This page intentionally left blank

CHAPTER 1

Market and Business Drivers for Big Data Analytics

1.1 SEPARATING THE BIG DATA REALITY FROM HYPE

There are few technology phenomena that have taken both the technical and the mainstream media by storm than "big data." From the analyst communities to the front pages of the most respected sources of journalism, the world seems to be awash in big data projects, activities, analyses, and so on. However, as with many technology fads, there is some murkiness in its definition, which lends to confusion, uncertainty, and doubt when attempting to understand how the methodologies can benefit the organization.

Therefore, it is best to begin with a definition of big data. The analyst firm Gartner can be credited with the most-frequently used (and perhaps, somewhat abused) definition:

> Big data *is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.*[1]

For the most part, in popularizing the big data concept, the analyst community and the media have seemed to latch onto the alliteration that appears at the beginning of the definition, hyperfocusing on what is referred to as the "3 Vs—volume, velocity, and variety." Others have built upon that meme to inject additional Vs such as "value" or "variability," intended to capitalize on an apparent improvement to the definition.

The ubiquity of the Vs definition notwithstanding, it is worth noting that the origin of the concept is not new, but was provided by (at the time Meta Group, now Gartner) analyst Doug Laney in a research note from 2001 about "3-D Data Management," in which he noted:

[1]Gartner's IT Glossary. Accessed from <http://www.gartner.com/it-glossary/big-data/> (Last accessed 08-08-13).

Big Data Analytics. DOI: http://dx.doi.org/10.1016/B978-0-12-417319-4.00001-6

> *While enterprises struggle to consolidate systems and collapse redundant databases to enable greater operational, analytical, and collaborative consistencies, changing economic conditions have made this job more difficult. E-commerce, in particular, has exploded data management challenges along three dimensions: volumes, velocity and variety. In 2001/02, IT organizations must compile a variety of approaches to have at their disposal for dealing with each.*[2]

The challenge with Gartner's definition is twofold. First, the impact of truncating the definition to concentrate on the Vs effectively distils out two other critical components of the message:

1. "cost-effective innovative forms of information processing" (the means by which the benefit can be achieved);
2. "enhanced insight and decision-making" (the desired outcome).

The second is a bit subtler: the definition is not really a definition, but rather a description. People in an organization cannot use the definition to determine whether they are using big data solutions or even if they have problems that need a big data solution. The same issue impedes the ability to convey a value proposition because of the difficulty in scoping what is intended to be designed, developed, and delivered and what the result really means to the organization.

Basically, it is necessary to look beyond what is essentially a marketing definition to understand the concept's core intent as the first step in evaluating the value proposition. Big data is fundamentally about applying innovative and cost-effective techniques for solving existing and future business problems whose resource requirements (for data management space, computation resources, or immediate, in-memory representation needs) exceed the capabilities of traditional computing environments as currently configured within the enterprise. Another way of envisioning this is shown in Figure 1.1.

To best understand the value that big data can bring to your organization, it is worth considering the market conditions that have enabled its apparently growing acceptance as a viable option to supplement the intertwining of operational and analytical business application in light of exploding data volumes. Over the course of this book, we hope to

[2]Doug Laney. Deja VVVu: others claiming Gartner's construct for big data, January 2012. Accessed from <http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/>.

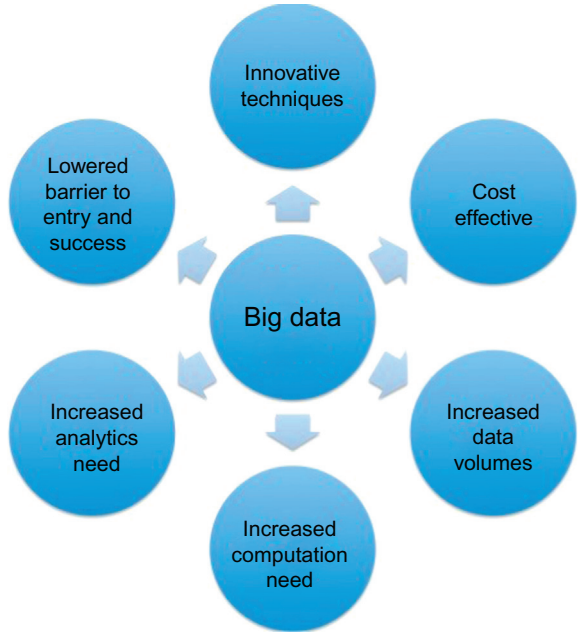

Figure 1.1 Cracking the big data nut.

quantify some of the variables that are relevant in evaluating and making decisions about integrating big data as part of an enterprise information management architecture, focusing on topics such as:

- characterizing what is meant by "massive" data volumes;
- reviewing the relationship between the speed of data creation and delivery and the integration of analytics into real-time business processes;
- exploring reasons that the traditional data management framework cannot deal with owing to growing data variability;
- qualifying the quantifiable measures of value to the business;
- developing a strategic plan for integration;
- evaluating the technologies;
- designing, developing, and moving new applications into production.

Qualifying the business value is particularly important, especially when the forward-looking stakeholders in an organization need to effectively communicate the business value of embracing big data platforms, and correspondingly, big data analytics. For example, a business

justification might show how incorporating a new analytics framework can be a competitive differentiator. Companies that develop customer upselling profiles based on limited data sampling face a disadvantage when compared to enterprises that create comprehensive customer models encompassing *all* the data about the customer intended to increase revenues while enhancing the customer experience.

Adopting a technology as a knee-jerk reaction to media buzz has a lowered chance of success than assessing how that technology can be leveraged along with the existing solution base as away of transforming the business. For that reason, before we begin to explore the details of big data technology, we must probe the depths of the business drivers and market conditions that make big data a viable alternative within the enterprise.

1.2 UNDERSTANDING THE BUSINESS DRIVERS

The story begins at the intersection of the need for agility and the demand for actionable insight as the proportion of signal to noise decreases. Decreasing "time to market" for decision-making enhancements to all types of business processes has become a critical competitive differentiator. However, the user demand for insight that is driven by ever-increasing data volumes must be understood in the context of organizational business drivers to help your organization appropriately adopt a coherent information strategy as a prelude to deploying big data technology.

Corporate business drivers may vary by industry as well as by company, but reviewing some existing trends for data creation, use, sharing, and the demand for analysis may reveal how evolving market conditions bring us to a point where adoption of big data can become a reality.

Business drivers are about agility in utilization and analysis of collections of datasets and streams to create value: increase revenues, decrease costs, improve the customer experience, reduce risks, and increase productivity. The data explosion bumps up against the requirement for capturing, managing, and analyzing information. Some key trends that drive the need for big data platforms include the following:

- **Increased data volumes being captured and stored**: According to the 2011 IDC Digital Universe Study, "In 2011, the amount of

information created and replicated will surpass 1.8 zettabytes, … growing by a factor of 9 in just five years."[3] The scale of this growth surpasses the reasonable capacity of traditional relational database management systems, or even typical hardware configurations supporting file-based data access.

- **Rapid acceleration of data growth**: Just 1 year later, the 2012 IDC Digital Universe study ("The Digital Universe in 2020") postulated, "From 2005 to 2020, the digital universe will grow by a factor of 300, from 130 exabytes to 40,000 exabytes, or 40 trillion gigabytes (more than 5,200 gigabytes for every man, woman, and child in 2020). From now until 2020, the digital universe will about double every two years."[4]
- **Increased data volumes pushed into the network**: According to Cisco's annual Visual Networking Index Forecast, by 2016, annual global IP traffic is forecast to be 1.3 zettabytes.[5] This increase in network traffic is attributed to the increasing number of smartphones, tablets and other Internet-ready devices, the growing community of Internet users, the increased Internet bandwidth and speed offered by telecommunications carriers, and the proliferation of Wi-Fi availability and connectivity. More data being funneled into wider communication channels create pressure for capturing and managing that data in a timely and coherent manner.
- **Growing variation in types of data assets for analysis**: As opposed to the more traditional methods for capturing and organizing *structured* datasets, data scientists seek to take advantage of unstructured data accessed or acquired from a wide variety of sources. Some of these sources may reflect minimal elements of structure (such as Web activity logs or call detail records), while others are completely unstructured or even limited to specific formats (such as social media data that merges text, images, audio, and video content). To extract usable signal out of this noise, enterprises must enhance their existing structured data management approaches to accommodate semantic text and content-stream analytics.
- **Alternate and unsynchronized methods for facilitating data delivery**: In a structured environment, there are clear delineations of the

[3]2011 IDC Digital Universe Study: extracting value from chaos, <http://www.emc.com/collateral/demos/microsites/emc-digital-universe-2011/index.htm>.

[4]The Digital Universe in 2020, <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>.

[5]See Cisco Press Release of May 30, 2012, <http://newsroom.cisco.com/press-release-content?type=webcontent&articleId=888280>.

discrete tasks for data acquisition or exchange, such as bulk file transfers via tape and disk storage systems, or via file transfer protocol over the Internet. Today, data publication and exchange is full of unpredictable peaks and valleys, with data coming from a broad spectrum of connected sources such as websites, transaction processing systems, and even "open data" feeds and streams from government sources and social media networks like Twitter. This creates new pressures for rapid acquisition, absorption, and analysis while retaining currency and consistency across the different datasets.

- **Rising demand for real-time integration of analytical results**: There are more people—with an expanding variety of roles—who are consumers of analytical results. The growth is especially noticeable in companies where end-to-end business processes are augmented to fully integrate analytical models to optimize performance. As an example, a retail company can monitor real-time sales of tens of thousands of Stock Keeping Units (SKUs) at hundreds of retail locations, and log minute-by-minute sales trends. Delivering these massive datasets to a community of different business users for simultaneous analyses gives new insight and capabilities that never existed in the past: it allows buyers to review purchasing patterns to make more precise decisions regarding product catalog, product specialists to consider alternate means of bundling items together, inventory professionals to allocate shelf space more efficiently at the warehouse, pricing experts to instantaneously adjust prices at different retail locations directly at the shelf, among other uses. The most effective uses of intelligence demand that analytical systems must process, analyze, and deliver results within a defined time window.

1.3 LOWERING THE BARRIER TO ENTRY

Enabling business process owners to take advantage of analytics in many new and innovative ways has always appeared to be out of reach for most companies. And the expanding universe of created information has seemed to tantalizingly dangle broad-scale analytics capabilities beyond the reach of those but the largest corporations.

Interestingly, for the most part, much of the technology classified as "big data" is not new. Rather, it is the ability to package these techniques in ways that are accessible to organizations in ways that up until recently had been limited by budget, resource, and skills constraints, which are typical of smaller businesses. What makes the big data

concept so engaging is that emerging technologies enable a broad-scale analytics capability with a relatively low barrier to entry.

As we will see, facets of technology for business intelligence and analytics have evolved to a point at which a wide spectrum of businesses can deploy capabilities that in the past were limited to the largest firms with equally large budgets. Consider the four aspects in Table 1.1.

The changes in the environment make big data analytics attractive to all types of organizations, while the market conditions make it practical. The combination of simplified models for development, commoditization, a wider palette of data management tools, and low-cost utility computing has effectively lowered the barrier to entry, enabling a much wider swath of organizations to develop and test out

Table 1.1 Contrasting Approaches in Adopting High-Performance Capabilities

Aspect	Typical Scenario	Big Data
Application development	Applications that take advantage of massive parallelism developed by specialized developers skilled in high-performance computing, performance optimization, and code tuning	A simplified application execution model encompassing a distributed file system, application programming model, distributed database, and program scheduling is packaged within Hadoop, an open source framework for reliable, scalable, distributed, and parallel computing
Platform	Uses high-cost massively parallel processing (MPP) computers, utilizing high-bandwidth networks, and massive I/O devices	Innovative methods of creating scalable and yet elastic virtualized platforms take advantage of clusters of commodity hardware components (either cycle harvesting from local resources or through cloud-based utility computing services) coupled with open source tools and technology
Data management	Limited to file-based or relational database management systems (RDBMS) using standard row-oriented data layouts	Alternate models for data management (often referred to as NoSQL or "Not Only SQL") provide a variety of methods for managing information to best suit specific business process needs, such as in-memory data management (for rapid access), columnar layouts to speed query response, and graph databases (for social network analytics)
Resources	Requires large capital investment in purchasing high-end hardware to be installed and managed in-house	The ability to deploy systems like Hadoop on virtualized platforms allows small and medium businesses to utilize cloud-based environments that, from both a cost accounting and a practical perspective, are much friendlier to the bottom line

high-performance applications that can accommodate massive data volumes and broad variety in structure and content.

1.4 CONSIDERATIONS

While the market conditions suggest that there is a lowered barrier to entry for implementing big data solutions, it does not mean that implementing these technologies and business processes is a completely straightforward task. There is a steep learning curve for developing big data applications, especially when going the open source route, which demands an investment in time and resources to ensure the big data analytics and computing platform are ready for production. And while it is easy to test-drive some of these technologies as part of an "evaluation," one might think carefully about some key questions before investing a significant amount of resources and effort in scaling that learning curve, such as:

- **Feasibility**: Is the enterprise aligned in a way that allows for new and emerging technologies to be brought into the organization, tested out, and vetted without overbearing bureaucracy? If not, what steps can be taken to create an environment that is suited to the introduction and assessment of innovative technologies?
- **Reasonability**: When evaluating the feasibility of adopting big data technologies, have you considered whether your organization faces business challenges whose resource requirements exceed the capability of the existing or planned environment? If not currently, do you anticipate that the environment will change in the near-, medium- or long-term to be more data-centric and require augmentation of the resources necessary for analysis and reporting?
- **Value**: Is there an expectation that the resulting quantifiable value that can be enabled as a result of big data warrants the resource and effort investment in development and productionalization of the technology? How would you define clear measures of value and methods for measurement?
- **Integrability**: Are there any constraints or impediments within the organization from a technical, social, or political (i.e., policy-oriented) perspective that would prevent the big data technologies from being fully integrated as part of the operational architecture? What steps need to be taken to evaluate the means by which big data can be integrated as part of the enterprise?

- **Sustainability**: While the barrier to entry may be low, the costs associated with maintenance, configuration, skills maintenance, and adjustments to the level of agility in development may not be sustainable within the organization. How would you plan to fund continued management and maintenance of a big data environment?

In Chapter 2, we will begin to scope out the criteria for answering these questions as we explore the types of business problems that are suited to a big data solution.

1.5 THOUGHT EXERCISES

Here are some questions and exercises to ponder before jumping headfirst into a big data project:

- What are the sizes of the largest collections of data to be subjected to capture, storage, and analysis within the organization?
- Detail the five most challenging analytical problems facing your organization. How would any of these challenges be addressed if the volume of data is increased by a factor of 10 and 100, respectively?
- Provide your own definition of what big data means to your organization.
- Develop a justification for big data within your organization in one sentence.
- Develop a single graphic image depicting what you believe to be the impact of increased data volumes and variety.
- Identify three "big data" sources, either within or external to your organization that would be relevant to your business.

This page intentionally left blank

CHAPTER 2

Business Problems Suited to Big Data Analytics

In Chapter 1, we identified some key market drivers for assessing how big data technologies might prove to be beneficial to an organization, including:

- the accelerating growth of data volumes to be consumed;
- the desire to blend both structured and unstructured data;
- lowered barrier to entry for enabling scalable high-performance analytics;
- reducing operational costs by leveraging commodity hardware;
- simplified programming and execution model for scalable applications.

In the past, the ability to acquire and deploy high-performance computing systems was limited to large organizations willing to teeter on the bleeding edge of technology. However, the convergence of the aforementioned market conditions has enhanced the attraction of high-performance computing to many different types of organizations now willing to invest in the effort of designing and implementing big data analytics. This is especially true for those organizations whose budgets were previously too puny to accommodate the investment.

2.1 VALIDATING (AGAINST) THE HYPE: ORGANIZATIONAL FITNESS

Even as the excitement around big data analytics reaches a fevered pitch, it remains a technology-driven activity. And as we speculated in Chapter 1, there are a number of factors that need to be considered before making a decision regarding adopting that technology. But all of those factors need to be taken into consideration; just because big data is feasible within the organization, it does not necessarily mean that it is *reasonable*.

Unless there are clear processes for determining the value proposition, there is a risk that it will remain a fad until it hits the disappointment phase of the hype cycle. At that point, hopes may be dashed

Big Data Analytics. DOI: http://dx.doi.org/10.1016/B978-0-12-417319-4.00002-8

when it becomes clear that the basis for the investments in the technology was not grounded in expectations for clear business improvements.

As a way to properly ground any initiatives around big data, one initial task would be to evaluate the organization's fitness as a combination of the five factors presented in Chapter 1: feasibility, reasonability, value, integrability, and sustainability. Table 2.1 provides a sample framework for determining a score for each of these factors ranging from 0 (lowest level) to 4 (highest level).

The resulting scores can be reviewed (an example of a radar chart is shown in Figure 2.1). Each of these variables is, for the most part, somewhat subjective, but there are ways of introducing a degree of objectivity, especially when considering the value of big data.

2.2 THE PROMOTION OF THE VALUE OF BIG DATA

That being said, a thoughtful approach must differentiate between hype and reality, and one way to do this is to review the difference between what is being *said* about big data and what is being *done* with big data. A scan of existing content on the "value of big data" sheds interesting light on what is being promoted as the expected result of big data analytics and, more interestingly, how familiar those expectations sound. A good example is provided within an economic study on the value of big data (titled "Data Equity—Unlocking the Value of Big Data"), undertaken and published by the Center for Economics and Business Research (CEBR) that speaks to the cumulative value of:

- optimized consumer spending as a result of improved targeted customer marketing;
- improvements to research and analytics within the manufacturing sectors to lead to new product development;
- improvements in strategizing and business planning leading to innovation and new start-up companies;
- predictive analytics for improving supply chain management to optimize stock management, replenishment, and forecasting;
- improving the scope and accuracy of fraud detection.[1]

[1]Center for Economics and Business Research Ltd. Data equity—unlocking the value of big data, April 2012. Downloaded from <http://www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf> (Last accessed 08-08-13).

Table 2.1 Quantifying Organizational Readiness

Score by Dimension	0	1	2	3	4
Feasibility	Evaluation of new technology is not officially sanctioned	Organization tests new technologies in reaction to market pressure	Organization evaluates and tests new technologies after market evidence of successful use	Organization is open to evaluation of new technology Adoption of technology on an *ad hoc* basis based on convincing business justifications	Organization encourages evaluation and testing of new technology Clear decision process for adoption or rejection Organization supports allocation of time to innovation
Reasonability	Organization's resource requirements for near-, mid-, and long-terms are satisfactorily met	Organization's resource requirements for near- and mid-terms are satisfactorily met, unclear as to whether long-term needs are met	Organization's resource requirements for near-term is satisfactorily met, unclear as to whether mid- and long-term needs are met	Business challenges are expected to have resource requirements in the mid- and long-terms that will exceed the capability of the existing and planned environment	Business challenges have resource requirements that clearly exceed the capability of the existing and planned environment Organization's go-forward business model is highly information-centric
Value	Investment in hardware resources, software tools, skills training, and ongoing management and maintenance exceeds the expected quantifiable value	The expected quantifiable value widely is evenly balanced by an investment in hardware resources, software tools, skills training, and ongoing management and maintenance	Selected instances of perceived value may suggest a positive return on investment	Expectations for some quantifiable value for investing in limited aspects of the technology	The expected quantifiable value widely exceeds the investment in hardware resources, software tools, skills training, and ongoing management and maintenance
Integrability	Significant impediments to incorporating any nontraditional technology into environment	Willingness to invest effort in determining ways to integrate technology, with some successes	New technologies can be integrated into the environment within limitations and with some level of effort	Clear processes exist for migrating or integrating new technologies, but require dedicated resources and level of effort	No constraints or impediments to fully integrate technology into operational environment

(*Continued*)

Table 2.1 (Continued)

Score by Dimension	0	1	2	3	4
Sustainability	No plan in place for acquiring funding for ongoing management and maintenance costs No plan for managing skills inventory	Continued funding for maintenance and engagement is given on an *ad hoc* basis Sustainability is at risk on a continuous basis	Need for year-by-year business justifications for continued funding	Business justifications ensure continued funding and investments in skills	Program management office effective in absorbing and amortizing management and maintenance costs Program for continuous skills enhancement and training

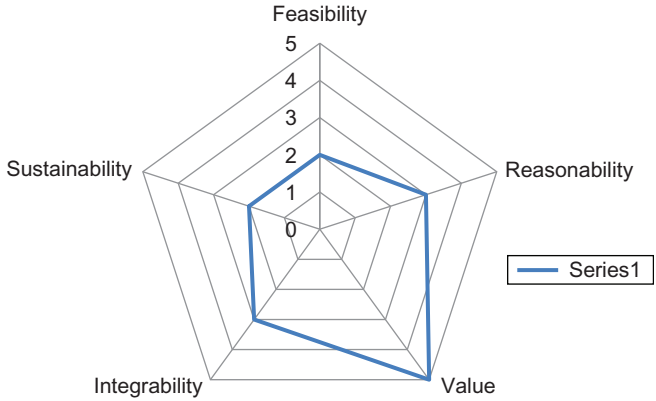

Figure 2.1 Radar chart example of readiness scores.

Curiously, these are exactly the same types of benefits promoted by business intelligence and data warehouse tools vendors and system integrators for the past 15–20 years, namely:

- Better targeted customer marketing
- Improved product analytics
- Improved business planning
- Improved supply chain management
- Improved analysis for fraud, waste, and abuse

Further articles, papers, and vendor messaging on big data reinforce these presumptions, but if these were the same improvements promised by wave after wave of new technologies, what makes big data different?

2.3 BIG DATA USE CASES

The answer must lie in the "democratization" of high-performance capabilities, which is inherent in the characteristics of the big data analytics application development environment. This environment largely consists of a methodology for elastically harnessing parallel computing resources and distributed storage, scalable performance management, along with data exchange via high-speed networks.

The result is improved performance and scalability, and we can examine another data point that provides self-reported descriptions using big data techniques, namely, the enumeration of projects listed at The Apache Software Foundation's PoweredBy Hadoop Web site (http://wiki.apache.org/hadoop/PoweredBy).

A scan of the list allows us to group most of those applications into these categories:

- **Business intelligence, querying, reporting, searching**, including many implementation of searching, filtering, indexing, speeding up aggregation for reporting and for report generation, trend analysis, search optimization, and general information retrieval.
- **Improved performance for common data management operations**, with the majority focusing on log storage, data storage and archiving, followed by sorting, running joins, extraction/transformation/loading (ETL) processing, other types of data conversions, as well as duplicate analysis and elimination.
- **Non-database applications**, such as image processing, text processing in preparation for publishing, genome sequencing, protein sequencing and structure prediction, web crawling, and monitoring workflow processes.
- **Data mining and analytical applications**, including social network analysis, facial recognition, profile matching, other types of text analytics, web mining, machine learning, information extraction, personalization and recommendation analysis, ad optimization, and behavior analysis.

In turn, the core capabilities that are implemented using the big data application can be further abstracted into more fundamental categories:

- **Counting** functions applied to large bodies of data that can be segmented and distributed among a pool of computing and storage resources, such as document indexing, concept filtering, and aggregation (counts and sums).
- **Scanning** functions that can be broken up into parallel threads, such as sorting, data transformations, semantic text analysis, pattern recognition, and searching.
- **Modeling** capabilities for analysis and prediction.
- **Storing** large datasets while providing relatively rapid access.

Generally, **Processing** applications can combine these core capabilities in different ways.

2.4 CHARACTERISTICS OF BIG DATA APPLICATIONS

What is interesting to note is that most of the applications reported by Hadoop users are not necessarily *new* applications. Rather, there are many familiar applications, except that the availability of a low-cost high-performance computing framework either allows more users to develop these applications, run larger deployments, or speed up the execution time. This, coupled with a further review of the different types of applications, suggests that of the limited scenarios discussed as big data success stories, the big data approach is mostly suited to addressing or solving business problems that are subject to one or more of the following criteria:

1. **Data throttling**: The business challenge has an existing solutions, but on traditional hardware, the performance of a solution is throttled as a result of data accessibility, data latency, data availability, or limits on bandwidth in relation to the size of inputs.
2. **Computation-restricted throttling**: There are existing algorithms, but they are heuristic and have not been implemented because the expected computational performance has not been met with conventional systems.
3. **Large data volumes**: The analytical application combines a multitude of existing large datasets and data streams with high rates of data creation and delivery.

4. **Significant data variety**: The + +data in the different sources vary in structure and content, and some (or much) of the data is unstructured.
5. **Benefits from data parallelization**: Because of the reduced data dependencies, the application's runtime can be improved through task or thread-level parallelization applied to independent data segments.

So what, how does this relate to business problems whose solutions are suited to big data analytics applications? These criteria can be used to assess the degree to which business problems are suited to big data technology. As a prime example, ETL processing is hampered by data throttling and computation throttling, can involve large data volumes, may consume a variety of different types of datasets, and can benefit from data parallelization. This is the equivalent of a big data "home run" application!

More examples are given in Table 2.2.

2.5 PERCEPTION AND QUANTIFICATION OF VALUE

So far we have looked at two facets of the appropriateness of big data, with the first being organizational fitness and the second being suitability of the business challenge. The third facet must also be folded into the equation, and that is big data's contribution to the organization. In essence, these facets drill down into the question of value and whether using big data significantly contributes to adding value to the organization by:

- Increasing revenues: As an example, an expectation of using a recommendation engine would be to increase same-customer sales by adding more items into the market basket.
- Lowering costs: As an example, using a big data platform built on commodity hardware for ETL would reduce or eliminate the need for more specialized servers used for data staging, thereby reducing the storage footprint and reducing operating costs.
- Increasing productivity: Increasing the speed for the pattern analysis and matching done for fraud analysis helps to identify more instances of suspicious behavior faster, allowing for actions to be taken more quickly and transform the organization from being focused on recovery of funds to proactive prevention of fraud.

Table 2.2 Examples of Applications Suited to Big Data Analytics

Application	Characteristic	Sample Data Sources
Energy network monitoring and optimization	Data throttling	Sensor data from smart meters and network components
	Computation throttling	
	Large data volumes	
Credit fraud detection	Data throttling	Point-of-sale data
	Computation throttling	Customer profiles
	Large data volumes	Transaction histories
	Parallelization	Predictive models
	Data variety	
Data profiling	Large data volumes Parallelization	Sources selected for downstream repurposing
Clustering and customer segmentation	Data throttling	Customer profiles
	Computation throttling	Transaction histories
	Large data volumes	Enhancement datasets
	Parallelization	
	Data variety	
Recommendation engines	Data throttling	Customer profiles
	Computation throttling	Transaction histories
	Large data volumes	Enhancement datasets
	Parallelization	Social network data
	Data variety	
Price modeling	Data throttling	Point-of-sale data
	Computation throttling	Customer profiles
	Large data volumes	Transaction histories
	Parallelization	Predictive models

- Reducing risk: Using a big data platform or collecting many thousands of streams of automated sensor data can provide full visibility into the current state of a power grid, in which unusual events could be rapidly investigated to determine if a risk of an imminent outage can be reduced.

2.6 FORWARD THINKING ABOUT VALUE

While we continue employing big data technologies for developing algorithms and solutions that are new implementations of old algorithms, we must anticipate that there are, or will be, opportunities for new solution paradigms using parallel execution and data distribution in innovative ways. Yet without proper organizational preparedness, neither approach is likely to succeed. In Chapter 3, we will discuss different aspects of corporate readiness in preparation for designing, developing, and implementing big data applications.

2.7 THOUGHT EXERCISES

Given the premise of considering the suitability of your business challenges, here are some questions and exercises to ponder:

- Using the organizational fitness criteria described in this chapter, assess the degree to which your organization is suited to evaluating big data.
- List what you believe to be the three business challenges in your organization that are most suitable candidates for big data.
- For each of those business challenges, list the characteristics that make it suited to big data.
- Who are the people in your organization with experience, knowledge, or training in big data? How well do they understand your business problem?

This page intentionally left blank