Part

I

Business Potential of Big Data

Chapters 1 through 4 set the foundation for driving business strategies with data science. In particular, the Big Data Business Model Maturity Index highlights the realm of what's possible from a business potential perspective by providing a road map that measures the effectiveness of your organization to leverage data and analytics to power your business models.

In This Part

CHAPTER 1

The Big Data Business Mandate

Having trouble getting your senior management team to understand the business potential of big data? Can't get your management leadership to consider big data to be something other than an IT science experiment? Are your line-of-business leaders unwilling to commit themselves to understanding how data and analytics can power their top initiatives?

If so, then this "Big Data Senior Executive Care Package" is for you!

And for a limited time, you get an unlimited license to share this care package with as many senior executives as you desire. But you must act NOW! Become the life of the company parties with your extensive knowledge of how new customer, product, and operational insights can guide your organization's value creation processes. And maybe, just maybe, get a promotion in the process!!

NOTE **All company material referenced in this book comes from public sources and is referenced accordingly.**

Big Data MBA Introduction

The days when business users and business management can relinquish control of data and analytics to IT are over, or at least for organizations that want to survive beyond the immediate term. The big data discussion now needs to focus on how organizations can couple new sources of customer, product, and operational data with advanced analytics (data science) to power their key business processes and elevate their business models. Organizations need to understand that they *do not need a big data strategy as much as they need a business strategy that incorporates big data.*

The *Big Data MBA* challenges the thinking that data and analytics are ancillary or a "bolt on" to the business; that data and analytics *are* someone else's problem. In a growing number of leading organizations, data and analytics are critical to business success and long-term survival. Business leaders and business users reading this book will learn why they must take responsibility for identifying where and how they can apply data and analytics to their businesses—otherwise they put their businesses at risk of being made obsolete by more nimble, data-driven competitors.

The *Big Data MBA* introduces and describes concepts, techniques, methodologies, and hand-on exercises to guide you as you seek to address the big data *business mandate.* The book provides hands-on exercises and homework assignments to make these concepts and techniques come to life for your organization. It provides recommendations and actions that enable your organization to start today. And in the process, *Big Data MBA* teaches you to "think like a data scientist."

The Forrester study "Reset on Big Data" (Hopkins et al., 2014)[1] highlights the critical role of a business-centric focus in the big data discussion. The study argues that technology-focused executives within a business will think of big data as a technology and fail to convey its importance to the boardroom.

Businesses of all sizes must reframe the big data conversation with the business leaders in the boardroom. The critical and difficult big data question that business leaders must address is:

> *How effective is our organization at integrating data and analytics into our business models?*

Before business leaders can begin these discussions, organizations must understand their current level of big data maturity. Chapter 2 discusses in detail the "Big Data Business Model Maturity Index" (see Figure 1-1). The Big Data Business Model Maturity Index is a measure of how effective an organization is at integrating data and analytics to power their business model.

[1] Hopkins, Brian, Fatemeh Khatibloo with Kyle McNabb, James Staten, Andras Cser, Holger Kisker, Ph.D., Leslie Owens, Jennifer Belissent, Ph.D., Abigail Komlenic, "Reset On Big Data: Embrace Big Data to Engage Customers at Scale," Forrester Research, 2014.

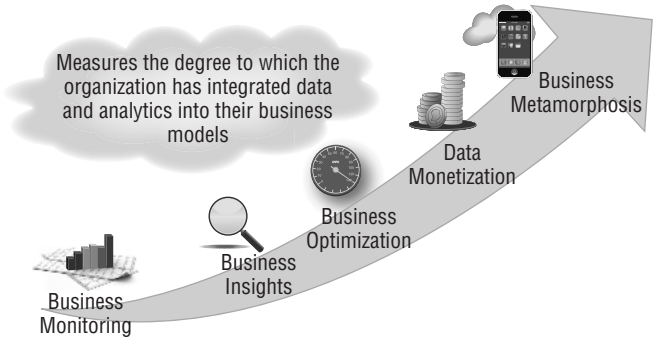

Figure 1-1: Big Data Business Model Maturity Index

The Big Data Business Model Maturity Index provides a road map for how organizations can integrate data and analytics into their business models. The Big Data Business Model Maturity Index is composed of the following five phases:

- **Phase 1: Business Monitoring.** In the Business Monitoring phase, organizations are leveraging data warehousing and Business Intelligence to monitor the organization's performance.
- **Phase 2: Business Insights.** The Business Insights phase is about leveraging predictive analytics to uncover customer, product, and operational insights buried in the growing wealth of internal and external data sources. In this phase, organizations aggressively expand their data acquisition efforts by coupling all of their detailed transactional and operational data with internal data such as consumer comments, e-mail conversations, and technician notes, as well as external and publicly available data such as social media, weather, traffic, economic, demographics, home values, and local events data.
- **Phase 3: Business Optimization.** In the Business Optimization phase, organizations apply prescriptive analytics to the customer, product, and operational insights uncovered in the Business Insights phase to deliver actionable insights or recommendations to frontline employees, business managers, and channel partners, as well as customers. The goal of the Business Optimization phase is to enable employees, partners, and customers to optimize their key decisions.
- **Phase 4: Data Monetization.** In the Data Monetization phase, organizations leverage the customer, product, and operational insights to create new sources of revenue. This could include selling data—or insights—into new markets (a cellular phone provider selling customer behavioral data to advertisers), integrating analytics into products and services to create

"smart" products, or re-packaging customer, product, and operational insights to create new products and services, to enter new markets, and/or to reach new audiences.

- **Phase 5: Business Metamorphosis.** The holy grail of the Big Data Business Model Maturity Index is when an organization transitions its business model from selling products to selling "business-as-a-service." Think GE selling "thrust" instead of jet engines. Think John Deere selling "farming optimization" instead of farming equipment. Think Boeing selling "air miles" instead of airplanes. And in the process, these organizations will create a platform enabling third-party developers to build and market solutions on top of the organization's business-as-a-service business model.

Ultimately, big data only matters if it helps organizations make more money and improve operational effectiveness. Examples include increasing customer acquisition, reducing customer churn, reducing operational and maintenance costs, optimizing prices and yield, reducing risks and errors, improving compliance, improving the customer experience, and more.

No matter the size of the organization, *organizations don't need a big data strategy as much as they need a business strategy that incorporates big data.*

Focus Big Data on Driving Competitive Differentiation

I'm always confused about how organizations struggle to differentiate between technology investments that drive competitive parity and those technology investments that create unique and compelling competitive differentiation. Let's explore this difference in a bit more detail.

Competitive parity is achieving similar or same operational capabilities as those of your competitors. It involves leveraging industry best practices and prepackaged software to create a baseline that, at worst, is equal to the operational capabilities across your industry. Organizations end up achieving competitive parity when they buy foundational and undifferentiated capabilities from enterprise software packages such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), and Sales Force Automation (SFA).

Competitive differentiation is achieved when an organization leverages people, processes, and technology to create applications, programs, processes, etc., that differentiate its products and services from those of its competitors in ways that add unique value for the end customer and create competitive differentiation in the marketplace.

Leading organizations should seek to "buy" foundational and undifferentiated capabilities but "build" what is differentiated and value-added for their customers. But sometimes organizations get confused between the two.

Let's call this the *ERP effect*. ERP software packages were sold as a software solution that would make everyone more profitable by delivering operational excellence. But when everyone is running the same application, what's the source of the competitive differentiation?

Analytics, on the other hand, enables organizations to uniquely optimize their key business processes, drive a more engaging customer experience, and uncover new monetization opportunities with unique insights that they gather about their customers, products, and operations.

Leveraging Technology to Power Competitive Differentiation

While most organizations have invested heavily in ERP-type operational systems, far fewer have been successful in leveraging data and analytics to build strategic applications that provide unique value to their customers and create competitive differentiation in the marketplace. Here are some examples of organizations that have invested in building differentiated capabilities by leveraging new sources of data and analytics:

- Google: PageRank and Ad Serving
- Yahoo: Behavioral Targeting and Retargeting
- Facebook: Ad Serving and News Feed
- Apple: iTunes
- Netflix: Movie Recommendations
- Amazon: "Customers Who Bought This Item," 1-Click ordering, and Supply Chain & Logistics
- Walmart: Demand Forecasting, Supply Chain Logistics, and Retail Link
- Procter & Gamble: Brand and Category Management
- Federal Express: Critical Inventory Logistics
- American Express and Visa: Fraud Detection
- GE: Asset Optimization and Operations Optimization (Predix)

None of these organizations bought these strategic, business-differentiating applications off the shelf. They understood that it was necessary to provide differentiated value to their internal and external customers, and they leveraged data and analytics to build applications that delivered competitive differentiation.

History Lesson on Economic-Driven Business Transformation

More than anything else, the driving force behind big data is the economics of big data—it's 20 to 50 times cheaper to store, manage, and analyze data than it is

to use traditional data warehousing technologies. This 20 to 50 times economic impact is courtesy of commodity hardware, open source software, an explosion of new open source tools coming out of academia, and ready access to free online training on topics such as big data architectures and data science. A client of mine in the insurance industry calculated a 50X economic impact. Another client in the health care industry calculated a 49X economic impact (they need to look harder to find that missing 1X).

History has shown that the most significant technology innovations are ones that drive economic change. From the printing press to interchangeable parts to the microprocessor, these technology innovations have provided an unprecedented opportunity for the more agile and more nimble organizations to disrupt existing markets and establish new value creation processes.

Big data possesses that same economic potential whether it be to create smart cities, improve the quality of medical care, improve educational effectiveness, reduce poverty, improve safety, reduce risks, or even cure cancer. And for many organizations, the first question that needs to be asked about big data is:

> *How effective is my organization at leveraging new sources of data and advanced analytics to uncover new customer, product, and operational insights that can be used to differentiate our customer engagement, optimize key business processes, and uncover new monetization opportunities?*

Big data is nothing new, especially if you view it from the proper perspective. While the popular big data discussions are around "disruptive" technology innovations like Hadoop and Spark, the real discussion should be about the economic impact of big data. New technologies don't disrupt business models; it's what organizations do with these new technologies that disrupts business models and enables new ones. Let's review an example of one such economic-driven business transformation: the steam engine.

The steam engine enabled urbanization, industrialization, and the conquering of new territories. It literally shrank distance and time by reducing the time required to move people and goods from one side of a continent to the other. The steam engine enabled people to leave low-paying agricultural jobs and move into cities for higher-paying manufacturing and clerical jobs that led to a higher standard of living.

For example, cities such as London shot up in terms of population. In 1801, before the advent of George Stephenson's Rocket steam engine, London had 1.1 million residents. After the invention, the population of London more than doubled to 2.7 million residents by 1851. London transformed the nucleus of society from small tight-knit communities where textile production and agriculture were prevalent into big cities with a variety of jobs. The steam locomotive provided quicker transportation and more jobs, which in turn brought more people into the cities and drastically changed the job market. By 1861, only 2.4

percent of London's population was employed in agriculture, while 49.4 percent were in the manufacturing or transportation business. The steam locomotive was a major turning point in history as it transformed society from largely rural and agricultural into urban and industrial.[2]

Table 1-1 shows other historical lessons that demonstrate how technology innovation created economic-driven business opportunities.

Table 1-1: Exploiting Technology Innovation to Create Economic-Driven Business Opportunities

TECHNOLOGY INNOVATION	ECONOMIC IMPACT
Printing Press	Expanded literacy (simplified knowledge capture and enabled knowledge dissemination and the education of the masses)
Interchangeable Parts	Drove the standardization of manufacturing parts and fueled the industrial revolution
Steam Engine (Railroads and Steamboats)	Sparked urbanization (drove transition from agricultural to manufacturing-centric society)
Internal Combustion Engine	Triggered suburbanization (enabled personal mobility, both geographically and socially)
Interstate Highway System	Foundation for interstate commerce (enabled regional specialization and wealth creation)
Telephone	Democratized communications (by eliminating distance and delays as communications issues)
Computers	Automated common processes (thereby freeing humans for more creative engagement)
Internet	Gutted cost of commerce and knowledge sharing (enabled remote workforce and international competition)

This brings us back to big data. All of these innovations share the same lesson: it wasn't the technology that was disruptive; it was how organizations leveraged the technology to disrupt existing business models and enabled new ones.

[2] http://railroadandsteamengine.weebly.com/impact.html

Critical Importance of "Thinking Differently"

Organizations have been taught by technology vendors, press, and analysts to think faster, cheaper, and smaller, but they have not been taught to "think *differently*." The inability to think differently is causing organizational alignment and business adoption problems with respect to the big data opportunity. Organizations must throw out much of their conventional data, analytics, and organizational thinking in order to get the maximum value out of big data. Let's introduce some key areas for thinking differently that will be covered throughout this book.

Don't Think Big Data Technology, Think Business Transformation

Many organizations are infatuated with the technical innovations surrounding big data and the three Vs of data: volume, variety, and velocity. But starting with a technology focus can quickly turn your big data initiative into a science experiment. You don't want to be a solution in search of a problem.

Instead, focus on the four Ms of big data: *Make Me More Money* (or if you are a non-profit organization, maybe that's *Make Me More Efficient*). Start your big data initiative with a business-first approach. Identify and focus on addressing the organization's key business initiatives, that is, what the organization is trying to accomplish from a business perspective over the next 9 to 12 months (e.g., reduce supply chain costs, improve supplier quality and reliability, reduce hospital-acquired infections, improve student performance). Break down or decompose this business initiative into the supporting decisions, questions, metrics, data, analytics, and technology necessary to support the targeted business initiative.

CROSS-REFERENCE **This book begins by covering the Big Data Business Model Maturity Index in Chapter 2. The Big Data Business Model Maturity Index helps organizations address the key question:**

How effective is our organization at leveraging data and analytics to power our key business processes and uncover new monetization opportunities?

The maturity index provides a guide or road map with specific recommendations to help organizations advance up the maturity index. Chapter 3 introduces the big data strategy document. The big data strategy document provides a framework for helping organizations identify where and how to start their big data journey from a business perspective.

Don't Think Business Intelligence, Think Data Science

Data science is different from Business Intelligence (BI). Resist the advice to try to make these two different disciplines the same. For example:

- Business Intelligence focuses on reporting what happened (descriptive analytics). Data science focuses on predicting what is likely to happen (predictive analytics) and then recommending what actions to take (prescriptive analytics).
- Business Intelligence operates with schema on load in which you have to pre-build the data schema before you can load the data to generate your BI queries and reports. Data science deals with schema on query in which the data scientists custom design the data schema based on the hypothesis they want to test or the prediction that they want to make.

Organizations that try to "extend" their Business Intelligence capabilities to encompass big data will fail. That's like stating that you're going to the moon, then climbing a tree and declaring that you are closer. Unfortunately, you can't get to the moon from the top of a tree. Data science is a new discipline that offers compelling, business-differentiating capabilities, especially when coupled with Business Intelligence.

CROSS-REFERENCE **Chapter 5 ("Differences Between Business Intelligence and Data Science") discusses the differences between Business Intelligence and data science and how data science can complement your Business Intelligence organization. Chapter 6 ("Data Science 101") reviews several different analytic algorithms that your data science team might use and discusses the business situations in which the different algorithms might be most appropriate.**

Don't Think Data Warehouse, Think Data Lake

In the world of big data, Hadoop and HDFS is a game changer; it is fundamentally changing the way organizations think about storing, managing, and analyzing data. And I don't mean Hadoop as yet another data source for your data warehouse. I'm talking about Hadoop and HDFS as the *foundation* for your data and analytics environments—to take advantage of the massively parallel processing, cheap scale-out data architecture that can run hundreds, thousands, or even tens of thousands of Hadoop nodes.

We are witnessing the dawn of the age of the *data lake*. The data lake enables organizations to gather, manage, enrich, and analyze many new sources of data, whether structured or unstructured. The data lake enables organizations

to treat data as an organizational asset to be gathered and nurtured versus a cost to be minimized.

Organizations need to treat their reporting environments (traditional BI and data warehousing) and analytics (data science) environments differently. These two environments have very different characteristics and serve different purposes. The data lake can make both of the BI and data science environments more agile and more productive (Figure 1-2).

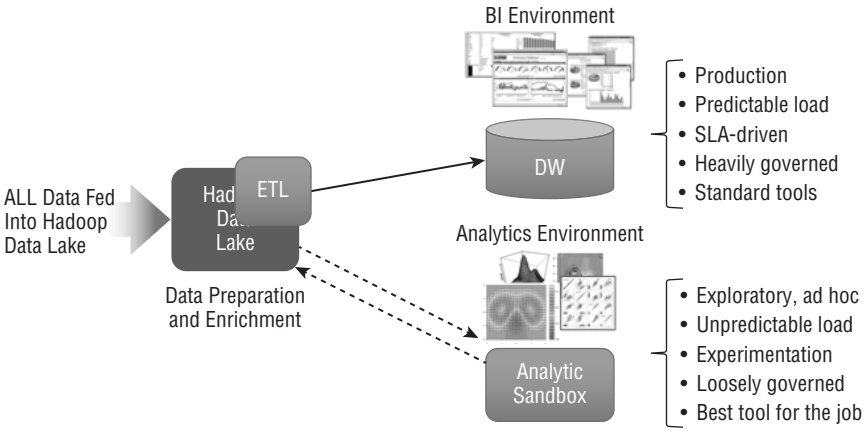

Figure 1-2: Modern data/analytics environment

CROSS-REFERENCE **Chapter 7 ("The Data Lake") introduces the concept of a data lake and the role the data lake plays in supporting your existing data warehouse and Business Intelligence investments while providing the foundation for your data science environment. Chapter 7 discusses how the data lake can un-cuff your data scientists from the data warehouse to uncover those variables and metrics that might be better predictors of business performance. It also discusses how the data lake can free up expensive data warehouse resources, especially those resources associated with Extract, Transform, and Load (ETL) data processes.**

Don't Think "What Happened," Think "What Will Happen"

Business users have been trained to contemplate business questions that monitor the current state of the business and to focus on retrospective reporting on what happened. Business users have become conditioned by their BI and data warehouse environments to only consider questions that report on current business

performance, such as "How many widgets did I sell last month?" and "What were my gross sales last quarter?"

Unfortunately, this retrospective view of the business doesn't help when trying to make decisions and take action about future situations. We need to get business users to "think differently" about the types of questions they can ask. We need to move the business investigation process beyond the performance monitoring questions to the predictive (e.g., What will likely happen?) and prescriptive (e.g., What should I do?) questions that organizations need to address in order to optimize key business processes and uncover new monetization opportunities (see Table 1-2).

Table 1-2: Evolution of the Business Questions

WHAT HAPPENED? (DESCRIPTIVE/BI)	WHAT WILL HAPPEN? (PREDICTIVE ANALYTICS)	WHAT SHOULD I DO? (PRESCRIPTIVE ANALYTICS)
How many widgets did I sell last month?	How many widgets will I sell next month?	Order [5,0000] units of Component Z to support widget sales for next month
What were sales by zip code for Christmas last year?	What will be sales by zip code over this Christmas season?	Hire [Y] new sales reps by these zip codes to handle projected Christmas sales
How many of Product X were returned last month?	How many of Product X will be returned next month?	Set aside [$125K] in financial reserve to cover Product X returns
What were company revenues and profits for the past quarter?	What are projected company revenues and profits for next quarter?	Sell the following product mix to achieve quarterly revenue and margin goals
How many employees did I hire last year?	How many employees will I need to hire next year?	Increase hiring pipeline by 35 percent to achieve hiring goals

CROSS-REFERENCE **Chapter 8 ("Thinking Like a Data Scientist") differentiates between descriptive analytics, predictive analytics, and prescriptive analytics. Chapters 9, 10, and 11 then introduce several techniques to help your business users identify the predictive ("What will happen?") and prescriptive ("What should I do?") questions that they need to more effectively drive the business. Yeah, this will mean lots of Post-it notes and whiteboards, my favorite tools.**

Don't Think HIPPO, Think Collaboration

Unfortunately, today it is still the HIPPO—the Highest Paid Person's Opinion—that determines most of the business decisions. Reasons such as "We've always done things that way" or "My years of experience tell me …" or "This is what the CEO wants …" are still given as reasons for why the HIPPO needs to drive the important business decisions.

Unfortunately, that type of thinking has led to siloed data fiefdoms, siloed decisions, and an un-empowered and frustrated business team. Organizations need to think differently about how they empower all of their employees. Organizations need to find a way to promote and nurture creative thinking and groundbreaking ideas across all levels of the organization. There is no edict that states that the best ideas only come from senior management.

The key to big data success is empowering cross-functional collaboration and exploratory thinking to challenge long-held organizational rules of thumb, heuristics, and "gut" decision making. The business needs an approach that is inclusive of all the key stakeholders—IT, business users, business management, channel partners, and ultimately customers. The business potential of big data is only limited by the creative thinking of the organization.

CROSS-REFERENCE **Chapter 13 ("Power of Envisioning") discusses how the BI and data science teams can collaborate to brainstorm, test, and refine new variables that might be better predictors of business performance. We will introduce several techniques and concepts that can be used to drive collaboration between the business and IT stakeholders and ultimately help your data science team uncover new customer, product, and operational insights that lead to better business performance. Chapter 14 ("Organizational Ramifications") introduces organizational ramifications, especially the role of Chief Data Monetization Officer (CDMO).**

Summary

Big data is interesting from a technology perspective, but the real story for big data is how organizations of different sizes are leveraging data and analytics to power their business models. Big data has the potential to uncover new customer, product, and operational insights that organizations can use to optimize key business processes, improve customer engagement, uncover new monetization opportunities, and re-wire the organization's value creation processes.

As discussed in this chapter, organizations need to understand that big data is about business transformation and business model disruption. There will be winners and there will be losers, and having business leadership sit back and wait for IT to solve the big data problems for them quickly classifies into which

group your organization will likely fall. Senior business leadership needs to determine where and how to leverage data and analytics to power your business models before a more nimble competitor or a hungrier competitor disintermediates your business.

To realize the financial potential of big data, business leadership must make big data a top business priority, not just a top IT priority. Business leadership must actively participate in determining where and how big data can deliver business value, and the business leaders must be front and center in leading the integration of the resulting analytic insights into the organization's value creation processes.

For leading organizations, big data provides a once-in-a-lifetime business opportunity to build key capabilities, skills, and applications that optimize key business processes, drive a more compelling customer experience, uncover new monetization opportunities, and drive competitive differentiation. Remember: buy for parity, but build for competitive differentiation.

At its core, big data is about economic transformation. Big data should not be treated like just another technology science experiment. History is full of lessons of how organizations have been able to capitalize on economics-driven business transformations. Big data provides one of those economic "Forrest Gump" moments where organizations are fortunate to be at the right place at the right time. Don't miss this opportunity.

Finally, organizations have been taught to think cheaper, smaller, and faster, but they have not been taught to think differently, and that's exactly what's required if you want to exploit the big data opportunity. Many of the data and analytics best practices that have been taught over the past several decades no longer hold true. Understand what has changed and learn to think differently about how your organization leverages data and analytics to deliver compelling business value.

In summary, business leadership needs to lead the big data initiative, to step up and make big data a top business mandate. If your business leaders don't take the lead in identifying where and how to integrate big data into your business models, then you risk being disintermediated in a marketplace where more agile, hungrier competitors are learning that data and analytics can yield compelling competitive differentiation.

Homework Assignment

Use the following exercises to apply what you learned in this chapter.

Exercise #1: Identify a key business initiative for your organization, something the business is trying to accomplish over the next 9 to 12 months. It might be something like improve customer retention, optimize customer

acquisition, reduce customer churn, optimize predictive maintenance, reduce revenue theft, and so on.

Exercise #2: Brainstorm and write down what (1) customer, (2) product, and (3) operational insights your organization would like to uncover in order to support the targeted business initiative. Start by capturing the different types of descriptive, predictive, and prescriptive questions you'd like to answer about the targeted business initiative. Tip: Don't worry about whether or not you have the data sources you need to derive the insights you want (yet).

Exercise #3: Brainstorm and write down data sources that might be useful in uncovering those key insights. Look both internally and externally for interesting data sources that might be useful. Tip: Think outside the box and imagine that you could access any data source in the world.