Transforming data into information.

How to make refining data as affordable as generating it.

CHAOSSEARCH

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 economics of the information economy: **CHAOS**SEARCH. Leveraging Amazon S3, this multi-model data technology supports text search and relational analytics on a single data set, scaling and simplifying data management to turn data into information with high performance at low cost.

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 IDC and Seagate report. 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 Solution:
CHAOSSEARCH is a data platform technology that leverages Amazon S3.

To address this dilemma, the team at **CHAOS**SEARCH has built a new distributed database from the ground up that is based 100% on Amazon S3 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 **CHAOS**SEARCH and Elasticsearch results.

ELB (1.35 billion)	**CHAOS**SEARCH	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 database model 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.

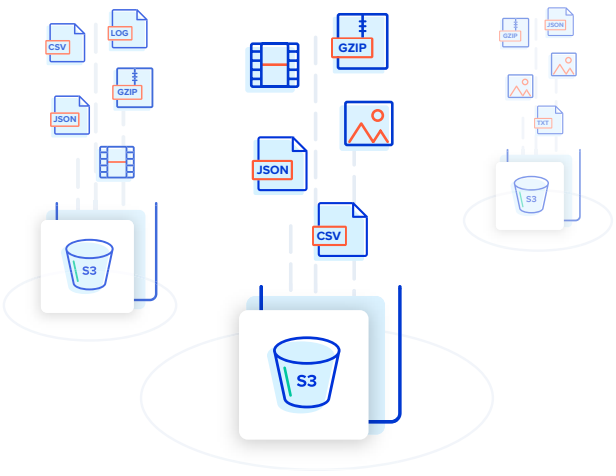
The Benefits:
Adios, cost and complexity.

Reduced Complexity

One of the unique benefits of **CHAOS**SEARCH is its ability to transform “your” S3 storage into a search and analytic database.

Over the last 10 years, S3 has become the go-to solution for storing the tsunami 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 **CHAOS**SEARCH come in.

With **CHAOS**SEARCH 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.) **CHAOS**SEARCH streamlines and automates data management in Amazon S3, dynamically and seamlessly discovering, cataloging, and indexing your data — regardless of size and type — without ever moving your data. The process of converting raw data



into actionable analytical insights is not only accelerated, but completely done for you. And since **CHAOS**SEARCH is a managed service, all the complexity of rolling your own stack or ETL-ing to a paid service is eliminated.

Also, because **CHAOS**SEARCH leverages AWS S3, customers can take advantage of the inherent S3 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.

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 S3 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. **CHAOS**SEARCH pricing is disruptive, but the platform and philosophy are even more disruptive.

Unlimited

Data retention with access to months and years of log data

50%

Half the cost of a comparable ELK cluster

No movement

No data movement or external storage of your private data

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, **CHAOS**SEARCH enables organizations to increase operational efficiency, build competitive advantage, and envision new universes of business opportunity.

That's what **CHAOS**SEARCH **can do.**

For more information on **CHAOS**SEARCH, contact us at **chaossearch.io/contact**.

About CHAOSSEARCH

CHAOSSEARCH is the creator of the first solution to make cloud storage searchable. The platform enables organizations to store, search, query, and visualize terabytes of data within their own Amazon S3 infrastructure with disruptive cost economics. **CHAOS**SEARCH gives customers a multi-tenant service model designed to accelerate innovation and drive business growth.

Founded in 2017, **CHAOS**SEARCH is a privately held, Boston, MA, company backed by .406 Ventures, Glasswing Ventures, and Stage 1 Ventures. For more information, visit us at chaossearch.io or follow us @CHAOSSEARCH.

CS-EB-03