

Introduction

Welcome to *Big Data For Dummies*. Big data is becoming one of the most important technology trends that has the potential for dramatically changing the way organizations use information to enhance the customer experience and transform their business models. How does a company go about using data to the best advantage? What does it mean to transform massive amounts of data into knowledge? In this book, we provide you with insights into how technology transitions in software, hardware, and delivery models are changing the way that data can be used in new ways.

Big data is not a single market. Rather, it is a combination of data-management technologies that have evolved over time. Big data enables organizations to store, manage, and manipulate vast amounts of data at the right speed and at the right time to gain the right insights. The key to understanding big data is that data has to be managed so that it can meet the business requirement a given solution is designed to support. Most companies are at an early stage with their big data journey. Many companies are experimenting with techniques that allow them to collect massive amounts of data to determine whether hidden patterns exist within that data that might be an early indication of an important change. Some data may indicate that customer buying patterns are changing or that new elements are in the business that need to be addressed before it is too late.

As companies begin to evaluate new types of big data solutions, many new opportunities will unfold. For example, manufacturing companies may be able to monitor data

coming from machine sensors to determine how processes need to be modified before a catastrophic event happens. It will be possible for retailers to monitor data in real time to upsell customers related products as they are executing a transaction. Big data solutions can be used in healthcare to determine the cause of an illness and provide a physician with guidance on treatment options.

Big data is not an isolated solution, however. Implementing a big data solution requires that the infrastructure be in place to support the scalability, distribution, and management of that data. Therefore, it is important to put both a business and technical strategy in place to make use of this important technology trend.

For many important reasons, we think that it is important for you to understand big data technologies and know the ways that companies are using emerging technologies such as Hadoop, MapReduce, and new database engines to transform the value of their data. We wrote this book to provide a perspective on what big data is and how it's changing the way that organizations can leverage more data than was possible in the past. We think that this book will give you the context to make informed decisions.

About This Book

Big data is new to many people, so it requires some investigation and understanding of both the technical and business requirements. Many different people need knowledge about big data. Some of you want to delve into the technical details, while others want to understand the economic implications of making use of big data

technologies. Other executives need to know enough to be able to understand how big data can affect business decisions. Implementing a big data environment requires both an architectural and a business approach — and lots of planning.

No matter what your goal is in reading this book, we address the following issues to help you understand big data and the impact it can have on your business:

✔ What is the architecture for big data? How can you manage huge volumes of data without causing major disruptions in your data center?

✔ When should you integrate the outcome of your big data analysis with your data warehouse?

✔ What are the implications of security and governance on the use of big data? How can you keep your company safe?

✔ What is the value of different data technologies, and when should you consider them as part of your big data strategy?

✔ What types of data sources can you take advantage of with big data analytics? How can you apply different types of analytics to business problems?

Foolish Assumptions

Try as we might to be all things to all people, when it came to writing this book, we had to pick who we thought would be most interested in *Big Data For Dummies*. Here's who we think you are:

✔ **You're smart.** You're no dummy, yet the topic of big data gives you an uneasy feeling. You can't quite get your head around it, and if you're pressed for a definition, you might try to change the subject.

✔ **You're a businessperson who wants little or nothing to do with technology.** But you live in the 21st century, so you can't escape it. People are saying, "It's all about big data," so you think that you better find out what they're talking about.

✔ **You're an IT person who knows a heck of a lot about technology.** The thing is, you're new to big data. Everybody says it's something different. Once and for all, you want the whole picture.

Whoever you are, welcome. We're here to help.

How This Book Is Organized

We divided our book into seven parts for easy reading. Feel free to skip about.

Part I: Getting Started with Big Data

In this part, we explain the basic concepts you need for a full understanding of big data, from both a technical and a business perspective. We also introduce you to the major concepts and components so that you can hold your own in any meaningful conversation about big data.

Part II: Technology Foundations for Big Data

Part II is for both technical and business professionals who need to understand the different types of big data components and the underlying technology concepts that support big data. In this section, we give you an understanding about the type of infrastructure that will make big data more practical.

Part III: Big Data Management

Part III is for both technical and business professionals, but it gets into a lot more of the details of different database options and emerging technologies such as MapReduce and Hadoop. Understanding these underlying technologies can help you understand what is behind this important trend.

Part IV: Analytics and Big Data

How do you analyze the massive amounts of data that become part of your big data infrastructure? In this part of the book, we go deeper into the different types of analytics that are helpful in getting real meaning from your data. This part helps you think about ways that you can turn big data into action for your business.

Part V: Big Data Implementation

This part gets to the details of what it means to actually manage data, including issues such as operationalizing your data and protecting the security and privacy of that data. This section gives you plenty to think about in this critical area.

Part VI: Big Data Solutions in the Real World

In this section, you get an understanding of how companies are beginning to use big data to transform their business operations. If you want to get a peek into the future at what you might be able to do with data, this section is for you.

Part VII: The Part of Tens

If you're new to the *For Dummies* treasure-trove, you're no doubt unfamiliar with The Part of Tens. In this section, Wiley editors torture *For Dummies* authors into creating useful bits of information that are easily accessible in lists containing ten (or so) elucidating elements. We started these chapters kicking and screaming but are ultimately very glad that they're here. After you read through the big data best practices, and the do's and don'ts we provide in The Part of Tens, we think you'll be glad, too.

Glossary

We include a glossary of terms frequently used when people discuss big data. Although we strive to define terms as we introduce them in this book, we think you'll find the glossary a useful resource.

Icons Used in This Book



Pay attention. The bother you save may be your own.



You may be sorry if this little tidbit slips your mind.



With this icon, we mark particularly useful points to pay attention to.



Here you find tidbits for the more technically inclined.

Where to Go from Here

We've created an overview of big data and introduced you to all its significant components. We recommend that you read the first four chapters to give you the context for what big data is about and what technologies are in place to make implementations a reality. The next two chapters introduce you to some of the underlying infrastructure issues that are important to understand. The following eight chapters get into a lot more detail about the different types of data structures that are foundational to big data.

You can read the book from cover to cover, but if you're not that kind of person, we've tried to adhere to the *For Dummies* style of keeping chapters self-contained so that you can go straight to the topics that interest you most. Wherever you start, we wish you well.

Many of these chapters could be expanded into full-length books of their own. Big data and the emerging technology landscape are a big focus for us at Hurwitz & Associates, and we invite you to visit our website and read our blogs and insights at www.hurwitz.com.

Occasionally, John Wiley & Sons, Inc., has updates to its technology books. If this book has technical updates, they will be posted at www.dummies.com/go/bigdatafdupdates.