

Data Scientist-Hot Victim of Short Term Shortage

Pooja Yadav*

Abstract

With ever increasing amount of data being generated every moment there is a strong need of a person who can not only read but can also interpret, analyze and can evaluate critically to bring out the desired result out of it. Data Scientist is one who can work on enormous data involving data complexity, can query huge databases with different aspects and extract meaningful data with different view using tools and applications. He is a mixture of business analysis, analytics modeling and data management and the scientist. These people are gaining importance and are becoming crucial assets as they give an edge to business in today's competitive era.

The purpose of this paper is to figure out the shortage of data scientist and to find out the remedies to cope up with such scarcity of the intellectual professionals in IT hiring market.

Keywords: Data scientist, hiring, internet, shortage.

Introduction

“By the end of the decade, 50 billion devices will be emitting information nonstop. Data scientist will help manage it all”, by Jessi Hempel.

Every major technology that comes, brings along with it a shortage of skilled labour in the IT hiring market. These shortages naturally put upward pressure on salaries. But this gap between the actual availability and perceived figures also compel or force companies to end up training the existing staff and only hire a few people with specialized skills.

Three decades down the line there was shortage of programmers, two decades same situation for website developers. And today it is the 'Data scientist'- the hot victim of such short term shortage. With petabytes of data being added to the pool everyday, there is real

shortage of skilled workforce who can manage and analyze this big data.

Big data refers to the idea that a company can mine its data and unlock potentially valuable insights. The term data scientist was first coined by Jeff Hammerbacher at Facebook in 2007.

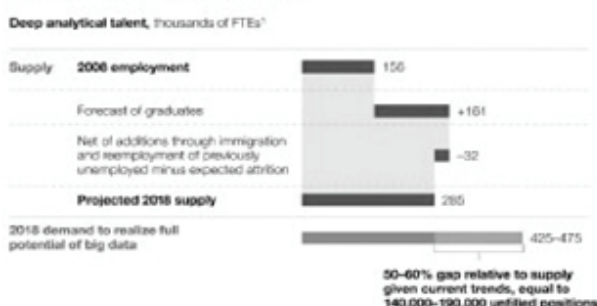
A research by the McKinsey Global Institute forecasts a 50 to 60 percent gap between the supply and demand of people with deep analytical talent. The study projects there will be approximately 140000 to 190000 unfilled positions of data analytics experts in the US by 2018 and a shortage of 1.5 million managers and analysts who have the ability to understand and make decisions using big data.

Other than McKinsey a potential shortfall of data scientists is also spotted by other companies. In a separate study of 500 data scientists and BI professionals, EMC discovered that 65% of the respondents expect demand for data scientists to outstrip availability over the next five years. Moreover, 83% of respondents believe that new tools and emerging technologies will exacerbate the demand for data scientists. Companies like flipkraft which deals with high volumes of transactions daily, generate very high data volumes that requires specialized skill professionals like data scientists.

Define A Data Scientist

“How Obama's data scientists built a volunteer army on Face book” explains one of the best and innovative

Demand in the United States for people with deep expertise in data analysis could be greater than its projected supply in 2018.



*Deep analytical talent are people who have advanced training with statistics or machine learning. FTE - full-time equivalent.

Source: Dun & Bradstreet, company interviews, US Bureau of Labor Statistics, US Census Bureau; McKinsey Global Institute analysis

*Assistant Professor, BCIPS, Dwarka, New Delhi

jobs done by a data science team. No matter how good your social media team is, the chances are it's never done anything like this. Rather than just using Facebook as a channel for posting messages and tracking its followers' feelings, the Obama for America data science team turned social media into a tool for efficiently recruiting the human resources it needed leading into the election's home stretch.

People have been doing data mining for years but that was on premise that the data was quite well organized and lived in big relational databases. But when it comes to handling messy, scrappy data including phone calls, videos and even e-mails, which is mostly unstructured, the DBA's could not be of much help.

While difficult to generalize, there are three main roles which a data scientist plays: data architecture, machine learning and analytics. "Data Scientist" is the new avatar of age old profession of statistician and mathematician. The professionals with core statistics or mathematics background coupled with good knowledge in analytics and data software tools can serve as data scientists.

"Basically, the role of data scientists or analyst is played traditionally by statisticians or mathematicians. Now, with the data growth their role has evolved bigger in organizations and has become more responsible than in past", says Sudeshna Datta, AbsolutData's cofounder and executive vice president. "Today they need to be very tech-savvy so that they can understand data complexity, queries as well as business aspects. It is not just that they will analyse and write reports, but they have to work on huge data volumes- understand and analyse it using software tools and applications. And provide a totally different insight of data", Datta explains.

Data science is a mixture of business analysis, analytics modeling and data management and the scientist should know how to look at large amounts of data to help businesses gain a competitive edge. According to Mike Driscoll, the founder of Big Data company, Metamarkets, "Data scientists must be able to straddle both the business and technical side of an organization. They must be able to take a large data set, model it, and ultimately tell stories from data usually the hardest piece."

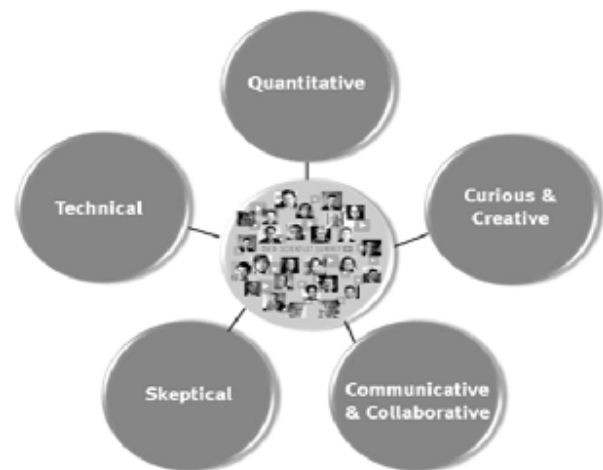
"What makes a data scientist unique is his ability to use technology and hacker skills to solve actual real-world problems, "said Geoff Domoracki, founder of

Dataweek. The skill set required by a data scientist need to be more extensive and refined as compared to a typical DBA (Database Analyst) because he should know the right questions to ask to bring the gem out of the data pool. Data science is also about asking questions interactively and iteratively to search for deep, buried patterns or to get signals from the noise.

Dr. Hadley Wickham cited an example of a business problem which only a fellow data scientist could solve. A data scientist at Progressive Car Insurance noted that in the run up to Halloween droves of people were searching for how to make a Flo costume. Flo is the pitchwomen in Progressive's commercials. The company set up a page dedicated to dressing up as Flo, which led to a huge spike in traffic and an increase in sales.

The data scientist are magicians who work on the data created by high tech systems and give or highlight new or uncovered facts out of it. The key challenge is data science requires higher education. A report by EMC find that 40 percent of data science professionals have an advanced degree, with most of them holding a doctorate.

Profile of a Data Scientist



Hilary Mason, chief scientist for the URL shortening service bit.ly, says a data scientist must have three key skills. "They can take a data set and model it mathematically and understand the math required building those models; they can actually do that, which means they have the emerging skills and finally they are someone who can find insights and tell stories from their data. This means asking the right questions, and that is usually the hardest piece."

Other qualities and traits possessed by data scientists are:

- Technical
- Quantitative
- Curious & Creative
- Communicative & Collaborative
- Skeptical

Technical: the data scientist should possess strong math skills, computer science and business degrees or other intellects in the hard sciences having technical depth to solve big data problems. So academically a data scientist could be from any field with some knowledge of computers, business and academic style research. Fluency with at least one scripting language (PHP or Python) would be an added advantage.

Quantitative: some of the best data scientists are PhDs in esoteric fields like ecology and systems biology, others from the backgrounds in physics and social science. Experience with large data sets makes one the ideal candidate. The role suits to the people with science oriented PhDs.

Curious and creative: the most dominant trait is an intense curiosity- a degree to go beneath the surface of a problem, find the questions and distill them into a very clear set of hypotheses that can be tested. A strong passion for finding and answering hard questions with data.

Communicative & collaborative: ability to communicate the findings is yet another important skill to be possessed. One should be capable of narrating stories with data verbally, visually or ideally both. People with strong social skills are found to be more effective.

Skeptical: A skeptical approach can always help one to look for new findings from the pool of data.

Is there a real shortage?

The October issue of Harvard Business Review portrays an article called “Data Scientist: The Sexiest Job of the 21st Century” by Thomas H. Davenport and D.J. Patil. This article highlights the scarcity of Data Scientists, which cannot be learned in universities, at least not yet. Organizations like Google, LinkedIn, Facebook, and the like can't survive without securing a large amount of data scientists.

So, we need to find out some way to cope up with the

present situation till some kind of courses are designed to train data scientists. Another topic for debate is that “Is there a problem that there are no qualified people or the problem is the qualifications themselves?”. The term data in 'Data Scientist' is so highlighted that at time or rather it is the one creating so much talked about hype and confusion. A data scientist is need not to be a computer expert, he can be from a variety of fields.

The shortage comes into picture when the recruiters confine themselves to people who have some kind of degree in data science as there are rarely any program or degree courses existing with any of the universities to train students or future data scientists. As stated by many experts these data scientists come from a wide range of scientific disciplines such as physics, biology, medicine and meteorology. The chief scientist of Kaggle, Jeremy Howard, has a degree in philosophy.

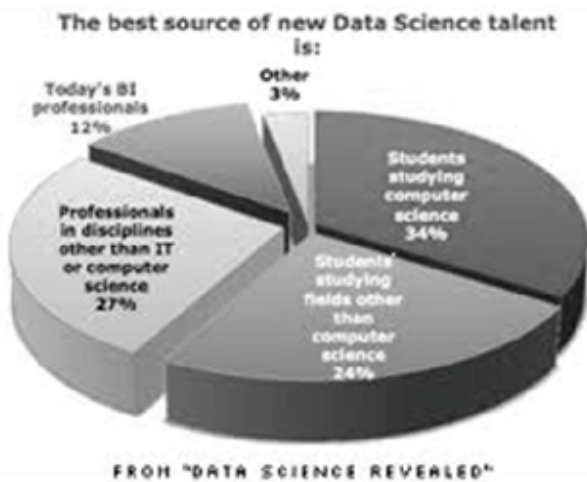
In 2004, India produced around 5900 science, technology and engineering PhDs, a figure that has now grown to some 8900 a year. The hope is that upto 20,000 PhDs will graduate each year by 2020. The number of science doctorates earned each year grew by nearly 40% between 1998 and 2008.

With such a huge number of PhD holders, having the ability to play with data and willingness to learn some computer skills and lots of curiosity, there should not be so much shortage of the talented pool to look from or recruit from. The much talked about McKinsey report projected the number as 190000 as the shortage of data scientist by 2018. And the report clearly states that their raw data is based on SOC code number from 2008, that was just one year after the term 'data scientist' was coined.

How to Hire A Data Scientist?

Data scientist jobs are hot in the era of big data. These scarce and intellectual professionals are responsible for manipulating petabytes of data into better decision-making, new streams of revenue and ultimately more business. A study by McKinsey Global Institute shows that a company using big data to its full potential could increase its operating margin by more than 60 percent, but there's a big shortage of talent.

Most of the companies have realized that this Big Data is more than a buzzword. If planned properly in a through manner than this Big Data can take any business to new, unforeseen and unimaginable heights



in this competitive world. But the most difficult part is, where and how to find the right persons to actually deploy such projects. Despite all the activities of generating data and making projects around the big data, there is a real crunch of locating such skilled labour who are defined as a hybrid of data hacker, analyst, communicator and trusted advisor (a real rare and powerful combination). Companies are under pressure to find and hire people with that perfect equilibrium of business acumen, database expertise and effective communication skills. "Everybody wants these people and it's harder to find them," says John Reed, senior executive director at Robert Half Technology, an IT staffing firm. "Because the demand far outweighs the supply, you will have to go to more sources for the right candidates."

You won't likely find many candidates the phrase "data scientist" on their resume. Some candidates may not even know they are a fit for the data scientist job. Here are a few tips for finding and hiring a data scientist.

Focus on the good universities

No university programs have yet been designed so far as to provide degree courses in data scientists. So, one has to look from what is available. The HR should try to figure out the data scientist from the research scholars of various good universities instead of someone having a good for nothing degree from a nearby study center.

Enlarge the search area

Too often, companies, in addition to looking for the exact skills for a job, are unwilling to allow workers to grow into jobs, as University of Pennsylvania professor Peter Cappelli has written.

Look for achievers in any field with a strong data and computational focus, which might take you as far afield from business as experimental physics or system biology. Candidates who love to play with data and get intimate with it are the ones to be looked for.

Look for a Team Instead of One Person

You may locate candidates with some skills and not others. So one can better concentrate on hiring a team instead of one person, to fulfill or cater all the requirements.

Tom Soderstrom, CTO of NASA's Jet Propulsion Laboratory (JPL), says that although he can clearly define the data scientist role at his company, he knows it may be impossible to find. "It's a special type of person and I've discovered that I don't think they exist." Instead, he says, "It could be a team of people. A data scientist could work with several interns and a community around them."

Link Networking and Technical User Groups

To find candidates for a data scientist job, you need to find out where the candidates spend time. It is recommended to make yourself and your company visible by joining LinkedIn groups, social media groups and technical user groups.

Hunt for internally available talent

DBAs can be trained and motivated to evolve into data scientists and chief data officers who focus on generating the most business value possible out of all the data they collect. DBAs are not the only group within IT that aspire to take on that role, but they do have the inside track when it comes to understanding how the company's data is currently managed which gives them a winning edge over the others.

Don't bother with technical expertise related to computers

The key job requirement in data science (as it is in many technical fields) isn't demonstrated expertise in some narrow set of tools, but curiosity, flexibility, and willingness to learn. Look for evidences that the candidate is self motivated and become skilled rapidly in latest technologies and methods.

Look for good communicators

A data scientist should have the fluency to communicate in a language which the stakeholders understand. In addition he/she should possess the special skills of storytelling out of the data both

verbally and visually. He/She should be good in communicating with numbers. At NASA JPL, Soderstrom sees the potential for data scientists to manipulate satellite data about oceans or weather to develop new types of science experiments. He says he's looking for someone who can both speak the language of business and work with big data technologies, such as Hadoop. "It's someone who can teach data to tell an interesting story that we didn't already know."

Apart from telling stories at times there would be situations when your data scientist needs to talk to your clients or some other business partners, so he/she should be smart enough to represent your business.

Follow up the creativity

Look for the winners of various competitions related to creativity and analytical skills. Organize different events to trace the creative talent and follow up to if you can train some of them to be a part of your organization.

Give lucrative offers

Attract the skilled personnel with money, good and ethically professional sound work atmosphere, central importance and at the same time some space to let them get intimate with the data.

Columbia University has put together its first course with "data science" in the title. In July, the school launched the Institute for Data Sciences and Engineering, according to instructor and course creator Rachel Schutt, a senior statistician at Google and an adjunct assistant professor in the Statistics Department. In addition, Cloudera Chief Scientist Jeff Hammerbacher, formerly head of Facebook's data team, and University of California at Berkeley computer science professor Mike Franklin taught an "Introduction to Data Science" course this past spring. A quick Google search uncovered a couple listing for schools ranging from Stanford and Stevens to Harvard (fall 2013) and the University of Cincinnati that offered "data scientist" courses. Few, though, use the term data scientist. Most are billed as advanced analytics degrees.

But until these and more universities produce more graduates with big data skills to address the shortage, CIOs will need to be creative to find people with the right skills and the right fit for their big data aspirations.

Conclusion

With ever increasing amount of data being generated every moment there is a strong need of a person who can not only read but can also interpret, analyze and can evaluate critically to bring out the desired result out of it. Data Scientist is one who can work on enormous data involving data complexity, can query huge databases with different aspects and extract meaningful data with different view using tools and applications. He is a mixture of business analysis, analytics modeling and data management and the scientist. These people are gaining importance and are becoming crucial assets as they give an edge to business in today's competitive era.

In absence of many or rather any university providing such degrees, the companies have got search very smartly for the skilled labor (Data Scientist) required to manage and utilize the petabytes of data generated everyday.

References

1. Thomas H. Davenport and D.J. Patil. "Data Scientist: The Sexiest Job of the 21st Century", HBR, October 2012.
2. <http://www.enterpriseappstoday.com/author/Herman-Mehling1800.htm>
3. Fogarty, Kevin. "Data Scientist: Hot Big Data Job", Information week - Online (Aug 13, 2012).
4. Fogarty, Kevin. "Data Scientist Jobs Hiding Under Less Sexy Titles", Information week - Online (Oct 1, 2012).
5. www.emc.com
6. <http://www.govtech.com/technology/Industry-Perspective-Is-there-a-Data-Scientist-Shortage-.html>
7. Scott Brave, Baynote." We don't need more data scientists - just make big data easier to use"
8. Bertolucci, Jeff "Hire A Data Science Team, Not A Data Scientist", Informationweek - Online (Nov 30, 2012).
9. Critical Shortage Of "Data Geek" Talent Predicted By 2018, Trends and Outliers (June 28, 2011)