soda foundation

# 2021 Data and Storage Trends

October 2021

In Partnership With

THE LINUX FOUNDATION | Research

# Contents

# Foreword

The year was 2015 when a few of us from different companies in the storage industry came together to discuss collaborating on a software project. It was a small but significant moment for an industry notorious for being highly proprietary.

The Linux Foundation OpenSDS project was launched as a result of that discussion, with the mission to address storage integration challenges. With increased interest, the OpenSDS project expanded into the SODA Foundation last year to better meet the needs of a growing and more diverse community. SODA Foundation's goals were many, including building solutions for end-users, data management standardization, cross-project integration, technology roadmapping, and gaining deeper insight into the industry as a whole.

To accomplish these goals, we needed a comprehensive study of the current data and storage landscape, and the role of open source in it. In addition to covering existing and emerging data and storage technologies, we also wanted this study to reflect how they are applied in the data center, the cloud, and the edge. This report is the culmination of the 2021 Data and Storage Trends Survey conducted in partnership with the Linux Foundation Research team. We also invited the following communities to be our survey partners, expanding reach across different domains, platforms, industries, and regions. These partners included Cloud Native Computing Foundation, Storage Networking Industry Association, Open Infrastructure Foundation, Storage Performance Council, Japan Data Storage Forum, China Open Source Cloud League, and Mulan Open Source Community.

The report reveals the top data and storage challenges faced by users, storage use trends (tape is not dead!), vendor product strategies, cloud storage use cases, and the important priorities for enterprises in the next three years, and beyond.

Some findings in this report such as the rise of cloud native, and the ubiquity of software-defined storage, are in line with general trends.

An interesting but not unusual finding is how the majority of vendors are now prioritizing open source. Traditionally, every vendor builds its ecosystem to achieve maximum lock-in. The business rationale behind that is understandable, and at times necessary for security, interoperability, and other objectives. The proliferation of open source technologies such as Kubernetes tests the limits of proprietary ecosystems, forcing vendors to rethink and embrace open source as a business strategy. That does not necessarily mean vendors are opening their tightly guarded ecosystems. Instead, it means that they typically engage open source as a means to extend their ecosystems. Regardless, more vendor involvement in open source ultimately benefits end-users.

This brings us to how and where SODA brings value to vendor and end-user organizations. This report validated the focus areas and direction of the Foundation. At the same time, it shows there is much work to be done, and many opportunities for improvement along the way. We hope it will help guide business and technology leaders in their decision-making and strategic approaches.

Finally, I would like to thank the LF Research team for assisting in this crucial research, our survey partners, and everyone that participated in the survey.

**- Steven Tan**
Chair, *SODA Foundation*
VP & CTO Cloud Solution - Storage, *Futurewei*

**74%** of end-users
**run database workloads all the time.**

**61%** of end-users
**rate DPA as the most important capability**
over the next three years.

**64%** of end-users identified
**storage capacity as their #1 storage infrastructure challenge.**

**55%:**
The share of total mentions
**identifying public cloud services as the foundation**
for application development and deployment.

**10-100X:**
The increased amount of
**annual data growth**
by the top 9% of end-users compared to the majority of mainstream enterprises.

**60%:** The number of end-users that identify
**Software Defined Storage as a key technology**
used in their storage infrastructure.

**Cloud application storage is the #1 use case**
for cloud storage.

**Backup and recovery is the #2 cloud storage use case**
reported by **58%** of end-users.

**Performance is the #1 cloud native storage pain point**
as identified by **49%** of end-users.

# Introduction

The SODA Foundation is an open source project under the Linux Foundation that fosters an ecosystem of open source data management and storage software for data autonomy. SODA offers a neutral forum for cross-project collaboration and integration and provides end-users quality end-to-end solutions.

In April of 2021, the SODA Foundation, in partnership with Linux Foundation Research, launched a survey to understand evolving data and storage trends that was conducted in English, Chinese, and Japanese-speaking markets—in particular, to identify the current data and storage activities, reliance on cloud services, workloads, challenges, and strategies going forward in the era of Cloud Native, Edge, IoT, AI, and 5G.

The intention of this survey data is to guide end-users and vendors on important issues, and help them be better equipped to make decisions, improve their products, and to assist the SODA Foundation in establishing new technical directions.

# Adjusting to the Needs of a Digital Economy

The phrase digital transformation is a catchphrase that describes the IT activities that enterprises engage in to compete more effectively in a digital economy. The historical triggers for the digital economy included mobile technologies, the cloud, social media, analytics, and big data. More recently AI/ML, containerization, functions, open source, and advanced analytics are facilitating and accelerating the rate at which enterprises are developing support for the digital economy. The pandemic is yet another unseen and unwelcome force accelerating the changeover to a digital economy.

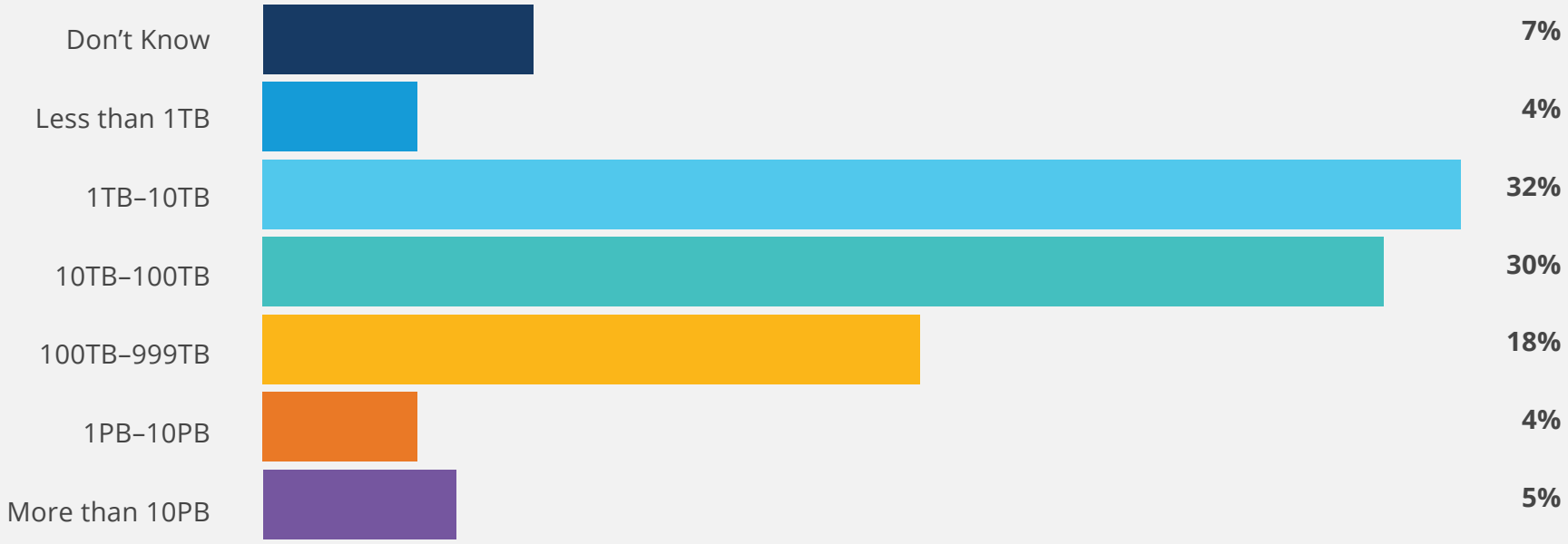
Becoming a data-driven enterprise is a necessity for enterprises looking to transition from being passive and reactive in their approach to markets and customers. Data is how enterprises will develop proactive and predictive methods to support their customers in an individualized way. To accomplish this, enterprises need data about pretty much everything: products, people, things, process, policy, behaviors, transactions, and state. Enterprises that used to think in terms of terabytes are now facing a new data management reality measured in petabytes. Attention to how data is collected, managed, and consumed is becoming the new competitive battlefield in IT. Consequently, all enterprises are on a data and storage journey that will impact how successful they are able to compete in a digital economy.

# Storage Requirements: From Terabytes to Petabytes

**Figure 1: Approximate Annual Data Growth for End-user Enterprises**

## How much is the approximate data growth per year for your organization?

**Sample Size = 97**



Data growth is a consequence of digital transformation and the rapid acceleration to supporting a digital economy. The result is that enterprises are becoming data-driven. Annual data growth is accelerating and this annual view into data growth is expected to increase and not stay constant.

Figure 1 shows data growth for end-user enterprises. Because the scale of the Y axis is non-linear, the average data growth by segment is primarily influenced by those metrics that are defined by petabytes (PBs).

Mainstream annual data growth is between 1-100 TBs per year, as reported by 62% of the sample. However, 9% of the sample is seeing annual data growth of 1PB or more. This is ten to 100 times greater growth than the mainstream and is likely a harbinger of where many enterprises will find themselves within a few years.

Due to the 9% of the sample that is seeing data growth of more than 1PB, the average data growth across all end-users is just over 1PB. This is an astounding finding and demonstrates not only the disparity across enterprises in their digital readiness, but also the rapid data growth being seen in the industry.

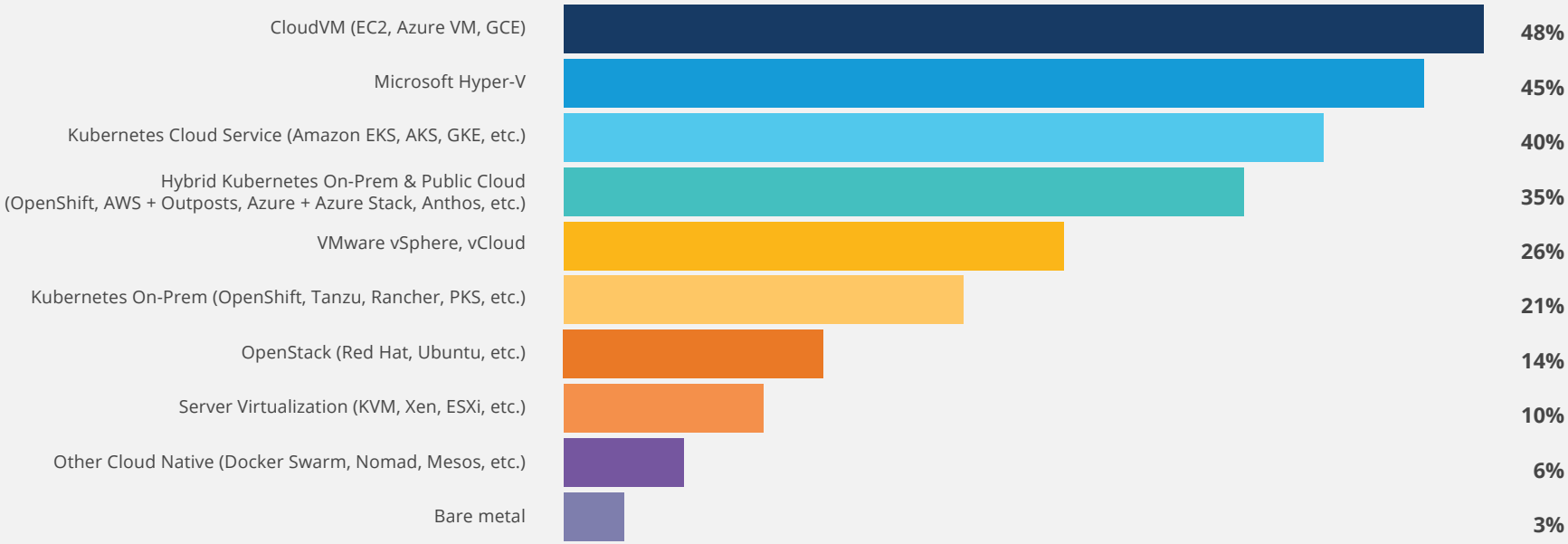
However, we expect overall annual data growth to rapidly increase because of IoT smart device data growth at the edge. The result is that enterprises should plan for data growth well beyond simple projections based on history.

# Enterprises Pivot to Cloud-based Containerized Environments

**Figure 2: Platforms in Use for Development or Production**

## Which platforms do you currently use in development or production?

**Sample Size = 96, Valid Cases = 96, Total Mentions = 238**



Over the last five years there has been a rapid transition from virtual machines (VMs) to containers. The advantages of containers, which include immutability, scalability, granularity, portability, and operation consistency, are compelling. A quick inventory of responses in Figure 2 shows an emphasis on the use of cloud resources and container-based tooling and orchestration. Because responses are sorted in descending order based on responses from the entire sample, container orchestration is performing strongly, although conventional VM-based technologies are still widely present.

The allure of the public cloud is its ability to spin up VMs or containers on demand and provide industry-leading tools like Kubernetes to automate, orchestrate, scale, and manage enterprise applications. The top three responses in Figure 2, which focus on public cloud capabilities, account for 55% of total mentions. This demonstrates the growing popularity of public cloud services in the agility that they provide to customers.

Hybrid cloud solutions, such as Red Hat/OpenShift, AWS/Outposts, Azure/Azure Stack, and Google/Anthos, showed usage across 35% of end-user enterprises. This is consistent with Kubernetes On-Prem, which shows a 21% penetration.

Together, this demonstrates the utility provided by cloud service providers and their efforts to develop tools to deliver seamless hybrid environments.

Bare metal solutions, at 3%, have yet to make inroads into end-user environments. While there can be a distinct performance advantage of bare metal solutions, there is a flexibility tradeoff that appears to users questioning the value proposition.

# Workloads: It's All About the Data

**Figure 3: The Importance of Data in Enterprise Workloads**

## What workloads do you run?

**Sample Size = 53–98**

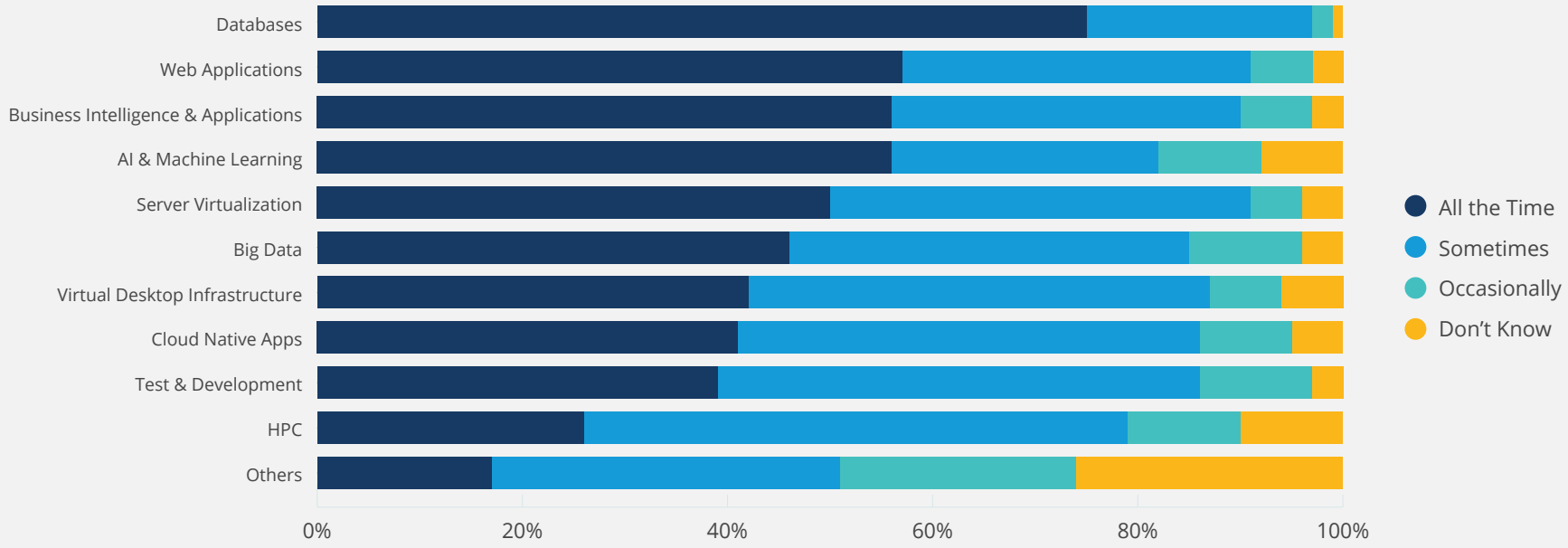


Figure 3, which looks at key workloads being run, reveals the importance and focus that end-user enterprises have on data. The 11 categories listed include four that are data-specific: databases, business intelligence, AI and machine learning, and big data. Based on total mentions, these data-centric workloads account for a 48% share of those workloads that run all the time. This is a significant commitment to data, despite not including the data-focused activities of web applications, cloud native applications, test and development, and high-performance computing.

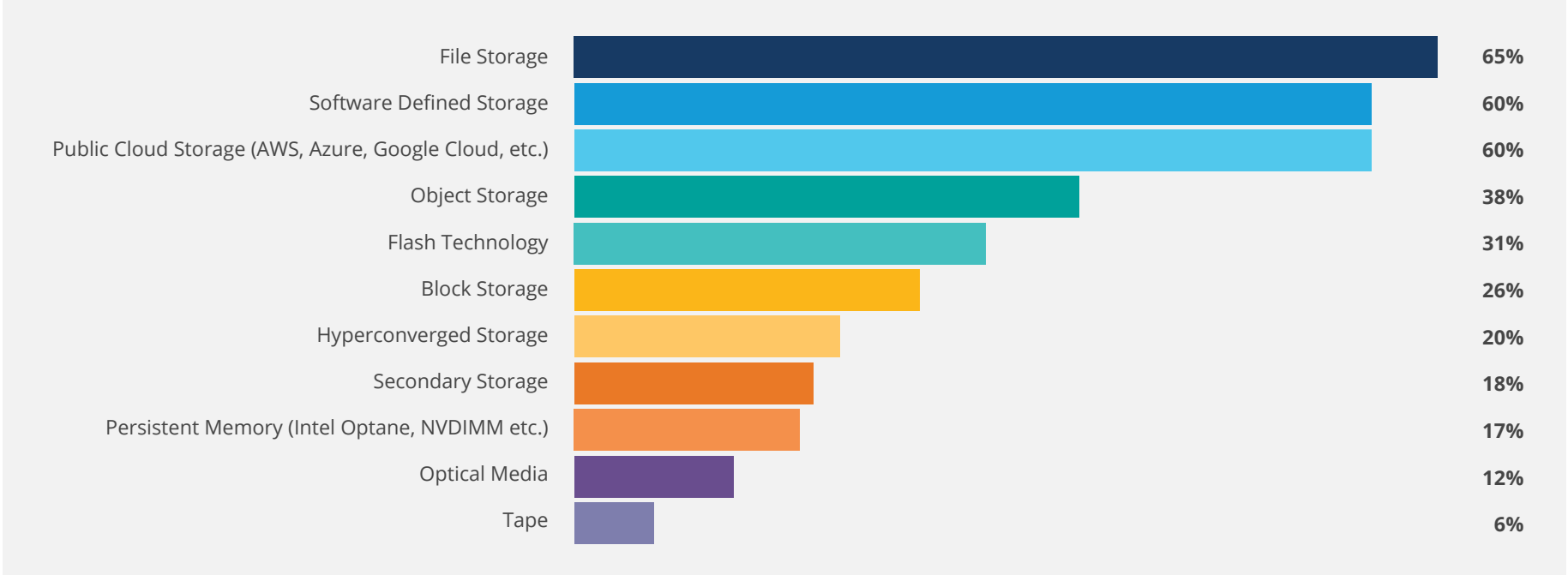
There was no surprise that the database workload, with 75% running workloads all the time, stands head and shoulders above all other workloads. While the volume of structured data continues to grow and drive added instances of SQL databases, prior research shows a growing affinity for NoSQL databases and the wide variety of use cases that they support. BI workloads (56% running workloads all the time) are expanding their scope as well, driven by advanced analytics, decisioning, and optimization. BI and data warehousing also cross over into big data, where 46% of enterprises are running these workloads all the time. Finally, AI/ML now represents a significant workload, with 56% of enterprises running workloads all the time. AI/ML and big data workloads also mean a significant growth in unstructured and semi-structured data growth.

# Software Defined Storage Could Revolutionize How We Manage Data

**Figure 4: End-user Enterprise Storage Technology Use Cases**

## What storage technologies are included in your infrastructure?

**Sample Size = 98, Valid Cases = 98, Total Mentions = 347**



When asked what storage technologies are included in the enterprise's storage infrastructure, three technologies shared a leadership position. File Storage (65%), Software Defined Storage (60%), and Public Cloud Storage (60%) are the storage technologies of choice for most enterprises, as shown in Figure 4. File storage is ubiquitous in IT, and its semi-structured approach to managing data provides a high degree of value because of its flexibility while delivering acceptable capacity, bandwidth, and latency. The same can largely be said for public cloud storage, with the Cloud Service Provider (CSP) managing availability, access, security, and performance across a wide variety of data storage form factors. The high penetration of file storage and public cloud storage reflects the importance of capabilities to address basic IT data management functions.

Software defined storage (SDS) could be a next-generation game changer in data and storage management. SDS provides a level of indirection between applications and data storage devices. SDS capabilities could include decision-making regarding how data is managed, directing data protection activities, and optimizing for cost and performance across multiple CSPs. SDS is also a great candidate for the application of ML and AI as a way to tune storage behavior.

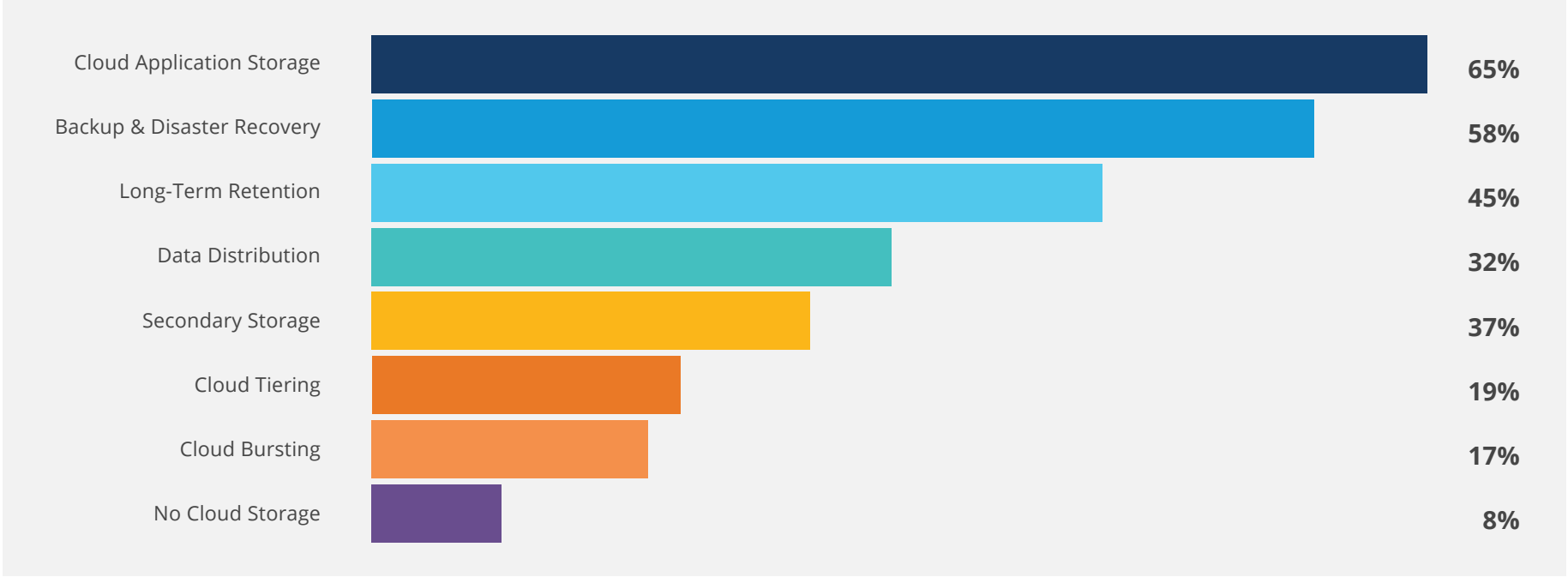
Somewhat surprising is the modest showing of hyperconverged storage, at 20%, and persistent memory, at 17%. These technologies offer significant potential, but without the benefit of longitudinal data, it's difficult to understand how fast these technologies are growing. This is an area that needs more research.

# Cloud Storage Use Cases Reflect Operational Imperatives

**Figure 5: Cloud Storage Use Cases**

## How do you use cloud storage?

**Sample Size = 98, Valid Cases = 98, Total Mentions = 266**



The leading use cases in cloud storage for end-user enterprises include application storage, disaster recovery (DR), and long-term retention. Figure 5 provides a list of eight cloud storage use cases, but only those use cases with a strong operational focus drove high penetration.

The leading use case identified by 65% of end-user enterprises was cloud application storage. Applications and the data they leverage and create together define the digital persona of the enterprise. Operational use cases such as DR, HA, and long-term retention are all dependent on the ongoing operation of applications. Therefore, it's not surprising to see cloud application storage in a leadership role.

Backup and disaster recovery was a leading use case and supported by 58% of end-user enterprises. Backup and DR facilitate continuing operations in the event of an outage, ransomware attack, or natural disaster. Also recognize that this is backup and DR in the cloud—which undoubtedly represents a significant change for enterprises managing this on premise.

Long-term retention (LTR), a use case supported by 45% of end-user enterprises, addresses audit and GRC concerns. LTR policies are often used to snapshot databases to ensure a database can be restored to any point within an N-day timeframe.

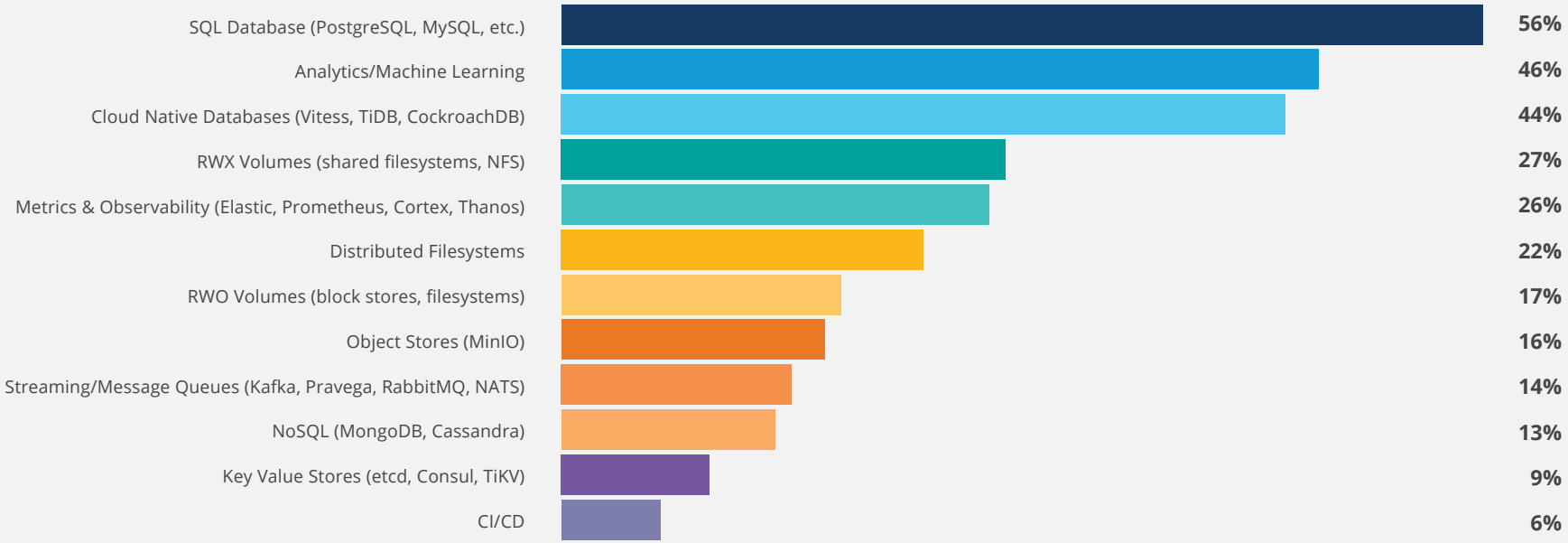
All three of these activities and use cases are considered mission-critical, which accounts for the priority that end-user enterprises attach to them.

# Container and Cloud Native Workloads Also Focus on Data

**Figure 6: Workloads Deployed Using Containers & Cloud Native Storage**

## What workloads or use cases do you deploy using container/cloud native storage?

**Sample Size = 98, Valid Cases = 98, Total Mentions = 291**



Cloud native means support for containers, dynamic orchestration of resources, microservices, and special attention to managing persistence and state. The highly distributed nature of microservice applications, especially those that need to scale, present management problems that can also scale. Special attention must be paid to observability, policy, orchestration, and automation.

We already know that data-centric workloads (database, BI, AI/ML, and big data) are frequently running all the time in end-user enterprises. When we asked specifically about container and cloud native workloads, we observed a similar pattern. Figure 6 lists a variety of workloads running in containers and using cloud native storage.

Figure 6 shows that SQL databases are the most frequently encountered workload across 56% of end-user enterprises surveyed. The ACID properties of SQL make it ideal for managing transactions and managing systems of record.

However, containers and cloud native storage have other qualities appealing to data-centric workloads. Analytics and machine learning, at 46%, was the second most common workload. These workloads can at times be compute and data-intensive, but this intensity can vary wildly. Containerized workloads are lighter weight than VMs and can be instantiated or destroyed much faster. The dynamic scalability of cloud native environments makes them a good match for intensive yet unpredictable workloads.

Cloud native databases, at 44%, are often described as NewSQL databases and are effective at supporting hybrid transactional and analytical processing workloads. These databases offer SQL-like capabilities while also retaining some of the best characteristics of NoSQL databases, which include High Availability (HA) and massive scalability. Consequently, NewSQL databases are a similarly good fit for containerized cloud native environments.

Regardless of database type (SQL, NoSQL, or NewSQL), there still remains a need for databases that are cloud native. The reason for this is that managing stateful applications in Kubernetes can still be challenging. Consequently, Distributed SQL databases have arrived to fill this niche, but it is also likely that every database vendor is currently also working on this problem/opportunity.

# Containerized Cloud Native Storage Is Not Without Pain Points

**Figure 7: Cloud Native Storage Pain Points**

## What are the most prevalent pain points when using container/cloud native storage?

**Sample Size = 98, Valid Cases = 98, Total Mentions = 271**

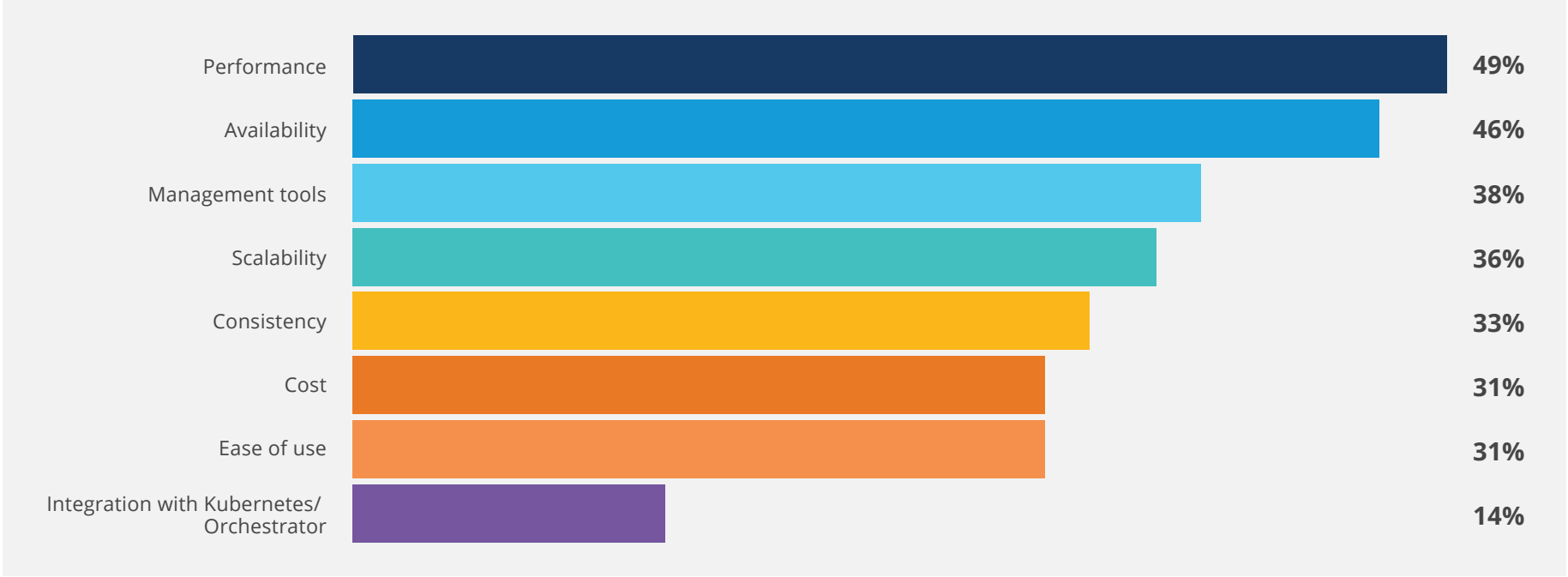


Figure 7 lists most prevalent pain points when using containers or cloud native storage. Performance was identified by 49% of end-user enterprises, as the number one pain point when using containers and/or cloud native storage. Performance was followed closely by availability, at 46%, as the second most common pain point.

Given the strong focus on SQL workloads as shown back in Figure 6, this is not surprising. SQL databases are known for becoming fragmented, needing periodic reorganizations, and balancing indexing with read/write performance and availability. The schema for NoSQL databases is designed around the query, virtually the opposite of SQL databases. This makes the transition from SQL to NoSQL difficult for many DBAs, and it is often said that NoSQL database design makes it easy for the user to break into jail. The results of not paying particular attention to NoSQL database design are performance problems that are hard, if not impossible, to resolve.

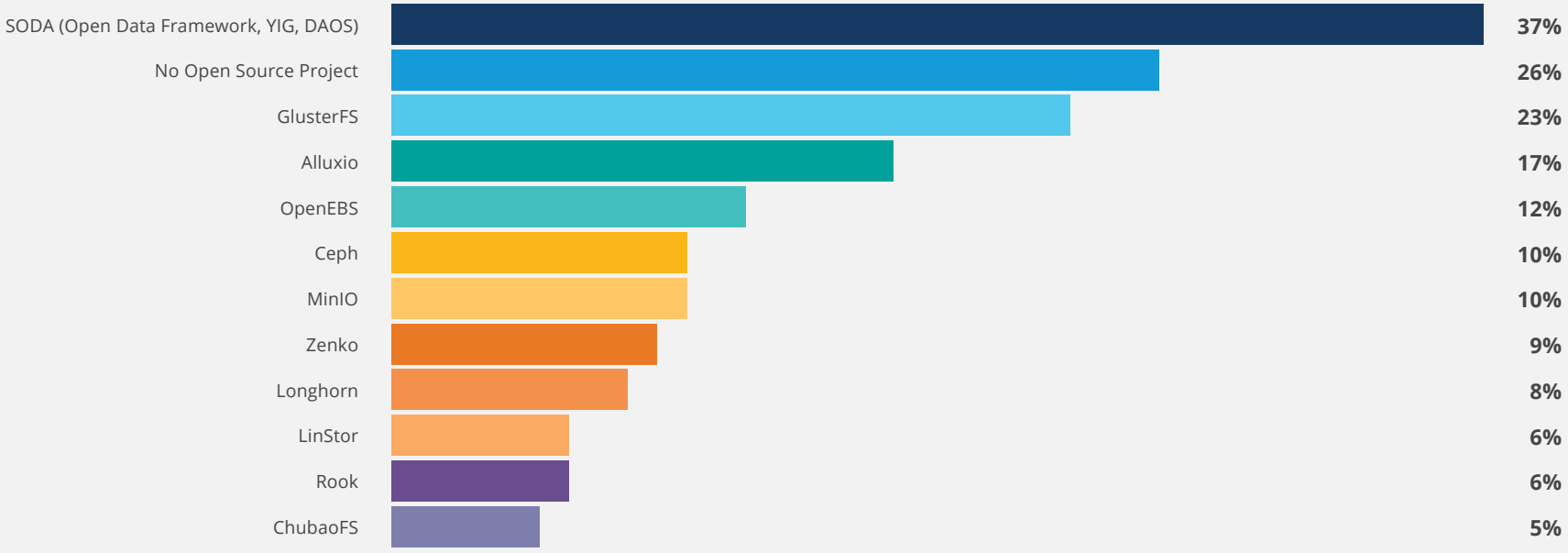
Management tools at 38% and scalability at 36% speak to the importance of intelligent, scalable, and data-driven management capabilities to address cloud native operations. While Kubernetes has become the standard for orchestration tooling to support observability, IT Operations Management (ITOM), AIOps, application performance management (APM), risk management, application security, data protection, and value stream management (VSM) all offer compelling capabilities for optimizing operations.

# Open Source Projects Are Shaping Data and Storage Infrastructure Use

**Figure 8: Open Source Projects in Use or Planned for Use in Data and Storage Infrastructure**

## Which open source project(s) are you currently using or planning to use in your data and storage infrastructure?

**Sample Size = 98, Valid Cases = 98, Total Mentions = 167**



Open source is an engine for innovation in most markets, including those driving data and storage infrastructure. The SODA Foundation, an open source project under the Linux Foundation, is seeking to establish an open, unified, and autonomous data management framework for data mobility from the edge, to the core, and to the cloud. The SODA Foundation is an incubator for data in storage infrastructure projects and is currently focused on data mobility, data protection, data security, and data analytics across cloud native, edge computing, IoT, big data, and ML technology stacks.

Figure 8 lists a number of established open source projects focused on data and storage infrastructure and shows their impact on end-user enterprises.

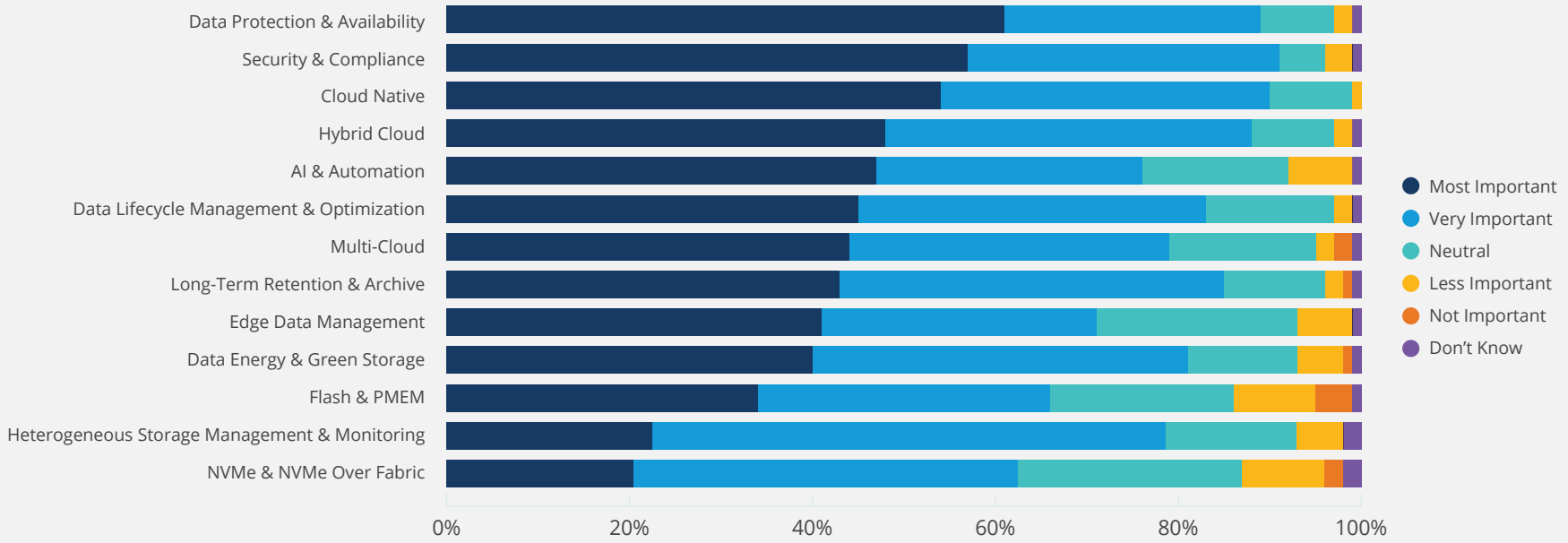
It is encouraging to see that 37% of end-user enterprises are keen to collaborate more with SODA-sponsored projects. This indicates that, SODA has an opportunity to play significant role in future data management deployments for end users. While it is disappointing to see that 26% of enduser enterprises were not engaged in any of the open source projects listed, this is by no means an exhaustive list. Based on the numbers, the 74% of enterprises who are involved in these open source projects participate on average in two to three projects simultaneously.

# Enterprises are Fixated on Data Protection and Security

**Figure 9: Capabilities and Technologies that Will Shape Data Management and Storage Plans**

## Rate the important of the following in planning your data management & storage for the next 3 years

**Sample Size = 98**



When enterprises were asked what capabilities and technologies are most important in shaping their data management and storage plans for the next three years, the results were encouraging. Figure 9 rates the importance of 13 specific capabilities or technologies specific to data management storage.

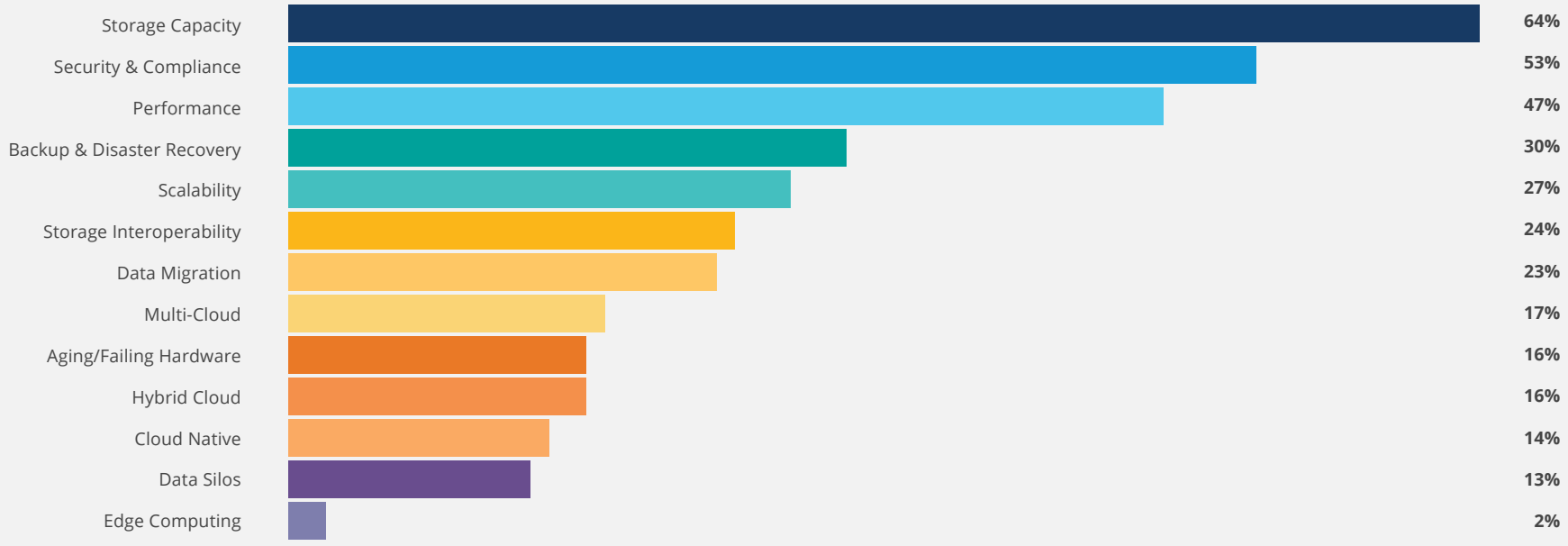
Data protection and availability, at 61%, as well as security and compliance, at 57%, were the top two capabilities identified by end-users. To be fair, it is surprising that data protection and security weren't identified as most important by even more end-users. The reason for this is that data protection and availability, as well as security and compliance, are capabilities that protect the enterprise against existential threats. The remaining capabilities and technologies listed here can be effective at improving operational performance, but do not approach the significance of both data protection and security.

# Storage Capacity and Security Are the Top Data and Storage Infrastructure Challenges

**Figure 10: Top Three Data and Storage Infrastructure Challenges**

## What are the top 3 major challenges in your current data & storage infrastructure? Please choose 3.

**Sample Size = 98, Valid Cases = 98, Total Mentions = 340**



The challenges that accompany data management and storage infrastructure grow in intensity and significance as data volumes grow. The complexity of managing data does not scale linearly, and the type of solution may need to change as volumes increase. This explains why Figure 10 identifies storage capacity as the most pressing challenge, as identified by 64% of end-users. In cases where organizations are nervous about the long-term ability of their current closed solutions to effectively address their storage capacity needs, involvement in corresponding open source projects may surface additional paths forward.

We know from Figure 9 that users are very concerned about security and compliance. While it is important that security and compliance factor heavily into data in storage planning, it is disconcerting that security and compliance are also being described as a major challenge. This suggests that the path forward in security and compliance is not well defined, and will be challenging for end-users to address.

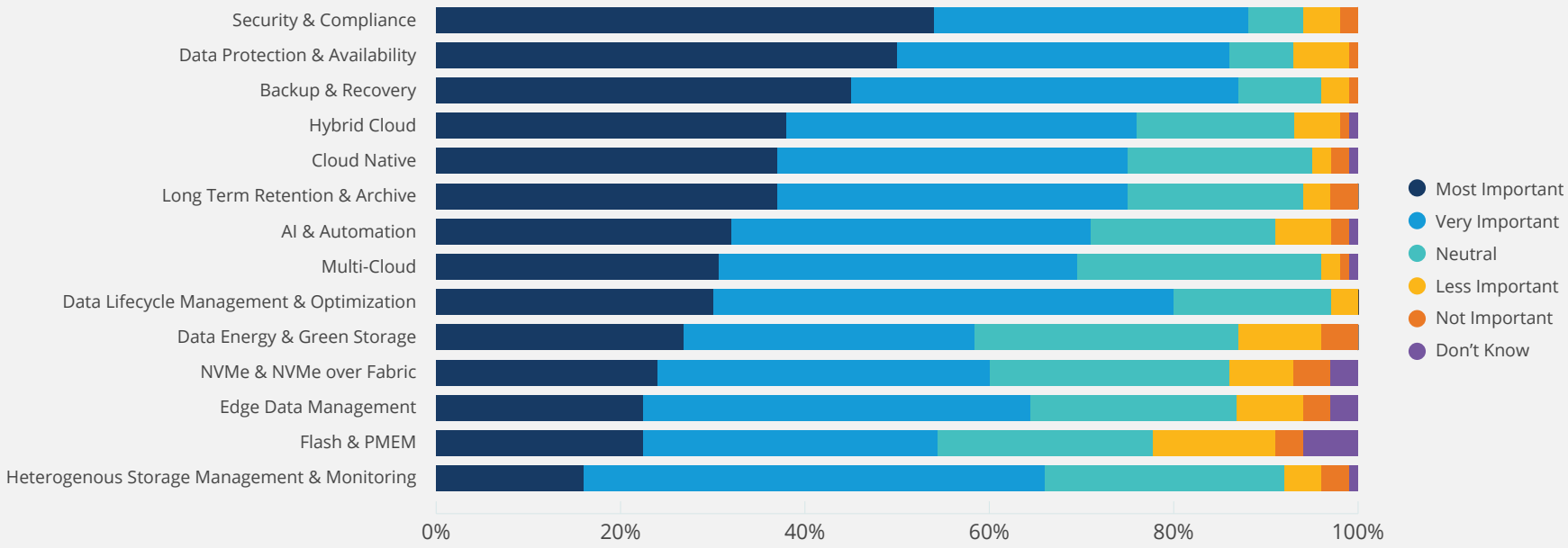
Performance, at 47%, is the third-ranked challenge. Performance management is an extensive topic and has always been challenging to address. Increasing data volumes will bring more complex objects, data models, and schemas. The challenges in optimizing the performance of SQL, NewSQL, and NoSQL databases are complex. Additionally, the challenge of linear scalability and performance on the journey from terabytes to petabytes is still a concern.

# Vendor Priorities in Planning New Data and Storage Product Capabilities Align with End-user Needs

**Figure 11: Vendor Priorities in Planning New Data and Storage Products**

## Rate the importance of the following in planning your data & storage products/solutions/ services for the next 3 years

**Sample Size = 98**



This survey included two questions that were specifically focused on vendors and service providers. The first, shown in Figure 11, asks these vendors to rate the importance of various capabilities and technologies. The assumption is that higher-rated capabilities and technologies provide more end-user value and will find their way into products, solutions, and services more quickly. Figure 11 shows a series of capabilities and technologies ranked from most important to least important by vendors.

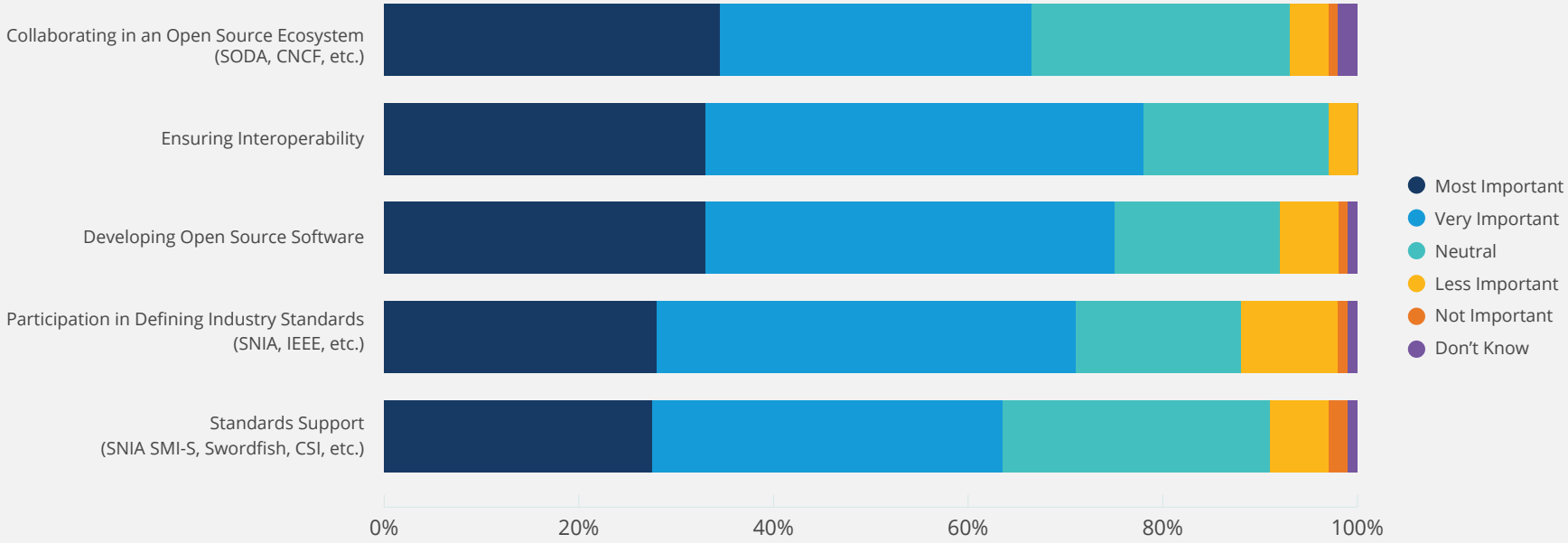
It is reassuring to see a strong alignment between end-user needs and capabilities perceived as important by vendors. Figure 9 shows that data protection and availability, as well as security and compliance for the leading capabilities, are shaping end-user plans for the next 3 years. Figure 11 shows that security and compliance, as well as data protection and availability, are likely also shaping vendor product plans. This is a key finding, especially because of the concern that end-users have regarding how to address security and compliance requirements, shown in Figure 10.

# Collaborating with Open Source Ecosystems Is a Vendor Priority

**Figure 12: The Importance of Ecosystems and IT Industry Organizations to Vendors**

## Rate the importance of the following in your product strategy

**Sample Size = 98**



Open source software provides unique opportunities to collaborate down the stack as a way of facilitating more extensive competition at the stack where the value is greater to users. A second question that we asked vendors and service providers was to rate the importance of IT industry organizations and issues that these organizations are particularly adept at solving. Figure 12 tests for the importance to vendors of various ecosystems and industry organizations.

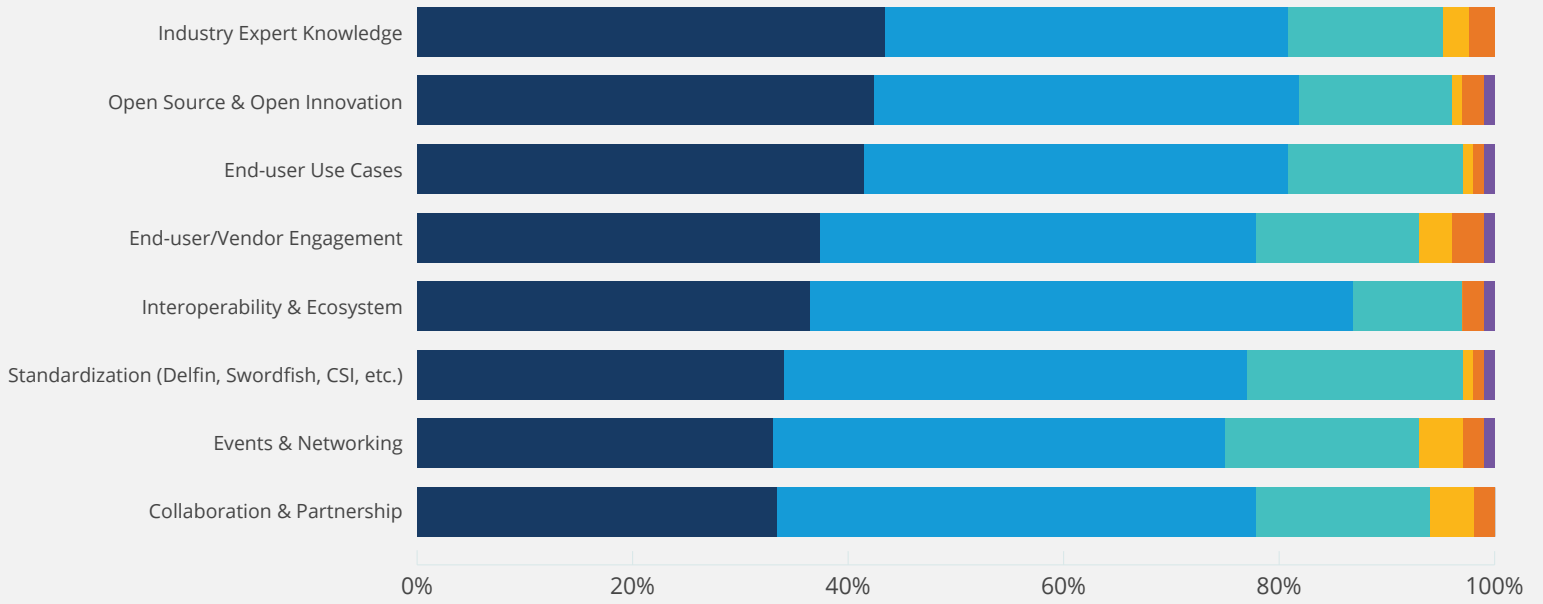
Scores across the organizational capabilities listed were uniform. While no more than one third of vendors identified any capability as most important, two thirds of the vendors in this sample ranked every capability as either most important or very important. This suggests that the majority of vendors are on board with the role that open source foundations, communities, and industry organizations provide, and value their contribution as an input to their product strategy.

# How SODA Brings Value to Your Organization

**Figure 13: How SODA Brings Value to Your Organization**

## Rate how SODA can bring value to you and your organization.

**Sample Size = 203**



The SODA Foundation is an open source project under the Linux Foundation that aims to establish an open, unified, and autonomous data management framework for data mobility from the edge, to core, to cloud. SODA brings together industry leaders to collaborate on building a common framework to promote standardization and best practices for data storage, data protection, data governance, data analytics, etc. to support IoT, big data, machine learning, and other applications.

SODA Foundation capabilities are shown in Figure 13. The responses included in Figure 13 include all of the end-user enterprises, vendors and services providers, and IT industry organizations surveyed.

Figure 13 is immensely encouraging because more than 75% of the sample embraced every one of the capabilities provided by SODA as being either most important or very important. This depth of support for the SODA Foundation speaks volumes about the focus, quality, and utility of the projects sponsored by SODA.

# Where SODA Can Bring Value to Your Organization

**Figure 14: Rate the Importance of Potential SODA Development Work**

## Rate which SODA development work would benefit you.

**Sample size = 203**

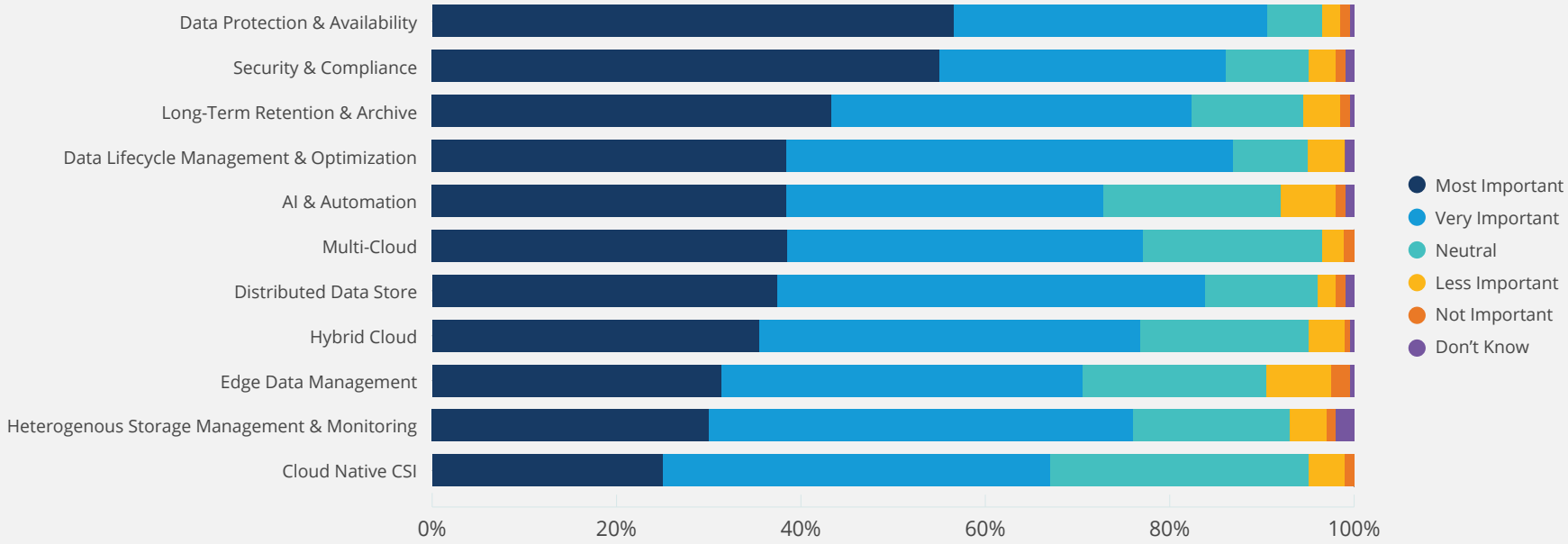


Figure 14 provides end-users, vendors, and IT Industry Organizations an opportunity to identify where SODA development work could benefit them the most. The results were predictable.

The strong affinity that end-users and vendors have for data protection and availability, as well as security and compliance, are the most important areas where potential SODA development work would add value. The scores cited as most important in these two areas are also significantly different than any of the other candidate areas. Finally, more than 85% of this sample see SODA Foundation involvement in data protection and availability, as well as security and compliance, as a most important or very important activity.

# Methodology and Demographics

From April 15 to August 19, 2021, SODA and the Linux Foundation fielded a worldwide survey of individuals at organizations on a range of questions related to trends and concerns about their data and storage environments. Employees in a range of different roles, including CxOs, developers, data & analytics professionals, enterprise architects, as well as R&D and product development were surveyed. Small, medium, and large enterprises were surveyed along with a cross section of end-user enterprises, vendors, and IT service providers, and academia/IT industry organizations.

The survey was promoted via social media, the Linux Foundation and Linux.com websites, the Linux Foundation Newsletter, and with the support of the following partners:

- Cloud Native Computing Foundation (CNCF)
- Storage Networking Industry Association (SNIA)
- Open Infrastructure Foundation (OIF)
- Japan Data Storage Forum (JDSF)
- China Open Source Cloud League (COSCL)
- Mulan Open Source Community
- Storage Performance Council (SPC)

A total of 247 respondents completed the survey. 27% of the respondents were sourced from the Linux Foundation members and 73% of the sample was acquired from a reputable panel vendor. The strong focus on collecting data from end-users and vendors with no a priori relationship to the Linux Foundation was an important way to ensure objectivity in the responses and avoid any LF sourcing bias.

Worldwide, 40% of the sample originated from North America, 25% from Europe, 17% from Latin America, 9% from India, 4% from Japan, 2% from China, and 3% from the rest of world (Middle East, Africa, Latin America, and Australia/Oceania).

Respondents came from organizations of all sizes including 15% from very small organizations (1-99 employees), 36% from small organizations (100-999 employees), 26% from medium organizations (1,000-4,999 employees), and 24% from large organizations (5,000 or more employees).

The distribution by role included 15% application developers, 10% R&D/product development, 21% data & analytics, 19% enterprise architects, 9% product management/marketing, 5% consultants, 13% sales/marketing staff, and 8% system integrators and CxOs.

The distribution of responses by organization type included 40% end-users (primarily consumers of IT products and services), 40% IT vendors and services

providers (Vendor/SvcP), and 20% members of standards organizations, open source organizations, and academia (Std/OSS/Ed). Given the diverse nature of these three constituencies, nearly all of the data in this report will be segmented and include a distribution for each of these groups as well as a total across all three groups. Where responses do not warrant an implicit order, we have sorted the responses in descending order based on total responses.

The analysis in this report generally focused on the finding for end-users and IT vendors/service providers. We frequently omit the responses of standards organization, open source organizations, and academia. The reason for this is because end-user organizations are the primarily consumers of IT products and services, and vendors/service providers are the producers of IT products and services.

For detailed information on demographics segmented by all organization types, please refer to Appendix A.

# Appendix A: Demographics

**Figure 15: Respondent Role Segmented by Organization Type**

## How would you best describe your role?

**Sample Size = 247**

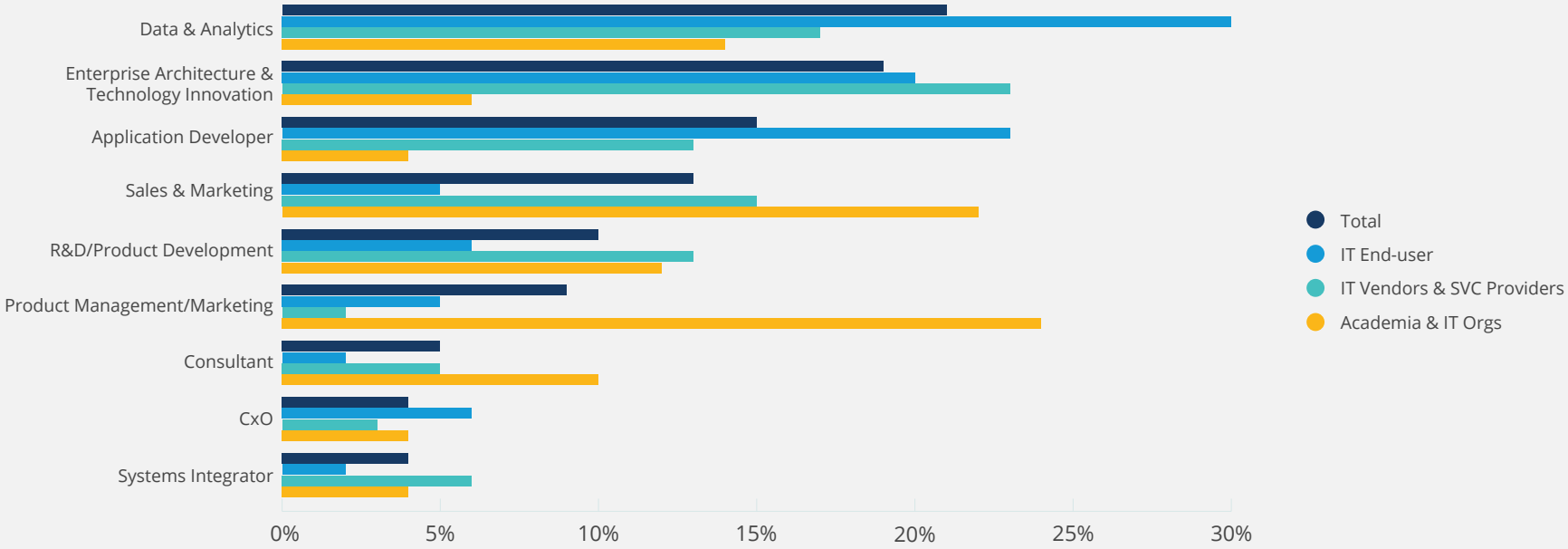


**Figure 16: Country/Region Segmented by Organization Type**

## Country/region?

**Sample Size = 246**

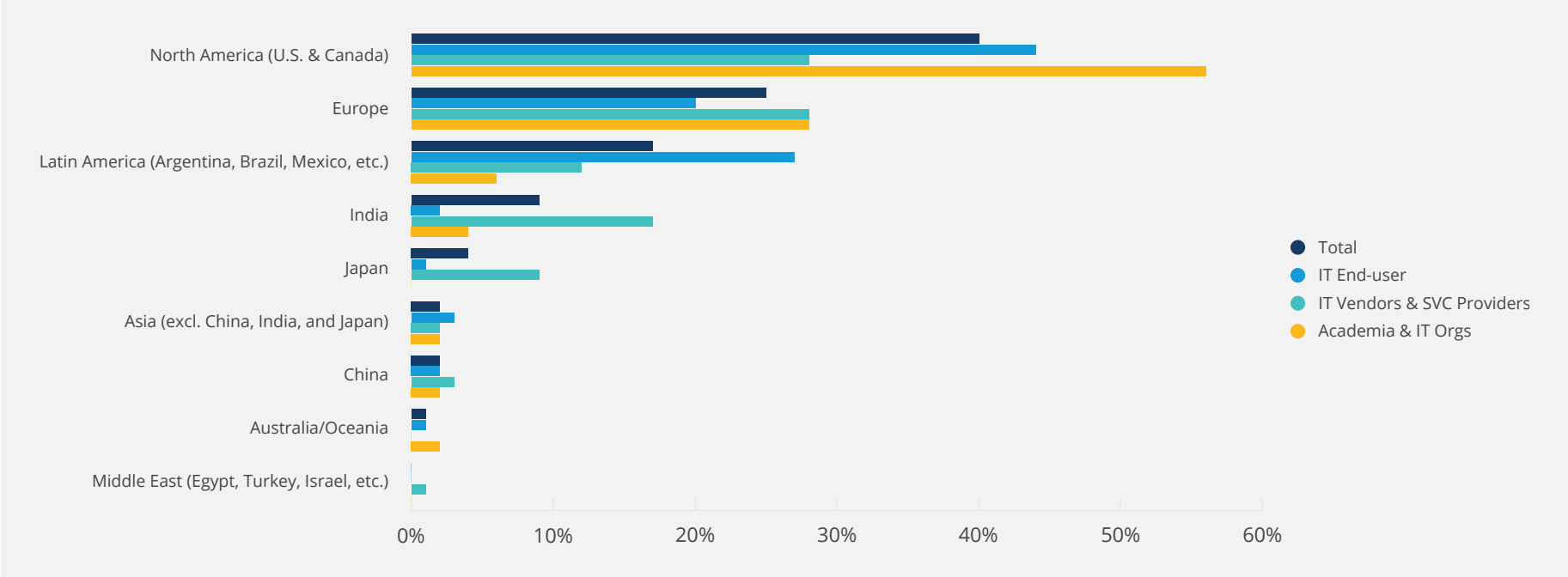


**Figure 17: Organization Size (Number of Employees) Segmented by Organization Type**

## What is the size of your organization (employees)?

**Sample Size = 247**

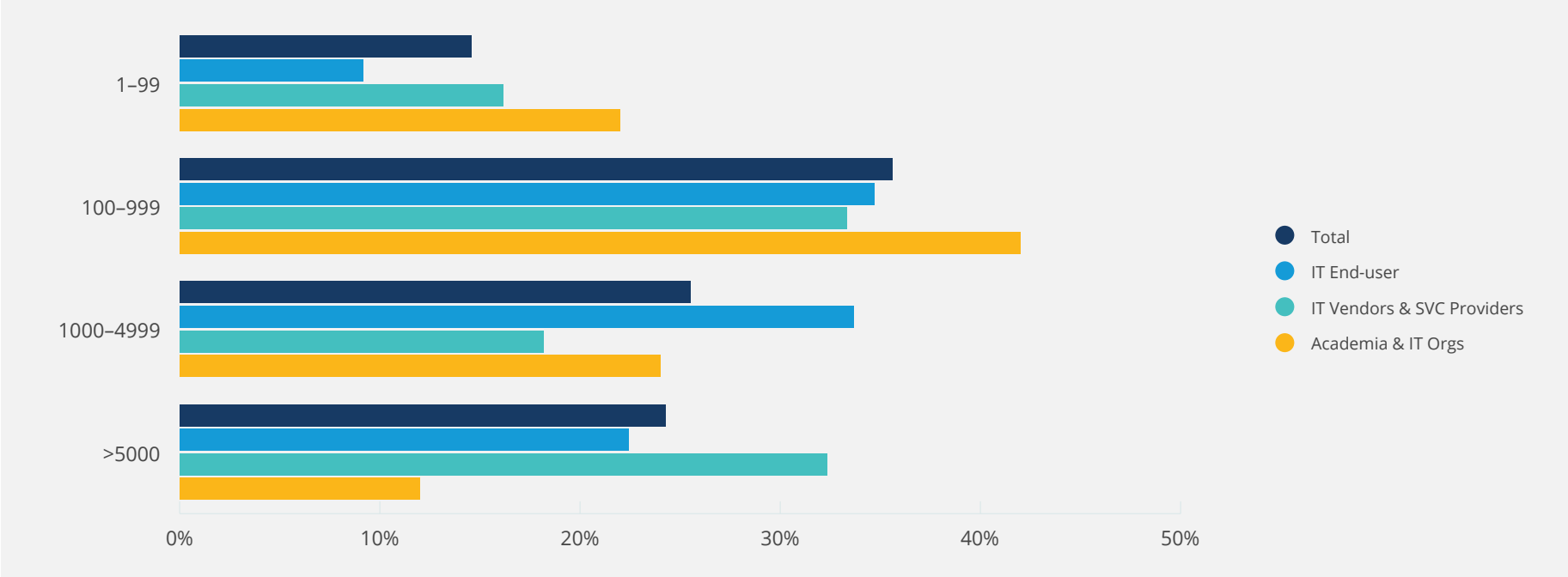


**Figure 18: Industry Segmented by Organization Type**

## Which industry does your organization belong to?

**Sample Size = 247**

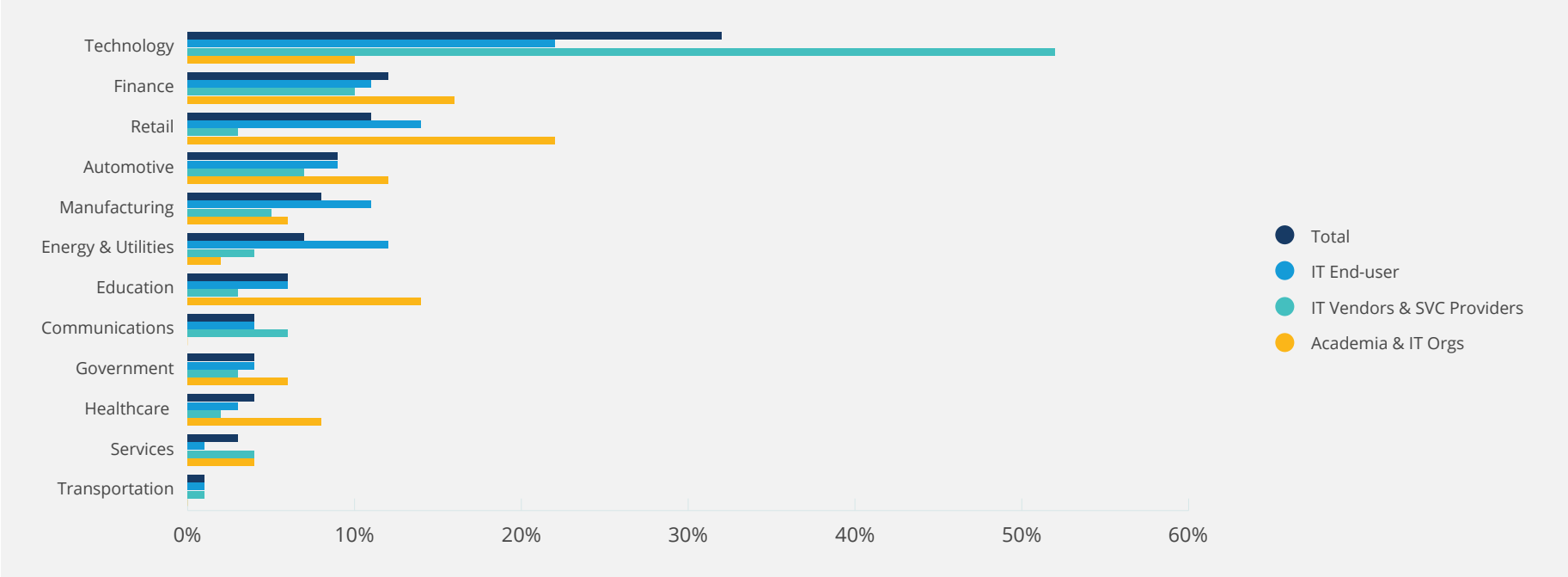
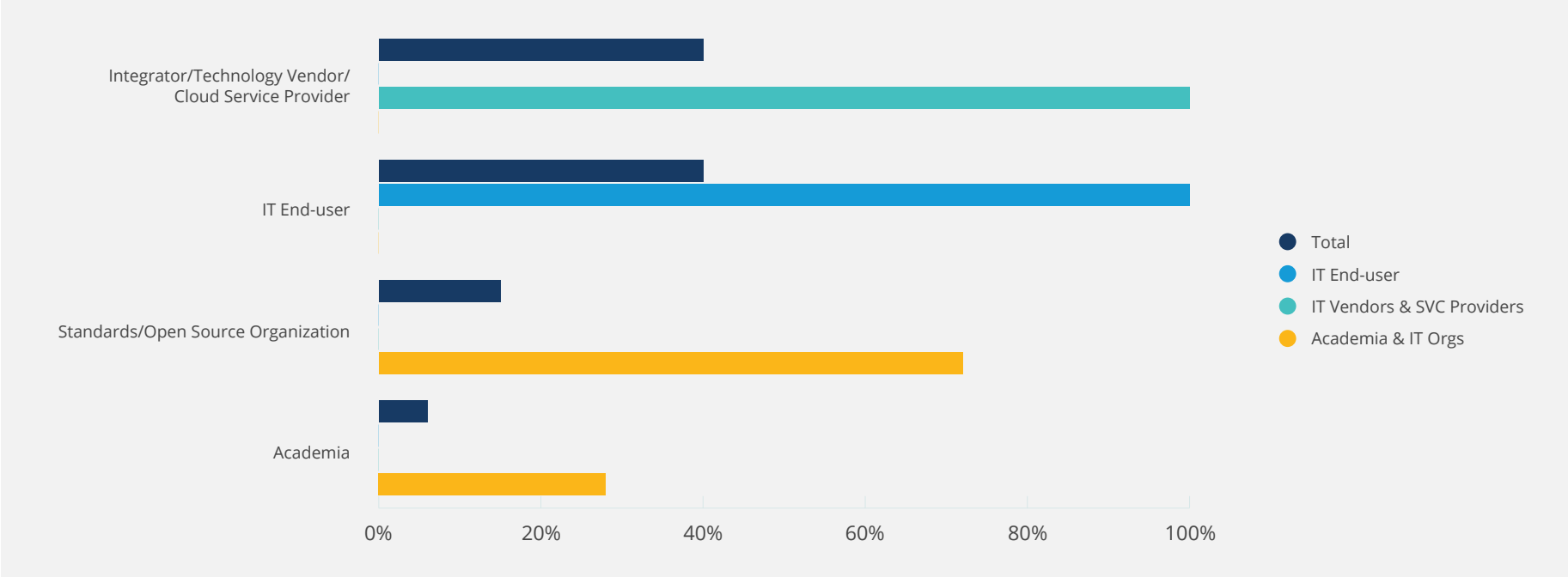


**Figure 19: Type of Organization Segmented by Organization Type**

## Which of the following most closely describes the type or function of your organization?

**Sample Size = 247**



# Disclaimer

This report is provided “as is.” The Linux Foundation and its authors, contributors, and/or sponsors expressly disclaim any warranties (express, implied, or otherwise), including implied warranties of merchantability, non-infringement, fitness for a particular purpose, or title, related to this report. In no event will the Linux Foundation and its authors, contributors, and sponsors be liable to any other party for lost profits or any form of indirect, special, incidental, or consequential damages of any character from any causes of action of any kind with respect to this report, whether based on breach of contract, tort (including negligence), or otherwise, and whether or not they have been advised of the possibility of such damage. Sponsorship of the creation of this report does not constitute an endorsement of its findings by any of its sponsors.

# Acknowledgments

This report and the research behind it would not have been possible without the contributions of many individuals. The SODA Foundation and Linux Foundation Research team in particular wish to thank Michael Dolan, Jason Perlow, and Alison Rowan (Linux Foundation); Anjaneya “Reddy” Chagam, Principal Engineer and Chief SDS Architect (Intel); Rakesh Jain, STSM and Researcher, Cloud, IoT, Analytics (IBM Research); Sanil Kumar D, Chief Architect (Huawei); Kei Kusunoki, Technology Development, Storage Engineer (NTT Communications); Kiran Mova, Co-Founder, Chief Architect (MayaData); Nicolas Trangez, Principal Architect (Scality); and Yusuf Yildiz, Solutions Architect (LinBit). Finally, thanks to all who contribute their time and talents to open source, especially in data and storage.

twitter.com/linuxfoundation

facebook.com/TheLinuxFoundation

linkedin.com/company/the-linux-foundation

youtube.com/user/TheLinuxFoundation

twitter.com/sodafoundation

linkedin.com/company/sodafoundation

youtube.com/channel/UCwHQwtFfinwD6PvutvQlfog/featured

soda foundation

SODA Foundation is an open source project under Linux Foundation that aims to establish an open, unified, and autonomous data management framework for data mobility from the edge, to core, to cloud. SODA brings together industry leaders to collaborate on building a common framework to promote standardization and best practices for data storage, data protection, data governance, data analytics, etc. to support IoT, big data, machine learning, and other applications. We are fostering collaboration and innovation across vendors, system integrators, cloud service providers, standards organizations, and consortiums across different industries, to provide quality end-to-end solutions to end-users.

To join, contact us at Contact - SODA Foundation

THE LINUX FOUNDATION | Research

Founded in 2021, Linux Foundation Research explores the growing scale of open source collaboration, providing insight into emerging technology trends, best practices, and the global impact of open source projects. Through leveraging project databases and networks, and a commitment to best practices in quantitative and qualitative methodologies, Linux Foundation Research is creating the go-to repository for open source insights for the benefit of organizations the world over.

October 2021

REVISED: This report has been updated since its original release on 14 October, 2021. This second version, released on 28 October, 2021, corrects errors found in the original text and graphics.