Business Analytics for Big Data

Unlock value to fuel performance

Contents

Introduction

Organizations today collect more data than ever before. Much of it is difficult to analyze, yet the insights contained in this data can be extremely valuable. New technologies for managing, exploring and analyzing data have advanced so far that it is now feasible for organizations of all sizes to capitalize on big data.

Admittedly, there is some disagreement about how to define the term *big data*. Some observers include the massive and newly usable sources of data as well as the supporting storage, management, governance and analytic technologies in their definitions.[1] Others consider the data as separate from the technologies that enable organizations to explore and use it in their decision-making.

Differences aside, there is widespread agreement among analysts and business leaders alike that applying analytics to big data is becoming a business imperative for many organizations in their quest to remain competitive.

This paper focuses largely on how analytics solutions support the use of big data in decision-making at all levels. It describes the benefits of analyzing big data and then points out some of the challenges involved. Finally, it provides an overview of how IBM technologies for business analytics meet these challenges and help clients provide analytics to business users and systems that help drive better outcomes across the entire enterprise.

Extracting insights from big data

Data is often described as a vital natural resource—sometimes, as "the new oil,"[2] with analytics the "refinery" that turns data into fuel. This fuel consists of insights that enable companies to know what action to take. A critical factor is getting that knowledge to the right people or systems in time to make a difference. That's where IBM Business Analytics software comes into play.

Organizations implementing big data initiatives may start by combining data they are already collecting—for example, data from internal transactional, ERP and CRM systems—with new types of data that previously were either unavailable or unusable for business decision-making. These data types include semi-structured machine sensor and RFID data and unstructured text data, social media data, as well as data stored in new systems that are designed for optimized analysis of large, complex data sources.

By combining analytics of big data with their traditional analytics, leading-edge organizations can:

- **Deliver new insights and new possibilities**—to do things they couldn't do before
- **Drive process and performance improvement**—to execute current initiatives more effectively
- **Create new revenue opportunities**—to generate new business models and ways to use data to generate cash flow

These efforts are dramatically improving organizations' ability to compete. However, they require analytics that are tuned specifically to the unique characteristics of big data—analytics that can deliver insights beyond "power users," so that decisions and actions can be consistently optimized at every level of the organization.

As beneficial as it can be to employ big data, analyzing it does present real challenges—in part, due to the unique characteristics of big data.

Big data is often described in terms of its **volume**, **variety**, **velocity** and **veracity**.

- Volume refers to the sheer amount of data being generated—nearly 2 zettabytes in 2011, according to IDC[3] and increasing to an estimated 35 zettabytes by 2020
- Variety indicates that analysts now look to text, images and other unstructured or semi-structured data, in addition to the data held in structured databases or other numeric-based file systems.
- Velocity reflects the fact that, with billions of sensors constantly collecting data on everything from weather to utility grid usage to financial trades to individual purchases and driving habits, the flow of data never ceases; and, with today's improved network capabilities, it arrives faster than ever before.
- Veracity acknowledges that much of this data may be incomplete or seemingly contradictory, requiring new techniques that deliver reliable insights despite such uncertainty.

The next sections of this paper describe three key challenges associated with analyzing big data and how IBM solutions help organizations meet those challenges.

Transforming healthcare with analytics

The Premier Healthcare Alliance serves more than 2,500 US hospitals and 73,000-plus other healthcare sites. Using IBM Business Analytics software, it has undertaken a groundbreaking initiative, combining previously siloed information from many different sources to help hospitals, doctors and other health providers **do something they could not do before** — identify across a network the treatments that benefit patients most. In one project, 157 participating hospitals saved an estimated 24,800 lives while reducing spending by $2.85 billion.

Challenges of analyzing big data

New ways of doing things always pose challenges, and applying analytics to big data is no exception.

Some organizations are still trying to determine which new sources of data will shed the most light on their business issues. Others find that they have implemented ad hoc big data projects without an overall plan for optimizing the impact of their efforts. In some cases, tools or techniques that seemed promising have turned out to have hidden costs or shortcomings. For example, specialized systems, such as those built on Apache™ Hadoop®, may lack adequate management tools or may require staff to have programming skills in order to extract insights.

In general, challenges to performing analytics with big data fall into three main areas:

- Those associated with **accessing** large amounts of data of various types and from a variety of new sources
- Those related to **interacting** with data in order to discover useful patterns and trends
- Those concerned with **performing advanced analytics**, and doing so at scale, to meet the need for timely insights, and integrating these with current analytics

As organizations have learned from years of building, maintaining and upgrading data warehouses, making data usable for analysis requires significant investments in infrastructure and in people with the skills to maintain and update those data stores.

The volume and velocity of big data sources—and the need for rapid analysis—make it impractical to create a massive new data warehouse before conducting analysis. In some cases, exploration and discovery may be required in order to understand what portions of the data are relevant. In many cases, analytics must be performed both on data in existing data warehouses and on data stored in specialized systems like those based on Hadoop.

Assessing credit risks more accurately

Equifax is a global leader in consumer and commercial information solutions. Using IBM Business Analytics software, Equifax developed a proprietary technique for "ensembling" multiple risk models into a single score. The technique weights and combines the significant dimensions of different risk models into a single score, which is 30 to 50 percent more accurate, depending upon the type of risk. Using this new technique, Equifax **gets better answers,** which its clients can use to make better decisions about their customer relationships.

The second challenge, being able to interact with big data to discover useful patterns and trends, is directly related to the sheer volume of information contained in big data. This "noise" makes it difficult to detect the "signal," the patterns and trends that are relevant to the business issue at hand. New visualization techniques are needed that can guide analysts during data exploration, particularly since the best visualization may not be one they are familiar with.

The third challenge, applying advanced analytics, requires organizations to examine both their technology infrastructure and their business processes. Advanced mathematical algorithms play a greater role when dealing with the complexity of big data. If additional varieties of data, such as text, are to be incorporated in the analysis, additional tools and techniques may be needed; however, in order to get the full picture, the results from text analysis need to be combined with results obtained from analyzing structured data. Organizations also need to develop processes and protocols so that the results of analytics performed on big data are comparable over time and embedded into operational systems and processes.

A related issue is the veracity of the data. Not surprisingly, data accessed from multiple sources is often "dirty." Differences occur in data formats, definitions, architecture, completeness and accuracy. Traditional data cleansing tools are not designed to handle such situations. However, analytics can be used to fill in missing values to make data more complete. And new techniques, such as entity analytics, can be used to resolve data inconsistencies so that analyses can be carried out and deliver valid results.

IBM Business Analytics for big data

IBM recognized the challenges inherent in analyzing big data and built our Business Analytics portfolio to empower organizations to surmount those challenges. With its advanced predictive analytics capabilities, this portfolio allows organizations to discover insights hidden in big data—and then use these insights to anticipate and shape business outcomes. As a result, organizations can more successfully align their information to their business objectives.

The IBM Big Data and Analytics solution connects Business Analytics capabilities to IBM's enterprise big data platform, as shown in Figure 1, below. The big data platform includes solutions for information integration and governance, data warehouse management, stream computing and an enterprise Hadoop solution. It also includes specialized systems and infrastructure solutions, such as the IBM PureSystems™ family of products, and IBM Smart Cloud solutions for cloud implementations.

The latest enhancements to this platform dramatically improve performance for organizations with data stored in IBM DB2® databases, including those using IBM Cognos® business intelligence software, IBM InfoSphere® Streams, and those exploring the potential of Hadoop to simplify their big data initiatives.

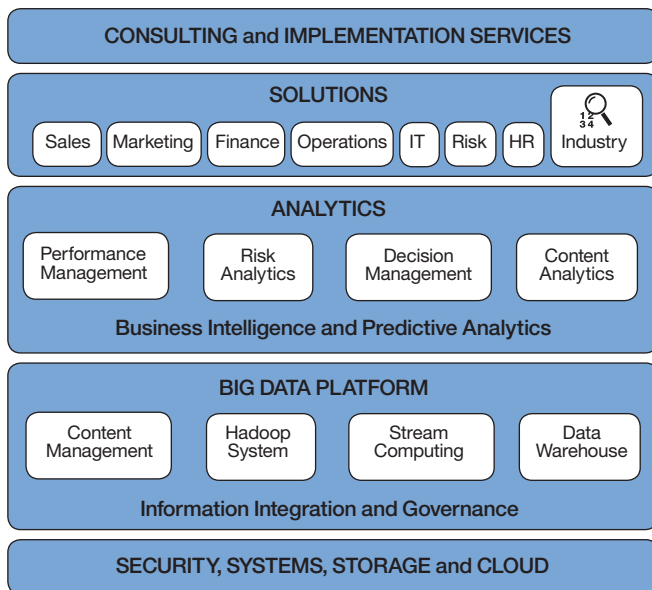**The IBM Big Data and Analytics Solution**

Figure 1: The IBM Big Data and Analytics Solution. For a more detailed treatment of products comprising the IBM platform for big data, download the white paper, *Architecting a Big Data Platform for Analytics*, from www.ibmbigdatahub.com.

IBM Business Analytics software provides organizations with new ways to combine the analysis of big data with traditional analytics, particularly in three areas: accessing big data, interacting with it visually to uncover patterns and trends, and then applying analytics that scale to meet big data and business performance requirements.

Access big data with confidence

IBM Business Analytics software enables organizations to access information of various types and from a variety of sources. Whether that data resides in relational databases, Hadoop-based systems, and whether it's inside or outside a firewall, it can be assembled and analyzed to obtain a more fully rounded view of business conditions.

Whatever data sources are chosen, with IBM Business Analytics software, users will enjoy a unified visual interface. In addition, analytic capabilities within the IBM Cognos business intelligence platform enable organizations to optimize access to large volumes of data in data warehouses or in external databases through in-memory acceleration, in-database processing and aggregate awareness.

"At rest" data can be stored either in traditional data warehouses or in specialized systems that are designed for handling large volumes of structured and unstructured data. These include IBM InfoSphere BigInsights™, which is IBM's implementation of an enterprise-level Hadoop system, and

the IBM PureData™ System for Analytics, a system designed for scaling big data analytics. IBM Business Analytics products are optimized to take advantage of these systems; in addition, IBM has optimized the way Hadoop-based information is accessed to deliver greater analytic efficiency.

Data that is "in motion," such as data on financial market trading behavior, RFID tags monitoring supply chains, or machine sensor data in manufacturing settings, can be analyzed using IBM InfoSphere Streams, a product that specifically designed to continuously look for patterns in this type of data in order to automate real-time decisions. For example, IBM predictive models can be embedded within Streams processing to identify patterns before the data is stored. Then results can be viewed in real time using IBM Cognos Real Time Monitoring, or actions can be automatically initiated.

After accessing and analyzing big data in combination with traditional data sources, analysts using IBM Business Analytics products can create interactive dashboards or pixel-perfect reports and deliver them efficiently to everyone in the organization for further analysis—even to mobile devices. Analysts can also generate predictive models to accurately determine what will likely happen in the future, and then deploy those models to optimize and automate decision-making.

Figure 2 illustrates how IBM Business Analytics products support the analysis of data, whether it is streaming, stored in traditional databases, or stored in specialized systems like Hadoop.
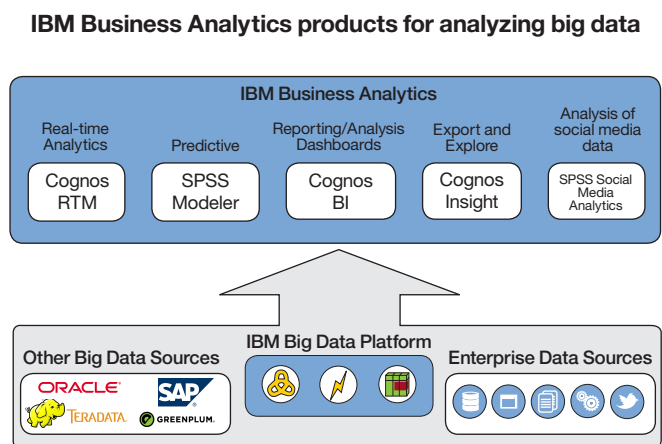
IBM Business Analytics products for analyzing big data

Figure 2: This diagram identifies the big data and enterprise data sources which IBM Business Analytics software can combine, spanning the breadth of data "in motion" and data "at rest." For data source versions, refer to the Cognos or SPSS system requirements on the **ibm.com**.

Utility plans for the future

Smart grids, renewable energy, electric vehicles: these are only some of the changes re-shaping the energy and utilities industries. One North American utility is using IBM Business Analytics software to analyze streaming data from grid sensors in near-real-time and using sophisticated scenario planning to uncover new revenue streams. This helps it make better decisions, such as understanding how to incorporate renewable energy sources in its portfolio while still meeting customers' needs for reliable energy.

Interact with big data visually

To effectively explore big data and discover insights hidden within it, advanced visualization capabilities are essential. Visualizations enable users to focus on the types of data or portions of datasets that merit further exploration. Because big data is inherently noisy—hiding patterns in masses of details—these capabilities are essential in picking up subtle signals.

The visualizations in IBM Business Analytics software go beyond traditional pie charts, bar graphs or scatter plots to present data in new ways that make patterns instantly meaningful.

These visualizations are augmented by intelligent discovery capabilities that guide users to the most appropriate visualization for their data—thus revealing insights that might otherwise have been hard to detect in large, complex datasets.

IBM has made a significant investment in visualization technologies, and continues to do so. Advanced techniques continue to be embedded in analytics solutions, making it easier for users to explore ever larger datasets with greater ease and confidence.

Social media analytics is a powerful tool for uncovering customer sentiment dispersed across countless online sources. Like IBM's Hadoop, text analytics and dashboard capabilities, IBM SPSS Social Media Analytics utilizes visualizations to help customize campaigns, predict customer behavior and identify influencers in the social media space, as shown in Figure 3.
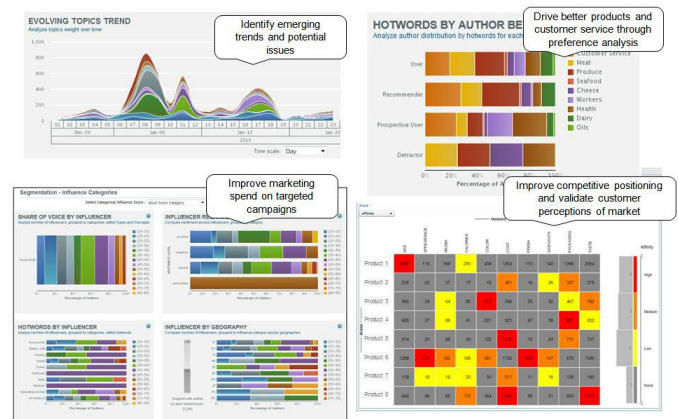
Visualizing patterns in social media

Figure 3: IBM SPSS Social Media Analytics presents visualizations which transform social media interactions into actionable insights.

Providing services based on analytics

A wholly owned subsidiary of Merck, Telerx provides customer care contact center services, primarily to clients in the pharmaceutical and consumer packaged goods industries. With more than 30 years' worth of data on consumer behavior, Telerx chose an IBM Business Analytics solution to improve current efforts. The company uses this software on text data from social media channels and from unsolicited customer comments to develop more precise predictive models that its clients can use to sharpen their marketing efforts.

Choose big data analytics that scale

As data volumes have grown significantly and data sets have become more complex, predictive analytics algorithms that leverage mathematics and computing power to discover patterns have become critical tools. These advanced algorithms need to scale efficiently so that organizations can discover connections and relationships in ever-larger data sets with ever greater accuracy.

IBM Business Analytics solutions offer advanced predictive analytics techniques for forecasting, clustering and categorizing data, as well as some that are critical for working with big data. For instance, entity analytics can be used to resolve some of the inconsistencies that may surface in big data and improve its veracity. Other algorithms are designed to analyze the concepts and insights hidden in text. Also important are algorithms that can help to understand relationships in social network data and reveal key influencers and how they affect other people's behavior.

Scaling these advanced analytic techniques often involves "bringing the analytics closer to the data." That is, running algorithms and procedures on data residing within a database, avoiding transferring the data to the analytical or predictive modeling tool, and using multi-core processing to speed analysis. IBM Business Analytics solutions integrate with several database engines that include in-database analytics; this significantly improves performance.

The latest enhancements to the IBM Business Analytics portfolio include IBM SPSS Analytic Server and IBM SPSS Analytic Catalyst. Analytic Server provides a framework that leverages the power of distributed file systems and databases such as Hadoop for running analyses on big data, enabling IBM SPSS Modeler to use sophisticated analytical algorithms against very large data in a timely and highly scalable fashion.

IBM SPSS Analytic Catalyst is an analytics tool that that makes predictive analytics and discovery more accessible to business analysts than ever before. IBM SPSS Analytic Catalyst uses advanced automation techniques to reveal key insights and the most statistically interesting relationships in big data and delivers results as interactive visualizations with plain—language summaries—without requiring programming skills or advanced statistical knowledge.

Conclusion

Analytics is the key to obtaining value from big data. With the advanced technologies now available to analyze big data, organizations can:

- **Gain a deeper understanding of the dynamics of their business** by accessing, visualizing and analyzing big data, and integrating the results along with their analyses of traditional information.
- **React faster**, since new perspectives on day-to-day operations can be obtained by using scalable analytics on large volumes of data and/or analyzing streams of data in real time or near-real time.
- **Deliver more complete and accurate answers** because all data, not just a sample, is analyzed, and because the analytics of more types—such as text and semi-structured data—can provide context to traditional data.
- **Automate and optimize decision-making** based on more accurate insights gained from analyzing big data.
- **Reduce IT costs** by using high-performance IBM technologies that support combining traditional data analysis with analyses of data held in less-costly systems such as those based on Hadoop.

Clearly, big data offers organizations real benefits. Although taking advantage of this vast new data resource poses challenges, a growing number of organizations are meeting these challenges with IBM Business Analytics.

Banks are increasing their customer focus by analyzing a variety of new data sources to create differentiated customer understanding, inform customer strategy and optimize customer interaction for competitive advantage.

Insurers are optimizing their claim-handling by accessing internal and external data sources to quickly understand the likelihood of fraud. This allows them to speed up the handling of simple claims and quickly investigate complex or likely fraudulent ones.

Government entities are looking at anomalies in submissions for social programs and taxes. They can uncover patterns of risk and fraud faster than ever before, and then determine the best course of action to follow.

Manufacturers and **wholesalers** are picking up signs of supply chain issues sooner so that they can take advantage of different logistical approaches to avoid the high costs associated with material delays, overstock or stock-out conditions.

Firms as diverse as **hotels**, **telecommunications companies**, **retailers** and **restaurants** are gaining a clearer view into customer preferences and building a more loyal and profitable customer base.

Cities and **public utilities** are more accurately gauging public usage of essential services, so they can optimize delivery mechanisms and avoid costly or dangerous bottlenecks.

As with many important business initiatives, when an organization decides it needs to expand its analytic scope to include big data, there is no simple one-size-fits-all solution. IBM Business Analytics software enables organizations to build upon existing analytics initiatives to meet the most urgent challenges that analyzing big data presents and transform the way they operate—to exceed expectations and outperform competitors.

About IBM Business Analytics

IBM Business Analytics software delivers data-driven insights that help organizations work smarter and outperform their peers. This comprehensive portfolio includes solutions for business intelligence, predictive analytics and decision management, performance management, and risk management.

Business Analytics solutions enable companies to identify and visualize trends and patterns in areas, such as customer analytics, that can have a profound effect on business performance. They can compare scenarios, anticipate potential threats and opportunities, better plan, budget and forecast resources, balance risks against expected returns and work to meet regulatory requirements. By making analytics widely available, organizations can align tactical and strategic decision-making to achieve business goals. For further information please visit **ibm.com**/business-analytics.

Request a call

To request a call or to ask a question, go to **ibm.com**/business-analytics/contactus. An IBM representative will respond to your inquiry within two business days.

1 *Converting Big Data Hype into Big Value with Analytics*, Colin White, BI Research, 2012.

2 http://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/ ©2012-2013 Forbes.com LLC™

3 http://www.datacenterknowledge.com/archives/2011/06/28/digital-universe-to-add-1-8-zettabytes-in-2011/ ©2006-2013 Data Center Knowledge

Please Recycle

YTW03329-USEN-01