EBOOK

Discovering Diamanti for Data Analytics

Why the first purpose-built, fully integrated Kubernetes platform is the right fit for Splunk

DIAMANTI

Discovering Diamanti for Data Analytics

INTRODUCTION

As we created this eBook, our aim was to give our readers (both technical experts and business decision-makers) a better understanding of both how and why Diamanti, a purpose-built platform for Kubernetes, is also an exceptional fit for Splunk.

Diamanti has a proven track record of delivering compelling benefits for Kubernetes. Since containerization of Splunk is an increasingly popular approach, it's natural that Diamanti could deliver significant advantages. We've confirmed, in real-world applications, that customers running Splunk on the Diamanti platform achieve 80% reduction in the total cost of ownership, a 6x footprint reduction, and 24x indexing acceleration compared with other approaches.

If those are the kinds of benefits you're looking for, this eBook is for you.

CHAPTER 1: WHY DOES SPLUNK MATTER?

What is Splunk? If you're already using it, you know the answer. It's a tool to discover problems, devise solutions, and act on opportunities.

And it all begins with **data**.

Splunk is a software platform that helps organizations make sense of machine data that's typically complex to understand, in an unstructured format, and not easily usable — because the data sets are too large, hard to read, or constantly changing. It helps organizations take data sets, often created by websites, applications, sensors and devices, and conduct searches, analysis, and create visualizations.

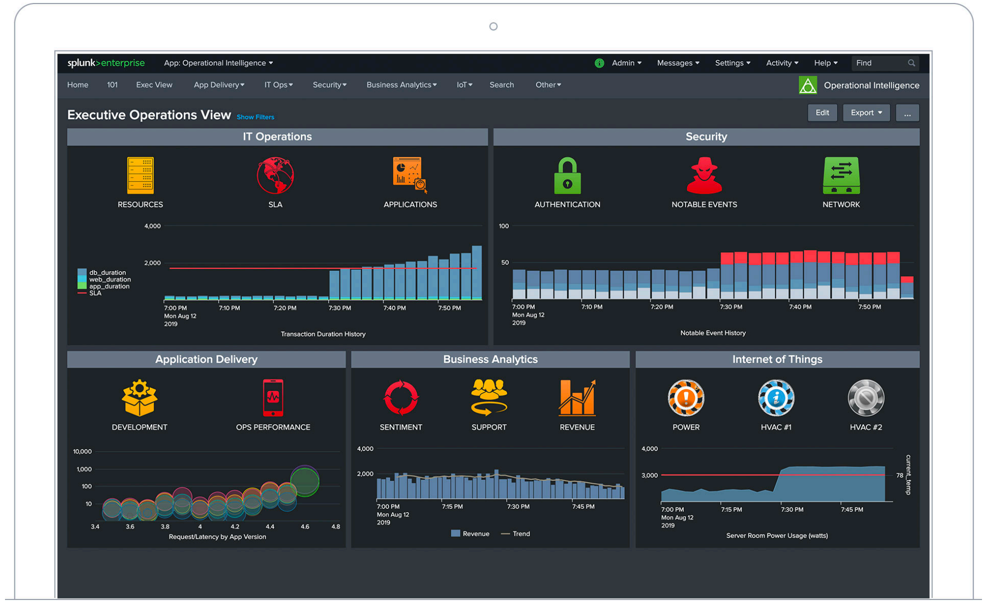

Figure 1. Splunk Enterprise Executive Operations Dashboard

Many people, when they think of Splunk, view it as a way to conduct data analytics. And it IS that — and that's where it began. Organizations who were creating massive amounts of unstructured data — from IoT sensors to data logs from modern data centers and networking firewalls — fed that data to Splunk in order to begin making sense of their data.

Originally, organizations used Splunk as a tool for descriptive and diagnostic analytics — describing issues and identifying problems. But increasingly, they also use the real-time capabilities of Splunk, coupled with sophisticated quantitative methods like statistical analysis, machine learning, and simulation, for predictive and prescriptive analytics, understanding possible outcomes and courses of action.

Splunk is a dominant, mature platform for business analytics and process mining. It helps organizations quickly identify any problems that keep them from meeting their goals.

1. With Splunk, leveraging machine data becomes easy. It's possible to build a Splunk solution that ingests data from thousands of stores, applications, or sensors and powers a consolidated monitoring dashboard in case challenges emerge.

2. It's also possible to watch business performance in real-time, providing better insight for business decision making across all departments, enhancing customer experiences.

To put it simply, Splunk helps organizations see the problems that detail business processes, find issues with service delivery, and keep on track against KPIs.

In the past organizations tended to keep Splunk platforms in a data science or business process improvement team. But over the past few years, they've discovered that the flexibility and power of Splunk applies to many domains outside outside traditional data analytics, including:

- **IT: O**rganizations can use predictive analytics to prevent IT incidents from impacting customers.
- **Security / SIEM: S**ecurity professionals can better understand issues and risks, improve incident analysis, and automate reporting.
- **DevOps: D**evelopers use Splunk for cloud monitoring, application performance monitoring, and incident response.

According to Gartner, Splunk has the highest market share worldwide in the ITOM Performance Analysis, AIOps, ITIM and Other Monitoring Tools subsegments of the IT Operations Management market[1]. 91% of the Fortune 100 use Splunk to improve their abilities to ask questions, make decisions, and take actions[2].

Of course adding Splunk to your environment isn't a trivial task, and organizations wouldn't use it unless it provided measurable, compelling benefits. A recent ESG report[3] says "organizations that place a strong strategic emphasis on data and its business value...achieve a number of key business and economic benefits." These include:

- Adding an average of 5.32% to their annual revenue.
- Removing an average of 4.85% from their annual operational costs.
- Meeting or exceeding customer retention targets.
- Making better, faster decisions than competitors.
- On the average, generating approximately $38.2 million by making smarter use of their data.

Being able to search, analyze, and visualize data is important. Splunk powers those capabilities for thousands of organizations around the world.

1. https://www.splunk.com/en_us/form/gartner-report-market-share-analysis-itom-performance-analysis-software-worldwide.html
2. https://www.splunk.com/en_us/about-us/why-splunk.html
3. https://www.splunk.com/pdfs/data-maturity/what-is-your-data-really-worth.pdf

CHAPTER 2: THE SECRETS OF SPLUNK

Splunk works by creating a cluster of dedicated resources running various parts of the Splunk software platform, traditionally on a set of dedicated servers. To provide all the functionality and advantages of Splunk, a Splunk cluster ingests data (you can call that input), parses the data, indexes the data (writing it to disk), and then conducts searches on the data.

None of this is particularly complicated or unusual. At a basic level, Splunk architecture looks like the scale-out architecture of any other distributed system. If you're familiar with Splunk architecture, you're probably familiar with a diagram like this.

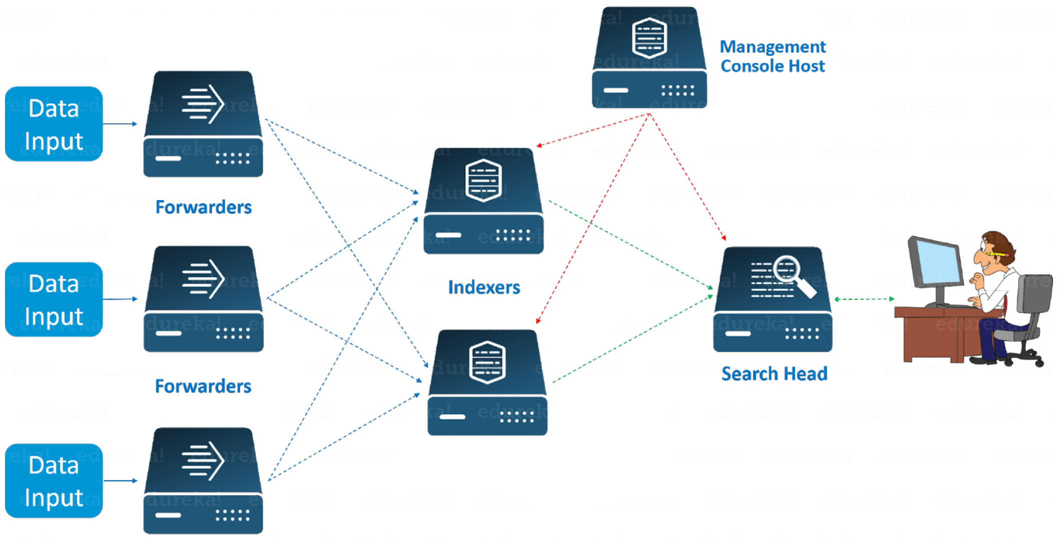

Figure 2. Example Splunk Architecture, courtesy https://www.edureka.co/blog/splunk-architecture/

At the simplest level, this is what a Splunk cluster looks like. It consists of three main components controlled by a management console. These are include Forwarders, Indexers, and at least one Search Head.

Forwarders are used to collect aggregated data from data sources, such as logs from a remote machine. You can install them on as many machines as you want and collect data, including real-time data, as you see fit. It's not unusual to put Forwarders on tens of thousands of remote systems because Splunk Forwarder uses very little CPU.

Many Forwarders simply forward raw data to indexers, but there's a special kind of Forwarder called a Heavy Forwarder that offers more intelligence -- conducting initial data processing and performing intelligent routing to Indexers. Heavy Forwarders reduce bandwidth and storage requirements, which, at scale, can be very advantageous as thousands of massive data streams from forwarders can clog up the network and server drive capacity. The downside of course is that a Heavy Forwarder puts more load on a server.

Indexers are the most important component of Splunk. They do the heavy lifting, indexing and storing the data coming from the forwarders. They transform incoming data, turning it into events, and store it in indexes for optimal search operations. They can also perform user-decision actions -- routing events to specific indexes or servers, masking sensitive data, identifying custom fields

Indexers also perform data replication. They keep multiple copies of indexed data within a cluster of Indexers that replicate each others' data. This improves resilience and performance.

Though it's possible to run Indexers on shared servers, traditionally, at scale, they run on dedicated servers to improve performance.

Finally, we have the Splunk **search head**. Users interact with Splunk through a graphical user interface running on the search head. Users can conduct searches and queries here. The Splunk search head can run with other Splunk components on a single server, or on dedicated servers.

Finally, for large scale deployments, it's normal to have a **management console host**. A management console host streamlines deployment and updates. It's a centralized management tool for distributing app updates, content updates, and configuration changes to the Forwarders, Indexers, and Search Heads.

Now that all these elements are clear, let's talk about the advantages and disadvantages of traditional Splunk architectures.

It's possible to run all these elements of Splunk on a single machine. You can have a Forwarder, an Indexer, and a Search Head running on a single server, and for small-scale, limited applications, that's a perfectly valid approach.

But most users deploy Splunk at scale. It's not unusual for a Splunk deployment to have thousands of users and index petabytes of data per day. You simply can't do that with a single server.

So inevitably organizations need a way to expand their Splunk environments, and Splunk is built to scale. The earlier diagram was a simple architecture. A more complex environment looks like this:

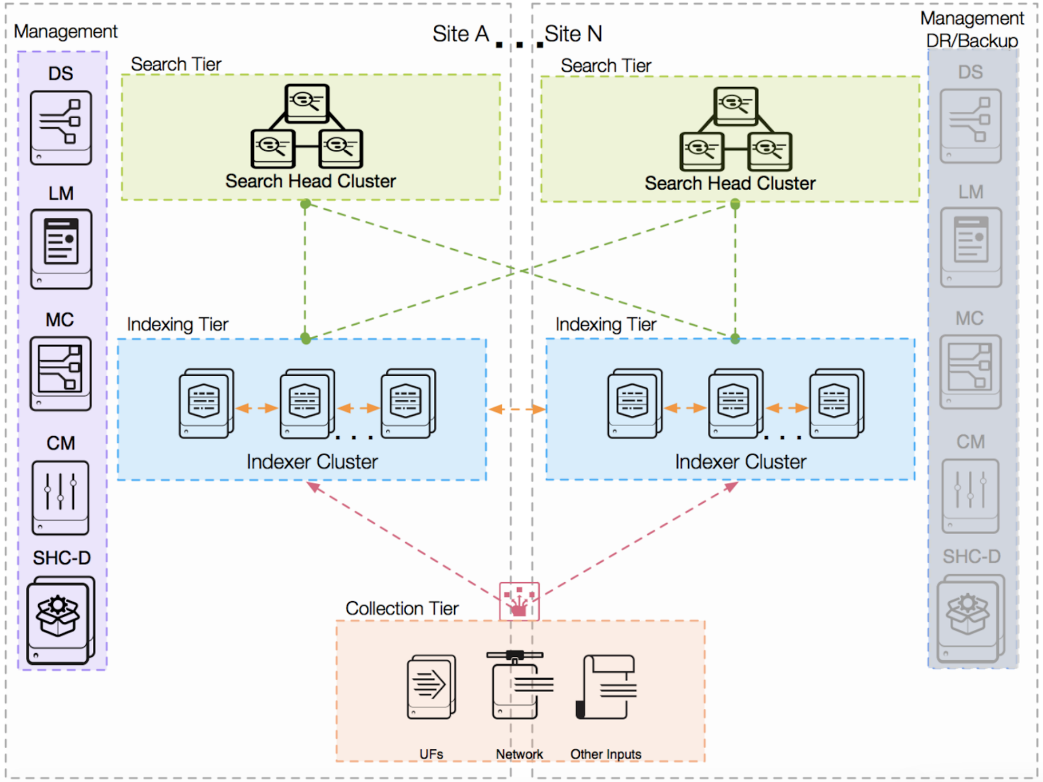

Figure 3. Splunk Validated Architecture for Multi-Site, Distributed Clusters (M3/M13)

As you can see, there's complexity at every tier. At the bottom, you have the collection tier made up of Forwarders, network bandwidth, and other inputs. One tier up, you have a collection of Indexing clusters. One level up from that, you can see Search Head clusters. And on the sides, you have management consoles, backup and recovery tools, etc.

Not only is this a complicated collection of infrastructure and code, Splunk Indexers and Search Heads are intensive processes -- they can use all available resources — CPU cycles, drive capacity and I/O, network throughput — so server capacity and performance are critical contributing factors to Splunk performance and scale.

Traditionally, determining how to build a Splunk infrastructure hasn't been easy. It made sense to put Splunk on dedicated servers and make those servers as powerful as possible, but determining the optimal configuration (with the best Total Cost of Ownership [TCO]) hasn't been a straightforward task.

For a long time, organization after organization struggled with Splunk configuration, sizing, and deployment, often oversizing or under sizing their infrastructures, with unfortunate consequences that included outsized costs, delayed time to insight, and diminished return-on-investment (ROI). As Splunk became relevant for different domains and cluster scales increased, Splunk recognized that additional guidance would be a good thing. So in 2018 Splunk released Splunk Validated Architectures (SVA)[4.]

The purpose of Splunk Validated Architectures is to "provide guidance on selecting proven reference architectures for stable, efficient, and repeatable Splunk deployments." SVAs provide a view of topologies, their characteristics, and deployment best practices. SVAs aim to:

- Provide step-by-step guidance and best practices
- Improve architecting, deployment, and operations
- Make deployments more stable and efficient
- Help deployments match infrastructure to requirements
- Shift resources from "analyze and fix" to "adopt and delight"
- Ensure success with optimal TCO

Splunk is a complicated platform, with many moving parts. SVAs aim to help organizations make better decisions. But there is another technology that has emerged in the last five years that makes modern, distributed applications like Splunk easier to manage and scale.

And that's what we're going to talk about in Chapter 3.

[4] https://static.rainfocus.com/splunk/splunkconf18/sess/1521672972553001juqQ/finalPDF/FN1151_HitchhikersGuideToSVAs_Final_1538666357995001 2mO7.pdf

CHAPTER 3: MAKING SPLUNK BETTER WITH CONTAINERS

Since the debut of Docker containers in 2013, the industry has been rapidly standardizing on new scale-out application architectures based on containers and Kubernetes. Splunk is one such application that can benefit from its many advantages.

That may not seem like an intuitive idea. We all know that containers are the newest approach to application development, deployment, and operations. These lightweight software packages, along with tools like Docker and Kubernetes, are becoming a commonly used. In a cloud-first world, leading-edge organizations are choosing containers because they offer a collection of compelling advantages, including:

- **Increased portability:** Since a container contains everything an application needs to run (libraries, runtimes, etc.), it's easy to put a container anywhere, whether on-premises or in the cloud.
- **Simple, fast deployment:** Containers make deployment simple, giving you a master image to deploy numerous replicas on demand. Spin up and spin down containerized applications as you desire.
- **Improved flexibility:** You can scale containerized applications quickly and easily by moving a container to additional hardware or cloud instances on the fly or scaling out with additional replicas.
- **Enhanced productivity** -- Containers cut down on development and deployment time, simplifying the install process and reducing errors.
- **Better security** -- Containers improve workload isolation and help protect from outside threats.

BACKGROUND INTO CONTAINERS & KUBERNETES

Application containers leverage Linux kernel concepts – namespaces and cgroups – that allow different applications to run simultaneously on the same operating system. In 2013, Docker was launched, adding tooling to make the creation and deployment of containers very simple.

Today, millions of developers are developing with containers and enterprises are rapidly adopting Kubernetes – the open source orchestration tool for managing containerized applications.

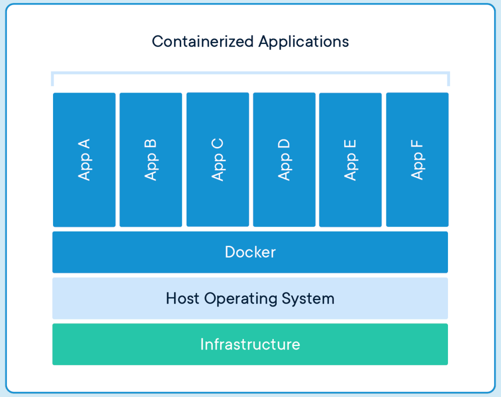

Figure 4. Docker Container Architecture

These advantages also apply to “off the shelf” applications like Splunk to greatly simplify management and operations while also delivering the following benefits:

- **Lower TCO:** A containerized deployment, when designed well, leads to a smaller footprint and more efficient resource utilization. For Splunk, that translates into faster forwarding and indexing with the same processing and I/O.
- **Faster Time-to-Value:** For Splunk, containers make deployment faster and easier. Organizations can test, patch and upgrade environments rapidly and begin analysis sooner.
- **Increased agility and flexibility**: Suddenly it’s easy to expand, select between on-premises and cloud infrastructures, and get the job done.

Combined, these advantages all integrate into something that’s very desirable in any Splunk environment — **accelerated time to insight**. Containers make Splunk better, which is why Splunk has, since 2019, accelerated the containerization of Splunk through both development efforts and official support.[5]

Splunk for containers is very popular. Their Splunk Enterprise and Universal Forwarder container images on Docker Hub are downloaded millions of times each month. And Splunk is continuing to invest in Splunk for containers with additional tools, including:

- **Splunk Connect for Kubernetes:** It’s a simple, secure, and scalable collector to help extract logs, metadata, and metrics from containerized applications. It runs natively in Kubernetes, includes filtering to optimize ingest, and it’s easy to deploy, use, and extend.
- **Splunk Connect for Docker:** It’s optimized for pulling logs from Docker containers into Splunk.

As Splunk itself puts it, “containers are easy, can be started and stopped quickly, have the ability to scale, can start up more instances when needed, and can shut down with a load subsides.”

And these capabilities translate into:

- Better overall performance
- Improved efficiency, faster deployment, expansion and contraction
- Accelerated time to insight for better business decision-making.

Or to put it another way, containerizing Splunk gives us better performance, more flexibility and portability, it’s easy to deploy, and it can be deployed across multiple infrastructures

But what infrastructure works best? That’s the focus of Chapter 4.

[5] https://www.splunk.com/pdfs/technical-briefs/splunk-for-containers-tech-brief.pdf

CHAPTER 4: INFRASTRUCTURE OPTIONS FOR SPLUNK

When Splunk emerged in the marketplace, performance was the primary concern. Data scientists built dedicated Splunk clusters for their analytics because Splunk, as a highly intensive workload, tends not to “play well with others.” Even on a small, single server Splunk environment, as data grows, forwarding and indexing demands increase, so servers experience higher and higher CPU utilization and drive utilization.

Once organizations reach a performance limit, they shift toward distributed Splunk environments. These add CPU and storage throughput at the cost of another problem — network congestion and contention. Since distributed Splunk environments are characterized both by extensive north-south network traffic as Forwarders route traffic to Indexers, and east-west traffic as data is shared across cluster nodes, network latency and congestion can become significant bottlenecks.

Dedicated Infrastructure

Combined, these problems drove deployers to use dedicated hardware for Splunk. It’s not uncommon to see Splunk running on dedicated servers connected by a dedicated cluster network. Dedicated infrastructure minimizes competition for CPU resources, storage throughput, and network bandwidth.

But this approach comes with problems. Dedicated infrastructure isn’t typically very agile.

Imagine that your data ingest requirements are growing rapidly, and your cluster is averaging 80% CPU utilization. You can add Forwarder or Indexer servers to their respective clusters, but to do so, you must purchase servers, wait for delivery, test, setup, and add Splunk. That takes time, and while you’re waiting, your CPU utilization is inching up, getting closer and closer to the dreaded 100% utilization. You lack the flexibility you need to succeed.

Dedicated infrastructure is also not entirely simple. Even with Splunk Validated Architectures, setting up a distributed Splunk environment isn’t the easiest thing to do. It takes time, specialist skill sets, and a degree of experience some organizations just don’t have.

Finally, dedicated infrastructure is also costly, especially if your organization wants to use Splunk in various ways, for SIEM or IT Monitoring as well as Operations. Your IT budget goes from supporting a single Splunk environment to several. You discover that you’re being asked to support extensive infrastructure purchases, setup, and maintenance, with all the added headaches multiple environments entail.

Splunk Cloud

Alternatively, organizations can take a “cloud-first” approach. With Splunk Cloud[6], you can sign up to a free trial and start analyzing your data without purchasing infrastructure, in minutes. You can expand your environments by submitting a request to the Splunk team instead of adding drives or switches. You can also rest assured that the Splunk team takes care of peripheral considerations, like the security of your data. And as long as you continue to pay your subscription fees, your Splunk environment continues to function.

For many organizations, Splunk Cloud is an outstanding option. But like any other cloud service, there are tradeoffs when compared to running on-premises. The most important tradeoff is **control.**

Unlike on-premises infrastructure, customizing the underlying infrastructure is not up to you and your team; it’s up to the cloud provider. And in this case, it’s important to understand the decisions[7] Splunk made about the underlying infrastructure.

1. Infrastructure options are fixed. If you wanted to move from 10Gb Ethernet to 40Gb Ethernet, or replace SSDs with NVMe drives, there’s no way to do that. Upgrades to the underlying infrastructure happen when Splunk (and their cloud partners, Amazon Web Services [AWS] and Google Cloud Platform [GCP]) decide to upgrade or add new services.
2. Splunk Cloud is highly performant and scales very simply as your needs grow, but the way they scale is fixed. With Splunk Cloud, as you grow, you add capacity in units called Splunk Virtual Cores (SVCs), which are “units of capability in Splunk Cloud that include compute, memory, and I/O resources.” Additional compute includes additional memory and I/O, whether or not you need those resources.
3. Uptime isn’t something you can control or directly address. Splunk provides “an uptime SLA for Splunk Cloud and will use commercially reasonable efforts to make the Services available.” Splunk Cloud is considered to be available if you can run a search, regardless of whether it’s ingesting data at the time.

Another factor is **cloud lock-in**. Splunk Cloud is available on either AWS or GCP and the features of Splunk Cloud depend on which underlying cloud you choose. Significant capabilities are only available with one cloud provider or the other.

To that point, your data is also tied to a single cloud provider. We all know that data gravity is a real thing, and organizations are increasingly choosing multiple clouds to avoid lock-in. Many organizations don’t want dozens or hundreds of terabytes of critical business data tied up on a single cloud, and they know that data migration is a headache they’d like to avoid. Data isn’t very portable, and you can’t control where your data resides.

[6] https://www.splunk.com/en_us/software/splunk-cloud.html

[7] https://docs.splunk.com/Documentation/SplunkCloud/latest/Service/SplunkCloudservice#Differences_between_Splunk_Cloud_and_Splunk_Enterprise

Finally, **costs** can also be an issue. Wasted cloud spend from idle or oversized resources totals billions of dollars and because cloud is so simple, it's easy to discover, over time, that you're overrunning your budgets, especially for a solution like Splunk that's becoming more widely utilized in your organization.

Infrastructure Trade-offs

The trade-offs between running Splunk on-premises and running Splunk Cloud are not so different from the debates that have been around since public cloud became a mainstream option:

- Dedicated hardware: high performance, highly flexible & controllable, but more complex to manage, less agile
- Cloud: Highly scalable and simple to use, but lower flexibility and control

But neither offer all the desired advantages. It's hard to get high performance, high simplicity, high flexibility and high control in a single package.

And it's also hard to manage costs.

	Dedicated Infrastructure	Splunk Cloud
Performance	High	Variable
Flexibility	High	Fixed
Control	High	Low
Ease of Management	Low	High
Dynamic Scalability	Low	High

So what do you do if you want everything? If you want a performant solution that's flexible AND under your control? If you want something that's easier? If you want to drive down TCO?

Well, you step away from conventional approaches to Splunk and turn to Diamanti. That's what we'll address in the next chapter.

.

CHAPTER 5: EXPLORING DIAMANTI FOR SPLUNK

Introducing the Diamanti platform.

We all know that Kubernetes applications can run on essentially any infrastructure. That's really the point of them.

But that doesn't mean that setting up Kubernetes infrastructure is easy. Doing so requires specialized knowledge and a new approach. Using traditional data center servers and building it yourself routinely results in infrastructure that doesn't perform as well as it might, costs more than it should, and is harder to set up, manage, and maintain than it needs to be.

Until now, it's been hard to choose an infrastructure platform that's purpose-built for Kubernetes. That's what Diamanti delivers. It aims to give enterprises a leading-edge infrastructure for containers with Kubernetes orchestration, that's turnkey, works out-of-the-box, and offers amazing benefits.

At a high-level, here's how it works.

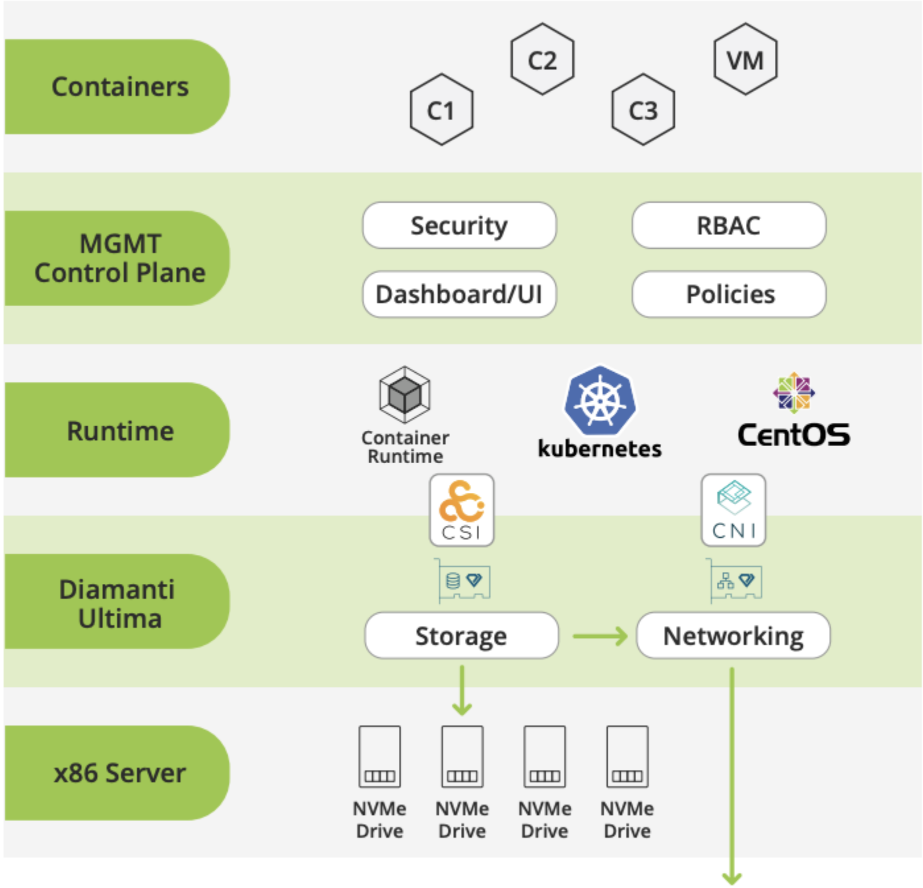

Figure 5. Diamanti platform overview

Comprised of a software platform, distinctive hardware offload accelerators, and x86 hardware, it's a hyperconverged infrastructure solution intended to strike a balance between the advantages of commodity hardware and open source software on the one hand, and the benefits of distinctive innovations on the other hand.

At the bottom of the stack is the **Diamanti D20X** family of hyperconverged hardware.

Figure 6. Three nodes of Diamanti D20X 1U HCI hardware

Diamanti D20X is leading-edge hardware that uses an x86 platform with right-sized configurations of that includes:

- Two Intel™ Xeon Scalable processors, up to 52 physical cores per socket
- Up to 768 GB of memory
- 960 GB of storage for software and ephemeral storage
- Up to 32 TB of NVMe solid state drives

Diamanti D20X hardware nodes aren't simply made up of commodity x86 hardware. Instead, Diamanti adds powerful **Diamanti Ultima** PCIe hardware accelerator cards which offload storage and networking I/O from the CPU.

- The Diamanti Ultima Network Card provides 4x10GbE ports via QSFP+ module and provisions virtual network interfaces on demand.

- The Storage Card accelerators offload storage processing, perform volume management and deliver enterprise-class storage services like snapshots, backup and restore, synchronous mirroring and asynchronous replication.

Diamanti Ultima cards are at the heart of Diamanti's distinctive capabilities. They free up system resources for dramatically improved performance by removing storage and networking processing from x86 CPUs. With full visibility into storage and networking traffic, Diamanti Ultima can deliver Quality of Service (QoS) guarantees that can't be met with commodity x86 hardware alone. These cards ensure that organizations like yours have greater control of network and storage resource allocations across the cluster, improving consistency, enabling multi-tenancy, and eliminating noisy neighbor problems.

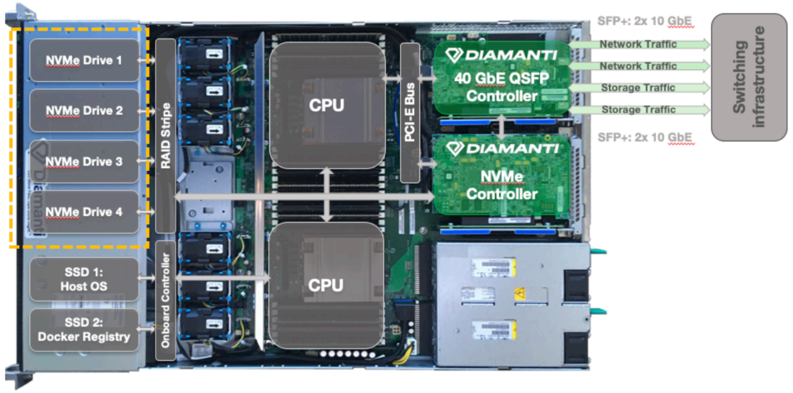

Figure 7. Diamanti Ultima PCIe offload cards

Let's turn to software now. **Diamanti Spektra** is the software that binds together Diamanti clusters with all the capabilities you need to succeed. Spektra includes:

- An underlying operating system (CentOS)
- Container runtimes (Docker and CRI-O)
- A certified Kubernetes distribution and alternatively, support for Red Hat® OpenShift®
- A security layer, including encryption, role-based access control, and integration to Active Directory and LDAP
- Integrated container storage interface (CSI) and container network interface (CNI) plugins designed for bare metal Kubernetes and optimized for use with Diamanti Ultima offload cards
- A management dashboard with enterprise visibility and control plane that can manage multiple Kubernetes clusters spanning on-premises and public cloud, giving you a new way to create a hybrid cloud
- Observability from the container level all the way up to multiple clusters
- Secure multi-tenancy to manage multiple isolated tenants and their respective resources
- Built-in data protection for stateless and stateful application migration and disaster recovery

As turnkey, hyper-converged container infrastructure, Diamanti environments easily scale on demand, integrate with your existing data center network, and provide a patented I/O-optimized architecture that delivers transformational performance for Splunk on containers.

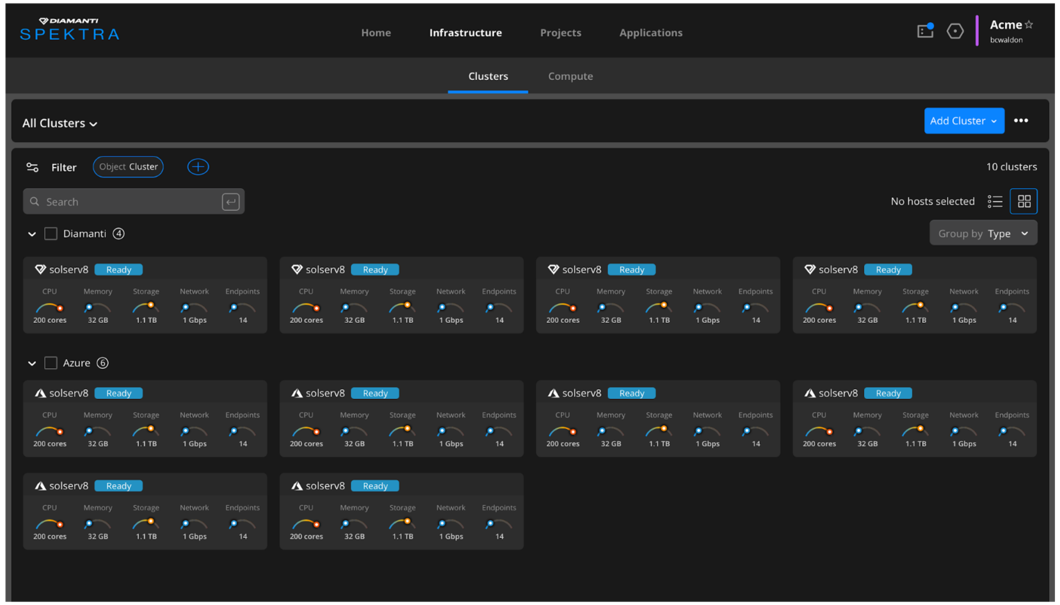

Figure 8. Diamanti Spektra management console for multi-cluster, multi-cloud Kubernetes management

Diamanti is fully compatible with Splunk Validated Architectures (SVAs). Right away, using Diamanti, you can create containerized Splunk at scale. One verified architecture is that of a distributed cluster deployment in a single site - known as a C1 or C11 SVA:

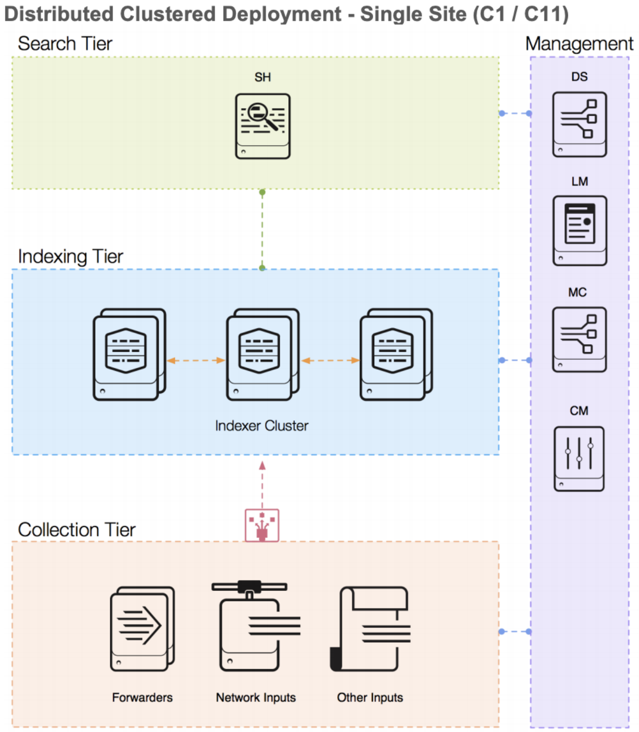

Figure 9. Splunk Validated Architecture for single site, distributed deployment (C1 / C11)

This topology includes a scalable indexer cluster in conjunction with an appropriately configured data replication policy. This provides high availability of data in case of indexer peer node failure. To simplify this deployment even further, Diamanti has collaborated with Splunk experts to deliver Helm charts for automating the implementation and configuration of Splunk Enterprise and Kubernetes per best practices to produce a consistent and stable environment.

With Diamanti, you have a new tool for creating stable, efficient, repeatable deployments without added complexity or compromise.

To make things more interesting, Diamanti has the power to run the Splunk Data Stream Processor (DSP) for real-time stream processing. You can run descriptive analytics and stream analytics on the same platform and infrastructure. You also have the capability to utilize DSP to apply artificial intelligence and/or machine learning on data before indexing and processing.

Finally, we support Splunk Enterprise for both container native virtualization and Kubernetes-native deployments. You can leverage the same persistent storage and advanced storage management for both at the same time, cutting operational complexity and costs.

We know this all sounds good, of course, but the rubber has to meet the road. What are the actual benefits? Let’s find out in Chapter 6.

CHAPTER 6: DISCOVERING THE BENEFITS OF DIAMANTI AND SPLUNK

Let's begin thinking about Splunk on Diamanti with the same qualities we've explored with other approaches to Splunk: performance, simplicity, flexibility, control, and of course, cost.

Performance

It's easy to say that your platform is faster than someone else's platform, but all too often, those advantages don't hold up in real life. Diamanti has, across a wide range of in-field use cases, determined that it can sustain 1,000,000 IOPS per 1U node, offers consistent 100μs latency, and delivers industry-leading application-level TPS. For Splunk, these performance numbers add up. We've seen 24x faster indexing on a 32 node Diamanti cluster with 200TB of storage. You can achieve an indexing rate of more than 1 TB/hour with just 16 indexers. Our platform offers so much performance that it empowers you to process more data and gather near-real-time insights by reducing index latency.

Simplicity

Diamanti is the only out of the box, ready to go Kubernetes solution. That gives organizations a way to sidestep the hours and hours it takes to set up an infrastructure for Kubernetes. Since we support SVAs and supply the Helm charts, there's no complex configuration, and our reference architectures make it easy to go from setup to success. Instead of spending weeks configuring your infrastructure and even more getting Splunk Enterprise deployed, you'll spend just hours.

As you continue to use and grow Splunk, unlike other infrastructures, we provide guaranteed QoS so that you aren't forced to second-guess your design choices. It's also easy to add more performance — just add a node to your cluster and it's automatically discovered, provisioned, and you can use Kubernetes to move containers to it.

Finally, you have 24x7 access to our Kubernetes and Splunk experts. Leaders in application containerization, it's like having a team of consultants at your beck and call, who can help you with your questions and accelerate your time to value.

Flexibility

It's easy to tune your cluster to your requirements. The Diamanti D20X family supports various configurations with more or less processing power, memory, and storage, so you can scale to what you need. We support container-native virtualization and Kubernetes-native Splunk and cloud-native architectures, giving you a bridge to hybrid cloud. In addition, since containers and Kubernetes are widely supported, there's no vendor lock-in -- you can leverage containers to make your Splunk environment more portable.

Control

You have complete control over the infrastructure, software, and Splunk layers. It's easy to tune, easy to setup, monitor and manage, and easy to maintain. You aren't bound by the limits of Splunk Cloud.

Costs

With 100% host utilization and 95% usable storage capacity, we've been able to improve on application performance and container density. You're able to do more with less — to cut down on TCO that's tied to infrastructure size and complexity. For Splunk, that's translated into a 70% reduction in TCO. That advantage gives you additional room to grow, accelerate performance, invest in other solutions, and just simply do what you need to do without worrying about unpredictable upgrade costs or cloud overruns.

In short, we deliver on our promises. But let's see what a real customer has done with Splunk on Diamanti.

	Dedicated Infrastructure	Splunk Cloud	Diamanti
Performance	High	Variable	High
Flexibility	High	Fixed	High
Control	High	Low	High
Ease of Management	Low	High	High
Dynamic Scalability	Low	High	Medium

CHAPTER 7: NBCUNIVERSAL, DIAMANTI, AND SPLUNK

BUSINESS PROBLEM

NBCUniversal (NBCU) is an international film and TV studio and distributor. To support the growing consumer shift to digital and mobile consumption of media and entertainment, NBCUniversal knew they needed to have better insight into their software development process and bring developer and operators closer together. They leveraged Splunk to deliver better insights on their internal processes, but soon ran into limitations on their architecture.

TECHNICAL CHALLENGE

NBCU's Splunk application was a critical tool for their DevOps efforts, but their existing environment was underperforming – it was limited in its architecture to ingest 1 TB per day. Like many organizations, their data sets were constantly growing. 1 TB/day wasn't enough — they wrestled with a huge backlog of unprocessed data and had lost the ability to drive actions in real time. The limits of their physical infrastructure and deployment methods were limiting scale.

DIAMANTI SOLUTION

NBCU deployed a 32-node Diamanti cluster with a total usable capacity of 200 TB and used the bare-metal platform to run Splunk. The Diamanti platform offered much greater performance per node than the previous platform. By removing hypervisors and other unnecessary layers of abstraction, the Diamanti platform — capable of delivering 1,000,000 IOPS per 1U node — delivered exceptional Splunk performance without expensive, complicated, and inflexible overprovisioning.

> *"Diamanti increased our application performance without code changes and allowed us to consolidate infrastructure while automating application deployment by our development team."*
>
> Ramana Mantravadi, SVP and CTO
> NBCUniversal Digital Products and Interactive Media

RESULTS

Faster ingest and indexing: NBCUniversal started processing approximately 1 TB/hour -- instead of 1 TB/day.

Decreased footprint: NBCUniversal reduced their physical Splunk infrastructure to 1/6th of its original size, cutting down on power and cooling, rack utilization, and monitoring/maintenance overhead.

Significant cost savings: NBCUniversal experienced an 80% savings in total cost of ownership.

CHAPTER 8: CONCLUSION

So we hope you've learned everything you needed to know about Splunk, Splunk on containers, and Splunk on Diamanti. Just to recap...

- In Chapter 1, we explored Splunk, what it does, and why it matters.
- For Chapter 2, we considered the secrets of Splunk, understanding the framework, some challenges, and what SVAs are.
- We used Chapter 3 to cover using Splunk in containers.
- Chapter 4 gave you a view of different ways to deploy Splunk, their advantages and difficulties.
- In Chapter 5, we gave you an overview of Diamanti: what it is and what it does.
- Chapter 6 we devoted to all the advantages of using Diamanti for Splunk.
- Finally, chapter 7 was a short case study, detailing how NBCUniversal used Splunk on Diamanti.

There are a range of ways you can learn more about what we do, but the simplest is to check out our Splunk and Diamanti landing page at https://diamanti.com/use-cases/splunk.

Other resources to check out:

- Hyperfast Data-to-Everything with Splunk and Diamanti (Solution Brief)
- Diamanti Spektra (Datasheet)
- Diamanti at Cloud Field Day 8 (Video playlist)
- What's New in Diamanti Spektra 3.0 (website)

DIAMANTI

About Diamanti

Diamanti delivers purpose-built infrastructure for modern applications. The Diamanti platform is the first and only Kubernetes solution integrated with a patented I/O-optimized architecture, delivering transformational application performance. With Diamanti, Kubernetes becomes an out of the box solution, allowing organizations to focus on deploying modern applications across on-premises and hybrid cloud infrastructure. Based in San Jose, California, Diamanti is backed by venture investors ClearSky, CRV, Engineering Capital, Goldman Sachs, GSR Ventures, Northgate Capital, Threshold Ventures (formerly DFJ Venture), and Translink Capital. For more information visit www.diamanti.com or follow @DiamantiCom

Contact sales for more information: https://diamanti.com/contact/