Intel IT Center

MARCH 2013

Predictive Analytics 101: Next-Generation Big Data Intelligence

A Crash Course on Operationalizing Predictive Analytics with Real-Time Data

You're sold on the potential of big data. But how do you make it work for your business? This brief provides you with a crash course on predictive analytics: why it matters, how businesses can operationalize it, the impact on IT, and how Intel can help.

Why It Matters

Big data derives most of its value from the insights it produces when analyzed—finding patterns, deriving meaning, making decisions, and ultimately responding to the world with intelligence. As big data technology continues to evolve, businesses are turning to predictive analytics to help them deepen engagement with customers, optimize processes, and reduce operational costs. The combination of real-time data streams and predictive analytics—sometimes referred to as processing that never stops—has the potential to deliver significant competitive advantage for business.

Predictive Analytics: Looking Ahead

Enterprises have long used business intelligence (BI) for competitive advantage, applying analytics to structured data (for example, transactions and customer information) stored in relational database management systems (RDBMSs). Analytics for big data is different. Big data is characterized by huge data sets and varied data types, both semistructured and unstructured (videos, images, audio, clickstreams, weblogs, text, and e-mail). Plus, big data is generated at a faster rate than most enterprises have had to handle before.

The massive scale and growth of data in general—and semistructured and unstructured data in particular—outstrip the capabilities of traditional storage and analytic solutions, which also do not cope well with the heterogeneity of big data. Organizations may be data rich, but new analytic processes and technologies are needed to unlock the potential of big data.

Analytics for big data is an emerging area, stimulated by advances in computer processing power, database technology, and tools for big data. Predictive analytics is a set of advanced technologies that enable organizations to use data—both stored and real-time—to move from a historical, descriptive view to a forward-looking perspective of what's ahead.

A Perspective on Predictive Analytics

Many experts use the term predictive analytics broadly to describe two types of future-oriented use scenarios for big data: predictive and prescriptive.

- **Predictive analytics** looks into the future to provide insight into what will happen and includes what-if scenarios and risk assessment. It can be used for forecasting, hypothesis testing, risk modeling, and propensity modeling.
- **Prescriptive analytics** is focused on understanding what would happen based on different alternatives and scenarios, and then choosing best options, and optimizing what's ahead. Use cases include customer cross-channel optimization, best-action-related offers, portfolio and business optimization, and risk management.

How Predictive Analytics Works

With big data, analytics is moving from traditional BI methods that use classic sorting with structured data to discovery techniques that utilize raw information. Traditional BI tools use a deductive approach to data, which assumes some understanding of existing patterns and relationships. An analytics model approaches the data based on this knowledge. For obvious reasons, deductive methods work well with structured data.

An inductive approach makes no presumptions of patterns or relationships and is more about data discovery. Predictive analytics applies inductive reasoning to big data using sophisticated quantitative methods such as machine learning, neural networks, robotics, computational mathematics, and artificial intelligence to explore all the data and to discover interrelationships and patterns.

Inductive methods use algorithms to perform complex calculations specifically designed to run against highly varied or large volumes of data. The result of applying these techniques to a real-world business problem is a predictive model. The ability to know what algorithms and data to use to test and create the predictive model is part of the science and art of predictive analytics.

The Rise of Real-Time Analytics

The desire to put real-time data to work in predictive models is on the rise. IT managers participating in Intel's 2012 survey on big data reported that while batch and real-time data processing currently is evenly divided, real-time processing would shift to two-thirds of the workload by 2015.[1]

Continuous streams of data can actually help enhance the efficiency and accuracy of predictive models:

- **Predictive models are informed by historical data that in real time or near real time may only be seconds or minutes old.** In the past, predictive models took time to build and test to exploit those patterns.
- **The higher the volume of data, the greater the opportunity to fine-tune and refine the model.** The relevance of a predictive model is dependent on the accuracy of key coefficients and factor variables—the ability to find correlations. In this way, data analytics differs from more traditional data mining.
- **Automated analytics algorithms, such as machine learning, continuously inform the predictive model and enable it to adjust.** Each adjustment can increase accuracy as the algorithm continues to process and analyze. Depending on the previous human or other machine action, the system can continue to generate new algorithms as needed, ensuring that the model remains relevant.

Companies are already using predictive analytics in real time. For example:

- Grocery stores use data from loyalty cards, ongoing promotions, and historical purchases to offer coupons for items you are likely to buy in the future.
- Predictive analytics applied to specific behaviors of customers browsing a web site enables delivery of a personalized experience.
- Google* Search and Microsoft* Bing* services use machine learning when they offer alternatives to queries during search.

Operationalizing Predictive Analytics

Organizations can use predictive analytics for decision making, solving business problems, and identifying opportunities, including:

- Optimizing business processes and reducing operational costs
- Engaging deeper with customers and enhancing the customer experience
- Identifying new product and market opportunities
- Reducing risk by anticipating and mitigating problems before they occur

While predictive analytics can be used to support major strategic decisions, it's also good for making smaller tactical decisions—the ones that people and machines make on a routine basis. Multiple variables impact any operational decision, and while the impact may be small for each decision (for example, "Should we stock this product?"), the aggregated value is great. Operational decisions are typically high volume and made in real time, but must be able to adapt quickly to market conditions.

Predictive analytics is most effective at the point of decision. Validated predictive rules, which can be enhanced with business rules management systems and complex event processing platforms,[2] can be embedded into enterprise applications such as customer relationship management (CRM), enterprise resource planning (ERP), or industry-specific software. Complex event processing can infer events or patterns based on multiple data sources in order to derive meaning and respond quickly.

Many of the actions resulting from predictive rules are associated with prescriptive actions and run unattended—for example, on web sites and at ATMs and other self-serve systems. Predictive analytics in these cases is strengthened by automated machine learning algorithms that continue to improve the relevancy and accuracy of the predictive model.

Use Cases: Industry Examples

Predictive analytics has the potential to deliver value in a diversity of industries. A few examples of use cases follow.

Industry	Use Cases
Energy and utilities	Energy consumption patterns and management
Financial services	Fraud identification, loan defaults, and investment analysis
Food and beverage	Supply chain demand prediction for creating, packaging, and shipping time-sensitive products
Healthcare	Rehospitalization and risk patterns in health-related data
Insurance	Fraud identification and individualized policies based on vehicle telemetry
Manufacturing	Quality assurance optimization and machine failure and downtime predictions
Transportation	Service and delivery route optimization
Marketing	Consumer behavior prediction, churn analysis, consumption, and propensity to spend
Travel	Buying experience optimization, upselling, and customized offers and packages

The Impact on IT

As if global economic uncertainty, changing market dynamics, and technological innovation aren't enough, IT is now under pressure to figure out how to make big data analytics work. Expanding the delivery of big data analytics services—by moving insight to the point of decision—requires new technologies, processes, and skills. IT can bridge the gap between the business knowledge needed to create relevant predictive models and the technologies needed to help create value.

The Right Infrastructure

The most obvious role for IT is to support advanced analytics infrastructure.

With analytics at the point of decision, high-performance, highly available infrastructure must be in place to handle the heavy demands of big data across users and systems. That includes emerging technologies that can handle grid computing, distributed file systems, tiered storage, and cloud-based data and processing. IT must be able to support sourcing, integrating, and processing data; data streaming; and alternative information architectures such as federated data warehouses and sandboxes for testing.

Big Data Intelligence Starts with Intel® Technology

Intel® technology-based infrastructure is at the heart of big data architecture:

- Intel Xeon® processors
- Virtualization technology
- Security technologies
- 10 gigabit Ethernet networking
- Intel Solid-State Drives (SSDs)
- Support for advanced storage technologies such as compression, encryption, automated tiering of data, data deduplication, erasure coding, and thin provisioning

Plus, enterprise-ready Intel Distribution for Apache Hadoop* software is designed to optimize big data management and processing on Intel architecture.

For more information about Intel technologies for big data, visit intel.com/big data. For more information about Intel Distribution for Apache Hadoop software, visit hadoop.intel.com.

Intel® Distribution for Apache Hadoop* Software at Intel

We use Intel® Distribution for Apache Hadoop* software (Intel Distribution) as part of our own big data and business intelligence (BI) platforms—and not just because it's ours. We tested it with two external Hadoop* distributions against a list of evaluation criteria, including aspects of platform architecture; administration, operations, and support; and the unique value proposition of each distribution.

The Intel Distribution best satisfied our criteria, including those related to reducing barriers to adoption by business groups, and demonstrated the lowest total cost of ownership—from initial costs to ongoing support. The Intel Distribution also supports integration with Intel's:

- Existing data warehouses and massively parallel processing (MPP) systems
- BI reporting tools and analytics engines
- Data tools, such as various extract, transfer, and load (ETL) tools
- Enterprise scheduling and access management tools
- Advanced analytics tools, such as Apache Mahout* software, a machine learning library that includes MapReduce algorithms and integration with the open-source R statistical programming language

In addition, with the Intel Distribution, we were able to design and implement the entire platform in just five weeks.

For more information about Intel Distribution for Apache Hadoop software, visit hadoop.intel.com. Intel Distribution is available worldwide today for evaluation. Technical support is provided currently in the United States, China, and Singapore, with other geographies expected later in the year.

Source: Araki, Assaf, et al. *Integrating Apache Hadoop* into Intel's Big Data Environment.* Intel (February 2013). intel.com/content/www/us/en/it-management/intel-it-best-practices/integrating-apache-hadoop-into-intel-big-data-environment.html

IT also must be able to:

- **Expand analytic capabilities as needed.** New innovative analytics capabilities include illustrative and interactive forms of visualization, complex event processing, rule engines, natural language query, mobile analytics, and gamification. As analytics become more available at the point of decision, the expanded set of users will require social, collaborative, and mobile tools to dig into business data.[3] IT has an important role to play here. For more on consumerization, see the *Planning Guide: Five Steps to Consumerization of IT in the Enterprise*.
- **Assess risk to data security and privacy.** Infrastructure must support compliance and government requirements and prevent leaks of intellectual property. Assessing risk is more difficult with unstructured information. Even if individual sources don't include personal information, triangulation of several data sources could expose individuals or corporate secrets.
- **Develop the skills to bring value to the business partnership.** Big data problems require people with new skills to manage the data and make sense of results—people in short supply. Companies are faced with the decision of whether to develop these skills internally, work with an analytics services provider to supplement in-house skills, or outsource completely. Analytics contest platforms, such as the Kaggle* platform, competitively "crowd-source" the solution to analytics problems, offering a unique other option.

Intel: Growing Big Data Skills In-House

At Intel, people now work in small big data analytics project teams of five multidisciplinary members on initiatives that are estimated to be six months in duration with a $10 million return on investment (ROI) target. This gives team members an opportunity to develop new knowledge and skills while helping to limit initial support of big data projects to those with a focus on higher-value initiatives. We are now moving to more-strategic and higher-business-value projects that can yield $100 million ROI.

Source: Lancour, Paul. "Inside IT: Intel CIO's Perspective on Big Data: Connected Social Media." Connected Social Media (podcast) (September 25, 2012). http://connectedsocialmedia.com/9108/inside-it-intel-cios-perspective-on-big-data/

How Intel Can Help

Intel has infrastructure technology and software that can support the most demanding big data analytics environments. In addition, Intel technology can be found throughout the enterprise, from the data center to the desktop, in mobile devices and machine sensors. Built to meet open-source standards, our technology delivers the performance, interoperability, and flexibility you need to support predictive analytics for a broad range of use scenarios.

To take the next steps with predictive analytics:

- Do your homework. Intel can provide you with practical tools to help you move forward with big data and other related initiatives, such as managing IT consumerization, a trend also referred to as employee BYOD (Bring Your Own Device), and cloud computing.
- Partner with business units to identify business potential.
- Develop a big data strategy to guide your initiatives and define relationships.
- Stay focused on a small set of business problems to create some wins and then grow your practice.
- Build the skills you need. Start with people who understand your business and complement their skills with deep technical and analytics knowledge.

Intel Resources to Learn More

The Intel IT Center provides straightforward, fluff-free information that addresses each of the ways Intel can help IT pros implement strategic projects like big data analytics. For planning guides, peer research, real-world customer references, solution spotlights, and live events about big data analytics, visit intel.com/bigdata.

[1] *Peer Research: Big Data Analytics.* Intel (August 2012). intel.com/content/www/us/en/big-data/data-insights-peer-research-report.html

[2] Gualtieri, Mike. "Let Big Data Predictive Analytics Rock Your World." Forrester (blog) (June 19, 2012). blogs.forrester.com/mike_gualtieri/12-06-19-let_big_data_predictive_analytics_rock_your_world

[3] "Gartner Lists 10 Disruptive Technologies for Business Information Management." Information Management (May 16, 2012). information-management.com/news/mobile-BI-analytics-cloud-Gartner-disruptive-strategy-10022486-1.html

Share with Colleagues

0313/RF/ME/PDF-USA 328807-001

Sponsors of Tomorrow.™