TCO Analysis for ELK Stack

Cost Consideration for Production Ready ELK Stack

By Vijay Upadhyaya
Director, Competitive Intelligence

Executive Summary

Modern applications generate colossal quantities of machine data. Skillful analysis has the potential to yield perceptive and actionable intelligence from all of this raw machine data. In response, open source solutions such as the ELK stack (also known as the Elastic Stack) have become very popular. Unfortunately, numerous enterprises are belatedly recognizing that placing the "free" Elastic Stack into production comes with abundant concealed costs and challenges.

What they've learned is that it's not just a matter of downloading Elasticsearch software and going live. Instead, these organizations are forced to shoulder significant unplanned expenses on hardware, infrastructure, consulting, security, training, and customization before they even attain production. Ongoing maintenance such as updates and upgrades adds yet another layer of expense and complexity. Worst of all, there are serious opportunity costs inherent when maintaining a customized Elastic Stack instance.

This paper highlights the hidden costs for Elastic Stack deployments and provides contrast with the advantages of selecting a best-of-breed SaaS machine data analytics solution such as Sumo Logic. Business executives managing the budget for IT infrastructure as well as technology leaders evaluating machine data analytics solutions will discover how Sumo Logic accelerates deployment, delivers higher value, and offers a lower total cost of ownership (TCO). Most importantly, it lets customers focus all their engineering and IT resources on their core business without wasting time on implementing and managing a machine data analytics platform.

Introduction

When properly interpreted, the deluge of machine data (logs, metrics, and events) that are now generated by organizations can yield tremendously valuable, far-reaching insights. These include identifying security breaches early in the cyber kill chain, correcting operations problems before they arise, and uncovering opportunities to boost profitability. This can also help predict and prevent application failures during periods of sudden surges in demand, such as when a mobile game goes viral, or a disruption in public transport suddenly quintuples demand for ridesharing services.

An entire ecosystem of machine data analytics tools has sprung up in response to the advent of ubiquitous machine data. One of the most popular offerings has been the Elastic Stack, which is composed of four open source projects and four commercial projects stitched together into a machine data analytics tool. Regrettably, organizations deploying it are realizing that attempting to move an assortment of separate open source projects into production presents a unique set of challenges, ranging from performance and scalability to administration and security. These complications with Elasticsearch add up to distinctly higher expenditures and unnecessary staff distractions.

Consequently, enterprises contemplating acquiring a machine data analytics platform for deriving maximum value from their machine data should evaluate best-of-breed integrated solutions capable of consuming numerous varieties of machine data (such as log, metrics, and events) to service a broad range of use cases (such as security and operations). A notable example is Sumo Logic's cloud-native, machine data analytics platform delivered as a SaaS service for log analytics and time series metrics powered by machine learning.

The Elastic Stack

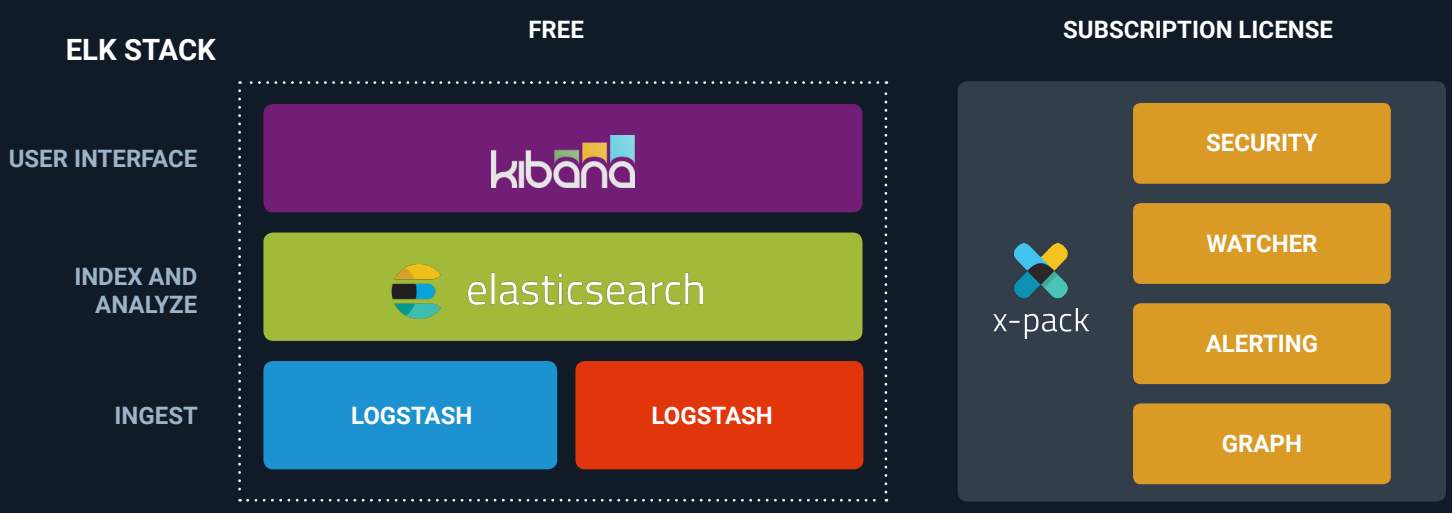
The original Elastic Stack was assembled from four separate open source projects:

1. Elasticsearch. A distributed, full-text search engine
2. Logstash. An aggregation service that collects and process logs
3. Beats. Collects and transports logs and metrics to Logstash or Elasticsearch
4. Kibana. Data visualization designed to work on indexed log content

The Elastic Stack may be extended beyond these core components for those customers willing to pay for a supplementary subscription. For example enterprise-grade features require X-Pack add-on bundles for security, alerting, monitoring, reporting, and graph capabilities.

While Elastic Stack – whether standalone or bolstered by X-Pack – is a popular approach for ingesting and analyzing log data, it has a number of unanticipated costs associated with it that should be contemplated prior to placing it into a production environment.

ELK Projects



Cost Considerations for Production-Ready Elastic Stack

As is the case with many multifaceted open source initiatives, the Elastic Stack presents a collection of daunting obstacles that must be overcome prior to achieving a successful deployment. Its overhead and costs can be allocated to one of three major lifecycle phases:

1. Infrastructure acquisition
2. Infrastructure support
3. Ongoing operations

The Elastic Stack may be deployed on-premise or hosted in the cloud at Amazon Web Services (AWS). It's important to note that these cost considerations are equally relevant in either case: on-premise outlays for servers and storage will normally be accounted as capital expenditures, while AWS deployments will be treated as operational expenses. To help demonstrate the true price tag of an Elastic Stack initiative over a three-year period, lets consider a typical mid-size ELK stack deployed in AWS:

- 160 GB/day of data ingestion from 100 monitored servers with ingest volume doubling every year
- 90 days of hot data retention
- 365 days of cold data retention
- 40 total users
- 20 concurrent users
- 20 concurrent searches

1. Infrastructure Acquisition

Given the dynamic nature of most machine data, it's very difficult to precisely determine the correct types and amounts of resources for the Elastic Stack. This means that enterprises confront two infrastructure acquisition risks: over-provisioning wastes money on under utilized assets; under-provisioning limits visibility at crucial times which increases the mean time to repair issues, violates SLAs, and inconveniences users.

To help estimate their infrastructure requirements, organizations must answer questions such as these:

- **How much machine data (in gigabytes) will be generated each day?**
- **How long will this information be retained and how many times the data will be replicated?**

Additional questions:

- How many simultaneous users will interact with the stored machine data?
- How many concurrent searches will be performed?
- How many servers will it be necessary to provision?
- What is the recommended configuration for each server (processor class, memory, storage, and so on)?
- How will the network infrastructure need to be expanded to support the Elastic Stack implementation?

In this example – which assumes 90 days of hot data retention – data ingestion ramps from 160 GB/day supported by a 32 TB storage cluster in the first year to 640 GB/day on a 127 TB cluster in year three. The corresponding number of Elastic Stack data node servers grows from 31 in the first year to 61 in the third year. Finally, these servers may be hosted on premise or rented in the cloud; this scenario assumes the latter.

	Year 1	Year 2	Year 3
Ingest Volume	160GB/Day	320GB/Day	640GB/Day
Cluster Size	32TB	63TB	127TB
Server	$222,171	$374,490	$476,036
Storage	$47,520	$95,040	$190,080

*Assumptions: retention period=90 days, replication factor =2, inflation factor= 1.1, Data Node:c4-8XL, Client Node: R4-2XL Cluster size = Ingest Volume*retention period*replication factor*inflation factor*

The infrastructure acquisition costs work out as follows:
The three-year total infrastructure acquisition cost for this configuration is **$1,405,337**. Note that this does not include outbound data transfer costs, which are charged by the gigabyte by AWS.

2. Infrastructure Support

After the initial rollout is complete, Elastic Stack customers need to carry out a serious – and constant – evaluation of their environment to ensure that it remains production-ready. This encompasses securing and monitoring the stack while continuously validating high availability and disaster recovery procedures. All of these efforts increase the cost and complexity of operating the Elastic Stack.

These are a subset of the questions that must be addressed to guarantee that the Elastic Stack supporting infrastructure will meet its obligations:

- **How will Elastic Stack outage planning, detection, and correction - including root cause analysis – be handled?**
- **Who will validate ongoing security certifications for the Elastic Stack?**

Additional Questions:

- What are the disaster recovery plans that will be necessary to support the Elastic Stack?
- Who will be responsible for designing, performing and testing the backup and recovery procedures?
- What are the costs to the business should the Elastic Stack suffer downtime?
- Will Elastic Stack data be encrypted at rest and in transit? If so, how? What will be the key management policy?
- How will expected data growth affect deployed hardware, storage, and network resources?
- What supplementary applications will need to be acquired, configured, and placed into production to augment the Elastic Stack?

Continuing the example scenario, three-year infrastructure support costs will be as follows:

	Year 1	Year 2	Year 3
DR/Backup	$0	$0	$0
Threat Intelligence	$80,000	$80,000	$80,000
Monitoring	$0	$0	$0

Assumptions: Threat intelligence service will range from $80K to $150K per year

The three-year total cost for supporting infrastructure is **$240,000**

Threat intelligence is $240,000 while data replication/backup costs are listed as $0 because AWS storage has robust data durability in place. Monitoring is also assumed here to be $0 because Elastic Stack Monitoring (previously known as Marvel) is included with Elastic support. It's important to note, however, that these costs, especially DR/Backup would be significant for an on-premise deployment.

3. Ongoing Operations

Even after accounting for initial rollout and further infrastructure tuning, the job of caring for the Elastic Stack can have a negative impact on the organization's overall goals. When all is said and done, the enterprise has constructed - and must maintain - a commercial-grade production-ready machine data analytics tool that can capture logs, events, and metrics. But that's not the business they're in.

Engineering time and talent are scare; labor expenditures on infrastructure detract from enhancing core application capabilities that are meant to serve their customers. Additionally, it's doubtful that staff members will be clamoring to perform these chores: unlike many other cutting-edge open source initiatives, these unavoidable Elastic Stack administrative and operational tasks such as upgrades and updates are tedious and mundane. Nevertheless, these are the concerns that must be resolved:

- **What are the professional services costs necessary to roll out the Elastic Stack?**
- **How many fulltime employees will need to set up, upgrade, update, and maintain the Elastic Stack?**

Additional Questions:

- What other applications or systems will need to be deferred or cancelled because of staff allocations to the Elastic Stack?
- How will machine data from unplanned sources – such as a threat intelligence feed – be integrated into the Elastic Stack production environment?
- What will be the expected capital outlays to cope with expected – and inadvertent – machine data growth?
- Who will be responsible for ongoing performance tuning?
- Who will be responsible for overseeing security audits?

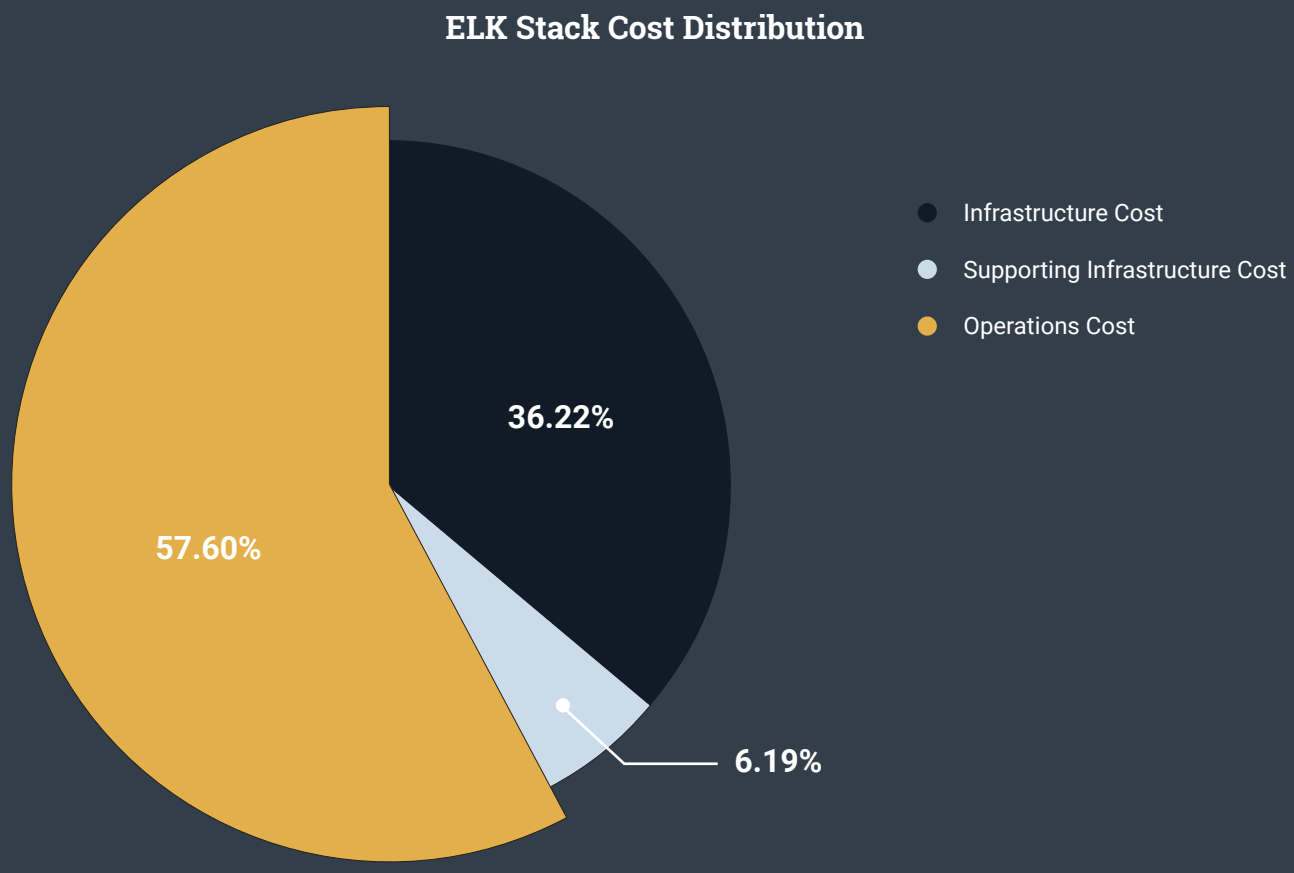
These expenses are noteworthy, and will only escalate over time:

	Year 1	Year 2	Year 3
Fulltime Employees	$290,000	$580,000	$660,000
Elastic Support and Maintenance	$155,000	$245,000	$305,000

Assumptions: Full time employee cost is $130K/year.

The total three-year outlay for ongoing operations will be **$2,235,000**.

In summary, the **three-year total cost of ownership to build out the Elastic Stack is approximately $3,800,000.** The breakdown of these expenses may be found below; more than half the cost is from operations.



Crucial Feature Deficiencies

As the following table illustrates, the Elastic Stack's capabilities are lacking in a number of critical areas. Each shortcoming degrades its utility while potentially increasing costs.

Deficiency	Implication
Lengthy mean time to remediate (MTTR)	Expanded risk of downtime or other major operational issues
No real-time data visibility	Delays in uncovering operational, security, or other issues
No improved DevOps processes (Livetail, Log Compare, Apps)	Diminished innovation: developers unable to easily access logs in real-time
No searchable cold data	Impossible to accurately view historical operational trends
No data encryption at rest	Exposure to data theft or other security incidents
No on-demand scalability	Increased hazard of wasting money by over-provisioning or failing SLA mandates by under-provisioning
No native metrics backend	Metrics ingested as logs are three times more expensive than native metrics
No security attestations	Possibility of compliance violations

The Sumo Logic Solution

Sumo Logic's cloud native Software as a Service (SaaS) alternative completely bypasses the business and technical obstacles inherent with an on-premise Elastic Stack implementation yet impose no operational overhead.

Deployment and Ongoing Administration

Launching a new Sumo Logic initiative is a quick, straightforward process that doesn't require support from third-party consultants or teams of internal employees. Its schema flexibility makes it possible for customers to immediately begin ingesting log data from any source without the need for writing parsers.

Once deployed, Sumo Logic is responsible for managing all backup and recovery procedures, along with applying software updates without imposing downtime. In situations where the logs generated by the customer's application indicate that an issue has taken place, Sumo Logic customers can employ its native, real-time dashboard and advanced analytics using machine learning to rapidly pinpoint and correct the problem.

Performance and Scalability

As a SaaS provider, Sumo Logic liberates customers from the burdens of resource provisioning. Instead, they're free to leverage its unlimited elasticity and scalability to match the rhythm of their own clients without any time lag. In practice this translates to instantly supporting bursting and handling exponentially larger data ingestion volumes.

Security

From the very beginning, Sumo Logic has treated security as a cornerstone of its solution. This is reflected in diverse product characteristics such as single sign-on, encryption of data at rest and in motion, and simplified, streamlined, and highly granular access control that makes it easy to provision and configure users. Sumo Logic also adheres to multiple, widely adopted industry security certifications including:

Innovation

As described earlier, installing the Elastic Stack introduces the very real risk of frittering away the time and efforts of valuable IT and engineering staff. By utilizing Sumo Logic, these team members are freed to focus on their principal obligations such as optimizing the software build pipeline or adding new features to their application. With mundane administration chores no longer part of the picture, enterprises can employ Sumo Logic's rich array of advanced analytics powered by machine learning, and out-of-the-box applications to derive maximum value and meaning from their machine data.

Sumo Logic Machine Data Analytics: Platform and Cloud Ecosystem Partners

The Sumo Logic machine data analytics platform accelerates the "monitor, troubleshoot and identify root cause" cycle from days to a few hours. Unified logs and metrics allow customers to monitor and troubleshoot issues early on and quickly get to the bottom of performance problems with applications or poor customer experiences due to latency.

Sumo Logic has numerous technology partners such as AWS, and Microsoft Azure, along with tight integration with other products in the ecosystem such as Content Delivery Networks (CDN), Identity and Access management platforms, and Application Performance Monitoring (APM) vendors. To further accelerate time-to-value, Sumo Logic offers applications that let customers visualize and monitor key performance indicators along with security-focused alerts.

Learn more about Sumo Logic

It's easy to launch a free Sumo Logic evaluation. To quickly begin ingesting gigabytes to petabytes of logs. metrics, and events, visit https://www.sumologic.com for free trial. To help prospective clients visualize its compelling technical and financial benefits compared with the Elastic Stack, Sumo Logic also offers a customized return-on-investment calculator. Please contact us at https://www.sumologic.com/contact-us.

sumologic®

Toll-Free: 1.855.LOG.SUMO | **Int'l:** 1.650.810.8700
305 Main Street, Redwood City, CA 94603

www.sumologic.com