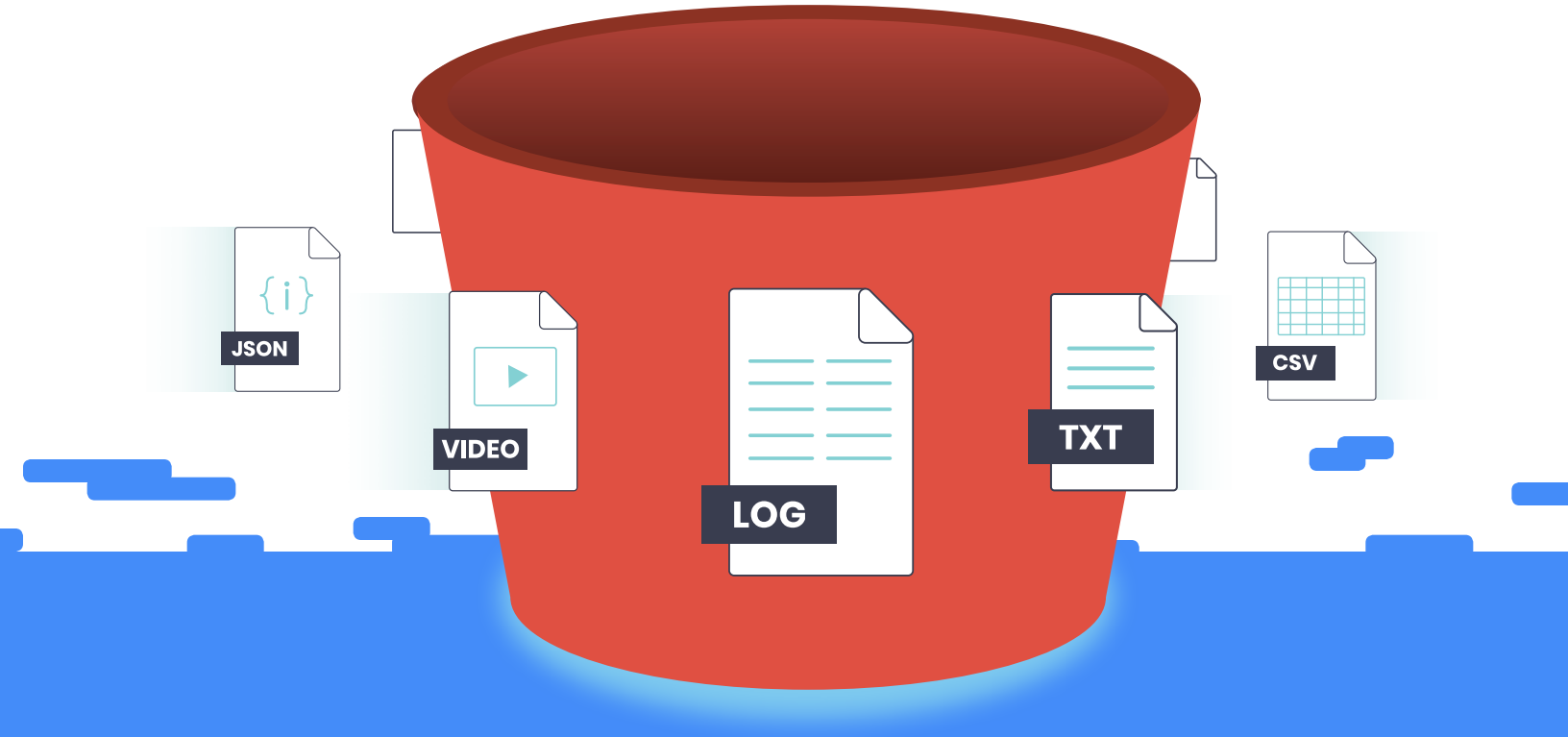
2018

STATE OF OBJECT STORAGE

The Emergence of the AWS S3 Data Lake



Industry Report on AWS S3 Blind Spots;
Historical Trend Analytics and Machine Learning
Emerging as S3 Data Lake Use Cases

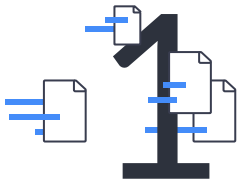
CHAOSSUMO.IO

SURVEY BACKGROUND

Chaos Sumo, emerging leader in object storage management and analytics, conducted an industry survey collecting over 120 responses among data science, analytics, engineering and DevOps/ IT professionals from a wide range of companies and industries. The survey was conducted from December 2017 to January 2018.

MAJOR REPORT FINDINGS

The growth in object storage such as Amazon S3 has surprised everyone and is now the preferred platform for collecting, analyzing and archiving today's growing mountain of diverse and disjointed data. With its simplistic, elastic, and cost efficient characteristics, object storage is fostering new trends in data lake architectures.

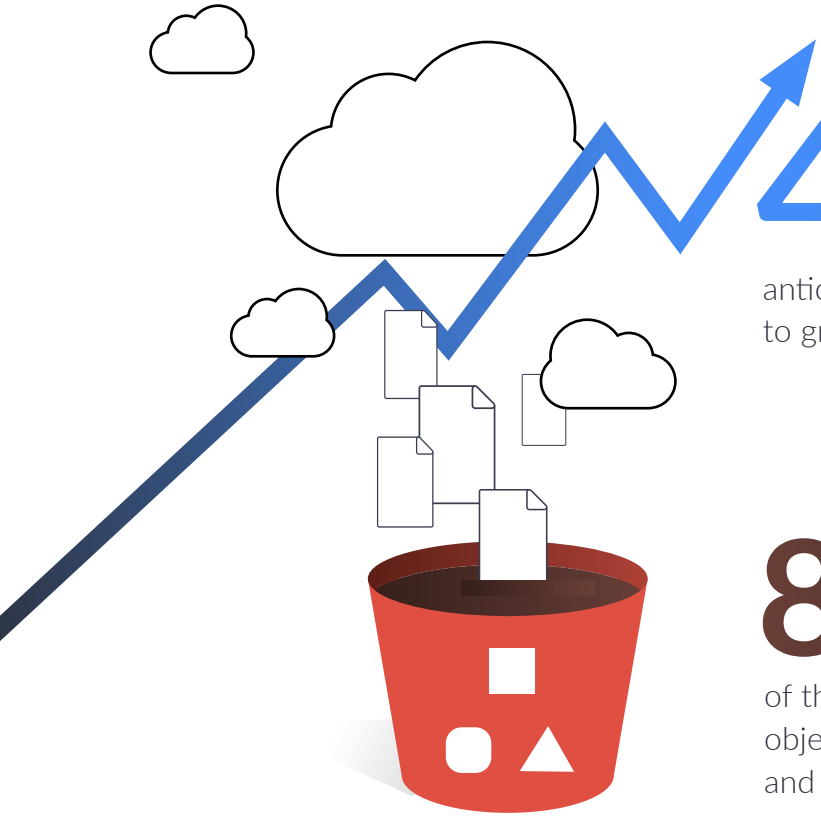


OBJECT STORAGE HAS INDEED GONE MAINSTREAM

Mainly as a cheap storage alternative but increasingly for application hosting and business analytics.

72%

of respondents say they use AWS S3 or another form of cloud-based object storage today.



40%

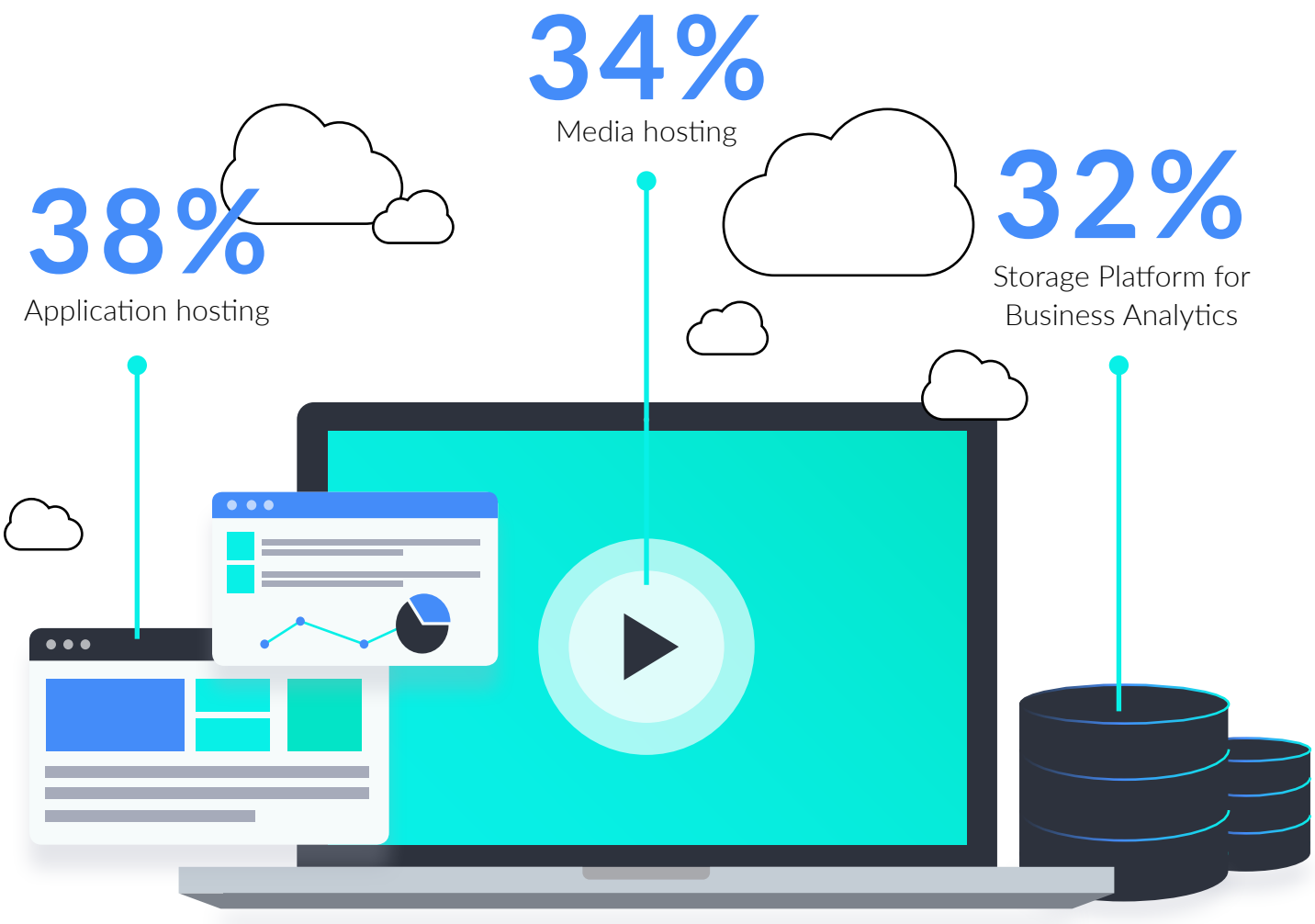
anticipate their investment in S3/object storage to grow over 50% in the next 12 months.

83%

of the respondents stated they use object storage mostly for backup, storage and archiving.

BUSINESS ANALYTICS IS BECOMING A STANDARD OBJECT STORAGE USE CASE

It is also gaining equal footing with application and media hosting. The survey respondents use AWS S3 for...

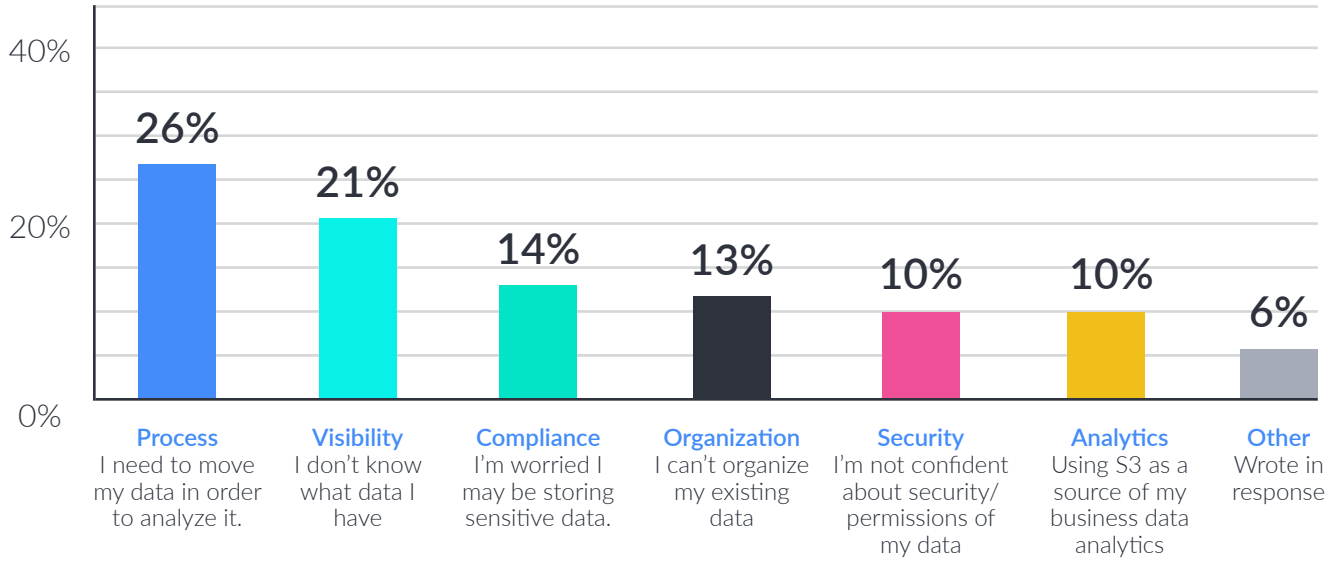


THE INDUSTRY IS FACING MOMENTOUS CHALLENGES

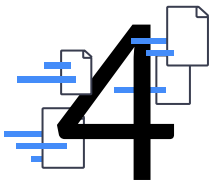
These challenges include cataloging and organizing the tsunami of data going into object storage, as well as the issues of normalizing and modeling this content for analytics and business insights.

As a new hot technology enjoying massive adoption, the biggest challenges with object storage and particularly S3 are visibility into the stored data and the ability to analyze the data right in S3, followed by concerns around increasing storage/compute/network costs as you expand your S3/object storage data, with **37%** sharing they are worried about increased costs.

What is your biggest challenge in using and managing AWS S3 / object storage today? (Select one)



The current inability of businesses to perform consistent, reliable, longitudinal and easy trend and predictive analysis in the object storage space (including log analytics) leads to business information being thrown away or archived in an inaccessible manner. One hidden culprit - the growing costs of storing data for real- or near-time analysis, is luring as the core impediment to doing more with the growing amount of data stored in object storage such as AWS S3.



OBJECT STORAGE HAS DATA VISIBILITY AND ANALYTICS ISSUES

These issues that are partly associated with object storage can be attributed to the myriad of analytics tools that only do part of the job, and never all of it.

42% share they are using home-grown solutions

while others quote using tools such as...

51%

RedShift

23%

Amazon Athena

7%

Elastic Logstash Kibana (ELK)

Object storage solutions that help to surface and virtualize that data will in turn make S3 the place for cost-effective trend analytics on critical event and log data and the promising new place to collect data for the much needed machine learning that will power up increased automation in the enterprise. Specifically for ELK users, the prohibitive storage costs associated with exabytes of data in S3 are compounded with additional costs, effort and resources needed for its scaffolding, which renders most of this rich data inaccessible.

5 THESE TOOLS ARE NOT ONLY INADEQUATE AT ADDRESSING THE JOBS THAT NEED TO BE COMPLETED

They also take a lot of time to set up and manage.

52%

of respondents share it took them more than **3 months** to build their current analytics architecture.

For most data management, science, engineering and infrastructure teams...

57%

of them says it takes more than six hours per week to clean, organize, and prepare data for further analysis.

37%

claims it is taking their teams more than 11 hours per week.

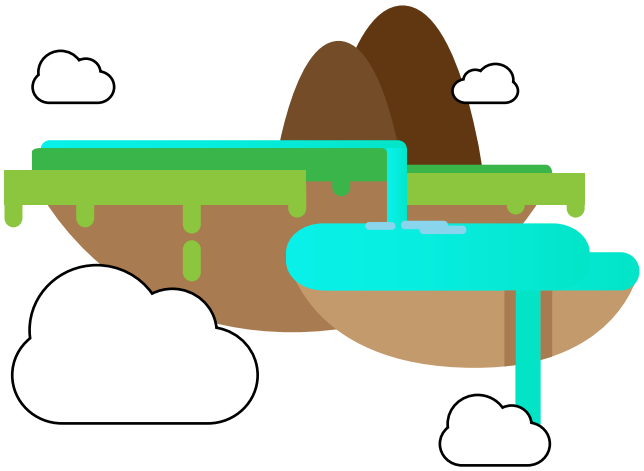
Overcoming these visibility and analytics barriers in object storage would be critical for those planning to use object storage as a repository for business analysis - which is what 50% of the respondents are aiming to use S3 for in the next year.

THE FUTURE OF OBJECT STORAGE

Streamlining and enabling data lake use cases for historical trend analysis and machine learning.

On the quest to developing scalable solutions that provide companies with the ability to perform deep analysis on such information, enabling business to optimize their operations and automate procedures using machine learning, data lakes are slowly gaining momentum within the enterprise.

28%
of respondents report having data lakes today



18%
plans to implement one in the next 12-18 months

Among the top challenges for launching data lake initiatives are the barriers to accessing and analyzing the data in them today -

only 36%
can access such data today

only 7%
say it is easy to analyze such data today

The ability to reduce data retention costs by 10x, streamline archiving information to object storage, simplify data and log archive management, while providing the same querying and visualization tooling used in today's offerings, will emerge as be a key enabler for enterprises to leverage the innovation that object storage avails. Such capabilities will enable enterprises to reimagine long-term, historical data management and data analytics based on object storage -- with an underlying data lake religion.

ABOUT

Chaos Sumo is a cloud-native data analytics service, enabling automated and cost-bending data scaling and long-term log and event data retention on AWS S3. Chaos Sumo extends the Elastic Stack (ELK) by automating the discovery, normalization and indexing of all of your log data types and sources. The service enables historical trend and machine learning analytics at a fraction of the cost of alternative solutions, and provides the ability to perform both relational and text-based analysis through a single integrated Kibana interface. Log data can be organized, managed, indexed and analyzed directly via REST-based S3 and Elasticsearch APIs, delivering value in minutes and enabling your DevOps teams, data engineers and data analysts to be more productive.

For more information and a list of career opportunities, please visit www.chaossumo.io or find us on Twitter, Facebook, or LinkedIn.