CHAOSSEARCH

WHITE PAPER

Total Cost of Ownership to Build and Operate an ELK Stack

TABLE OF CONTENTS

01 EXECUTIVE SUMMARY

Managing log data has never been more important or complex. With modern applications routinely generating terabytes of machine data per month, businesses urgently need an effective and affordable way to monitor and analyze log data at scale. Many IT organizations use the Elasticsearch-Logstash-Kibana (ELK) stack tool set for this, thinking that because it's open source, it's simple and cost-effective. Unfortunately most of these organizations go on to find that this is actually not the case.

While it may be reasonably easy to download the ELK stack and stand up a simple deployment, the solution is not free. As your data grows, so do your commitments and costs – often to surprisingly high levels, due to the complexity associated with it's highly distributed architecture in which its data is partitioned and stored across numerous shards and separate servers responsible for their portion of the data to support scalabilityThis whitepaper offers a path to estimate the true cost of building and maintaining a do-it-yourself (DIY) production ELK stack.

By understanding how these costs are generated and what the Total Cost of Ownership (TCO) is, you can make informed decisions about whether to embark on – or grow – an ELK stack deployment. You'll also be able to more accurately compare an ELK stack's TCO with that of commercial solutions, and determine the wisest investment for your organization.

02 WHY LOG DATA IS VITAL TO BUSINESS

Businesses simply can't afford their systems and applications to sputter or become inaccessible. These occurrences can damage your brand, stifle revenue streams, and result in significant lost opportunity. To minimize and avoid these problems altogether, log data should be monitored.

Virtually all systems and applications produce logs. Organizations that monitor and analyze logs on a continual basis can proactively ensure their systems and applications are performant during peak usage. When implemented effectively, log analysis can help prevent disruptions, optimize operational performance, minimize security vulnerabilities, ensure regulatory compliance, reduce required user support, better understand customer usage, and improve the bottom line.

03 THE ELK STACK AND ITS CHALLENGES

While there are a number of log analysis approaches that organizations can choose from today, many have opted for installing and managing their own ELK stack. As its name implies, the ELK stack is comprised primarily of these three separate open source projects:

- **E**lasticsearch - a search database
- **L**ogstash - a log ingestion and processing pipeline
- **K**ibana - a visualization tool for log search analytics

Some people also include Beats in the list of core ELK open source projects. Beats is a set of agents that collect and send data to Logstash. In addition, organizations can pay for a subscription to components that extend the ELK stack's capabilities to include security, alerting, reporting and graphs.

The ELK stack can be deployed on-premise or in the cloud. Marketed as a free, easily accessible, and simple-to-install toolkit, it gets tens of thousands of downloads every month. However, while the software itself is free, running it is not, and that's where the challenge of predicting the true TCO lies. The infrastructure required to support growing volumes of log data and the staffing needs required for ongoing tuning, configuration, and patching often result in a TCO

that is higher than expected by organizations with ELK stack deployments. This paper and calculator have been created to help you reliably predict a true ELK stack TCO.

04 ELK STACK TCO ANALYSIS

Our TCO analysis is based on a the following scenario:

- Year 1 daily log data ingest volume of 500GB
- Year 2 daily log data ingest volume of 875GB
- Year 3 daily log data ingest volume of 1.5TB
- Annual growth rate of 75%
- 12 months searchable data (60 days active, 305 days inactive)
- Industry standard ELK configuration
    - Replication factor of 1
    - All in a single cluster
    - 1 FTE for every 5TB of daily data ingested

Note: This analysis leverages Amazon's Elasticsearch Sizing guidelines.

To run an ELK stack that meets typically acceptable SLAs in a production environment, you need to carefully consider all known and hidden costs. For simplicity's sake, we'll bucket these costs into two categories: Infrastructure, Initial Build and Ongoing Operations. Let's dive into each of these 2 categories:

1. Infrastructure

The first cost to consider when calculating TCO is the infrastructure you need to run the ELK stack. Since the infrastructure varies based on the amount of data you expect to generate, estimating your log data volume accurately is very important. If you over-provision, you waste money. But if you under-provision, you can lose log data, miss critical insights and impede business performance.

Here are the key factors to consider when estimating your capacity needs:

- Log volume generated daily by your applications, systems, and networks.
- When/if your organization typically experiences spikes in log volume. You need to ensure that your environment can scale with ease so that influxes of log data don't become bottlenecks.
- How long you need to retain log data for 1) indexing and 2) archiving.
- How much your log volume will grow year over year.

In addition to the above, it's also important to estimate:

- The number of concurrent users and concurrent searches.
- Fault-tolerance and redundancy requirements. A 1:1 ratio between server and a replicated backup is generally considered best practice.
- What additional infrastructure will be required to expand the stack over time.
- How many additional servers you'll need and their configurations, e.g., processor class, memory, storage. You should expect to add servers as log volumes grow to maintain search performance.

Keep in mind that Elasticsearch and Logstash require significant capacity, availability, and redundancy. Because of this, you should run Logstash and Elasticsearch on different and multiple servers. Kibana requires high availability so that users can reliably interact with and perform analysis on log data. To ensure that you don't slow performance for the user, it's also best to run Kibana on dedicated servers.

Infrastructure Cost Breakdown

The cost of deploying an ELK stack on-premise or in the cloud like an Amazon Web Services (AWS) is relatively similar. The main differences are the speed of deployment and on-demand scalability benefits of cloud computing versus the ability to treat on-premise deployments as a capital expenditure. Our analysis focuses on the more typical cloud deployment model.

In our scenario, the costs for infrastructure, based on an observed 25% discount off of the AWS Elasticsearch service published pricing, is as follows:

COST CATEGORY	YEAR 1	YEAR 2	YEAR 3	TOTAL
Hosting	$688,000	1,195,000	$2,065,000	$3,948,000

2. Initial Build and Ongoing Operations

The amount of work needed to build an ELK stack before it goes into production cannot be overstated. Many organizations dedicate full-time employees to handle the complexities inherent in developing and deploying an ELK stack. Its distributed shard-based architecture creates issues of consistency and durability due to the complex dependencies and failure modes across shards, which in turn makes shard management a non-trival operational task. Additionally, it's not uncommon to run into problems with Logstash not running or not shipping data, and Kibana not fetching mappings or not connecting with Elasticsearch.

To build a production-ready solution, your team needs to:

- **Configure the stack to ingest and parse logs** from all logging components – and maintain what could be hundreds of configurations needed to accommodate the large variety of logging frameworks, data formats, and log sources.
- **Build a resilient data pipeline** and ensure that you don't lose log data if your system generates events faster than Elasticsearch can index them. This typically requires placing a buffer in front of Logstash that acts as the entry point for log events. Doing so will enable you to accumulate the data until it can be pushed to Elasticsearch. Some organizations use Apache Kafka, Redis, or RabbitMQ for buffering logs. However, this requires you to host and maintain yet another piece of software.
- **Handle mapping exceptions.** To ensure that Elasticsearch indexes documents instead of returning failure messages and dropping logs that don't fit into the automatically generated mapping, you have to keep log formats consistent and consistently monitor Elasticsearch exceptions.
- **Ensure log data consistency.** Applying relevant parsing abilities to Logstash is critical to ensuring you have correct fields for Elasticsearch and Kibana. But it is also challenging. It's easy to make mistakes using Logstash, so you need to devote time to testing all log configurations before your ELK stack goes into production.
- **Implement monitoring and alerting capabilities** that notify you of performance and potential security issues. You must research and evaluate the many open source and commercial solutions on the market, and then dedicate resources to implementing and integrating them into your ELK stack. And of course, if you choose commercial solutions, you need to factor them into your costs.

Your organization's work on the ELK stack doesn't end after it's rolled out in production. As log data volumes increase – which they will – more resources are consumed and new complexities and issues arise. ELK stack experts must be on hand to respond to these issues and perform the day-to-day maintenance that self-hosted software implementations require. The number of people required to handle it will grow as your stack expands, and potentially cause you to commandeer engineers from other priorities.

This undertaking is as critical as it is enormous, tedious and time-consuming. It includes:

- **Maintaining your infrastructure and planning capacity increases.** It's critical to plan ahead, otherwise your stack is likely to hit the wall and experience failures. Adding or removing servers from an ELK cluster, for instance when data volumes grow, is also a non-trivial task and may require the process of rebalancing shard allocations.
- **Reindexing outdated indices so that you stave off potential failures and log data losses.** Remember that logs are dynamic. Their formats change over time and require configuration adjustments. In addition, the number of indices handled by Elasticsearch impacts performance, so you'll need to remove or freeze old and unused ones.
- **Monitoring cluster health and responding to failures.** You need to track information about the status of the cluster, the number of nodes, and the counts of active shards. Also monitor counts for relocating shards, initializing shards, and unassigned shards. All of this is needed to tune the cluster for better performance.

- **Handling software upgrades.** Upgrading an ELK stack can be a large undertaking, given the size and complexity of its deployment. Even though new versions are released on a regular basis you need to thoroughly research what the changes are for each ELK component before deciding whether to expend the effort required to implement them. To make sure you don't lose data during Elasticsearch upgrades, run tests in a non-production environment first. When upgrading Logstash, pay attention to compatibility between it and Elasticsearch. And if you're upgrading Kibana, it's important to note that plugins often break and visualizations sometimes need total rewrites. Backup your objects first and test the Kibana upgrade process before rolling it out.
- **Support.** Because log and event analysis are critical to business performance, some organizations take advantage of Elastic's professional support to help manage their ELK stack. There are Gold, Platinum and Enterprise options. While Gold is the least expensive, it doesn't provide 24-hour support or emergency patches. Gold and Platinum are priced per-node, meaning support costs grow along with your stack.

Operational Cost Assumptions

A good rule of thumb is to assign one full-time employee (FTE) for every 5 TB of log data ingested daily. As your system grows more complex, Elasticsearch optimization activities and the related tasks of managing your ELK stack will require more resources.

Your full-time employee must be an expert in ELK stack and know how to manage the performance and indices under various production workloads. We assume the cost of a full-time employee is $130,000.

Other costs

Support - Elasticsearch may be free, but support is not. Many enterprises will also see the value in a support plan from Elastic. Elastic offers Basic, Gold, Platinum, and Enterprise subscription plans, which include SLA-based support and dedicated support contacts. If you want 24/7 coverage, then you must factor in this cost as well.

Remember, pricing varies based on the number of nodes you have. So if your system is continuously growing, then your cost of support will grow too.

Training - Elastic has produced many instructional videos on how to use and manage ELK stack. The company also offers in-person and virtual classroom training for both administrators and end users. Such offerings may cost you approximately $5,000 per participant, per training. You can choose to save money by developing and maintaining your own training materials in-house and conducting internal training sessions for the rest of the team, but that would be another task that distracts your ELK experts from managing your production-grade system.

In our scenario, the costs for initial build and ongoing operations is as follows:

COST CATEGORY	YEAR 1	YEAR 2	YEAR 3	TOTAL
Operations	$138,000	$135,000	$135,000	$408,000
Support & other	$76,000	$132,000	$228,000	$436,000
Total	**$214,000**	**$267,000**	**$363,000**	**$844,000**

05 ELK STACK TCO OVER A THREE-YEAR PERIOD

The total cost of ownership for a do-it-yourself ELK stack implementation can easily reach $4 million over the course of three years. This price tag will continue to grow as your logging volume increases due to the complexities inherent in, and resources required to, scale and maintain the stack.

COST CATEGORY	YEAR 1	YEAR 2	YEAR 3	TOTAL
Hosting	$688,000	$1,195,000	$2,065,000	$3,948,000
Operations	$138,000	$135,000	$135,000	$408,000
Support & other	$76,000	$132,000	$228,000	$436,000
Total	**$902,000**	**$1,462,000**	**$2,428,000**	**$4,792,000**

06

PUTTING ELK STACK COSTS IN PERSPECTIVE

While the open source ELK stack software toolkit is free, it requires infrastructure and resources to build, grow, and maintain. Over a three-year period, the cost of hosting, customizing, scaling, and maintaining this increasingly complex infrastructure skyrockets – while the drain and strain on your engineering team grows. That's why many organizations are turning to solutions like ChaosSearch. ChaosSearch is an entirely new approach to log analytics. ChaosSearch is a revolutionary, cloud-native, big data platform built for massively scalable search and analytics at a fraction of the cost and complexity of existing solutions, like the ELK stack. It is a new paradigm and approach that helps businesses retain and extract more value and insight from data, faster and at low cost.

ChaosSearch developed a patent-pending indexing technology deployable directly on cloud object storage as a backing store. ChaosSearch is built as a cloud-native, stateless, API-based data fabric to support search directly on cloud storage. ChaosSearch's mathematical approach to indexing results in indexes which are 25x smaller and 60x faster to build than Elasticsearch, greatly simplifying and reducing the cost of it's deployment. With Elasticsearch API support and an integrated Kibana, ChaosSearch also delivers the functionality of an ELK stack with zero overhead, nothing to install - and at a price up to 80% less expensive.

With ChaosSearch, all the decisions and work that go into building, managing and scaling an ELK stack are handled for you. It scales linearly into 100's of TB of data, and since it's backed by scalable cloud object storage, you never need to worry about data security or integrity. ChaosSearch doesn't require extensive training or a staff of dedicated DBA's.

Unlike ELK stack deployments that can take weeks or months, fully managed ChaosSearch environments are up and running in minutes. The result is accelerated troubleshooting, lower costs, and more time spent focusing on your core business.

Get started now

In addition to significant cost savings, ChaosSearch enables you to get up and running in minutes, not days or weeks. Let us know if you want to learn more.