

THE REAL WORLD OF THE DATABASE ADMINISTRATOR

By Dr. Elliot King, Research Analyst
Produced by Unisphere Research,
a Division of Information Today, Inc.
March 2015

Sponsored by

Quest

Produced by

 **UNISPHERE**
RESEARCH

TABLE OF CONTENTS

<i>Introduction and Key Findings.....</i>	3
<i>The Database Infrastructure</i>	4
<i>Data Under Management</i>	10
<i>DBAs' Responsibilities.....</i>	14
<i>The Future</i>	21
<i>Conclusion.....</i>	25
<i>Appendix A.....</i>	26

INTRODUCTION AND KEY FINDINGS

In the world of IT, perhaps nothing is more exciting than the introduction of radically new technologies. And the information management arena may be on the verge of one of those moments.

The past 10 years have been a golden age for generating and capturing a virtual avalanche of new kinds and quantities of data. With the growth of cloud computing and associated storage technologies, social media, low-cost video, and other hot new technologies, companies have access to an abundance of new data types.

These developments have led to new ways of thinking about data. Approaches like those represented by Hadoop and NoSQL seemingly point the way to a post-structured-data world. And concepts like data lakes paint a picture of assembling huge pools of highly varied data that can be accessed and analyzed on demand. The idea is to get to better data analysis, quicker and less expensively using much larger datasets to advance everything from fraud detection to predictive analysis.

The ability to capture, store, retrieve, analyze and save data in new ways holds the potential to gain new insights and new ways of doing business. If information is power, to paraphrase an old adage, more information can lead to more power. Great information can guide companies to greater business success.

But the attention to the new developments in information management should not obscure the fact that structured data in relational databases still provides the foundation for the information infrastructure in many, if not most, companies, and will do so in most organizations for the foreseeable future. The ways in which companies can exploit structured data is far from exhausted. And database administrators (DBAs) still represent the front-line personnel for data management in most enterprises.

To understand the current role of the DBA and the way that role is changing, Quest commissioned Unisphere Research, a division of Information Today, Inc., to survey database administrators and others charged with managing corporate data. The 300 respondents came from a wide range of companies in terms of size and industry. Nearly two-thirds of the respondents came from organizations with more than 1,000 employees and more than a dozen industries were represented. One-quarter of

respondents' organizations are running more than 500 databases. Details about the respondent pool can be found in the appendix.

Among the key findings are the following:

- While Hadoop and NoSQL are exciting new technologies, their use currently is confined primarily to large companies. The traditional database management system still provides the foundation for the information management infrastructure in most organizations. Oracle and Microsoft SQL Server are the most common platforms to support mission-critical data.
- Most enterprises believe that more familiar “new” technologies such as virtualization and cloud computing will have more impact on their organization over the next several years than “newer” emerging technologies such as Hadoop. In fact, Hadoop and NoSQL do not factor into many companies' plans over the next few years.
- Information infrastructures are not static. Most companies run multiple databases and are open to adding new database platforms if there is a need to do so. The most common motivating factor for adding a new database management system is the need to support new analytical applications.
- Most DBAs are responsible for multiple database instances from multiple vendors.
- Structured data remains the bedrock of the information infrastructure in most organizations.
- While maintenance and performance are the top responsibilities for most DBAs, security is becoming an increasingly important item on their agendas. However, currently, DBAs spend less time on security issues than they do on supporting database development.
- The key challenge for DBAs is learning new technologies.

THE DATABASE INFRASTRUCTURE

Although a lot of attention has rightly been lavished on emerging new information management technologies, mainstay database management systems still play the critical role in the information infrastructure. As Figure 1 shows, Microsoft SQL Server and Oracle are found in the overwhelming majority of organizations, with MySQL, IBM DB2, and MongoDB representing the next most popular database management systems.

The appearance of MongoDB on this list provides evidence for a growing acceptance of NoSQL technology, particularly in larger companies. Around 70% of all MongoDB users are running more than 100 databases and 30% are running more than 500 databases. And, nearly 60% of the MongoDB users are in companies with more than 5,000 employees.

While there is relative diversity in the use of database platforms—more than 20 different brands were mentioned by respondents—Oracle, Microsoft SQL Server, IBM DB2 LUW, and SAP Sybase ASE were the dominant RDBMSs running mission-critical data, with Oracle and Microsoft SQL Server clearly the leaders. Indeed, approximately 78% of the respondents indicated that they were running mission-critical data on Oracle and 72% said they were using Microsoft SQL Server as a platform for their mission-critical data. Almost one-quarter of the respondents said that they had turned to IBM DB2 LUW for mission-critical activities and almost 12% have opted for SAP Sybase ASE.

In contrast, roughly 6% of the respondents are using MongoDB, the most popular NoSQL database technology in this pool, for mission-critical data. As a whole, the survey results provide a good barometer for the current state of the adoption of NoSQL technology. NoSQL is beginning to penetrate larger companies but is not yet routinely a platform of mission-critical data. Moreover, as Figure 2 shows, many companies have not yet factored NoSQL into their plans in the next few years.

Along the same lines, although Hadoop has justifiably generated attention as a next-generation information management, its use is still largely confined to large companies. In fact, 60% of respondents who are currently using Hadoop are running more than 100 databases and 45% are running more than 500 databases. Moreover, approximately two-thirds of the respondents using Hadoop are employed by companies with more than 1,000 people.

Furthermore, as Figure 3 shows, Hadoop does not yet factor into most companies' plans. At this time, only 15% of the respondents are actually using Hadoop (and around another 5% are in the process of deploying it), and nearly 60% currently have no plans to use it in the future.

Clearly, traditional RDBMSs shoulder the lion's share of data management in most organizations. And since more than

85% of the respondents are running Microsoft SQL Server and about 80% use Oracle, the evidence is clear that most companies support two or more DBMS brands. As Figure 4 shows, by far, the most important reason companies use more than one database platform is to support different applications, with the need to support multiple user groups the second most common reason. Supporting increased data volume was only the fifth most important reason for adding a new DBMS platform.

Under the right circumstances, organizations are open to adding new DBMS platforms. The most common reason given for adding a new brand is the emergence of a new analytical use case. But other factors can also motivate a company to add a new DBMS as well, including the need to improve performance and flexibility as well as to better manage costs. (See Figure 5).

Several key insights are reflected in these results. First, most companies run multiple databases and are open to adding new database platforms if there is a perceived need to do so. The most common motivational factor to add a new database management system is to support new analytical applications. This scenario bodes well for Hadoop if it can deliver on its promise to make analysis more flexible and less costly. At this moment, however, the use of Hadoop and NoSQL technology is largely confined to bigger companies and many companies have no plans to explore these newer technologies at all.

The Role of the DBA

As the information infrastructure changes due to both the growth of data under management as well as the introduction of new technologies, including cloud-based computing, it could be expected that the role of the DBA, long on the front lines of data management, would change as well. In this survey, 75% of respondents' companies had between 1 and 25 people with the specific title of DBA, and 15% had more than 25 people with the title of DBA (although in many companies, people who have other titles will perform tasks associated with database administration). As might be expected, most DBAs are responsible for multiple database instances. As Figure 6 shows, almost half of the DBAs manage more than 25 database instances each, and almost 10% manage more than 100 database instances each.

Not surprisingly, the number of database instances for which each DBA is responsible is growing. (See Figure 7.) The number of databases for which each DBA is responsible is decreasing at only 5% of the respondents' companies.

Not only are DBAs responsible for managing multiple database instances, they manage database management systems from multiple vendors as well. As indicated in Figure 8, approximately 70% of the respondents said that the DBAs in their organizations



were responsible for managing databases from at least two vendors, and 7% of respondents said the DBAs at their companies were responsible for managing databases from five or more vendors.

The growth or potential growth of non-relational data management technology such as NoSQL and Hadoop raises an interesting question. Will DBAs be responsible for non-relational information management technology? Or, will a different job classification emerge to claim stewardship of those technologies? The answer to that question is not yet clear but the indication is that over the long term, DBAs will be responsible for non-relational information management platforms as well. Figure 9 shows that roughly two-thirds of the respondents that had deployed either Hadoop or NoSQL said that DBAs were responsible for managing those technologies.

Drilling down, in 62% of the organizations that have already deployed a NoSQL platform—but not Hadoop—

DBAs are responsible for managing the NoSQL technology. In organizations that have deployed Hadoop—but not NoSQL—only 52% of the organizations have DBAs managing Hadoop. And, in companies that have both Hadoop and NoSQL installed, DBAs are responsible for managing the non-relational technologies 72% of the time.

Clearly, the long-term trends that DBAs have faced are still in place. DBAs are the front-line data management administrators. Each DBA is responsible for multiple database instances from multiple vendors and the number of databases each DBA is expected to manage is increasing. But, there is a twist. Over the next few years, as Hadoop and NoSQL become more common in the enterprise, DBAs are also likely to be the ones responsible for the administration of non-relational data infrastructure technologies.

Figure 1: What database brands do you have running in your organization?

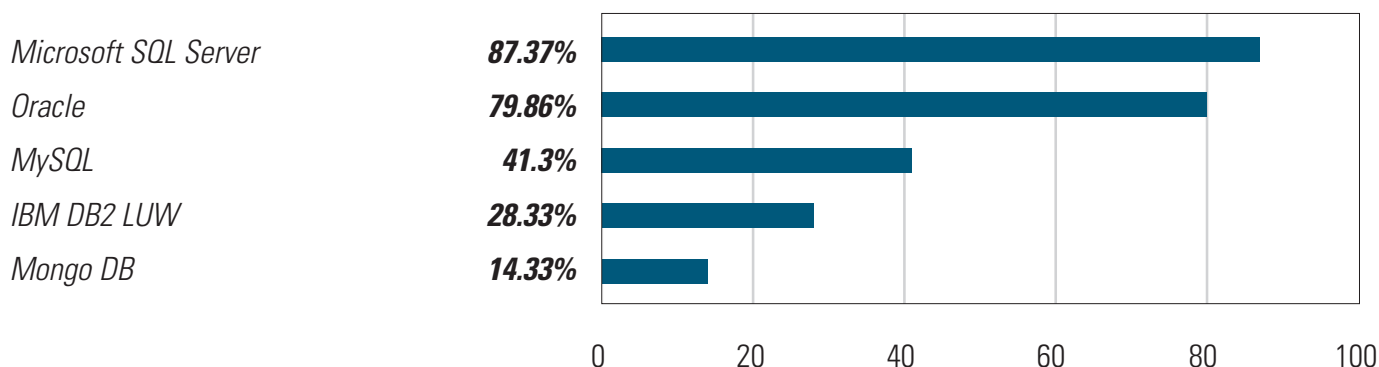


Figure 2: Does your company currently have plans to adopt NoSQL database technology?

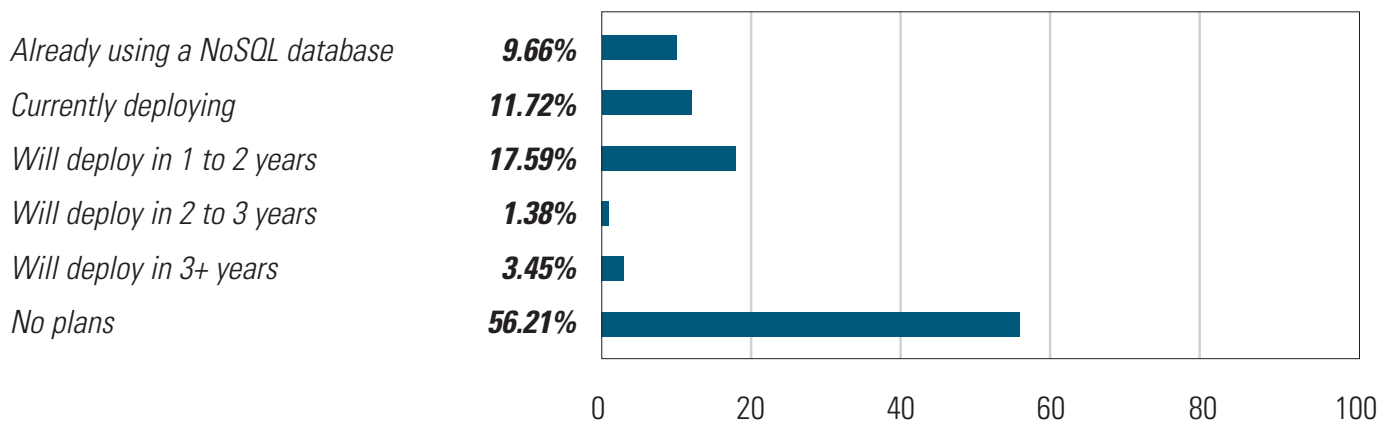


Figure 3: Does your company use or have plans to use Hadoop?

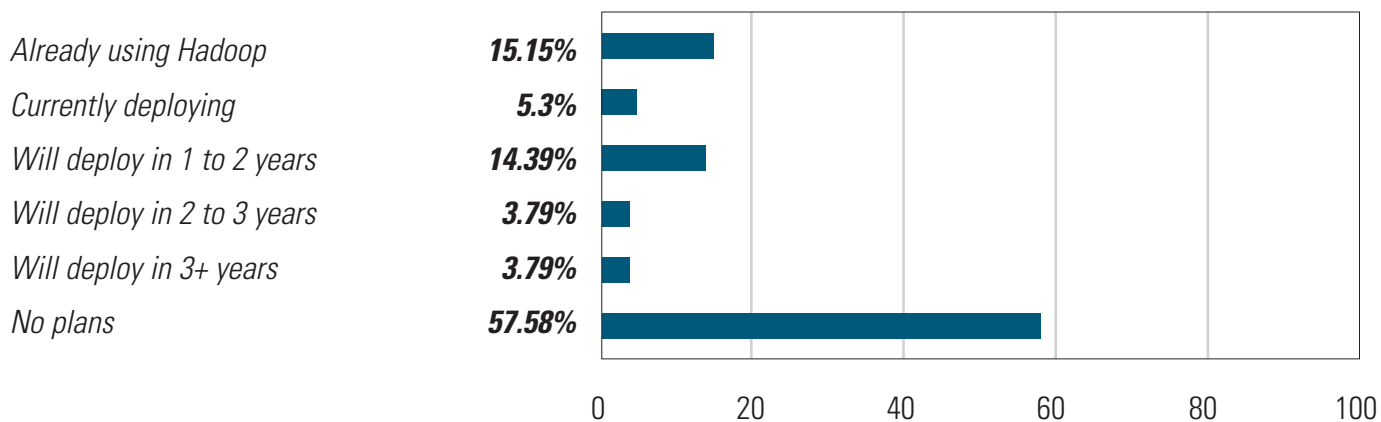


Figure 4: What is the most important reason for using multiple database platforms in your organization?

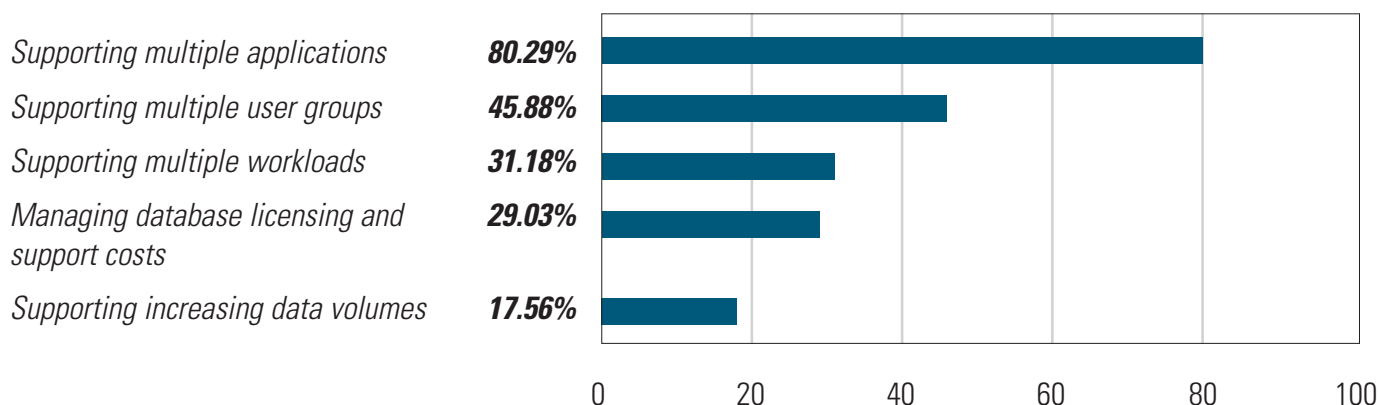


Figure 5: What are the most important reasons for adopting a new database management system in your organization?

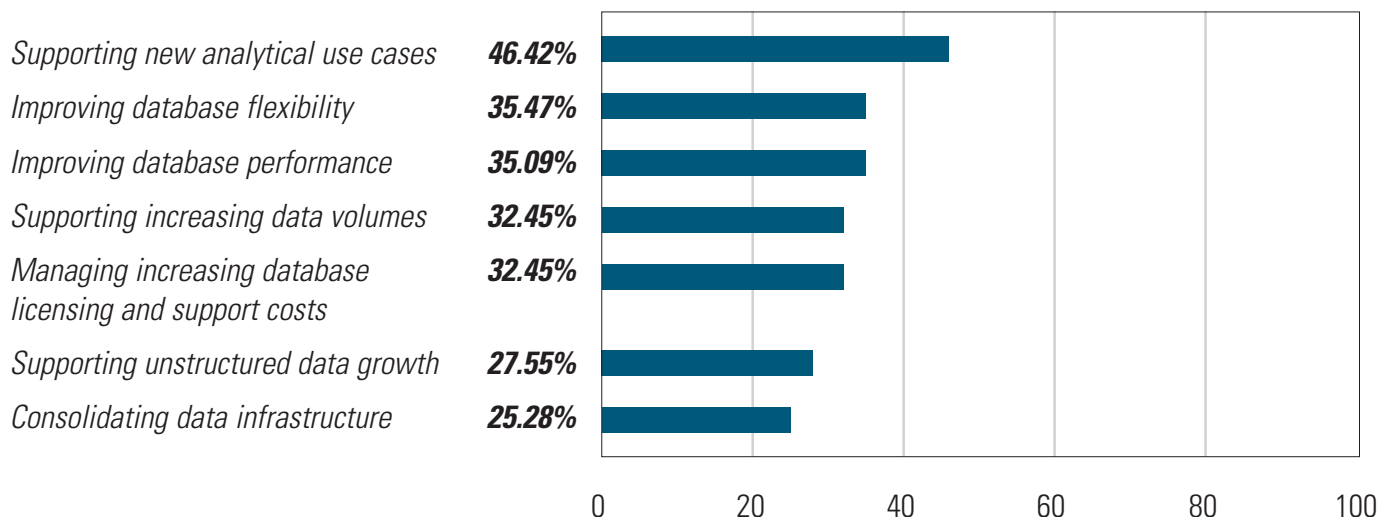


Figure 6: Approximately how many database instances does each DBA manage?

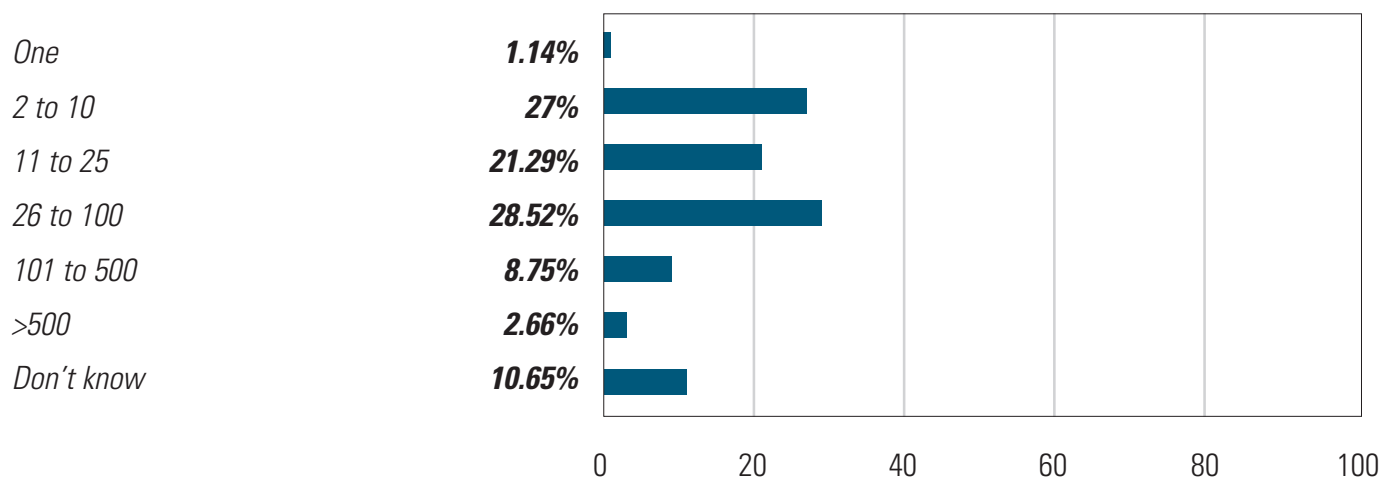


Figure 7: Is the number of databases for which each DBA is responsible increasing, decreasing, or staying the same?

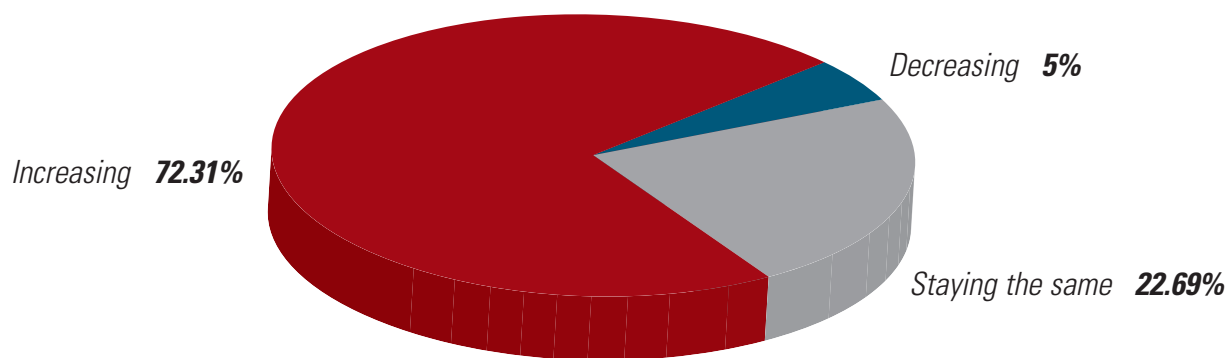


Figure 8: How many database platforms (i.e., platforms from different vendors) is each DBA responsible for managing in your organization?

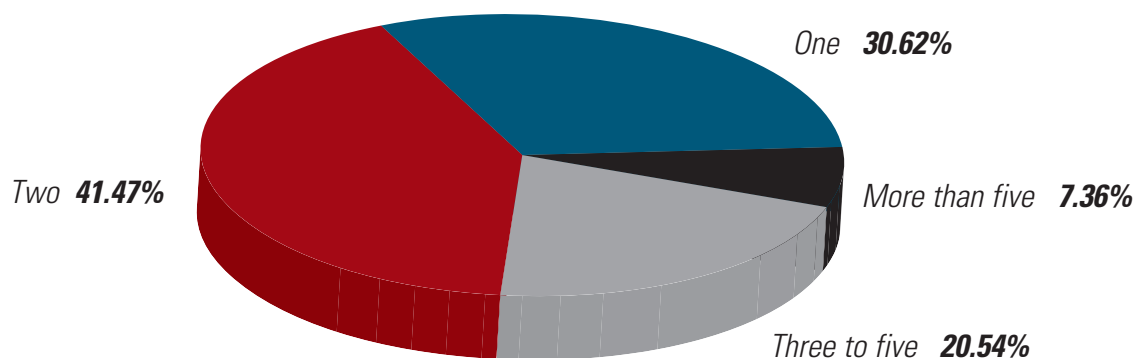
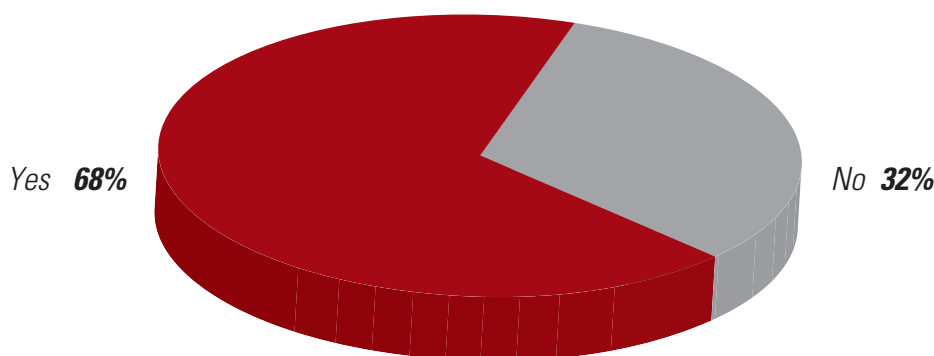


Figure 9: Are the DBAs that are responsible for managing relational database management systems also responsible for managing non-relational systems (such as NoSQL and Hadoop)?

(Only respondents who had deployed Hadoop or NoSQL technology)



DATA UNDER MANAGEMENT

What seems to be explosive data growth has been a constant feature of information technology and management for a long time. The sheer volume of data under management seems staggering and yet what was once considered a lot of data is now seen as trivial. The respondents in this survey manage a high volume of data, with close to 30% managing more than 500TB.

Over the past 5 years, however, the raw growth of data has been only part of the story. As importantly, more data types are being captured, stored, and made available for analysis. Moreover, companies have access to more data that they themselves did not create. Those two drivers—new data types and new data sources—have led to the interest in what is called big data.

For all the interest in how to capture and manage unstructured and semi-structured data, structured data remains the bedrock of the information infrastructure in most companies. As Figure 10 shows, two-thirds of the respondents indicated that structured data represented at least 75% of the data under management. Nearly one-third said that their organizations do not actively manage unstructured data at all.

Interestingly, the reported rate of growth for the total amount of data under management is substantially less than is commonly reported. Slightly less than half of the respondents reported that the total amount of data under management—both structured and unstructured—has been growing at a rate of 25% or less annually. Looked at another way, however, around two-thirds of the respondents said their overall data is growing between 10% and 50% annually. (See Figure 11.)

The growth rate of structured data generally parallels the growth rate of data overall. As Figure 12 shows, about two-thirds

of the respondents indicate that their structured data is growing between 10% and 50% annually.

Since some organizations are not yet managing much unstructured data, the rate of growth for unstructured data overall as reported in this survey is not as great as might be expected. Figure 13 shows that less than 12% of the respondents believe that the growth rate of unstructured data under management in their organization is more than 50% annually. But this chart also reveals another important point. About one-third of the respondents simply do not know how fast the amount of unstructured data under management is growing. This can be taken as evidence that in many organizations, unstructured data may still be outside normal data management processes.

Figures 14 and 15 show the sources for data growth. Figure 14 asked about the sources of structured data growth and Figure 15 posed the same question for unstructured growth. Both returned interesting results. The primary driver for the growth in structured data is transactional data, including e-commerce data, followed by data from ERP systems and then financial data. Structured data from the internet outside of e-commerce, such as clickstream data, represented a significantly less important source of structured data growth.

As for unstructured data, despite all the attention that has been lavished on social media as a potential new data source for analytics, the most important driver for the growth of unstructured data is internally generated documents, followed by email. Clearly, when it comes to unstructured data, the challenges of managing internally generated data have a top priority.

Figure 10: What percentage of this data is structured (i.e., stored in traditional relational database management systems)?

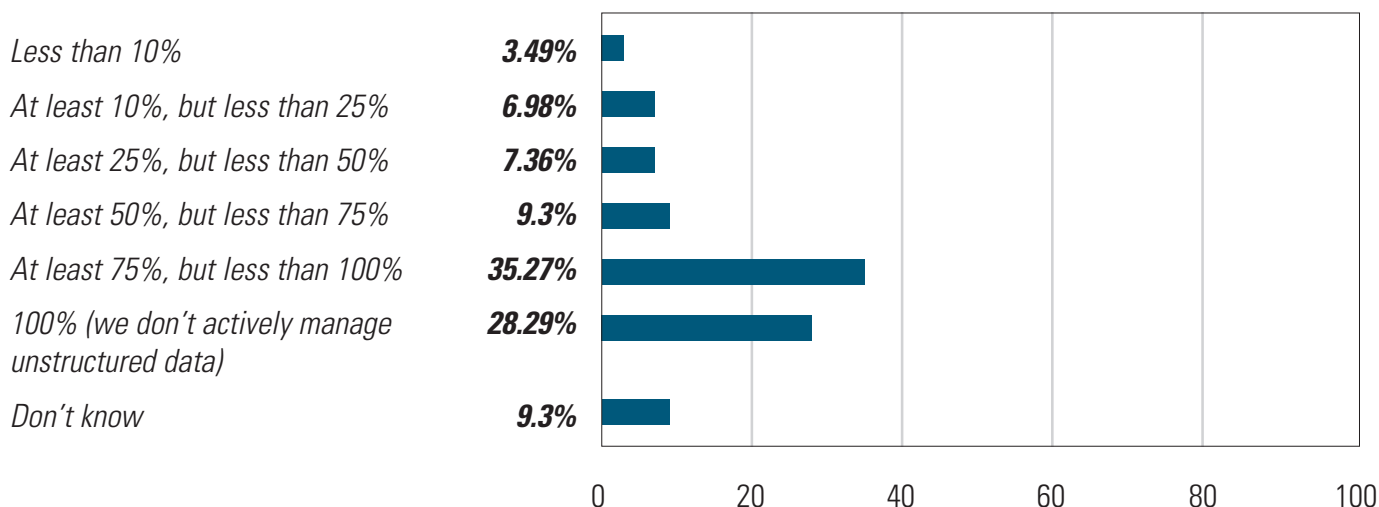


Figure 11: How fast is the total amount of data under management at your organization growing annually (both structured and unstructured)?

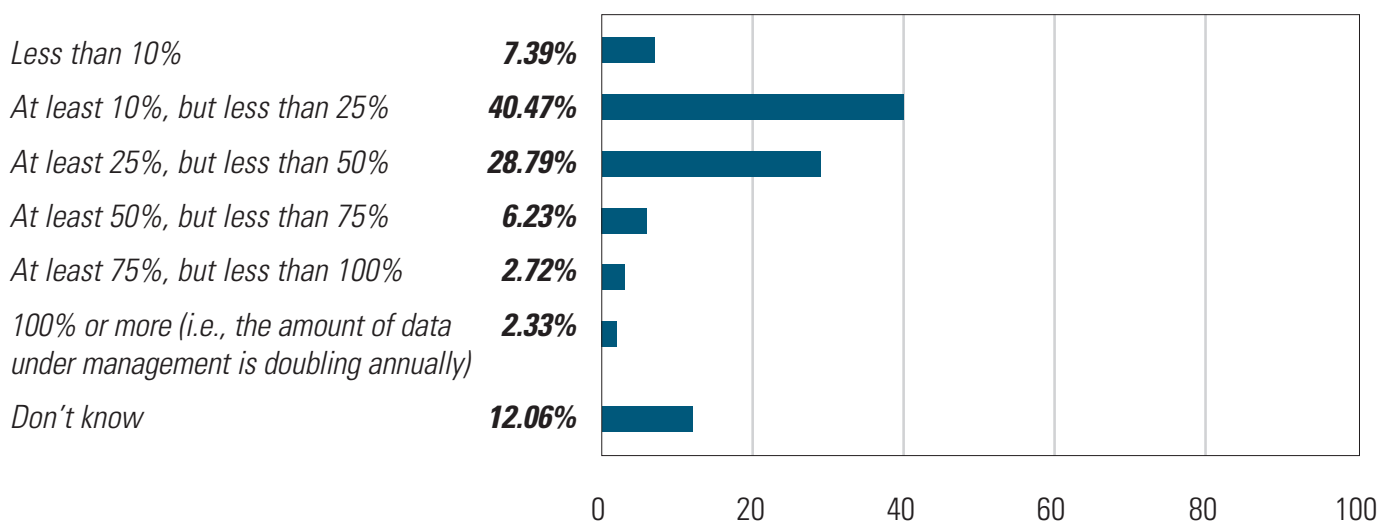


Figure 12: How fast is the amount of structured data growing annually?

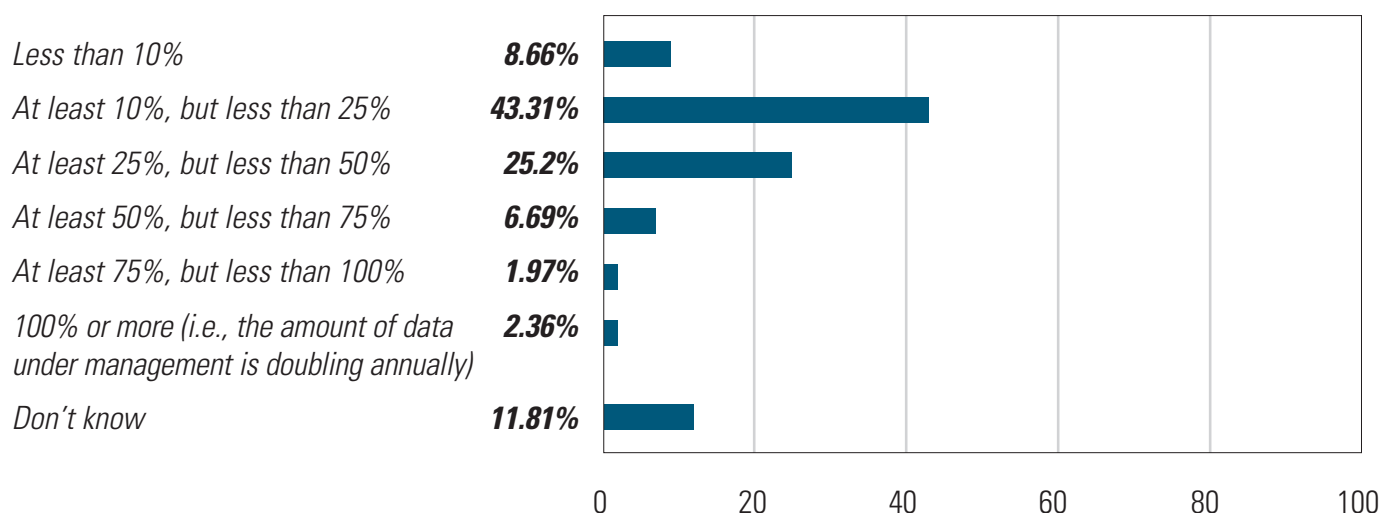


Figure 13: How fast is the amount of unstructured data growing annually?

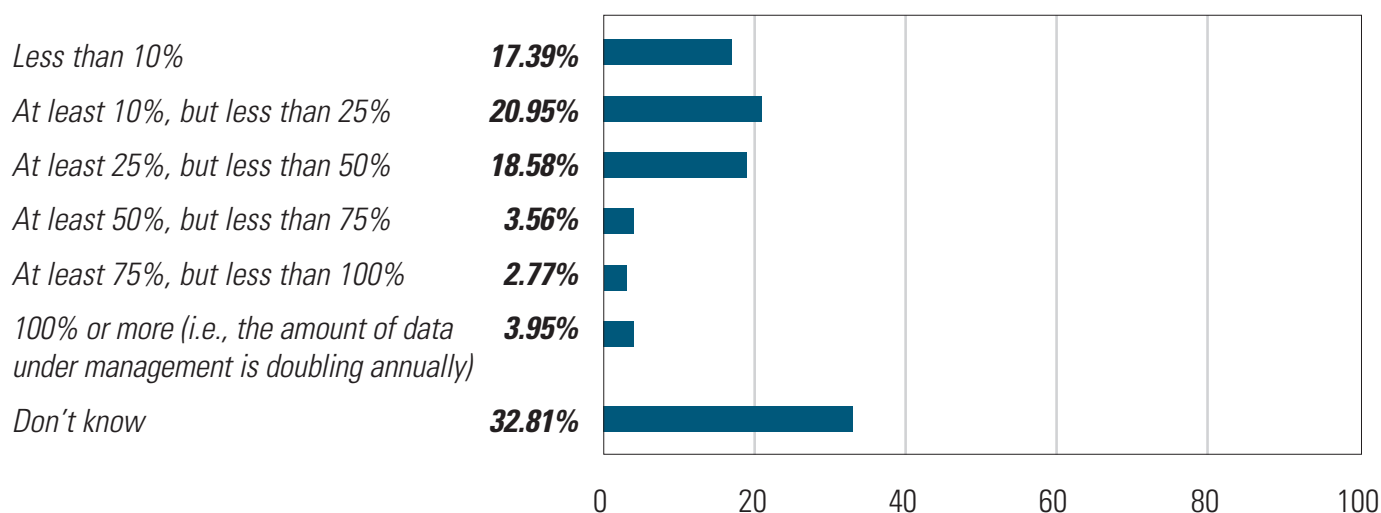


Figure 14: What are the most important sources of *structured* data growth at your organization?

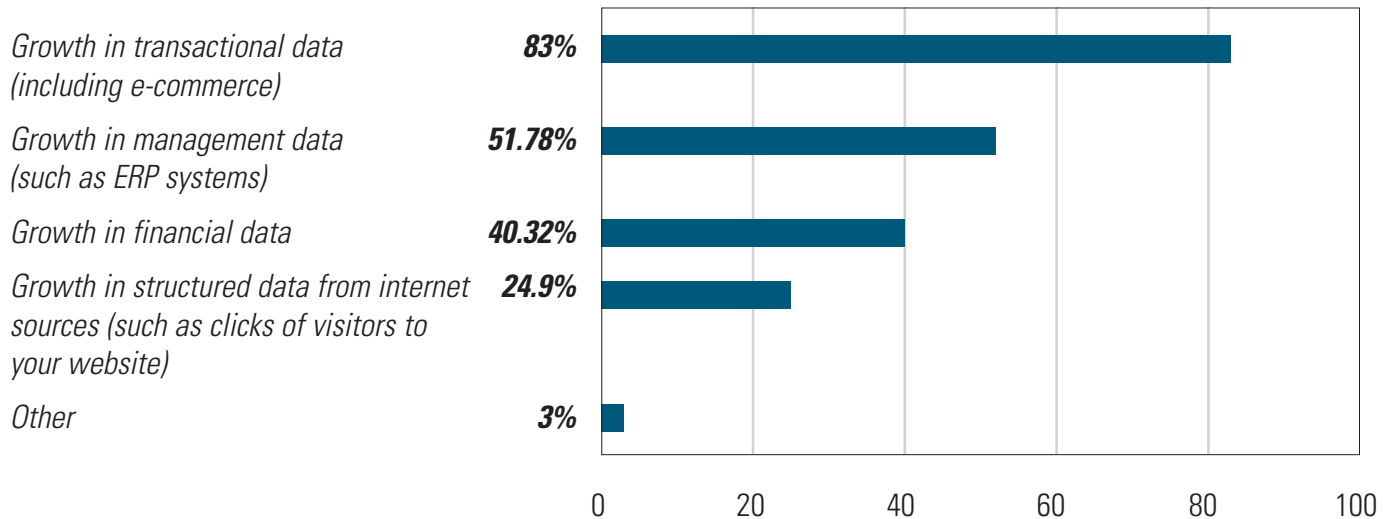
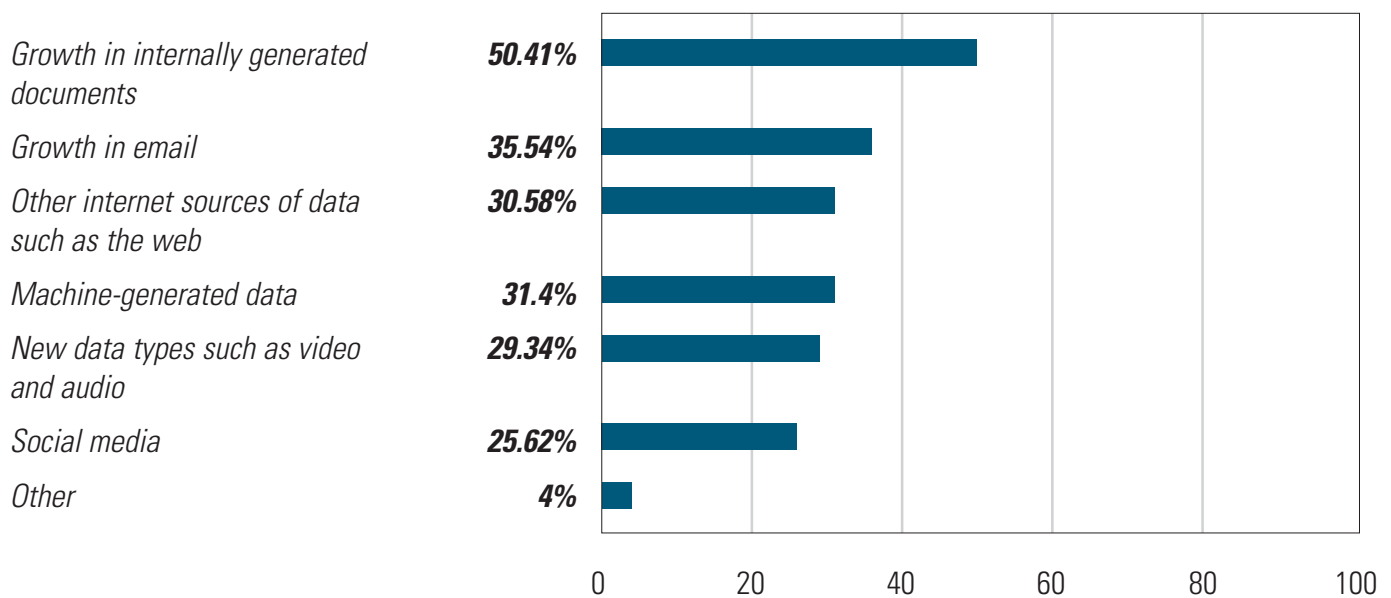


Figure 15: What are the most important sources of *unstructured* data growth at your organization?



DBAS' RESPONSIBILITIES

Given the importance of data and databases to the overall functioning and success of every organization, effective database administration is critical. DBAs perform many tasks, and the demands on their time have changed as database technology has evolved and new capabilities have been introduced. Some tasks that were once performed manually are now automated. And many database administration tools have been introduced that help DBAs perform their essential tasks.

Figure 16 shows the tasks DBAs perform that respondents consider the most important. The results are not surprising. The DBAs' top responsibilities are maintenance, performance, and configuration. In other words, DBAs must make sure that the databases are up and running, performing well, and are configured to function appropriately. Beyond that, DBAs must be concerned with a wide range of tasks ranging from capacity planning to security.

The way DBAs spend their work days largely aligns with their perception of their most important responsibilities. (See Figure 17.) Interestingly, however, DBAs may spend less time on security issues than they do on supporting development even though security is generally a higher priority. They also seem to spend less time on data integration and capacity planning than might be expected.

Given that DBAs view systems uptime and performance as their two most important responsibilities, it is appropriate that those are the two metrics by which they measure themselves. Figure 18 shows the top five criteria by which DBAs evaluate their own performance.

Although the way DBAs judge their performance is consistent with what they perceive as their top priorities, they believe that their managers evaluate them differently. As Figure 19 reveals, survey respondents felt that the speed of resolving problems was a more important issue for managers in evaluating DBA performance than system performance, and system performance was slightly more valued than system uptime.

Of course, the way DBAs spend their time changes as new responsibilities are added, as new technology is incorporated

into the infrastructure both at the data management and data administration levels, and as staffing changes and organizations generally evolve. Figure 20 presents some of the granular tasks DBAs perform and assesses which of those tasks now require more time and which require less.

For most DBAs, the amount of time spent on specific tasks is not changing much. But there are some notable exceptions. More than 60% of the respondents indicated that they are spending more time on performance tuning and more than 53% said cloning or provisioning databases for test and development required more attention. Very few tasks seem to have required less attention. Indeed, the only tasks tabbed by more than 20% as requiring more attention were verifying that all scheduled jobs had run (22.08%), and verifying all instances are up (20.6%).

One approach to making a DBAs' job more efficient is to automate routine tasks. There are several different approaches to automating tasks, including using third-party tools or writing custom scripts. Figure 21 provides insight into which DBA tasks are automated.

Clearly, at most organizations, many fundamental administrative tasks are automated. More than 80% of the respondents say that verifying that all the database instances are up has been automated. And more than 70% report that verifying all jobs have run, backups have been successful, and there is adequate disk space, is currently automated as well. The tasks that are most commonly performed manually are performance tuning (78.85%) and provisioning or cloning databases (67.98%). That finding is consistent with the earlier finding that those tasks require an increasing amount of DBAs' attention.

Finally, as more tasks are automated DBAs are sometimes called on to do database development tasks as well. In Figure 22, respondents indicated which development tasks they thought required the most time, regardless of whether development was done by a DBA or a database developer.

Even as new non-SQL-based technologies begin to gain traction, SQL is still very much the primary focus for most DBAs and database developers.

Figure 16: What do you consider the most important responsibilities for DBAs?

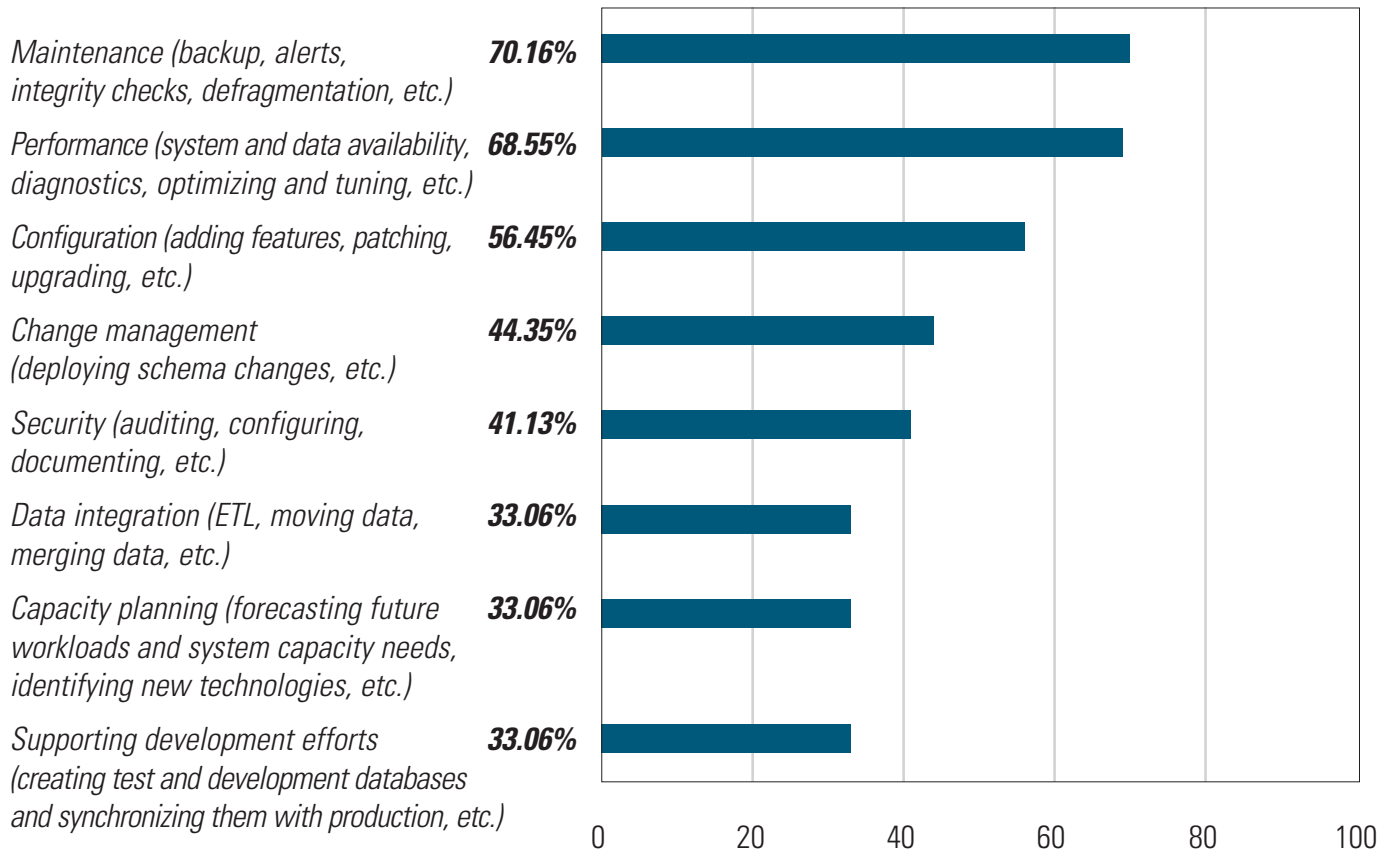


Figure 17: To which of these responsibilities do you think DBAs typically devote the most time?

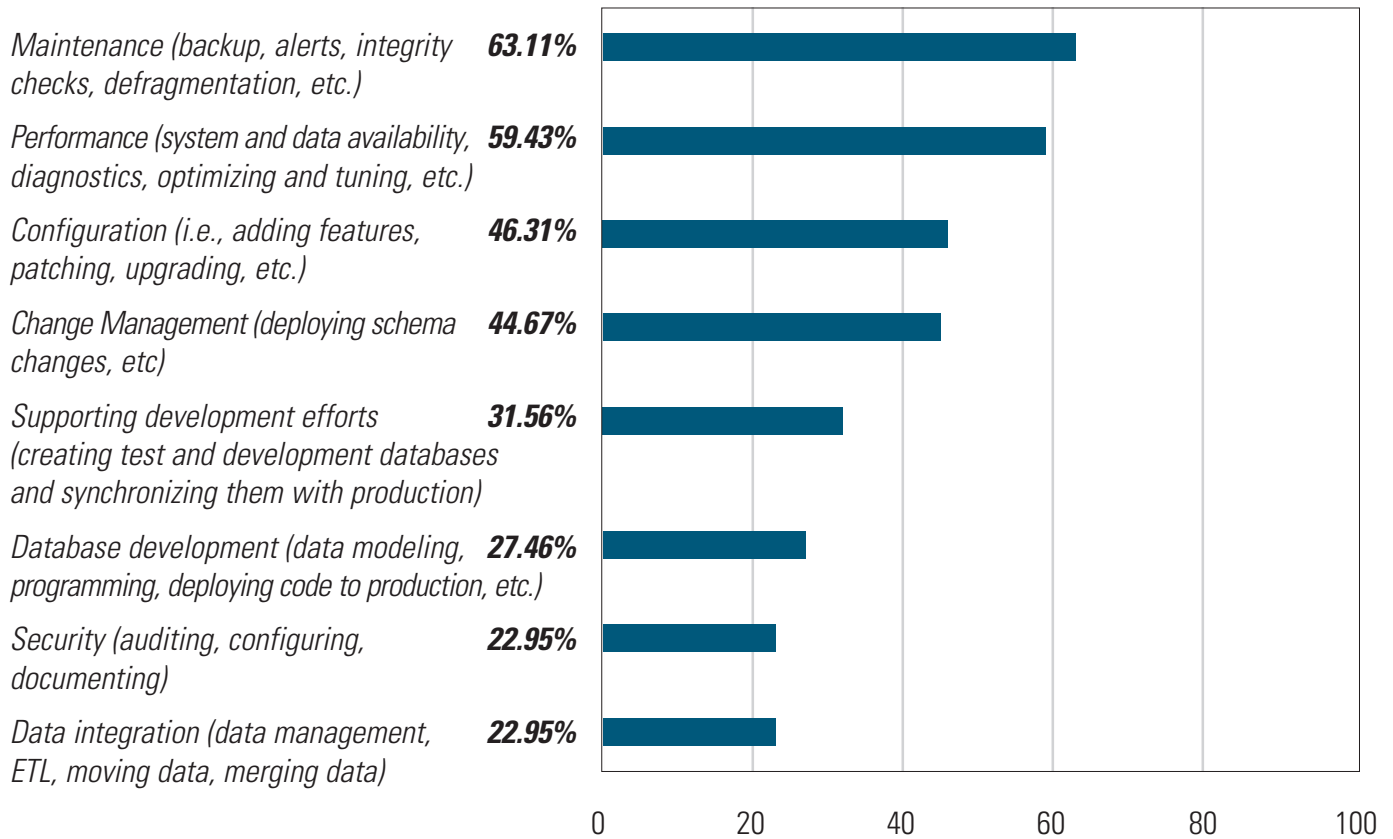


Figure 18: What are the most important metrics by which *you* measure your performance?

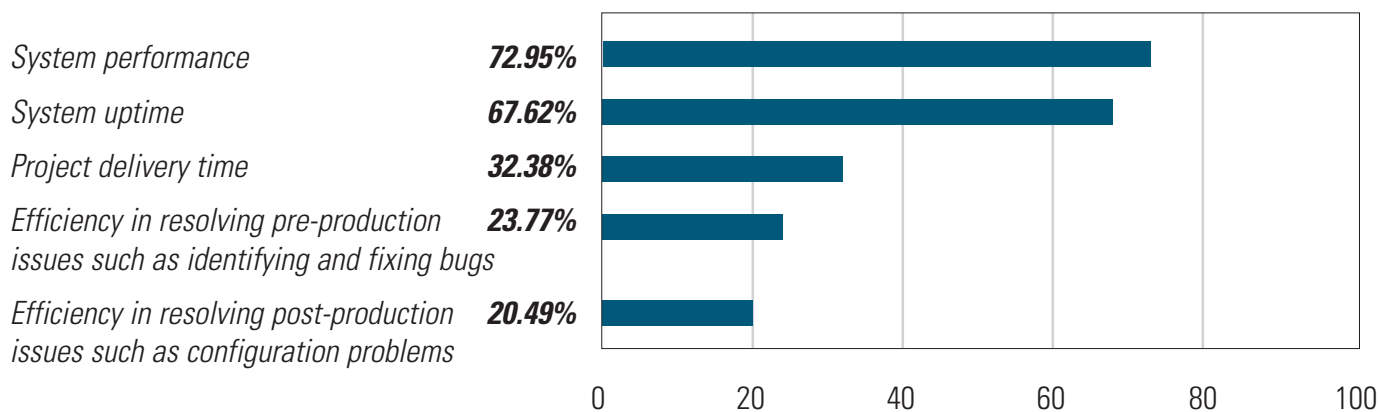


Figure 19: What are the most important metrics by which *your manager* measures your performance?

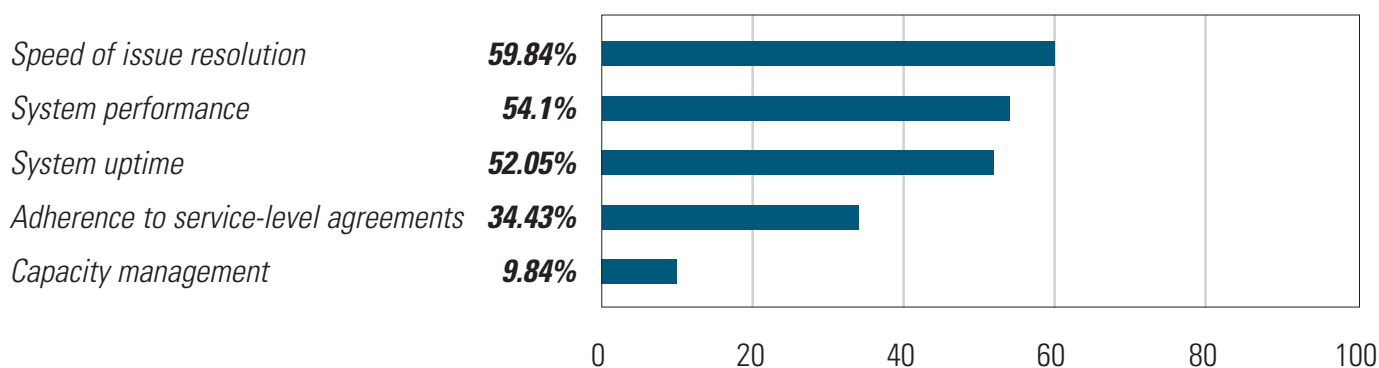


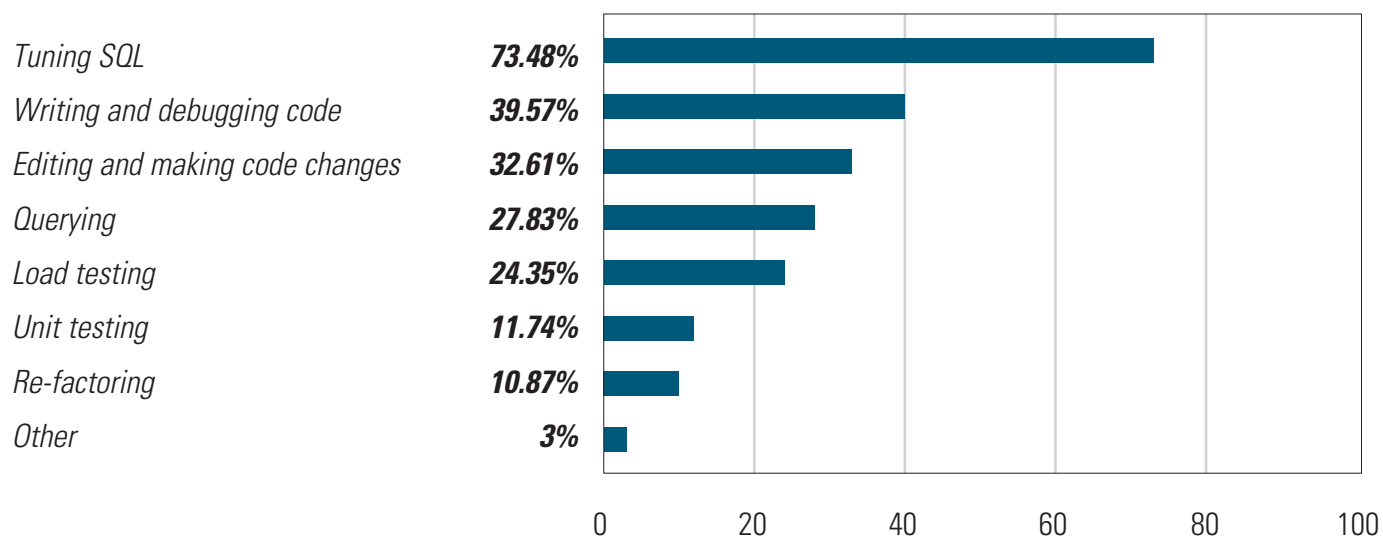
Figure 20: Among these responsibilities commonly assigned to DBAs, is the amount of time you devote to each responsibility increasing, decreasing, or staying the same?

	<i>Increasing</i>	<i>Decreasing</i>	<i>Staying the same</i>
<i>Verify all instances are up</i>	24.89%	20.6%	54.51%
<i>Inspect error logs for unusual events</i>	29%	17.32%	53.68%
<i>Verify that all scheduled jobs have run successfully</i>	16.88%	22.08%	61.04%
<i>Verify success of database backups</i>	23.38%	19.48%	57.14%
<i>Monitor disk space</i>	30.3%	19.48%	50.22%
<i>Review database size and growth settings</i>	37.83%	12.61%	49.57%
<i>Cloning or provision databases for test and development</i>	53.68%	10.39%	35.93%
<i>Verify that replicated databases are synchronized</i>	26.09%	13.91%	60%
<i>Performance tuning for database workload (rewriting SQL statements, adding indexes, etc.)</i>	61.21%	4.31%	34.48%

Figure 21: Of the tasks that many DBAs perform, which are automated, performed manually or not generally performed?

	<i>Automated</i>	<i>Manual</i>	<i>Not generally performed</i>
Verify all instances are up	81.14%	15.35%	3.51%
Inspect error logs for unusual events	46.49%	45.61%	7.89%
Verify that all scheduled jobs have run successfully	70.93%	23.35%	15.73%
Verify success of database backups	71.18%	25.33%	3.49%
Monitor disk space	72.49%	24.45%	3.06%
Review database size and growth settings	40.79%	50.88%	8.33%
Cloning or provision databases for test and development	23.68%	67.98%	8.33%
Verify that replicated databases are synchronized	42.92%	36.28%	20.8%
Performance tuning for database workload (rewriting SQL statements, adding indexes, etc.)	15.42%	78.85%	5.73%

Figure 22: Regardless of who is specifically responsible for database development, which of the following development tasks do you feel occupies the most time?



THE FUTURE

The relational database has been at the foundation of large-scale information management for several decades and should continue to be so for the foreseeable future. That said, interesting new information management, storage, and analytical technologies are making their way into many enterprises and DBAs will play a role in their implementation and use. At the same time, as companies use more data types from more sources, both complexity and risk are escalating. Both of those developments will present challenges to DBAs. Figure 23 reveals what respondents believe will be the most significant challenges facing DBAs over the next 3 years in terms of the developments in data itself.

Managing data growth and improving security are on most people's radars as areas that need to be addressed in the future. Beyond that, there is widespread disagreement about where DBAs need to focus over the next several years. This variety of opinions can be seen as evidence that we are heading into a period of change and the demands those changes will place on DBAs is not yet clear.

The same wide array of opinions exists when respondents are asked to anticipate the key changes in the data management infrastructure over the next several years. As Figure 24 suggests, no single technology or change has truly come into sharp focus. No single technology or changes was tabbed by at least 50% of the respondents.

Interestingly, in terms of new technologies, DBAs do not believe that Hadoop and NoSQL are going to have the biggest

impact on their organizations over the next several years. They are more focused on the continued growth of cloud computing and virtualization. (See Figure 25.)

Obviously, changes in the data and data management infrastructure will put pressure on DBAs and their organizational teams. As Figure 26 shows, DBAs realize that they will have to learn new technologies and they may have to do so in an era of smaller budgets.

Change, of course, represents both opportunities and threats. DBAs will be critical for enterprises to be able to maximize the benefits of the opportunities and minimize the threats.

Over the next year, however, the most pressing items on the agenda of DBAs will be more evolutionary than revolutionary. When asked to identify their most important project for next year, the most common answer was upgrading their database systems or enterprise applications, followed by consolidating their information infrastructure. That consolidation could come as the result of an upgrade or move to the cloud, or it could be independent of other projects. Cloud technology is on the radars of many DBAs. Many would like to automate more tasks associated with database administration. No single project or project area emerged as a consensus issue from this survey. However, one of the most challenging aspects of database administration is that each setting is unique.

Figure 23: In terms of data itself, over the next 3 years, what do you anticipate will be the top challenges facing DBAs?

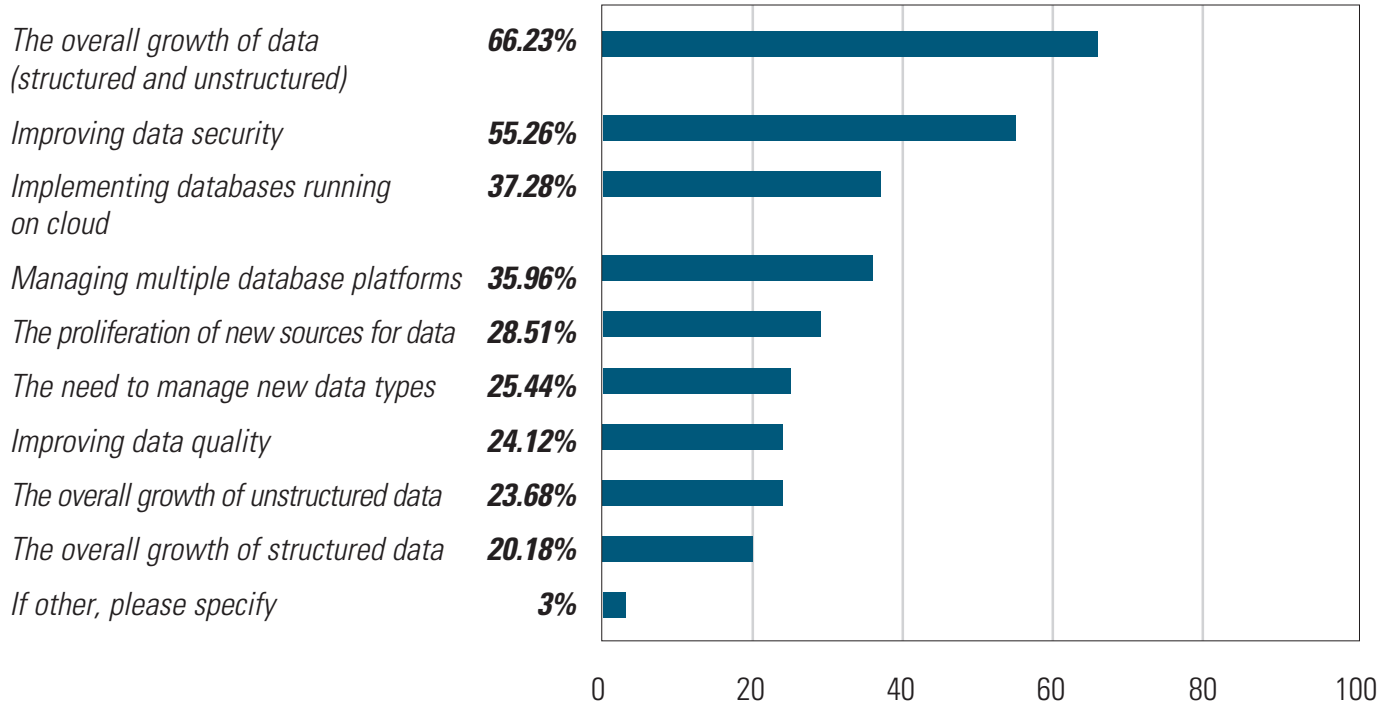


Figure 24: In terms of the data management infrastructure, what do you think will be the top challenges facing DBAs over the next few years?

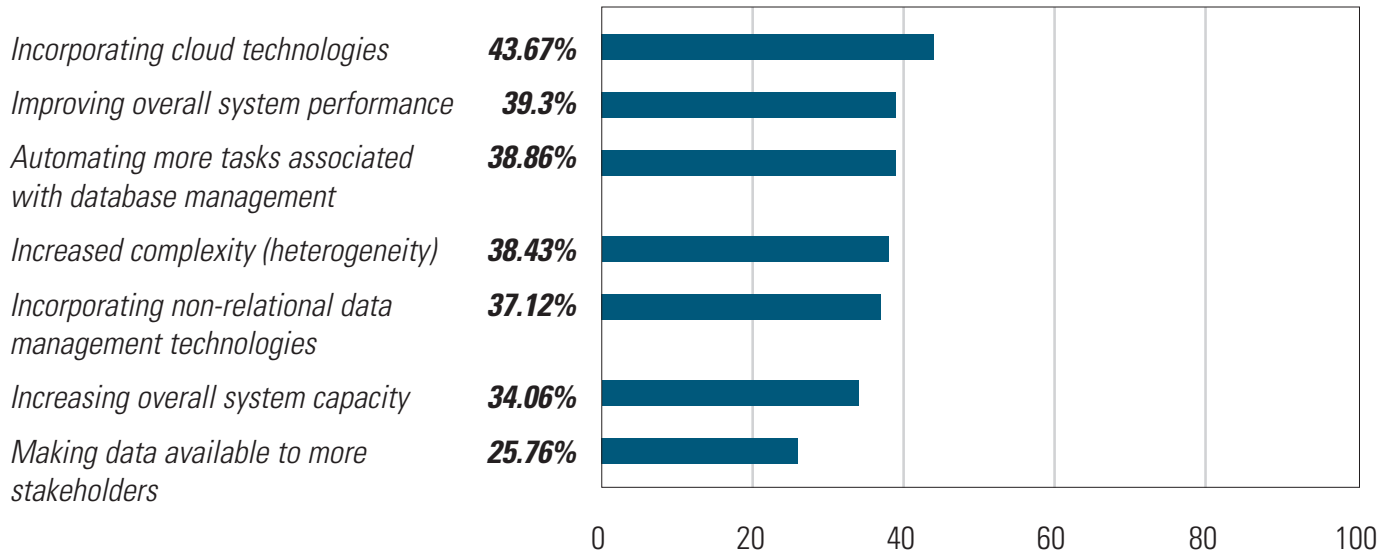


Figure 25: What technology trends do you think will have the most impact on database administration over the next 3 years?

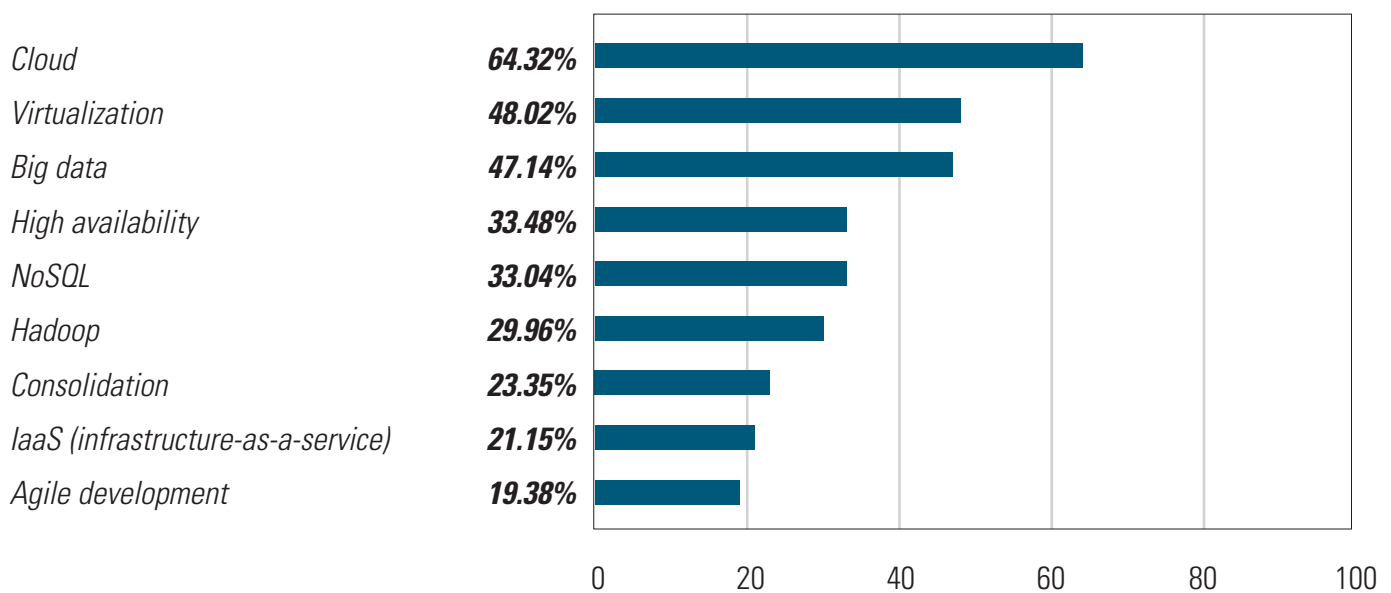
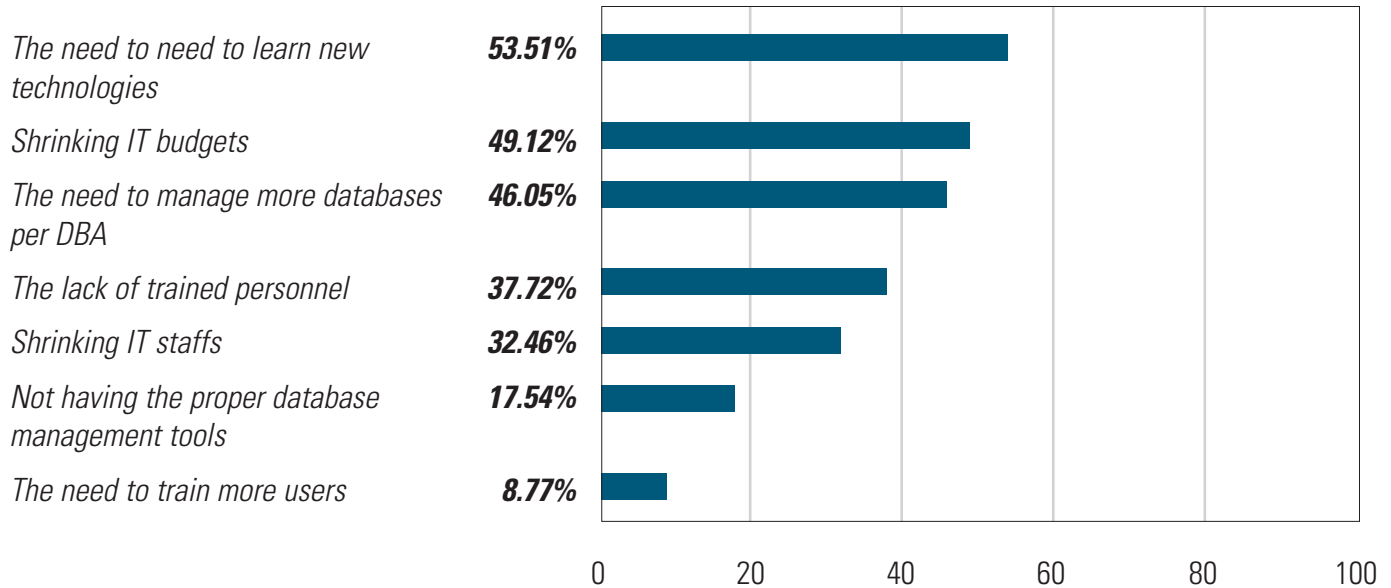


Figure 26: In terms of database administration itself, over the next 3 years, what do you think will be the top challenges facing DBAs?



CONCLUSION

While exciting new information management technologies such as Hadoop and NoSQL are beginning to make inroads in large companies, by and large, traditional database management systems still represent the critical informational foundation in most companies. DBAs are the front-line personnel maintaining that foundation, responsible for ensuring that performance meets the needs of the organization. With that in mind, several long-term trends are still in place. DBAs are being asked to manage more database instances and DBMSs from different platforms.

The amount of data under management continues to grow at a crisp pace. And while some routine administrative tasks have been automated, performance tuning, and provisioning are still often done manually.

Perhaps as importantly, as the use of new technologies grows, DBAs will be the ones responsible for integrating those opportunities into their organizations. To that end, they have to continue to learn new technologies and continue to find ways to do more with less.

APPENDIX A

The following figures show the characteristics of the pool of respondents.

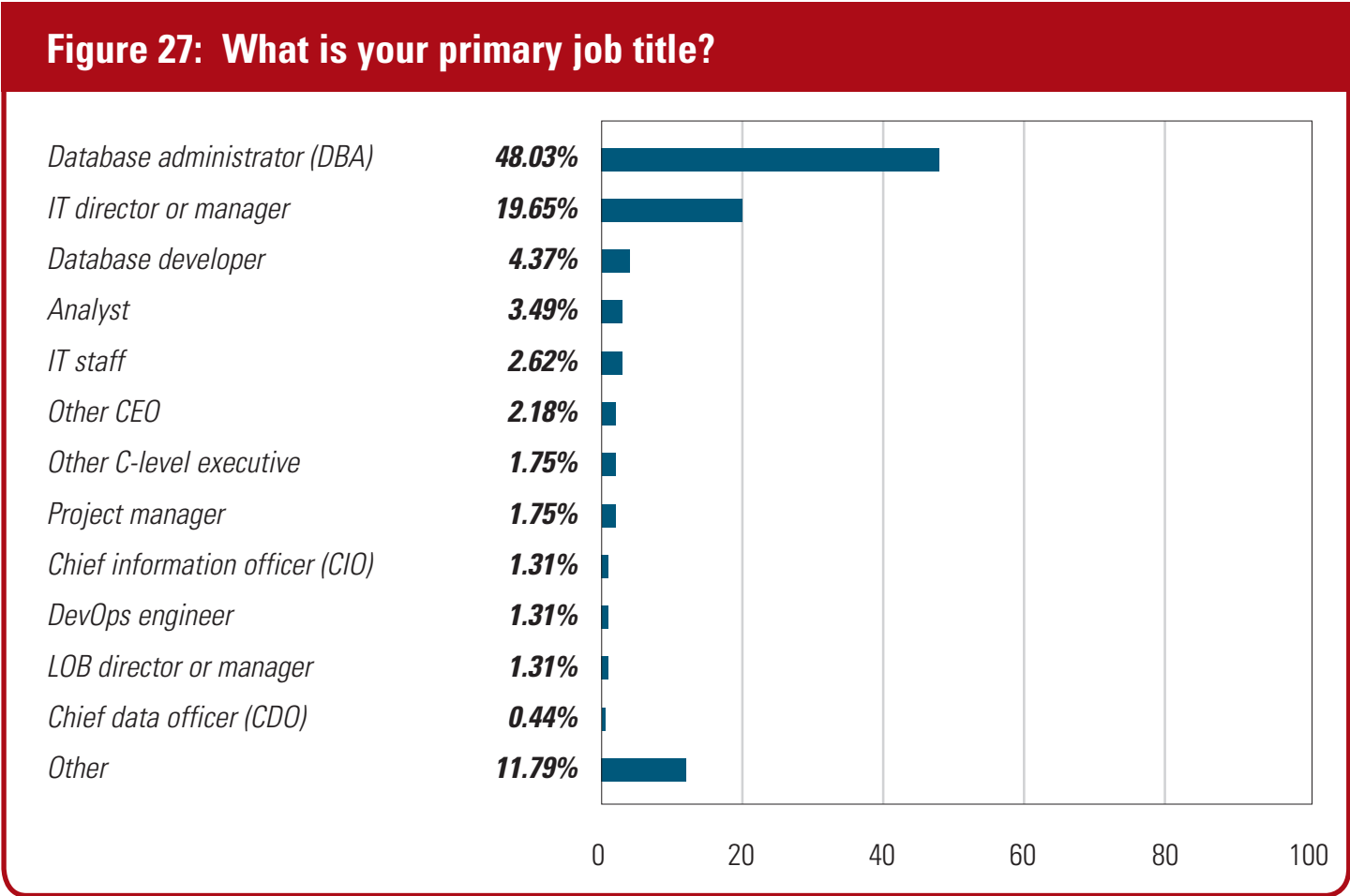


Figure 28: What is your organization's primary industry classification?

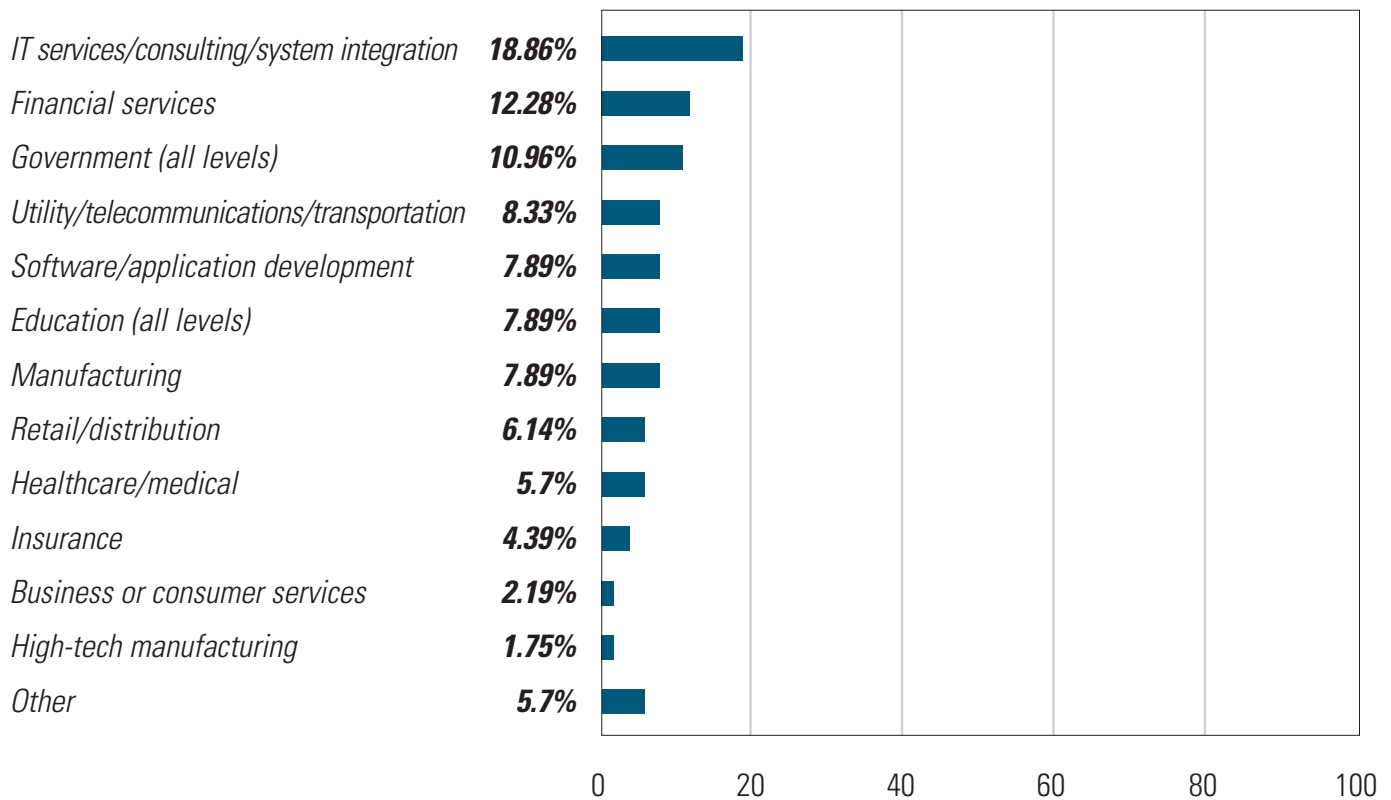


Figure 29: Approximately how many employees (including all locations, branches, and subsidiaries) are in your organization?

