

The Sexiest

Job of the 21st Century

IT'S CALLED 'DATA SCIENTIST,' AND IT'S IN HIGH DEMAND THROUGHOUT CORPORATE AMERICA.

BY JERRY ROCHE

This year, the demand for data and analytics resources will reach 4.4 million jobs globally, but only one-third of those jobs will be filled, according to researchers at Gartner. The emerging role of data scientist is meant to fill that skills gap.

And the reason for this explosion in what *Harvard Business Review* terms “data scientists”? Big Data — a concept and an approach to business management that most large corporations are now embracing.

One of the nation's largest companies, IBM, believes that “new skills are needed to fully harness the power of Big Data. Though courses are being offered to prepare a new generation of Big Data experts, it will take some time to get them into the workforce. Meanwhile, leading organizations are developing new roles, focusing on key challenges and creating new business models to gain the most from big data.”

A MOUNTAIN OF DATA

In this day and age — when most corporate functions depend on the technology of the Internet and/or the Cloud for easy but secure access anytime, anywhere — data collection is no problem. Consider these statistics:

- >> Twitter generates 1,000 “tweets” per minute;
- >> Google handles 700,000 searches every hour;
- >> Facebook features 700,000 status up-

dates per hour;

- >> YouTube handles 600 new video uploads every minute;
- >> And email messages fly through the technosphere at the rate of 168 million per second.

Big Data is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big Data is arriving from multiple sources at an alarming velocity, volume and variety.

Early adopters of Big Data analytics gained a significant lead over the rest of the corporate world.

“We now live in an age of Big Data,” says Cynthia M. Wong, a senior researcher for Internet and Human Rights. “Our communications and activities routinely leave rich digital traces that can be collected, analyzed and stored at low cost. In parallel, commercial imperatives drive a range of companies to amass vast stores of information about our social networks, health,

finances and shopping habits. The plummeting cost of storage and computing means that such data can be retained for longer and mined for future, unforeseen purposes.”

Know, data collection is no problem but data management is. What do you do with mountains of data available to your learning organization? How do you sift through it? How do you determine what is relevant to learning? How do you incorporate it into an overall learning plan and, subsequently, into individual learning initiatives?

Those are daunting questions, even for the most astute and technologically oriented learning professional. To extract meaningful value from Big Data, you need optimal processing power, analytics capabilities and skills.

RISKS VS. BENEFITS

When mountains of proprietary data are being stored, security is always a question. Because corporate investments are put on the line when the decision is made to integrate Big Data into a corporate philosophy, the first question that C-level executives might ask is, “Is building an advanced analytics capability really worth it?” Until now, the answer to that question has not been resolved.

But a recent Bain & Co. study found that early adopters of Big Data analytics gained a significant lead over the rest of the corporate world. Examining more than 400 large companies, Bain found that “those with the most advanced analytics capabilities are outperforming competitors by wide margins.” Early adopters are:

‘Organizations should already be tapping into social media streams. If you’re not listening, then you won’t be leading.’

—Douglas Laney, Gartner

- >> 2x as likely to be in the top quartile of financial performance within their industries;
- >> 5x as likely to make decisions much faster than market peers;
- >> 3x as likely to execute decisions as intended; and
- >> 2x as likely to use data very frequently when making decisions.

BIG DATA AND LEARNING

The first thing that managers of corporate learning programs must realize is that you are not in this quest alone. You, your business leaders and your information technology (I.T.) leaders must join forces to realize the value of the data at hand. And it won’t hurt to have one of those sexy “data scientists” on loan from the business side.

“Learning adds veracity and value to Big Data,” says Eric Bruner, chief technologist at GP Strategies Corp. “Learning data can include the time of day during which it is referenced, the format/device from which it is consumed, where it’s being consumed, what kind of ratings the different lessons are receiving, how long it’s on the user’s screen, and how many starts and stops the user is making to consume it. It can tell you what your learners like and dislike.”

Where do you find existing Big Data? According to Douglas Laney of Gartner, unused sources of Big Data can be found beneath your feet in the form of operational data you collect during the normal course of business. You can start with data that your learning management system (LMS) already spits out.

“Increasingly,” Laney says, “organizations are looking to extend this data with additional sensors or instrumentation. Many organizations are also finding value in the intersection of operational data with externally available data. This external data comes from a growing cadre of syndicated

data providers and public data, made available by government open-data policies and initiatives by many Western countries over the past few years.”

Of course, Laney adds, organizations should already be tapping into social media streams. “If you’re not listening, then you won’t be leading,” he notes.

Meanwhile, Bruner contends that good learning data:

- 1 is relevant to the user;
- 2 comes in smaller chunks;
- 3 is simple;
- 4 is always available (just-in-time);
- 5 is flexible;
- 6 fits in the work stream; and
- 7 provides a means of interaction.

Furthermore, IBM believes that “insights from Big Data can enable all employees to make better decisions, deepen customer engagement, optimize operations, prevent threats and fraud, and capitalize on new sources of revenue. But escalating demands require a fundamentally new approach to architecture, tools and practices.

Andy Price, CEO of Tamr, claims that data analysis must start with asking aspirational or transformational analytic questions. He contends that the process includes:

- >> Getting a broad and dynamic inventory of all your data, and embracing ambiguity and variety of enterprise data.
- >> Matching workload to appropriate engine and technology.
- >> Using distributed systems, which can radically lower your cost. “You need to build an enterprise infrastructure that is

as inexpensive, scalable and persistent as that of modern Web companies. Put tight spending limits on storage and systems infrastructure.”

- >> Expecting modern and dynamic visualization; that is, iterative versus reporting.
- >> Treating the Cloud as a first-order resource, not just an ancillary one. It can provide the world’s largest high-performance computing and persistence infrastructure through on-demand rental.
- >> Establishing better collaboration between development and operations teams. “Dev-Ops is to the Cloud as systems management was to client/server computing.”
- >> Making sure that you have the appropriate tools to manage the data.
- >> Perhaps using JSON (JavaScript Object Notation), a lightweight data-interchange format, easy for humans to read and write, easy for machines to parse and generate. “Sources will proliferate; embrace them. They are the ultimate evolution of relational and object-oriented technologies coming together that provide a loose, flexible coupling between data access and applications.”
- >> Incorporating bottom-up data/metadata management like curation and integration that incorporates both internal and external data. “We need to get to the point [like Google] that we don’t care where data comes from,” Price says.

CONCLUSION

Big Data’s most important corporate function is to increase the effectiveness of decision-making within the organization and, more specifically, within your learning/training department. It will help you determine the programs and initiatives that are best for your learners, how to proceed with their implementation, and how to adjust them over time.

“Big Data analytics is the application of analytic capabilities (descriptive, diagnostic, predictive and prescriptive) on enormous, varied or rapidly changing datasets,” concludes Gartner’s Neil Chandler. “The application of analytic capabilities combined with the increased scope, content and context of big data — particularly when merged with more traditionally structured datasets — has drastically increased the variety of use cases for decision support, and in some cases, decision automation.” 