Progress®

The 2020 Data Connectivity Survey Report

by David Loshin, Knowledge Integrity, Inc.

Table of Contents

Progress®

Executive Summary

As in prior years, the 2020 Progress Data Connectivity report is based on solicited feedback from information management professionals about the state of their enterprise information management landscape. The 2019 survey, which collected responses from over 1,800 respondents worldwide across more than 13 distinct industries, was put under the microscope to drill even further down into that data.

Our goal? To seek out patterns and insights that can guide your ongoing data management strategies, whether exploring trends across regions and industries or other factors such as laws, regulations and industry standards.

Here are some key findings from this year's survey:

- Moving to the cloud: As an integral part of everyday business operations, the cloud is here to stay. Organizations continue to gain comfort with the adoption of cloud-based technologies, with the use of cloud platforms, tools and services continuing to increase. Not only that, but respondents noted cloud migration as the top modernization project, and there is even more significant prominence for this in several individual industries.

- Object storage is key: While cloud-based Relational Databases such as Microsoft SQL Azure, Amazon Redshift and Amazon Aurora continue to increase in popularity, Amazon's S3 object storage solution remains the most popular of all the big data platforms.

- Continuing need for connectivity: The need for secure, reliable and scalable data connectivity solution gateways continues to increase, with a special focus on the challenge of integrating cloud data with on-premises data.

- Data quality is king: Of the challenges presented to respondents in past surveys, data quality demonstrated the most dramatic increase in year-over-year selection. Clearly, increased demand for data and analytics has exposed the pitfalls of flawed data.

- System diversity is normalizing: While Microsoft Windows remains the predominant operating system, there are a healthy number of organizations that employ systems using Unix/Linux-derived operating systems, and 55% of the respondents indicated that both Windows and Linux are being used.

- Adoption of NoSQL is still early: Across respondents, adoption of alternative database technologies such as NoSQL, Graph and Time Series is still limited. However, there is some uptake on the use of the MongoDB NoSQL solution.

- Data protection in the cloud: In general, there is increased concern regarding compliance with data privacy laws, especially among users of cloud environments.
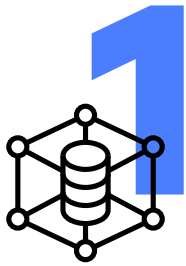
Introduction: Modernization Is Driving Data Strategy

Over the past five years, three technology trends have triggered a dramatic upheaval in organizational data strategy:

- Open-source frameworks for high performance computing that leverage commodity components, including Apache projects Hadoop, YARN and Spark, have drastically lowered barriers to entry for all businesses to implement big data infrastructures for operations and analytics.

- The economics and simplicity of cloud computing, bolstered by the growing array of cloud provider host-native services, has influenced many organizations in migrating their data and applications to the cloud.

- Better end-user reporting and analytics tools, coupled with increased end-user sophistication, have resulted in the emergence of the citizen analyst role. The citizen analyst is a business-oriented problem solver who is knowledgeable in the ways that analytics and machine learning algorithms are wielded to identify profitable business opportunities.

As a result, many organizations are launching modernization programs to consider how current and future business needs can be met through legacy system renovation. This involves moving data to cloud platforms, integrating that data in the cloud with applications deployed using cloud computing resources and native services, and leveraging a microservices architecture to simplify application development and deployment.

Progress®

Naturally, it is impossible for an organization to simultaneously modernize all of its applications, as that type of transition is likely to take a number of years. At the same time, there are a number of cloud hosts and cloud configurations available for organizations to consider and select. Organizations will increasingly need to manage an enterprise information infrastructure that can be deployed across a hybrid architecture that blends multicloud and on-premises systems.

This background provides context for the 2019 Data Connectivity report. This year, we focused on how respondents replied to questions about:

- Modernization and connectivity
- Data challenges
- Computing platforms (such as cloud platforms, platform choices, BI and reporting tools)
- Big data platform choices
- Database choices (including the traditional relational model, databases used for analytics and newer types such as time series, NoSQL and graph databases)
- BI and analytics
- Influencing factors (regulations, biggest data integration challenges)

When reviewing the responses to the 2019 survey, there are some clear observations that can be made about the critical aspects of system and data interoperability within an evolving hybrid on-premises/multi-cloud environment.

First, the mechanical aspects of modernization and cloud migration are still impacted by known data challenges such as data quality and data sprawl, as well as data volume, velocity and variety. In retrospect, combining the results of questions about challenges, connectivity, and regulatory compliance suggest that perhaps the list of data challenges might be expanded in upcoming data connectivity surveys.

Second, as modernization goes hand-in-hand with cloud migration, it is valuable to drill down into the responses about modernization in the context of cloud migration and choices of computing platforms in general, and cloud vendors in particular.

Third, with a growing pool of citizen data analysts thirsting for simplified analytics using massive data volumes, it is valuable to review big data technology, database technology and BI tool choices, especially when sliced by industry and location.

Finally, the results of the 2019 survey reflect general trends. Many organizations have already begun their migration to the cloud and organizations that have not should consider their existing impediments and begin preparing for that transition.

Baseline Demographics

2

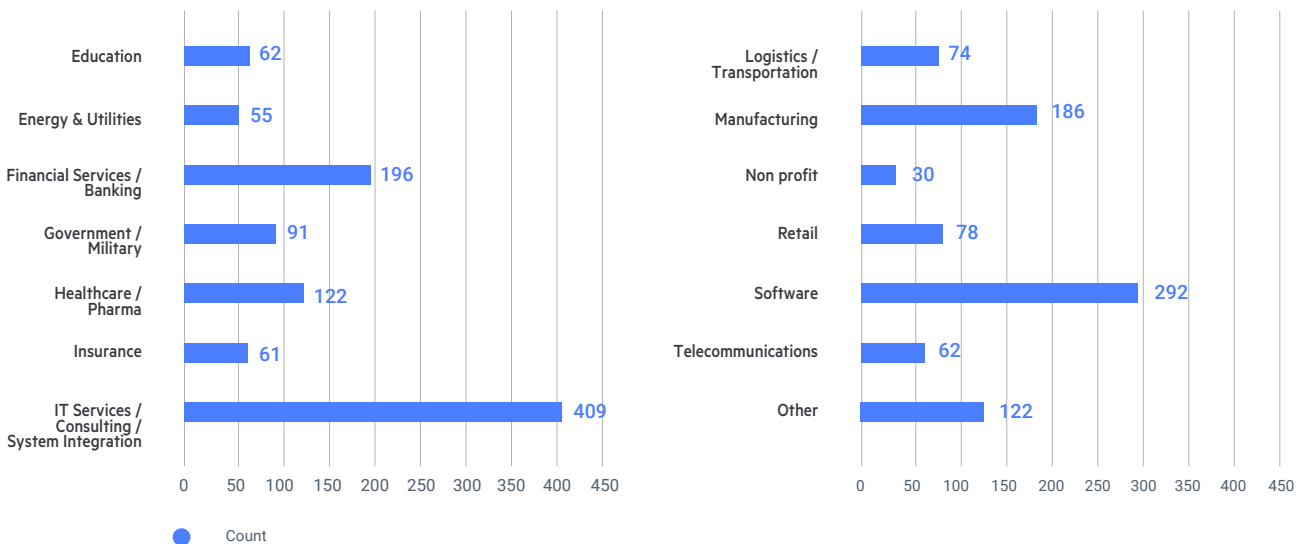As a way of providing context for the results of this year's survey, consider these basic demographics.

This year's survey was completed by 1,840 respondents across the globe (see Figure 1), with nearly half located in North America and a third in Europe/Middle East/Africa.

Figure.1
Number of respondents by geographic region.

Region	Percentage
Asia Pasific	15,65 %
Europe, Middle East and Africa	33,37%
North America	42.93%
South America	8.04%
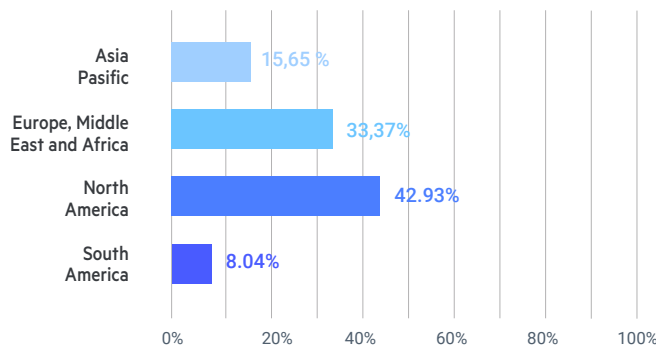
The respondents represented a wide variety of industries (Figure 2), with IT Services/Consulting/System Integration representing the largest cohort (22.23%) and Software a relatively close second (15.87%).

Figure.2
Breakout by industry.

Industry	Count
Education	62
Energy & Utilities	55
Financial Services / Banking	196
Government / Military	91
Healthcare / Pharma	122
Insurance	61
IT Services / Consulting / System Integration	409
Logistics / Transportation	74
Manufacturing	186
Non profit	30
Retail	78
Software	292
Telecommunications	62
Other	122

Count

Progress®

Because of the large number of respondents in IT Services or Software, it is not surprising to see that more than 24% of the respondents related that their job function was Software Development (Figure 3).

Breakout by job function.

Architect	19.08%
BI Specialist	3.53%
Business Analyst	4.95%
CEO	3.86%
CIO / CTO	7.23%
Consultant	8.80%
Data Engineer	2.72%
Data Scientist	1.96%

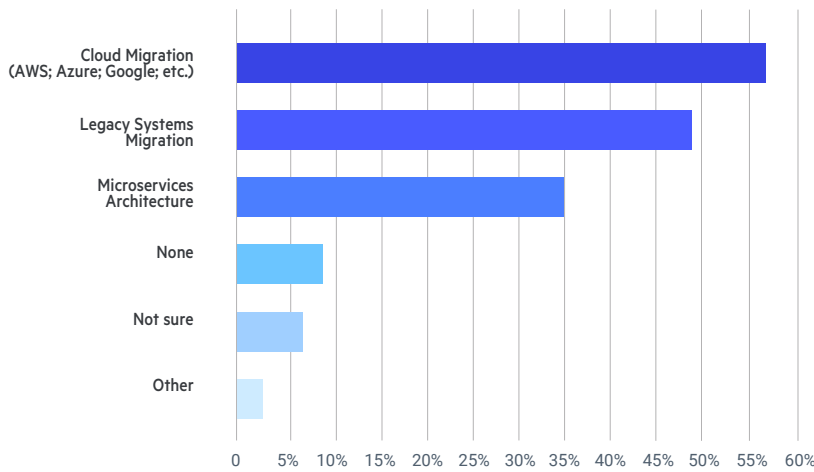| Database Administrator | 3.53% |
Database Analyst	1.68%
DevOps	2.28%
Product Management	2.72%
Sales Engineer	0.76%
System Administrator	5.43%
Software Development	24.62%
Other	6.85%

● Percentage
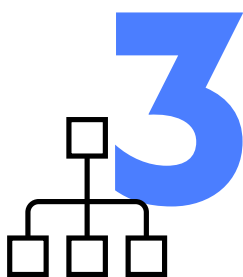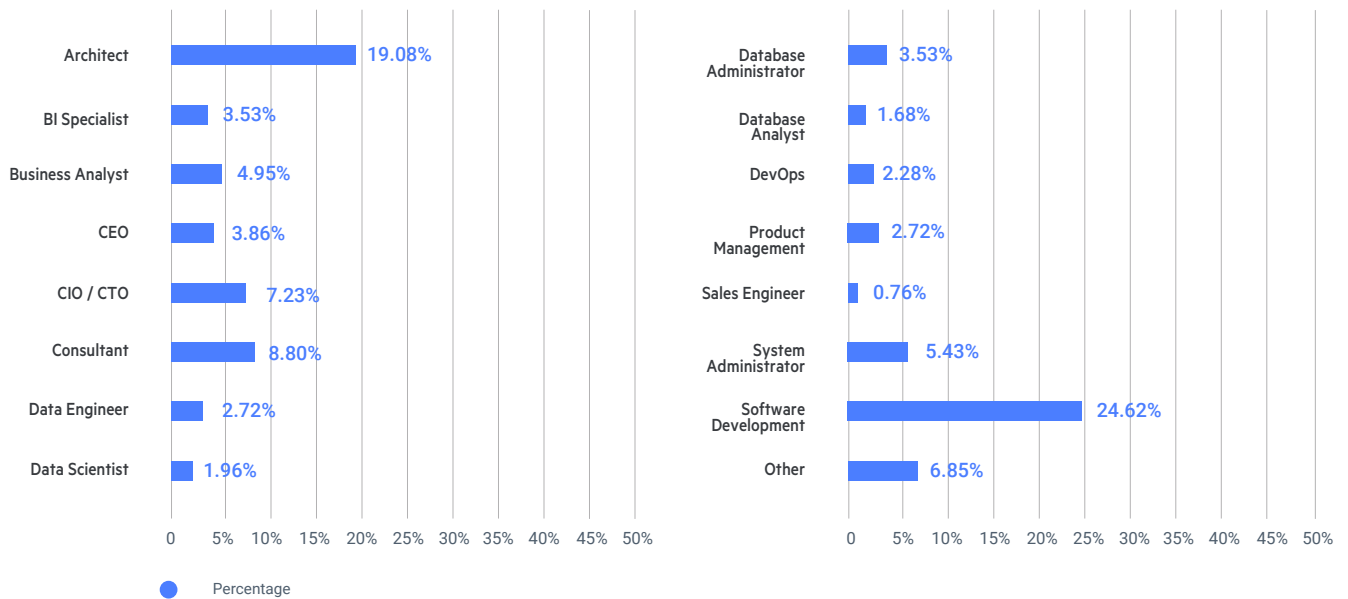
3 Modernization and Connectivity

3.1   Modernization

A new question in the 2019 survey asks, "What kinds of modernization projects is your company currently doing or planning to do?" and in the context of the industry trends, it is valuable to consider the responses in relation to the industry trends noted in section 1.

At first glance, our respondents appear to be aligned with the industry trends of organizations leveraging public cloud IaaS, PaaS and SaaS services. With respondents able to choose multiple selections, 56% noted their current and/or planned activity in cloud modernization, and nearly half (49%) indicated a legacy systems migration. Number three was microservices architecture, indicated by 36% of the responses. Curiously, 16% indicated they have no modernization projects underway or that they were not sure if they did.

Progress®

Percentages of modernization projects.

Cloud Migration (AWS; Azure; Google; etc.) — Legacy Systems Migration — Microservices Architecture — None — Not sure — Other

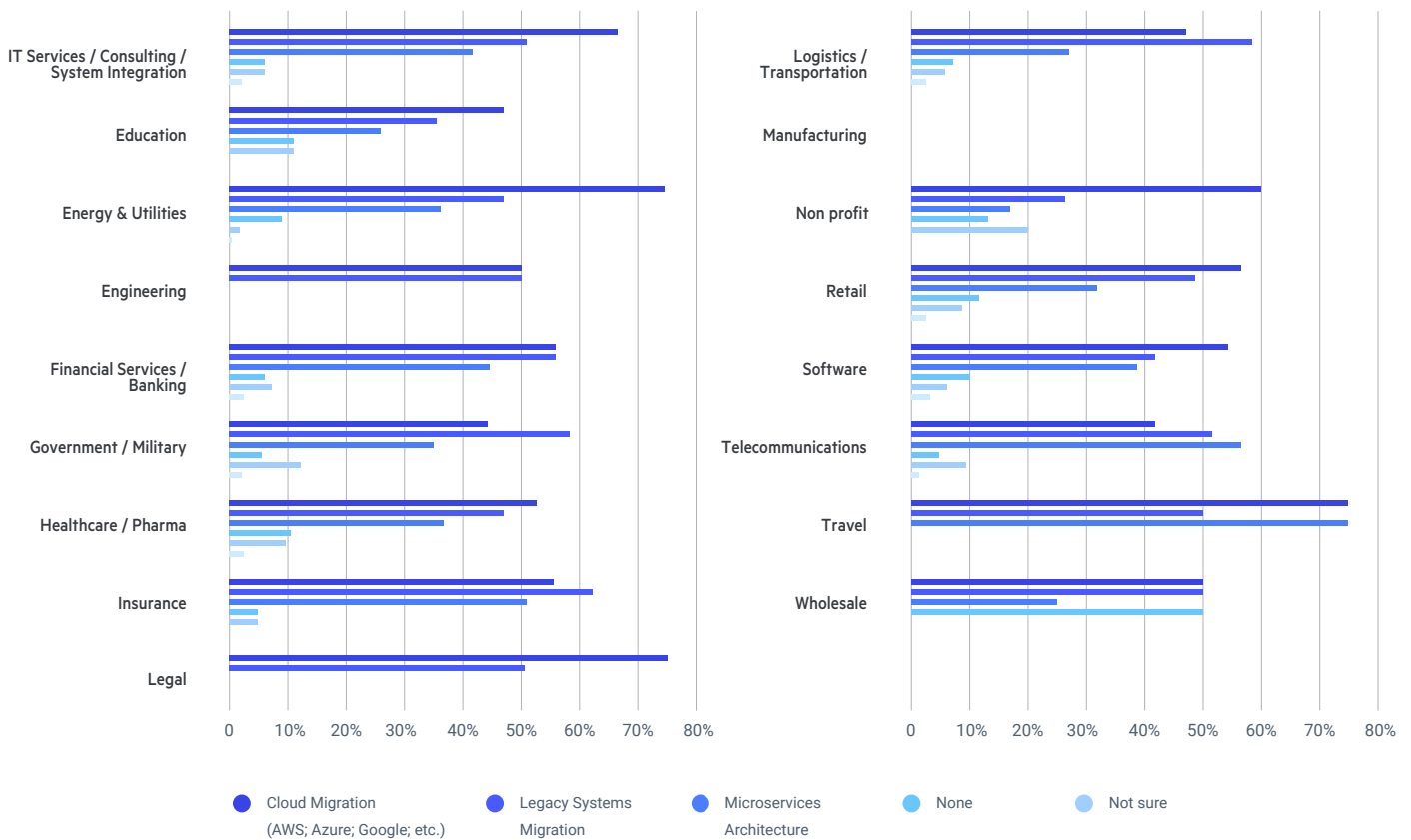0 5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60%

Analyzing the data by industry reveals some interesting patterns. While for most industries (e.g. IT services, healthcare, retail, legal, education, energy and utilities), cloud migration is the clear leader in modernization projects, there are some noteworthy exceptions. When we examine responses of individuals in the government sector, legacy systems migration is the leading modernization project, with 58% of those indicating they worked in the government/military making that choice as well. Cloud migration (AWS, Azure, Google, etc.) followed at 44%, and microservices at 35%.

This is not necessarily unexpected, as there is a predictable cadence for system renovation in government agencies, but these agencies may be slower to move to new technology platforms like the cloud as a result of a combination of risk avoidance (particularly around concerns of data protection) with the elongated cycle time for government procurement and contracting. We will continue to monitor government cloud adoption in the future as initiatives like FedRamp (for blanket approval of technologies) and improved data security work to reduce risk concerns.

Some industries are still grappling with older legacy systems and need to reengineer those systems to more modern ones before they can move to the cloud, and in these cases, legacy migration seems to be the priority. In the insurance industry, legacy systems migration is in the lead at 62%, with cloud at 56%. 58% of respondents identifying with the logistics/transportation industry chose legacy systems migration, followed by cloud migration at 47%.

Progress®

Motivations for modernization

Legend:
- Cloud Migration (AWS; Azure; Google; etc.)
- Legacy Systems Migration
- Microservices Architecture
- None
- Not sure

3.2 Connectivity

As with the 2018 report, this year's survey considered the question of data connectivity in the context of growing variety and volume of data across a modern organization's extended information enterprise. Today's data professionals and their business analyst clients are faced with increased numbers of data sources, a wider variety of methods by which datasets are accessed and a demand for standard drivers that reduce the need to develop new APIs and thereby simplify integration.

This year's survey asked respondents, "How do you currently connect your cloud apps to on-premises data that is behind the firewall?" As with the 2018 report, with increased cloud adoption, enterprises need to be able to integrate with data sources both on-prem and in the cloud across the firewall. Because of security requirements, there are several fundamental access solutions (such as access using a VPN, or connecting via SSH, etc.). As shown in Figure 5, 40% of the respondents indicated their use of a VPN, about 21% selected SSH tunneling, followed by reverse proxy with 9.5%.

Figure.6
Percentage of respondents indicating their methods of connecting cloud apps to on-premises data
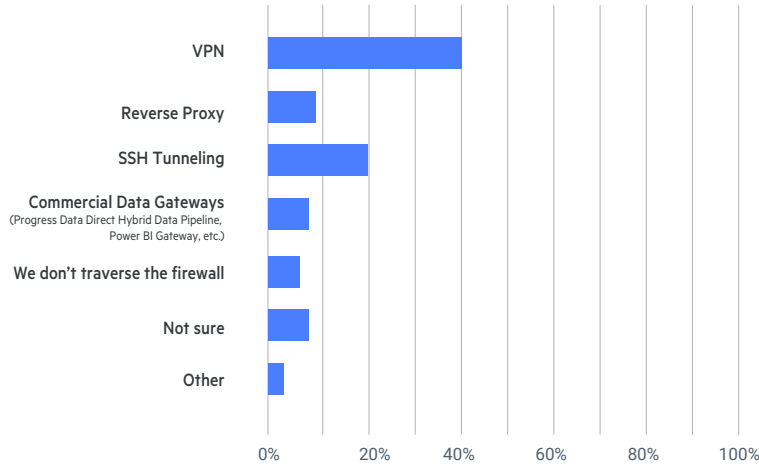
However, managing these solutions can become increasingly complicated in a hybrid environment, especially when multiple firewalls are involved and scalability is a priority. Hence the need for hybrid connectivity solution gateways.

This reinforces the importance of enterprise-ready, standards-based drivers to simplify development and implementation of modernized business intelligence, analytics, as well as operational applications that pull data from across a hybrid, multi-cloud environment. The demand for these hybrid connectivity solution gateways (e.g. Progress® DataDirect® Hybrid Data Pipeline and Power BI Gateway) also accounts for the notable increase of data gateways from 9% in 2018 to 13% in 2019.

Challenges

While there have been vast improvements in the breadth of solutions and tools supporting the modern hybrid environment, there are persistent issues and challenges that continue to plague organizations. Similar to 2018, the 2019 survey solicited information from respondents about their biggest data integration challenges (see Figure 1).

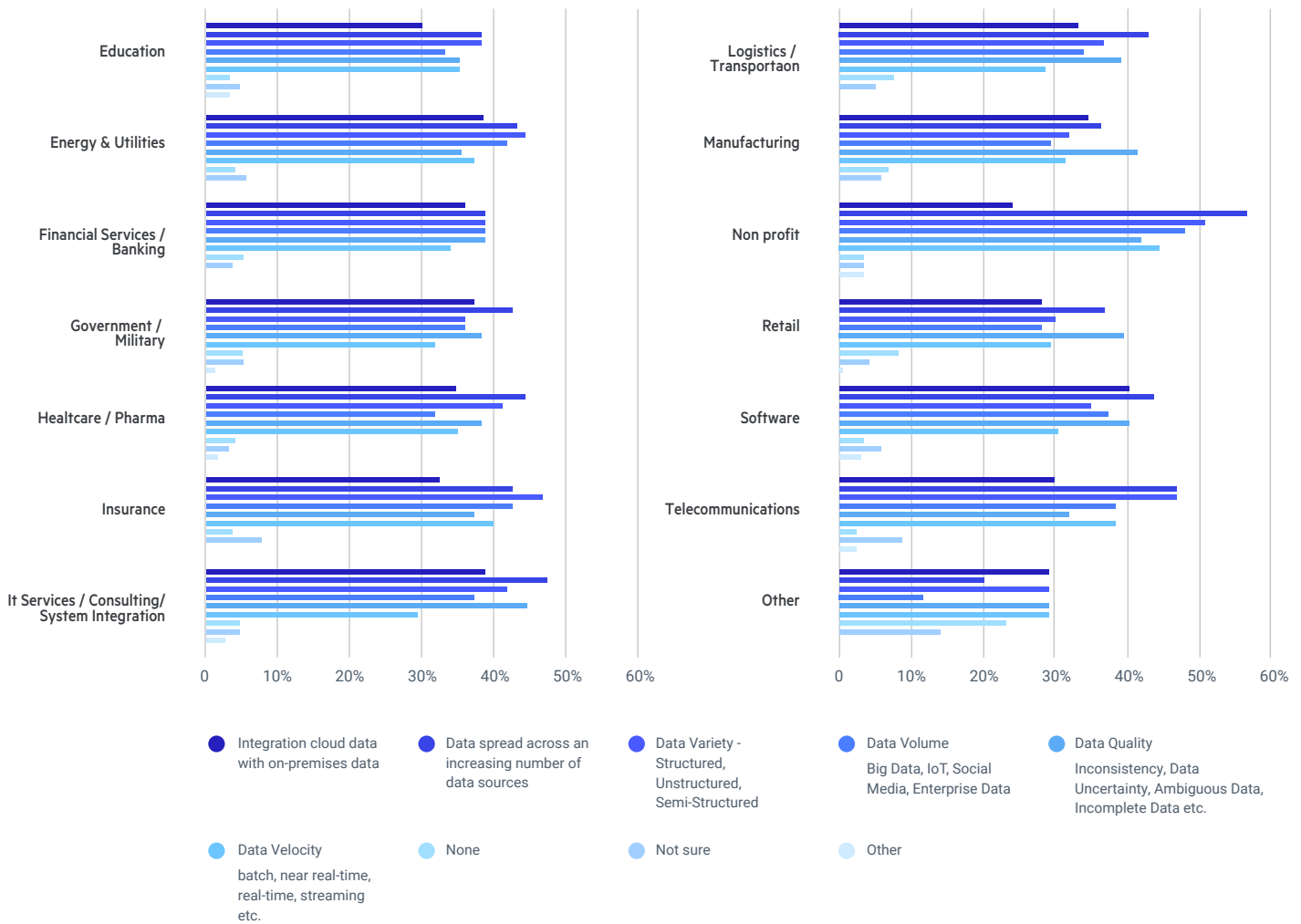**Percentage of respondents indicating their choice of data integration challenges.**
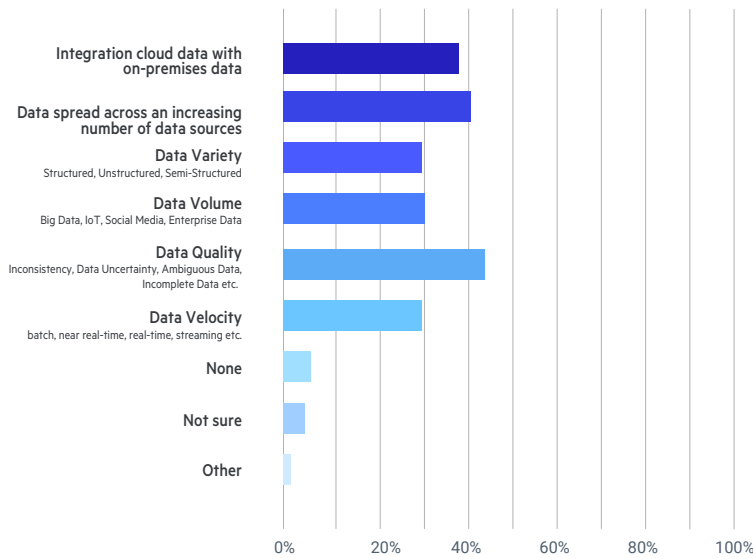
Data quality seems to be increasing in criticality. In 2019's survey, data quality/veracity/inconsistencies/incomplete data moved from third position in 2018 (30% of respondents) to the most frequent selection, chosen by 44% of the respondents—and an eyebrow-raising increase from 14% of the respondents in 2017!

With the exception of data velocity, all the other options seem to have decreased in frequency from 2017 to 2018, with the number of responses stabilizing in 2019. Data velocity increased from 20% in 2017 to 26% in 2018 and slightly increased to 29% in 2019.

Other responses remained fairly static. "Data spread across an increasing number of data sources" was the second most frequent selection (40%) and "Integrating cloud data with on-premises data" was the third-most popular (38%). The remaining options were distant with "Data Volume" at 30% and "Velocity" and "Variety" both at 29%. Some of the challenges submitted by respondents indicating "Other" included themes revolving around data security, regulatory compliance and data privacy—issues that are reflected in a subsequent question about the organization's subjection to regulations.

One interesting consideration was revealed after slicing the results by industry (see Figure 8): While data quality was the most frequently chosen selection over all of the responses, in individual industries, data sprawl ("Data spread across an increasing number of data sources") and data variety topped the responses!

Figure.8
Challenges, broken out by industry.

Legend:
- ● Integration cloud data with on-premises data
- ● Data spread across an increasing number of data sources
- ● Data Variety - Structured, Unstructured, Semi-Structured
- ● Data Volume — Big Data, IoT, Social Media, Enterprise Data
- ● Data Quality — Inconsistency, Data Uncertainty, Ambiguous Data, Incomplete Data etc.
- ● Data Velocity — batch, near real-time, real-time, streaming etc.
- ● None
- ● Not sure
- ● Other

5 Computing Platforms

5.1 Hybrid/Services Environment

This year's survey solicited information about the platforms and operating systems in use by the respondents or their customers to support their applications (see Figure 5). Like last year (and not unsurprisingly!), Microsoft Windows not only continued to be the leading operating system, its selection rose from 84% of respondents in 2018 to 87% in 2019. Unix/Linux-based operating systems also experienced an increase in popularity over the prior year, with Mac OS X rising from 15% in 2018 to 22% in 2019, and "generic Linux" experiencing a substantial increase from 55% of 2018 respondents to 65% in 2019.

Progress®

Figure.9
Percentage of respondents reporting use of platform/OS.
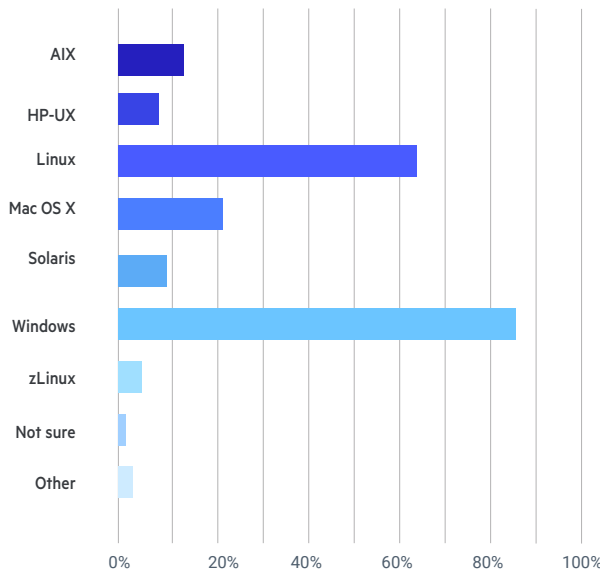
The growth of Linux-based systems may be attributable to a few potential drivers. First, in general, adoption of open-source applicationware is growing, and many open-source products are rooted in the Linux OS, suggesting a correlation. Second, increased use of cloud platforms, which often use Linux distributions as their core operating system, may also account for that growth.

How common is it for organizations to indicate their commitment to a single operating system technology?

Among the responses, approximately 22% indicated that Windows was their only operating system, while 5.5% only selected Linux. On the other hand, over 55% chose both Windows and Linux. These results reinforce the presumption that organizations are increasingly working in a hybrid environment that requires visibility of data and reporting/analytics across multiple end-user platforms. This implies the need for streamlined connectivity—creating transparent layers across the different components of the hybrid platform demands seamless self-service for users as well as low-latency data movement and exchange in a virtualized manner to satisfy reporting and analytics needs.

5.2   Cloud Adoption

Our 2018 report highlighted how the increasing pace of business has influenced large enterprises to expedite the migration of their data and their workloads from on-premises environments to the cloud. In the 2018 report, we saw increased adoption among the top three cloud providers (AWS, Azure and Google Cloud) over the prior year, and that trend continued in 2019.

Azure and Google Cloud demonstrated the most significant growth and remain dominant players in the cloud space. Compared to the 2018 numbers, the percentage of respondents in this year's survey indicating that they used AWS increased from 44% to 48%, Azure increased from 39% to 47% and Google from 18% to 25%. Amazon and Azure are still the biggest players in the space, with Amazon adoption nearly twice that of the third player (Google), according to respondents.

According to our survey, AWS, Azure and Google continue to dominate the market, although it appears that Azure is closing the gap, nearly catching up to AWS. In the 2018 survey, Google's popularity trailed VMware's, but their positions are reversed in 2019, with the percentage of respondents selecting VMware decreasing from 22% in 2018 to 20% in 2019.

Figure.10
Cloud adoption percentages.

Progress®
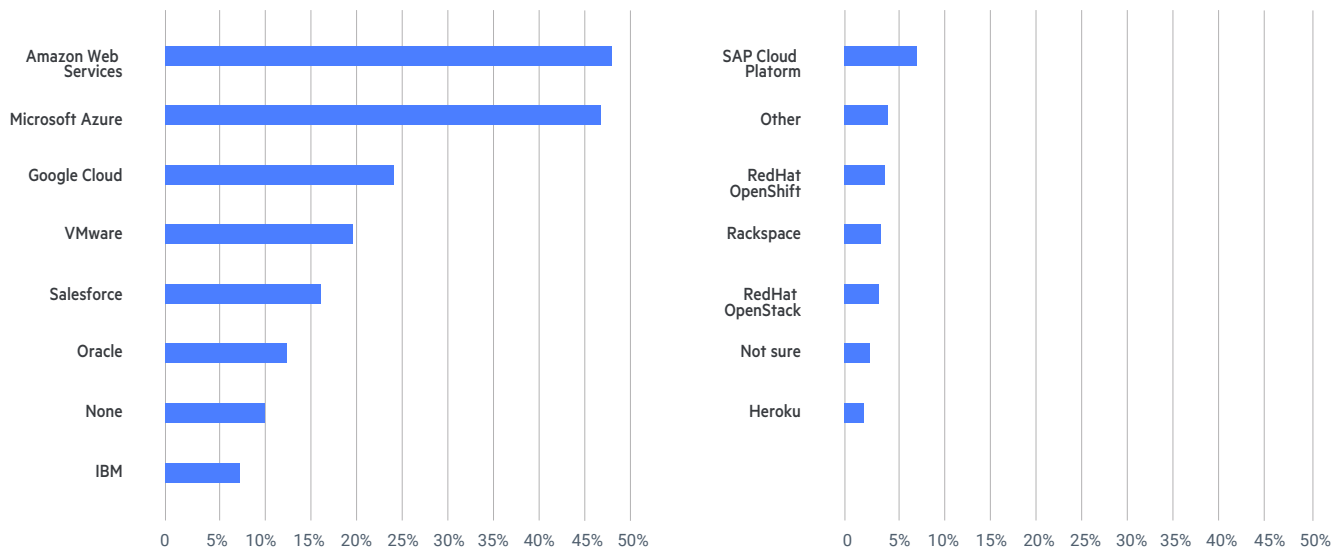
According to the 2019 survey, AWS and Azure maintained their positions as the most frequently selected public cloud infrastructure (see Figure 6). Their popularity might be attributable to the breadth of their services offerings. Both vendors provide Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) offerings as well as a variety of value-added services (ranging from serverless computing to integrated machine learning) that enable clients to use them to not only "lift and shift" existing on-premises applications but also enable development of full end-to end solutions. Additionally, Microsoft's cloud offerings include Office 365, and many customers who have used Microsoft desktop products are migrating to the SaaS version of those products through Office 365.

Looking at sectors with significant representation among survey respondents, we see some interesting patterns emerge when we break down the findings by industry. Similar to the overall results of the survey, 57% of the respondents in the software industry opted for AWS, 46% for Azure and 29% for Google Cloud. However, there were some industries, such as energy and utilities, insurance and nonprofit, in which Azure was chosen at a higher rate than AWS (see Table 1).

Industry	% choosing AWS	% choosing Azure	% choosing Google
Energy & Utilities	29%	56%	22%
Insurance	42%	49%	16%
Non-Profit	23%	47%	27%

Table 1: Percentages of respondents within each industry selecting a cloud provider.

Other variations from the "big three" are seen. In the telecommunications and logistics/transportation industries, VMware is chosen more frequently than Google. In telecommunications, 27% chose VMware (vs. 21% for Google), and in logistics/transportation, 31% chose VMWare (vs. 26% selecting Google). Respondents from the government and military space indicate that AWS (46%) and Azure (47%) are almost even, with only 20% of the respondents choosing Google. In the distribution industry, AWS takes the lead at 67% while Google and Azure run neck and neck at 33%. Finally, Salesforce seems to have some popularity in specific industries, including IT services/consulting/system/integration (24%), energy and utilities (22%), insurance (20%) and manufacturing (19%).

When we consider respondent choice by region, adoption of cloud platforms generally reflects the overall distribution, with AWS the leading choice, Azure a close second and Google a distant third. However, there is one notable exception—EMEA respondents indicated that Azure is used more frequently, chosen by 47% of the respondents vs. 39% choosing AWS.

Progress

Figure.11
Selections of cloud vendors by region.
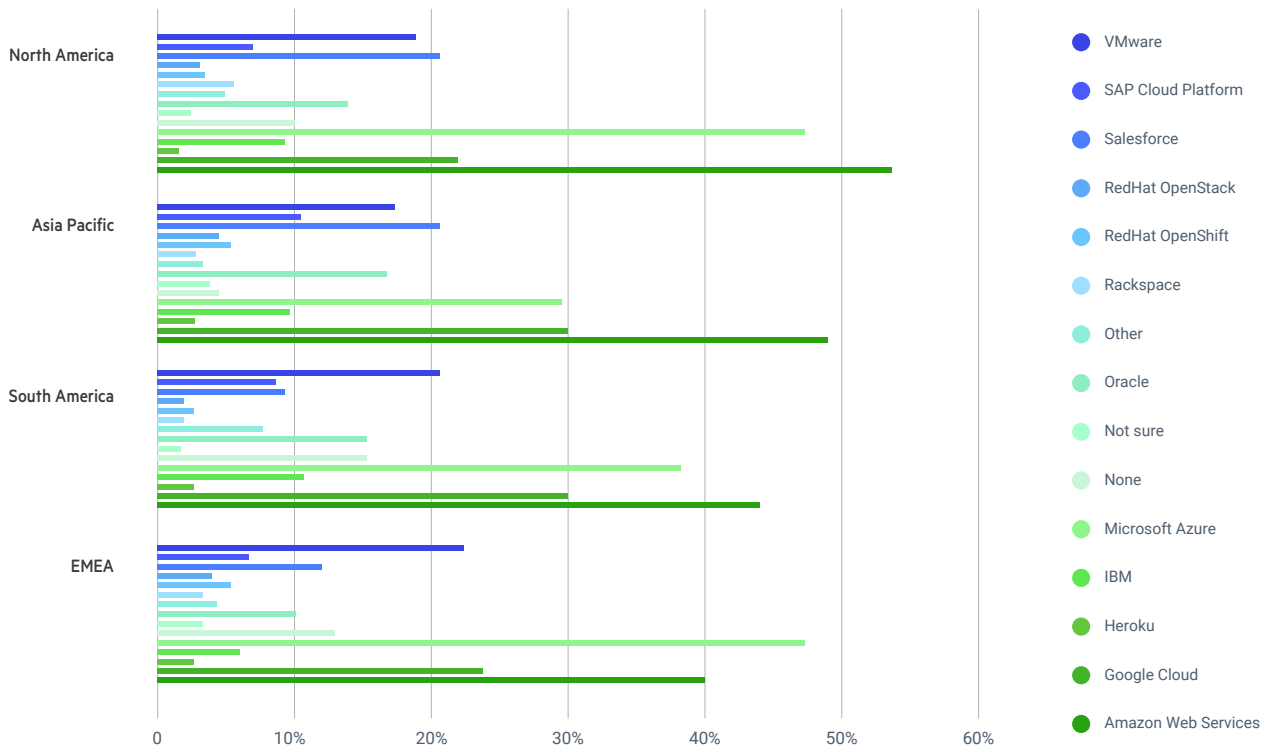
Cloud adoption is strong in Asia Pacific. This region had the lowest percentage of respondents (5%) indicating they were not using cloud services, followed by North America at 10%, EMEA at 13% and South America at 15%. This suggests that there are still organizations on the fence regarding cloud adoption.

Table 2: Percentage of respondents within each region indicating use of three top cloud vendors.

An interesting pattern we considered is where respondents indicated the use of all of the top three cloud vendors, ranging from 9% to 13% across each region (see Table 2). This is indicative of a growing collection of organizations that are adopting a multi-cloud strategy.

Region	Percentage using AWS, Azure and Google Cloud
North America	9%
Asia Pacific	13%
South America	11%
Europe and Middle East	10%

Conversely, we can look at where there are clear preferences for top vendors. What about those who use Amazon but not Azure? As shown in Table 3, more North American and South American respondents indicated a preference for AWS over Azure than the other way around, with the opposite being true in EMEA.

Region	Percentage using AWS but not Azure	Percentage using Azure but not AWS
North America	25%	18%
Asia Pacific	22%	21%
South America	20%	15%
Europe and Middle East	14%	21%

5.3   Big Data Platforms

Overall, the big data platform results of the 2019 survey are very much the same as last year, with the percentages for the selections for most platform products close to those registered in the 2018 survey. It is interesting to note that 31% of the respondents indicated they are not using any big data platform, while 12% are not sure. This suggests that there is still a large number of organizations that have not yet identified a significant role for big data as part of their data environment.

That being said, among the selected options, both in 2018 and in the 2019 survey, Amazon S3 (17% of respondents) and Hadoop Hive (15% of respondents) are the most frequently selected platform products. The relative popularity of these choices is reflective of the different big data adoption strategies: taking advantage of scalable object storage (S3) or relying on traditional Hadoop ecosystem components for structured data analysis (Hadoop Hive). See overall percentages in Table 1 and a comparison in Figure 12.

When looking at organizations' future plans, we see that the number of respondents selecting "none" and "not sure" are much lower, suggesting that more organizations plan to adopt big data technologies in the near future (see Figure 13).

Apache Spark SQL is the next most popular selection, nearing 10% in this survey. And although overall current adoption of big data platforms is not widespread, there is a small number of respondents (approximately 5%) who indicated they plan to adopt either S3 or Hive within the next two years. Otherwise, the collective responses did not suggest a coming wave of adoption of other technologies within the next two years.
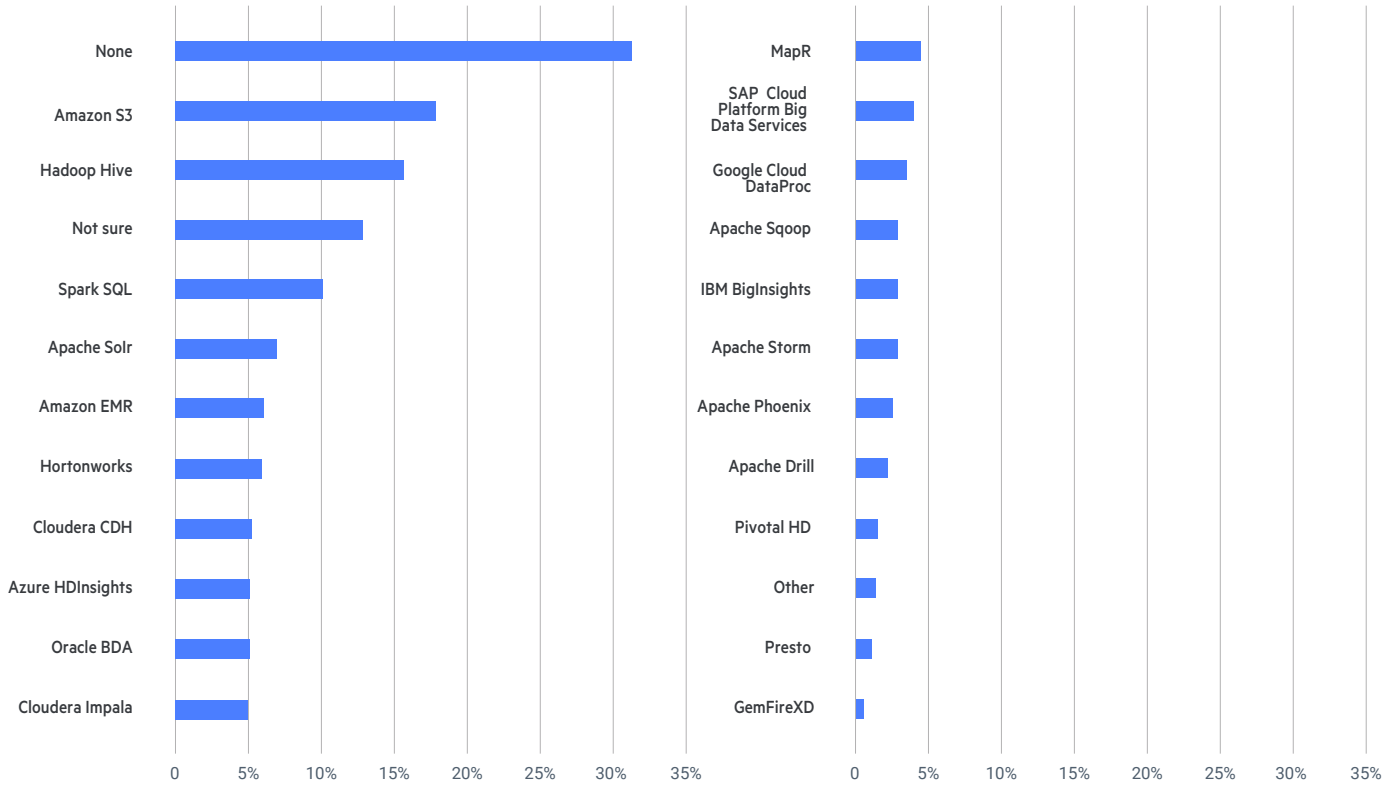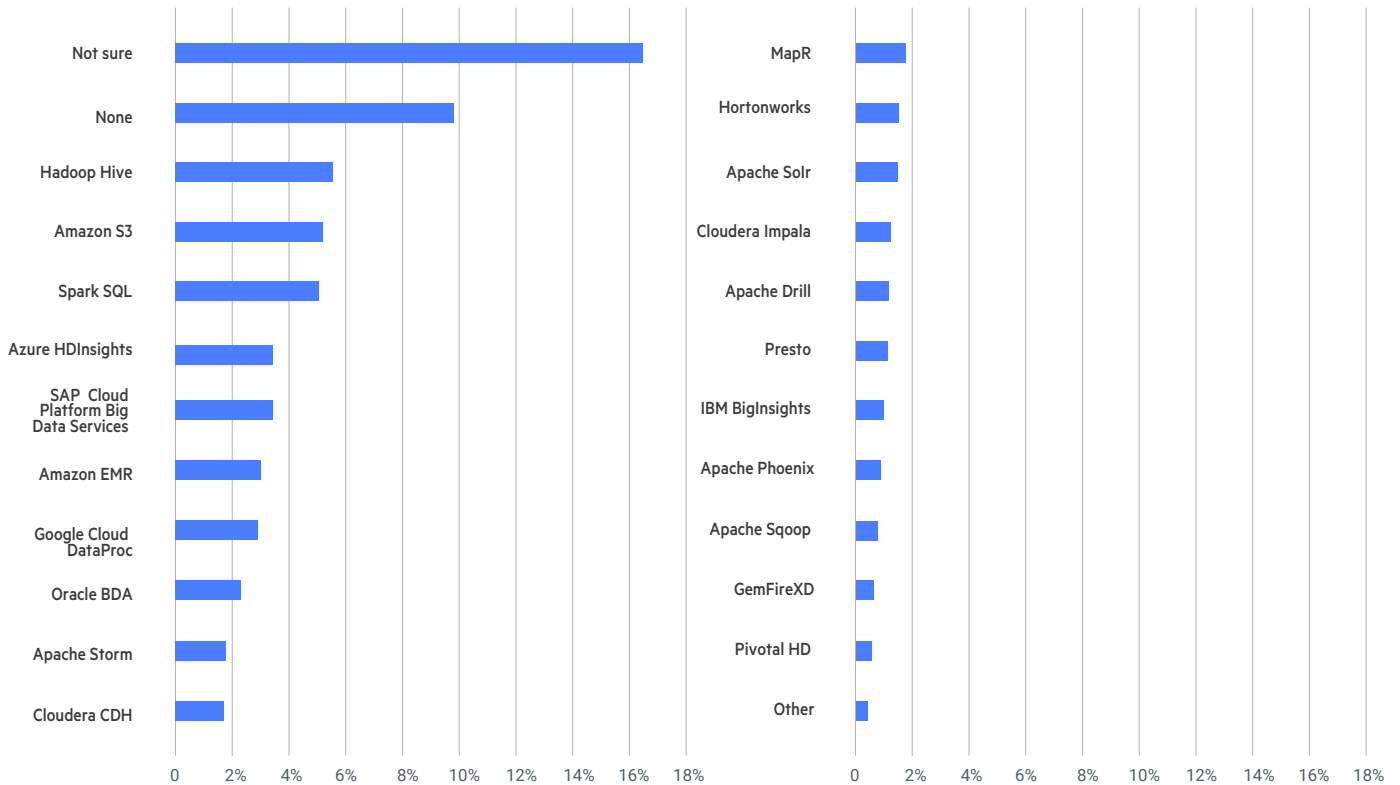
Progress

Progress®

Current big data platform use.

**Percentages of respondents expecting to
adopt big data platforms in the next 2 years.**

We also considered the selections of big data platforms in relation to the choices by respondents indicating use of a cloud platform. While, overall, the most frequently selected cloud vendors were AWS, Azure and then Google Cloud, it should not be surprising that when respondents selected big data platforms provided by a particular cloud vendor (such as AWS S3), the choice of cloud vendor tended to overwhelmingly be the vendor providing that service. For example, individuals reporting that they used AWS S3 typically selected AWS as one of their cloud vendors.

5.4   Big Data Platforms by Industry

Slicing the big data platform choices by industry is somewhat revealing, especially when looking at those reporting that they use no big data platform. While the overall percentage was 31%, 49% of those working in manufacturing, 33% in nonprofit and 33% in government indicated they use no big data platform. The industries with the lowest reported numbers of no big data platform are telecommunications (19%) and energy (20%).

Object storage in AWS S3 was reported to be used by all industries, with the highest levels reported in IT services (27%), education (21%), software (20%) and financial services (19%).

Hadoop Hive seems to be most frequently selected by those indicating they work for the insurance industry (23%), followed by those working in the financial industry (22%), and then the IT services/consulting industry (21%). There is limited adoption among those in energy, telecom and government, in which Hive was chosen between 14%-16% of the time.

When considering Hadoop implementations, Hortonworks is selected in the retail and financial Industry 9% of the time, and in the telecom industry 8% of the time, followed by insurance, IT services and software at 7% of the time.

Cloudera's Hadoop implementation appears to be more popular than Hortonworks in the education, government/military and IT services/consulting/system integration industries. In the telecom, financial service insurance, retail, logistics, energy and utilities, manufacturing and logistics/transportation industries, Hortonworks appears to be chosen more frequently (see Table 4 below). In the past year, though, Cloudera and Hortonworks have merged, and it will be interesting to review how these selections will change in the upcoming year's survey.

Progress

Region	CDH %	HDP %
IT Services/Consulting/System	8.801956	6.845966
Education	8.196721	3.278689
Government/Military	7.692308	4.395604
Software	7.534247	7.191781
Telecommunications	4.83871	8.064516
Financial Services/Banking	4.081633	8.673469
Healthcare/Pharma	3.278689	4.918033
Insurance	3.278689	6.557377
Energy & Utilities	1.818182	3.636364
Manufacturing	1.604278	3.208556
Logistics/Transportation	1.351351	4.054054
Retail	1.282051	8.974359

Table 4: Hortonworks vs. Cloudera, by industry.

Apache Spark, which is essentially the follow-on to Hadoop, is gaining adoption, especially because of its integrated ability for executing SQL queries. When looking at the selection of SparkSQL by industry (see Table 3), IT services/consulting/ system integration and energy and utilities are the most frequent adopters (both just under 15%), followed by government/military, financial services/banking and software (at just under 11%).

Table 5: Spark SQL selection by industry.

Region	PCT
IT Services/Consulting/System Integration	14.66993
Energy & Utilities	14.54545
Government/Military	10.98901
Financial Services/Banking	10.71429
Software	10.61644
Retail	8.974359
Telecommunications	8.064516
Healthcare/Pharma	7.377049
Logistics/Transportation	6.756757
Manufacturing	5.347594
Insurance	4.918033
Nonprofit	3.333333
Education	3.278689

Progress®

5.5 Big Data Platforms by Region

Despite what is reflected in the technology media, there was a surprisingly high percentage of respondents indicating that their organizations do not currently use a big data platform. When segregated by region, South American respondents most frequently indicated that they were not using a big data platform (37%), while 34% of EMEA respondents and 30% of North American respondents said the same. Meanwhile, only 24% of the APAC respondents indicated that they were not using a big data platform.

There are similarities in the most frequently selected choices in North America, APAC and EMEA: the top three selections for big data platforms are Spark SQL, Hadoop Hive and Amazon S3, although they appear in different orders across those three regions. Interestingly, a higher percentage of the APAC respondents selected each of those top three choices than the respondents from either North America or EMEA. In South America, the top three tools are Hadoop Hive (11.5%) Amazon S3 (10%) and Oracle BDA at 9%. In general, APAC respondents indicate a higher percentage adoption rate of big data platforms than the other regions.

The overall percentages of big data platform choices are shown in Figure 14, Figure 15, Figure 16 and Figure 17.

Figure.14
Percentage of EMEA respondents selecting different big data platforms.
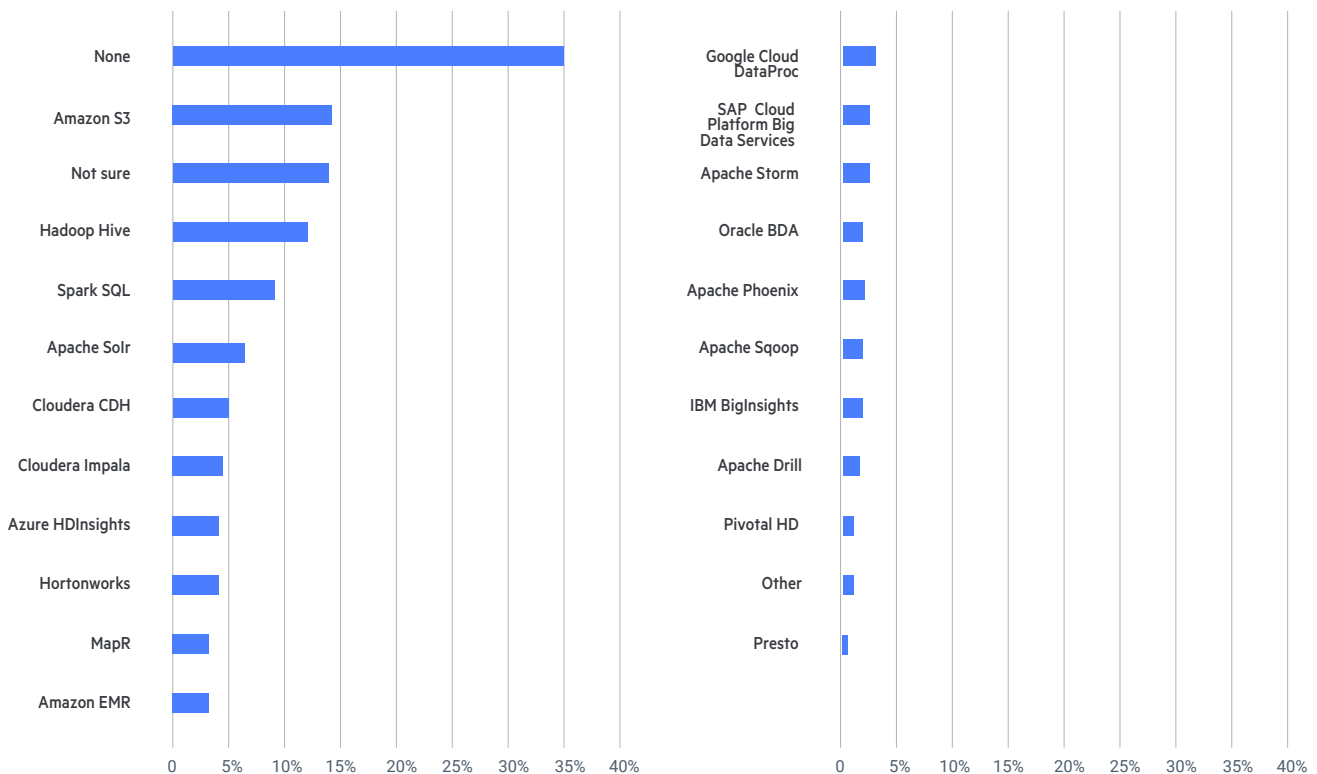
Figure.15
Percentage of South American respondents selecting different big data platforms.
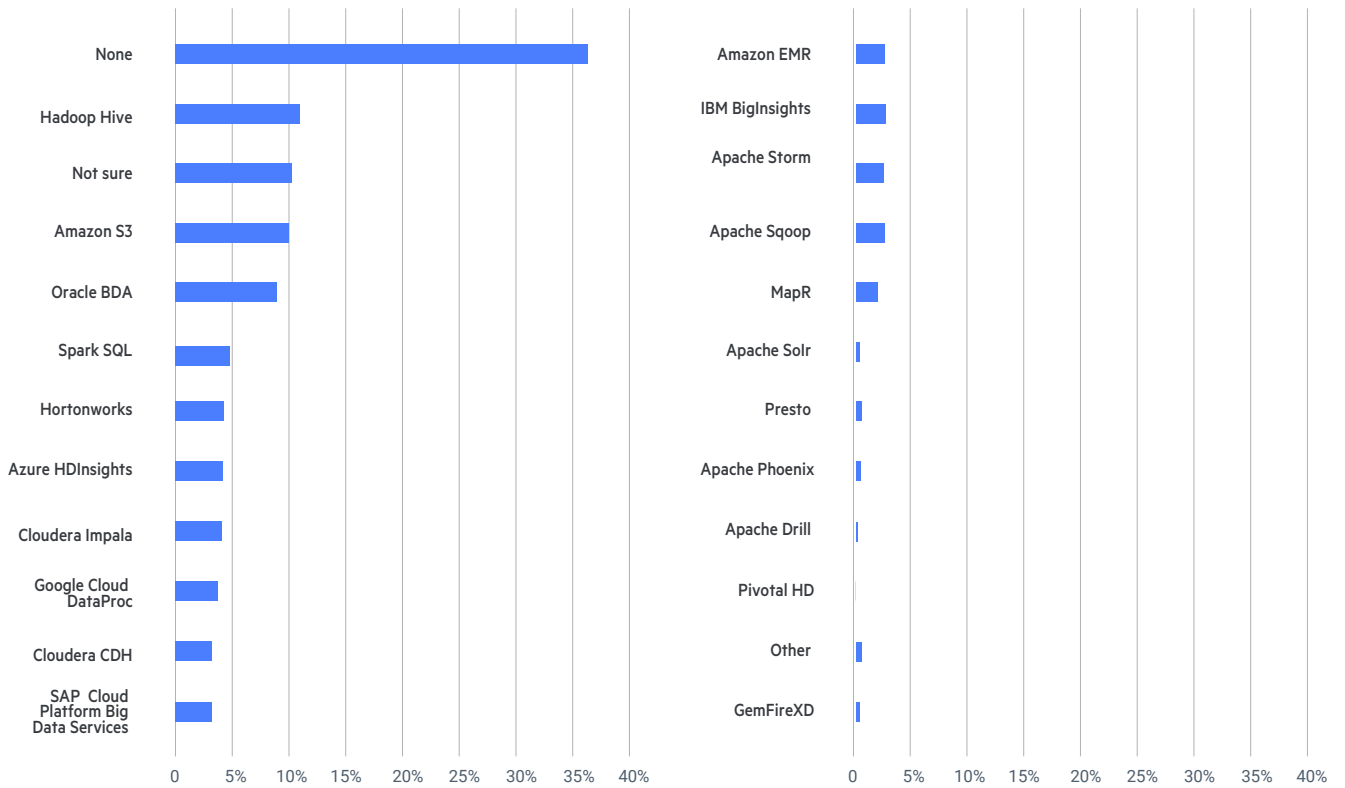

Figure.16
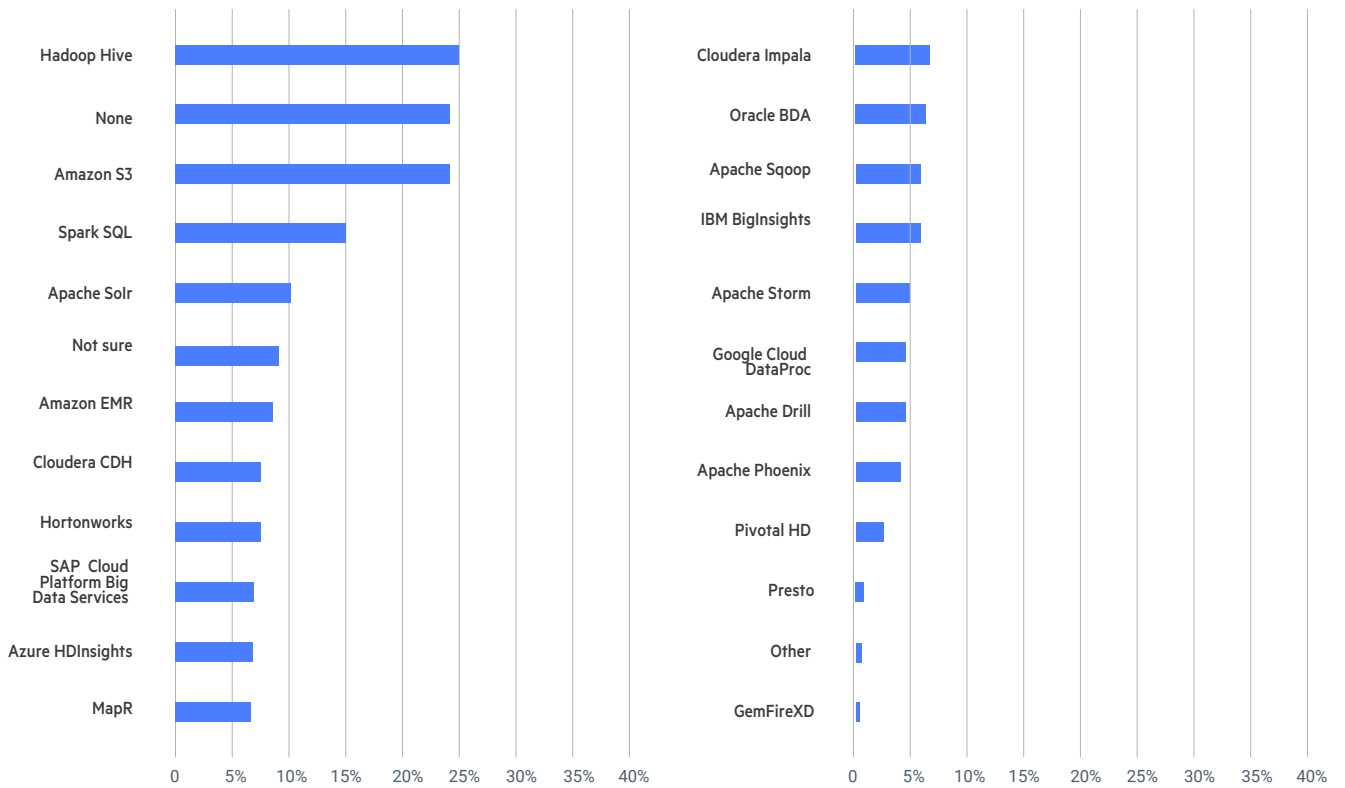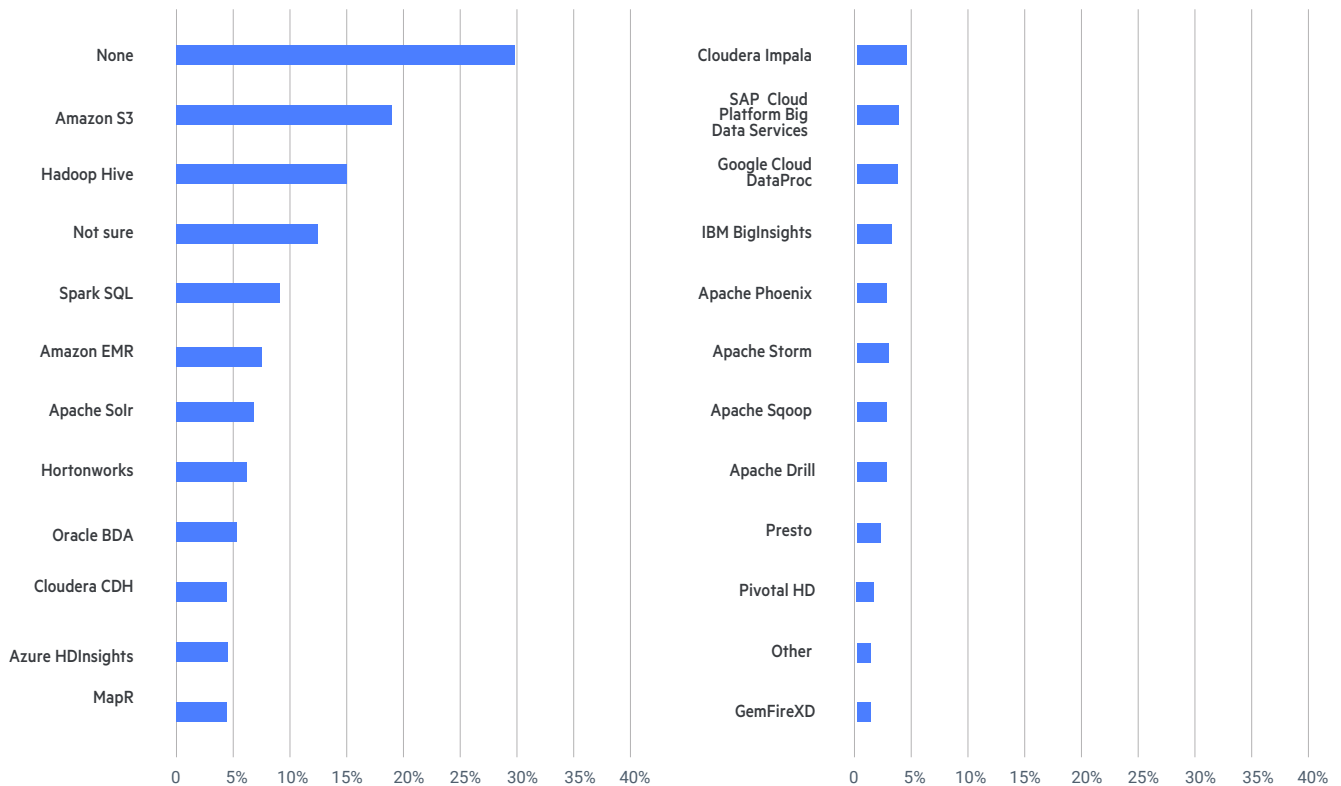**Percentage of APAC respondents selecting different big data platforms.**

Percentage of North American respondents selecting different big data platforms.

Databases

In this year's survey, the top three relational databases remain the same as the 2018 survey. As can be seen in Figure 18, a number of the popular choices from 2018 remained in the top ten, including the leader, SQL Server at 57%, MySQL at 41%, Oracle at 37%, PostgreSQL at 24% and Microsoft Access at 22%. The number of respondents indicating they are using MariaDB has increased since 2018, which can be explained in the context of MySQL's being controlled by Oracle while MariaDB remains open source.

In accordance with evidence of increased cloud adoption, Microsoft SQL Azure has jumped to third place, increasing from 19% of the respondents in 2018 to 27% in the 2019 survey. The use of databases engineered for the cloud has grown since 2018 as well, with Amazon Redshift increasing from 7% in 2018 to 10% in this survey and Amazon Aurora growing from 5% to 8% in the same time span.

The interest in adopting cloud-based databases is reinforced by those reporting their future plans. Microsoft SQL Azure is the most popular planned database technology (9%), followed by Amazon Redshift and Amazon Aurora both at 6%.

Figure.18
Percentage of respondents choosing their databases.

● Amazon Aurora	● PostgreSQL
● Amazon Redshift	● Progress OpenEdge
● FireBird	● SAP HANA
● Google BigQuery	● Snowflake
● Google Cloud SQL	● SQL Server
● IBM DB2	● SQLite
● Infomix	● Sybase ASE
● MariaDB	● Sybase IQ
● MemSQL	● Teradata
● Microsoft Access	● Vertica
● Microsoft SQL Azure	● VoltDB
● MySQL	● None
● Netezza	● Not sure
● Oracle	● Other
● Pivotal Greenplum	
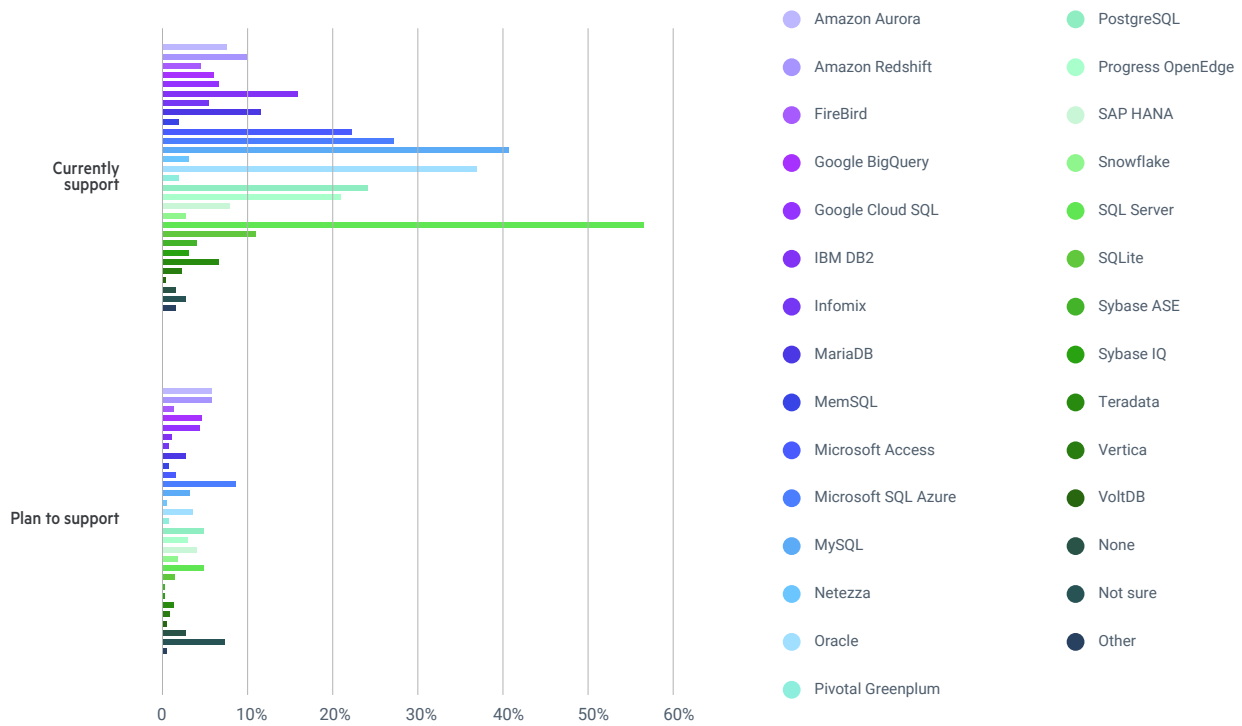
6.1   Database Selections by Industry

When examining the selections broken out by industry, there are some interesting observations. While overall, the most popular database choice is SQL Server (chosen nearly 57% of the time), it is used less frequently in some industries. For example, 24% of respondents working in the IT services industry selected SQL Server, 16% in the software industry chose SQL Server, 11% in manufacturing and financial service each and only 6% in healthcare/pharma.

We see a similar pattern for the next most popular database, MySQL. While overall it was selected 41% of the time, it was selected by those in IT services at 28%, software at 16%, financial services at 9%, manufacturing at 9%, and healthcare/pharma at 6%. Likewise for Oracle (number three on the list)—while it was chosen 37% of the time overall, IT services chose it 25% of the time, software 16%, financial services 13%, manufacturing 8% and healthcare/pharma 7% of the time.

◢ Progress®

Interestingly, we see some variation from the general pattern of SQL Server, MySQL and Oracle. For example, the most frequently chosen databases for those reporting that they work in the logistics/transportation industry are SQL Server (66%), MySQL (41%), Progress® OpenEdge® (31%) and then Oracle (28%). Similarly, in the manufacturing industry, the order of the top choices differs from the overall numbers, with 61% choosing SQL Server and 36% selecting Progress OpenEdge, which is then followed by MySQL (35%), Microsoft Access (30%) and Oracle at 29%. Finally, in the finance industry, while SQL Server is the most popular at 56%, Oracle (45%) appears to be more popular than MySQL (35%). Several DB technologies are then tied at 22%: IBM DB2 (which may be attributable to the prevalent use of IBM mainframes in the finance industry), Microsoft SQL Azure and PostgreSQL.

When restricting the analysis to those who indicated they use cloud services, we generally see the same top three choices (SQL Server, MySQL and Oracle). However, among the 244 respondents who indicated they use Oracle Cloud services, the most frequently chosen relational database was Oracle (77%), followed by 53% choosing MySQL (which is also an Oracle product).

We see another interesting example of brand consistency among the 860 respondents who indicated they use Microsoft Azure, in which 71% said they use SQL Server, followed by 50% using Microsoft SQL Azure, which is then followed by MySQL (48%) and Oracle (41%).

And while you might expect that of the 878 respondents who indicated they use AWS Cloud service to prefer the AWS database services (Amazon Redshift or Amazon Aurora), it turns out that SQL Server is still most frequently selected (60%), MySQL at 50% and Oracle at 42%.

6.2   Alternative Databases

NoSQL databases provide alternative models for representing structured and semi-structured data and are growing in popularity, especially in areas where there is net new application development. It is surprising, though, that in our 2019 survey, 32% of the respondents indicated they don't use a NoSQL database.

Among those reporting that they use a NoSQL database, MongoDB is by far the most popular, coming in at the first position of a named technology at 24% of respondents. That is a lower percentage than in 2018 when 27% selected MongoDB. The next most frequently selected option, Cassandra (a little less than 10%) is also lower than last year's percentage (11%). Some NoSQL choices that gained popularity between 2018 and now include Amazon DynamoDB, which rose from 5% in 2018 to 8%, and Redis, which rose to 8% from 7% in 2018. There

Progress®

were only a few more options that rated higher than 5% of the current survey's respondent selections, including Google Cloud Datastore (6%), Azure Cosmos DB and Oracle NoSQL (both 5.4%), and HBase at 5% (see Figure 19).

Figure.19
Percentage of NoSQL databases
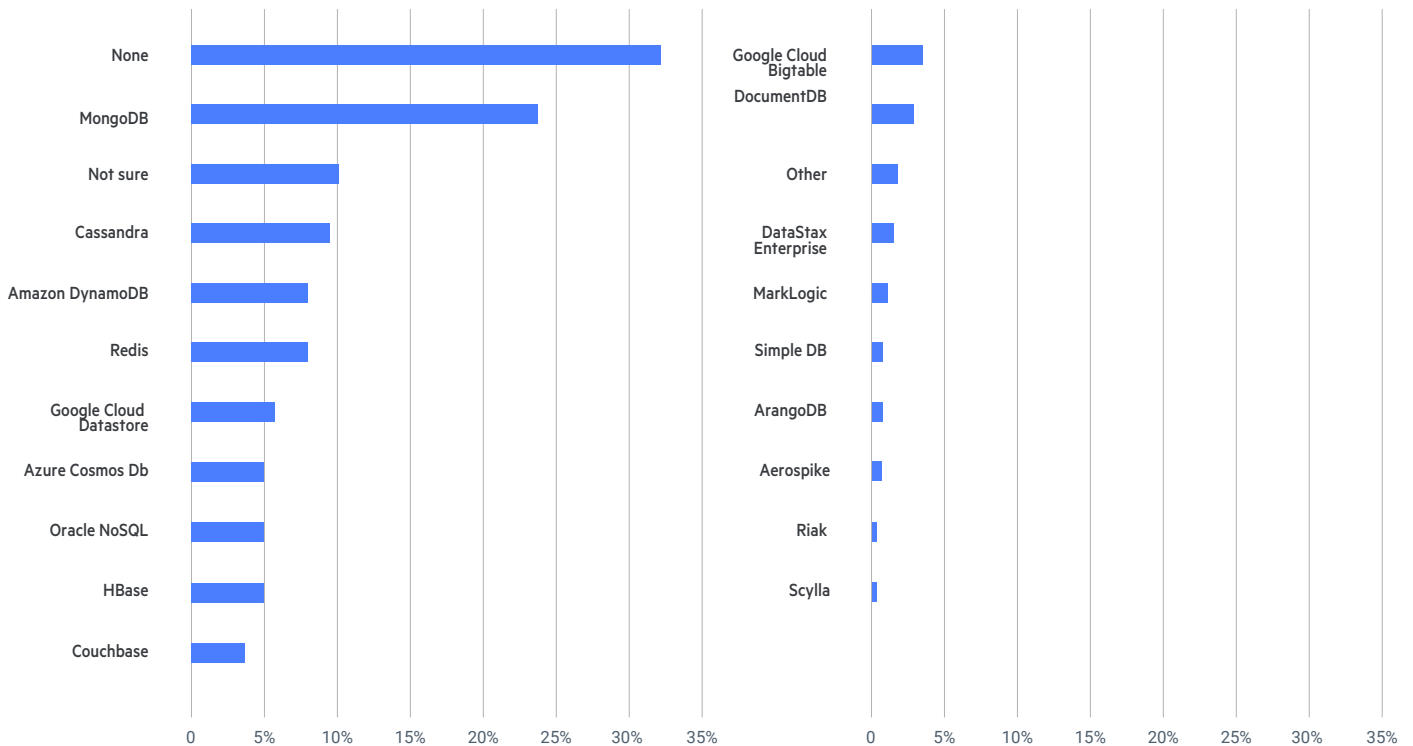
Another class of non-relational databases is time series databases, which are still in their infancy and have not yet reached significant adoption. As Internet of Things (IoT) and other real-time data streams (either from automated machine-generated data, sensors and devices embedded within both residential and commercial "smart environment" technologies, autonomous vehicles and sensor-connected manufacturing machinery) become more commonplace, there will be a growing need for database systems that can scale to support real-time ingestion of high-speed streaming data. Consequently, we can expect that time series databases will be piloted and introduced more frequently.

In the 2019 survey, over 51% of the respondents indicated they do not use a time-series database. The most frequently named selection of a time series database, selected by 5% of the respondents, was Prometheus. Other tools like InfluxDB (4%), Amazon Timestream (4%) and Graphite (3%) had similarly low adoption rates (see Table 6).

Progress

26

Answer	Q10.current	Q10.planned
Amazon Timestream	3.641304348	4.184782609
Graphite	3.043478261	2.173913043
InfluxDB	3.967391304	2.010869565
Kdb+	0.869565217	0.706521739
None	50.70652174	16.0326087
Not Sure	23.69565217	22.06521739
OpenTSDB	1.847826087	1.684782609
Other (please specify):	0.869565217	0.326086957
Prometheus	4.673913043	2.608695652
RRDTool	2.445652174	0.923913043

Table 6: Percentage selections for time series databases

Graph databases are another alternative NoSQL approach that have gained some general attention because their storage models and data representations enable more sophisticated analytics. Graph databases can be used for building relationship "networks" (such as representing connections in social networks), and are used for security, fraud analysis and for retail recommendation and marketing applications. As interest in these analyses grows, expect to see greater adoption of graph databases.

However, similar to the time series databases, there is limited adoption among this year's survey participants. According to respondents, 51% do not use any kind of graph database and an additional 20% are not sure. The leading choice at a little below 6% was GraphDB, followed by Azure Cosmos DB, Neo4j and AWS Neptune (see Table 7).

Table 7: Percentage of those using graph databases.

Graph DB Technology	PCT
None	50.70652
Not Sure	20.05435
GraphDB	5.869565
Azure Cosmos DB	5.597826
Neo4j	5.326087
Amazon Neptune	3.532609
Other (please specify):	1.141304
JanusGraph	1.141304
TigerGraph	1.086957

Progress®

7 BI and Analytics

The expanding pool of citizen data scientists has expectations for using a suite of reporting and analytics capabilities and services that simplify their analysis processes while increasing accessibility to a wide spectrum of available enterprise data assets. This constant need to support a wide array of business intelligence and analytics consumers focuses attention on the types of tools required to drive operational and strategic decision-making.

Our current survey asked respondents to indicate vwhich business intelligence (BI) and reporting tools they used (see Figure 20), and which were planned to be adopted within the next two years. As might be expected, many people indicated their continued reliance on desktop end-user tools, with Microsoft Excel being selected most frequently by 40% of the respondents. The second most frequently chosen option was Microsoft BI Platform, selected by 27%. This year, Power BI moved into the third most popular slot (chosen by 26% of respondents), followed by Tableau (22%). This represents a change from 2018, where Tableau was the third most frequently selected choice. The popularity of Power BI is clearly growing—in the 2018 survey, only 18% selected Power BI, and that product also led the list of platforms planned to be adopted within the next two years, with 6% of respondents selecting that choice, followed by Tableau (5%).

We sliced the results by industry, and as might be expected, Excel continues to be the most frequently selected tool across all industries. In the financial services/banking, healthcare/pharma, government/military, retail, insurance, education, energy and utilities and nonprofit industries, Microsoft BI Platform is the second most frequently selected option. In the IT services/consulting/system integration, software, manufacturing and logistics/transportation industries, the second most selected BI tool is Power BI.

Despite the apparent popularity of the two Microsoft BI platforms, there are some instances where Microsoft BI Platform and Power BI didn't make it into the second or third spot. In the financial Industry, Microsoft BI Platform is in second place (31%), Tableau (26%) occupies the third position and Power BI was selected as the fourth most frequent choice at 24%. In both healthcare/pharma and government, Tableau displaces Power BI and among the respondents in the retail industry, Power BI is selected less frequently than Tableau, SAP IBM Cognos and QlikView. In the telecom industry, Tableau (23%) is the second most frequently selected product, with Microsoft BI Platform in third place (18%), SAP in fourth place (16%) and MicroStrategy and Power BI tied in fifth place at (15%).

Progress

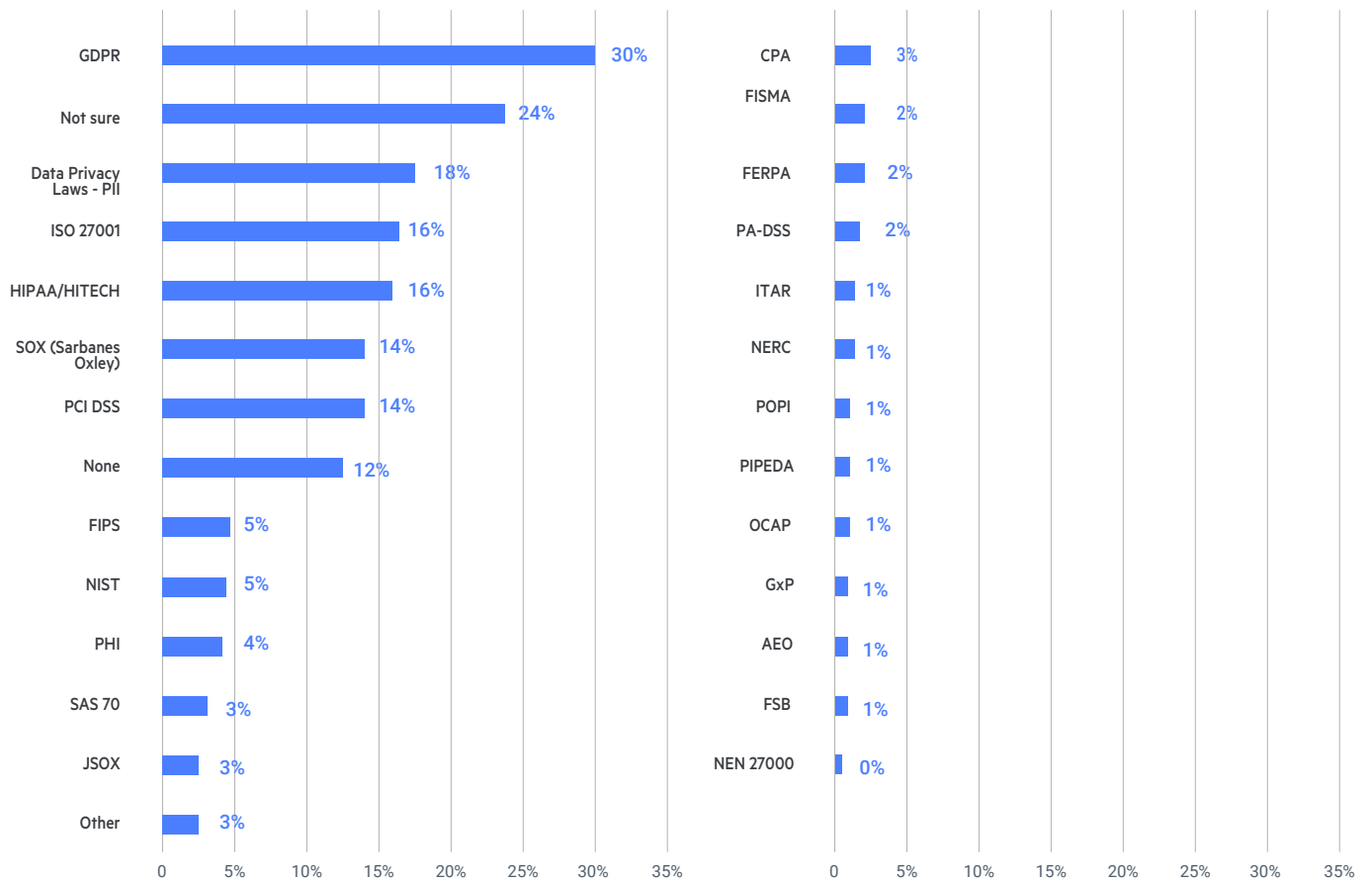**Percentage of respondents and their selected BI and reporting tools.**
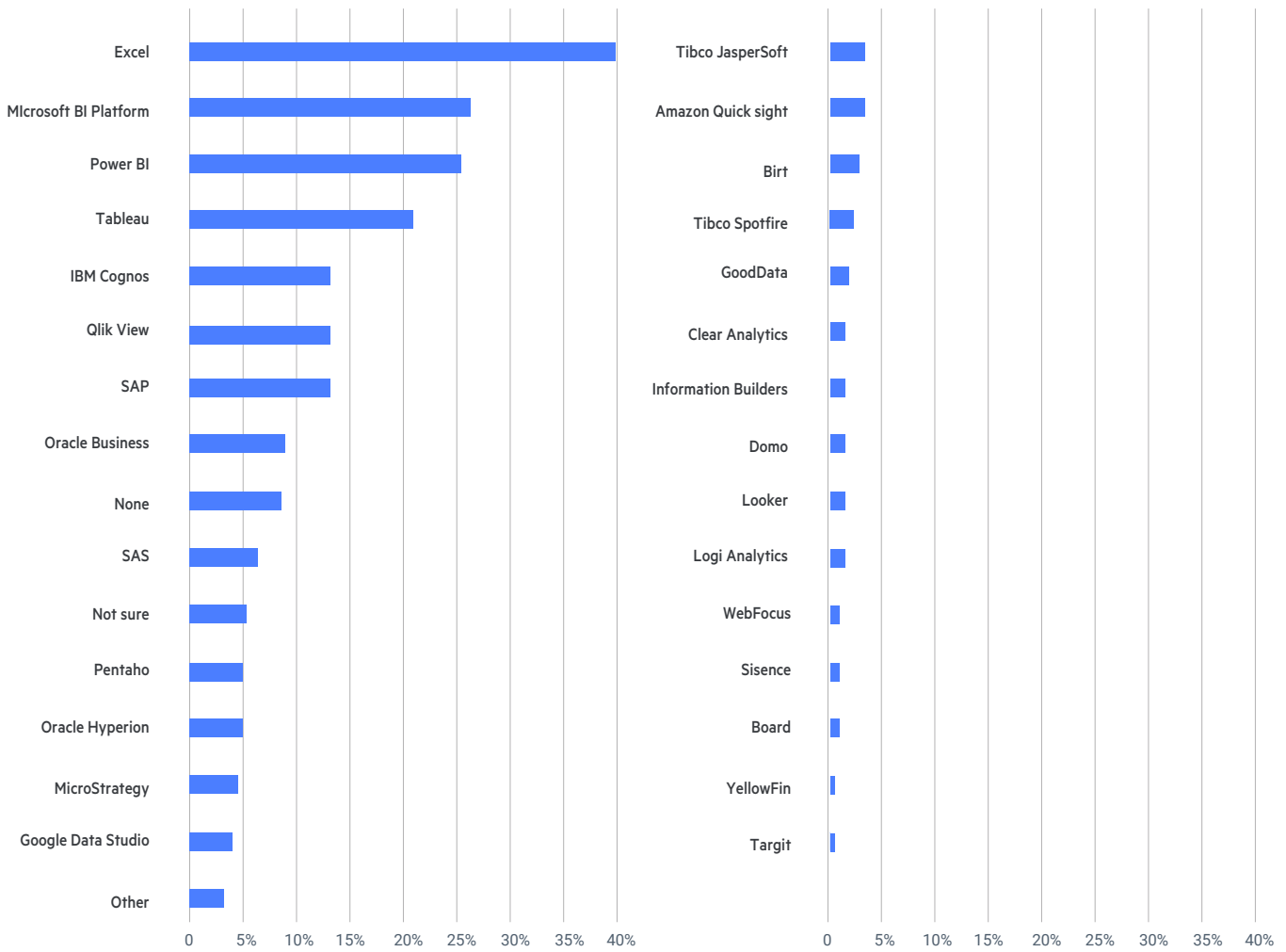
Influencing Factors of Regulations

8.1 Basic Statistics

Organizations are subject to many different regulations, and similar to 2018, this year's survey asked about which regulations organizations are subject to. Other than a few exceptions, the results for this report were largely in line with the prior year.

One exception is the EU's General Data Protection Regulation, or GDPR, which went into effect in May of 2018. While in 2018, 15% of respondents indicated that their organizations were subject to GDPR, in the current survey, that percentage doubled to 30%! Another notable change in general is that Data Privacy Laws-PII rose from sixth place at 12% in 2018 to second place at 18%. With GDPR's meteoric rise coupled with increases in the ISO 27001 standard for information security, HIPAA/HITECH and the Payment Card Industry Data Security Standard (PCI DSS), it is clear that the criticality of sensitive data protection cannot be denied.
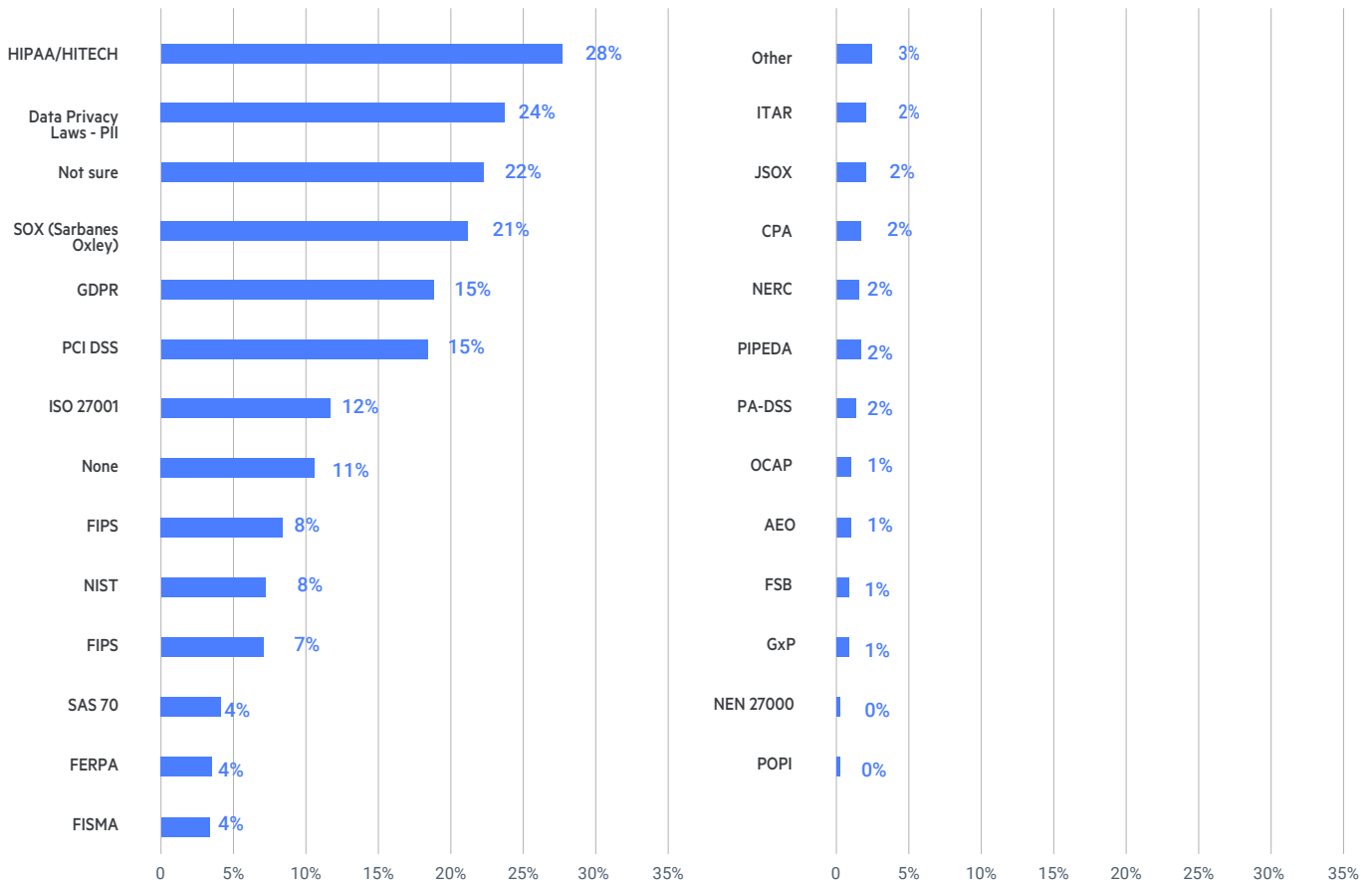
Figure.21
Percentage of respondents subject to regulations.

Regulation	%		Regulation	%
GDPR	30%		CPA	3%
Not sure	24%		FISMA	2%
Data Privacy Laws - PII	18%		FERPA	2%
ISO 27001	16%		PA-DSS	2%
HIPAA/HITECH	16%		ITAR	1%
SOX (Sarbanes Oxley)	14%		NERC	1%
PCI DSS	14%		POPI	1%
None	12%		PIPEDA	1%
FIPS	5%		OCAP	1%
NIST	5%		GxP	1%
PHI	4%		AEO	1%
SAS 70	3%		FSB	1%
JSOX	3%		NEN 27000	0%
Other	3%			

Progress®

8.2 Regulations by Geography

One would typically expect that organizations are more likely to be bound by regulations within their own geopolitical regions, and this is reflected in the responses to the survey. For North American respondents, while GDPR is definitely important, it was eclipsed by HIPAA/HITECH and Data Privacy Laws-PII. This might reflect anticipation of the California Consumer Privacy Act (CCPA), which is effective as of January 1, 2020. Of course, Sarbanes-Oxley (SOX) compliance was selected relatively frequently (see Figure 22).

Figure.22
North American regulatory expectations.

Regulation	%
HIPAA/HITECH	28%
Data Privacy Laws - PII	24%
Not sure	22%
SOX (Sarbanes Oxley)	21%
GDPR	15%
PCI DSS	15%
ISO 27001	12%
None	11%
FIPS	8%
NIST	8%
FIPS	7%
SAS 70	4%
FERPA	4%
FISMA	4%
Other	3%
ITAR	2%
JSOX	2%
CPA	2%
NERC	2%
PIPEDA	2%
PA-DSS	2%
OCAP	1%
AEO	1%
FSB	1%
GxP	1%
NEN 27000	0%
POPI	0%

When we turn our attention to EMEA, it is no surprise that a whopping 53% of the respondents selected GDPR. We also see ISO 20071 is in second place with 20%, followed by Data Privacy Laws at 12% (see Figure 23).

In the Asia Pacific region, GDPR is the most selected named regulation at 26%, ISO 27001 is in second position at 20%, with Privacy Laws-PII a close third at 19% (see Figure 24).

In South America, GDPR and Data Privacy Laws-PII are less frequently selected than ISO 27001 which is chosen by 20%, followed by SOX at 14% (see Figure 25).

North American regulatory expectations.

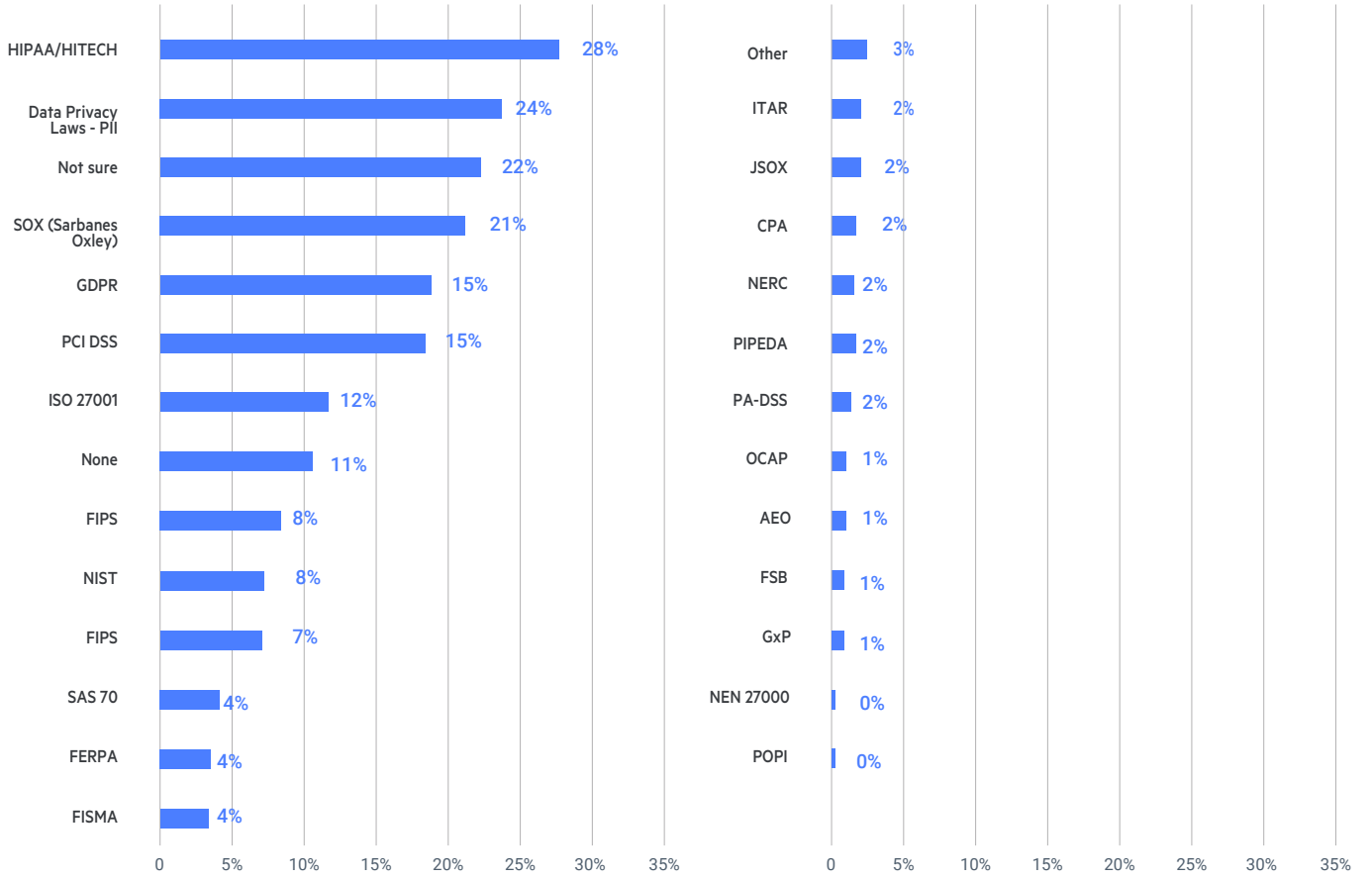Figure.22

North American regulatory expectations.

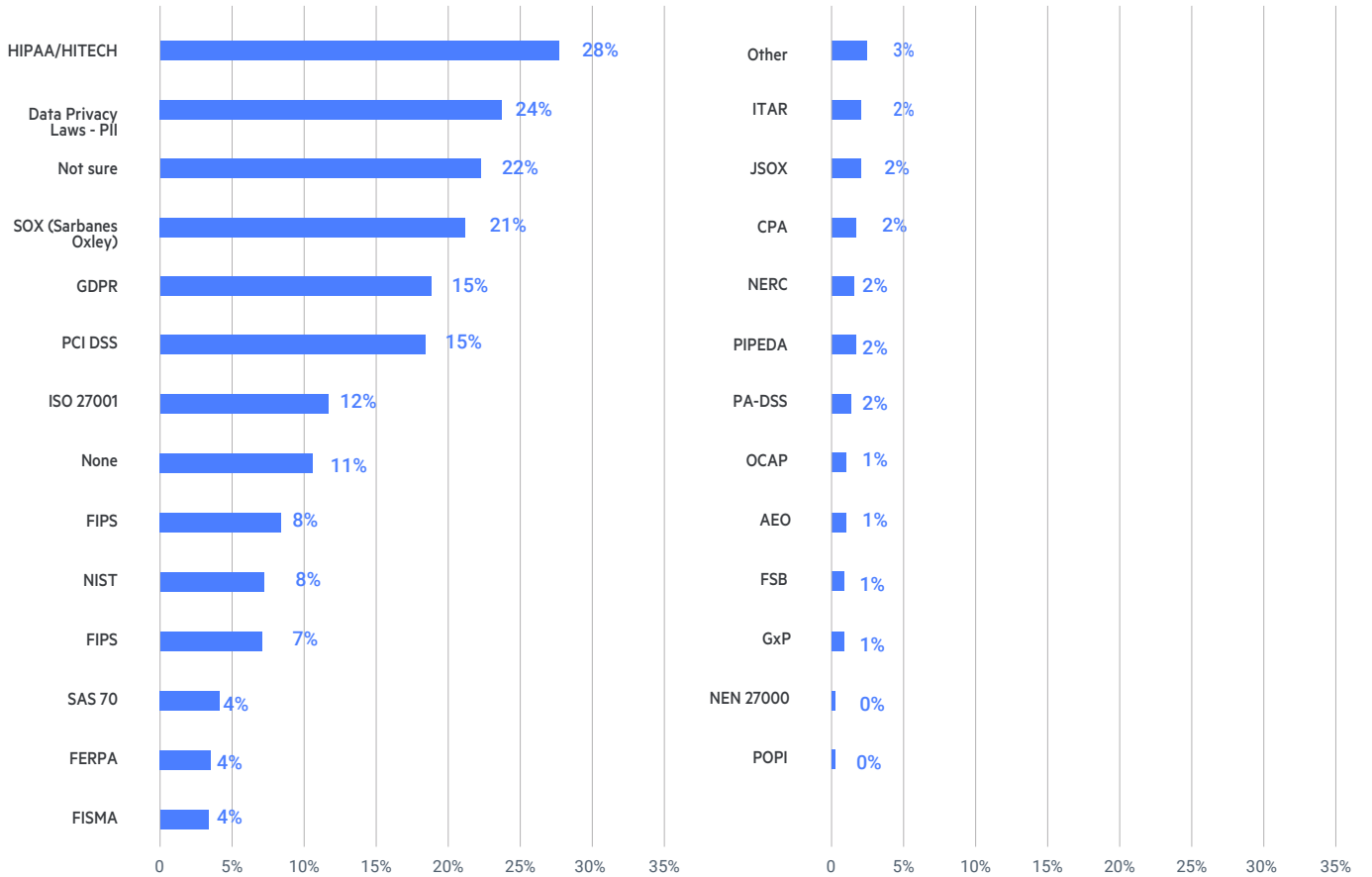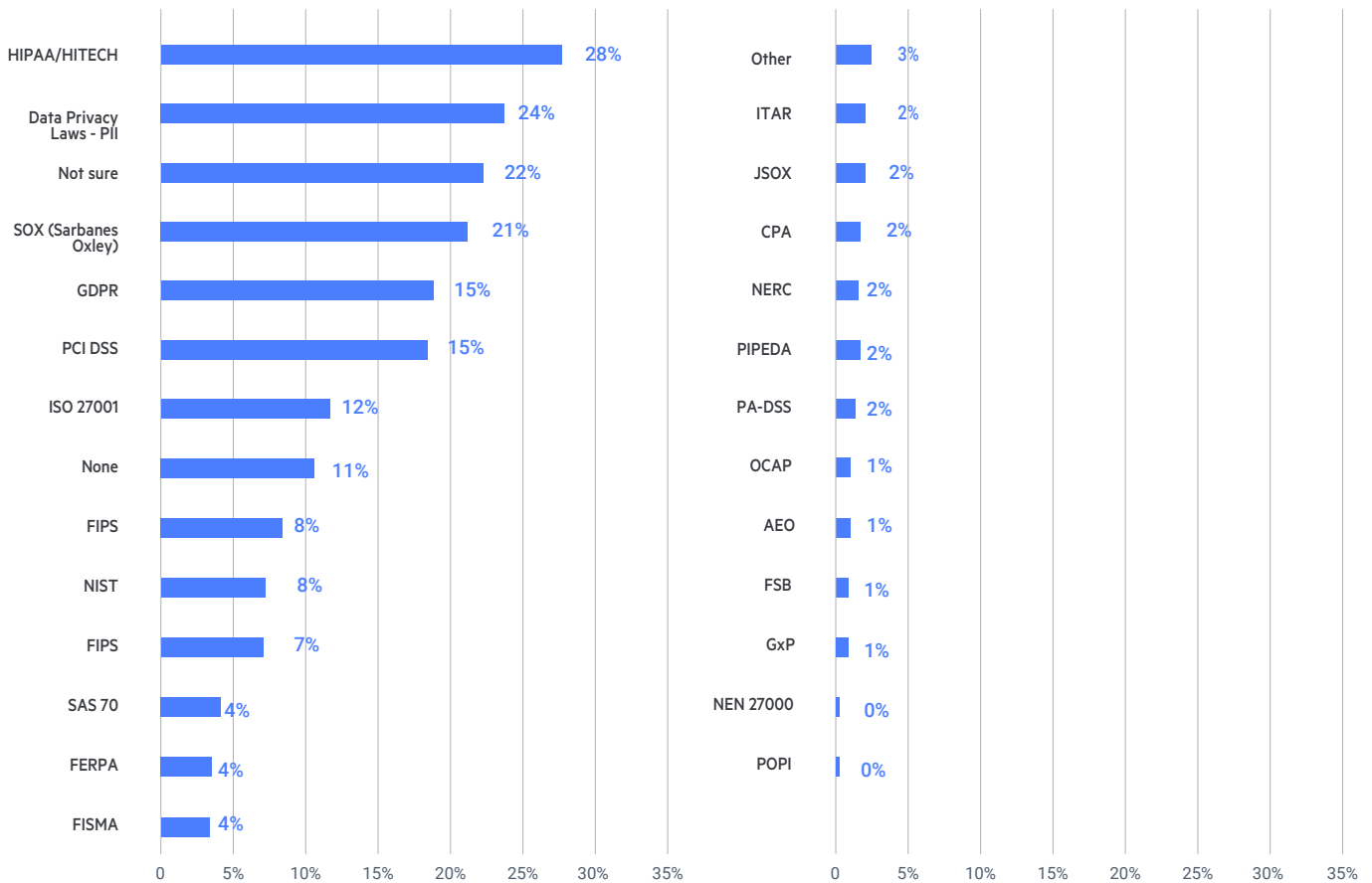Progress

North American regulatory expectations.

One interesting observation is the number of respondents in each region indicating that their organization has no regulations or that they are not sure (as high as 31% in APAC and 28% in South America)! With over 80 countries with privacy laws, sensitive data protection is a global phenomenon, and we expect that in the future the number of "None" and "Not Sure" responses will decrease.

8.3 Regulations by Industry

In this analysis, we look at the top five regulations for the industries of the survey respondents. For this, we look at industries where there were greater than 50 respondents in that industry.

In IT services, software, manufacturing and insurance, GDPR was the most frequently chosen selection. Not surprisingly, HIPAA/HITECH and PHI (Protected Health information) were most frequently cited in the healthcare industry, and Sarbanes-Oxley (SOX) was the most frequently cited concern in the financial industry. For those in the education industry, it is not surprising to see Data Privacy

Progress®

and FERPA among the top choices, and the appearance of HIPAA/HITECH might be attributable to the healthcare data management concerns for universities that are connected to hospitals and other health-related organizations.

8.4   Regulations and Cloud Adoption

In the past, there has been some reluctance to using cloud technology due to concerns that cloud environments are more likely to be breached or compromised by hackers. This is particularly the case when storing data that might contain sensitive information (such as personal/private individual data or otherwise sensitive financial data).

However, in this year's survey, fewer than 11% of the respondents indicated that their organizations had no use of public cloud infrastructure. This suggests that organizations are overcoming their earlier resistance to moving sensitive data to the cloud.

One might think that industries without a legal mandate for private data protection might be more inclined to move to the cloud. However, among respondents reporting their organizations have no regulatory requirements, only 32% indicated they use AWS and 36% indicated they use Azure. This might lead one to conclude that data protection concerns are not the only impediments to cloud adoption.

- For those respondents with regulatory requirements, an interesting pattern emerges:

- Of the respondents who indicated that they deal with payment information and as a result need to meet Payment Card Industry Data Security Standard (PCI DSS) requirements, 68.9% indicate they use AWS while 61.5% indicate they use Microsoft Azure. As a percentage, that is nearly twice those with no regularity requirements.

- Of the respondents indicating they deal with health data and need to comply with regulations associated with PHI (Protected Health Information) requirements, we see that 62.8% report using AWS and 62.8% also report using Microsoft Azure.

- Of the respondents who need to meet the data protection requirements of HIPAA and HITECH, 64.6% reported using AWS while 58.5% are using Microsoft Azure.

Progress®

- For the respondents needing to comply with GDPR, 56.4% use AWS and 57.1% use Azure.

- For FISMA respondents 55.8% use AWS and 58.1% use Azure.

- For those indicating that their organizations are subject to SAS70, 75.9% are AWS users and 60.3% use Azure.

- Of the respondents with NIST requirements, 69% use AWS and 64.3% use Azure.

These percentages seem to indicate that organizations with regulatory requirements are not only overcoming their prior concerns about the security and data protection capabilities of cloud vendors, they are aggressively migrating their data and applications to the cloud. One potential explanation is that these organizations have determined that a combination of improved data security and a wealth of benefits of the cloud (such as scalability, elasticity and only paying for the computing and storage resource they use) outweigh the perceived risks of exposure.

Connectivity is the Key

If there are any common themes in this year's survey, they would be modernization, cloud migration, data protection and data quality. As organizations are assembling their system modernization strategies, the boundaries around the enterprise information environment are dissolving. The on-premises data center protected by systemic perimeter security is rapidly transforming into a hybrid environment that spans on-premises, multiple cloud platforms and a collection of SaaS/PaaS services, and that hybrid architecture is bound to be the norm for the next decade, if not longer.

That being said, the breadth and scope of this extended information enterprise can only be enabled with a foundational approach to data and system connectivity. The more citizen analysts demand real-time access to a growing inventory of different types of enterprise (and extra-enterprise) assets that are distributed across different platforms, the greater the need for seamless data connectivity. The 2019 survey results bear this out—with growing cloud adoption, a mix of different operating systems and plans for modernization in concert with cloud migration, getting tripped up when it comes to data connectivity and data quality will impede any organizations data strategy.

Progress®

Ready to learn more?

About Progress

Progress (NASDAQ: PRGS) offers the leading platform for developing and deploying strategic business applications. We enable customers and partners to deliver modern, high-impact digital experiences with a fraction of the effort, time and cost. Progress offers powerful tools for easily building adaptive user experiences across any type of device or touchpoint, the flexibility of a cloud-native app dev platform to deliver modern apps, leading data connectivity technology, web content management, business rules, secure file transfer, network monitoring, plus award-winning machine learning that enables cognitive capabilities to be a part of any application. Over 1,700 independent software vendors, 100,000 enterprise customers, and two million developers rely on Progress to power their applications. Learn about Progress at www.progress.com or +1-800-477-6473.

Progress®