

CHAOSSEARCH



TRANSFORMING DATA INTO INFORMATION

How to Make Refining Data as Affordable as Generating It

It's been said that data is the new oil and, when refined, can fuel today's information economy. The dilemma: as data volumes continue to grow exponentially, refining data has become incredibly expensive and complex.

CONTENTS

Introduction	3
The Issue	3
Existing Technologies Are Obsolete	4
The Solution.	5
The Benefits	6
Reduced Complexity	6
Reduced Costs	7
Conclusion	7

INTRODUCTION

It's been said that data is the new oil and, when refined, can fuel today's information economy.

The dilemma: as data volumes continue to grow exponentially, refining data has become incredibly expensive and complex. In this eBook, we'll explore this challenge and introduce a radical, disruptive new approach to extracting value and business intelligence from data that has the power to completely change the trajectory of the information economy: ChaosSearch. Leveraging cloud object storage such as AWS S3 and Google Cloud Storage (GCS), the ChaosSearch Data Lake Platform indexes customers' cloud data, making it completely searchable and primed for large-scale analytics. ChaosSearch supports text search and relational analytics on a single data set, scaling and simplifying data management to turn data into information—all while drastically reducing time, cost and complexity.

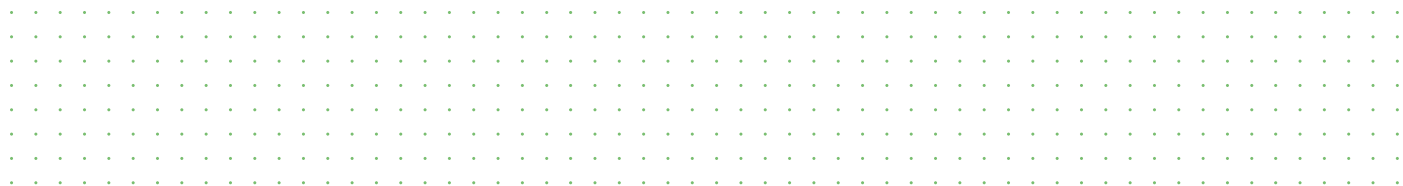
THE ISSUE: Data is Cheap. Information is Expensive.

With the advent of the Internet, the cloud, and all things connected, data has become the lifeblood of companies' external communication, internal operations, and overall business value. To keep everything running smoothly, data must be stored and analyzed to optimize processes, inform decision-making, and deliver competitive advantage.

The challenge is in refining raw data into valuable information. Data becomes information when questions can be asked of it and insights can be derived from it. Yet the more data there is, the harder and more expensive it is to refine.

The volume of data is exploding beyond the ability of human minds to fathom it. It's estimated that data will grow from 33 zettabytes in 2018 to 175 zettabytes by 2025, according to the most recent [Seagate/IDC report](#), "Rethink Data: Put More of Your Business Data to Work—From Edge to Cloud". The reason is simple: data can easily be generated via cheap compute and easily saved via cheap storage. There's simply no cost barrier to generating ever larger mountains of data.

However, the cost of transforming data into accessible/actionable information has been enormous. Refining data is much more expensive than generating data. And as data increases in volume, variety, and velocity, the amount of compute and storage required to extract value from it increases exponentially.



EXISTING TECHNOLOGIES ARE OBSOLETE

There are some technologies and databases that have tried to address the cost of extracting value from data. A variety of computer science algorithms/structures have been developed to implement advanced database solutions, but they all pretty much do the same thing: store data in a way that analysis can be performed more efficiently than manually analyzing the raw source.

Combining compute and storage, these solutions require data to be moved into systems where it can be algorithmically refined for optimized access. Specific types of database solutions solve different analytic problems: text search databases are used when searching for a needle in a haystack, and relational databases are employed when data relationships must be correlated.

While these approaches yield some success, the compute and costs associated with these databases can still be intense, especially as companies often need to use several of these databases in concert.

With the growth in volume, variety, and velocity, these solutions are simply no longer viable or affordable. The capacity to generate data will always outpace the capacity to analyze it. In other words, the “cost of a question” continues to rise as data volumes continue to grow. To truly reduce the cost of refining data into information, innovation is needed.

The capacity to generate data will always outpace the capacity to analyze it.



THE SOLUTION: ChaosSearch Transforms Cloud Object Storage into an Active, Multi-Model Data Lake

To address this dilemma, the team at ChaosSearch has built a new data lake platform from the ground up that is based 100% on cloud object storage. Our UltraHot™ universal data format (index) and associated architecture allows for direct and accelerated analytics on cloud object storage.

We started with the simple goal of making information smaller. If information is at its theoretical minimum, you can store less, move less, and process less—which are all great attributes when trying to reduce the cost and complexity of big data.

To illustrate this point, we've highlighted the metrics of indexing Elastic Load Balancer (ELB) logs—a common use case. The table below demonstrates both the ChaosSearch and Elasticsearch results.

ELB (1.35 BILLION)	CHAOSSEARCH	ELASTICSEARCH
Index – Time	1.5 Hours	4.2 Days
Index – Size	150 GB	2.1 TB
Index – Cost	c4.2xlarge (<\$1.00)	c4.2xlarge (\$38.00)

Our solution is a powerful new data lake platform that reduces information beyond today's compression algorithm ratios. It's a new file format that supports search and relational analytics and removes the need for manually partitioning data at scale. It's a single, intelligent, and holistic solution for all data management and analytics needs that provides:



The ability to simply, quickly, and inexpensively store all data at any scale.

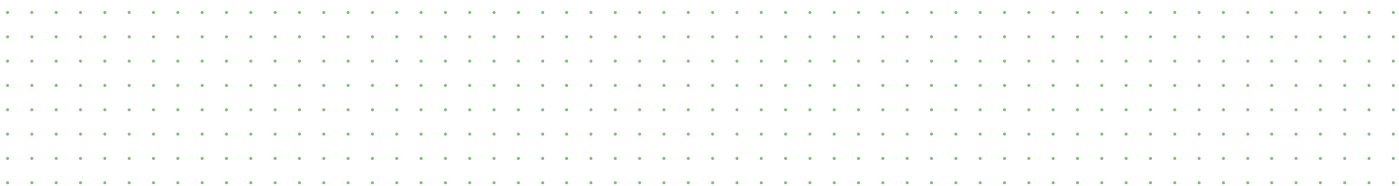


The removal of complexity and external systems for management/analytics.



Data management and analytics at a disruptive price.

ChaosSearch boils down to the idea that anyone can store everything and ask anything of their own data. It's not just an evolution in storage and analytics convergence—it's a revolution in business opportunities.



THE BENEFITS: Adios, Cost and Complexity.

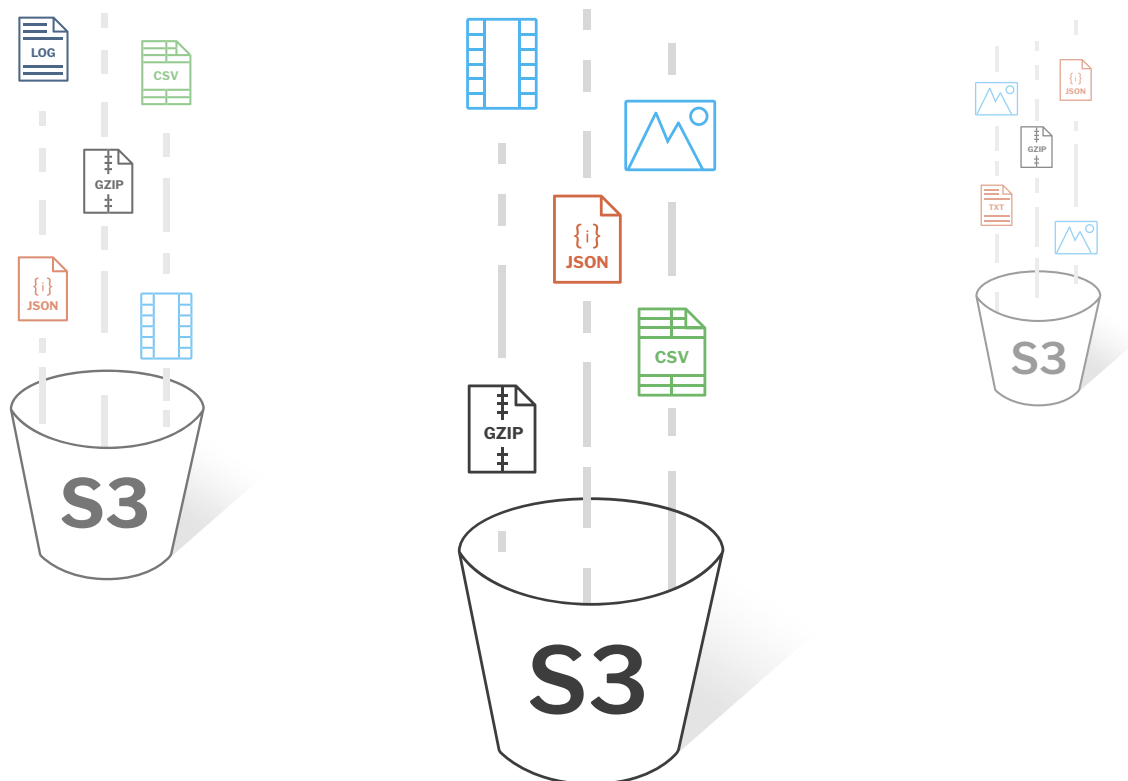
Reduced Complexity

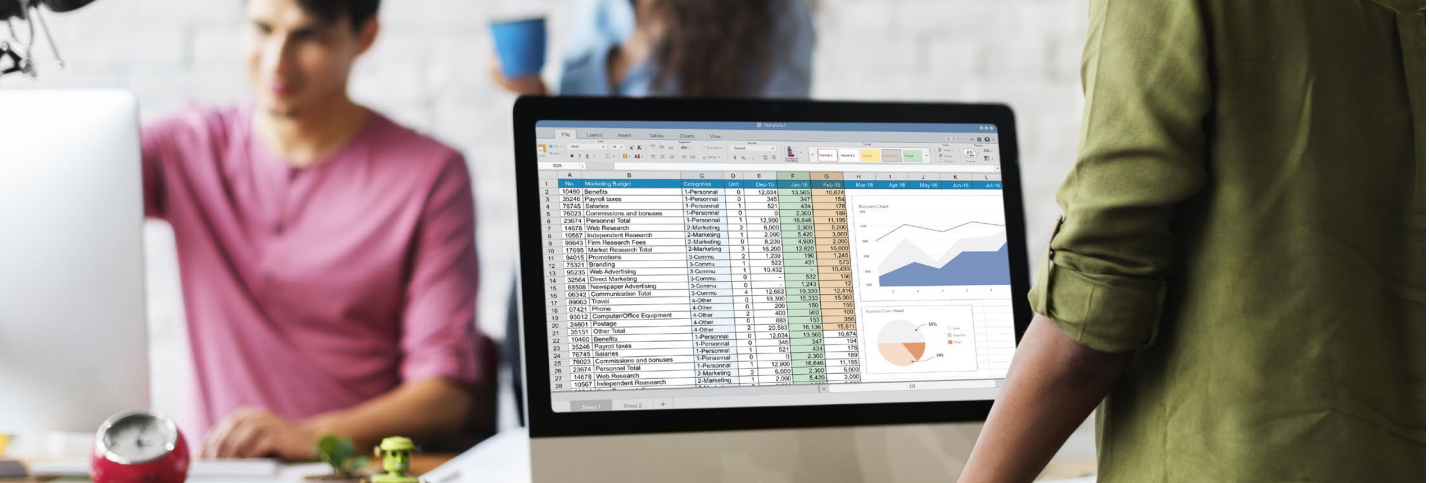
One of the unique benefits of ChaosSearch is its ability to transform “your” cloud object storage into a single, unified search and analytic database.

Over the last 10 years, cloud object storage has become the go-to solution for storing the tsunamis of log and event data that companies are generating. But to analyze it and realize its value, companies must extract, transform, and load (ETL) data into external database solutions or pay for analysis from log management services—and the cost and complexity for either approach is high. Standing up a database such as Elasticsearch, Logstash, Kibana (ELK) may help reduce some costs. Offloading management to a 3rd party solution might address complexity, but it will break the bank at scale. That’s where the benefits of ChaosSearch come in.

With ChaosSearch there is no data movement, transformation, or schema management. (Anyone who has ever built or used a log management system knows how much headache and frustration this can eliminate.) ChaosSearch streamlines and automates data management in Amazon S3 and GCS, dynamically and seamlessly discovering, cataloging, and indexing your data—regardless of size and type—without ever moving your data. The Chaos Refinery® allows for “virtual” data transformation without the need to manipulate or move the underlying original log data, accelerating time to actionable analytical insights. And since ChaosSearch is a managed service, all the complexity of rolling your own stack or ETL-ing to a paid service is eliminated.

Also, because ChaosSearch leverages cloud object storage, customers can take advantage of the inherent storage capabilities such as data replication, high availability, and 11-9’s of data durability. This ensures that your data is redundant, secure, and always available.





Unlimited

Data retention with access to months and years of log data

80%

Up to 80% less expensive than a comparable ELK cluster

~~**Movement**~~

No data movement or external storage of your private data

Reduced Costs

ChaosSearch has taken a new approach to search and analytics data management. Making storage a first-class citizen has eliminated complexity. And starting with object storage as the “only” backing store (there are no HDDs or SSDs involved) has drastically reduced costs. Our ability to transform cloud object storage into a search and analytic database or data lake with built-in data management features removes all the external scaffolding required to extract, transform, and load data.

The ability to streamline and automate data management and analytics while providing cost savings beyond data-under-management pricing is a game changer. ChaosSearch pricing is disruptive, but the platform and philosophy are even more disruptive.

When information is less expensive, companies can view their businesses in a brand new way. By unlocking the value of mountains of data stored in S3 and GCS, ChaosSearch enables organizations to increase operational efficiency, build competitive advantage, and envision new universes of business opportunity.

CONCLUSION

When information is less expensive, companies can view their businesses in a brand new way. By unlocking the value of mountains of data stored in S3, ChaosSearch enables organizations to increase operational efficiency, build competitive advantage, and envision new universes of business opportunity.

That’s what ChaosSearch can do.

For more information on ChaosSearch, contact us at ChaosSearch.io/contact.

ABOUT CHAOSSEARCH

ChaosSearch empowers data-driven businesses like Blackboard, Equifax, and Klarna to Know Better™, delivering data insights at scale while fulfilling the true promise of data lake economics. The ChaosSearch Data Lake Platform indexes a customer's cloud data, rendering it fully searchable and enabling data analytics at scale with massive reductions of time, cost and complexity. The Boston-based company raised \$40M Series B in December 2020 and is hiring to support its hyper growth.

For more information, visit ChaosSearch.io or follow us on Twitter [@ChaosSearch](https://twitter.com/ChaosSearch) and LinkedIn.

info@chaossearch.com | www.chaossearch.io