, you may work for an Internet company that is pretty advanced in bringing various data sources together using new technology. However, that doesn't mean that your analytics or governance processes are mature. Maturity involves a combination of people, processes, and technology across a series of dimensions.

An Evolving Ecosystem

There has been a flurry of activity from vendors offering infrastructure and analytics tools to support big data and big data analytics. This includes new forms of data warehouses and other data stores, new ways to integrate data, new ways to query it, and new ways to analyze it. This involves technologies such as analytic platforms, Hadoop, NoSQL databases, data virtualization, in-memory analytics, and refactored (i.e., changing the internal code) analytics algorithms to move beyond desktop analytics to include massively parallel processing capabilities. In fact, an ecosystem is growing around big data and big data analytics.

This ecosystem is developing as companies try to wrap their arms around big data. For example, in a recent TDWI Best Practices Report on big data management, author Philip Russom notes that only 10% of survey respondents in the study reported having deployed a special solution for managing big data today.[1] Most of these are very new (7%), whereas only a few are relatively mature (3%). Another 10% of respondents said they have a big data management solution in development. If users' plans pan out, Russom notes, dedicated big data management solutions will jump from rare to mainstream within three years.

On the analytics front, technologies for advanced analytics, advanced data visualization, and text mining show strong potential for growth as organizations build out their big data analytics programs. Newer technologies such as stream mining are on the radars of more companies that seek to analyze data with millisecond response times. Geospatial analytics is becoming extremely important for a variety of use cases.

It is an exciting and a challenging time for organizations as they consider big data opportunities.

TDWI Big Data Maturity Model: The Context for Benchmark Scores

Big data maturity can be described as the evolution of an organization to integrate, manage, and leverage all relevant internal and external data sources. It means creating an innovative ecosystem, delivering insightful business value, and enabling impactful transformations. In other words, big data maturity is not simply about having some technology in place to deal with high volumes of data. Nor is it simply about using social media to analyze buzz about your brand. It is a journey that involves building an ecosystem that includes technologies, data management, analytics, governance, and organizational components.

A maturity model for big data is useful for a number of reasons for any organization that is considering or in the process of implementing a big data project. First, it helps to create structure around a big data program and determine where to start. It also helps identify and define the organization's goals around the program and creates a process to communicate that vision across the entire organization. A maturity model will provide a methodology to measure and monitor the

[1] Philip Russom [2013]. *Managing Big Data*, TDWI Best Practices Report, Q4. Available at tdwi.org/bpreports.

state of the program and the effort needed to complete the current stage, as well as steps to move to the next stage of maturity. It serves as a kind of odometer to measure and manage the speed of your progress and adoption within the company for a big data program.

The TDWI Big Data Maturity Model consists of five stages: nascent, pre-adoption, early adoption, corporate adoption, and mature/visionary. As organizations move through these stages, they should be gaining more value from their investments. Figure 1 illustrates these stages and the movement from one stage to the next. There is also a chasm as organizations move from early adoption to corporate adoption. Although big data and big data analytics are relatively new and few companies are in the mature or visionary category, many large and larger midsize companies, as well as smaller Internet firms, are moving out of the nascent stage to begin to experiment with and even embrace big data and big data analytics. In fact, at some point, we will be able to draw a curve to depict big data stages of adoption.

In the figure, you'll note that—currently—the amount of time that users spend in the early adoption phase is often longer than the other phases. We discuss this in more detail below. In addition, although the stages in the figure are divided into discrete blocks, the reality is that companies don't necessarily move from one phase to the next so cleanly. For example, an organization might be advanced in one aspect of big data, such as its infrastructure, but relatively immature in its management strategy.

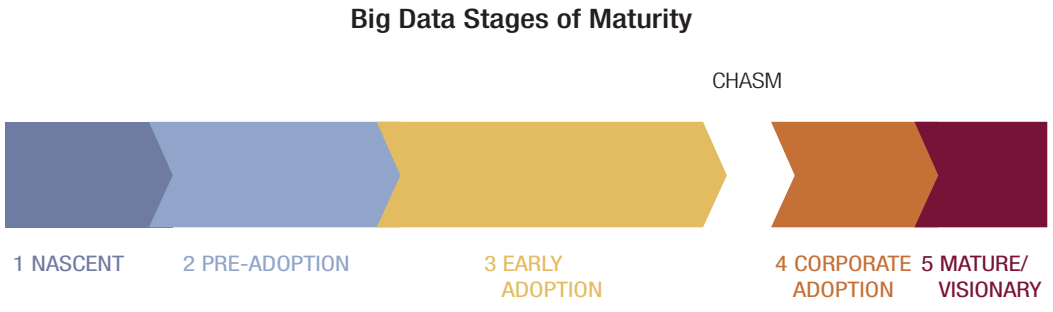
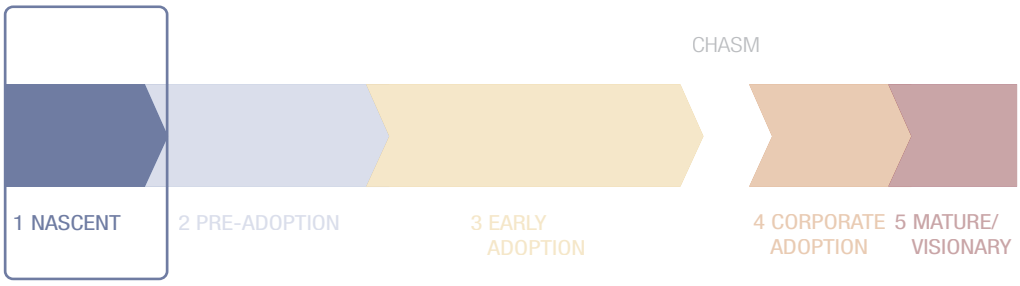


Figure 1. *Stages of maturity in the TDWI Big Data Maturity Model.*

The first section of the benchmark guide provides an overview of each of these five stages of the TDWI Big Data Maturity Model. This description provides a context to use to interpret your scores when you take the assessment. The TDWI Big Data Maturity Model Assessment Tool measures the maturity of a big data and big data analytics program in an objective way across various dimensions that are key to deriving value from big data analytics. These dimensions are organization, infrastructure, data management, analytics, and governance. (See Figure 2, page 14, for more information.)

Stages of Maturity



Stage 1: Nascent

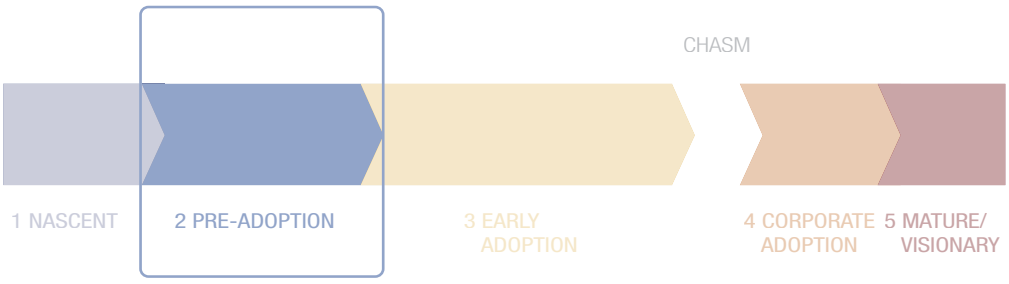
The nascent stage represents a pre–big data environment. In this stage, most companies have a low awareness of big data or its value across much of the business. There is no real executive support for the effort, although there are pockets of people spread throughout the company who may be interested in the potential value of big data. Often, the organization has bought into the concept of analytics and it may have a data warehouse, for instance, but it has not yet started to explore advanced analytics or begun its big data journey. This may also mean that its governance strategy is more IT centric than business-and-IT centric. Some key characteristics of the nascent organization include:

Data management. Often, despite having some sort of data warehouse (especially in the case of the enterprise), a nascent organization will have also collected data as files with different formats, but without naming standards, and with storage structures that are minimally defined. A nascent Internet-based business may have the same problem in terms of collecting data with no real data management strategy in place or idea of what to do with the data. In fact, data that could be useful may be discarded in this phase because no one knows what to do with it. In addition, the organization typically does not have much in the way of an end-state data architecture. Its data strategy and data life cycle management strategy are not strong and are often more about immediate results and silos of information than an overall plan.

Nascent organizations are also often not that far along in terms of an overall unified data strategy, nor do they see a reason to have one. These organizations may not yet have thought about what it means to be data driven.

Infrastructure. A key piece of the enterprise and its information management standards lies with the infrastructure of the corporation. Big data management requires an in-depth understanding of the new infrastructure to support critical big data components such as data types, data volume, user and data security, access to information, processing requirements, and information storage. The nascent organization often misunderstands the need to differentiate between infrastructure and data, leading to short-lived success in their big data journey.

Analytics. Nascent organizations may or may not make use of advanced analytics. However, the nascent organization is often pushing the boundaries of what it can do analytically. If the company has an analytics group in place, it is generally within a department or line of business and targeted at a specific function such as marketing. In other words, analytics is occurring in pockets and silos in the organization. Typically, if the organization is utilizing advanced analytics such as predictive analytics, it is working with structured data only.



Stage 2: Pre-Adoption

During the pre-adoption stage, the organization is starting to do its homework regarding big data analytics. Staff may be reading up on the topic and attending conferences. The organization may have invested in some new technology, such as Hadoop, in support of big data. It knows that it will be implementing big data analytics in the near term, although the effort is usually departmental in scope.

Organization. In this stage, the organization is just beginning its big data journey. In traditional companies, typically, one executive sponsor is on board, but companywide support can be spotty. Often the sponsor is in the CIO organization and not in the business. The mindset is generally around experimentation. The team charged with exploring big data is trying to determine the top business problems to solve and may have brought in some business partners to help with this. They realize that identifying the right business problem is critical for success. The company typically utilizes analytics as part of its decision-making process in various departments, but the company itself isn't necessarily analytics driven. In some companies at this stage, there may even be skepticism around big data and analytics.

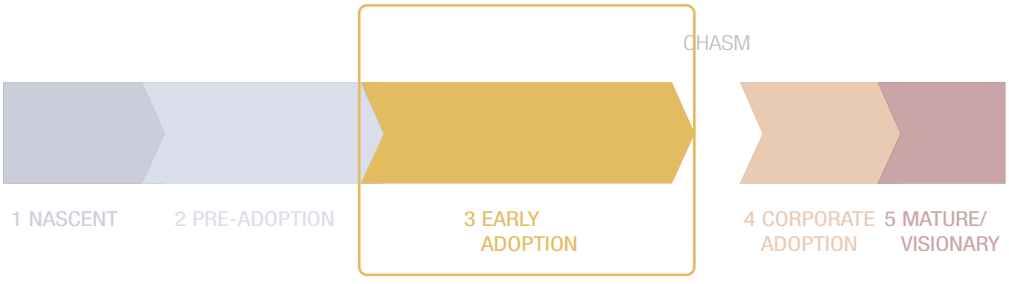
In Internet-based companies, where the mindset might be to get it done as quickly as possible, the goal might be to get the first application up and running and worry about the rest later. Typically the CIO team might be working with a business sponsor who is the driver behind the effort, but who may not yet be totally engaged.

Infrastructure. In this stage of maturity the company may be trying out Hadoop or some other big data technologies as part of the experimentation phase leading to a proof of concept. The installation, configuration, and maintenance are defined, and there is a level of compliance to an enterprise standard, although the class of infrastructure may not be close to production ready or may utilize a basic use case. For example, some companies may stand up a Hadoop cluster or invest in a NoSQL database, but this is likely only supporting one kind of workload to support a single initiative or department.

Data management. In this stage of maturity, the organization may have started to identify and collect some big data sources (these are typically internal) as part of the experimentation process. However, these files typically have different formats, with minimal naming standards and minimally defined storage structure. From a data strategy perspective, there is often metadata attributed at the departmental level and a defined end state data architecture, whether it is in a semantic layer, in the platform architecture, or in a database layer. There is no defined data life cycle management and no data auditability and lineage.

Analytics. Typically, the organization has explored some kind of advanced analytics. There may be some groups of individuals, especially at traditional companies, who are adept at predictive modeling, for example, but they may be aligned at the department or line-of-business level. These statisticians or business analysts may be part of the overall voice of the organization that sees the need for a big data implementation because they see the value. There may also be some outside organizations helping to set the analytics strategy or involved with performing some kind of advanced analytics.

Governance. At this stage of maturity, organizations might have a steering committee overseeing the program from a governance perspective, with representatives from the department providing reports on progress and compliance. However, most do not. For data governance, a similar effort may be in place with departmental data strategy, management, and a corporate data management steering committee overseeing the integration of data.



Stage 3: Early Adoption

This stage of maturity is typically characterized by one or two proofs of concept (POCs) which become more established and production ready. Organizations tend to spend a long time in this stage, often because it is hard to cross the chasm that leads to corporatewide adoption of big data and big data analytics.

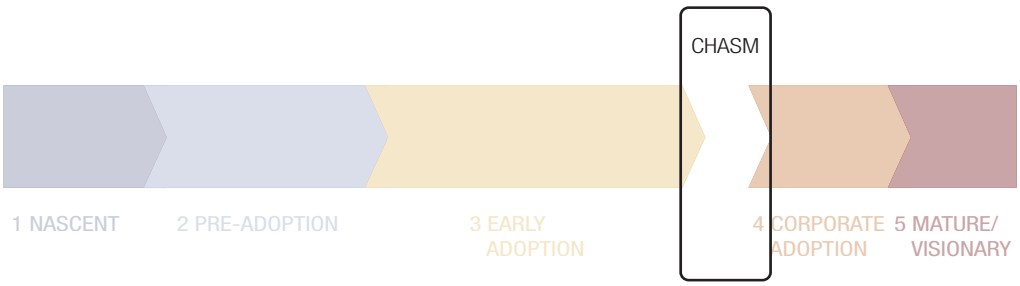
Organization. At this point, there is generally at least one executive sponsor involved. However, it is also at this stage when more executives might start to become interested in the program as companies show some wins in the POCs. As the organization gets excited about the prospects of big data, more people start to come on board. This often means that a team is established to start to plan and strategize for a wider big data scope. This also means that politics may start to kick in, especially in cases where the goal of the big data project is cost containment rather than competitive advantage.

Infrastructure. During the early adoption stage, there may be various kinds of big data technology in place. This might include an appliance or a Hadoop cluster. NoSQL databases might be in place. However, the notion of a unified architecture or ecosystem is not yet widespread and these technologies are not operationalized. A typical infrastructure is a tier 2 production-class cluster that is installed and maintained in the company's data center or even in the cloud. The installation, configuration, and maintenance are defined per enterprise standard. If a NoSQL database is in place, its use is generally offline and not part of the operational infrastructure. In some companies where cost savings are paramount, Hadoop may be used to take advantage of growing data volumes without expanding the data warehouse. Although this is a valid hybrid approach, it may not be well managed and the data may not be used yet for meaningful analysis.

Data management. In this stage of maturity, organizations will have data collected as files of different formats, potentially with division or enterprise standards for naming and storage management. From a data strategy perspective, there may be metadata attributed at the division level and a defined end

state data architecture, whether it is in a semantic layer, in the platform architecture, or in a database layer. At this stage, the organization is not throwing data out unless there is a specific purpose for doing so. However, a solid companywide big data management strategy is typically not in place. Ideally, if the organization is moving at a fast pace to implement its big data initiatives, they are aware of issues related to data quality and security.

Analytics. The organization may be utilizing descriptive or even predictive analytics in its projects. In early adoption, the kind of analytics tools will depend on the problem that the company is trying to solve. Typically, organizations are still using one kind of data, although that may vary among organizations in this phase. For example, some companies in early adoption are utilizing large volumes (i.e., more than 10 TB) of structured data that is stored in an appliance. The company may be running some sort of predictive model on this data. This is an established and production-ready implementation, but it still may be departmental in scope and has not moved on to other forms of data. Alternately, a company (e.g., a publisher) might be mature in managing and utilizing large amounts of content but not strong analytically. Some companies might be utilizing different kinds of data but not in an integrated fashion. For instance, some organizations may be primarily utilizing internal structured data but making use of unstructured social media data in another part of the company. In addition, sometimes one department in a company has a specific use for big data analytics. This might be the network monitoring department in the case of a wireless carrier. That department might be advanced in using location data as well as other kinds of data for its analysis, but the deployment is isolated to that department.



The Chasm

As organizations try to move from early adoption to corporate adoption, there is generally a series of hurdles they need to overcome. This is often why companies spend a large amount of time in this phase. There is the obvious challenge of obtaining the right skill set. Hadoop skills and advanced analytics skills may not be present in the organization. There may also be political issues. For example, one organization may have been driving the company's big data effort and has brought other departments on board. However, when it comes time to extend the platform or put more stringent standards and governance in place, departments begin to fight for control over who owns the data or who may want their particular vision in place. In the case of smaller companies that want to be nimble, there comes a point after several projects or as they start to grow when they realize they might need to put some structure in place and deal with issues such as data security or management.

To successfully cross the chasm, you'll need to address the following challenges:

Funding. Many early big data projects are bootstrapped or driven by a visionary executive IT champion. Of course, it is critical to establish wins with these early projects in order to secure future funding. IT funding is often not enough; business involvement is critical. Some companies are unable to get past the prototyping or POC phase because they don't have business funding and

buy-in. There needs to be a bridge between IT and the business in big data initiatives. Business involvement is needed because big data projects must have business value with tangible business outcomes.

Data management and data governance. In order to move forward to corporate adoption and sharing of data, a solid data management and governance plan needs to be in place. Some organizations believe this is the single most important key to big data maturity. To get to corporate adoption, data is going to need to be shared across the organization. This means that some sort of unified information architecture or companywide analytics platform, or at least a coherent way to get at data for analytics, is in place. That means data management and governance are critical. Even for smaller and arguably more nimble organizations, there comes a point when process becomes important to growth.

Architecture. Some companies in early adoption try to put together best-of-breed approaches, but this often results in an architecture that is too heavy before it gets off the ground. Big data is complex. Companies that have been successful take it one step at a time. They are often surgically precise in how they deploy big data technologies.

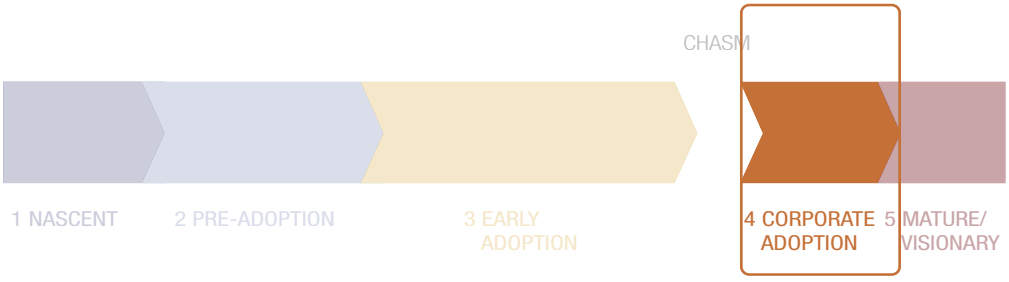
Skill sets. A big barrier for big data projects moving past the chasm is developing the skill set for new technologies such as Hadoop or NoSQL databases. If the company can afford it, they will go out and hire staff for this. For instance, some companies find university hires to be effective. However, these hires may not understand the business, so there is often a significant learning curve. There can also be a disconnect between the Hadoop development team and the traditional data warehouse team. It is important to get these two teams together. Some organizations have found that moving teams together helps develop skills across the board and makes it easier to drive an ecosystem approach going forward. For example, the ETL team understands data quality and data models and can provide that insight, along with a knowledge of the business, to the Hadoop team. Companies also need the skills to operationalize the big data technologies and might move to commercial-grade Hadoop clusters or enterprise NoSQL to cross the chasm, if they weren't in place before. More mature organization will have a view about the kind of staff they need to run these systems.

Cultural and political issues. Often it is the cultural and political issues that can stop big data analytics from becoming more pervasive throughout the organization, especially when the motivation for big data is more about cost containment than competitive advantage. For example, a company might have built a solid POC that is being used to drive competitive advantage. However, an executive might get wind of this and think that the data management infrastructure needs to go into the cloud, although not everyone agrees. If there is no solid plan and road map or business case in place with executive sponsorship (that staunchly supports the plan) or a determined set of product managers, this may stall the project. Of course, it might get stalled anyway while the discussions are under way. This kind of roadblock is often more of an issue in large, more established companies than in small or midsize companies.

Governance. Moving to corporate adoption requires a big data governance team. Some companies may feel that they are mature because they have corporate sponsorship and a big Hadoop cluster in place. However, they are now faced with data management, life cycle, and governance issues. At this stage of maturity, organizations should have program governance in place, with program management office (PMO) guidance for the program and a steering committee that oversees the program from a division's perspective. The program should be executed as a budgeted and planned initiative from the division's perspective and be treated on par with other data integration programs. Of course, Internet companies might feel that it is more important to be agile than to be weighed

down by issues such as governance because of the push to get products out the door. Company growth will most likely be the determinant here. If companies grow quickly, they will need to put governance processes in place or deal with the consequences.

Crossing the chasm can be a long exercise in the big data journey unless the organization has both top-down and bottom-up team strategies in place or is compelled by a strong business reason to make big data happen. To cross the chasm, companies need to ensure that the right governance, data architecture, data life cycle management security strategies, and organizational structures are in place. In our observations to date, we have seen many organizations spend extra time as they near the chasm and require additional staffing to cross it.



Stage 4: Corporate Adoption

Corporate adoption is the major crossover phase in any organization's big data journey. During corporate adoption, end users typically get involved, gain insights, and transform how they do business. For instance, they may change how decisions are made by operationalizing big data analytics in the organization.

Most organizations trying to reach this stage of maturity might have repeatedly addressed certain gaps in organization, infrastructure, data management, analytics, and governance.

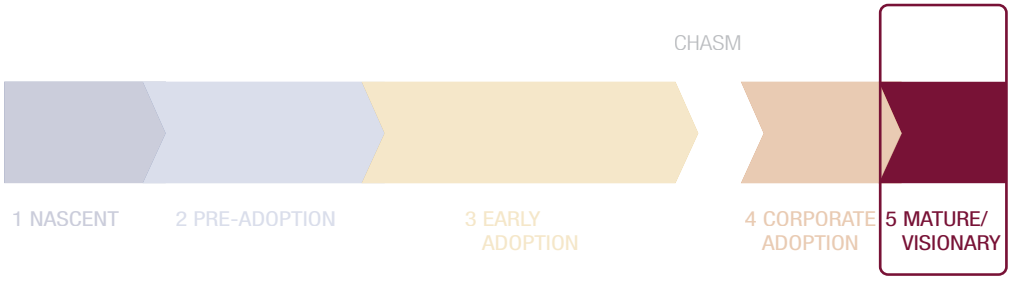
Organization. At this stage, the company has usually come to realize that analytics is a competitive differentiator. Innovation in data and data analysis is a core value, and an analytics culture prevails. The business strategy is generally top-down and bottom-up, with a data infrastructure that can support this. The funding process is secured, and an ROI is in place for big data analytics.

Infrastructure. In this stage of maturity, the typical company infrastructure is a tier 1 production-class cluster that is installed and maintained in the data center, which might include the cloud. A range of technologies might be used, including enterprise NoSQL databases, Hadoop, and appliances or a data warehouse, but the information architecture is unified in a way that underpins the analytics. A Hadoop cluster might include 50 to 100 nodes at this phase (although cluster size is not always tied to maturity). This might also be a commercial distribution because as the company moves out of early adoption, it is actually concerned about managing the clusters. Another sign of operational maturity is that the company can perform multiple workloads on a cluster. The installation, configuration, and maintenance of the infrastructure are defined per enterprise standards. The deployment is driven at an enterprise level in the organization with the complete support of IT teams and line-of-business participation. The infrastructure and architecture of the big data ecosystem comply with backup and recovery or disaster recovery procedures, which are in place. A unified architecture that takes an ecosystem approach is also in place.

Data management. The more mature a company is, the better it is able to manage and make use of its data. In this stage of maturity, organizations can make use of many forms of data. Data sharing is a collaborative activity that is well managed through strong data governance policies. Data siloing is minimized in this phase, and the company sees value from all components of the data infrastructure. The company is always looking for new data to enhance its analytics. From a data strategy perspective, there is metadata attributed at the division or company level and a defined end state data architecture, whether it is in a semantic layer, in the platform architecture, or in a database layer. There is a defined data life cycle management and a data auditability and lineage process or framework.

Analytics. Any company might collect a lot of data but not make use of it. It's a sign of maturity when new data coming into the organization can be analyzed quickly and made part of the logical infrastructure. In this stage, analytics supports the organization. Typically, a company at this stage also has a center of excellence (COE) in place that serves different parts of the organization. The COE includes the data science team, which might even train other groups in the use of analytics in different forms. The COE is generally also tasked with evangelizing big data, and its members might also be looking for commonalities in analytic needs to determine what might be reusable in other parts of the company. At this point, analytics might be operationalized as part of a business process. In other words, analytics might be automated or integrated with the business process. An example of this might be using machine-generated data together with other data about customers to determine a next best offer. These companies are typically using new capabilities and not just existing BI infrastructure.

Governance. The truly mature company understands that big data can be a liability waiting to happen. This company is concerned with answering questions such as, "Whose data was it? Whose data is it? Where is it going? How long will it last?" At this stage of maturity, organizations will have program governance in place, with PMO guidance for the program and a steering committee that oversees the program from a company perspective. For data governance, a similar effort will be in place with a well-defined data strategy and management and steering committee overseeing the progress of data. The overall executive sponsor is involved from a monthly update perspective. The program is executed as a budgeted and planned enterprise initiative and treated on par with other data integration programs.



Stage 5: Mature/Visionary

Only a few companies can currently be considered visionary in terms of big data and big data analytics. At this stage, organizations are executing big data programs as a well-oiled machine using a highly tuned infrastructure with well-established program and data governance strategies. The program is executed as a budgeted and planned initiative from the company perspective. In the visionary stage, there is excitement and energy around big data and big data analytics.

Organization. The visionary company exhibits several characteristics. First, executives have bought into analytics and big data analytics and view it as critical, de facto standard for how to do business. Analytics is seen as a competitive weapon and the mindset is creative. Second, analytics is not simply used to drive strategy or insight; instead, the business is always looking for opportunities to use analytics in new ways. In fact, a key characteristic of a visionary company is that it is continuously determining new ways to use and create value from analytics. Finally, collaboration is a big part of the culture.

Infrastructure and data management. The way you manage complexity is key to big data maturity. The visionary company has deployed a coherent analytics infrastructure that is fully operational and can be used in the mission-critical aspects of the business. Part of the infrastructure includes the ability to integrate new sources of data for analytics, whether they are internal or external to the company. The infrastructure leverages commercial Hadoop and enterprise NoSQL databases and there is security, disaster recovery, backup and recovery, performance management, and proactive infrastructure monitoring in place—even if the public cloud is used. The program is treated as mission critical and given the right amount of staffing and skills. Data is shared across the organization.

Analytics. In the visionary company, the mindset is to continually develop analytics. Typically, this kind of company makes use of all kinds of data, including real-time data, and uses this as part of its decision making and incorporates into business processes. The visionary company can connect the dots between new data and existing assets. The business can tap into this to enhance product offerings and bring new products to market. The infrastructure, data, and analytics are completely intertwined. There are COEs in place, and teams are working to deliver new and exciting forms of analytics. Some visionary organizations build an innovation team consisting of business and IT that innovates on the technologies, brings them back to the business, and takes them into production.

Evaluating Benchmark Scores

The benchmark survey has approximately 50 questions across the five categories that form the dimensions of the TDWI Big Data Maturity Model (see Figure 2).

- **Organization:** To what extent does the organizational strategy, culture, leadership, and funding support a successful big data analytics program? What value does the company place on analytics?
- **Infrastructure:** How advanced and coherent is the architecture in support of a big data initiative? To what extent does the infrastructure support all parts of the company and potential users? How effective is the big data management approach? What technologies are in place to support a big data initiative, and how are they integrated into the existing environment?
- **Data management:** How extensive are the variety, volume, and velocity of data used for big data analytics, and how does the company manage its big data in support of analytics? This includes data quality and processing as well as data integration and storage issues.
- **Analytics:** How advanced is the company in its use of big data analytics? This includes the kinds of analytics utilized and how the analytics are delivered in the organization. It also includes the skills required to make analytics happen.
- **Governance:** How coherent is the company's data governance strategy in support of its big data analytics program?

These dimensions should seem familiar because they are the same categories we have been referencing throughout this guide. In addition, the assessment looks at data relating to the maturity of a company's BI deployment, whether or not the company is outsourcing its big data capabilities, and how the company is using the cloud. These factors and others are used to explore relationships in the data in order to help determine best practices for big data analytics.

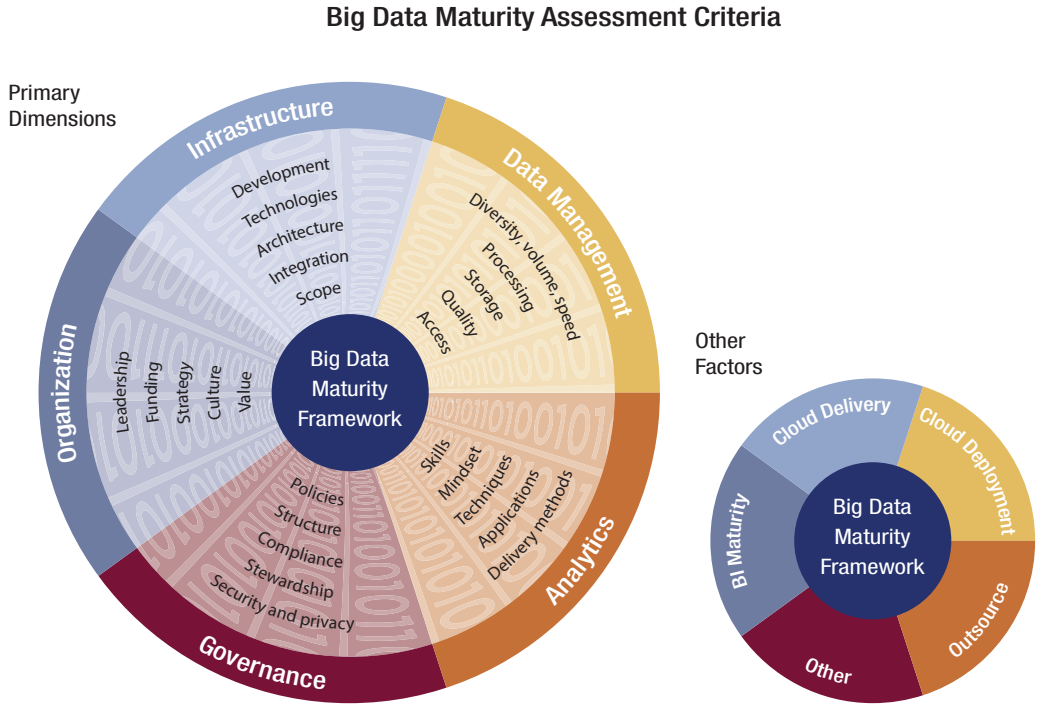


Figure 2. *Scores are based on five primary dimensions of big data maturity, plus several other factors.*

Of course, organizations can be at different stages of maturity in each of these five categories, and most are. In addition, as we've mentioned throughout this guide, there are different kinds of companies. For instance, an Internet-based company without an established information infrastructure may have a different experience than an enterprise with a large geographic footprint.

Scoring

There are approximately 50 questions in the assessment—10 in each of the five primary dimensions. Questions may be weighted differently depending on their relative importance. Each dimension has a potential high score of 50 points. Because organizations can be at different levels of maturity in the five dimensions, we score each section separately as well as provide an overall score. There are also questions that aren't scored, but rather used for best practices guidance.

The output of the assessment is a score in each dimension and the total score as well as a gap recommendation that provides advice and best practices for getting to the next stages of maturity.

Interpretation

Once you complete the survey, a report-based interface will show how your responses compare to those of your peers. The breakdown of scores for each dimension is as follows:

SCORE PER DIMENSION	STAGE
<15	Nascent
16–25	Pre-Adoption
26–35	Early Adoption
36–45	Corporate Adoption
46–49	Mature
50	Visionary

For instance, if you receive a score of 17 in the organization dimension of the assessment, you are in the pre-adoption stage for that particular dimension. You should expect to see your scores for each dimension vary, as big data programs don't necessarily evolve at the same rate across all of the dimensions. For example, your company might be more advanced in terms of bringing big data sources together than it is in analyzing them or governing this data.

The chasm can be overlaid between the early adoption and corporate adoption stages.

When you complete the assessment, you might see scores like this:

DIMENSION	SCORE	STAGE
Organization	17	Pre-Adoption
Infrastructure	37	Corporate Adoption
Data Management	27	Early Adoption
Analytics	20	Pre-Adoption
Governance	10	Nascent

Total Score: 111/22.2 (i.e., the average)

This means that you are mature in your infrastructure but still have a ways to go to manage, analyze, and govern your information assets. You will receive feedback in those areas.

Summary

The TDWI Big Data Maturity Model Assessment Tool provides a quick way for organizations to assess their maturity in big data analytics and compare themselves in an objective way against others with big data initiatives. The assessment is based on the TDWI Big Data Maturity Model, which consists of five maturity stages with a chasm between stages 3 and 4.

The assessment serves as a preliminary measure of your big data maturity. It consists of approximately 50 questions across five categories; this merely touches the surface of all of the complexities involved in building out your big data analytics ecosystem. To gauge precisely where you are, it may also make sense to work with an independent source to validate your progress.

Cloudera

www.cloudera.com

Founded in 2008, Cloudera pioneered the business case for Hadoop with CDH: the world's most comprehensive, thoroughly tested, and widely deployed 100% open source distribution of Apache Hadoop in both commercial and non-commercial environments. Now, the company is redefining data management with its platform for big data, Cloudera Enterprise, empowering enterprises to Ask Bigger Questions™ and gain rich, actionable insights from all their data, to quickly and easily derive real business value that translates into competitive advantage. As the top contributor to the Apache open source community and leading educator of data professionals with the broadest array of Hadoop training and certification programs, Cloudera also offers comprehensive consulting services. Over 700 partners across hardware, software, and services have teamed with Cloudera to help meet organizations' big data goals. With tens of thousands of nodes under management and hundreds of customers across diverse markets, Cloudera is the category leader that has set the standard for Hadoop in the enterprise.

IBM

ibm.com/bigdata

Big data can address a multitude of problems if you know how to use it. The insights gleaned from big data can help organizations deepen customer engagement, optimize operations, prevent threats and fraud, and capitalize on new sources of revenue. But to be successful, you need a clear understanding of the characteristics of big data and the skills required to manage it.

IBM is unique in having developed an enterprise-class big data platform that allows you to address the full spectrum of big data business challenges.

Why a platform versus individual products to address big data pain points? The real benefit of the platform is leverage—the ability to start with one capability and easily add others over your big data journey. IBM is the only vendor with this broad and balanced view of big data with the needs of a platform—the benefit is pre-integration of its components to reduce your implementation time and cost. To learn more, visit ibm.com/bigdata.

MarkLogic

www.marklogic.com

For more than a decade, MarkLogic has delivered a powerful and trusted next-generation enterprise NoSQL database that enables organizations to turn all data into valuable and actionable information. Key features include ACID transactions, horizontal scaling, real-time indexing, high availability, disaster recovery, government-grade security, and more. Organizations around the world rely on MarkLogic's enterprise-grade technology to make better decisions faster. MarkLogic is headquartered in Silicon Valley with field offices in Washington, D.C., New York, London, Frankfurt, Utrecht, and Tokyo. For more information, please visit www.marklogic.com.

Pentaho

pentaho.com

Pentaho, Inc., is well known for its unified, open, embeddable, pluggable business analytics platform architected for the future of analytics. Pentaho Business Analytics tightly couples data integration (DI) with business intelligence (BI) and analytics to support the entire big data analytics process. Pentaho Business Analytics provides a simplified yet powerful analytics experience across the broadest spectrum of analytics, from traditional reporting and dashboards to visualization tools and predictive analytics. Pentaho's enterprise-ready big data integration includes the adaptive big data layer, which provides the broadest and deepest support for big data across all the top Hadoop, NoSQL, and analytic database systems today, and delivers native integration to each while allowing portability between them for full insulation from risk. Pentaho's architected data blending allows big data and traditional data to be blended together at the source, maintaining governance, accuracy, and timeliness, for agile access and use by all the analytics in its suite. Pentaho's Visual MapReduce tool eliminates the need for the complex coding normally required with big data, increasing productivity with current staff and skills. From integration to blending to visualization to powerful analytics, Pentaho brings your big data to life.

TDWI RESEARCH

TDWI Research provides research and advice for business intelligence and data warehousing professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of business intelligence and data warehousing solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

555 S Renton Village Place, Ste 700
Renton, WA 98057-3295

T 425.277.9126
F 425.687.2842
E info@tdwi.org

tdwi.org