

Five Guiding Principles for Realizing the Promise of Big Data



Bhargav Mantha is a manager at ZS Associates and a leader in the firm's global business intelligence practice. bhargav.mantha@zsassociates.com

Bhargav Mantha

Abstract

Most organizations characterize big data in terms of volume, velocity, and variety, but it is more useful to consider big data in the same way we look at information management, analytics, and how they impact business decisions. After all, *big data* is a sweeping term that includes a variety of enterprise concerns, from managing and securing data sets to technologies that can analyze the data more quickly and thus enhance business value.

In this article, we outline five guiding principles to help companies make prudent investments and realize the promise of big data. Businesses should use these guidelines to help them think hard about when, where, and how to best realize big data's value within their organizations.

Introduction

Hardly a day goes by without some mention of big data in our lives. The hype-versus-hope debate of big data will continue for some time as organizations across industries grapple with the questions of why big data is important, what to do with it, and how to get started.

Although big data is most easily characterized in terms of high volume, velocity, and variety, it is more practical to define big data by the way we think about information management and analytics and how they impact business decisions.

One of the biggest obstacles organizations face is thinking big data is an initiative when, in fact, big data is an umbrella term that covers many problem spaces, data sets, technologies, and opportunities for enhancing business value.

Here are five guiding principles to help your enterprise avoid becoming overwhelmed by the hype and focus instead on making prudent investments that will help you realize the promise of big data.

Principle #1: Determine the business case first.

A critical step for key executives to ensure big data adoption is to identify the business initiative and quantify tangible business value. This involves pinpointing which parts of the business would benefit from expanding available data to provide more complete answers. For example, a brand manager investigating a decline in sales may want to augment the analysis by integrating insight from call center records, Web logs, and consumer sentiment through social media commentary on quality, functionality, or price.

Key executives may also determine if big data analytics can help monetize a portion of their business. For example, they may use analytics to immediately make a relevant offer after a credit card is used to initiate another transaction instead of storing the transaction for later reporting.

The business cases for investing in big data vary. They can be business-process-specific, such as improving the customer experience, optimizing R&D, or managing IT. They can be industry-specific, such as optimizing price or channels for technology firms, detecting fraud for the financial services industry, managing intellectual property for media companies, or improving treatment outcomes for healthcare providers.

Organizations often take a misstep by thinking that big data is just another source for business intelligence (BI). For example, one organization confessed to using their big data pilot to build Facebook and Twitter interfaces to gather social media data, but said the effort was unsuccessful because executives failed to consider what to do with that data. They didn't determine at the outset how to process the data, what questions it could answer, and what analytics were required to make sense of it (sentiment analysis, monitoring evolving topics, or uncovering networks and relationships).

Finding a business-driven initiative with measurable outcomes—whether improved customer retention, increased revenue from improved sales/channel productivity, or even cost reduction—will improve your organization's success rate with big data initiatives.

A critical step for key executives to ensure big data adoption is to identify the business initiative and quantify tangible business value.

Principle #2: Use the right tools and technologies.

Organizations should consider four main capabilities to expand their existing BI and analytics initiatives to support big data analytics.

The most important capability is **advanced analytics** to uncover previously hidden patterns. With new types of data comes the need to apply new types of algorithms, such as entity analytics, network analytics, text analytics, and real-time scoring. Scalability is important because improved accuracy and trust in your data means your users are more likely to want to integrate additional data sources or increase data volumes. Analytics must be able to push these algorithm processes to interpret text, images, and video streams.

Visualization and exploration can help your enterprise find more complete answers to business questions. New types of data (and greater volume) increases the need for new forms of visualization (such as heat maps) to present the data to users and highlight important patterns. Tools such as Tableau Software and Datameer enable interactive, iterative, search-like, visual data discovery.

The third capability is to turn **insight into action** to drive a decision—either with a manual step or an automated process. Applying analytics to streaming big data requires technology that uses predictive models and

business rules to automate decisions and identify outliers where business judgment is needed.

Finally, analytics tools must assemble the **right mix of information** in a way that makes sense to the business. This may include:

- Tooling to compose fast-performing queries on very large data sets or to access high-performance analytic databases such as Aster Data, EMC Greenplum, or IBM Netezza
- Analytic processing capabilities to ingest data in motion, apply filters, and surface relevant real-time data
- Query and process returned data from unstructured data (for example, in HDFS, the Hadoop Distributed File System)

Big data requires more than just Hadoop. Although that open source software framework has the greatest name recognition, big data is too varied and complex for a one-size-fits-all solution. Other classes of technologies are equally well suited to managing big data, such as NoSQL (not only SQL) and MPP (massively parallel processing) stores.

Again, what matters is which of the three Vs (volume, velocity, variety) poses the greatest challenge for you and which of these technologies supports the business case. In fact, there is no requirement for you to invest in your own infrastructure. Instead, you might explore options for a cloud-based service, such as Google BigQuery, and save on infrastructure costs.

Principle #3: People, people, people.

After you've developed the business case for big data, begin a thorough skills assessment, because newer analysis techniques and technologies may require different skills or talent. There are three particular roles (and associated competency models) that you can define for a big data initiative:

- The **data scientist**, who applies his or her statistical, mathematical, and computer science skills to work on large, complex data sets to find, interpret, and distribute statistically significant information. He or she will also ensure that significance is easily understood and acted upon by others.
- The **business analyst**, who blends business understanding with data acumen to determine what information is important for the business and how to bridge the IT or data science gap.
- The **technologist**, who has the skills needed to identify and assemble the best set of big data technologies and developers (for example, Hadoop and Hive) to deliver on the business initiative.

Notice that the skills required don't all need to be about Hadoop and advanced algorithms. One of our clients admitted to feeling overwhelmed by the hype, leading them to think of big data initiatives as beyond the company's technology skills. In fact, all the client was looking to do was gain insights from clickstream data, which did not require Hadoop or the skills of a data scientist. Mapping the business case, determining the technology needed, and obtaining the appropriate skill sets helped the client overcome their fear and make the right investments toward big data analytics.

Principle #4: Start thinking social.

Big data could be an important component of your social media strategy, especially when it comes to understanding customers, prospects, and key influencers. Social media allows for ongoing engagement that can provide near-real-time insight into customer attitudes and behavior. Analysis of social media data can help you rapidly identify trends: who uses your solutions, what customers and prospects think about your and your competitors' brands and solutions, and what emerging markets are developing.

Recognizing the value of and leveraging social media data sources are relatively new challenges for many organizations. A social intelligence effort will require rethinking and redesigning existing information manage-

ment ecosystems. New analytical platforms, techniques, tools, and governance processes are needed to unlock customer insights.

The implementation of a socially enabled business through big data includes three main steps:

1. **Listening** to consumer dialogue on social networks, sites, and communities and collecting the data
2. **Analyzing** the gathered information (mostly opinions) and applying natural language processing algorithms to extract actionable meaning and the most recurrent themes
3. **Engaging** with customers by closing the loop and taking quick, decisive, and appropriate actions based on gathered insights

Principle #5: Don't treat big data as mission critical right away.

Although big data quality will become increasingly important, don't treat social media data or wiki data like mission-critical financial data right away. Apply the appropriate level of control to its use and exposure.

Your initiative may well be stifled from day one if you apply the rigor of initiating and managing traditional data warehouse projects. Instead, help the process be iterative and collaborative: let the business and IT explore interesting sources of data, refine what is important, and apply the appropriate algorithms. Better outcomes are possible when an organization conscientiously allows big data initiatives to be iterative, exploratory, and even transient in some cases.

The Big Data Opportunity

Big data presents a growing opportunity to understand and change interactions with customers. It allows companies to improve existing business processes, to launch new lines of business, and to reevaluate how and why data can improve decision-making processes. Using these guidelines, think hard about when, where, and how to best realize big data's value within your organization. ■