COLLABORATIVE APPROACHES NEEDED TO CLOSE THE BIG DATA SKILLS GAP

STEVEN MILLER

Abstract: The big data and analytics talent discussion has largely focused on a single role – the data scientist. However, the need is much broader than data scientists. Data has become a strategic business asset. Every professional occupation must adapt to this new mindset. Universities in partnership with industry must move quickly to ensure that the graduates they produce have the required skills for the age of big data. Existing curricula should be reviewed and adapted to ensure relevance. New curricula and degree programs are needed to meet the needs of industry.

Keywords: Big data, data science, organiztion design, big data jobs, big data talent/skills gap

A recent McKinsey Global Institute study forecasts a significant shortfall in big data skills in the U.S.: "By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions" (Manyika et al., 2011: 3). The big data skills shortage is global, and every region will face similar challenges. These numbers do not tell the whole story, however, because the lack of skilled big data practitioners of all types is limiting the ability of business to derive value from big data (Kelly, 2013). There is a talent shortfall as well in data strategy and in a wide variety of technical data management positions (Olavsrud, 2013; Rowe, 2013; Yerak, 2013), largely due to the shortfall in university, professional, and executive education programs designed to produce the talent needed to fill the growing demand for every type of big data professional.

Although the data scientist role gets the most attention, rebuilding businesses to strategically use data requires redefining existing jobs, creating new specialty roles, and perhaps creating new C-suite or redefining existing C-suite roles. Businesses will need highly skilled and experienced information strategists, data architects, data governance professionals, and visualization experts, as well as chief analytics officers (Advani, 2012) or chief data officers (Aiken & Gorman, 2013), all with broad data skills and most with highly specialized skill sets. In this article, I describe the needed new big data specialists and offer a set of recommendations about how to develop them.

WIDENING THE FOCUS BEYOND THE DATA SCIENTIST

Exactly what a data scientist is and what skill set is required is a subject of intense debate (Harris, 2013). No official standards exist yet, and consequently anyone can call himself or herself a data scientist. Universities and industry need to collaborate to define the profession and set basic standards. Questions that need to be answered include: Do data scientists need to be PhDs, or is data science an applied role? Is it reasonable to expect that individuals possess all required skills or are team approaches preferred? Must every data scientist be an expert in machine learning? What are the core skills required by a data scientist (individual or team), and how are they different from what is being taught in existing statistics or business analytics programs? Until these and other questions are asked and answered, both recruiting

Journal of Organization Design
JOD, 3(1): 26-30 (2014)
DOI: 10.7146/jod.9823

and educating data scientists will be major challenges for industry and academia. Additional considerations include what leadership and management skills (IT and non-IT) are required to ensure that firms and other organizations are leveraging their data assets. Even more important than adding specificity to the data scientist's job and role is a wider focus on other big data jobs. These include information strategists, information systems professionals, and data governance and ethics professionals.

The Information Strategist

What makes some companies more adept than others at leveraging data as a strategic asset (Deutscher, 2013)? One example is Amazon's decision to retain every customer's entire purchase history for operational use while many, if not most, other companies simply store the data for a brief period and then delete it forever. This affords Amazon deep customer insight which enables superior recommendations for future purchases. Information strategy is an emerging discipline that concerns itself with all aspects of data as a business asset (Adler, 2013). Today these business-focused data skills are largely developed on the job as university data management courses concern themselves with technical issues such as data models. Strategic business-level data management is not prominent in most curricula, but there is a pressing need for graduates with both business and industry acumen.

Big Data Information Systems Professionals

Most information systems curricula provide students with a solid introduction to traditional database concepts such as structured query language, relational database, and data warehousing. Businesses, however, need data professionals with broader and deeper analytics skills that enable them to tackle the full breadth of today's data management technologies and challenges, including security, privacy, master data, Hadoop, real-time streaming data, real-time predictive analytics, cloud, and mobile. Businesses need, for example, data professionals who understand when ACID (ACID, 2014) databases are mandatory and when compromises can be made for scalability because eventual consistency is good enough (Bailis & Ghodsi, 2013).

One of the fastest-growing big data jobs is that of data engineer (Data Engineer Job Trends, 2014; Data Engineer Jobs, 2014). The data engineer has deep knowledge of relational databases and NoSQL databases such as Hadoop, can integrate data from diverse data sources, and can design data-driven services. Data engineers work in tandem with data scientists (Walker, 2013). Big data architects (van Rijmenam, 2013) are senior technical staff capable of designing large-scale big data solutions that often span legacy systems and newer systems that handle real-time machine and sensor data (Machine-Generated Data, 2014). Information systems curricula need to evolve quickly to better prepare students for these emerging jobs and roles.

Data Governance and Ethics Professionals

These days failures in data security and governance regularly create public embarrassments for companies (Duhigg, 2012). Even those thought of as among the most tech-savvy have made major blunders. Big data introduces ethical challenges as well. Just because you can obtain the home address of everyone who owns a gun, should you publish that information for all to see (Maas & Levs, 2012)? De-identification is often used to protect the privacy of individuals by deleting or obscuring actual names, addresses, social security numbers, and other data elements which uniquely identify individuals. However, simple de-identification techniques are often insufficient to protect privacy in the age of big data since it is often possible to re-identify individuals by combining data from multiple public sources with your de-identified data (Re-identification, 2014). Data governance is a major challenge for every organization, public or private (Aiken, Allen, Parker, & Mattia, 2007). Unfortunately, data governance is largely absent from academic curricula. Data ethics courses, in particular, are rare and when offered are simply an elective course even though the argument can be made that the ethical handling of data should be a skill for every professional who works with big data. The skills for common data governance jobs, such as data steward and data quality

analyst, are usually acquired on the job, as the data governance profession has not been fully developed by universities.

A COMPREHENSIVE TALENT DEVELOPMENT APPROACH

Business and academia must collaborate to clearly define the big data knowledge and skill sets required across the organization. Every profession, whether business or technical, will be impacted by big data and analytics (Laster, 2010; Miller, 2010; Parry, 2010; Vaidhyanathan, 2010). Top academic business programs, such as the Kellogg School of Management at Northwestern University, are embracing the big data phenomenon (Zettlemeyer, 2013). Law schools and law firms are responding to the need by changing curricula and practices (IPIC Academic Program, 2014). Some universities, such as Lehigh University, have created laboratories where students use technology to analyze data specific to the field of study (Financial Services Laboratory, 2014). Laboratories like these are vital to ensure that students develop a practical understanding of how to apply data and analytics skills in the real world. But focusing solely on data-specific jobs is not broad enough. Data and analytics literacy must become an expectation across all curricula, regardless of the ultimate field or degree pursued. Graduates without big data skills will not be prepared for the business challenges they will face upon entering today's workforce.

Given the growing emphasis on data science and business analytics, businesses will move to aggressively recruit talent and re-train existing staff with a focus on analytics. However, analytics skill alone is not sufficient to succeed. Differentiating your business will require a comprehensive strategy that considers data as a core business asset. The pressures to innovate and differentiate will only become more intense, and consequently the big data talent shortage will become even more acute. A major challenge for many universities is that the skills required by all big data and analytics professionals will require cross-program collaboration to produce the needed "T-shaped professionals" (Brooks, 2012). The vertical stem of the T is a foundation of deep disciplinary skill. The horizontal bar of the T adds the breadth of skill necessary to work across an organization with the ability to influence others, collaborate across disciplines, and develop creative solutions to complex business problems. Universities must respond by adapting curricula to the needs of business. Big data professionals will need deep and broad skills spanning both technical and business domains. Big data information systems graduates will need math and statistics, machine learning, predictive analytics, decision management, computer science and programming, data ethics, information law, information privacy, data security, data and information theory, and visualization and communication (the arts), in addition to the core information systems, database, data warehousing, and data mining education they receive today. For example, not only will business firms and other organizations require big data professionals who understand basic machine learning concepts and can apply existing algorithms to a solution, but those organizations striving to differentiate themselves in an increasingly competitive market will need hard-to-find, highly skilled, machine-learning data scientists who can invent innovative machine-learning algorithms to underpin groundbreaking solutions. Lastly, the challenge is not simply about IT and technology. Jobs spanning the entire business spectrum, including legal, sales, marketing, finance, product development, manufacturing, and operations, will be impacted by the big data phenomenon.

RECOMMENDED ACTIONS

Academic leaders in partnership with industry and government need to assess the rapidly changing technological landscape and create new curricula and programs to develop talent for the increasing number of big data jobs. By working collaboratively with industry partners, academic leaders must evaluate all existing curricula and programs to determine where and how data and analytics knowledge and skills can be infused into the curriculum to ensure graduates have the skills industry requires to compete in a big data world. Academia, industry, and government should join together to create a national consortium to address the big data and analytics skills challenge. That consortium would:

- Create formal definitions of prioritized jobs such as data scientist, information

strategist, big data architect, and data governance professional.

- Establish curriculum requirements and accreditation standards for programs designed to produce the required knowledge and skills for specialty jobs. Use workforce analytics (Ringo, 2012) to provide actionable feedback to ensure that curricula rapidly evolve to meet the evolving needs of industry.
- Set minimum standards for data and analytics literacy required by all students in the age of big data. Create and deliver literacy training via massive online open courses (MOOCs).
- Create open online communities around shared interests to engage industry, government, and academia.
- Partner with industry organizations such as IT-oLogy and the National Consortium for Data Science (National Consortium for Data Science, 2014) to establish strong internship programs and increase collaboration between business and academia.
- Foster the creation of textbooks and courseware to address both literacy and specialized skills at all levels from undergraduate to executive education.
- Establish working groups to address key data policy issues such as information security, individual privacy, and the ethical use of big data.

By working collaboratively, industry, academia, and government can begin to close the knowledge and skills gaps outlined in this article and better prepare students, managers, and professionals for big data jobs of the future.

REFERENCES

ACID. 2014. *Wikipedia*. http://en.wikipedia.org/wiki/ACID (Accessed March 29, 2014)

Adler S. 2013 (February 1). I am an information strategist. *IBM Data Magazine*. http://ibmdatamag.com/2013/02/i-am-an-information-strategist/

Advani D. 2012 (July). Chief Analytics Officer: The newest member of the C-suite. The Atlantic: Smarter CMO Solutions Blog. http://www.theatlantic.com/sponsored/ibm-power-data/archive/2012/07/chief-analytics-officer-the-newest-member-of-the-c-suite/260570.

Aiken P, Allen MD, Parker B, Mattia A. 2007. Measuring data management practice maturity: A community's self-assessment. *Computer* 40(4): 42–50. http://www.computer.org/csdl/mags/co/2007/04/r4042.html

Aiken P, Gorman MM. 2013. *The Case for the Chief Data Officer: Recasting the C-Suite to Leverage Your Most Valuable Asset* (1st ed.). Morgan Kaufmann, San Francisco, CA.

Bailis P, Ghodsi A. 2013 (April 9). Eventual consistency today: Limitations, extensions, and beyond. *ACM queue*. http://queue.acm.org/detail.cfm?id=2462076

Brooks K. 2012 (April 19). Career success starts with a "T." *Psychology Today*. http://www.psychologytoday.com/blog/career-transitions/201204/career-success-starts-t

Data Engineer Job Trends. 2014. Indeed.com. http://www.indeed.com/jobtrends?q=%22data+engineer%22&relative=1 (Accessed March 30, 2014)

Data Engineer Jobs. 2014. *LinkedIn*. https://www.linkedin.com/job/q-data-engineer-jobs (Accessed March 29, 2014)

Deutscher M. 2013 (March 11). IBM's CEO says big data is like oil, enterprises need help extracting the value. *SiliconANGLE*. http://siliconangle.com/blog/2013/03/11/ibms-ceo-says-big-data-is-like-oil-enterprises-need-help-extracting-the-value/

Duhigg C. 2012. How companies learn your secrets. *NYTimes.com*. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0

Financial Services Laboratory. 2014. *Lehigh University*. http://fsl.lehigh.edu/ (Accessed March 29, 2014)

Harris D. 2013 (July 11). Kaggle now has 100K data scientists, but what's a data scientist? *GigaOM.com*. http://gigaom.com/2013/07/11/kaggle-now-has-100k-data-scientists-but-whats-a-data-scientist/

IPIC Academic Program. 2014. Michigan State University College of Law. http://www.law.msu.edu/ipic/program.html (Accessed March 29, 2014)

Kelly J. 2013 (September 19). Enterprises struggling to derive maximum value from big data. *Wikibon*. http://wikibon.org/wiki/v/Enterprises_Struggling_to_Derive_Maximum_Value_from_Big_Data

Laster J. 2010 (May 28). Scholars scale up music studies. *The Chronicle of Higher Education*. http://chronicle.com/article/Scholars-Scale-Up-Music/65709/

Maas K, Levs J. 2012 (December 25). Newspaper sparks outrage for publishing names, addresses of gun permit holders. *CNN.com*. http://www.cnn.com/2012/12/25/us/new-york-gun-permit-map/index.html

Machine-Generated Data. 2014. *Wikipedia*. http://en.wikipedia.org/wiki/Machine-generated_data (Accessed March 29, 2014)

Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C. 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Miller MH. 2010 (May 28). The spoken word, earchable for scholarship. *The Chronicle of Higher Education*. http://chronicle.com/article/The-Spoken-Word-Searchable/65706/

National Consortium for Data Science. 2014. http://data2discovery.org/ (Accessed March 29, 2014)

Olavsrud T. 2013 (September 25). Enterprises confident about tackling big data initiatives. *CIO.com*. http://www.cio.com/article/740321/Enterprises_Confident_About_Tackling_Big_Data_Initiatives

Parry M. 2010 (May 28). The humanities go Google. *The Chronicle of Higher Education*. http://chronicle.com/article/The-Humanities-Go-Google/65713/

Re-identification. 2014. Electronic Privacy Information Center. http://epic.org/privacy/reidentification/ (Accessed March 29, 2014)

Ringo T. 2012 (March 23). Workforce analytics isn't as scary as it sounds. Harvard Business Review Blog. http://blogs.hbr.org/2012/03/workforce-analytics-is-not-as/

Rowe A. 2013 (July 6). The big search for big data candidates. Wanted Analytics Blog. http://www.wantedanalytics.com/insight/2013/07/06/the-big-search-for-big-data-candidates/ (Accessed March 29, 2014)

Vaidhyanathan S. 2010 (June 3). Crunching words in great numbers. *The Chronicle of Higher Education*. http://chronicle.com/article/Crunching-Words-in-Great/65777/

van Rijmenam M. 2013 (May 6). Big data solutions architect job description. Big Data Startups Blog. http://www.bigdata-startups.com/job-description-big-data-solutions-architect/.

Walker M. 2013 (July 2). Data scientists vs. data engineers. Data Science Central. http://www.datasciencecentral.com/profiles/blogs/data-scientists-vs-data-engineers

Yerak B. 2013 (August 25). In growing field of big data, jobs go unfilled. *Chicago Tribune*. http://articles.chicagotribune.com/2013-08-25/business/ct-biz-0825-big-data-20130825_1_big-data-data-sources-data-analysts

Zettlemeyer F. 2013. The four elements of a big data mindset. Kellogg School of Management, Northwestern University. http://www.kellogg.northwestern.edu/news_articles/2013/06272013-big-data-mindset.aspx

STEVEN MILLER

Talent & Skills Ecosystem Strategist
IBM
E-mail: steven.miller@us.ibm.com