logdna

THE TRUE COST OF THE ELASTIC STACK

Costs Associated with Running the Elastic Stack in Production

Summary

Considering the various costs associated with hosting, customizing and maintaining a homegrown Elastic Stack, the true cost of this "free" software quickly becomes apparent.

Contact

For more information, please contact us at outreach@logdna.com or visit us at www.logdna.com

View Our Whitepaper

TABLE OF CONTENTS

EXECUTIVE SUMMARY

With modern applications becoming increasingly complex, organizations need comprehensive log management tools to better understand them through access to actionable data. Although several choices exist in the market, the Elastic Stack (also known as the ELK Stack) has emerged as one of the most popular solutions. As a collection of open source products with an extensive feature set, it can be a great option for teams that are just beginning their journey to better log management practices. However, as they quickly scale, teams realize that the Elastic Stack's free price tag comes with many hidden costs.

The Elastic Stack has a low barrier to entry; it is easy to get started. As you grow from proof of concept and start onboarding your users, have you accounted for all of the costs (people, time, infrastructure) and the potential failure scenarios that can result if you fail to take these variables into account?

These costs are measured in a variety of ways such as time, money, and resources. They include considerations involved with infrastructure, consulting, training, and support throughout the Elastic Stack lifecycle. Engineering teams must also spend time learning how to effectively use, as well as constantly maintain the new stack, taking precious time away from engineering priorities. Finally, measuring "failure" costs, or the costs of your service being unavailable to your users, needs to be a part of your planning work.

This whitepaper presents the costs involved in using the Elastic Stack as a log management solution. Understanding the true cost of the Elastic Stack allows organizations to more easily compare it with other options, including commercial solutions such as LogDNA.

Choosing a log management solution is crucial in allowing developers to deliver products faster. It also has big implications on an organization's bottom line. Using our total cost of ownership (TCO) analysis, you can determine for yourself whether the Elastic Stack is the right investment for your organization.

LOG MANAGEMENT IS A MUST FOR MODERN DEVOPS

Modern applications have become more distributed and complex, and as a result generate truly enormous amounts of log data. Logs provide crucial insights into operational performance, software flaws, and security vulnerabilities. When used effectively, the data provided by logs help development and DevOps teams better understand, debug and troubleshoot their applications and systems. This leads to improved products, optimized deployment pipelines, reduced support time, and significant cost savings.

The adoption of microservices, containers, and serverless computing in particular is transforming the way software organizations architect, build and deploy their solutions. This, in turn, has transformed the way these applications generate log data, leading to increased logging complexity and an acute need for better log management solutions. Log management platforms must now be powerful enough to ingest terabytes of log data per day, flexible enough to parse dozens of different log formats, and reliable enough to tolerate failures and unexpected spikes in log volume. They must also be easy to use, and empower developers to "self serve" their access to the logs that are generated by their applications, without slowing down their development cycles through context switching.

Organizations have several options for modern log management, including a mix of open source and commercial solutions. Each presents its own unique benefits and challenges. Historically, the open source Elastic Stack has been a popular option due to its free cost and extensibility. However, many teams who build and maintain their own Elastic Stack soon realize that its perceived benefits are greatly overshadowed by its shortcomings, especially as they begin to scale their efforts.

WHAT IS THE ELASTIC STACK?

Originally known as the ELK Stack, the Elastic Stack is comprised of four independent open source projects:

- Elasticsearch, a search and analytics engine
- Logstash, a log ingestion and processing pipeline
- Kibana, a data visualization tool for Elasticsearch
- FileBeat, a set of agents that collect and send data to Logstash

Deployed together, these projects create a basic approach to collecting, ingesting, and visualizing logs. There are a number of companies that offer additional functionality via commercial packages and add-ons. Additionally, each project can be further extended using third-party add-ons, plugins, and libraries.

The Elastic Stack's open source nature and ability to run both on-premise and in the cloud has made it a popular option for companies seeking a log management solution at a low cost. However, there are many long-term and hidden costs associated with the Elastic Stack that are not obvious at first glance. Organizations looking to implement a log management solution should be aware of these costs, both from a budgetary and time/resource perspective, before considering the Elastic Stack as a possible option.

THE COST OF AN ELASTIC STACK DEPLOYMENT

The costs associated with an Elastic Stack deployment can be split into four distinct categories:

- Infrastructure
- Customization
- Operations
- Support

The following sections will explore each of these in depth, and what costs are typically associated with them.

CATEGORY ONE

INFRASTRUCTURE: HOSTING THE ELASTIC STACK

The exact infrastructure requirement for the Elastic Stack will vary depending on your organization's logging requirements. When estimating your overall capacity, you should consider factors associated both with capacity and deployment.

Capacity refers to the ability to collect and store log data in a meaningful way.
Topics to consider here include:

- The daily log volume of all of your applications and servers
- The growth of your total log volume for each year (either historical or anticipated)
- The amount of time you wish to retain your logs for indexing and archiving
- The ability to handle spikes and surges in log volume

When considering deployment of an Elastic Stack, it's important to take these factors into account:

- The location of your logging components relative to your logging infrastructure
- The number of engineers who will be accessing your logging infrastructure
- The ease of upgrading and expanding your stack over time
- Built-in fault tolerance and redundancy for when your stack experiences failures

Remember, the Elastic Stack is made up of various components. Elasticsearch (the mechanism that stores and indexes your log data) along with Logstash (the log ingestion engine) require high levels of availability, capacity, and redundancy to prevent log data from being dropped or deleted. Kibana uses fewer resources, but still requires high availability as it is the primary tool for users to interact with log data and to perform log analysis. If your log management stack goes down, then you lose visibility into what's happening at every level of your application. Downtime associated with any of these components presents serious risk not only to your pipeline, but more importantly to your business as a whole.

When calculating the expected costs of hosting (along with the subsequent costs outlined in this whitepaper), we assume a mid-size Elastic Stack deployment on AWS. We will then calculate the costs over a three-year period with the following requirements:

- Enough storage to ingest the average amount of log data an organization requires access to, factoring in regular growth over time
- 30 days of short-term log retention (also known as hot storage)
- 1 year of long-term log retention (or cold storage)

Also included in this cost estimation are nodes provisioned for redundancy and load balancing, as well as extra storage capacity to accommodate spikes.

	Year 1	Year 2	Year 3	Total
Storage	1 TB	5 TB	5 TB	
Number of Nodes	3	5	12	
Infrastructure Hosting Cost	$24,000	$76,000	$182,000	$282,000

CATEGORY TWO

CUSTOMIZATION: ADDING VALUE TO THE ELASTIC STACK

At its core, the Elastic Stack provides a basic mechanism for collecting, indexing, and visualizing logs. However, the real value in log management platforms comes from the ability to make that data insightful and actionable by making sense of everything that is being ingested. For teams, that means extra work customizing their solution in a number of ways before it can be deemed production ready.

First, the stack must be configured to ingest and parse logs from each of your logging components. The enormous variety of logging frameworks, data formats and log shippers means having to maintain dozens to hundreds of different possible configurations. Your stack must be capable of ingesting all of these log types while simultaneously resolving any conflicts between fields. Without this parsing ability, teams would be overwhelmed with the sheer amount of data produced by their logs, without any direction on where to find what they're looking for.

Second, your stack must have a system in place for pipelining logs in the event your logging components generate events faster than Elasticsearch can index them. Logstash provides some queueing capabilities, but surges in log volume could overwhelm your stack, resulting in unresponsive nodes or dropped messages. Many organizations use Apache Kafka as a solution for buffering logs, especially for high-volume surges. This results in additional infrastructure spend, as well as additional support costs to maintain your logging capability.

Finally, you must implement monitoring and alerting to notify the appropriate teams of performance issues or downtime. There are countless solutions available ranging from open source software to commercial services. Researching, implementing, and integrating these services takes time and expertise. In addition, understanding when you get alerted and how to configure your thresholds can only reliably be done with trial and error.

For all of these considerations, solutions don't come ready "out of the box." They require resources to build and configure, which are often full time engineers. We will consider this cost as part of the next section.

	Year 1	Year 2	Year 3	Total
Pipeline Hosting	$30,000	$60,000	$120,000	$210,000
Monitoring / Alerting	$17,000	$34,000	$68,000	$119,000
Total Customization Costs	$47,000	$94,000	$188,000	$329,000

CATEGORY THREE

OPERATIONS: MAINTAINING THE ELASTIC STACK

The Elastic Stack is not a "set it and forget it" solution. Simply considering the above customization needs necessitates a conversation around who will be responsible for building out a solution from scratch. And as you scale your development efforts, log volume increases, and issues such as slower queries and significantly increased resource usage will begin to emerge. Engineering staff must be available to respond to these issues otherwise the stack will experience failures.

With all of these auxiliary needs, many organizations often hire full time employees (often as dedicated Elastic Stack engineers) to help develop, deploy and ultimately maintain the stack. Considering all of the hosting and customization work outlined thus far, hiring one full time employee to achieve this is a conservative estimate. And even if you start with a single dedicated resource, you will find those needs quickly increase as the stack grows in size and complexity.

During the stack's deployment, this resource needs to spend time on:

- Maintaining the infrastructure that the stack is hosted on
- Performing software upgrades and reindexing outdated indices
- Monitoring cluster health and responding to failures
- Planning capacity increases

This results in more engineering resources to continually configure and maintain. With the average salary of an engineer starting at $150,000, that resource cost adds up quickly and adds significant costs. As it scales, the Elastic Stack requires time and money that could have been dedicated to your core business, as both must now be directed towards building and maintaining a complex ancillary system. In an attempt to simplify business operations, you end up creating an entirely new operation with its own staff, requirements, and objectives.

Additionally, another surprise operations cost in maintaining your own log management stack comes from the fact that in doing so you are assuming total risk around it's availability. That means that if the stack is degraded, developers are unable to effectively access that data, which means more time spent trying to locate and resolve issues in real time. This impacts their ability to deliver on their core function—writing new code. We quantify this impact as a "productivity loss," in which development is severely hindered, or in some cases brought to a halt due to a lack of data insights.

To calculate this, we assume that a development team (1 FTE salary at $150,000) averaging 10 releases per week can expect to find about 5 issues per release. We then factor in how long it takes to find and remediate an issue when the log stack is degraded, and subtract that time from the engineer's hourly rate. Note that this calculation is static, does not account for multiple development teams, nor does it account for any annual growth. So this cost can add up incredibly quickly!

	Year 1	Year 2	Year 3	Total
Employee Salary	$150,000	$300,000	$450,000	$900,000
Productivity Loss	$34,000	$34,000	$34,000	$102,000
Total Operations Cost	$184,000	$334,000	$484,000	$1,002,000

CATEGORY FOUR

SUPPORT: HIDDEN COSTS OF THE ELASTIC STACK

Beyond the ongoing operational costs outlined in the previous section, there are some other ongoing (and often hidden) costs when building and maintaining an Elastic Stack. These additional considerations, when not taken into account, can adversely impact the overall goals of the organization in several ways.

To ensure the long term success of introducing a log management solution,employees outside of the Ops team (i.e. development teams) will need to learn how to use the new stack effectively. Some organizations may choose to hire outside services to deliver official training in best practices. But even without training, there is still a cost associated with employees taking time away from their core tasks to learn a new system.

Additionally, maintaining a complete logging stack is a daunting operation. Logs are essential to development and operations teams in providing the insights they need to troubleshoot application errors and production problems. However, there are no set rules for deploying and maintaining the Elastic Stack. It requires a lot of trial and error, which may result in frequent downtime, lost log data, slow query performance, and a longer setup process. All of these make it harder for engineers to resolve problems, resulting in

increased risks around product quality and deployment velocity. Both of these directly impact the businesses' bottom line.

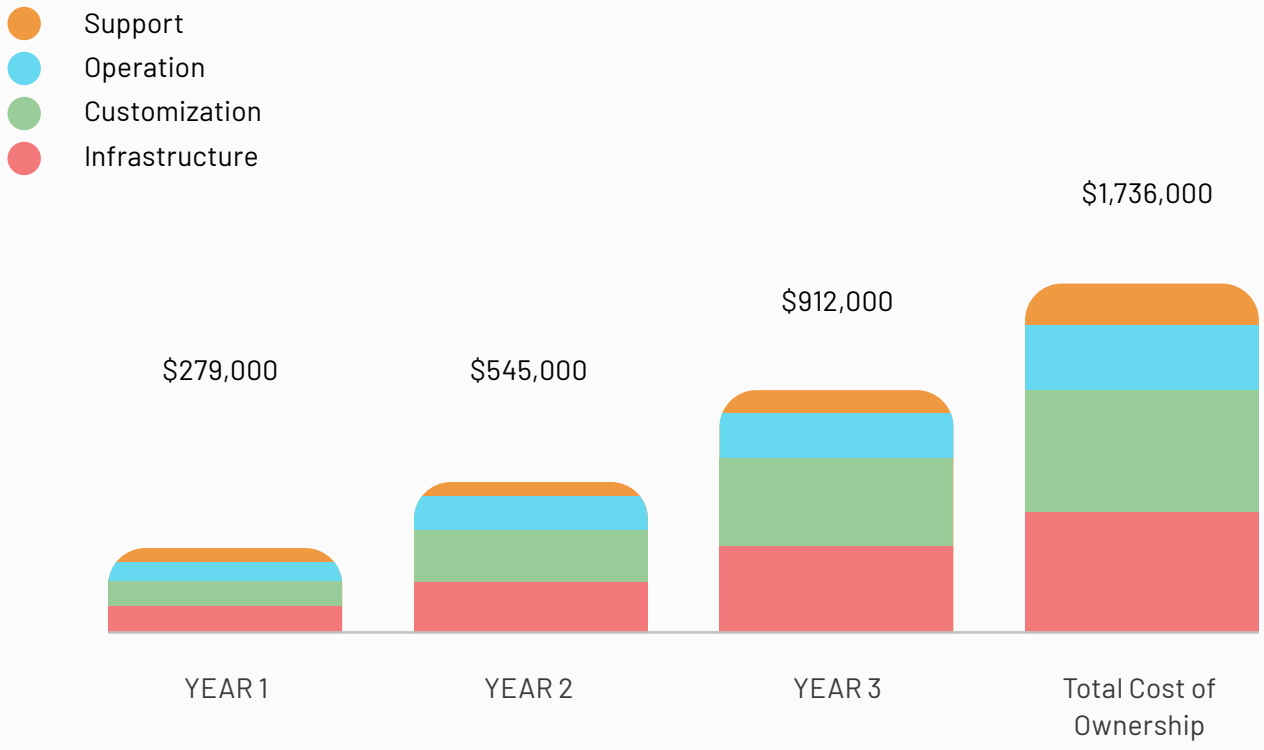
To help mitigate these risks, teams may purchase support packages or professional services engagements that provide dedicated resources to help troubleshoot issues more quickly. While this can provide a low cost safety net as you build the stack, as you scale your solution so will the cost of these ancillary services (which are often priced on a per-node basis).

	Year 1	Year 2	Year 3	Total
Training	$2,000	$4,000	$8,000	$14,000
Elastic Support	$22,000	$37,000	$50,000	$109,000
Total Support Costs	$24,000	$41,000	$58,000	$123,000

As you scale your development efforts, log volume increases, and issues such as slower queries and significantly increased resource usage will begin to emerge. Engineering staff must be available to respond to these issues otherwise the stack will experience failures.

THE THREE YEAR TCO OF THE ELASTIC STACK

Bringing these four cost considerations together, our example shows that the total cost of ownership of the Elastic Stack deployment reaches $1,736,000 in the first three years — and that's with only reaching 1 TB/day in year three. From the costs associated with building infrastructure to host the stack, along with the need to hire and train engineers to continually update and support, it's easy to see how everything adds up so quickly. As time goes on and your logging volume increases, and your requirements become more complex, you can expect these costs to continue to increase.



CONCLUSION

We have seen how the Elastic Stack can grow from a free solution to a $1.7 million solution in just three years.

Considering the various costs associated with hosting, customizing and maintaining a homegrown Elastic Stack, the true cost of this "free" software quickly becomes apparent.

Solutions like the Elastic Stack, when built to scale with the needs of your DevOps teams, leave you with a complex infrastructure, a large demand on engineering resources, and an expensive bill. But perhaps most importantly, it can leave you with an inherent risk to your business. If any of these areas are under resourced, or left out of your budget, it can result in a faulty log management solution. And the downstream effects not only impact your development teams, but ultimately your business as a whole.

Because of this, many organizations that are looking to scale log management beyond the Elastic Stack often look to commercial solutions, such as LogDNA, to provide similar functionality and benefits at a much lower TCO.

With LogDNA, the considerations you would have to make in building your own log management solution are already baked into our platform, resulting in a complete logging solution right out of the box. Our SaaS offering provides a scalable platform that ingests, analyzes and visualizes data to make it more consumable and actionable for your teams. With features that allow for meaningful customization, along with a number of controls that give you power over your log data, LogDNA empowers everyone in your organization with data that makes it easier to identify, remediate, and even proactively prevent issues in your modern applications. And best of all, LogDNA's intuitive interface and simple setup make it the developer's choice for insightful log management, as it does not require extensive onboarding or training efforts. With LogDNA, your teams have real-time access to the most meaningful log data, without the need to disrupt their workflow or take time away from core business functions.

UNLEASH THE POWER OF YOUR LOGS FROM DAY ONE

To see how quickly you can set up LogDNA and start seeing real insights into your log data, sign up for our 14-day free trial. If you are interested in seeing the platform in action, please contact our solutions team.

PR & marketing inquiries:	marketing@logdna.com
Sales inquiries:	sales@logdna.com
Technical inquiries:	tech@logdna.com
Customer service:	outreach@logdna.com