Data science is a team sport. Do you have the skills to be a team player?

IBM

1 Introduction

2 The data scientist

3 The data engineer

4 The developer

5 The business analyst

6 Data science teams: The new agents of change

7 Resources

Introduction

The greatest challenge of the big data revolution is making sense of all the information generated by today's vast digital economy. It's well enough for an organization to collect every slice of data it can reach, but how does it extract value from this massive volume of information?

The more data you have, the better the quality of your reports and strategic recommendations, right? Sure…if you can analyze that data intelligently and quickly, and make it actionable with valuable insights. Otherwise, more data can mean more problems: messy data, storage woes, security risks, frustrated business teams and overloaded IT staff.

So how do you figure out what you really want from your data—and which data can get you those answers?

How do you turn raw metrics and records into information that has (and drives) actual business value? Do you have the skills your business needs to drive results?

Successful organizations build data science teams that incorporate different skill sets and responsibilities, instead of relying on a few elite individuals.

In practice, several people work on a team to build data products. Your analyses will only be as good as the team that is responsible for collecting, building and analyzing the underlying data.

What does the data science team bring to the table? The deeper programming types, such as data engineers and front-end developers, bring consumability to data science. The mathematics and data scientist types use statistical algorithms to find patterns in data. Throughout the process, everything needs to align with the preferred business outcome, which is guided by the eye of the business analyst. Working together, the data science team can outthink today's challenges and problems to create new opportunities and possibilities for tomorrow.

Which talents and abilities define the members of a data science team, and how do they complement each other? **Read on and find out.**

The data scientist

Solving problems and answering questions through data analysis is quickly becoming the norm in today's data-driven world. As real-world experiments become ubiquitous in modern business, the data scientist is evolving into the role that stokes, tweaks and fuels this operational engine.

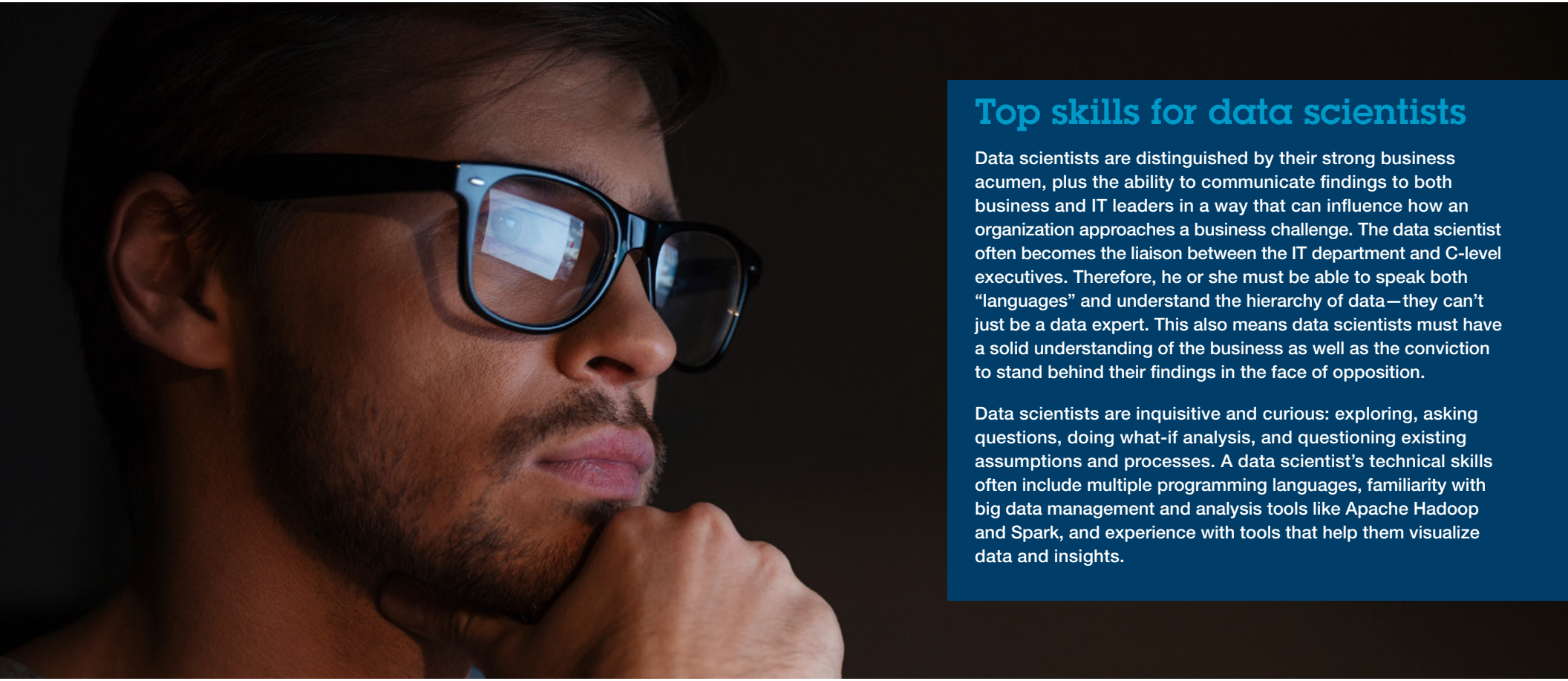
A good data scientist will not just address business problems; he or she will zero in on the problems that have the most value to the organization.

Data scientists are often referred to as "unicorns" because they have a rare combination of talents: they handle a variety of responsibilities and skill sets covering mathematics, statistics, domain expertise, communications and more. Basically, the job of the data scientist is to look for hidden patterns. They accomplish this by applying advanced analytics techniques including (but not limited to) machine learning, modeling, statistics and visualization. Often, data scientists will construct models to predict outcomes or discover underlying patterns; their game plan is to produce actionable insights that can be used to improve future outcomes.

Because data scientists are involved in each step of the journey in building data products, they tend to bring a holistic view to solving problems with data. However, they can't be experts in everything—this is where their team can help.

They experiment continuously by deploying new predictive models, business rules and orchestration logic into next-best-action-powered applications.

A skilled data scientist explores and examines data from multiple disparate sources. They will pore through all incoming data with the goal of linking new information to historical data to find a relationship or trend that offers a crucial competitive advantage or addresses a pressing business problem. They don't just collect and report on data; they look at it from many angles, determine what it means and then recommend ways to apply the findings. They need to make sure their queries are correct and must be able to back up their conclusions with sound models and trusted data—as the data scientist is often expected to present recommendations to management and executive teams.



Top skills for data scientists

Data scientists are distinguished by their strong business acumen, plus the ability to communicate findings to both business and IT leaders in a way that can influence how an organization approaches a business challenge. The data scientist often becomes the liaison between the IT department and C-level executives. Therefore, he or she must be able to speak both "languages" and understand the hierarchy of data—they can't just be a data expert. This also means data scientists must have a solid understanding of the business as well as the conviction to stand behind their findings in the face of opposition.

Data scientists are inquisitive and curious: exploring, asking questions, doing what-if analysis, and questioning existing assumptions and processes. A data scientist's technical skills often include multiple programming languages, familiarity with big data management and analysis tools like Apache Hadoop and Spark, and experience with tools that help them visualize data and insights.

The data engineer

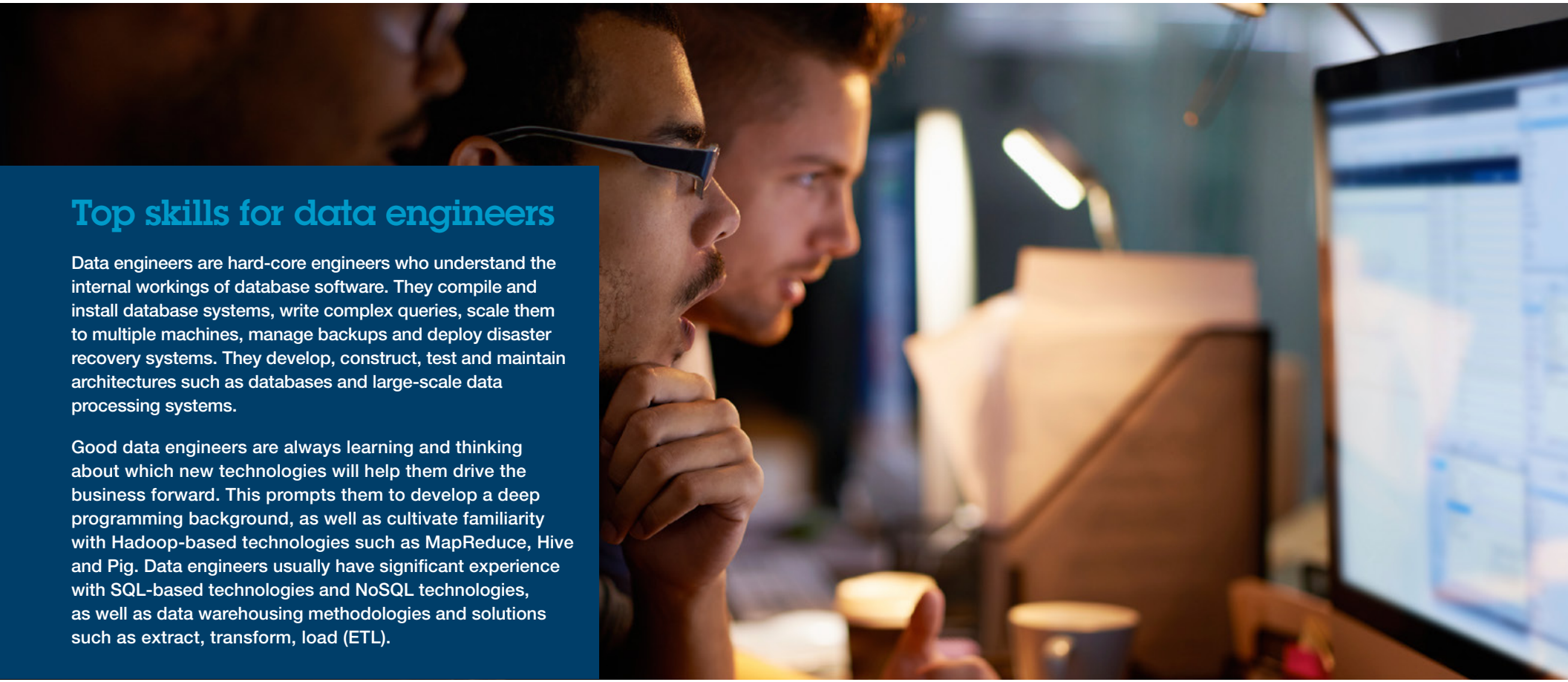
The data scientist may be responsible for uncovering hidden patterns in data, but where do you think they get their data, and what happens when their solutions need to scale to thousands of users or handle sensitive information? Data scientists can only go so far without proper support to operationalize their work.

Enter the data engineer. At a high level, data engineers help gather, organize and clean the data that data scientists will ultimately use to build their analysis. If data scientists extract value from data, data engineers make sure data flows smoothly from source to destination so it can be processed.

Data engineers are responsible for setting up systems and processes that other data workers—including data scientists—use and rely on to work with data. Data engineers must understand how to finesse the flow of data to minimize movement latencies and bring agility to analytics. They also work with front-end developers when moving data science projects into production.

In many organizations, a data engineer will be in charge of integrating data, including designing, building and measuring data ingestion and integration pipelines for large volumes of temporal data from different sources. Examples include database extracts, application server logs, scanned images, voice recordings, Twitter streams, websites and health sensor data. Once continuous pipelines are installed—to and from these huge "pools" of filtered information—data scientists can pull relevant data sets for their analyses.

Data engineers are often tasked with laying the groundwork for a data analyst or data scientist to easily retrieve the data needed for their evaluations and experiments.



Top skills for data engineers

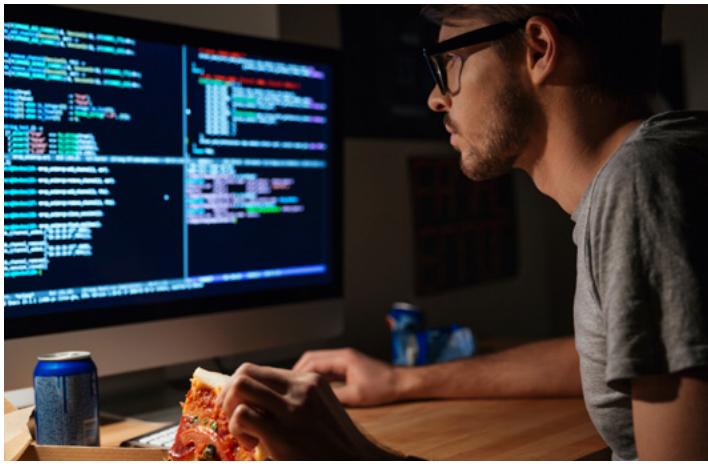
Data engineers are hard-core engineers who understand the internal workings of database software. They compile and install database systems, write complex queries, scale them to multiple machines, manage backups and deploy disaster recovery systems. They develop, construct, test and maintain architectures such as databases and large-scale data processing systems.

Good data engineers are always learning and thinking about which new technologies will help them drive the business forward. This prompts them to develop a deep programming background, as well as cultivate familiarity with Hadoop-based technologies such as MapReduce, Hive and Pig. Data engineers usually have significant experience with SQL-based technologies and NoSQL technologies, as well as data warehousing methodologies and solutions such as extract, transform, load (ETL).

The developer

Although developers don't always have a fully dedicated position within a data science team, they are an essential piece of the puzzle. A successful project usually productizes the data science work so it can serve an internal stakeholder or external customers. Baking data science into business processes is how companies build competitive advantage through data.

Developers often come in at the end of the data science workflow. They are responsible for building the applications where the models will sit—applications that leverage the insights and data gathered from the rest of the data science team. This requires a fair amount of programming time, and it can be a challenging job. Lack of integrated technologies can stifle developers' progress, making it difficult to embed the data science team's analysis.



Developers bring consumability to data science, allowing internal and external stakeholders to take advantage of data and analytics in their day-to-day work.

Another requirement: developers must have a full component of programming skills at the ready. This includes expertise in building web services, front-end development skills and strong knowledge of user interface functionality and features. In addition, they should be familiar with application programming interfaces (APIs) and using them to integrate various data products and sources into applications. When done right, the developer's skill creates an application so solid that you don't even realize there's a data product underneath.

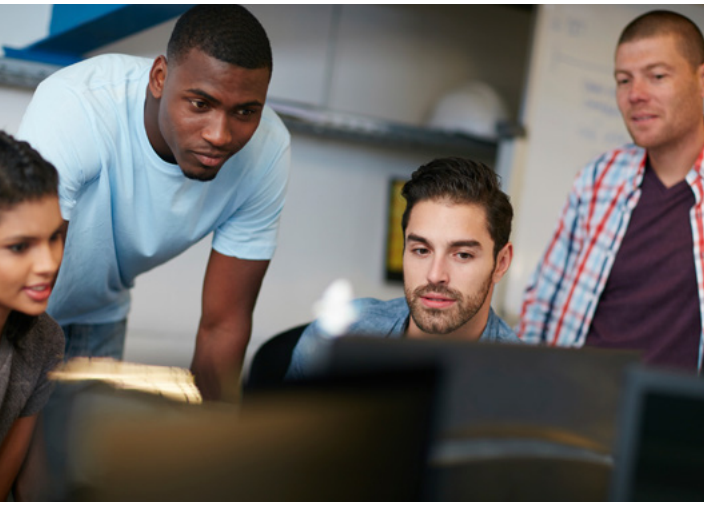
To build game-changing mobile, web and enterprise applications that will disrupt markets, developers need the data—and the tools—to turn their vision into reality. The developer must collaborate with the data scientist, data engineer and business analyst to ensure alignment between the business objectives and the analytics back end.

The business analyst

The business analyst, sometimes referred to as the citizen analyst, provides business-level expertise and guidance to the data science teams. Their responsibility is to apply domain knowledge and make data-informed decisions.

The business analyst understands what the business needs, but doesn't have the technical background to develop a detailed analysis. Improvements in technology and deeper understanding of the user profile have turned them into "analytical consumers," allowing them to make decisions based on data insights. Flexible and user-friendly technologies allow them to develop some business-level analysis without coding and without having to ask the data scientist. This gives business analysts more independence, increasing the chance they will use the data science team's finding in their everyday analyses.

Data science teams: The new agents of change



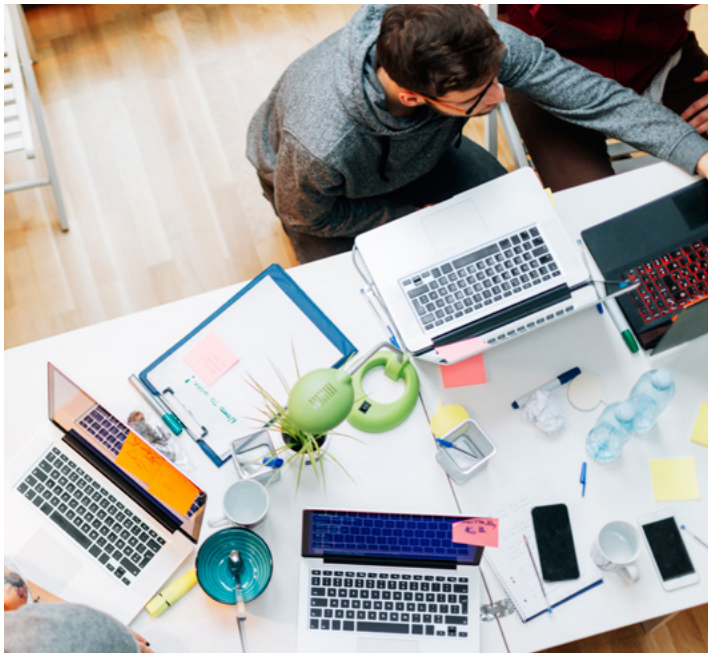
Skilled data scientists, data engineers, developers and business analysts are transformative figures in modern business. They are the beating heart of the big data economy. It's not just that they are designing new systems; they are going to bat for new sources of data and new ways to use that data. Of course, IT still has to build the system, but the data science professionals are the ones who help departments collaborate to solve problems and speed innovation.

The best data products are the ones that the end user doesn't even notice. This requires careful thought about the interface, the data that's captured and the automation enabled by the analytics back end. The technology to collect and analyze massive volumes of business data is available now, and you can exploit it to your company's benefit. Are you ready to be a team player in the high-stakes game of data science?

Training the data science professionals of the future

To remain prepared and equipped to be a solid team player, you need to improve your stats by building knowledge and enhancing your skills. IBM is one of many companies that is helping data science professionals get better and faster at doing their jobs.

Any rapidly growing field needs professionals with new skills and expertise. IBM has been an active supporter of the data science community, and plans to continue this support in online and in-person educational forums. Through events, meetups, courses, content, contributions to the open source community and more, IBM supports today's soon-to-be and current data professionals, helping them prepare to meet the high-volume, high-speed data demands of the future.



Resources

Ready to boost your data science skills? Here are resources to get you started:

1. Build your data science skills with Big Data University
2. Get started with the tools you need with the IBM Data Science Experience

IBM Analytics
Route 100
Somers, NY 10589

Produced in the United States of America
June 2016

Please Recycle

CDM12348-USEN-00