accenture

consulting | technology | outsourcing

Institute for
High Performance

The Team Solution to the Data Scientist Shortage

By Jeanne G. Harris, Nathan Shetterley,
Allan E. Alter and Krista Schnell

Point of View

High performance. Delivered.

Today's data scientist shortage won't go away soon, yet companies need them more than ever to leverage the value of big data. The solution is to build teams of data scientists instead of seeking soloists.

If you are looking for brilliant, top-tier data scientists to take your company to big data nirvana, we have news for you. Unless you run a hotshot Silicon Valley company, chances are enough of them aren't going to walk through your door. There's a more realistic approach for the rest: divide up the job and conquer.

Data scientist is the most common term for the often PhD-level experts who operate at the frontier of analytics, where data sets are so large and the data so messy that less-skilled analysts using traditional tools cannot make sense of them. But they are more precisely described as data engineer-scientist-manager-teachers.

Like any curious scientist, they test theories by exploring and running experiments with data. As computer scientists and programmers, they design the intricate models, algorithms and visualizations that can help companies distill insights from huge volumes of chaotic data. Data scientists also do the work of engineers. They acquire external data sets to supplement internal data, and often manage the data they use and maintain the systems which host it. And as their companies' top experts on generating analytical insights, they guide, train and sometimes manage other quantitative professionals, and help general managers understand what they need to know about big data-era analytics.

"Being a data scientist is not only about data crunching. It's about understanding the business challenge, creating some valuable actionable insights to the data, and communicating their findings to the business," says Jean-Paul Isson, the global vice president of predictive analytics and business intelligence at Monster Worldwide, Inc.

Data scientists' behind-the-scenes influence can be massive. LinkedIn Corp.'s data scientists figured out how to recommend "people you may know" to the site's users, thereby launching a feature that has boosted its page views by millions.[1] The data scientists at Monster Worldwide improved customer retention among job-listing employers by 15 percent within a year and boosted spending too, in part by benchmarking the performance of their listings against similar job postings and offering guidance on how to improve performance.[2] Zest Finance, a financial services startup led by former Google Inc. and Capital One Financial Corp. executives, claims it can reduce underwriting risk for credit issuers by 40 percent through big data analysis.[3]

Data scientists are in high demand, but research by the Accenture Institute for High Performance has found the world is facing a severe shortage. There is simply not enough PhD talent to fill the jobs. The shortage is especially severe in the U.S. where 80 percent of new data scientist jobs created between 2010 and 2011 have not been filled, according to our analysis.

And the shortage is getting worse. One reason is that data scientists require a scarce combination of skills. They must master advanced statistical and quantitative methods and tools, along with the new computing environments, languages and techniques for managing and integrating large data sets. Data scientists must also possess industry knowledge and business acumen to create models and solve real-world problems. And they need excellent communication and data visualization abilities in order to explain their models and findings to others. That combination is hard to find.

"Being a data scientist is not only about data crunching. It's about understanding the business challenge, creating some valuable actionable insights to the data, and communicating their findings to the business."

Jean-Paul Isson, Monster Worldwide, Inc.

What's more, the number of graduates with the requisite technical skills isn't keeping up with rising demand. The U.S. is expected to create around 400,000 new data science jobs between 2010 and 2015, but is likely to produce only about 140,000 qualified graduates to fill them. While emerging economies are continuing to produce STEM talent, the shortage in the U.S. alone will still exceed the combined surpluses in China and India combined.[4]

The analytics talent shortage increases the workload on any company's data scientists. Add to that their mentoring work, and any project coordinating they must do, and a scarce resource is stretched like taffy.

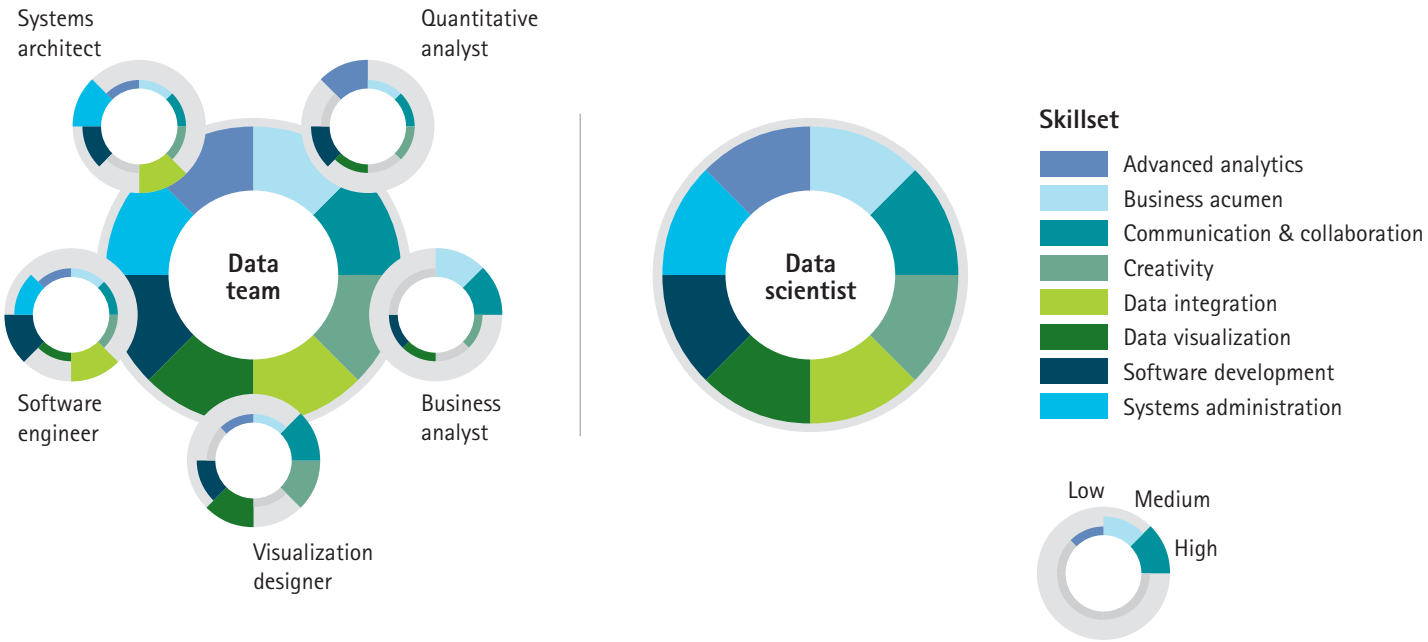
The U.S. is expected to create around 400,000 new data science jobs between 2010 and 2015, but is likely to produce only about 140,000 qualified graduates to fill them.

Training quants to become data scientists and turning to outside sources can help, but both have their limits: training can take years, and many companies cannot accept the risks of opening up their analytical innards to outsiders.

But there is another solution which holds real promise for offsetting the shortage: Create a team of people who individually lack the full skill-set of a data scientist, but as a group possess them all. (See Figure 1.) When physicists take on a big project, they do not build super-colliders and analyze the data all by themselves. They bring together a team to design the equipment, run experiments, do the math and analyze the data. Likewise, it makes sense to divide the labor of a data scientist rather than search for that rare combination in a single person.

Monster Worldwide, LinkedIn and PayPal already employ data scientist teams. Monster has built up a team of data crunchers, statisticians, business analysts, computer scientists and "navigators" who can explain findings to managers. LinkedIn organizes its data scientists as

Figure 1: The 8 Skills of Data Scientists
Executives are struggling to find individuals who possess all eight data scientist skills and abilities. A data scientist team can serve as an alternative.



a product team that includes product marketers, designers and web developers along with mathematicians and engineers.[5] Mok Oh, PayPal's former chief scientist, sought a team of computer science PhDs, leavened with statisticians and MBAs. Other companies are taking this team-building path. American International Group, Inc. is assembling a "Science" team of statisticians, business analysts, project managers, systems architects and engineers for its Property and Casualty business.[6]

The size of these teams will vary: from a handful of people for pilots and short tactical projects, to ten, twenty or more for longer projects and ongoing analytical work. A small team could include one or more software engineers and quantitative analysts who know Hadoop (an open source computing environment often used for big data processing) and can write scripts in the languages used to prepare, integrate, clean, run and analyze big data such as Hive and Pig. It could also include a systems architect to maintain the systems that host the data, and ensure these systems can communicate with one another.

Larger teams add specialists to supplement these roles, dividing the work between them into narrowly defined tasks: programmers in Java and Python that write the commands which prepare data for use, quantitative analysts who can dig deeper into the findings to find the insights, and data visualization specialists who can turn findings into easy-to-understand graphics for sharing outside the team. Larger teams also require project managers/liaisons to oversee their efforts and coordinate the team's work with its sponsors.

On the business skill side, data scientist teams can also include quantitatively oriented business analysts and visualization designers. Business analysts act as business function experts who understand what information is most valuable to the business, and can communicate findings back to functional managers in language they understand. They are the most suitable member of the team to serve as the liaison with different departments. Visualization designers are specialists who can effectively use data to tell stories through graphics. Another option is to create what Isson calls a "navigator," a professional communicator who shares the data and the team's conclusions through straightforward language and graphics. At Monster, each business function is assigned a navigator who serves as its liaison with the team.

Between them, these data scientist teams will have the necessary knowledge of the company's business needs, and the ability to:

- design statistical models for getting desired insights out of the data that is being collected,
- create text mining algorithms for analyzing unstructured data,
- create machine learning algorithms for embedding analytics into business processes,
- clean and convert raw data into formats that can be used by other tools,
- carry out quality assurance testing to ensure the models deliver insights accurately, and
- design easy-to-grasp ways to display insights through data visualization.

Creating teams that are a melting-pot mixture of complementary hard and soft skills is the approach Accenture is taking to build up the scale of its own data science capacity.

Some of these roles may already exist within the company under other names or guises. Companies already have people who clean data and operate systems, and "data stewards" who work with business analysts to manage data and help ensure it's used well. Still, it can be difficult to find even technologists who are familiar with Hadoop and other Big Data technologies. But these teams can be aided by tools that simplify these tasks. These tools are still quite new, and may not necessarily cover all the work that data scientist teams need to do. But as they mature, they will increasingly allow business analysts with less technical know-how to be part of a data scientist team.

Consider a team in the retail industry that's been charged to improve sales from recommendations made to its online customers. The retailer may have stored massive amounts of data in Hadoop, but new tools make it possible for business analysts to still work with the dataset. New tools such as Pivotal's HAWQ, Microsoft Corp.'s PolyBase, Teradata Corp.'s Aster SQL-H and Cloudera, Inc.'s Impala

enable programmers who know SQL, a widely used query language, to prepare data in Hadoop for analysis. The eponymous tools from Datameer, Inc., Jaspersoft Corp. and Pentaho Corp. provide familiar interfaces such as spreadsheets, or reports for analyzing data in Hadoop. These tools remove the need to learn new programming languages. Business analysts could use such tools to discover that people who purchase certain products are much more likely to buy certain others, and then display the results through an easy-to-use data visualization application, such as Tableau Software's Tableau Desktop, QlikTech International AB's QlikView and TIBCO Software Inc.'s Spotfire.

Most organizations already have a great deal of experience managing teams and projects. But executives should keep the following points in mind for creating effective data engineer-scientist teams:

Widen the recruiting pool: Don't just look for people who already have these roles and skills in competing companies. Search outside your industry, and even outside the business world. It can sometimes be easier to find people who have used tools to analyze data in the sciences than highly trained and specialized developers in your own industry. Restless academics with strong analytical skills may also be able to find a new home on a data scientist team. The same is true with physics majors, artists and graphic designers. Physics majors are worth seeking for their mathematical imagination and highly speculative minds. Artists, in particular graphic designers, can bring creativity and imagination to data visualization. Consider as well the possibility of including outside specialists on the team—for example, by teaming up with a university or tapping a crowdsourcing service.

When everyone is learning new skills from their teammates and thus furthering their careers, team members have more reasons to stay put.

Communicate, collaborate, but don't necessarily co-locate: These teams need to work closely together. In an ideal world, all members work together in the same location and even room. But companies should not give up on teams if they cannot co-locate them. Videoconferencing is just the start. Remote workers can be paired up and set up to share the same screen on their computers, so each can enter text on the same command line and see the same windows. And Monster's dispersed data scientists follow a common framework when they produce their findings, says Isson. "We make sure our people around the world have common goals, methods and processes, and a common view of our market and customer base."[7]

Boost effectiveness and retention through team learning: On a data scientist team, it's helpful to encourage members to pick up skills from other members. Over time, this creates flexibility; when one member is unavailable, others can pick up the slack. It can also create a unit that is more resistant to attrition. When everyone is learning new skills from their teammates and thus furthering their careers, team members have more reasons to stay put.

In addition, the time-proven wisdom about managing teams bears repeating: Data scientist teams, like others, flourish best when there is effective leadership, a strong mandate from above and clear goals. They require a path for taking projects from design through implementation. Like many projects in the IT world, they benefit from working in rapid, iterative sprints of preparation, analysis and review. We also recommend starting with short, low-risk projects to learn the ropes before tackling longer, more complex ones.

Businesses are long on experience with managing teams. They will remain short on data scientists. Why shouldn't businesses use what already know to compensate for what they lack?

About the authors

Jeanne G. Harris is the managing director for IT and analytics research at the Accenture Institute for High Performance, and the co-author of Competing on Analytics (Harvard Business Review Press, 2007). She is based in Chicago. Her email is jeanne.g.harris@accenture.com

Nathan Shetterley is an R&D manager with the Accenture Technology Labs. He is based in San Francisco. His email is nathan.shetterley@accenture.com

Allan E. Alter is a research fellow with the Accenture Institute for High Performance. He is based in Boston. His email is allan.e.alter@accenture.com

Krista Schnell is an R&D developer with Accenture Technology Labs. She is based in San Jose. Her email address is krista.schnell@accenture.com

The authors wish to thank Andrew Musselman and Allan Enemark for their contributions.

Notes

1 Thomas H. Davenport and D.J. Patil, "Data Scientist: The Sexiest Job of the 21st Century, " Harvard Business Review, October 2012. http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/

2 Interview with Jean-Paul Isson, Global Vice President, Predictive Analytics and Business Intelligence, Monster Worldwide Inc. and author, Win With Advanced Business Analytics: Creating Business Value from Your Data" (John Wiley & Sons, Inc.; October 2012), September 19, 2013.

3 Jessica Leber, " In a Data Deluge, Companies Seek to Fill a New Role," MIT Technology Review, May 22, 2013 http://www.technologyreview.com/news/513866/in-a-data-deluge-companies-seek-to-fill-a-new-role/

4 Elizabeth Craig, Robert J. Thomas, Charlene Hou and Smriti Mathur, Where Will All the STEM Talent Come From?", Accenture, May 2012 http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Where-Will-All-the-STEM-Talent-Come-From-FINAL.pdf

5 Patil, ibid.

6 AIG job postings on Monster.com, September 19, 2013, jobview.monster.com

7 Isson, ibid.

About Accenture

We are one of the world's leading organizations providing management consulting, technology and outsourcing services, with approximately 275,000 employees; offices and operations in more than 200 cities in 54 countries; and net revenues of $28.6 billion for fiscal 2013. Our homepage is www.accenture.com.

About the Accenture Technology Labs

Accenture Technology Labs, the technology research and development (R&D) organization within Accenture, has been turning technology innovation into business results for 20 years. Our R&D team explores new and emerging technologies to create a vision of how technology will shape the future and invent the next wave of cutting-edge business solutions.

About the Accenture Institute for High Performance

The Accenture Institute for High Performance develops and publishes practical insights into critical management issues and global economic trends. Its worldwide team of researchers connects with Accenture's consulting, technology and outsourcing leaders to demonstrate how organizations become and remain high performers through original, rigorous research and analysis.