CHAOSSEARCH

2021 BENCHMARK REPORT

Log Management and Analytics

Log management and analytics remains a dynamic field. Companies who value log data and invest in capturing data will harness insights that propel their businesses forward.

1. INTRODUCTION

Most companies want to add new data sources and expand how they use log data to support business operations.

Meanwhile, the volumes of log data continue to grow rapidly, putting pressure on organizations to scale their log management systems to handle the growth. Given these dynamics, it's no wonder that analysts expect the global market for log management solutions to grow from $1.9 Billion in 2020 to $3.7 Billion by 2025.[1]

In 2021, ChaosSearch conducted a research project to better understand how customers are managing their ever-growing volumes of log data, and how they are harnessing it to drive their daily operations.

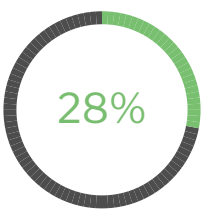
Through detailed phone interviews and an online survey, we collected data on 50 medium-to-large enterprise organizations that make heavy use of log data management in their daily operations. Participants were selected based on their role in managing their log data management environment. They represent a wide range of industries and organization sizes. See page 18 for details on survey demographics.

The research confirms that organizations continue to recognize the value of tapping into their log data to achieve better outcomes across a range of business and IT functions. While facing continued challenges in handling the scale requirements that the ever-growing flood of data presents, organizations recognize there remain greater opportunities for them to expand the scope and discipline around their practice of log management and they plan continued growth and investment in the foreseeable future.

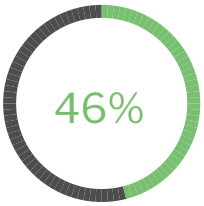
[1] Log Management Market, Region-Global Forecast to 2025, MarketsandMarkets, 2020.

KEY FINDINGS:

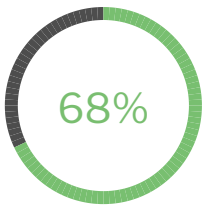
Top Use Cases:



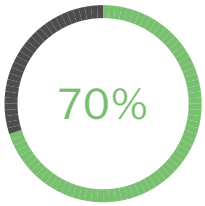
Business Operations



BI & Analytics

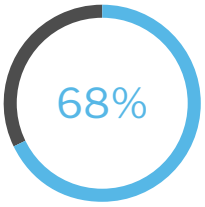


IT Monitoring

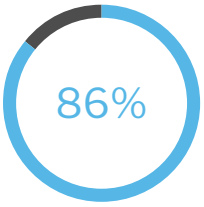


Security

In the coming months, as it pertains to log management:

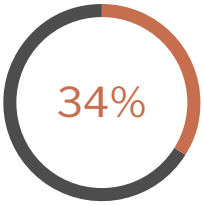


Plan to extend the usage to new functions

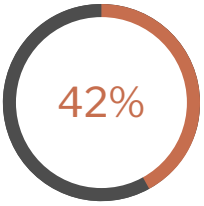


Expect to increase their budgets

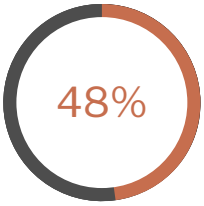
Top Challenges:



Managing Growth



Managing Costs

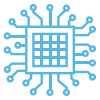


Infrastructure Complexity

Top Priorities:
(Log Management)

Increasing the centralization of their log management function

Improving the ability to correlate multiple data streams



Adding more data sources for log management and analysis

Top Priorities:
(Security Operations)

Threat Hunting

Forensic Investigations

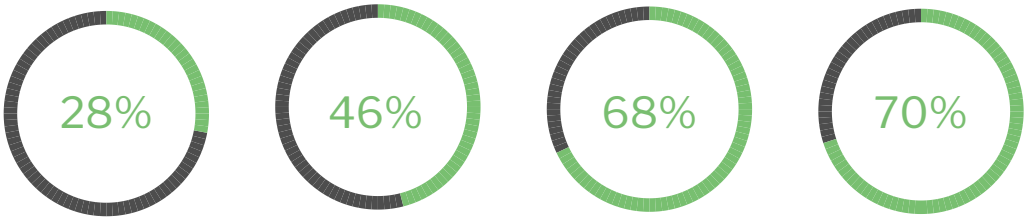
2. OVERVIEW OF LOG MANAGEMENT TODAY

The study confirms that organizations are leveraging log data across a range of use cases today.

Over 90% of survey respondents reporting using log management and analytics for at least two primary use cases. Unsurprisingly, the top two use cases remain the two original drivers of the market: security (70%) and IT monitoring (68%).

Interestingly, almost half of the respondents reported that Business Intelligence / Analytics is a primary use case, and 28% use log management in their daily business operations.

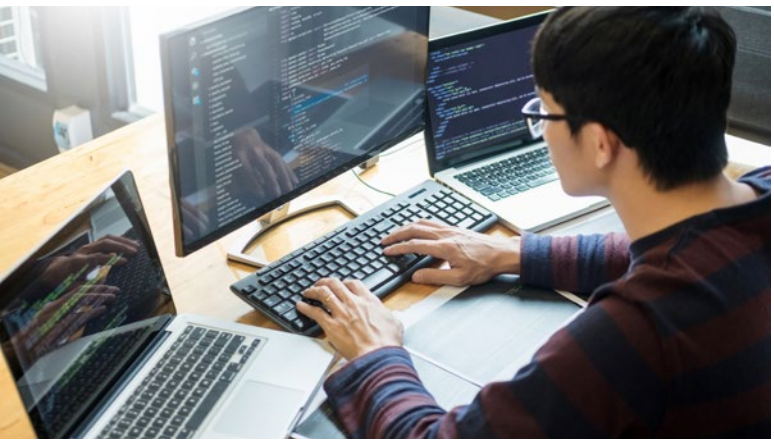
Which of the following are primary use cases for your log management and analytics function?



Within the security domain, participants demonstrate a wide range of SecOps uses for log data and log analytics, with investigations and insider threats being the most prominent.

Which of the following security operations are performed today in your organization, using log analytics?

While IT monitoring and security remain the leading uses of log data, the survey uncovered a wide range of interesting use cases within the category of business operations. At right are a few examples of how companies are using log analytics to drive better business outcomes and improve their competitiveness in the market:



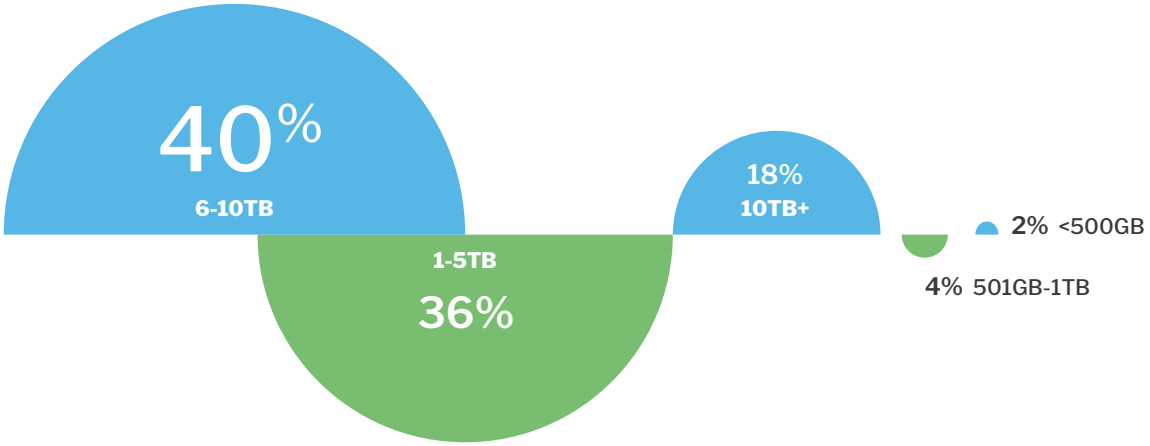
- A global industrial equipment manufacturer attaches IOT devices to farm equipment in order to track weather, temperature, humidity, time-of-day, depth of planting, soil conditions, seeds planted, and geo-routing on farms to deliver farm productivity reporting and recommendations daily.
- A biopharmaceutical firm tracks cell culture logs with alerts on temperature and humidity, and leverages the logs for various trend analyses.
- A large industrial printing manufacturer tracks usage of its high-resolution large format color printers to ensure uptime, cost control, timely maintenance, and unauthorized usage. This is a critical aspect of their custom service delivery, and seen as a competitive differentiator.
- A large online retailer leverages log data from web visits and other customer activities, to develop 360-degree profiles of customers, and analyze the steps in the "customer journey" to identify opportunities for improved efficiency and increased revenue capture.

Log Data Ingest and Retention

Many variables drive the volume of log data an organization generates, including the number of users, devices, applications, IT environments, and infrastructure elements and our survey demonstrates a wide range of data capture across the participants. Two things were universal — data growth rates are high and are a source of pain, and data retention is an increasing priority. Our survey shows that companies recognize the potential value of log data, a concern about untapped potential of log data, and a strong desire to do more with that data.

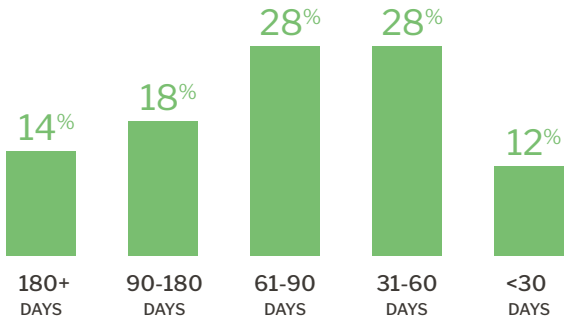
Among the participants, 94% ingest at least 1 TB or more per day, with 18% in the 10TB + range. The average daily ingest volume for the participants is 7.9 TB. The chart shows a wide distribution amongst participants.

What is the average amount of log data ingested per day?



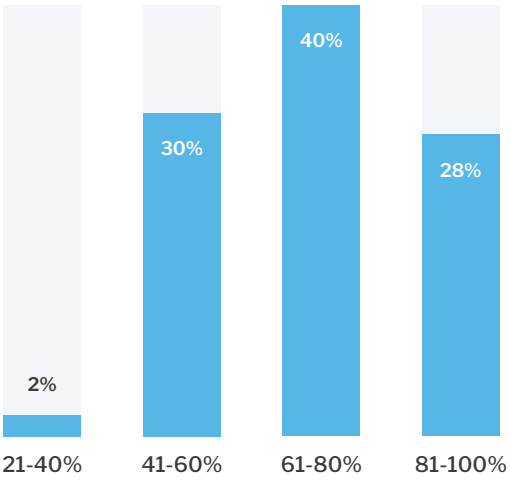
While logs play a major role in providing a real-time view of an organization, and are critical for use in real-time alerting, many use cases rely on data that is more than just a few days old. This is particularly true for the security ops team, that leverages historical data for anomaly detection, forensic investigations, and threat hunting. Our survey shows recognition of the value of historical data with 88% of participants retaining at least one month of log data, and 60% retaining log data for 2 months or more.

What is the average retention rate for log data in your log management platform?



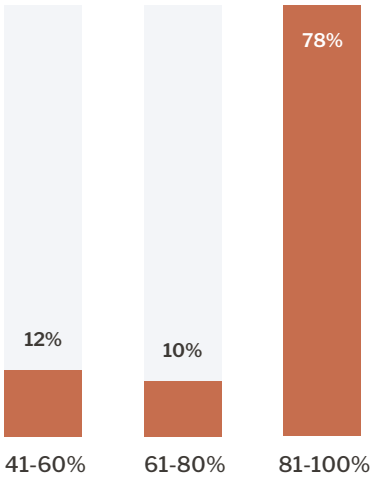
Interestingly, participants acknowledged various limitations that prevent them from capturing all of the log data generated daily and expressed the desire to increase their rate of log data capture, as the charts below exemplify. Whereas only 28% of customers are capturing 80% of log data or more today, 78% express the desire to capture 80% or more.

What percentage of the total log data generated do you capture for use today?



*0% for 0-20%

What is the ideal log data volume capture rate?



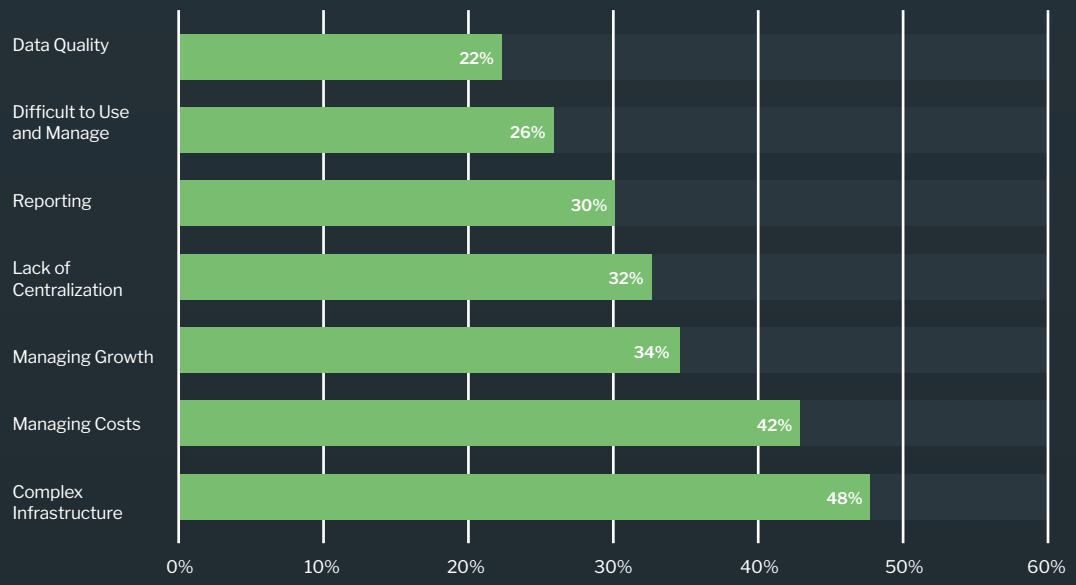
*0% for 0-40%

3. TOP CHALLENGES

The survey reveals a wide range of potential challenges that the team managing the log analytics function must wrestle with. 48% identified infrastructure complexity as the top log management challenge.

Managing cost and growth are both in the top 3 challenges, which follows from the responses seen in the earlier sections. Interestingly, nearly a third of participants highlighted the lack of centralization in their log management function.

Which of the following are top log management challenges for your organization today?



4. BEST PRACTICES

Through live interviews and research, we identified a list of the most prominent log management best practices used by organizations today.

The table below introduces the top 10 identified best practices, and a short definition of each.

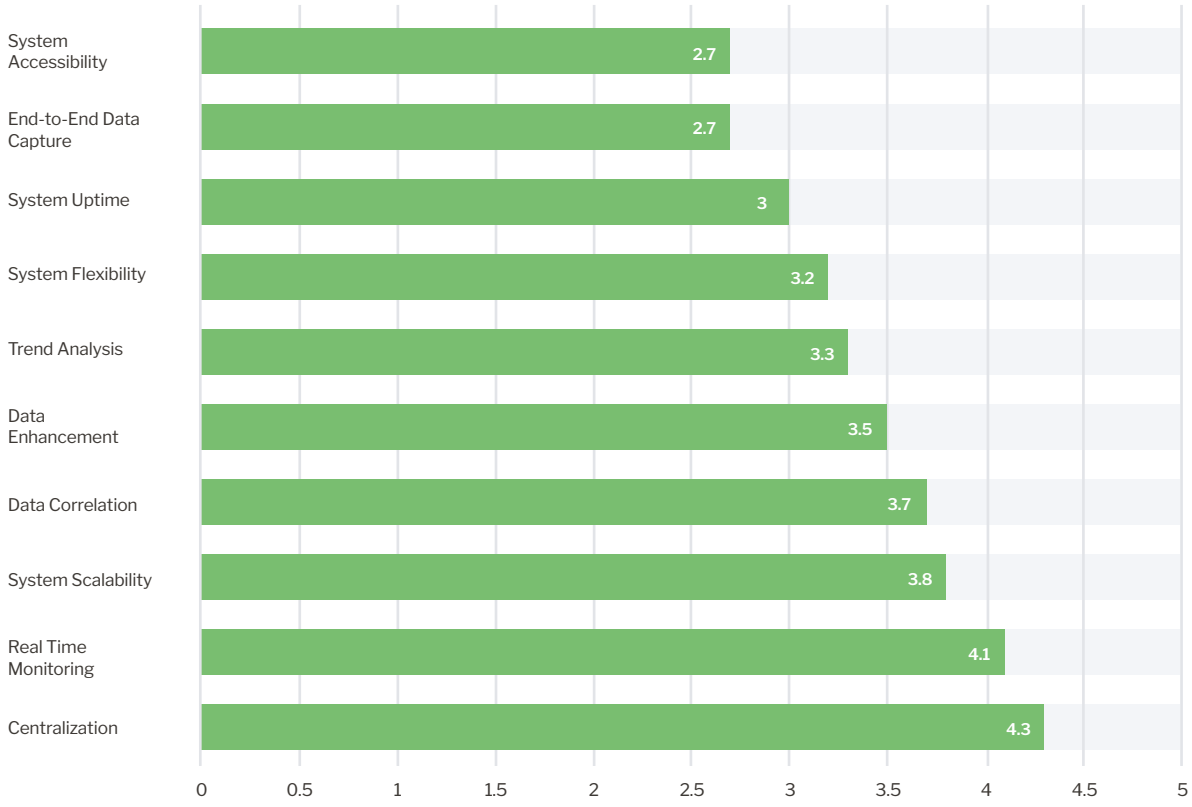
BEST PRACTICE	DEFINITION
Centralization	Consolidation of data streams (and associated reduction of tools) into a centralized repository
Data Correlation	Correlation allows the analysts to derive key insights that cut across multiple data sources, leading to better outcomes for each of the main use cases. Correlation requires the ability to aggregate the relevant data streams, and make them available for use in queries, reports and visualizations.
Data Enhancement	Ability to make available relevant, contextual data available. Depending on context, this often includes directly updating individual data records to include the additional relevant data. For example, security analysts often seek to append relevant command line log data to the log data showing certain events that could be indicators of a threat.
End-to-End Data Capture	This refers to collecting logs across all system components, including relevant metrics and events from the underlying infrastructure, application layers and end user clients.
Real Time Monitoring	Many key use cases in security and infrastructure management rely on real time access to log data. This typically includes proactive alerting on certain events or thresholds.
System Accessibility	This refers to the ability for a wide range of employees form different functions to access and make use of the log data.
System Flexibility	This broadly refers to the ability to make changes without significant disruption. This usually entails adding new data streams from new sources, changing data formats, or integrating with 3rd party products without significant disruption.
System Scalability	This refers to the centralized log management system's ability to scale across two key dimensions: scaling up to handle increased volume of daily data, scaling out to allow for longer retention periods.
System Uptime	Given the usefulness of log data to many core functions, maintaining log management system uptime is critical. While the metric is easy to understand, the operations required to keep a large and growing log management platform online and accessible at all times can by trying.
Trend Analysis	The ability to conduct advanced analytics on historical data, in order to spot trends allowing the analyst to anticipate challenges, spot opportunities, and generally support business decisions.

5. TOP PRIORITIES

We then explored the relative importance of each best practice to the participants, as well as their own assessment of how their organizations rate with regard to each.

Participants identify centralization as the top priority, with an average score of 4.3 out of 5. Real time monitoring and system scalability round out the top 3 with scores of 4.1 and 3.8 respectively.

On a scale of 1-5, rate the following in terms of your organization's log management priorities in the next 12 months.



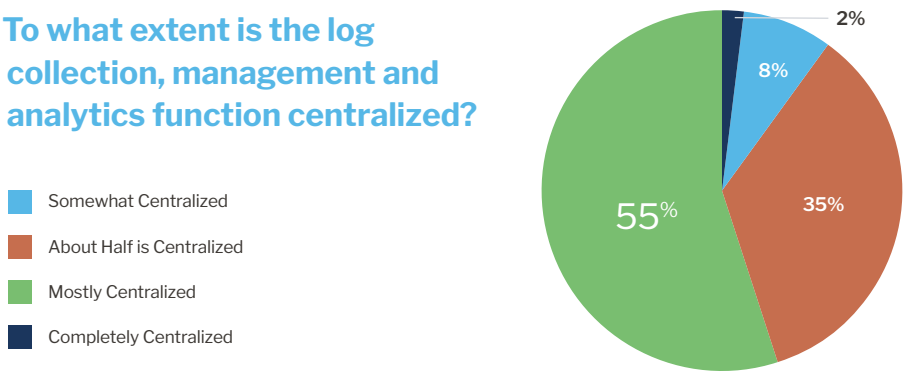
6. BENCHMARKS: TODAY'S LOG MANAGEMENT OPERATIONS

This section shows the aggregated results of how the participants rate their organizations with regard to each identified best practice.

Centralization

While 37% of participants report that their log data is mostly or completely centralized, the survey shows there is a long way to go for the majority of organizations. 55% of participants report that about half of their log data is centralized, and 8% report that log data is only somewhat centralized.

To what extent is the log collection, management and analytics function centralized?



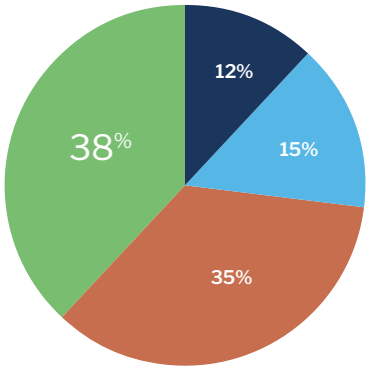
Data Correlation

Participants' responses show that their organizations demonstrate an understanding of the need for data correlation, with 50% showing strong or excellent correlation capabilities.

To what extent do you enable correlation of data from multiple sources?

- Limited
- Moderate
- Strong
- Excellent

***0% Not Done**

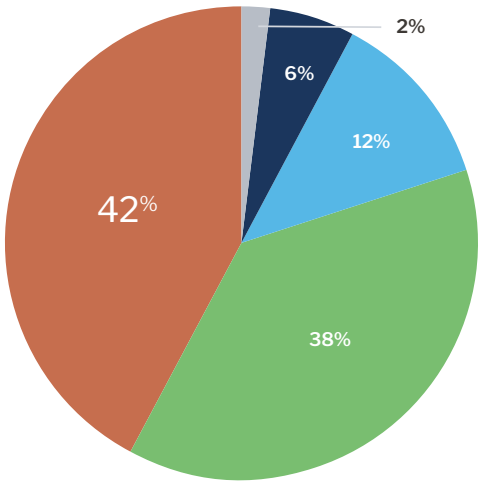


Data Enhancement

The results show a split with slightly less than half (44%) reporting strong or excellent data enhancement capabilities, and 54% reporting moderate or limited capabilities.

To what extent do you enable adding contextual data and/or enhancing log data files with additional data from other sources?

- Not Done
- Limited
- Moderate
- Strong
- Excellent



End-to-End Log Data Management and Analysis

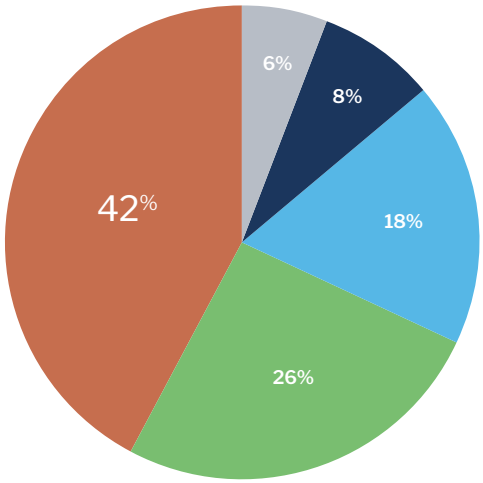
The results show that most companies have not yet achieved mastery in capturing and leveraging the logs from end-to-end, with only 14% rating their organizations as strong or excellent. This likely reflects the complexity of pulling all data from all sources necessary to achieve this. It also demonstrates the tradeoffs that organizations are forced to make, as exemplified in the introduction section which showed the disparity between the actual and the optimal log data capture rate.

To what extent do capture and use end-to-end log data?

Real Time Monitoring

76% of participants rate their organization moderate or better in their use of logs for real-time monitoring, with 34% giving a strong or excellent rating. These results here are somewhat lower than expected given that real time monitoring and alerting is one of the primary roles for log management, but they likely reflect the reality of challenge organizations face in setting up and running real time monitoring and alerting effectively. Many participants highlighted the problem of “alert fatigue” caused by false-positive alerts, which is likely a factor in those that rate their organizations with a moderate or lower score.

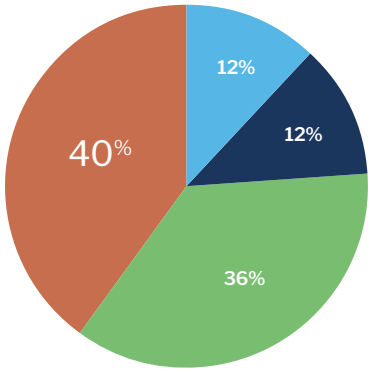
To what extent do you enable real time monitoring of log data?



System Accessibility

Results on this question show recognition of the need to make log data analytics accessible to a number of different groups within the overall organization, with 88% scoring moderate or above, and 48% of participants rating their organizations as either strong or excellent.

To what extent is the log analytics function widely accessible to a wide range of employees, from across different groups?



***0% Not Done**

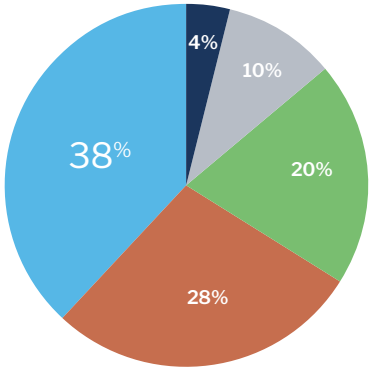
System Flexibility

IT professionals recognize the importance of flexible and dynamic systems. However, when it comes to their log management systems, the survey shows that organizations today generally struggle. Only 24% rate themselves as strong or excellent in this area, with 48% scoring their organization's ability to maintain flexible log management systems as either moderate or poor.

Discussions with participants highlighted a few areas that make change management difficult including:

- **Difficulty of managing disparate data formats**
- **Difficult ingesting and parsing new data sources**
- **Complex data pipelines**
- **Need for frequent reconfiguration of the underlying databases**

To what extent can you add / change data sources or make changes to data formats easily and without disruption?



System Scalability

Scalability and flexibility are related, given that scaling a system is one type of change. Thus, it's not surprising to see similar results when participants score their organizations' ability to scale the log management environment. 50% score this as limited or poor, and only 20% of participants rated this dimension as good or excellent for their organization.

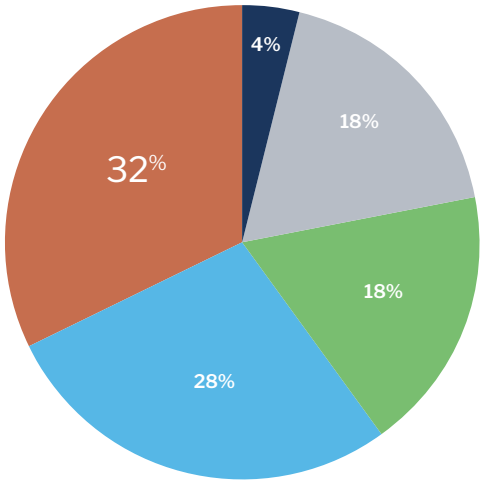
In discussing this challenge with survey participants, they highlighted the frequent need for reformatting or database reconfiguration as a problem with scaling, and many highlighted the associated challenge of the mounting infrastructure expenses that result from scaling their systems.

Please rate the scalability of your log management system.

System Uptime

The large majority of participants highlighted the challenge of keeping their central log management system up and available at all times as a significant pain point. This was particularly true for several of the larger organizations that had global operations and required round-the-clock access to log management for monitoring, reporting and analysis. Only 22% rate their organizations as good or excellent, whereas 46% give themselves a score of limited or poor.

Please rate your ability to maintain log management system uptime.

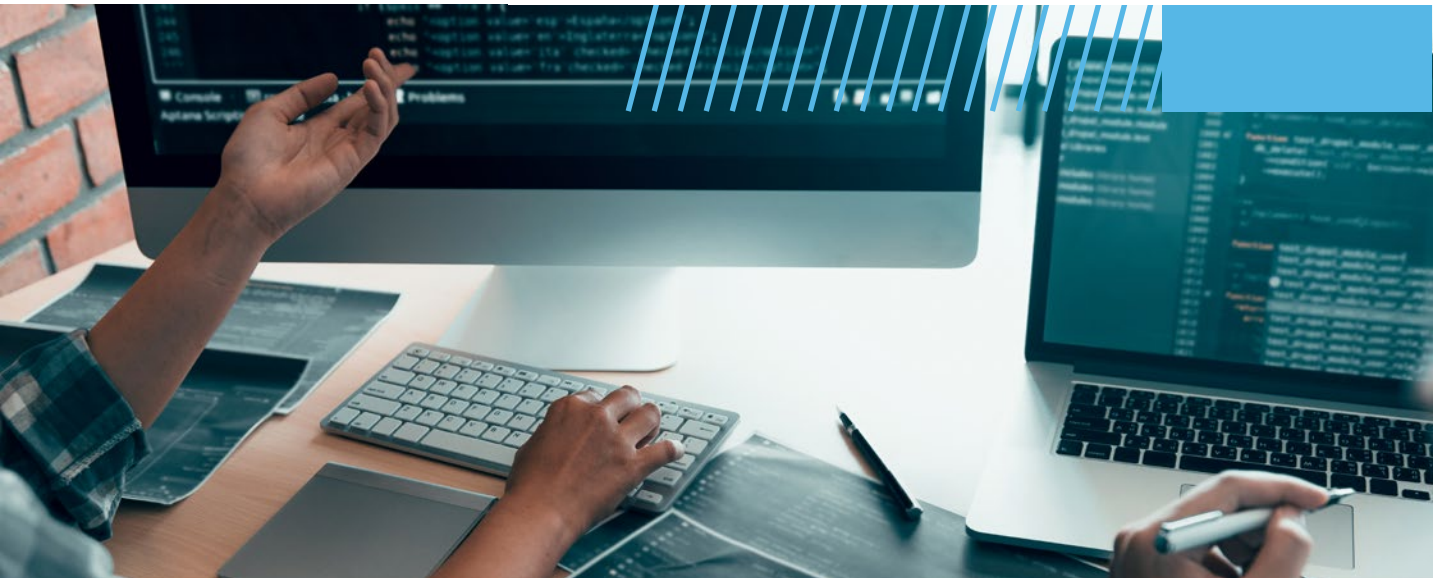


Trend Analysis

Ratings on the capability to perform trend analysis using log data were fairly evenly distributed amongst the participating organization, with 22% performing at a good or excellent level, 30% at a moderate level, 30% at a limited level, and 18% of organizations doing no trend analysis with their log data.

Trend analysis is dependent on the use of long-term historical data, so the results here reflect the fact that 68% of participants have retention periods of 90 days or less.

To what extent does your organization enable trend analysis of log data?

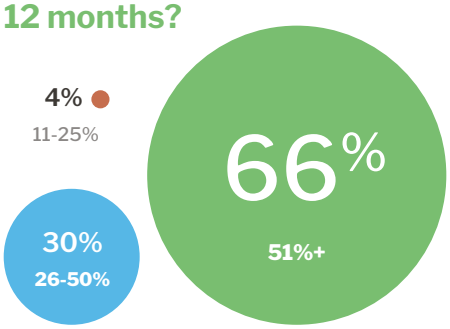


7. GROWTH AHEAD

In assessing the near-term future, participants see more data growth, longer retentions, and increasing budgets in the 12 months ahead.

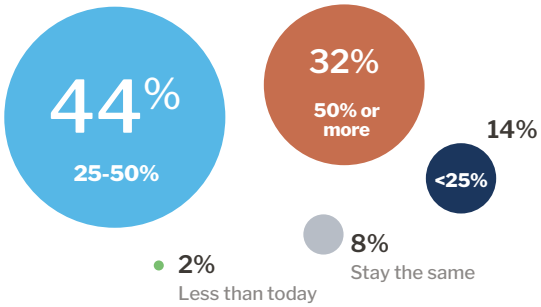
Almost all participants expect growth in both log data volumes and their log data retention. 60% reported that they expect at least 50% log data growth year-over-year, and 32% expect to increase their average retention rate by at least 50%.

At what rate do you expect log data to grow in the next 12 months?



*0% for <10%

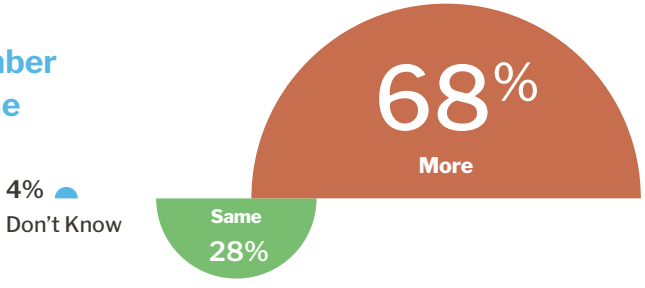
How do you expect your log data retention to change in the next 12 months?



In addition to anticipating growth in data and retention, organizations also plan to expand the scope of how log data is used in their day-to-day operations. A large majority (68%) plan to add more use cases in the coming 12 months, whereas no survey respondents expect to reduce the scope of the log management function.

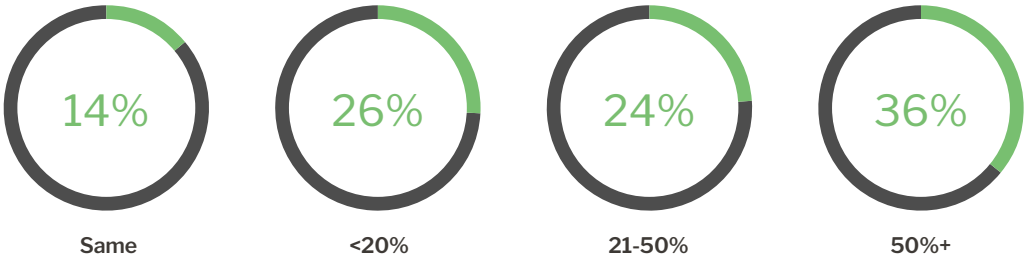
How do you expect the number of use cases to change in the next 12 months?

*0% for Less



Given the above insights, it is unsurprising that participants expect their budgets to increase as well, with 86% anticipating at least 20% growth and 36% expecting their budgets to grow by 50% or more.

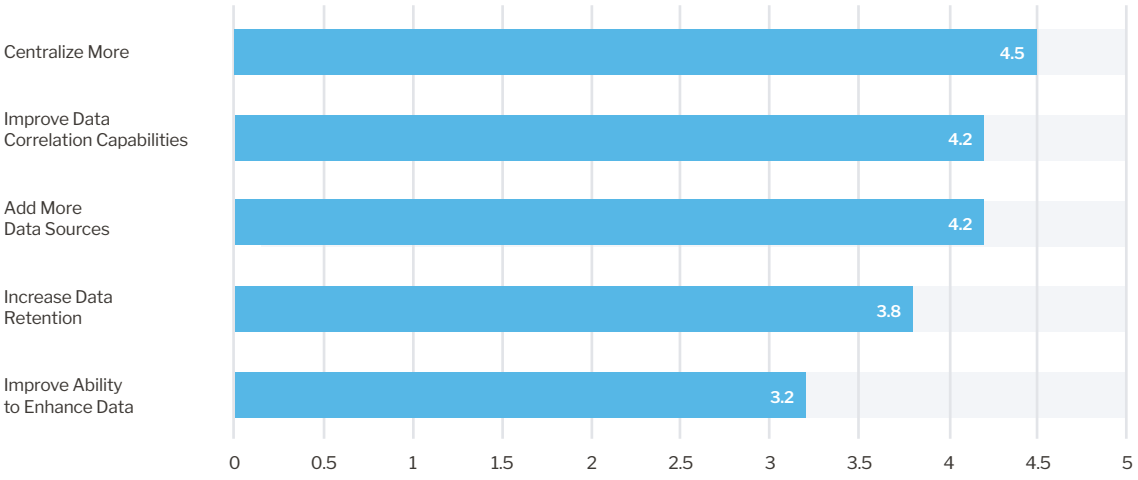
What do you anticipate for budget growth for the log data management function in the next 12 months?



General Log Management Priorities

Asked about their top priorities going forward, respondents again reflected the theme of scalability, scoring it a 4.5 out of 5. Interestingly, the other four of the top five priorities reflect organizations' growing requirement to derive critical insights from their log data. Participants are looking to add more data sources and retain data longer, and in parallel, they will be looking to bolster their advanced analytical capabilities with better data correlation and better data enhancement.

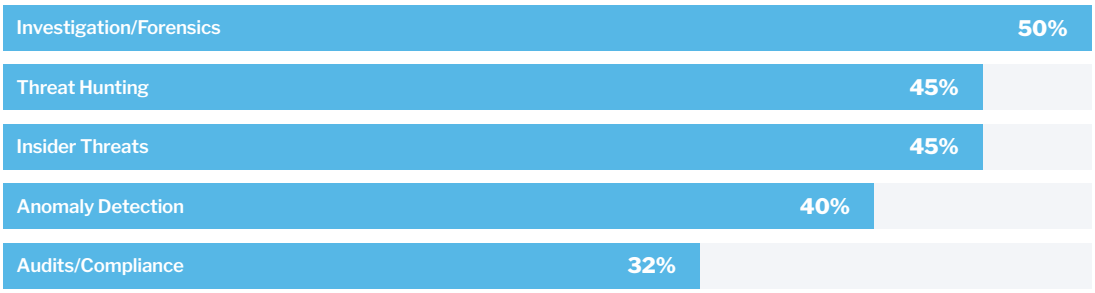
What are your organization's top priorities for your log management and analytics function in the next 12 months?



Security Log Analytics Priorities

Within the realm of log analytics for security operations, threat hunting stands out as a growing priority, with 45% rating it as a top priority in the next year, whereas only 32% identified threat hunting as a use of log data in their organization today (see section 2). Anomaly detection also shows movement with 40% identifying it as a priority in the next year vs. 30% who are performing it today.

What are the top areas of improvement and/or investment for security operations that rely on log data analytics?

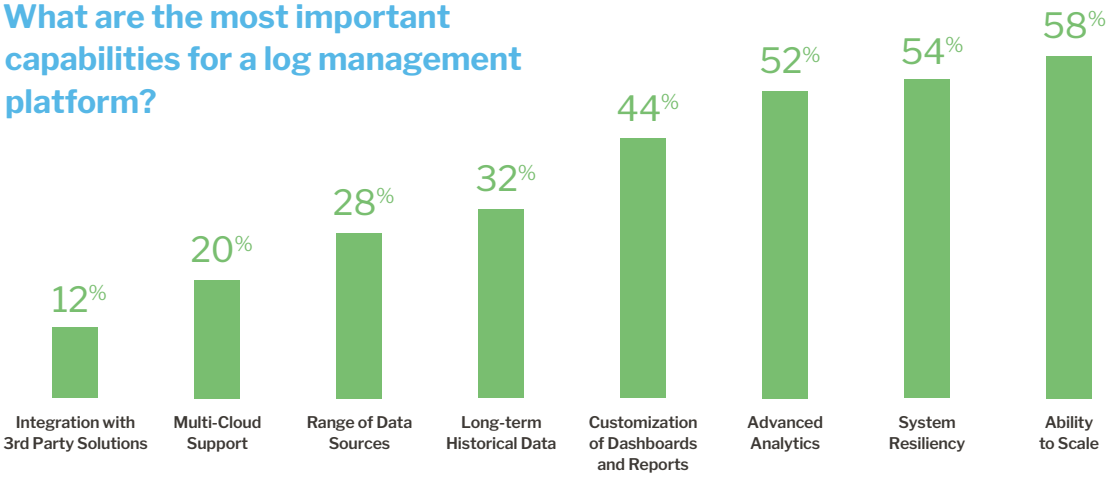


Log Management and Analytics Platform

Discussions with participants revealed that while satisfaction with their existing solution is generally high, many of them expect to bring in a new solution for log management and analytics within the next 3 years. Drivers include the need to scale better and the need to move toward an all-cloud solution.

When asked to name the most important capabilities for a log management platform, participants identified scalability (58%), resilience (54%), and advanced analytics (52%) as the top three.

What are the most important capabilities for a log management platform?



8. METHODOLOGY AND DEMOGRAPHICS

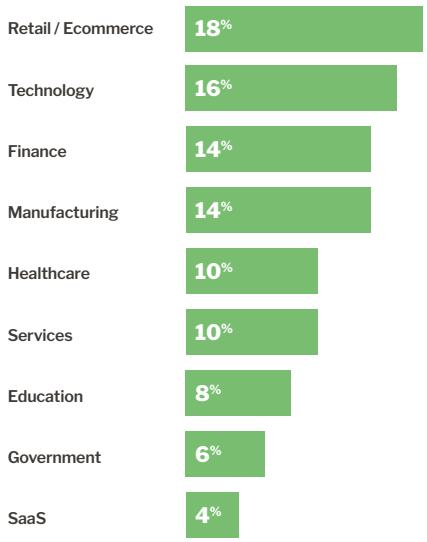
This report is based on results of a comprehensive phone-based survey of IT professionals that are responsible for deploying and managing their organizations' log management solution(s), conducted in February and March 2021.

The respondents include both executives and hand-on practitioners. The survey includes a mix of mid-to-large enterprise companies across a range of industries (F1000 or equivalent), as well as younger, fast-growing companies.

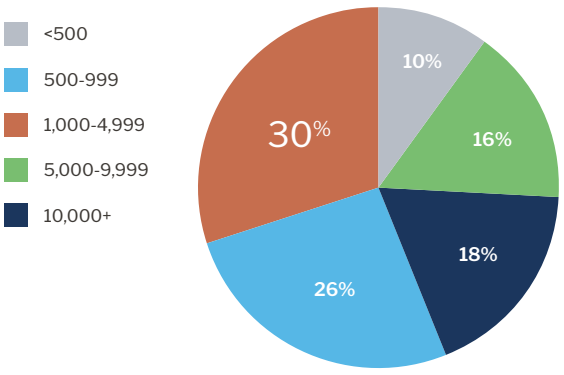
Industry and Organization Size

The chart on the left demonstrates that our survey spanned a wide range of industry verticals, with retail, tech, finance and manufacturing as the top 4 represented industries. Given that the amount of log data roughly correlates with an organization's size, the survey targeted medium-to-large enterprise companies (or equivalent sized organizations). As the chart on the right shows, 64% of participating organizations have at least 1000 employees, and 34% have at least 5000.

What is the industry of your organization?



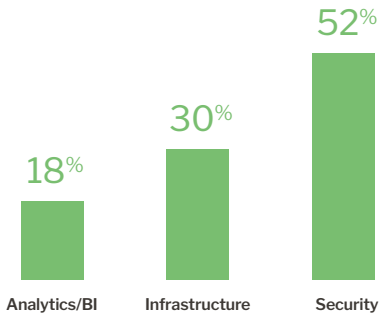
What is the size of your organization?



Primary Role

Participants were selected based on their responsibility for the log management function in their organization. While most participants wear many hats, they were asked to identify just one primary role, reflected in the chart below.

What is your responsibility for the log management function in your organization?



9. CONCLUSION

This research provides a snapshot of log management and analytics today, confirming that organizations continue to recognize the value of their log data, and indeed see opportunities to invest further.

Companies plan to add more data sources, capture and retain more data for longer durations, and are looking for ways to scale their systems to enable the growth. While they will continue to invest in the core operations that log management traditionally serves—security and IT monitoring—most companies plan to expand. Others are finding interesting ways to leverage their log data today, increasing market competitiveness, customer satisfaction, and ultimately, top line results.

ABOUT CHAOSSEARCH

ChaosSearch enables customers to Know Better™, delivering data insights at scale while achieving the true promise of data lake economics. The ChaosSearch Data Platform connects to and indexes data within a customer's cloud storage environment, rendering it fully searchable and available for analysis with existing data tools – all with unlimited scale, industry-leading resiliency, and massive cost savings.

Based on these capabilities, ChaosSearch is an ideal replacement for the commonly deployed ELK stack today. With ChaosSearch, customers can perform scalable log analytics on AWS S3, using the familiar ElasticSearch API for queries, and Kibana for log analytics and visualizations while reducing costs and improving analytical capabilities.

We'd like to hear from you about your log management challenges and priorities. For any questions or requests, or to simply learn more, visit us online or send us an email.

info@chaossearch.com | **www.chaossearch.io**