Coralogix

THE COST OF DOING ELK ON YOUR OWN

Contents

Introduction 3

< part 1 >
Breaking Down Cost of Ownership 4

< part 2 >
Estimating Cost of Ownership for the “Free” ELK Stack 5

< part 3 >
Elastic Cloud 12

< part 4 >
ELK Stack and Elastic Cloud TCO Summary 16

< part 5 >
Full-Stack Observability, Without Paying For The Noise 18

<PART 1>

Introduction

Logs have long been the bedrock of modern software observability. No deployment is confirmed without a healthy log output. This reliance on logs has created a challenge of scale.

A mid-size company's applications will output 100-200 GB daily. That's 500,000-1 million log lines every day. Some of those are duplicates that need to be filtered, and only a fraction of these logs are critical and need to be handled immediately. As such, the importance of having efficient log management cannot be overstated.

In this guide we will explore the observability offerings of Elastic, the company that introduced Elasticsearch, Logstash and Kibana, also known as the ELK stack. We will look at running either an ELK stack or Elastic Cloud, Elastic's newer PaaS offering, on your own with a focus on engineering overhead, infrastructure costs, and ease of log management.

<PART 1>

Breaking down cost of ownership

For any tool, open-source or commercial, there is more to consider than the initial price tag. While the purchase price is often visible and straightforward, many other associated costs still need to be considered.

- What hardware or infrastructure is required to support the tool?
- How long will it take to set up and onboard the team using the tool?
- Will the tool require ongoing maintenance and/or support staff?

Cost of Ownership can be broken down into visible costs and hidden costs. Visible costs are easily identified and quantifiable, while hidden costs are not immediately obvious nor easily quantifiable.

A tool may have a reasonable price tag but still have significant costs hiding beneath the surface. That's why it's important to consider all of the following cost areas when evaluating a new tool.

- Purchase cost
- Infrastructure
- Initial setup and customization
- Ongoing operational costs
- Intangible Cost

<PART 2>

Estimating cost of ownership for the “Free” ELK stack

Building your own log management solution means setting up, customizing, and managing an ELK stack instance.

The associated costs will vary depending on several aspects such as: how much log data is generated by your system? How much is your log volume likely to increase each year? How long do you need to retain that data? How accessible does your data need to be?

We’ll break down the costs associated with building an ELK stack solution for a typical company based on the following assumptions:

- Produces approximately 100GB/day of log data for the first year (increases ~50% each year)
- Requires 14-day data retention
- Requires high data availability

Let’s look at how each of the previously mentioned costs associated with ELK ownership play out over a 3-year period.

Purchase cost

This is the tip of the iceberg, the most visible and “obvious” cost associated with any tool - its price tag. For a free ELK Stack instance, the upfront cost is $0 (surprise!), easily setting it apart from commercial solutions. It’s important to remember that there are likely additional costs hiding beneath the surface. For example, Elasticsearch’s security and machine learning features are largely limited to the paid tiers (Platinum and Enterprise).

ELK Paid Subscription
(Platinum and Enterprise)

- Anomaly detection
- Endpoint protection notifications
- Multi-stack monitoring
- Support services
- Machine Learning
- Cross-cluster replication
- etc.

Underlying infrastructure

For the company described, the requirements will include 3 primary servers, 2 data servers, and disk space for data storage. Three primary servers are the minimum needed for a production offering since redundancy is required, and an uneven number of servers is recommended to avoid the split-brain problem. These can also use them for Kibana and Logstash. Two data servers are required for redundancy. Lastly, disk space should be sufficient so the system never crosses the 85% disk watermark to ensure there is enough disk space for incoming shards. We’ll make use of AWS pricing to provide a useful estimate.

We will use a c5n instance for the primary server since this type is a cost-effective, high-performance instance specialized for heavy computing.

One primary server instance (c5n.large) costs $0.108 (at on-demand rates), which comes out to about $946 each year. This large size was chosen based on the data requirements stated above. For data instances, we’ll use the i3en.large servers with an upper limit of 1.25TB SSD per machine.

	Year 1	Year 2	Year 3
Primary Servers	$2,838	$2,838	$2,838
Data Servers	$21,742	$25,965	$34,409
Disk Storage (General Purpose SSD on EBS)	$6,300	$12,600	$25,258
Total Hosting Cost	**$30,880**	**$41,403**	**$62,505**

The i3 instances are storage optimized for high transaction, low latency workloads such as Elasticsearch. For the first year, one machine covers our stated data requirements plus one copy for data redundancy. In the second year, 2 instances are required plus redundancies, and in the 3rd year 4 instances will be needed (again, with redundancies). We will also use a dedicated instance on AWS. i3en.large instances cost $0.241 hourly on demand. There is also a dedicated per region fee based on up time of the instance which amounts to $730 per instance per month for 100% uptime. This amounts to $1811.86 monthly, or $21,742 annual cost for the first year. In subsequent years, the hourly on-demand rate increases with the number of servers, but the dedicated per region fee remains the same assuming only 2 regions are used. In terms of disk space, the cost is $0.226 per GB of data per month, plus one copy for redundancy, plus an additional consideration for system overhead. That works out to $6,300 per year for the first year with linear increases for the larger data capacity in years two and three.

Setup & engineering costs

Unlike out-of-the-box commercial solutions, engineering costs for setting up an ELK stack are significant. These include the EC2 servers, mappings, Kibana and collectors. It is often disregarded as a sunk cost (i.e. engineering salaries are being paid regardless of how that time is used). But this approach can derail an engineering team’s focus and end up being extremely expensive. Although getting an ELK stack deployed can take a few minutes, ELK requires a significant investment of time and resources to make it production-ready.

Above all else, the system must be configured to ingest and parse logs coming from multiple systems and locations, with potentially different logging frameworks and formats. This data pipeline, feeding into Logstash, must also be optimized to ensure no data is lost. In some cases, this requires additional investment in hardware.

Completing the setup for our “typical” company would take the average engineer (who is familiar with the ELK Stack) about 5 full working days. The following values are based on the average engineer’s salary of $150K/year. This translates to $2,900 just for the initial setup, not including the investment in training engineers to use the tool.

Getting Your ELK Stack Production-Ready:

- Create your deployment and customize to best fit your use case
- Configure index management
- Configure Kibana, Elasticsearch clients, Beats, Logstash and APM Agents
- Add additional plugins for desired functionality
- Secure your deployment
- Plan for scale and availability

	Year 1	Year 2	Year 3
Setup & Customization	$2,900	$5,800	$11,600
Training	$8,650	$17,300	$25,950
Total Setup Cost	**$11,550**	**$23,100**	**$37,550**

If we assume that the training will take an additional 3 days, with 5 engineers being trained, that comes out to $8,650 for the first year. As the stack becomes more complex and personnel at the company changes, this number will likely grow in years 2 and 3. Due to the complexity of the ELK Stack architecture, companies will often dedicate at least one full-time employee to building and running it.

Ongoing operational costs

Engineering costs don’t end once your ELK Stack is set up. Additional considerations include monitoring performance of the ELK Stack and troubleshooting system errors and outages. As your system grows, logs tend to evolve and it’s important to ensure that the logs are ingested and indexed properly. The tooling itself has a free version, but Elastic support and many features only are available at an additional cost. Elastic offers tier-based subscriptions that include varying levels of support.

Ongoing ELK Maintenance Tasks:

- Performance monitoring and tuning of ELK clusters
- Performing ELK upgrades when new versions are released
- Maintaining proper ingestion and indexing of data
- Creating and maintaining dashboards in Kibana
- Creating and maintaining alerting
- Handling change requests from teams and Executives
- Troubleshooting system errors and/or outages

	Year 1	Year 2	Year 3
Ongoing Operations (i.e. Engineering Time)	$50,000	$75,000	$150,000

Higher tiers also come with additional capabilities, including out-of-the-box integrations, alerting based on anomaly or SLO, endpoint security, machine learning search and analysis capabilities, etc. These features also reduce engineering costs by making your observability stack more effective in reducing MTTR and engineering effort. However, Elastic uses a resource-based pricing model for their subscriptions, so as your system scales, cost will scale up along with it.

The engineering effort required for version upgrades is also a serious cost consideration. To simplify the upgrade, Elastic synchronizes the releases of Elasticsearch, Kibana, and Logstash, the tools that make up the ELK stack. However, there is still a fair amount of time and effort required to smoothly upgrade versions, including checking for deprecated features or any other breaking changes on each tool, backing up data, testing in pre- production environments and much more. Conventional wisdom in the software engineering world is that in the first year of using ELK, one engineer dedicating a third of their time should be sufficient for necessary maintenance and scaling requirements. As the stack grows and becomes more complex, more time must be invested.

Assuming consistent stack growth, by the second year, that engineer will likely need to dedicate half of their time to maintaining the ELK Stack, and by the 3rd year, one full-time engineer should be allocated to managing the ELK stack.

Intangible costs

Intangible costs are those that we simply can’t see coming. This could include things like the cost of system downtime or critical issues, missed opportunities/ value, etc. For example, the difference between Elastics’s paid solution with Machine Learning capability compared to their free ELK stack, which is equipped only with static alerting thresholds, can be very significant. An anomaly in error rates would likely be flagged in the paid version with ML-powered anomaly detection. In contrast, in the free version, if the actual volume of errors doesn’t pass a certain threshold, the static alerting would miss it.

If that piece of information is the difference between 1 hour of downtime and 1 day of downtime, the cost difference can be huge in terms of customer satisfaction and retention, SLA violations, etc.

These are things that ultimately cost companies money, though they may not be easily quantified or attributed to tooling investments. Although not easily measured, intangible costs cannot be ignored when considering cost of ownership and return on investments. Ultimately, any trade-off that impacts the return you see on your investment in a tool IS part of the cost of ownership.

Machine learning mechanisms and training is required to set up anomaly detection in an effective way. Without it, you may face both false positive and negative alerting. Most of the machine learning capabilities of Elasticsearch are locked behind the Platinum and Enterprise tiers and will add to the cost presented in this guide.

<PART 3>

Elastic Cloud

Elastic also offers a managed solution for deploying, operating, and scaling Elasticsearch, Kibana, and other Elastic brand tools. This solution reduces the need to manually set up and scale up the required infrastructure. However, engineering effort is still required to properly size, monitor and update the cluster. Tiered subscriptions are also still present, so designers should understand what features will be used to decide on the appropriate subscription. Let's look at the cost of running your observability tool in Elastic Cloud.

Purchase costs

Purchase cost would remain $0 since we will remain in the standard tier for comparison. While not necessarily paying for the ELK subscription, Elastic Cloud does associate tiered costs with infrastructure. These costs are discussed in the next section. For this reason, We will assume the purchase cost is still $0 but associate costs with the infrastructure chosen instead.

Infrastructure costs

Elastic Cloud allows users to choose from different cloud providers for infrastructure. Instance types and sizes are also customizable. We will use AWS in the us-east-2 region to make a direct comparison with the custom Elasticsearch setup discussed earlier. Elastic Cloud's lowest tier starts at $95/month. We will use the standard tier feature set, but configure the infrastructure for the use case outlined earlier. This setup selects 3.4TB storage with 116GB RAM in hot storage. As recommended, this is spread across two availability zones to avoid data loss. We will not provide warm or cold storage for this example. This gives an hourly rate of $3.0069. We will also include 16GB Kibana implementation across two availability zones. This Kibana setup yields an hourly rate of $0.6816. Several other offerings, including machine learning and coordinating and ingest instances, would be useful for an observability stack. Coordinating and ingest instances process all incoming requests. Using this feature will result in better performance and reduce the chances of failure. Machine learning instances automate the analysis of time-series data. They create baselines of normal behavior that are used to detect anomalies. We will select the smallest 1GB RAM option for each of these over only a single availability zone. With this setup, these servers yield no extra hourly rate.

	Year 1	Year 2	Year 3
Infrastructure Cost	$32,311	$57,543	$87,111

Elastic Cloud is billed hourly regardless of use. Since our data will grow over three years, we will scale the Elastic Cloud hot storage infrastructure by increasing the hot storage to the next higher size annually and recalculating the cost with the new hourly rate. Using the hourly rates listed above, and assuming this Elastic Cloud set up will always be running, we can budget the following for infrastructure costs annually.

Setup costs

Since Elastic Cloud is an out-of-the-box commercial solution for Elasticsearch, minimal engineering is required to get the system running. The EC2 servers, mappings, Kibana, and collector are all configured in the infrastructure and require little work beyond understanding your system's requirements. The system will still need a data pipeline configured to collect logs from different locations and ingest them into Logstash. This pipeline will also need optimization and may require additional hardware investment beyond what is set up in Elastic Cloud. We will assume that an average engineer makes a salary of $150K USD and will require 1 day for setup.

	Year 1	Year 2	Year 3
Setup & Customization	$577	$577	$577
Training	$8,650	$10,385	$12,115
Total Setup Cost	**$9,227**	**$10,962**	**$12,692**

In subsequent years, engineers may need to update the infrastructure as the system scales. We will assume one day is sufficient for these updates. Training engineers to use the Elastic tools should take an additional 3 days. We will assume we will train 5 engineers to use the Elasticsearch tools. As the system becomes more complex, employees must spend more training with the tool, and the cost will increase.

Ongoing operational costs

Operational costs for Elastic Cloud are less than those for the custom ELK solution since the infrastructure itself does not need to be considered. Even though this is a managed solution, engineers must still monitor the system for performance and errors. If these degrade, a change to the Elastic Cloud configuration may be needed.

The Elastic Cloud also will not upgrade automatically. Engineers must still check for deprecated features and breaking changes before upgrading. Data should also be backed up prior to upgrade to avoid outages and data loss.

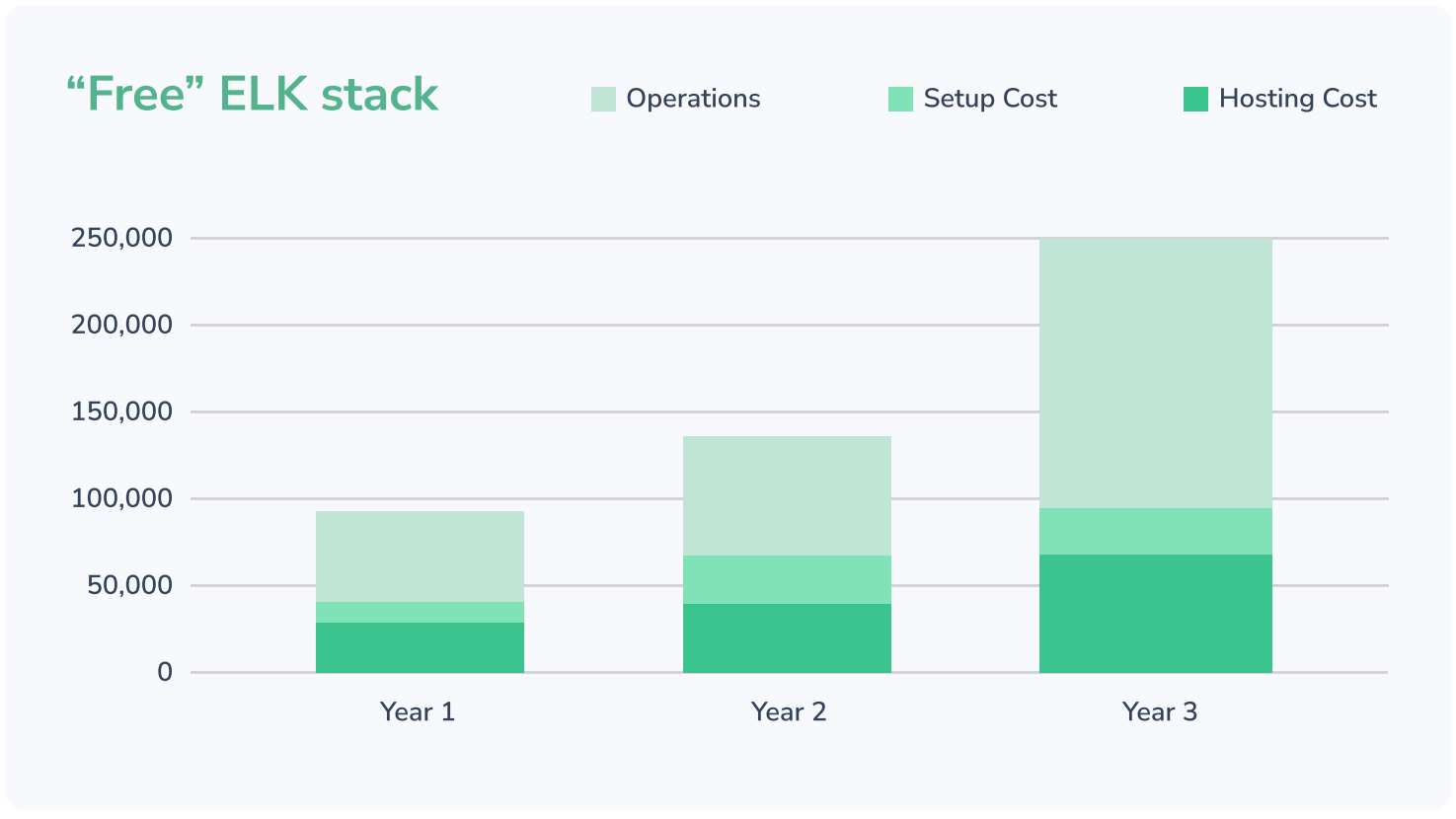
	Year 1	Year 2	Year 3
Ongoing Operations (i.e. Engineering Time)	$50,000	$60,000	$70,000

An engineer will likely need to dedicate one third of their time to maintaining the Elastic Cloud. This will increase over time as the stack becomes more complex, though not as much as a custom deployment since much of the infrastructure maintenance requirements are handled.

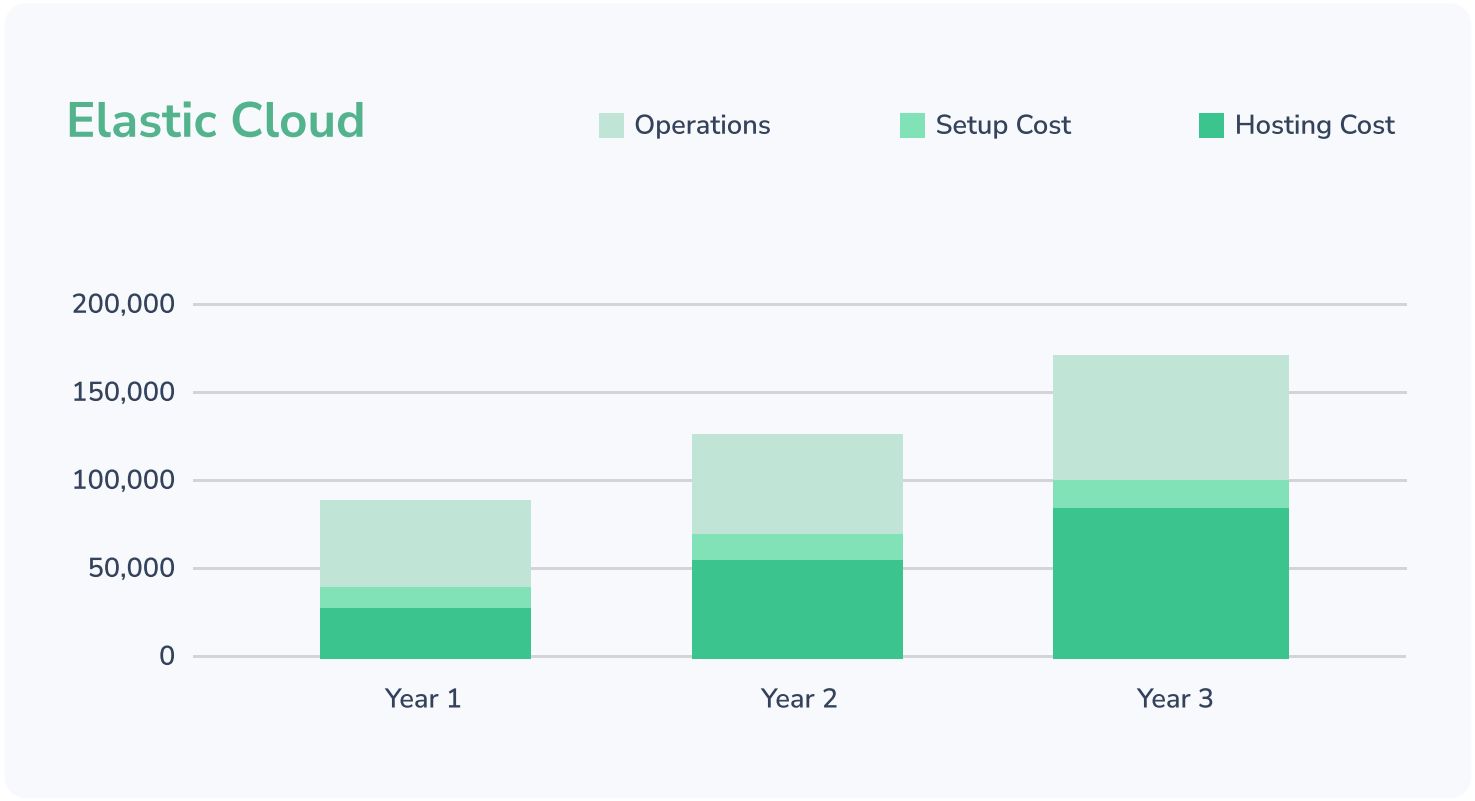
<PART 4>

ELK stack and Elastic Cloud TCO summary

At first glance, **free log management** is hard to ignore, but remember that with open-source tooling, the price will likely come in the form of engineering hours. There are many commercial log management solutions on the market, each with their own advantages and disadvantages. Aside from looking for a tool that matches your team's functional requirements, cost is still a major factor to be considered.



Managed tools like Elastic Cloud come with engineering cost, but that cost is reduced compared to the completely custom solution. Instead, the fees are built into the cost of the managed infrastructure and should be expected to grow year-over-year as your solution becomes more complex.



	Year 1	Year 2	Year 3
Total Cost of Ownership for "Free" ELK Stack	$92,430+	$139,203+	$248,105+
Total Cost of Ownership for Elastic Cloud	$91,538+	$128,505+	$169,803+

<PART 5>

Full-stack observability, without paying for the noise

Most log data is never queried or analyzed and only contributes to high data storage costs (as seen above), but at the same time, logs do contain important information. It just doesn't make sense for you to pay the same price for critical log data that you pay for irrelevant log data.

Coralogix provides tools to shed expensive storage for unnecessary data and a means to monitor without the expensive hot indexing required by an ELK stack. In-stream log analysis provides comprehensive monitoring and alerting without needing to hot index data. Full dashboarding, alerting, fast querying from archive and more are available without a tiered paywall. This capability allows Coralogix to provide data triaging, called the Total Cost of Ownership (TCO) Optimizer. Customers can choose where data is routed - whether to hot or cold storage immediately after ingestion. Retention is unlimited as all data is stored in S3 or similar archival storage.

Take ownership of your observability spend by deciding what data is needed for lightning-fast queries, what won't be frequently needed, and what is only stored for rare occurrence requirements (e.g. compliance related logs). Maximize not only the amount of data you can ingest, but also the amount of data that you can generate comprehensive insights for.

Schedule a Demo