All About Analytics

Hugh J. Watson, Terry College of Business, University of Georgia, Athens, Georgia, USA

ABSTRACT

To understand and be successful with analytics, it is important to be precise in understanding what analytics means, the different targets or approaches that companies can take to using analytics, and the drivers that lead to the use of analytics. For companies that use advanced analytics, the keys to success include a clear business need; strong, committed sponsorship; a fact-based decision making culture; a strong data infrastructure; the right analytic tools; and strong analytical personnel in an appropriate organizational structure. These are the same factors for success for business intelligence in general, but there are important nuances when implementing advanced analytics, such as with the data infrastructure, analytical tools, and personnel. Companies like Amazon.com, Overstock.com, Harrah's Entertainment, and First American Corporation are exemplars that illustrate concepts and best practices.

Keywords: Analytics, Benefits, Best Practices, Business Intelligence, Case Study, Keys to Success, Trends

INTRODUCTION

There is considerable "buzz" about analytics. It is the topic of numerous articles, books, web seminars, white papers, and research reports. Many academic and practitioner conferences are focusing on analytics. There is growing evidence that analytics is becoming an important component of organizational success. Despite the attention, it is difficult to get a holistic understanding of how analytics can and should be used in organizations. Of particular interest is what organizations need to do in order to be successful with advanced analytics.

A good starting point is to explore how the analytics term is used and its relationship to business intelligence (BI). We will see that analytics is an umbrella term that includes several kinds of analytics. There are also different approaches that organizations can take to analytics. For example, some are developing a single or a few applications while others have business models that are dependent on the use of analytics. Some companies are more analytics-based than others, because of, for example, the nature of the industry.

Most of our attention is focused on what is necessary to succeed with advanced analytics. While many of the factors are similar to those for BI in general, there are considerations within the factors that are critically different, such as the technology and people needed.

BEING PRECISE ABOUT ANALYTICS

The analytics term is used imprecisely. Sometimes it is employed interchangeably with business intelligence. Another interpretation is that if you view BI as "getting data in" (to the warehouse) and "getting data out" (data access and analysis), analytics is the analysis part of

DOI: 10.4018/jbir.2013010102

BI. Or finally, analytics is the "rocket science" algorithms (e.g., neural networks) and methods used to find patterns in data (e.g., customer segmentation analysis) or to optimize performance (e.g., revenue management).

It is useful to think of descriptive, predictive, and prescriptive analytics. With *descriptive analytics*, the objective is to describe *what has occurred*. With this view, reporting, OLAP, dashboards/scorecards, and data visualization are all examples of descriptive analytics. These are the core and most common BI applications.

Predictive analytics focuses on *what will occur* in the future. The algorithms and methods for prescriptive analytics include regression analysis, machine learning, and neural networks. These techniques have been around for some time and have traditionally been called data mining. While these methods continue to evolve, the most significant development is their inclusion in analytical workbenches and applications that make them much easier to use.

Prescriptive analytics is intended to show *what should occur*. It is used to optimize system performance. Revenue management, which strives to optimize the revenue from perishable goods, such as hotel rooms and airline seats, is a good example. Through a combination of forecasting and mathematical programming, prices are dynamically set for the good over time to optimize revenues.

Another perspective is that the progression from descriptive to predictive to prescriptive analytics is a movement from *hindsight to insight to foresight* (Barnes et al., 2012). First companies want to understand the past, then they want to predict the future, and then they want to optimize what they do.

In most cases, imprecise use of the analytics term does not cause difficulties. There is a problem, however, when discussing the requirements for success with analytics. The requirements for descriptive analytics are different in important ways to predictive and prescriptive analytics. We will refer to predictive and prescriptive analytics as *advanced analytics.*

Returning to the issue of the different interpretations of the analytics term, this article uses analytics to describe the analysis of data and advanced analytics as the "rocket science" algorithms and methods of predictive and prescriptive analytics. With this interpretation, analytics is a subset of BI rather than an alternative term.

DIFFERENT TARGETS FOR ANALYTICS

Companies can have different "targets" or approaches to analytics. No one target is better for all firms, and each target can be best for a particular company depending on its situation. All of the targets can potentially deliver significant business value. These are the same targets as for BI (Wixom & Watson, 2010).

One target is to develop a *single or a few analytic applications*. These applications are typically departmental solutions and satisfy specific business needs. For example, a company may use analytics to identify customer segments for more targeted marketing campaigns. These applications are not necessarily developed in-house. There are a growing number of analytic applications and services that are offered by third parties, either as a service over the Internet or on a consulting basis. The services approach is an especially appealing option for smaller companies that do not have the necessary in-house resources for advanced analytics.

A single or a few applications is a common starting point for analytics in most companies. While satisfying a business need, the initial applications can also serve as a proof of concept for analytics. Over time, there are more point solutions and management becomes aware of the need to take a more holistic approach.

With *enterprise-wide analytics*, a company puts resources, organizational structures, and processes around analytics. The infrastructure (e.g., data, software) is created to do analytics on a company-wide basis. Given this infrastructure, analytics is used throughout the organization and is often a key to business success. Later we will discuss the various component parts of this infrastructure. From a BI maturity curve perspective, companies that are doing analytics enterprise-wide have progressed to an advanced stage of maturity (Eckerson, 2004).

The final target, *organizational transformation,* is brought about by either opportunity or necessity. The company adopts a new business model in reaction to market realities and the model cannot be executed without the extensive use of analytics. The business and analytics strategies are intertwined and analytics is a requirement for business success. We will later see examples of this with Harrah's and First American Corporation.

THE PAYOFF FROM ANALYTICS

There is growing evidence that analytics generates considerable business value. It is likely there would not be much excitement about analytics if companies were not benefiting from it. Even the *New York Times* is writing about analytics based decision making (Lohr, 2012). Tom Davenport and Jeanne Harris did much to raise awareness of how companies are using, benefiting, and competing on analytics with their best selling books (Davenport & Harris, 2007; Davenport, Harris, & Morison, 2010). Case studies have documented how companies are using analytics (Wixom & Watson, 2010; Wixom, Watson, & Werner, 2011). Recently, empirical research is documenting the payoff.

In 2010, the IBM Institute for Business Value and the MIT Sloan Management Review surveyed nearly 3,000 executives and business analysts about their companies' use of analytics (LaVallel et al., 2010). Among the study's findings was that top performing companies use analytics five times more than low performing ones. Also, top performing companies are two times more likely to shape future business strategies and guide day-to-day operations based on analytics. A follow up study in 2011 of more than 4,500 executives, managers, and business analysts discovered that the number of companies using analytics to create a competitive advantage had surged by 57 percent in the past year and the performance gap had widened between those companies that rely on advanced analytics and those that don't (Kiron et al., 2011).

Brynjolfsson, Hitt, and Kim (2011) studied 179 large publically traded firms and found that firms that emphasize decision making based on data an analytics have output and productivity that are 5-6% higher than what would be expected without this emphasis. They also found that the relationship between the use of data and analytics appears in other performance measures such as asset utilization, return on equity, and market value.

ANALYTICS ORGANIZATIONS

To varying degrees, companies have used analytics for many years. What has changed is that for many firms the use of analytics has evolved from a "nice to have" to a "requirement for competing in the marketplace." Such firms can be thought of as *analytics organizations* because they rely on advanced analytics for competitive advantage and possibly even survival.

This movement to analytics is driven by multiple factors. At the top of the list is business need. To illustrate, companies are especially interested in understanding customers' behavior, preferences, likely responses to offers and promotions, profitability, and more. More data is now being collected and available for analysis than ever before, including Big Data, such as Web traffic; social media; image, audio, and video; and machine-generated data. A variety of analytical platforms (e.g., appliances) are providing the performance needed to analyze and make use of the available data. Government reporting and compliance requirements associated with the Sarbanes Oxley Act (SOX), the Health Insurance Portability and Accountability Act (HIPPA), and Basel II encourage the use of analytics.

It can be argued that analytics is potentially a sustainable source of competitive advantage (Eckerson, 2011). It helps employees throughout the organization make better decisions. For example, senior executives can see simulations of the economic consequences of possible acquisitions under varying assumptions. Brand managers can investigate the sources of a product's profitability problems. Customer care center operators can be given complete customer profiles and buying preferences that improve customer satisfaction and company profitability. Smarter decisions should lead to more profitable and sustainable organizations.

CONDITIONS THAT LEAD TO THE USE OF ANALYTICS

There are multiple reasons why some companies rely heavily on analytics. Three are especially important: the nature of the industry, seizing a market opportunity, and responding to a problem.

The Nature of the Industry: Complex Systems vs. Volume Operations

Geoffrey Moore, author of *Crossing the Chasm* (2002), *Inside the Tornado* (2004), and other books, makes a distinction between complex systems and volume operations. The distinction helps in understanding what kinds of organizations are most likely to depend on analytics.

Companies that operate in a *complex systems* environment offer products and services for complex problems and provide individualized solutions. The products and services are organized around the needs of individual customers, there is considerable interaction with each customer, and the dollar value of each interaction is high. Examples include Boeing, World Bank, and Halliburton.

Companies with *volume operations* provide standardized products and services. The volume is high but the dollar value of each interaction is low. Because of the large number of customers and standardized products and services, customer interactions are generally conducted through technology rather than person-to-person. Amazon.com, Coke-a-Cola, and Hertz provide examples of volume operations companies.

While analytics is important to complex systems organizations, it is critical to volume operations. It is only through analytics that companies can operate efficiently and have personalized interactions with customers.

Volume Operations Example: Online Retailers: Amazon.com and Overstock.com

Online retailers like Amazon.com and Overstock.com provide excellent examples of high volume operations that rely on analytics to compete. To illustrate, as soon as you enter their web sites, a cookie is placed on your PC and all clicks are recorded. Based on your clicks and any search terms, recommendation engines decide what products to display. After you purchase an item, they have additional information that is used in marketing campaigns. Customer segmentation analysis is used in deciding what promotions to send you. How profitable you are to the retailer influences how the customer care center treats you. A pricing team helps set prices and decides what prices are needed to clear out merchandise. Forecasting models are used to decide how many items to order for inventory. Dashboards monitor all aspects of organizational performance.

I had the opportunity to interview Patrick Byrne, the founder and CEO of Overstock.com, and I asked him to describe his company. I expected him to say Overstock is an "online retailer" or perhaps an "information-based company". Instead, he said that it is a "business intelligence company." Analytics is how Overstock.com competes in the marketplace.

Seizing an Opportunity Example: Harrah's

Harrah's was originally the "blue collar" casino where "everyone knew your name" (Watson & Volonino, 2002). In 1993, the gaming laws changed to allow gambling on Indian reservations and river boats. Harrah's seized the opportunity and adopted a new business model that included expansion (largely through acquisition of other casinos), creating a brand identity, and Total Rewards, the industry's first loyalty program. Gary Loveman, a Harvard marketing professor, was brought in to drive this new business strategy. Analytics was essential to Harrah's approach. Analytics allows Harrah's to know its customers, where they gamble, what games they play, their profitability, and what offers to make to get them to return.

The managers of the Harrah's properties used to run their casinos as private fiefdoms and decisions were made based on "Harrahisms" – things that the managers believed were true. With the new business model, decision making became much more centralized and was based on "the facts" and constant experimentation to see what works best (i.e., fact based decision making). The success of this approach is seen in Harrah's now being the largest gaming company in the world. The "blue collar" casino bought the more upscale Caesars (in 2004) and changed its corporate name to Caesars Entertainment in 2010. Interestingly, many of the people who successfully implemented analytics at Harrah's were hired by other casinos and spread the use of analytics throughout the gaming industry.

Responding to a Problem: First American Corporation

First American was a regional bank headquartered in Nashville, TN (Cooper et al., 2000). It does not currently exist by that name as it has been bought and renamed several times. The bank had lost $60M and was operating under letters of agreement with regulators. It had serious problems. A new management team was brought in that cut costs and "stopped the bleeding." Management knew, however, that it needed a new business strategy if the bank was to survive. As it explored its options (e.g., being a low cost provider), it realized that its best option was a customer intimacy strategy (supported by analytics) because of the information it had about its customers that other banks didn't. This strategy was called Tailored Client Solutions (TCS).

A heavy dose of consultants and professional services were brought in to quickly build a customer-centric data warehouse and create customer and product profitability models. Small teams also worked on developing applications in each of the TCS application areas. For example, one team implemented a customer loyalty program that rewarded customers on the depth and breadth of bank usage (e.g., the number of products used and account balances in each). Another team analyzed and redesigned the products and services offered (e.g., requiring a minimum balance on the unprofitable Seniors Account). The distribution channel of hubs, branches, and ATMs was analyzed and redesigned to increase customer satisfaction and reduce costs. Each of the applications provided financial "lift" and gave credence to the new approach to running the bank. The bank was transformed from "banking by intuition" to banking by "information and analysis." It moved from impending financial disaster to being a leader in the financial services industry through its innovative use of analytics.

THE REQUIREMENTS FOR BEING AN ANALYTICS ORGANIZATION

In most cases, an organization does not become analytics-based overnight. From a *stages-of-growth or maturity curve perspective*, organizations evolve over time. For example, The Data Warehousing Institute's (TDWI) BI maturity model uses a human metaphor with prenatal, infant, child, teenager, adult, and sage stages (Eckerson, 2004). The use of advanced analytics appears in the adult and sage stages. Each stage is defined by a set of characteristics, and movement to the next stage usually requires a balanced movement across all of the characteristics. Continuing with the TDWI model, the characteristics include scope, analytic structure, executive perceptions, types of analytics, stewardship, funding, technology platforms, and change management and administration. The movement to the use of advanced analytics (the "types of analytics" characteristic) should be accompanied by positive changes in the other characteristics in order to help ensure success.

A few companies move quickly to advanced analytics, most typically driven by necessity. This was the case at FAC, where survival of the bank was at stake. To make the jump to an advanced stage, FAC invested heavily in consultants and vendors' products and professional services. The providers of these resources knew what needed to be done and FAC was willing to make the necessary financial and other commitments. Another example is the Cosmopolitan of Las Vegas, the latest hotel/casino to open on the Strip. It entered a market where Harrah's and other leading casinos already had outstanding analytics programs in place. It was able to quickly move into advanced analytics by using vendors that already had gaming specific industry offerings (e.g., the Teradata hospitality/gaming data model) and hiring people with analytics experience in other casinos.

An *analytics readiness or success factors perspective* provides another lens for considering what is required in order to become a successful analytics-based organization. This perspective identifies what is already or should be put in place in order to be successful. Drawing upon Williams' (2004) BI readiness factors, the following can be identified for advanced analytics:

- A clear business need;
- Strong, committed sponsorship;
- Alignment between the business and IT strategy;
- A fact-based decision making;
- A strong data infrastructure;
- The right analytical tools;
- Strong analytical personnel in an appropriate organizational structure.

Let's consider each of them, drawing upon the definitions, concepts, and cases introduced before. Special emphasis is given to advance analytics.

A Clear Business Need

Almost every information technology project should be a response to a business need rather than a technology solution. At Harrah's, the need was supporting its opportunity to expand into new markets, encourage cross play at Harrah's casinos, and increase customer loyalty and profitability. At First American, it was to "save the bank" through a customer intimacy strategy supported by a data warehouse and analytics. In both of these cases, and in many others where analytics is used, the business need is to focus on better understanding and responding to customer needs and preferences.

The driver for analytics in many companies is a specific "organizational pain," a term used by many consulting firms. This pain is often in the marketing, finance, or operations areas. Because finance controls the purse strings, often pays for the data warehouse, and has a need for analytics, it is common for finance to be a major user of analytics.

Strong, Committed Sponsorship

It is well known that if you don't have strong, committed sponsorship, it is difficult to succeed. If the target is a single or a few analytical applications, the sponsorship can be departmental. When the target is enterprise-wide analytics, especially with organizational transformation, the sponsorship needs to be at the highest levels and organization wide. Senior executives must believe in fact-based decision making and constant experimentation rather than intuition.

Strong, committed sponsorship was evident in the case studies. The CEO at Overstock.com considers Overstock to be a BI company. The CEO at First American Corporation "bet the bank" on analytics. The CEO at Harrah's made analytics the basis for competing in the marketplace.

Alignment Between the Business and IT Strategy

The case studies also provide exemplars of close alignment between the business and IT strategy. At Overstock.com, Harrah's, and First American Corporation, the business and analytics strategies were so intertwined that it is impossible to discuss one without the other. The connection might not always be this close, but it is important to make sure that the analytics work is always supporting the business strategy.

A clear business need and strong, committed sponsorship help drive alignment. It is also important to have IT and analytics leaders who understand and are committed to helping drive the business. Another key ingredient is analytics governance. This involves putting people, committees, and processes in place to support alignment. The highest level committee of a multi-tier structure should bring senior management and IT and analytics managers together to discuss the business strategy and how it can be supported by analytics (Watson, Fuller, & Ariyachandra, 2004).

A Fact-Based Decision Making Culture

In companies that rely on analytics, the numbers rather than intuition or supposition drive decision making. Individuals and groups are judged on performance metrics tied to business objects. To illustrate, it was previously mentioned that analytics replaced "Harrahisms." It is also said that three things will get you fired at Harrah's – stealing, sexual harassment, and not basing your decisions on facts. Gary Loveman, Harrah's CEO, is famous for saying, "Do we think, or do we know?"

Another aspect of a fact-based decision making culture is constant experimentation to see what works and what doesn't. For example, at Harrah's different offers to different market segments are tested. Do older customers who live in Jackson, Mississippi respond better to (1) offers of free chips or (2) a complimentary room and tickets to a show at the Tunica property? Experiments like this one are run thousands of time each year, and the findings are integrated into marketing campaigns. At Overstock.com different web site designs and email campaigns are constantly tested. For example, are people more likely to open an email containing an offer if the subject line contains capital letters? Basing marketing campaigns on facts and experimentation is better than the old "spray and pray" approach, where the same generic offer is broadcast to all possible customers.

Many times the culture of an organization emanates from the top. If senior management is basing decisions on data and facts, this approach is likely to filter down throughout the organization. Sometimes, however, top management needs to "kick start" a cultural change, and there are several actions that can be taken.

Recognize that some people can't or won't adjust to the demands of analytical jobs. In these cases, the people either need to be let go or placed in positions that better meets their skills and interests. At First American Corporation, prior to becoming an analytics-based organization, there were 12 people in Marketing. Afterwards, there were 12, but none were the same. As the CEO explained it, "The original group thought that marketing was giving out balloons and suckers along the teller line and running focus groups." They either left the bank or took other positions and were replaced by people who could do the analytical marketing work.

There are other actions that senior management can take. Be a vocal supporter of analytics and stress that outdated methods must be discontinued. Ask to see what analytics went into decisions and link incentives and compensation to desired behaviors.

Not all companies are ready for analytics and a fact-based decision making culture. Trust in analytics is an issue. The 2011 Blumberg Businessweek study found that that just 58% of respondents believe that executive management trusts the results of analytics.

A Strong Data Infrastructure

In the 1990s, I was a consultant to the World Bank and helped build their first executive information system (EIS) (Watson, Houdeshel, & Rainer, 1997). People were surprised and impressed that we were able to roll out the initial version within 90 days from when the project was approved. This was possible because much of the required data was already understood and in place. Though there was not a data warehouse at the time, the previous year the EIS project manager had worked on a bank-wide reporting project. He knew what data was available, who "owned" it, who to work with to access it, and how to access it. This story illustrates the well-known, but always important point that being able to access needed data is critical to a decision support project. For many projects, 70-80 percent of the time and cost is spent on preparing the data. The Bloomberg Businessweek Research Services survey (2011) found data to be the number one challenge that companies find in the adoption of analytics.

Another aspect of the EIS story is germane. An important application within the EIS allowed users to view human resources data in a report format and "slice and dice" the data using dimensions and measures. For example, executives could identify the number of Japanese women who were at a grade level of 20 or above and compare the number this year to last. Because performance (i.e., system response time) was a concern, the decision was made to place the human resources data in a multidimensional database. The lesson is that it may be necessary to use a special-purpose platform in order to get good performance and learn the answers to business questions.

The reason for this example is that it illustrates several truths that apply to data management and analytics: (1) a strong data infrastructure is important, and (2) special-purpose technology may be needed in order to satisfy performance requirements. Let us consider how this plays out in terms of data management to support advanced analytics.

- **Access to Warehouse Data:** In many organizations, users are limited to specific, predefined views of warehouse data (Barnes, 2012). There are tight controls over what data users can see and access. This level of control does not serve analytic modelers (i.e., people who develop advanced analytics models) well. They need access to underlying data structures and the ability to join, transform, and aggregate data in any way necessary for their work. They also need the ability to enter new data to the warehouse. These needs can result in resistance from the "keepers" of the data warehouse, but are important to successful advanced analytics. A possible solution to the potential conflict over control versus flexibility is an analytical sandbox, which is discussed later.
- **Big Data, Advanced Analytics, and New Technologies:** Organizations have long stored large amounts of data in data warehouses, but we have moved into the era of Big Data, which is characterized by high volume, variety, and velocity (the three Vs) (Russom, 2011). The amount of data is so large that it is difficult to work with in traditional database management systems. Big Data is the result of capturing and storing more transaction details, social media, sensing devices on machines (e.g. utility meters), web sites, GPS, image and video, and more. This data is stressing the capabilities of existing technologies and spawning the emergence of new ones. The opportunity is to apply advanced analytics on this data in order to improve organizational performance. For example, rather than running focus groups to understand reactions to a newly introduced product, companies can monitor social media in near real time and make adjustments on the fly. For example, Starbucks introduced a new coffee and within hours lowered the price based on the analysis of social media data.

For analytic modelers, Big Data and the new technologies are welcome developments because they can work with larger data sets, analyze new kinds of data, and include a wider variety of variables in their models. The challenge is to accommodate all of this data while providing good system performance, not only for advanced analytics but also for users of descriptive analytics. To illustrate, many advanced analytics algorithms are very CPU intensive and also consume the entire disk input/output in the system. Consequently, advanced analytics can negatively impact other users of the same system. An executive's dashboard shouldn't refresh slowly because a data mining algorithm is running. In response to the need to handle more data and provide excellent performance, vendors and BI Directors have turned to a variety of new approaches and technologies. Let's consider the most interesting ones.

- **Different Technology Options and Approaches:** One approach is to create an "*analytical sandbox*." As the name implies, this is a place where data is stored and modelers can "play" (i.e., develop and run) their models without disrupting other applications. In many instances, this is a physical implementation where there is a separate server with data loaded from the data warehouse and other sources. (This is the potential solution to the control/flexibility issue mentioned previously). In some instances the sandbox is logical rather than physical. In this case a logical partition or set of tables is created in the data warehouse. The effectiveness of this approach is highly dependent on the power and capabilities of the data warehouse technology.

There are advantages and efficiencies from not having to move the data from the warehouse to a separate server. This has spawned *in-database analytics*. With this approach, the analytical algorithms are integrated into the database software. The best example is the partnership that SAS has with Teradata and Oracle to integrate SAS' predictive modeling algorithms into their database software. Not only is the modeling done in the warehouse, but other functions, such as customer scoring (using the results of the modeling), are performed as well. Another advantage of in-database analytics is that provides modelers with access to all of the data in the warehouse rather than possibly subsets or aggregations.

Mixed load software is another way of managing the needs of different kinds of users and tasks. It allows database administrators to assign priorities to different database queries. Like a fast freeway, there are high speed lanes, normal speed lanes, and truck lanes for heavy loads. The quality of mixed load software varies, but the good products ensure short tasks and high priority users have a fast response time and advanced analytics applications execute at a reasonable speed.

A growing number of BI tools provide *in-memory* analytics. These tools store up to a terabyte of data in memory (i.e., RAM) and provide exceptional speed. For example, Qliktech and Tibco Spotfire are in-memory online analytical processing (OLAP) tools that allow users to change dimensions and measures, drill down, and roll up, with no need to access a database because the data is stored in the tool's memory. However, the data is for "a point in time," and depending on the application, needs to be periodically refreshed from the data warehouse.

Analytical platforms are another development. These platforms are designed from the ground up and combine hardware (e.g., servers) and software (e.g., operating system, database) to accelerate query performance against large volumes of data, and provide order of magnitude increases in performance. *Data appliances* are the oldest of these "built for speed" platforms. They were pioneered by Netezza (now owned by IBM), and are now offered by both large (e.g., Oracle) and small (e.g., ParAccel) vendors. They offer a convenient way of offloading complex queries from the data warehouse. Some platforms are in the form of *software-only databases*. An increasingly popular kind are *columnar databases* that reverse the rows and columns to provide greater speed (most queries use only a small number of columns in a row) and offer greater data compression (providing greater storage capacity). Many of the data warehouse vendors are now incorporating the columnar approach into their core products.

Third party hosted platforms are yet another option. This cloud computing approach can take a variety of forms. One alternative is software-as-a-service (SaaS) where a BI vendor makes its products available on demand in a cloud. For example, MicroStrategy offers MicroStrategy Cloud and touts its speed of deployment, reduced project risks, and reduced costs. Another alternative is when a company wants to move quickly and inexpensively to develop and test a new analytical application and chooses to create it in the cloud rather than in house. Many vendors (e.g., Teradata, Microsoft) offer private and/or public cloud-based solutions.

Much of Big Data is semi-structured (e.g., formatted documents) or relatively unstructured (e.g., image) and technologies have emerged for processing it. For example, *non relational database* vendors combine text indexing, natural language processing techniques, and traditional relational database technology to support the execution of ad hoc queries against semi-structured data.

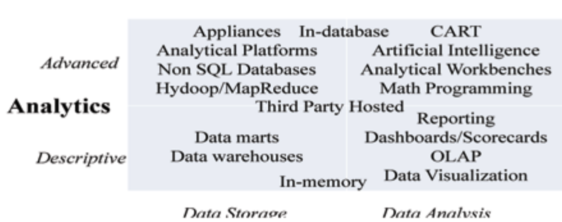
Apache Hadoop is an open source (from the Apache Software Foundation) software framework for processing large amounts of data across potentially massively parallel clusters of servers. A key component of Hadoop is the Hadoop Distributed File System (HDFS) that manages the data spread across the various servers. HDFS is file based and does not need a data model to store and process data, and as a result, can store data of any structure. The Hadoop infrastructure exists to run MapReduce programs (using programming or scripting language such as Java, Python, C, or Perl) in parallel. MapReduce programs take large datasets, extracts and transforms useful data, distributes the data to the various servers where processing occurs, and assembles the results into a smaller, easier to analyze file.

There are many related Apache projects that help create a Hadoop ecosystem. For example, Pig is a high-level parallel processing programming language that is used to write MapReduce programs to run within the Hadoop framework. HBase is a distributed columnar database modeled after Google's Big Table that gives Hadoop a data storage option for large tables. Hive supports SQL-like queries and data summarization. Mahout is a library of data mining algorithms for clustering, classification, and filtering. Collectively, these and other "Apache projects" provide an ever growing set of capabilities for processing and analyzing Big Data.

There are three major ways that companies are using Hadoop/MapReduce (Eckerson, 2011). With the first, companies employ Hadoop/MapReduce as an online archive because of its expandable storage capacity and low cost. The second is as a source system for a data warehouse. In this situation, Hadoop/MapReduce processes semi-structured and/or unstructured data and then passes the output file to a data warehouse where it can be analyzed along with other data. In the third use, either the MapReduce programs or data analysis tools that work with the output file are used to analyze the data. This latter use is likely to grow as vendors see the opportunity to capitalize on companies' desire to analyze semi and unstructured data. Figure 1 summarizes the different database technologies used with descriptive and advanced analytics

- **The Emerging Architecture for Analytics:** In the future, Eckerson (2011) envisions an analytics architecture where the data warehouse is used primarily for structured data and descriptive analytics. Because of its real-time capabilities, it is also used for operational BI. In addition to the warehouse, analytical modelers have access to analytical sandboxes, analytical

Figure 1. The Database and Analysis Technologies Used with Different Kinds of Analytics



platforms, and non relational data repositories like Hadoop/MapReduce that store multi-structured data (i.e., data of any structure). This emerging environment is definitely not "your mother's BI architecture" of only a few years ago.

The Right Analytical Tools

Most organizations have a variety of descriptive analytics tools; however, a different set is needed for advanced analytics. A way of understanding this is to think in terms of hypotheses generation and hypothesis testing. With predictive analytics, the objective is to find hidden relationships in the data (e.g., market segments). The tools and algorithms generate hypotheses about possible relationships in the data. By way of contrast, with descriptive analytics the user provides the hypotheses and explores the data to test whether the hypotheses are supported.

Descriptive and predictive analytics are complimentary, however. To illustrate, descriptive analytics can be used to understand the data, think about relationships before using predictive analytics, and further explore the relationships found through the use of predictive analytics. Also, data visualization is very useful in helping interpret the output from predictive analytics. Some vendors are starting to integrate predictive analytics capabilities into their descriptive analytics products. For example, some dashboards can now not only display time series data but also make projections into the future.

There is not always a clear distinction between the analytical tools and the hardware and software used for storing data. This is the case with in-database analytics and in-memory analytics. In both cases, data storage and analytics are intertwined. The same intertwining is also found in specific application software. For example, predictive analytics is also tightly integrated into campaign management and fraud detection software.

Companies prefer to standardize on a limited number of BI tools, but moving into advanced analytics only increases the number. Analytic modelers often have preferences for tools they are familiar with and are well suited for a specific task. They also like to experiment with new ones, often from the open-source community. It is a balancing act between the benefits of standardization and getting the best tool for the task at hand.

SAS and SPSS (now an IBM company) have been market leaders in predictive analytics for many years. Their products are best described as *analytical workbenches*. They provide a variety of algorithms to use; allow data mining processes to be created, saved, and reused; and assist in exploring, preparing, and combining data to be used with the algorithms.

R, a programming language and software environment for statistical computing and graphics, has quickly become very popular with analytical modelers. It is also at the core of many open source products.

Strong Analytical Personnel in an Appropriate Organizational Structure

Advanced analytics often requires personnel with specialized skills. A 2011 Bloomberg BusinessWeek study found that many organizations lack the necessary talent needed for advanced analytics. With proper training some existing personnel will be able to accept the challenge, at least for more structured analytics supported by appropriate software. For the "rocket science" work on new, challenging applications, new personnel will often have to be brought in, either through new hires or professional services contracts. This latter was the case at both Harrah's and First American Corporation.

Another option for access to advanced analytics is to outsource it to service providers such as Mu Sigma and MethodCare or to companies that offer specific analytics as a service over the Internet. Companies that take these approaches do not have to invest time and money in hiring, training, and organizing their own analytic teams. On the other hand, many companies do not feel comfortable turning their data and potentially strategic applications over to a third party, and it can get expensive when there is considerable analytics work to be done. Another limitation is that it eliminates the potential benefits that can accrue from in-house employees becoming intimate with the data over time and finding new and unanticipated analytic applications. A third party can deliver great value, but eventually their results become repetitive, not exploratory.

Figure 1 summarizes different data analysis technologies used with descriptive and advanced analytics. In-memory and in-database analytics span the database and data analysis categories. Third party hosted platforms can be in any of the cells depending on what is being hosted.

- **The Skills Needed for Advanced Analytics:** Three skills are critical for doing advanced analytics: data, domain, and modeling knowledge. It is rare when someone is strong in all three of these areas. A solution is to create teams where the various people collectively have the necessary skills (Laney & Kart, 2012). For example, a business person has domain knowledge, someone from IT possesses data knowledge, and an analytical modeler has modeling skills. This team approach to advanced analytics is common (EMC, 2011).

Let's consider several examples that illustrate the importance of the various skills. As for data, depending on the company's data infrastructure and the nature of the analytics, the data may be in a data warehouse, in operational systems, from a third party provider, and more. It is necessary to be familiar with what data is available, how to access it, and how to make it available to for the analytics tool.

Domain knowledge is important when choosing the right variables to include in models and interpreting the analytic output. Putting every conceivable variable into a predictive model to see what is statistically significant doesn't work well. You will get too many spurious findings. Rather, it is important have insights into what variables are likely to be related. And once you have findings, domain knowledge is necessary to understand how the findings can be used. Consider the hoary story of the relationship between beer and diapers in the market basket of young males in convenience stores. You still have to decide (or experiment to discover) whether it is better to put them together or spread them across the store (in the hope that other things will be bought while walking the isles). Domain knowledge is important here (and perhaps experimentation).

And finally, it is necessary to have modeling knowledge. To illustrate, for every analytic application, there are modeling alternatives that are promising options but many that are not. After the potential models are run, it is still necessary to interpret the output and decide which model is best for the particular application. All of this requires advanced modeling skills. For predictive analytics in particular, an understanding of multivariate statistics is mandatory.

- **Power Users, Business Analysts, and Data Scientists:** So who are the people who do analytics? While there is no clean, simple answer to this question, it is useful to consider power users, business analysts, and data scientists. *Power users* are business people who become very familiar with an analytical tool (e.g., MicroStrategy), use it in their work, develop simple applications, and help other users. Most often their expertise is with descriptive analytics. They are high in domain knowledge but lower in data skills and much lower in modeling skills.

Business analysts are also business people but their job is very analytics intensive. Examples include financial analysts and market researchers. They often have BBAs and MBAs, are very analytical by nature, work with descriptive analytics tools or applications that have advanced analytics integrated into them, and are located in a specific business unit, such as marketing. They are high in domain knowledge, and have moderate data and modeling skills.

Much has been written recently about *data scientists* (Laney & Kart, 2012). These are the "rocket scientists" mentioned previously. They typically have advanced degrees in statistics, operations research/management science, computer science, engineering, mathematics, or econometrics. One survey found that 40 percent have a master's degree and an additional 17 percent have a doctorate (EMC, 2011). They work with advanced analytical tools such as SAS, SPSS, R, and open-source software and build models based on neural networks, genetic algorithms, classification and regression trees (CART), machine learning, and the like. Rather than being associated with a particular business unit, they work on whatever applications are given the highest priority. They are especially high in modeling knowledge, strong in working with data, but may not always have business domain knowledge (will depend on their work experience).

There is currently a shortage of data scientists and it is projected to become worse. A 2011 study by the McKinsey Global Institute predicts a shortfall of between 140,000 and 190,000 data scientists by 2018 (Manyika et al., 2011). Other surveys also suggest that the demand will outstrip the supply (EMC, 2011). The dearth of analytical talent is a bigger issue in the use of advanced analytics than the available technology (Laney & Kast, 2012).

Many companies will turn to universities for the needed talent. The good news is that universities are gearing up to help. For example, North Carolina State University (with the support of SAS), Saint Joseph's University, and Northwestern University have recently created specialized masters degrees in BI and analytics, and more programs will certainly emerge. The bad news is that advanced analytics is a challenging field of study that is not easily or quickly completed, thus limiting the number of graduates even when degree programs are in place. Other options for acquiring talent are to hire it away from other companies or to train your own people. Deloitte Consulting is following the latter path. They have worked with the Kelly School of Business at Indiana University to create an analytics certificate program for their consultants.

It is also possible that some business analysts can do advanced analytics work when they are given appropriate software and training. Advanced analytics takes someone who is inquisitive, analytical, and enjoys solving problems. For the more structured analytical tasks, such as identifying market segments for campaign management, software vendors have developed solutions that facilitate the analytical process. The software provides user-friendly graphical interfaces and partial automation of the model building, testing, and implementation process. The software often builds multiple models using different algorithms and suggests which one is best based on the data. Over time we can expect continuing automation of analytical work, reducing the need for data scientists. Of course, training of the business analyst on the software is critical.

It is helpful to recognize that not only the backgrounds but the interests, motivations, and career paths for business analysts and data scientists tend to differ (Brohman & Watson, 2006; Chambers, 2012). Most fundamentally, business analysts focus on the business and see their career path as progressing through a business unit. Data scientists, on the other hand, focus on their technical expertise and see their path as working on a variety of interesting, challenging analytics projects across the organization. Data scientists often expect more flexibility and freedom in the workplace (e.g., reflected in less tolerance for meetings), respond differently to incentives (e.g., may respond well to new training opportunities), and feel like they are the smartest people in the room (they may be, but not necessarily in a business or organizational sense). Some of the most salient differences between business analysts and data scientists are summarized in Figure 2.

- **Organizational Placement:** In most organizations, business analysts and data scientists are hired by department heads and are pooled in pockets of an organization (Eckerson, 2011). This is most likely

Figure 2. The Differences between Business Analysts and Data Scientists

	Business Analyst	Data Scientist
		
Education:	BBA, MBA	MS, PhD
Tools:	Cognos, Hyperion	R, SAS
Analytics:	OLAP	CART
Focus:	Business	Analytics
Scope:	Departmental	Enterprise-wide
Value:	High	Exceptionally high

the case in organizations that are relatively new to advanced analytics or haven't found the need to take an enterprise-wide approach.

Over time, the realization often grows that there is a need to optimize the use of human resources, reuse models or integrate models in other areas, share best practices, prioritize the overall work to be done, and the like. When this is the case, some organizations create an Analytics Competence Center (ACC). The ACC can be either centralized, hybrid, or virtual.

With a *virtual* ACC, people maintain their current responsibilities but also take on the duties required by the ACC. The benefits of this approach include low initial cost, ability to get representation from multiple parts of the organization, better integration with the business, and the ability to tap into existing knowledge. The potential risks are low accountability of resources and potential lack of executive buy-in.

A *centralized* ACC is staffed with people whose sole responsibility is for the center's deliverables. The ACC reports directly to the center's executive sponsor. The benefits include providing greater accountability for analytics, allowing resources to specialize, and providing a well-defined structure for doing analytics. The risk is that it will be poorly connected to the business.

A *hybrid* ACC has a small centralized team with a percentage of the team's time allocated and under the direct control of the business units. The benefits include the singular focus of the team and the ability to deploy resources to specific projects. The potential risks include increased initial costs due to resource allocation and potential resource conflicts due to part-time roles from key people.

While all approaches can work, the hybrid structure is especially appealing. Business analysts remain in the business units and data scientists can be allocated on a project-by-project basis.

CONCLUSION

As we have seen, there are several potential drivers behind analytics organizations and the use of advanced analytics – the nature of the industry, seizing an opportunity, or responding to a problem. Becoming advanced analytics focused is not easy, however. It requires a movement along the analytics maturity curve with attendant changes in organizational culture; the kinds of data captured, stored, and analyzed; the data infrastructure, including technologies and approaches; the analytics tools used; and organizational personal who perform analytics. Most companies make the move along the maturity curve gradually. If the need is dire, however, some companies have committed the resources to bring in new technology and outside help in order to move forward quickly. For companies who successfully complete the journey to advanced analytics, empirical research is now confirming the organizational benefits.

REFERENCES

Barnes, S., Ring, T., Yamunan, V., Franks, B., Hiltbrand, T., Taylor, J., & Williams, S. (2012). BI experts' perspective. *Business Intelligence Journal*, *17*(1), 20–27.

Bloomberg Businessweek Research Services. (2011). The current state of business analytics: Where do we go from here? Retrieved from http://www.informationweek.com/whitepaper/Business-Intelligence/Analytics/the-current-state-of-business-analytics-where-do-wp1318437706?articleID=191703582

Brohman, M. K., & Watson, H. J. (2006). Maximizing the return on OLAP and data mining analysts. *Business Intelligence Journal*, *11*(3), 30–36.

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decision-making affect firm performance? *Social Science Research Network (SSRN), April 2011.*

Chambers, R. (2012). How to manage analytic resources. *Information Management*. Retrieved from http://www.information-management.com/newsletters/analytics-resources-PM-BI-leadership-Chambers-10021862-1.html

Cooper, B. L., Watson, H. J., Wixom, B. H., & Goodhue, D. L. (2000). Data warehousing supports corporate strategy at first American corporation. *Management Information Systems Quarterly*, *24*(4), 547–567. doi:10.2307/3250947.

Davenport, T., & Harris, J. (2007). *Competing on analytics: The new science of winning*. Boston, MA: Harvard Business Press.

Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at work: Smarter decisions, better results*. Boston, MA: Harvard Business Press.

Eckerson, W. (2004). Gauge your data warehouse maturity. *DM Review*, *51*, 34–37.

Eckerson, W. (2007). *2007 TDWI benchmark report*. Seattle, WA: The Data Warehousing Institute.

Eckerson, W. (2011). Big data analytics: Profiling the use of analytical platforms in user organizations. BeyeNetwork. Retrieved from http://www.beyeresearch.com/executive/15546

EMC. (2011). Data science revealed: A data-driven glimpse into the burgeoning new field. Retrieved from http://www.emc.com/collateral/about/news/emc-data-science-study-wp.pdf

Gartner. (2009b). *Howard Dresner*. Retrieved from http://www.gartner.com/research/fellows/asset_79427_1175.jsp

Kiron, D., et al. (2011). Analytics: The widening divide," IBM, *MIT Sloan Management Review*. Retrieved from http://sloanreview.mit.edu/feature/achieving-competitive-advantage-through-analytics

Laney, D., & Kart, L. (2012). Emerging role of the data scientist and the art of data science. *Gartner Report*. Retrieved from http://www.gartner.com/id=1955615

LaValle., et al. (2010). Analytics: The new path to value. IBM, *MIT Sloan Management Review*. Retrieved from http://public.dhe.ibm.com/common/ssi/ecm/en/gbe03371usen/GBE03371USEN.PDF

Lohr, S. (2012, February 12). The age of big data. *New Your Times*. Retrieved from http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=1&emc=eta1

Manyika, J., Chui, M., Brown, B., & Bughin, J. Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*. Retrieved from http://www.mckinsey.com/Search.aspx?q=data scientists

Moore, G. (2002). *Crossing the chasm: Marketing and selling high-tech products to mainstream customers*. New York, NY: Harper Business Essentials.

Moore, G. (2004). *Inside the tornado*. New York, NY: Harper Business Essentials.

Power, D. J. (2007). A brief history of decision support systems. *DSSResources.COM*, version 4. Retrieved May 20, 2009, from http://DSSResources.COM/history/dsshistory.html

Russom, P. (2011). *Big data analytics. TDWI Best Practices Report*. Seattle, WA: The Data Warehousing Institute. Retrieved from http://tdwi.org/research/2011/09/best-practices-report-q4-big-data-analytics.aspx

Watson, H. J. (1997). *Building executive information systems and other decision support applications*. Hoboken, NJ: John Wiley & Sons.

Watson, H. J., Fuller, C., & Ariyachandra, T. (2004). Data warehouse governance: Best practices at Blue Cross and Blue Shield of North Carolina. *Decision Support Systems*, *38*(3), 435–450. doi:10.1016/j.dss.2003.06.001.

Watson, H. J., & Volonino, L. (2002). Customer relationship management at Harrah's entertainment. In G. A. Forgionne, J. N. D. Gupta, & M. Mora (Eds.), *Decision-making support systems: Achievements and challenges for the decade*. Hershey, PA: IGI Global. doi:10.4018/978-1-59140-045-5.ch010.

Williams, S. (2004). Assessing BI readiness: A key to BI ROI. *Business Intelligence Journal*, *9*(3), 15–23.

Wixom, B. H., & Watson, H. J. (2010). The BI-based organization. *International Journal of Business Intelligence Research*, *1*(1), 13–25. doi:10.4018/jbir.2010071702.

Wixom, B. H., Watson, H. J., & Werner, T. (2011). Developing an enterprise business intelligence capability. *The Norfolk Southern Journey*, *10*(2), 61–71.