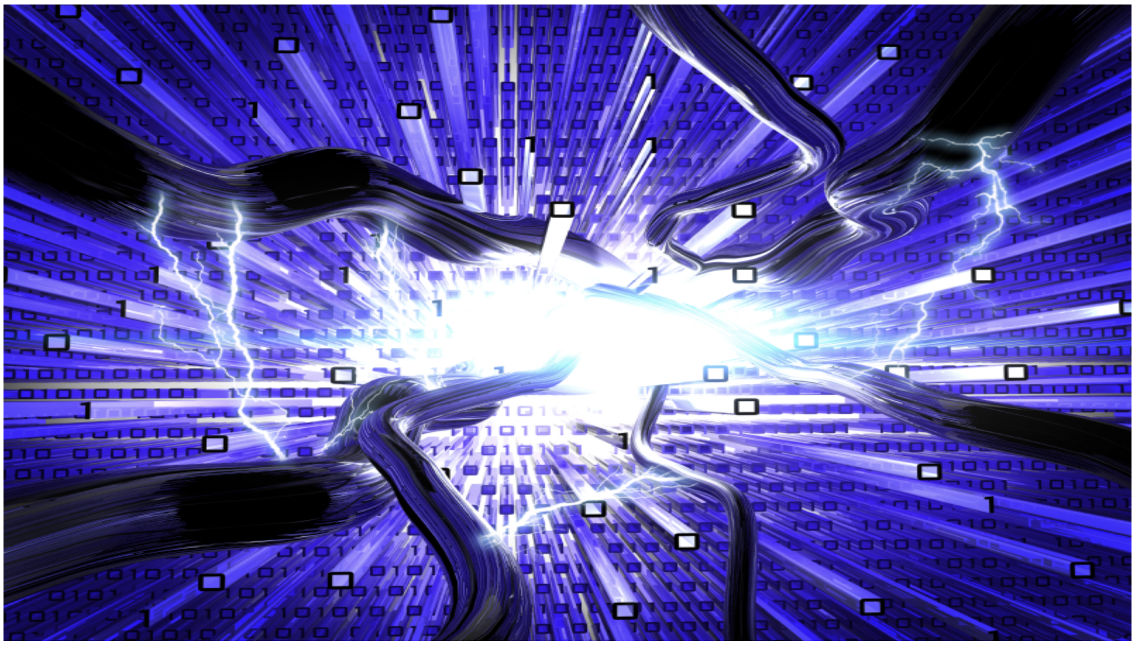


Data Center Downtime at the Core and the Edge: A Survey of Frequency, Duration and Attitudes

Sponsored by Vertiv

Independently conducted by Ponemon Institute LLC
Publication Date: January 2021

Data Center Downtime at the Core and the Edge: A Survey of Frequency, Duration and Attitudes

Ponemon Institute, January 2021

Table of Contents

Data Center Downtime at the Core and the Edge: A Survey of Frequency, Duration and Attitudes

Ponemon Institute, January 2021

Introduction

Edge computing is expanding rapidly and re-shaping the data center ecosystem as organizations across industries move computing and storage closer to users to improve response times and reduce bandwidth requirements.

While forms of distributed computing have been common in some sectors for years, this current evolution is distinct in that it is enabling a broad range of new and emerging applications and has higher criticality requirements than traditional distributed computing sites.

At the same time, core data center managers are dealing with increased complexity and balancing multiple and sometimes conflicting priorities that can compromise availability.

As a result, today's data center networks are more vulnerable to downtime than ever before. In an effort to quantify that vulnerability, the Ponemon Institute conducted a study of downtime frequency, duration and attitudes at the core and the edge, sponsored by Vertiv.

The study is based on responses from 425 participants representing 132 data centers and 1,667 edge locations. All core and edge data centers included in the study are located in the United States/Canada and Latin America (LATAM).

The study found data center networks vulnerable to downtime events across the network. Core data centers experienced an average of 2.4 total facility shutdowns per year with an average duration of more than two hours (138 minutes). This is in addition to almost 10 downtime events annually isolated to select racks or servers. At the edge, the frequency of total facility shutdowns was even higher, although the duration of those outages was less than half that of those in core data centers.

The study also looks at the attitudes that shape decisions regarding core and edge data centers to help identify factors that could be contributing to downtime events. More than half (54%) of all core data centers are not using best practices in system design and redundancy, and 69% say their risk of an unplanned outage is increased as a result of cost constraints.

Leading causes of unplanned downtime events at the core and the edge included cyberattacks, IT equipment failures, human error, UPS battery failure, and UPS equipment failure.

Finally, the study asked participants to identify the actions their organizations could take to prevent future downtime events. They identified activities ranging from investment in new equipment to infrastructure redundancy to improved training and documentation.

Key Findings

Facility Size

Edge data centers aren't necessarily defined by size but by function. For the purpose of this research, edge data centers are defined as facilities that bring computation and data storage closer to the location where it is needed to improve response times and save bandwidth. Nevertheless, as seen in Figure 1, edge data centers were on average about one-third the size of the core data centers.

The extrapolated size for core data centers that participated in this study is 15,153 square feet/1,408 square meters. For edge computing facilities, the average size is 5,010 square feet/465 square meters.

Figure 1: Average square footage of core and edge computing facilities

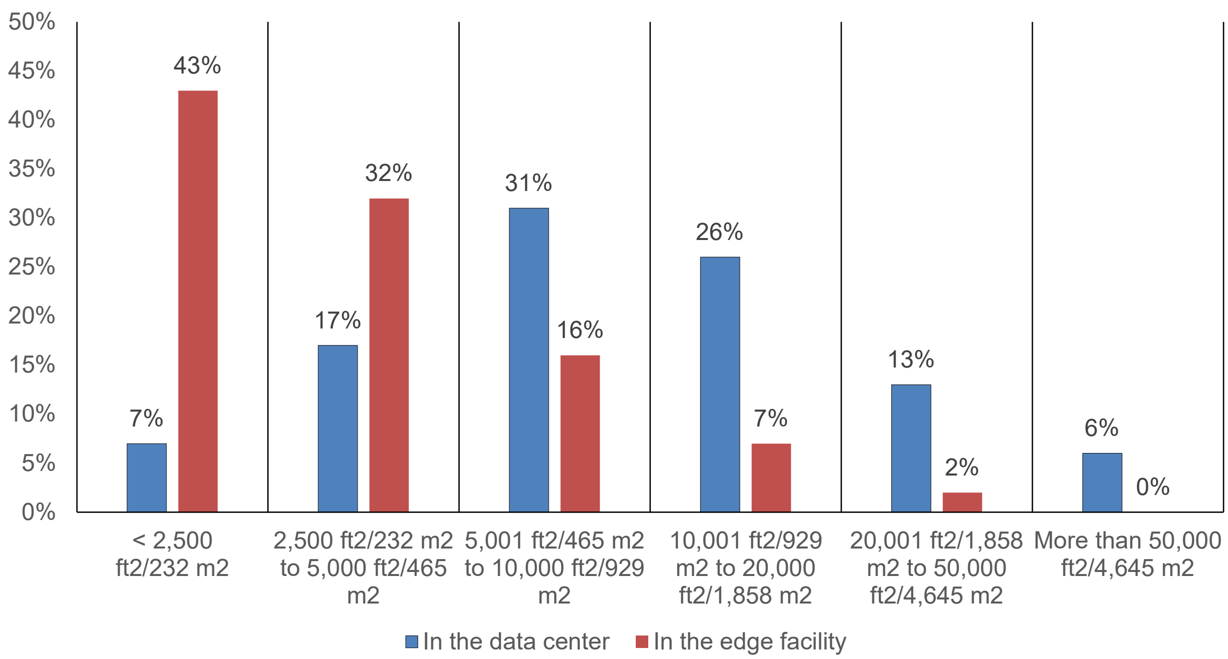


Figure 2 shows average data center size, by quartile in U.S./Canada and LATAM. Across all quartiles, the U.S. and Canadian data centers studied were significantly larger than those in Latin America.

Figure 2: Data center size by quartile

Quartile	Overall Size	U.S./Canada Size	LATAM Size
Qtrl 1	4,001 ft²/372 m2	5,040 ft²/468 m2	2,962 ft²/275 m2
Qtrl 2	8,452 ft²/785 m2	10,732 ft²/997 m2	6,172 ft²/573 m2
Qtrl 3	15,898 ft²/1,477 m2	19,005 ft²/1,766 m2	12,791 ft²/1,188 m2
Qtrl 4	32,400 ft²/3,010 m2	40,500 ft²3,763 m2	24,300 ft²/2,258 m2

Frequency of Core and Edge Downtime

Figure 3 shows the shutdown experience of participating data centers over the past 24 months. As can be seen, total data center shutdown has the lowest frequency (4.81). However, these events are also the most disruptive, and the 4.81 unplanned total facility shutdowns over a 24-month period would be considered unacceptable for many organizations.

Partial outages of certain racks in the data center have the highest frequency at 9.93, followed by individual server outages at 9.43.

It can be difficult to directly compare the total number of downtime events in edge and core facilities due to the higher complexity generally found in core data centers and the increased presence of personnel in these facilities. However, it is possible to compare total facility shutdowns for core and edge data centers. Edge data centers experienced a slightly higher frequency of total facility shutdowns at an average of 5.39 over 24 months. As edge sites continue to proliferate, reducing the frequency of outages at the edge will become a high priority for many organizations.

Figure 3: Downtime frequency over 24 months

Type of Event	Frequency
Core Data Center	
Primary utility power outage	7.19
Total facility shutdown	4.81
Local shutdown of certain racks	9.93
Outage limited to individual servers	9.43
Edge Data Centers	
Total facility shutdown	5.39

Figure 4 shows that data centers in LATAM were more likely to experience all types of outage at both the core and the edge than data centers in the U.S. and Canada.

Figure 4: Frequency of outages in the past 24 months: U.S./Canada versus LATAM

Type of Outage	Americas	U.S./Canada	LATAM
Core Data Center			
Primary utility power outage	7.19	5.60	9.22
Total data center shutdown	4.81	4.00	5.84
Local shutdown of certain racks	9.93	8.26	12.07
Outage limited to individual servers	9.43	7.68	11.66
Edge Data Center			
Edge computing facility shutdown	5.39	4.98	5.92

Duration of Core and Edge Downtime Events

Figure 5 reports the average duration of various types of outage events for the period studied. Total shutdowns of the core data center have the longest duration at 138 minutes. In contrast, partial outages of certain racks in the data center have the highest frequency at 9.93 but the shortest duration at just less than an hour.

Downtime of edge facilities has a considerably shorter duration than similar events in the core data center, despite the limited technical resources generally available at these sites. This is likely due to the more focused functionality of these sites that limits complexity.

Figure 5. Downtime duration by type of event

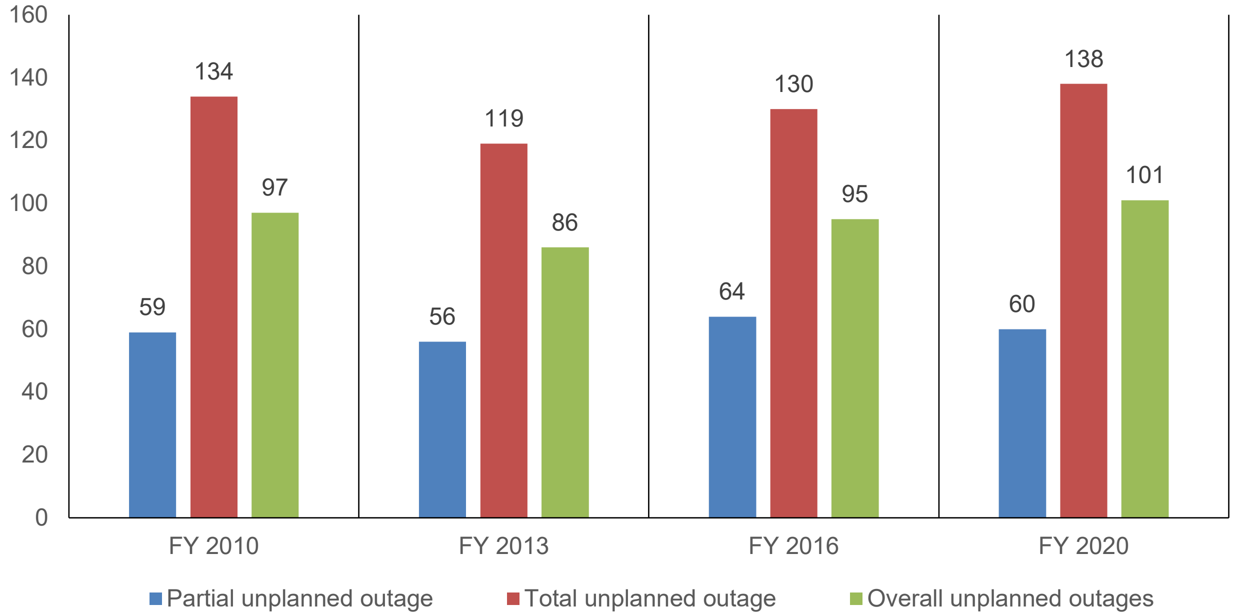
Type of Event	Duration (min)
Core Data Center	
Primary utility power outage	99.80
Total data center shutdown	137.87
Local shutdown of certain racks	59.98
Outage limited to individual servers	60.05
Edge Data Center	
Total facility shutdown	45.40

Figure 6 puts the current downtime duration data in historical perspective by comparing the average total and partial unplanned outage duration for core data centers with previous studies using the same methodology. (Frequency and duration for edge data centers were not included in previous studies.)

All studies show that the duration of total unplanned outages is more than twice the length of time as partial outages over the past 10 years. For example, in 2020 total unplanned outages lasted 138 minutes and partial unplanned outages lasted 60 minutes.

The data also shows that the duration of unplanned outages has risen steadily over the last three studies. The average duration for all unplanned outages rose from 86 minutes in 2013 to 101 minutes in 2020 and total facility outages from 119 minutes in 2013 to 138 minutes in 2020.

Figure 6: Trends in the duration for partial and total shutdown (measured in minutes)
Comparison of 2010, 2013, 2016 and 2020 results



Attributes of Edge and Core Data Centers

When asked to compare attributes for edge and core facilities, participants in the study showed relatively minor differences in the risks associated with each (Figure 7). Interestingly, availability was seen as a higher priority at the edge than in the core data center: Sixty-two percent of participants consider availability the highest priority, including cost minimization, compared to 55% for core data centers. This may be due to the limited technical resources available at these sites to deal with downtime events and the number of edge sites organizations expect to be supporting over the next five years.

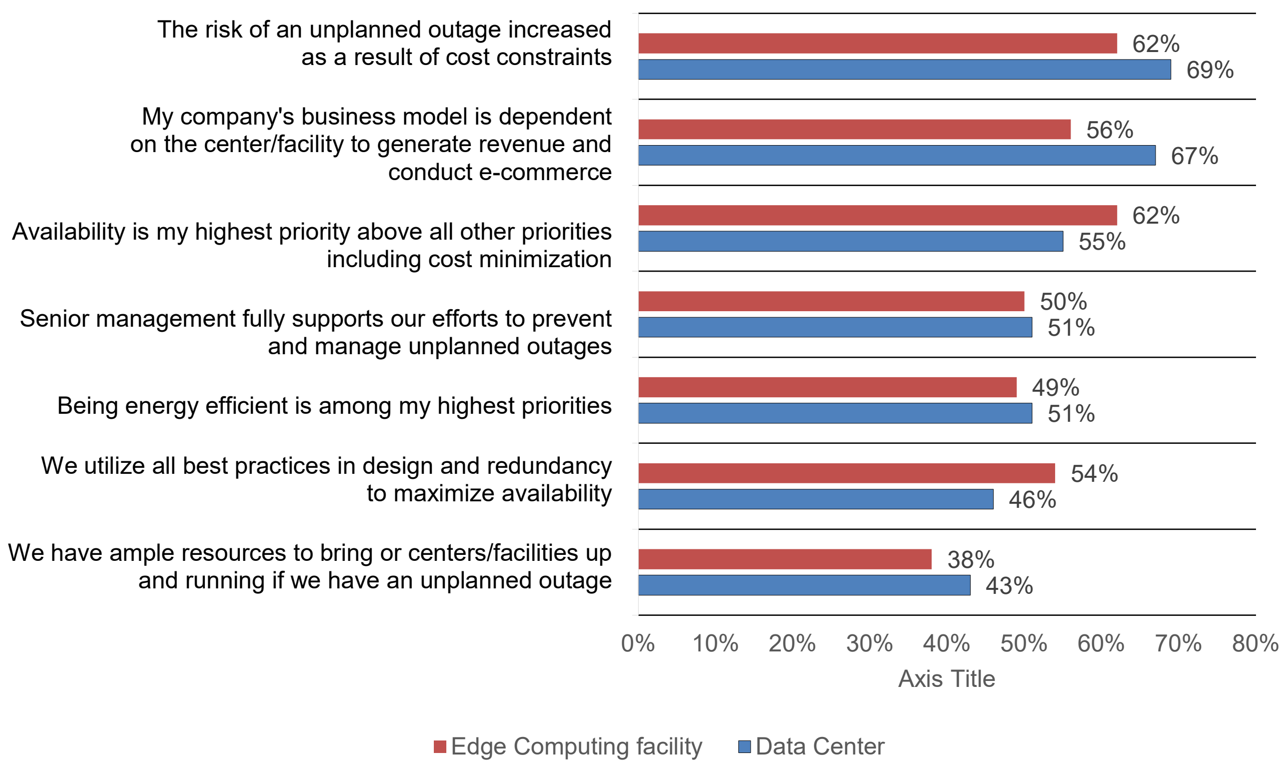
Best practices in design and redundancy were also employed more consistently at the edge. However, even at the edge, only slightly more than half of participants in the study (54%) said they were employing best practices. This indicates that participants are aware of best practices but are limited in their application either due to cost constraints, failing to prioritize availability, or some combination of the two.

A substantial majority of participants cite cost constraints as increasing the risk of unplanned outages for edge and core data centers, indicating that even some facilities that prioritize availability over cost minimization are not making the necessary investments to reduce downtime risks. This perception is reinforced by the finding that only half of participants say their senior management fully supports efforts to prevent unplanned outages.

Energy efficiency, which is driven by both cost and environmental concerns in today's data centers, was cited as among the highest priorities for both edge and core facilities by about half of participants (51% for core data centers and 49% for edge), indicating that organizations are taking a similar approach to energy efficiency at the core and the edge.

Core data centers are more likely to report that the business they support is dependent on systems that generate revenue and conduct e-commerce (67% of data centers versus 56% of edge locations).

Figure 7: Attributes for edge and core data centers



Each percentage shows the combined Agree and Strongly Agree response (using a five-point agreement scale).

Causes of Edge and Core Data Center Downtime

Cyber attacks, IT equipment failures, failures due to human error, UPS battery failures, and UPS failures were the leading causes of outages cited by participants in the study. Edge facilities were slightly more vulnerable to cyberattacks, IT equipment failures, and failures from human error than core data centers.

Other causes of outages cited included automatic transfer switch (ATS) failure, generator failure, UPS capacity exceeded, weather-related failures, heat-related failures, and water incursion.

Actions to Prevent Downtime Events

According to the Uptime Institute 2020 Global Data Center Survey, three of four participants said their most recent downtime event was preventable. While it may not be possible to transfer those results directly to this study, it does suggest that there is an opportunity to reduce the frequency of downtime events in many data centers.

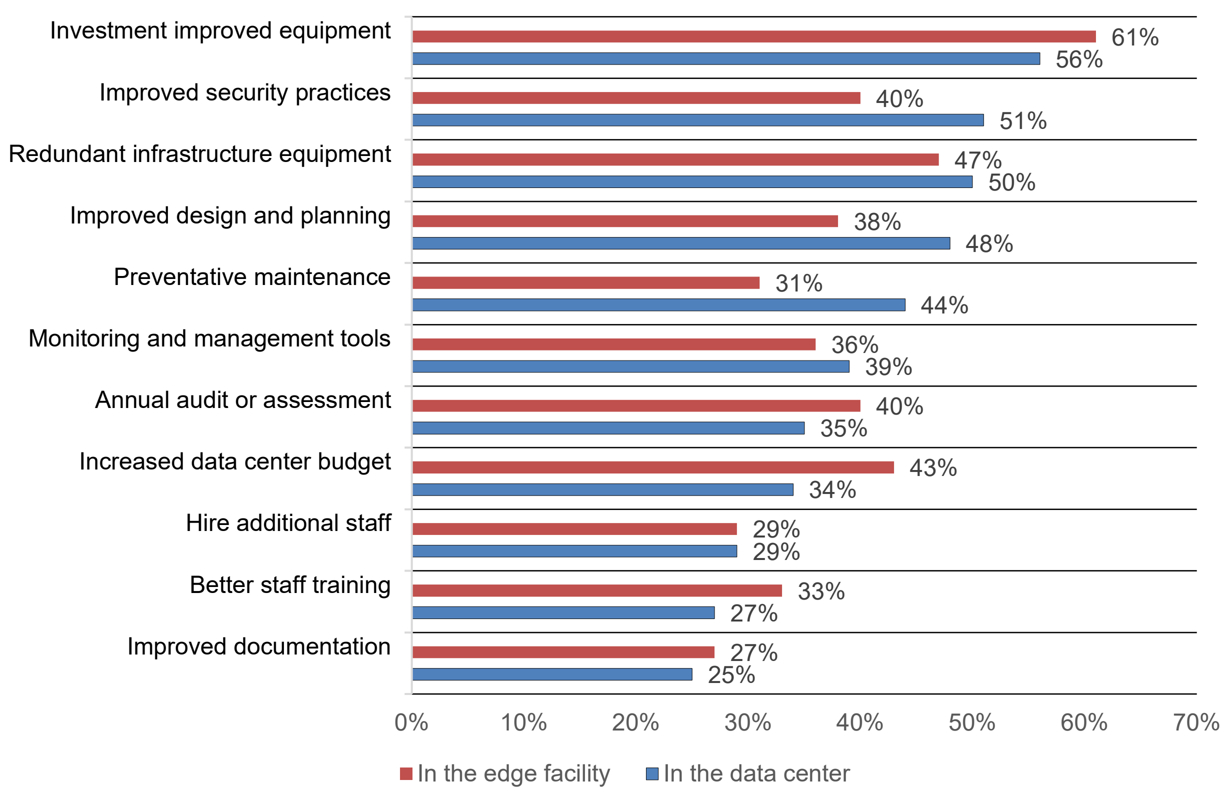
Figure 9 shows the actions participants said could be taken to prevent unplanned outages in the future. The number one action cited at both the core and the edge was investment in IT equipment. This is likely due to the high frequency of downtime events related to individual servers.

Similarly, 51% of core data centers and 40% of edge locations select improved security practices as a primary step to preventing unplanned outages, reflecting continued concern about the impact of denial-of-service attacks on data center availability.

Other steps participants listed could have an impact on reducing long-duration total facility outages, including redundant infrastructure equipment, improved design and planning, preventive maintenance, management and monitoring tools, audits or assessments, increased budget, hiring additional staff, and improved staff training.

Figure 8: What can be done to prevent unplanned outages in the future?

Comparison of data centers and edge locations



Participant Profile

The study is based on responses from 425 participants representing 132 data centers and 1,667 edge locations. All core and edge data centers included in the study are located in the United States and Canada and Latin America and Mexico (LATAM).

The following table summarizes the distribution of companies and separate data centers participating in the study. A total of 15 industry sectors are represented in the final sample. Our final sample includes a total of 108 separate organizations representing 132 data centers and 1,667 edge facilities.

Figure 9: Sample distribution of data centers located In the U.S./Canada and LATAM

Industries	Companies	Data centers	U.S./Canada	LATAM	U.S./Canada Edge locations	LATAM Edge locations
Financial services	14	14	8	6	125	87
Healthcare & pharmaceuticals	11	10	7	3	78	31
E-commerce	9	11	8	3	234	112
Industrial & manufacturing	9	12	7	5	67	32
Education	9	9	4	5	33	20
Media & entertainment	9	9	3	6	27	13
Government	8	13	5	8	89	56
Retail	8	7	6	1	138	61
Colocation	7	9	6	3	24	5
Utilities & energy	5	8	4	4	52	35
Services	5	9	5	4	53	42
Communications	5	7	3	4	63	52
Consumer products	3	4	2	2	36	20
Hospitality	3	5	3	2	20	17
Transportation	3	5	3	2	24	21
Total	108	132	74	58	1,063	604

Figure 10 summarizes the sample of participating companies' core data centers according to 15 primary industry classifications. Financial services and healthcare are the two largest industry segments representing 13% and 10% of the sample, respectively. Financial services companies include retail banking, payment processors, insurance, brokerage, and investment management companies.

Figure 10: Distribution of participating organizations by industry segment

Computed from 132 benchmarked data centers

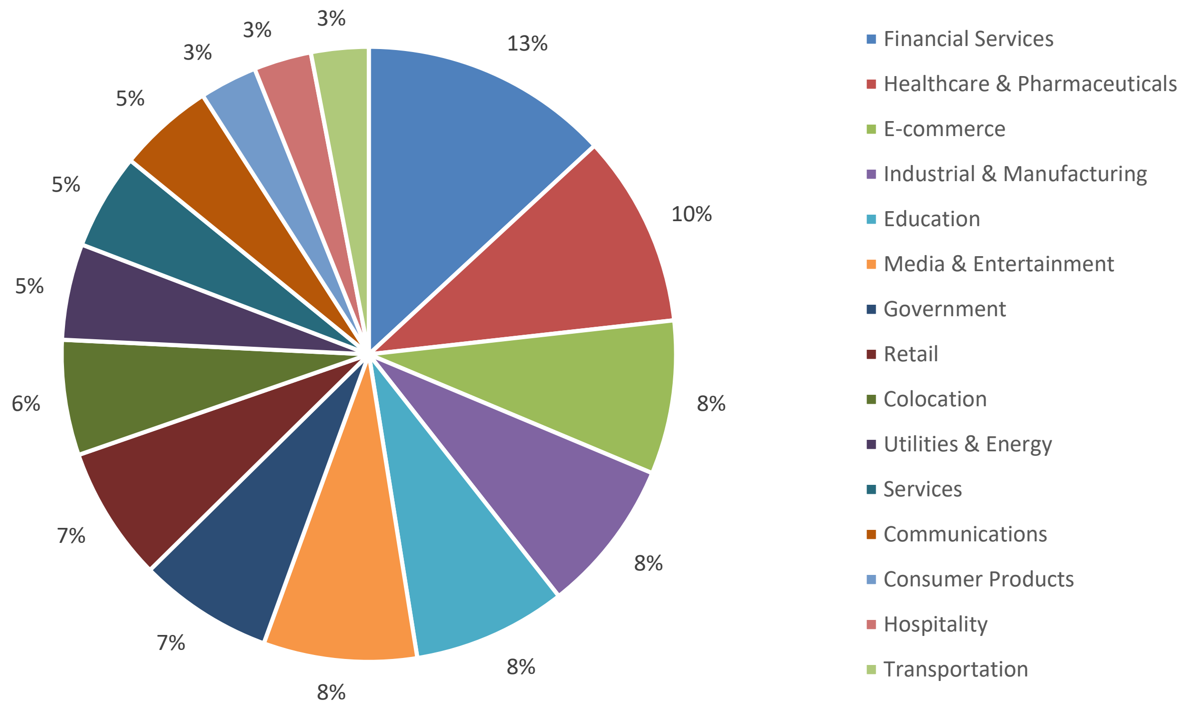
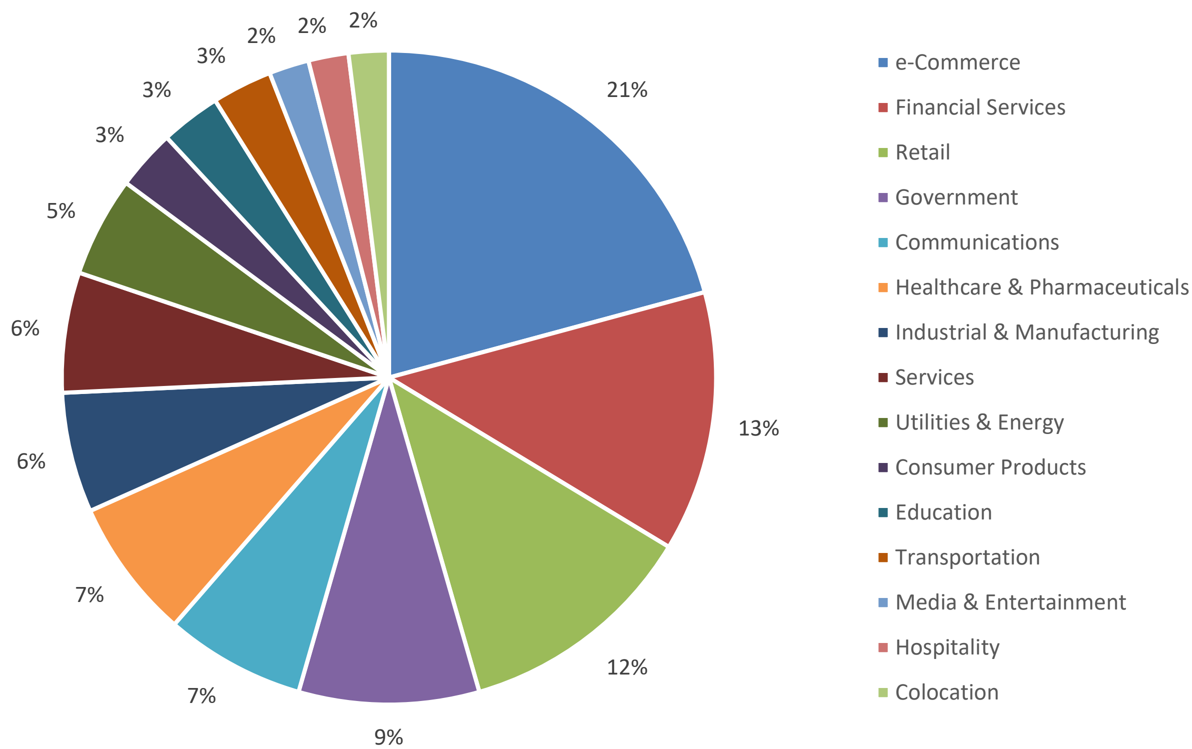


Figure 11 reports the percentage frequency of 1,667 edge locations by industry classification. At 21%, e-commerce is the largest segment, followed by financial services at 13%.

Figure 11: Distribution of edge computing facilities by industry

Computed from 1,667 benchmarked edge locations



Following are the functional leaders within each organization who participated in the benchmarking process:

- Facility manager
- Chief information officer
- Data center management
- Chief information security officer
- IT operations management
- IT compliance and audit
- Operations and engineering
- Cloud administrator

Caveats

This study utilizes a confidential and proprietary benchmark method that has been successfully deployed in earlier Ponemon Institute research. However, there are inherent limitations to benchmark research that need to be carefully considered before drawing conclusions from findings.

- **Non-statistical results**: The purpose of this study is descriptive rather than normative inference. The current study draws upon a representative, non-statistical sample of data centers, all experiencing at least one unplanned outage over the past 12 months. Statistical inferences, margins of error, and confidence intervals cannot be applied to these data given the nature of our sampling plan.

- **Non-response**: The current findings are based on a small representative sample of completed case studies. An initial mailing of benchmark surveys was sent to a benchmark group of more than 600 organizations, all believed to have experienced one or more outages over the past 12 months. One hundred and thirty-two data centers provided usable benchmark surveys. Non-response bias was not tested so it is always possible companies that did not participate are substantially different in terms of the methods used to manage the detection, containment and recovery process.

- **Sampling-frame bias**: Because our sampling frame is judgmental, the quality of results is influenced by the degree to which the frame is representative of the population of companies and data centers being studied. It is our belief that the current sampling frame is biased toward companies with more mature data center operations.

- **Company-specific information**: The benchmark information is sensitive and confidential. Thus, the current instrument does not capture company-identifying information. It also allows individuals to use categorical response variables to disclose demographic information about the company and industry category. Industry classification relies on self-reported results.

- **Unmeasured factors**: To keep the survey concise and focused, we decided to omit other important variables from our analyses such as leading trends and organizational characteristics. The extent to which omitted variables might explain benchmark results cannot be estimated at this time.

If you have questions or comments about this report, please contact us by letter, phone call or email:

Ponemon Institute LLC
Attn: Research Department
2308 US 31 North
Traverse City, Michigan 49686 USA
1.800.877.3118
research@ponemon.org

Ponemon Institute

Advancing Responsible Information Management

Ponemon Institute is dedicated to independent research and education that advances responsible information and privacy management practices within business and government. Our mission is to conduct high quality, empirical studies on critical issues affecting the management and security of sensitive information about people and organizations.

We uphold strict data confidentiality, privacy and ethical research standards. We do not collect any personally identifiable information from individuals (or company identifiable information in our business research). Furthermore, we have strict quality standards to ensure that subjects are not asked extraneous, irrelevant or improper questions.