Unlock Big Value in Big Data with Analytics

An IBM Redbooks Point-of-View publication

By John Hagerty, BA Market Strategist, and
Tina Groves, BA Product Strategist for Big Data

Highlights

Big data expands and evolves analytics that were not previously possible because of lack of available information, technology limitation, or prohibitive cost. Companies that extend analytics to big data can:

► Gain more complete answers and create new perspectives by relating and analyzing all sources of data

► Make better business decisions by anticipating business outcomes and uncovering hidden patterns and relationships across all sources of data

► Uncover new opportunities and enable new business models by differentiating customer service or reducing operating costs

Ask any group of business or IT leaders, and a clear majority reports that they are actively engaged in discussions about the influence of big data in their organizations or industry sector. For many organizations, however, "discussions" are as far as they have progressed on this topic. There is now a widespread recognition that big data unto itself has limited business impact unless you can unlock the value in that data through innovative analytics techniques and capabilities. Analytics for big data is simply an evolution of business analytics, with the emergence of myriad sources of data that can fuel richer insights, discover new opportunities, and enable game-changing results.

Today's business leaders recognize that this data is too large and complex and changes too quickly for decision makers to manually analyze it. Decision makers need analytics to address the following big data trends:

► Use all information by integrating existing and new sources of data to reveal new insights, identify new opportunities, and deliver new business value.

► Take advantage of all perspectives by analyzing data to assess what happened, to understand what is happening now, to predict what will happen, and to automatically take action based on recommendations.

► Empower people with analytics so that new sensor, geographical, and location data can fuel analytics in new segments of the organization.

► Embed analytics into decisions to enable fact-based decisions at the point of impact so that organizations employ analytics-driven operations that automate decisions whenever possible.

Big data encompasses more than analytics. It represents a full ecosystem around gathering information from many sources, preparing that information for use, processing the information to derive insight, and then delivering that insight throughout the enterprise. But how do you move beyond the hype to realize the value of big data?

Delivering analytics value from big data requires an evolution of thinking. Although you might still be in the early stages of adoption, you can gain an advantage by following a logical approach to enhance your analytics capabilities.

Redbooks

If you want to capitalize on your big data needs, consider the following key questions:

► What are the key business analytics capabilities that big data can extend?

► What are the essential business models and services needed?

► What business cases do you need to add or modify to be successful?

How big data expands business analytics capabilities

The business intent for analytics has remained constant—to help organizations monitor, manage, diagnose, predict, and optimize performance in every corner of the business. To date, analytics has largely been applied to internally generated, highly structured business data. In most cases, this data is captured by enterprise applications that track business and operational processes.

Today's innovative analytics capabilities for big data go far beyond simple data gathering to provide the following opportunities:

► Deep sources of machine or sensor data can be effectively tapped to identify when equipment needs to be serviced before it breaks down and disrupts the business process.

► Customer service records can be interpreted to detect shifts in client opinion about recently introduced products to capture more market share.

► Social media commentary can be analyzed to understand potential buyer concerns and preferences that might impact production and distribution decisions in near real time.

► Granular transaction activity can be scanned and scored in real time to identify fraud, waste, and policy abuse to help prevent it from occurring.

These examples are just a few of the many opportunities that can be realized when analytics are applied to big data. Organizations that understand these connections can apply analytics at the point of decision impact, having all relevant information at hand.

So what has changed? Big data represents advances in technologies and capabilities that make analyzing massive amounts of data and analyzing unique types of data both feasible and affordable for a broad range

of applications. Specifically, big data fuels business analytics with three distinct benefits:

► Delivers more complete answers and new insights

► Improves processes and performance

► Creates new business models and differentiated services

Delivers more complete answers and new insights

Previously, the cost of retaining data needed to be weighed against the value it provided to the organization. Data with known usefulness was typically summarized in a data warehouse or data mart. The source data was maintained for a period of time and then archived or deleted. Asking a new business question became a challenge because the source data might no longer be accessible, if it was retained at all. Data that was difficult to analyze or whose value was uncertain, such as log files or video feeds, was typically deleted daily or weekly.

Technology advances, such as Hadoop, massively parallel processing (MPP) systems, or columnar data stores, enable storing and processing massive amounts of data at an acceptable cost. Keeping more data allows you to:

► Analyze data over longer time periods, which better identifies trends and excludes anomalies

► Include larger sample sizes, which reduces the effect of incorrect data and increases the confidence factor

► Keep all available properties, which could identify new causal relationships[1]

Data platforms have also embraced diverse sources of content—extending from structured data to unstructured sources, ranging from social media feeds to telemetry data to video files.

To take advantage of big data, complementary advances in analytics capabilities effectively tackle the breadth and depth of data that is generated and accumulated. Some examples of these advances are adaptive predictive models, automated decisioning, network analytics, analytics on data-in-motion, and new visualizations.

---

[1] In this context, the term *properties* also refers to *attributes*, *fields*, *columns*, or *variables*.

Using these types of analytics, technologies can provide the following insights:

- Assemble different mixes of information that were not previously not available, for example using click stream patterns from web server logs combined with purchase data when analyzing shopping cart abandonment rates

- Discover patterns or relationships that were previously hidden, for example analyzing physician notes on patient records to discover a relationship between having a "mother" at home and lower readmission rates

- Visualize and explore all relevant information together and in context to gain a more complete view, for example using interactive bubble charts to highlight key points from a complex budget proposal[2]

- Take action or automate processes to improve outcomes, for example analyzing phone calls in real time to calculate a customer's loyalty and churn score to cue a client service representative with the right promotion

Improves processes and performance

Armed with more complete answers and insights, organizations can look to streamline processes and measure performance improvements in new ways. Achieving operational efficiencies can result in cost savings, reduced financial losses and litigation risk, improved customer satisfaction, or improved product quality.

For example, one car manufacturer identified a trend where a brake light was being replaced sooner than normal use in some cars through the analysis of warranty service records. Further analysis ruled out a specific car make or model. Investigation into the production line and parts did not identify a root cause for the situation. However, when the analysis was expanded to the car purchase records, an unusual insight was discovered. All cars that had brake lights burn out also had a sunroof. In the end, relocating a portion of the wiring around the sunroof mechanism corrected the brake light issue.

This new insight, the relationship between the brake light service and the sunroof, enabled the car manufacturer to change its manufacturing process before incurring significant warranty costs.

A specific example from the telecommunications industry was implemented by XO Communications, a large communications service provider that offers comprehensive telecommunications and Internet services to more than 90,000 customers.[3] The company wanted to improve the retention of its small business customers and needed to identify which of its small business customers were at the highest risk of switching to a competitor.

XO Communications created an internal data mart and used IBM® SPSS® Modeler to create a predictive model that evaluated more than 500 potential variables for predicting voluntary customer defection within 90 days. The company's customer intelligence team then built a multiple-regression model that focused on the 25 most relevant variables. Monthly churn risk scores were assigned to all customers, with client services executing "proactive save" communications on the top 10%. These process improvements provided the following results:

- 35% reduced customer churn within the first year
- 60% increased retention rates
- 60% improved billed revenue retention rate
- 376% ROI achieved in five months

In addition, the company was able to decrease the number of client service agents needed for the same level of customer contact.

Creates new business models and differentiated services

Forward thinking finance professionals are partnering with business experts to evaluate new sources of revenue and business model innovation from big data assets.

What kinds of data are prompting new services?

- Global positioning system (GPS), telemetry, or other sensor data

  Vehicle manufacturers that use telemetry data for diagnostics can leverage those signals to create an intelligent on-board concierge system as a new service. When issues are detected, a service appointment at the dealership and a car rental reservation can be made automatically for the customer.

  Another example includes municipalities that implement intelligent parking services. Sensors

---

[2] For an example of an interactive bubble chart, see:
http://www.nytimes.com/interactive/2012/02/13/us/politics/2013-budget-proposal-graphic.html?_r=2&

[3] For more information about this IBM case study, see:
http://www-01.ibm.com/software/success/cssdb.nsf/CS/STRD-8SSF3T

detect the availability of a parking spot, and the resulting information is shared through street-side signboards, an interactive website, and mobile applications.

► Customer contact information

Financial entities, such as credit card companies or banks with access to both customer information and vendor purchases, are experimenting with offering promotions to customers, paid by the vendors. For example, a coffee purchased by credit or debit card triggers the financial entity sending a vendor offer, such as a 10% discount for a second cup, to the customer's mobile phone or email.

► External data combined with an internal source

Agencies in the travel business are exploring methods to cross-reference weather data with travel plans. These new insights allow the agencies to assess service opportunities to customers by ensuring that umbrellas or sunscreen are available and that itineraries are revised automatically.

► Click-through data

On-demand engagement marketing providers can analyze emails to provide guidance to customers on the best dates and times to send emails for maximum response. Focusing a customer email campaign can increase performance, thus enhancing the value of this service.

Building the business case for your big data project

The promise of finding new, more complete, or hidden insights is alluring but often is not enough to justify spending precious resources. Prudent business owners can increase the probability of success by having a clear understanding of the anticipated outcome and by eliminating any unknowns at the beginning of a project. When developing a big data project, you need to either develop a new analytics model problem on known data or implement a known analytics problem on challenging data due to volume or the number of variables.

To get started on analytics and big data projects, here are examples of proven drivers—from simple to more sophisticated, with examples—where you can generate greater value in your organization:

► High value analyses or reports that are hard to update due to manual steps

Driver: Process efficiencies

Example: Marketing dashboards with cost-per-lead or sales conversion metrics typically combine data from different applications, such as leads management, customer relationship management (CRM), order management, and finance. Many companies assemble separate extracts manually, which can often be a time-consuming process. Hadoop likely will not be needed to solve this problem. However, combining data from different applications, where the customer identifiers differ, is a necessary analytics skill for most big data applications.

► Outcomes requiring cross-functional or inter-departmental views

Drivers: Process efficiencies, public transparency, tax reduction

Example: New York City used a predictive model on data from 19 departments to score which properties were most likely to contain illegal conversions and be a fire risk.[4] By focusing building inspectors on the top 5% high-risk buildings each week, the inspectors' effectiveness in issuing vacancy orders went from 13% to 70% of cases.[5]

► Monitoring related to risk or safety, such as employee fatigue

Drivers: Liability exposure, reduced staff turnover, differentiated service offering

Example: Transportation specialist FleetRisk Advisors provides a unique range of services that are designed to reduce accident rates and driver turnover for trucking and logistics organizations and commercial fleets.[6] Its services use predictive models to assess drivers' monthly workload (for example, distance driven, opportunities to sleep, or salary) against vehicle telematic data to score

---

[4] An *illegal conversion* is when a dwelling zoned for, say, six people is converted into many smaller units, housing as many as 10 times the number of people for which it was designed. Illegal conversions represent a public safety hazard for fire, crime, and epidemiology.

[5] *Predictive data analytics is saving lives and taxpayer dollars in New York City*, interview with Mike Flowers by Alex Howard, O'Reilly Strata, 26 June 2012

[6] For more information, see:
http://public.dhe.ibm.com/common/ssi/ecm/en/ytc03278usen/YTC03278USEN.PDF

drivers for fatigue and job satisfaction. These scores enable its clients to intervene with the highest-ranked drivers, helping to prevent accidents before they occur and reduce employee turnover.

► Coordination of business processes involving multiple systems

Drivers: Improved customer service, price optimization, process efficiencies

Example: Pricing specialists set around hundreds to thousands of prices within a brand portfolio each season. Traditional pricing systems require updating huge volumes of data from partners to combine with customer data, buying history, and other systems. Commodity price optimization analyzed this variety of data to calculate optimal price points based on the desired margin and corresponding parameters specified by the pricing specialist, which accelerates time to results and increases consistency.

Checklist for success

Before you begin your big data project, consider the following checklist for success. These factors can influence the success of any analytics projects but are more crucial when tackling big data challenges.[7]

☐ Culture

The culture within your organization needs to use data in its decision making in order to realize the benefits of any new insights. To assess your organization's analytics maturity, take IBM Analytics Quotient (AQ) quiz.[8] You can find recommendations about how to increase your organization's AQ at the end of this paper.

☐ Skills

To be able to discover new insights from big data, your organization will need the math and technical skills to access and model data and also a genuine interest in rooting through data as well as a willingness to find new approaches.

To develop the big data technical skills, IBM sponsors the Big Data University, an online learning

resource that was created by experts around the world.[9] When skills are more immediately needed, IBM services are available, from lab services for setting up IBM software to Business Analytics and Optimization Services to aid business transformation through analytics with leading technologies.[10]

☐ Architecture

Adopting big data means that data is being shifted from or supplementing existing data stores. In some cases, the big data technology might be viewed as a replacement for another data technology. However, because different data technologies offer different performance and support for analytics processing, optimizing data storage costs needs to consider the organization's analytics requirements.[11]

☐ Data

Data availability refers to its suitability for processing. Data with low availability, such as scanned files, data that is manually collected from old-style meters, or data belonging to another entity, increases the cost of a project. At the outset, scoring the availability of all data inputs will contribute to estimating costs.

☐ Capabilities

Tools or applications for big data extend from descriptive and diagnostic elements that describe past or current scenarios to correlate, score, and model data to forecast what will likely happen. Those predictive insights can then drive automated decision-making to streamline process and put analytics at the point of impact. IBM Cognos® and SPSS fulfill this full range of capabilities, on all your data, from files to relational databases to Hadoop sources.

As with all enterprise-wide projects, executive sponsorship is a must. When proposing a new way of looking at the business, leaders must be onboard to support potential change. If you do not have executive sponsorship, consider waiting to implement your project until it is in place.

---

[7] The factors listed here were influenced by the following comprehensive surveys: *The criteria for the Ease of Use Index*, pp. 124-125, "Big Data: The next frontier for innovation, competition and productivity" report, *The McKinsey Global Institute*, May 2011. Adoption patterns described in "Analytics: The real-world use of big data" 2012 *Big Data Survey*, IBM Institute for Business Value, Oct 2012.

[8] To determine your AQ, visit:
http://www-01.ibm.com/software/analytics/aq/

[9] You can find more information about Big Data University at:
http://www.bigdatauniversity.com/

[10] For information about Business Analytics and Optimization services, see:
http://www-935.ibm.com/services/us/gbs/business-analytics/

[11] For more information, refer to the paper *Architecting A Big Data Platform for Analytics* by Mike Ferguson of Intelligent Business Strategies, which is available at:
http://www.ibmbigdatahub.com/whitepaper/architecting-big-data-platform-analytics

Next steps: How IBM can help

IBM has a long history of addressing big data challenges, from the SABRE system in the 1960s to inventing stream computing in 2009.[12] These advances have included investments in developing analytics prowess from reports and interactive analyses to complex algorithms, support for very large data sets and sophisticated industry-specific applications.

Analytics for big data is a journey, and each organization will progress at a different speed. Value drivers, coupled with proven success factors, dictate which path or paths you pursue and when.

Keep in mind these principles when deciding the next steps in expanding your analytics quotient to incorporate big data.

Think big, start small, brainstorm with others

Like any business decision, it is important to understand context in the scheme of the broader picture. Therefore, take a step back and evaluate the implications of an analytics on big data program on the enterprise and the existing infrastructure. That way, decisions on which projects to pursue first, second, and so on have the context of the bigger analytics strategy.

Need help getting started? IBM Business Analytics and Optimization Services provide both industry and analytics expertise to help you find the first project.[13]

Get it "right"

The right information to the right people at the right time in the right consumption model—embedded in a process, delivered interactively, and so on—is far more successful than simply giving everyone access to data and leaving each person to figure out what to do with it. Part of getting it "right" is making sure that the full range of analytics capabilities—from descriptive to diagnostic to predictive and prescriptive—are brought to task in the most appropriate manner to generate that business value.

Need help matching need to capabilities? Book a visit to an IBM Briefing Center or an IBM Analytics Solution Center to learn more about IBM Business Analytics for big data.[14]

---

[12] From the IBM Icons of Progress series available at:
http://www-03.ibm.com/ibm/history/ibm100/us/en/
[13] For more information, see:
http://www-935.ibm.com/services/us/gbs/business-analytics/
[14] To book a visit, refer to:
http://www.ibm.com/smarterplanet/ca/en/business_analytics/
solution_centers/index.html

Tap information that you have not been able to access

Many organizations begin their analytics on big data journey with internally generated data that they have not been able to cost-effectively manage or analyze. But there is a wide range of possible structured and unstructured data sources within and outside the enterprise that are now at your disposal. Many of these data sources will complement internal data, generate more complete insight, and create value for your enterprise.

Resources for more information

For more information about the concepts highlighted in the paper, see the following resources:

► *Converting Big Data Hype into Big Value with Analytics* by Colin White

   http://www.ibmbigdatahub.com/whitepaper/conve
   rting-big-data-hype-big-value-analytics-bi-re
   search

► IBM Smarter Analytics™

   http://www.ibm.com/analytics/us/en/what-is-sm
   arter-analytics/big-data-analysis.html

► IBM Smarter Analytics solutions

   http://www.ibm.com/analytics/us/en/solutions/
   business-need/

► IBM Business Analytics for Big Data

   http://www-01.ibm.com/software/analytics/solu
   tions/big-data/index.html

► IBM Big Data Platform

   http://www-01.ibm.com/software/data/bigdata/

► IBM Big Data Hub

   http://www.ibmbigdatahub.com/

► IBM Many Eyes® v2

   http://www-958.ibm.com/software/analytics/lab
   s/manyeyes/

► IBM Cognos

   http://www-03.ibm.com/software/products/us/en
   /business-intelligence

► IBM SPSS Modeler

   http://www-01.ibm.com/software/analytics/spss
   /products/modeler/

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document, REDP-5026-00, was created or updated on June 3, 2013.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (or), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at `http://www.ibm.com/legal/copytrade.shtml`

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:
Cognos®
IBM®
Many Eyes®
Redbooks®
Redbooks (logo)
Smarter Analytics™
SPSS®

The following terms are trademarks of other companies:

Other company, product, or service names may be trademarks or service marks of others.