



## Extreme Competition

**Peter Fingar**  
Executive Partner  
Greystone Group

[pfingar@acm.org](mailto:pfingar@acm.org)

Latest book:  
Enterprise Cloud  
Computing: A Strategy Guide  
for Business and Technology  
Leaders Meghan-Kiffer Press  
[www.mkpress.com/ECC](http://www.mkpress.com/ECC)

### ***Big Data + Predictive Analytics = Actionable Business Insights:*** **Consider Big Data as the Most Important Thing for Business since the Internet**

*Adapted from the forthcoming book, Business Innovation in the Cloud  
([www.mkpress.com/bic](http://www.mkpress.com/bic)).*

The IT analyst firm, Gartner, provides a description of big data, “Big data refers to the volume, variety and velocity of structured and unstructured data pouring through networks into processors and storage devices, along with the conversion of such data into business advice for enterprises.” These elements can be broken down into three distinct categories: volume, variety and velocity.

- *Volume (terabytes, petabytes and eventually exabytes):* The increasing amount of business data—created by both humans and machines—is putting a major hit on IT systems, which are struggling to store, secure and make accessible all that information for future use.
- *Variety:* Big data is also about the increasing number of data types that need to be handled differently from simple email, data logs and credit card records. These include sensor- and other machine-gathered data for scientific studies, health care records, financial data and rich media: photos, graphic presentations, music, audio and video.
- *Velocity:* It’s about the speed at which this data moves from endpoints into processing and storage.

Hmm? Nice definition, but what does Big Data really mean as far as its business impact? Ken Rosen, Managing Partner at Performance Works isn’t the least bit happy with Gartner’s definition as he thinks it misses the real point, “That’s like saying ‘New ideas come from electricity moving among brain cells.’ It’s correct, but the emphasis is wrong. It makes sense an IT-oriented firm like Gartner would focus on speeds, feeds, and infrastructure, but executives need a different view.

“Let’s be clear, Big Data is *not* simply dealing with a lot of data! I sympathize with colleagues who hate the name. After all, 1,000 movies is a Petabyte. But 1,000 movies is not a Big Data problem. On the other hand, I just spoke with the CTO of a top Pharmaceutical firm. They have a serious Big Data initiative, and the total data fits on a single hard drive.

“So what IS Big Data? *New meaning from new sources.* Big Data is finding new meaning from new data sources. New meaning that was never practical to find before—because of scale, data format, distribution of data in many locations, the fact that no one thought of looking before, etc. Sources from Lego store purchase patterns to iPhone GPS info. From automobile traffic patterns to Internet data traffic patterns. From weather to earthquakes. From tech support response times to medication response times. It is *easily* as much a new mindset as new technology. Again, *new meaning from new sources.*..

“Why should you care? Because just for a start, businesses can learn what to offer and to whom.

When to offer something new and through what channels. Which employee can best solve a problem and when to get outside help. Which competitor will win and when their stock price will reflect the victory. I'll go out on a limb here: I consider Big Data the most important thing for business since the Internet."

Indeed Big Data is all about delivering new insights to decision makers. Here's an example. As reported in a *Forbes* article, Walmart wanted to find out the biggest-selling items people bought before a hurricane hits. The No. 1 answer—batteries— was not a surprise. But the unexpected No. 2 item was Kellogg's Pop-Tarts. They last a long time, don't require refrigeration or preparation, and are easy to carry and store. As a result of this intelligence, Walmart can now stock up on Pop-Tarts in its Gulf Coast stores ahead of storm season. This is where the reach of new-generation business analytics tools shine by directly helping enterprises make smart decisions.

Historically, data analytics software hasn't had the capability to take a large data set and use it to compile a complete analysis for a query. Instead, it has relied on representative samplings, or subsets, of the information to render results. That approach is changing with the emergence of new big data analytics engines, such as the open-source Apache Hadoop. Hadoop and other such systems provide complete looks at big data sets. Instead of a team of analysts spending days or weeks preparing the parameters for data subsets, and then taking 1, 2 or 10 percent samplings, all the data can be analyzed at one time, in real time. Why bother? Because data sitting in storage arrays and Cloud accounts represents unrefined value in its most basic form. If interpreted properly, the stories, guidelines and essential information buried in storage and databases can open the eyes of business executives as they make strategic decisions for their company.

Apache Hadoop, open-source software, has proved to be the data prospector with the most market traction in the last several years. Hadoop processes large caches of data by breaking them into smaller, more accessible batches and distributing them to multiple servers to analyze. It's like cutting your food into smaller pieces for easier consumption. Hadoop then processes queries and delivers the requested results in far less time than old-school analytics software—most often minutes instead of hours or days.



*Doug Cutting, Apache Foundation Chairman,  
named his new creation Hadoop,  
after his son's big stuffed elephant.*

IBM, the first large systems maker to use the engine, provides its Hadoop-based InfoSphere BigInsights. CEO Sam Palmisano revealed in an August 2011 presentation, "In about a year from now, you'll be starting to see the fruits of our 'big bet' on big data. The work we've been doing for the last several years with Watson [the IBM computer that won Jeopardy! matches against two human champions] will move into products that will be used for a great many purposes, including health care, science and financial applications. Our engineers say they're not far away from building a supercomputer about the size of a human brain that can fit into a shoebox."

What IBM is talking about is Big Data "predictive analytics." The term "business intelligence (BI)" was coined way back in 1958 and companies have relied on BI ever since. But most of the results

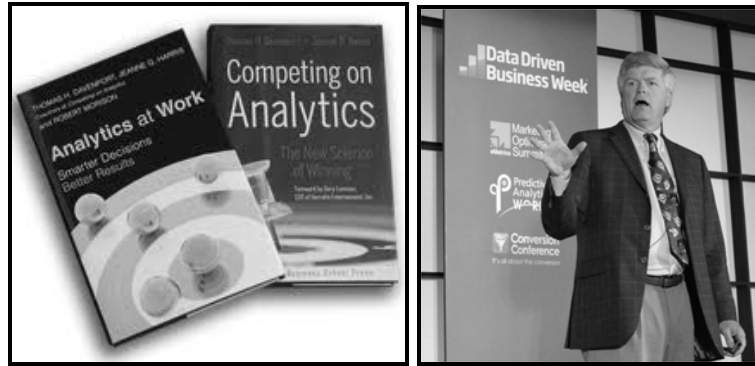
of BI only show *what has happened*, not *what's most likely to happen next*. That's where predictive analytics come in.

Predictive analytics is an area of BI that deals with extracting information from data and using it to *predict* future trends and behavior patterns. The core of predictive analytics relies on capturing relationships between explanatory variables and the predicted variables from past occurrences, and exploiting it to predict future outcomes. While traditional BI can tell us where we've been, predictive analytics can tell us where we are likely to be going so that more rapid and insightful decisions can be made.

The term predictive analytics is used to mean predictive modeling, "scoring" data with predictive models, and forecasting. *Predictive models* analyze past performance to assess how likely a customer is to exhibit a specific behavior in the future in order to improve marketing effectiveness. Amazon serves up a simple example of predictive models that most of us can relate to when we visit its site and it recommends what we might like based on our individual buying history and patterns.

*Decision models*, on the other hand, describe the relationship between all the elements of a decision—the known data (including results of predictive models), the decision and the forecast results of the decision—in order to predict the results of decisions involving many variables. These models can be used in optimization, maximizing certain outcomes while minimizing others. Decision models are generally used to develop decision logic or a set of business rules that will produce the desired action for every customer or circumstance.

Deep analytics may use several quantitative methods broadly grouped into regression techniques and machine learning techniques. These methods apply to many industries, including financial services, insurance, telecommunications, retail, travel, healthcare, pharmaceuticals. The book, *Competing on Analytics*, serves up many stories of exemplars other companies can learn from in their pursuit of competitive advantage through next-generation analytics.



*Tom Davenport, Jeanne Harris and Robert Morison*

Never mind 20th-century focus groups and marketing research surveys. Let the Big Data + Predictive Analytics games begin, for they are not just about computing and databases, they are about a new generation of analytics driving business insights for business innovation in the Cloud.

Of course, to *act* on this new insight requires mastery of business process management, for BPM is how stuff actually gets done in light of new insights. Forget BPMN swim lanes going from as-is to to-be. Start with “What the hell for?”

That’s what big data with deep analytics is all about.

Show me the patterns and trends I couldn’t otherwise begin to fathom.

Give me what IBM-emeritus guru James Martin calls alien *intelligence* that only *deep analytics across big data* can provide.

Then I can give you *to-be processes* that can cope with today’s new world of uncertainty.

**BPTrends LinkedIn Discussion Group**

We created a BPTrends Discussion Group on LinkedIn to allow our members, readers and friends to freely exchange ideas on a wide variety of BPM related topics. We encourage you to initiate a new discussion on this publication, or on other BPM related topics of interest to you, or to contribute to existing discussions. Go to LinkedIn and join the [BPTrends Discussion Group](#).