Chapter 1

Introducing Big Data and Analytics

In This Chapter

- Defining big data and analytics
- Gaining insight into the practice of analytics
- Learning about established vendors in the big data market

Big data is an increasingly valuable asset to business, but its potential will only be realized through effective analysis. This chapter defines big data and analytics, and highlights key vendors in this space.

Defining Big Data and Analytics

Sometimes a quantitative change leads to qualitative change. With more data, you need more efficient data processing. This phenomenon is apparent in information management where the increasing volume and variety of types of data require fundamental change in the way we manage data. In addition to traditional data sources, such as financial management and customer relationship management tools, we now have data available from:

- Online browsing activities and click stream analysis
- System logs capturing information about events across networks, devices, and applications running on IT infrastructure
- Social media sentiment and activity

- A constant stream of product feedback in the form of online reviews
- Mobile device information, including location data and activity of billions of cellphone users
- Third-party vendors who are building more complex and rich sets of data about individuals and organizations

Traditional methods of collecting, filtering, integrating, and analyzing data are insufficient for the scale and complexity of the data available to today's enterprises. Big data demands new kind of infrastructure.

Big data and analytics entails not just a change in volume of data, but also a change in the way we work with the data.

Looking at the emergence of big data and analytics

It is reasonable to ask why traditional business intelligence (BI) type reporting is not sufficient to extract maximum value from big data. After all, BI is a well-established domain with best practices and a 20-plus-year track record of providing value across a range of industries.

BI reporting systems do a number of things well. They are responsive to ad hoc queries against structured data. For example, if a sales analyst needs to know the current quarter sales volumes of a product line compared to last year's sales for the same quarter, she could easily get the data from a data warehouse or data mart reporting tool.

Similarly, if a regional manager wanted to drill down into performance data to identify underperforming stores, he could work with an online analytic processing (OLAP) cube or dimensional data warehouse report to find that data. Each of these scenarios requires isolating a subset of data, such as sales by quarter or performance by store, and then aggregating that subset of data, such as summing sales over time or averaging margins across stores.

Certainly, you could run reports such as these over big data sets. However, you would not uncover correlation and connection within those data sets. To extract more value, you

have to deploy two additional types of analysis: statistics and machine learning.

You can use existing business intelligence tools with big data, but you won't realize the full potential of big data without tools that exploit the unique properties of big data.

Describing and predicting with statistics

Statistics is a branch of mathematics that focuses on describing populations of data and using data to make predictions about future events. Descriptive statistics are widely used in traditional BI reporting. Any time you run a report that shows minimums, maximums, means, or standard deviations, you're working with descriptive statistics. Some examples include

- Personalized recommendations in retail, insurance, and other industries, offering an array of options from which customers can choose
- Predictive maintenance, such as estimating the likely time to failure of a critical component of an industrial machine

The other type of statistics is predictive statistics. These methods use data to build models, which are mathematical formulas for making predictions based on set of data.

Descriptive statistics help understand the characteristics of data sets, while predictive statistics use data sets to make inferences about new instances of similar populations. Describing the size of customer segments and their average marginal revenue, which is a type of descriptive statistics. Predicting the impact of a different sales offers on changes in product sales is an example of predictive statistics.

Finding patterns with machine learning

Machine learning is a field of computer science dedicated to developing algorithms that can identify a wide range of

patterns in data. These patterns come in many forms and include the following:

- Detecting products frequently bought together
- Classifying fraudulent transactions
- Making recommendations to a customer based on the choices of similar customers
- Discerning subgroups, or clusters, of similar customers, transactions, events, or other entities within a large population

The emerging field of big data and analytics employs additional techniques to those found in business intelligence platforms, and this means new tools are needed.

Working with new tools for big data and analytics

In later chapters, we drill down into detail about different big data and analytics tools, but for now it's sufficient to highlight different types of big data tools and their uses. Here are several you'll be hearing more about throughout this guide:

- **Hadoop:** A big data storage and processing platform that is almost synonymous with big data. The term *Hadoop* is also used to describe the broad ecosystem of data management, processing, and visualization projects that work with the core Hadoop platform.
- **NoSQL:** A NoSQL (originally referring to "non-SQL" or "nonrelational") database provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases.

The big data and analytics field is rapidly changing, and new tools are being released frequently. Watch for specialized tools that can help solve your particular problems.

Deploying Big Data and Analytics in the Enterprise

Enterprises have a number of options for building their big data and analytics platform, including both open-source and commercially developed tools. Many organizations will have multiple business use cases that could benefit from big data and analytics. Before deploying big data and analytics in the enterprise, it's important to consider the full range of platform options and the business uses cases that will initially benefit from the new capabilities.

The market for big data platforms and analytics tools is maturing as indicated by the growing number of established vendors with product offerings in this area. In addition, open-source projects continue to create well-designed and efficient applications and support tools for big data and analytics.

Many companies began developing big data and analytics tools to meet their specific internal requirements and then contributed them to the open-source community. That applies to everything from Hadoop itself to NoSQL platforms (such as MongoDB and Cassandra) to analysis and visualization tools.

Infrastructure is a crucial element to the success of big data and analytics. Enterprises should consider platforms that enable highly scalable and highly available infrastructure for big data and analytics applications.

Cisco big data and analytics solutions offerings

Cisco and its partner ecosystem can offer comprehensive solutions for end-to-end big data and analytics.

Cisco UCS Integrated Infrastructure for Big Data integrates industry-leading compute, storage, connectivity, and management capabilities into a unified, fabric-based architecture optimized for big data and analytics workloads. Cisco UCS Integrated Infrastructure for Big Data is a highly efficient, scalable, high-performance solution that can help your organization grow quickly and cost-effectively. Use it to deliver insights faster, and reduce total cost of ownership (TCO). The Cisco innovations allow you to unlock the intelligence in your data to help you create a sustainable, competitive business advantage.

As enterprise big data workloads increase in size and complexity, the network will play a crucial role in ensuring workloads are completed and insights are delivered in a timely fashion. Cisco offers an ideal solution to solve network constraints caused by increasing network traffic: Cisco Application Centric Infrastructure (ACI) can apply policies and dynamically load balance across the application infrastructure for optimal performance.

Chapter 2

The Business Case for Big Data and Analytics

In This Chapter

- Identifying the common use cases of big data and analytics in your enterprise
- Integrating big data and analytics into your application ecosystem

Studies show that big data and analytics can significantly improve business process to customer retention. To get the maximum benefits of big data and analytics, you need to understand the business, understand the customers, and have a long-term plan to monetize ever-growing data.

In this chapter, we offer sample use cases that can enhance customer-focused decision-making. We also describe key aspects of successful implementations and deployments.

Looking at Some Popular Use Cases

From forming sales predictions to offering purchase suggestions to customers, big data can inform a multitude of business concerns. Here are a few use cases to give you an idea of what big data can do for business.

Cutting costs with predictive analysis

Businesses love saving money and making money. One of the best ways to do that with big data is by using predictive analysis. By analyzing large business data sets with predictive analysis, companies can predict, with increasing accuracy, how well certain products will sell, when they will need more of a certain product, and so on. This can cut down on waste, and reduce unsold products sitting in a warehouse for years at a time.

You can take this one step further, however. What would happen if you let your customers know what you were predicting?

Personalizing recommendations for customers

Businesses can use individual customer data, such as purchasing history or location data, to personalize certain marketing practices. For instance, when a customer adds an item to an online shopping cart, the vendor may want to immediately recommend a cross-sell product.

Similarly, when a mobile device user passes a restaurant, the sales generation service may push a coupon to increase the chances of the person stopping in for lunch.

This sort of marketing can also evolve into stopping *customer churn* (customers stopping their patronage of your business). When you can offer personalized product suggestions or coupons, a customer may be more likely to stick around and purchase from you again.

Figure 2-1 shows an example of predictive analytics through data points that correspond to particular values along the *x*-axis and *y*-axis. These could represent the price of a commodity over time, the average sales margin over time, or some other measure that changes over time and is of interest to a business. A method known as *linear regression* can find a line that best fits the historical pattern of data.

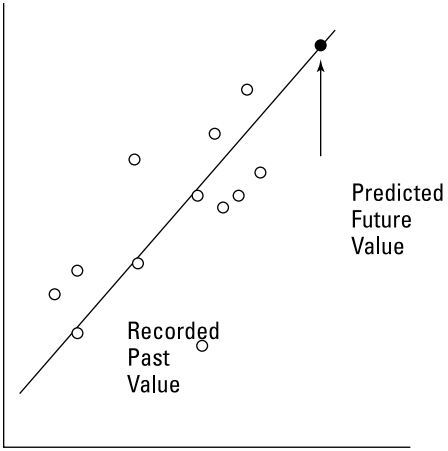

Figure 2-1: Predictive statistics is applied to big data to help make projections about future events and values.

However, it's important not to try to find too close a fit to historical data. The goal is to find a pattern that is useful for predicting future values, not have a model that exactly describes the past. The model that best fits the past may not give enough room to account for measurement errors or variations in your customer data.

Finding an optimal data model is challenging and not always obvious. Be sure to understand your algorithms and their limitations before deploying models to production.

Enhancing patient care with analytics

A major hospital system and research organization started using big data and analytics to improve healthcare delivery. Initial efforts were successful, but the big data and analytics team soon found itself outgrowing its homegrown cluster. To continue to enable leading-edge clinical research, the BI leadership team turned to a Cisco solution that included UCS servers and a Hadoop vendor.

As a result of upgrading their homegrown cluster, clinicians were able to receive information that helped improve patient

outcomes, while IT was able to store and process more clinical data without incurring additional costs.

Big data and analytics can revolutionize the healthcare industry. For example, it can

- Improve quality, safety, efficiency, and reduce health disparities
- Engage patients and family
- Improve care coordination and public health
- Maintain the privacy and security of patient health information
- Enhance clinical outcomes
- Improve population health outcomes
- Increase transparency and efficiency
- Empower individuals
- Generate more robust research data on health systems

Additional popular use cases

Here are some more use cases for big data and analytics:

- Financial services firms can assess risk and achieve compliance by combing data from different sources.
- Pharmaceutical firms can combine data from multiple places to get product to market faster.
- Manufacturing firms can manage inter-train forces to better manage acceleration and braking.
- Food and beverage firms can evaluate online sales data across thousands of stores to gain business and marketing insights.