

Closing the Data Skills Gap



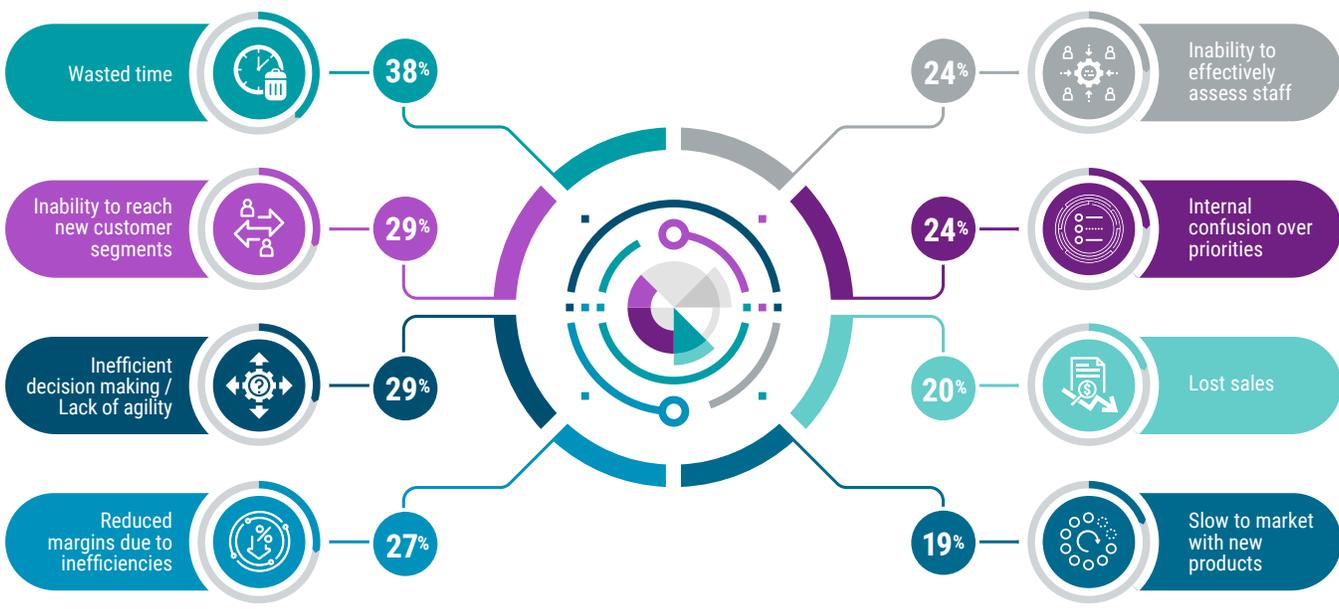
Much has been made of the importance of data in the digital economy. Digital data has obviously been present for decades, but it has taken on a greater priority in an age of artificial intelligence, predictive analytics and customer personalization. Data has been called the “new currency” and the “new oil” of business—an asset so critical that it undergirds the entire operational structure. Yet for all the value being ascribed to data, companies are still in the early stages of building their data discipline and skills.

Why is data so important to business today? Simply put, poor data management or insufficient data analysis is impacting the bottom line. Organizations have been collecting data, but there has typically not been a comprehensive operational approach or a focus on the necessary skills. This has led to abundance without intelligence. Most companies have data silos in every department, limiting the ability to build a holistic view of corporate data. Data is streaming in from new sources such as social media and smart devices, but there is no structure for translating this data into business decisions. Although there is a profusion of historical records, there is a clear

shortage of expertise in using these records to build future plans.

Companies are starting to feel the effects of poor data management or insufficient data analysis. Chief among these is wasted time. As different pieces of an organization have to hunt for the data they need, it consumes time that they could be using to focus on their core functions and new innovations. Increased efficiency, another common goal for technology, comes as a result of well-designed systems and workflow. Data management and analytics should be treated as a comprehensive program, not point tools for specific purposes.

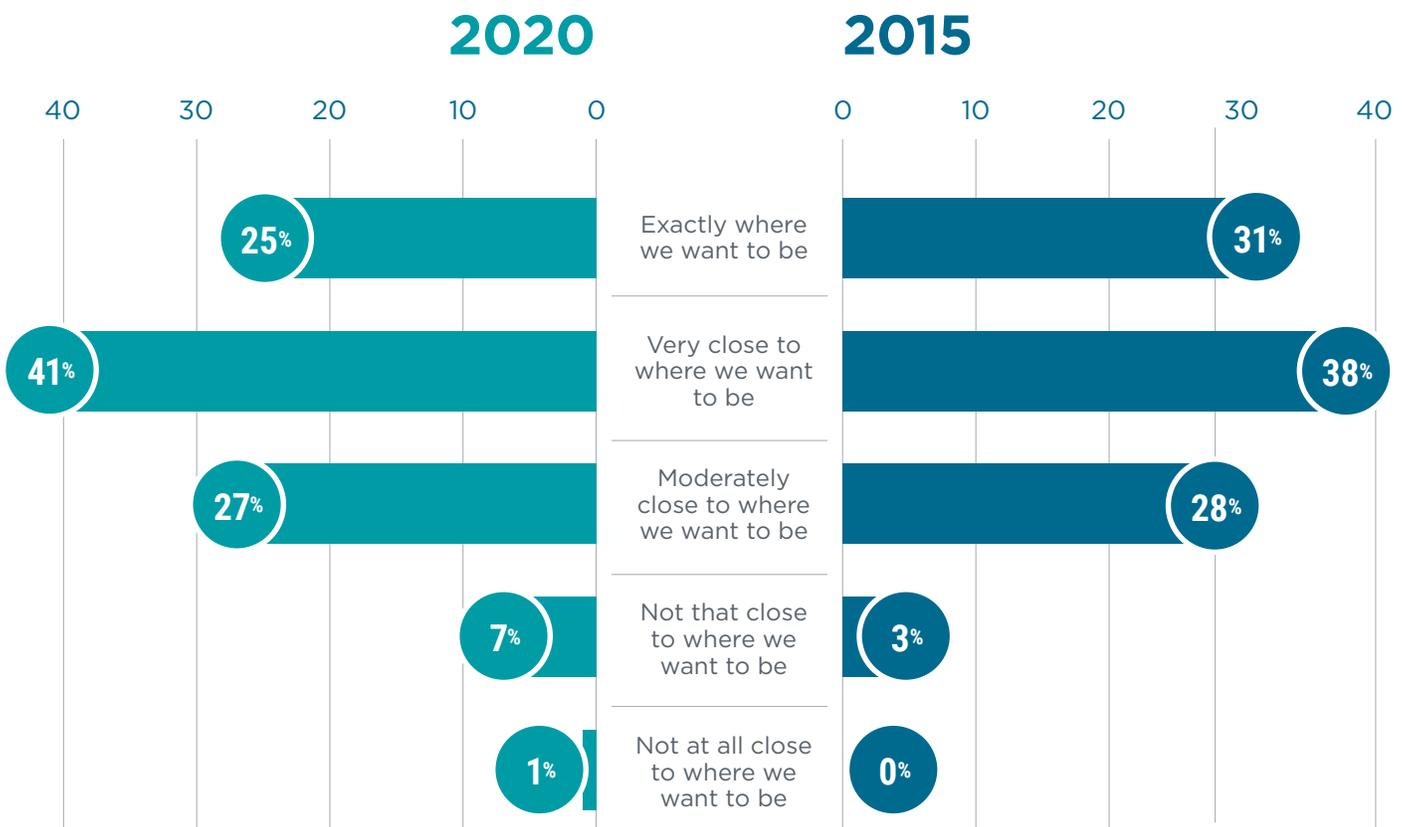
Shortcomings Due to Poor Data Usage



Only one-quarter of companies feel they are where they need to be when it comes to managing and analyzing data.

As with most technology trends, recognizing data as a high priority does not automatically lead to improved capabilities. In fact, there is often a step backwards as companies realize how much ground they have to make up. CompTIA's most recent data shows that only a quarter of companies feel that they are exactly where they need to be when it comes to managing and analyzing their data. Given the pace of digital transformation, this number may drop further before best practices are established.

Views on Current Data Capabilities



In order to get the highest value from real-time data and large data sets, organizations need to focus on their data skills. As more and more companies use data to improve their internal operations and to better understand their customers, new and improved skills will drive data success. These skills address a wide range of business problems, from building a resilient data architecture to improving the speed of data analysis to mining the data for new insights. In essence, the new currency of business requires new specialists to extract value for stakeholders.

If data is truly a critical resource, then it needs to be handled properly. If satisfaction with data capability is relatively low, then there is plenty of room for improvement. A solid understanding of the data function provides the context for building a corporate strategy, and an appreciation for the primary data job roles guides the process for building skills.

As more companies use data to improve internal operations and better understand their customers, new and improved skills will drive data success.



Focusing on the Data Function: Why Do Businesses Need Data?

CompTIA's Functional IT Framework describes four pillars of enterprise technology—areas that companies are managing as standalone functions throughout the organization rather than throwing everything technical into one big bucket inside an IT department.

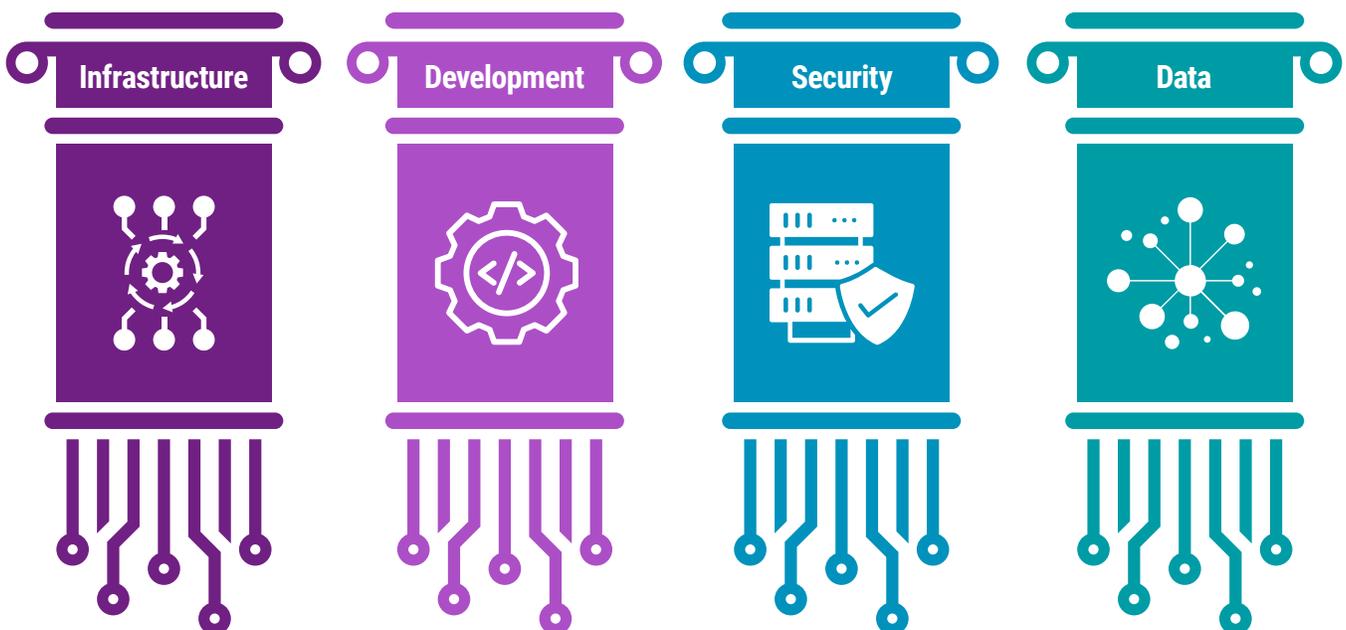
The most foundational pillar is infrastructure—the physical (and virtual) hardware components that host systems and connect employees. On top of this foundation, software development builds the applications that provide functionality. These two disciplines have had distinct skill requirements and operational processes for decades.

More recently, cybersecurity has become a specialized field as companies have focused more on securing digital assets and managing risk. Historically, cybersecurity was handled within the infrastructure function, so there is a natural connection between these two areas even as cybersecurity develops into a standalone function.

Data is the youngest of these four pillars, but it is growing fast. Industry observers believe that the first chief data officer (CDO) was appointed in 2002, and a 2018 study from NewVantage Partners showed that the number of firms employing a CDO had reached 62.5% after hovering at a mere 12% in 2012.

Data science is one of the hottest topics in technology, but it is also a recent creation, born during the big data craze. Earlier, there was not an equivalent “data engineer” role in the same way that companies employed network engineers or server administrators.

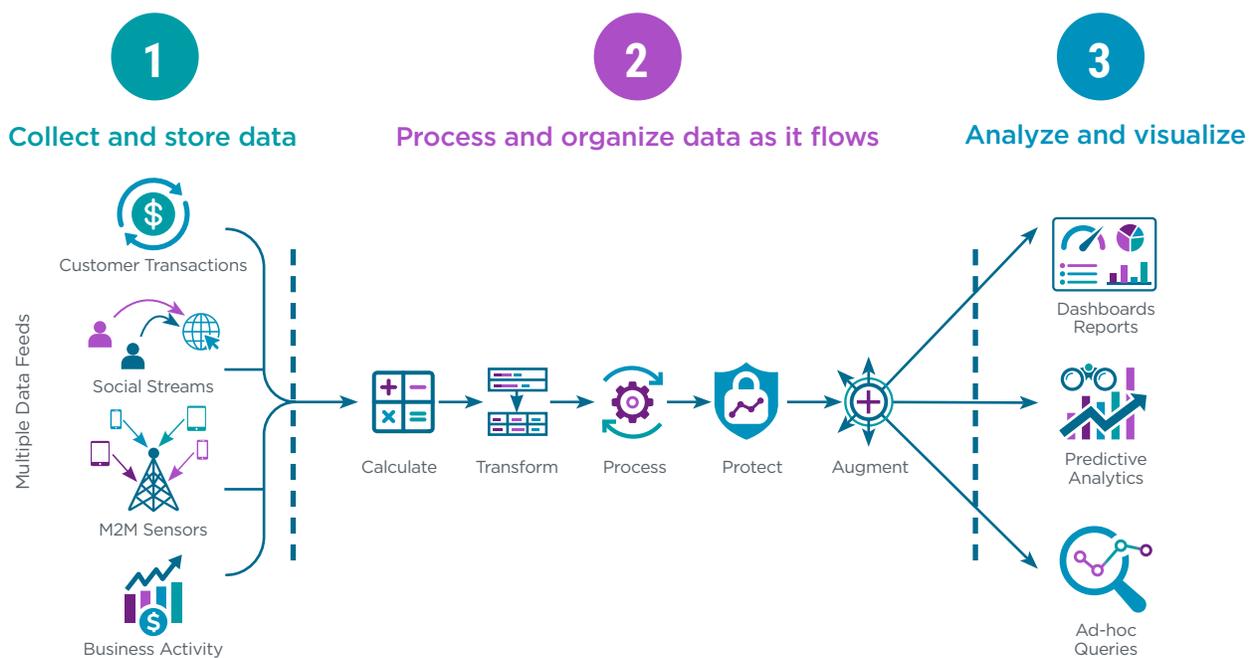
The Functional IT Framework



In many firms, the data function starts off as an offshoot of software development. The skills and critical thinking needed in development translate well to data, where there is an abstract component of dealing with bits and bytes. In addition, many data specialists use the same tools as software developers, including programming languages such as Python or Javascript. The recent acceleration in the amount of data and the types of data that a company can manage has brought focus to certain specialized skills, but there is a foundation that must be built before moving to more advanced applications.

The recent acceleration in the amount of data and the types of data that a company can manage has brought focus to certain specialized skills.

Three Stages of Data Usage



That foundation is the basic management of data, including an understanding of all data throughout an organization, general database proficiency and most critically the analysis of corporate data. Breaking overall data activity into three foundational stages allows organizations to ask pertinent questions: Where does our data come from? How quickly do we want to process different data streams? What insights need to be provided as the data is analyzed? Digging into these questions helps companies understand the technical challenges involved in modern data management and analysis.

Many organizations, especially in the SMB space, do not have strong database skills and instead rely on Excel spreadsheets or other simple tools for analysis. As companies grow in data sophistication, they improve both their tools and their skills to efficiently manage the flow of data and perform analysis. The starting point for serious data activity involves two well-established tools: relational databases and SQL (structured query language).

Relational databases are the traditional mechanism for storing data in a useful format. Data records are placed into the database with defined relationships between certain parts of each record. In order for these relationships to be understood, the data must be structured consistently within the database. This limits the type of data that can be stored and the analysis that can be performed. The most common tool for manipulating the data in a relational database is SQL. SQL has a long history and is well-known among developers for its ability to operate on centrally managed database schemas and indexed data, but it also has its limits.

As data volumes grow, the architecture of SQL applications that operate on monolithic relational databases becomes untenable. Furthermore, the types of data being collected no longer fit into standard relational schemas. Two classes of data management solutions have cropped up to address these issues: NewSQL and NoSQL.

Just as the data toolset is growing,
so is demand for data skills
throughout the organization.

The data function, like software
development, may have outposts
within business units.

NewSQL allows developers to utilize the expertise they have built in SQL interfaces, but directly addresses scalability and performance concerns. SQL systems can be made faster by vertical scaling (adding computing resources to a single machine), but NewSQL systems are built to improve performance of the database itself or to take advantage of horizontal scaling (adding new machines to form a pool of resources and allow for distributed content). Along with maintaining a connection to the SQL language, NewSQL solutions allow companies to process transactions that are ACID-compliant, which ensures database validity.

To handle unstructured data, many firms are turning to NoSQL solutions, which diverge further from traditional SQL offerings. Like NewSQL, NoSQL applications come in many flavors: document-oriented databases, key value stores, graph databases, and tabular stores. The foundation for many NoSQL applications is Hadoop. Hadoop is an open-source framework that acts as a platform for big data—a lower-level component that bridges hardware resources and end user applications.

Just as the data toolset is growing, so is demand for data skills throughout the organization. The data function, like software development, may have outposts within business units. Different departments may have different needs for data and may use their own employees to sift through information and determine a course of action. This distribution of data talent can create data silos and lead to other organizational challenges as a business determines how to best allow flexibility while maintaining overall technology guidelines.

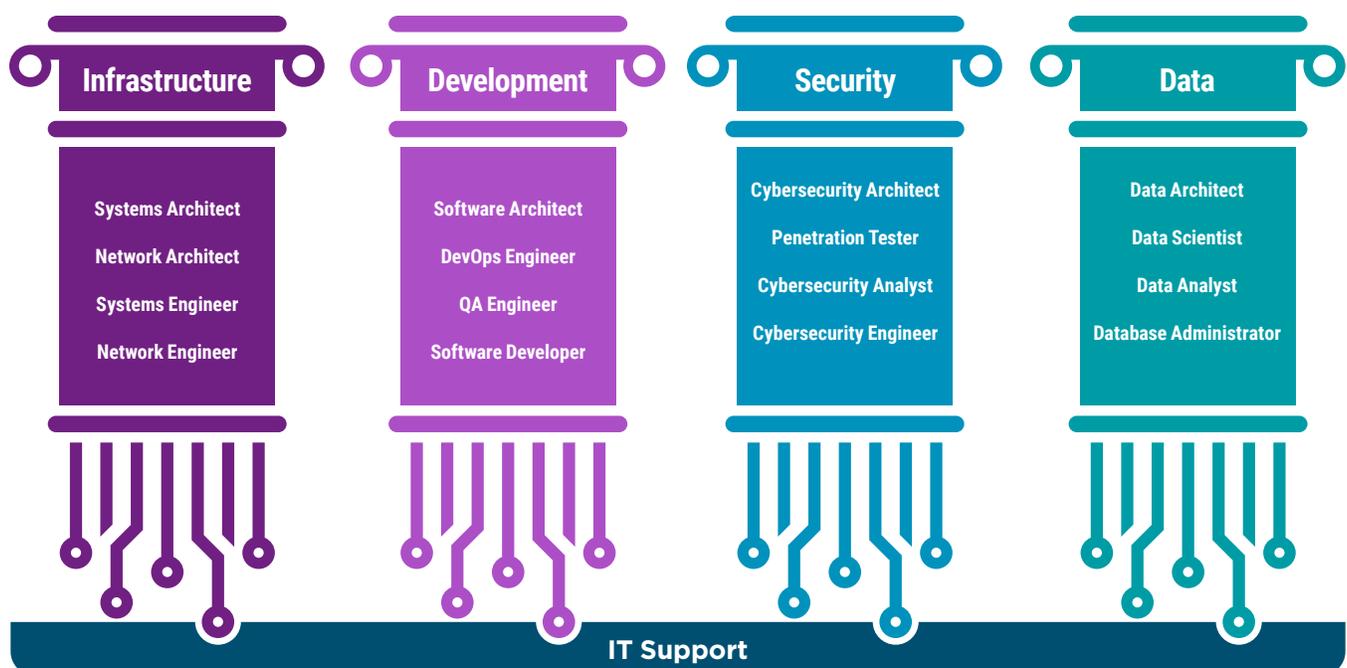
Forming Data Teams: Job Roles and Required Skills

Whether all data skills are centrally located within an IT department or spread across multiple business units, companies are taking steps toward establishing data teams. The concept of data teams is relatively new. Given that the CISO position was created in the mid-1990s and the CDO position was created in the early 2000s, it makes sense that data teams would be lagging behind security teams. Only 44% of companies say that they have internal employees who are dedicated to data management or data analysis. Among those companies that already have dedicated employees, there is still high demand to develop skills around analysis that drives new business value.

When building a data team, the obvious question is which players need to be added. Using data from Burning Glass Labor Insights, CompTIA has defined 17 distinct technology roles that exist today. This definition is based on volume and skill set. There is endless variation in the titles used from company to company as each firm has

individual needs based on their industry vertical or existing expertise. However, examining the key skill sets used across job postings leads to consolidation around a smaller set of job roles that make up the vast majority of current postings. The 17 roles not only describe unique job responsibilities but also convey scope and complexity.

Foundational Technology Roles



For the data function, there are four distinct roles that can help fill out an organization's data team:

- **Database administrator:** The most traditional of the data roles, many database administrators (DBAs) started out within a software development function (and many DBAs may still sit in that function). Although the role focuses heavily on using SQL and scripting to manipulate data residing in relational databases, there is often an infrastructure component to this role as well since DBAs need to understand the systems that are hosting the data structures. Common skill clusters include database administration, SQL databases and programming and systems administration.
- **Data analyst:** While the DBA may currently be the most common role since it is the most well established, data analysts lead the pack in terms of demand. Most companies are ready to move beyond the basic insights that come from standard database manipulation, and they are looking for connections between the data and business operations. The lines are blurred between "data analyst" and "business analyst," but at the end of the day businesses are seeking to build core skills in both IT departments and business units. Common skill clusters include data analysis, business process and analysis and data visualization.
- **Data scientist:** Just as the field of data is the most nascent of the four technology pillars, data scientists are the newest entry among the 17 major job roles. The key differentiator between data analysts and data scientists is the use of statistical modeling to drive predictive insights for the future. Recently, machine learning algorithms have started playing a significant role in building data models. This role is often cited as one of the fastest-growing, but that is largely because

When considering the areas of the data function where businesses are seeking improvement, there is a clear priority: Companies want to improve analytics capability.

the current base of true data scientists is so small. Common skill clusters include statistics, big data and machine learning.

- **Data architect:** Across all four pillars, the main characteristics of the architect roles are very similar. These individuals may come from any of the specialized roles, but they have at least a working knowledge of each data discipline along with some familiarity with other technology functions. Since the role of the architect is to plan and build the appropriate systems, they are more familiar with business tradeoffs and leadership. Common skill clusters include project management, system design and implementation and communication.

When considering the areas of the data function where businesses are seeking improvement, there is a clear priority: Companies want to improve analytics capability. Although some of the desired improvements fall into the category of data management or database administration (search capabilities or remote data access), the lion's share of the wish list revolves around better analysis. Speeding up analytics, applying analysis to specific business activities and finding patterns in customer behavior are all examples of problems that a data analyst can help solve to improve operations and decision making.

Businesses Seek Improvement in Many Areas

Data Capability	Currently Doing Well	Doing, but Want to Improve	Want to Start Doing
Real-time analysis of incoming data	40%	42%	13%
Search capabilities across corporate data sources	39%	37%	16%
Email marketing campaign effectiveness	38%	37%	13%
Social media monitoring	37%	33%	16%
Visualization capabilities	35%	37%	17%
Website traffic patterns	34%	38%	15%
Customer profiling and segmentation analysis	33%	47%	11%
Remote or mobile access to corporate data	33%	43%	14%
Predictive analytics	33%	42%	17%
Metrics and key performance indicators (KPIs)	33%	38%	17%
Pattern recognition	32%	38%	18%
Relationship analysis	29%	42%	16%

Generally speaking, data mining is the activity that connects the two worlds of data management and data analysis. While database administrators hold primary responsibility for building the environment, analysts also need to understand these concepts, especially in a business where all the disparate data sets have not yet been assembled into a cohesive data lake. Once analysts understand the structure, they can begin digging through the data to find connections and patterns.

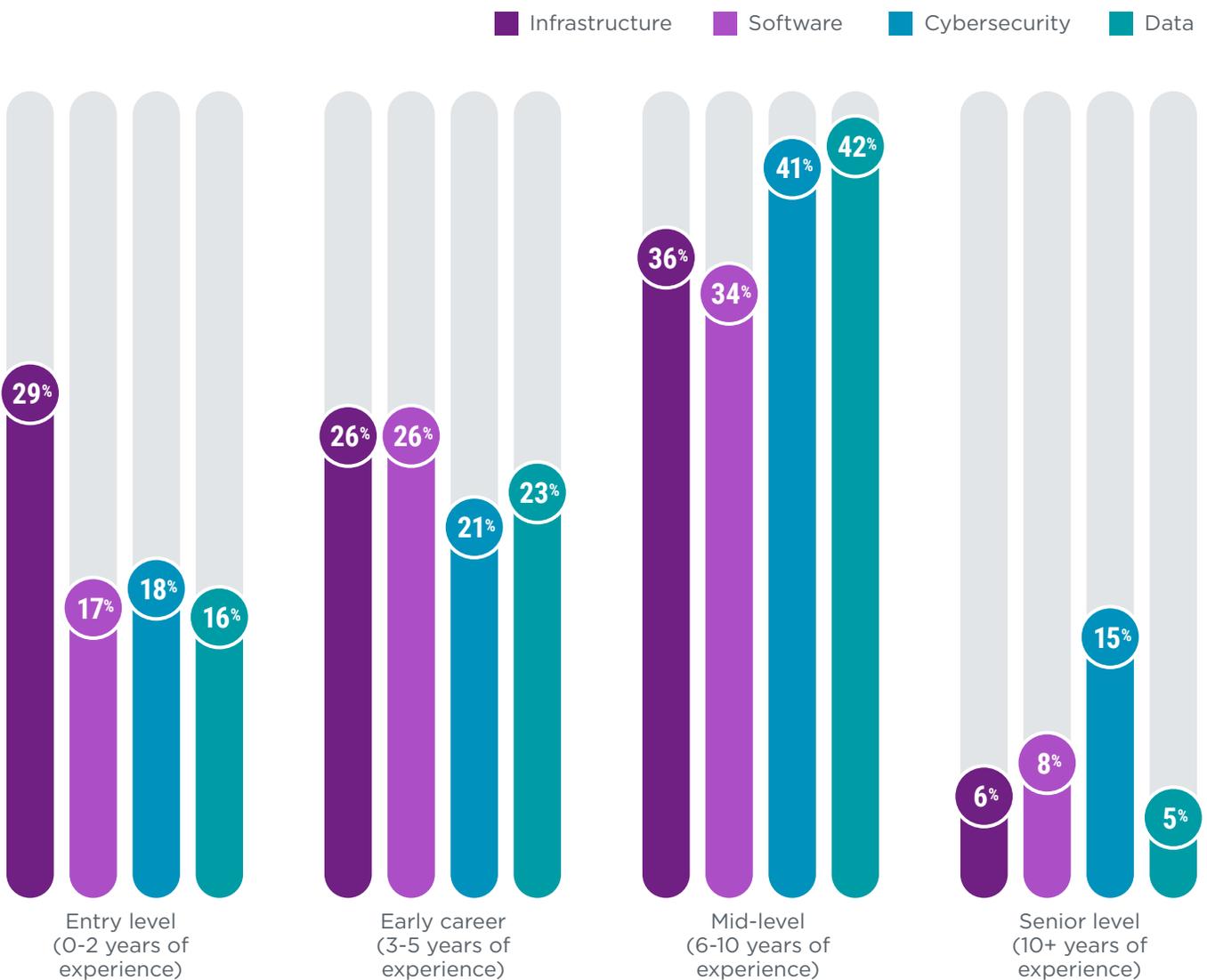
Data visualization is an especially interesting task. It combines technical knowledge with business savvy and communication skills. In a similar vein, data governance requires expertise in the regulations around digital data and the growing differences between regulations across state or national borders. In a rapidly changing digital economy, finding all of the unique skills required for data analysis is becoming a monumental challenge.

Data visualization is an especially interesting task. It combines technical knowledge with business savvy and communication skills.

Finding Data Specialists: Is There a Data Skills Shortage?

Given the relative novelty of the data function, one would expect that companies have a strong demand for entry-level positions. This is not exactly the case. In fact, companies with hiring plans in 2021 are looking for more mid-level data specialists than in any other field. Part of the explanation for this is that teams in the emerging fields of data and cybersecurity are often created from existing software and infrastructure teams. However, the high demand for specialized skills indicates a pipeline problem that has no easy solution.

Desired Level of Expertise When Hiring IT Pros



Without well-defined entry-level roles, companies have less ability to rely on historical means of skill development, where they are able to obtain a qualified candidate from a traditional pipeline like a four-year degree program and then give that candidate the job experience and training opportunities that build more advanced skills. The explosion in demand for the advanced skills exacerbates the problem.

Organizations are showing a willingness to consider candidates that do not come from a traditional four-year degree program. Interestingly, the willingness to consider data candidates without a four-year degree is higher than any other field besides IT support, and IT support as an entry-level position is already a field with different pathways such as bootcamps.

Tech Hiring Considerations for Non-4-Year-Degree Holders

	Unlikely to accept candidates w/o 4-yr-degree or unsure	Possibly would accept candidates w/o 4-yr degree	Likely to accept candidates w/o 4-yr degree	Net likely + possibly
IT support/help desk positions	17%	40%	43%	83%
Data analyst/database positions	21%	47%	33%	80%
Software developer/web developer positions	25%	42%	33%	75%
Cybersecurity positions	27%	40%	34%	73%
Network engineers	28%	41%	31%	73%
CIO/IT management	31%	37%	33%	69%

As companies move beyond the mindset of a traditional pipeline, they should be sure that any workforce development efforts address both short-term candidate discovery and long-term corporate success. Especially in rapidly evolving technology environments like the field of data, there are several important considerations for a skill development process that involves many different participants, including job seekers, HR professionals and hiring managers.

First, any process should include proof of expertise around a common set of technical areas. As companies gain a better understanding of their own requirements, it would be preferable to have a standard indicator of expertise, such as a four-year degree or an IT certification. These credentials often establish foundational knowledge, so candidates may still require training on proprietary or vendor-specific technology. In the long run, though, the ease of finding candidates with a strong foundation can shorten the overall time to having a productive employee.

Second, the development process should apply to a broad range of previous experience. In a field like data analytics, the ideal candidate may have a variety of skills, including technical skills such as data structures or analytic techniques and business skills such as operational practices or financial measurements. Training or certification can bridge the gap for a candidate that may have some of the prerequisite information but is missing certain domain expertise.

In the data field, time is of the essence. Companies have massive amounts of data—now the question is what to do with it?

Finally, any development process should be geared toward long-term success. Ideally, the business would find candidates who have higher levels of comfort with job responsibilities and advance more quickly in their careers. In the fast-paced environment of digital organizations, these employees help move the business forward by embracing their roles and finding opportunities for growth.

In the data field, time is of the essence. Companies have massive amounts of data—now the question is what to do with it? The ability to rapidly and effectively analyze data can improve time to market, enhance customer satisfaction and drive growth for the future. Treating data as a discrete function and building data specialists will accelerate this process. As organizations place a priority on speeding up data analysis, they should also place a priority on speeding up skill discovery and development.



CompTIA.org

Copyright © 2021 CompTIA, Inc.. All Rights Reserved.

CompTIA is responsible for all content and analysis. Any questions regarding the report should be directed to CompTIA Research and Market Intelligence staff at research@comptia.org.