Control Virtual Chaos

A New Approach for Managing Virtual Environments

WHITE PAPER

splunk>

Abstract:

Virtualization has revolutionized the modern datacenter. Whether it be network, server, application or desktop virtualization, each offers numerous benefits such as cost savings, physical server consolidation, dynamic load balancing, ease of migrations and more. While these benefits are compelling, virtualization has also introduced a new level of complexity to managing the datacenter. Visibility, or a lack thereof, is probably the biggest challenge.

With virtualization, datacenter administrators lack the necessary visibility to help them solve problems faced by their application owners. Capturing and storing all the relevant data at full fidelity is critical to truly understanding application performance, especially when you have mission-critical applications running in virtualized environments. Visualizing this data within the context of data from other technology tiers is critical to understanding exactly which events in which tier are causing problems and impacting performance. Correlating, trending and analyzing virtualization data and data from other technology tiers such as storage, networks, and operating systems is a big data problem. Only a big data technology can deliver the visibility required for this complex and transient technology landscape.

This whitepaper discusses the visibility challenges facing today's administrators and highlights a new approach to managing, reporting and analyzing the silos of big data generated by virtualized infrastructures.

Is Virtualization Leaving You in the Dark?

Although virtualization eases the creation, change and movement of infrastructure, it can create numerous blind spots in the environment.

For example: an application owner might find his or her virtual machine reporting a lack of resources. However, given that metrics are stored in full fidelity for only 1-2 hours in traditional virtualization management tools, the virtualization admin may not have the right visibility into hypervisor-level metrics to detect the associated problem.

In another example, a virtualized host might run into latency problems while accessing back-end storage and the problem only surfaces as a "very slow application," the root cause of which may only surface upon inspection of logs on the ESX (for example) host.

Some of the typical challenges faced by administrators of virtual environments include:

- **Lack of performance data in full fidelity** – While vCenter (for example) is a great administrative tool, granular performance data (metrics for 20 second time intervals) is available for only 1-2 hours, after which it is aggregated into five minute and hourly summaries. With VMs frequently in motion, historical data becomes essential during troubleshooting. Access to granular historical data becomes essential for problem resolution, trending and analytics.
- **Lack of visibility into virtual machines and other infrastructure components** – Performance data from the virtualization layer only provides a part of the answer—application, operating system, server & storage performance and event data often play a critical role in understanding problems, thresholds and trends.
- **Absence of an operational analytics engine** – Troubleshooting a virtualized environment requires a different class of data management and analysis technology. The highly dynamic and complex nature of virtual environments requires a holistic perspective—one that goes wide across the hosts within the virtual cluster or datacenter and also goes top to bottom to capture metrics from the operating systems, applications, network, storage, I/O and capacity. Trending, analyzing and correlating across these diverse sources of data requires a powerful big data engine.

Even though virtualization brings about exciting new dynamics to infrastructure management, it creates new challenges in effectively managing and planning the physical and virtual infrastructure. Gaining insight into your virtual deployment and making essential correlations with the applications and other parts of the infrastructure is vital to efficiently managing your resources and gaining the benefits of virtualization.

Traditional Tools Fall Short

The tools that are designed for physical infrastructure typically do not work well with the dynamic and decentralized nature of virtual environments. The biggest challenge with the virtual environments is getting information from vCenter Server (VC). While VCs may contain abundant information about the virtual environment, accessing information from the VC can potentially clog and slow down the VC itself. The VC also does not contain the necessary level of information required to intelligently correlate information from the virtual environment to the physical host.

Specialized virtualization tools focus exclusively on the virtualization layer and are unable to add context about other technology tiers in your environment. As an example, an application that uses cache memory for performance optimization might find that the hypervisor does not report cache memory as "used memory." Virtualization admins will deliberately undercount and under-provision in these scenarios, because they don't have visibility into operating system metrics.

It becomes essential to look for technology agnostic tools that will:

- Consolidate information from all tiers of your distributed architecture
- Retain granular performance metrics from the virtualization hosts when necessary and summary metrics where appropriate
- Augment performance metrics with logs and asset inventory information
- Correlate machine data coming from various sources such as Virtual Centers, hypervisors, applications, network, infrastructure, operating systems, storage, etc.
- Discover, alert and report on the virtual assets to monitor usage and performance, plan capacity, track changes, optimize assets and control the lifecycle intelligently and automatically
- Deliver advanced operational analytics with the ability to adapt and customize easily

Virtualization and Splunk

Splunk software helps make sense of machine-generated data by capturing and analyzing massive amounts of structured and unstructured time-series data from any source, such as network traffic, custom applications, application servers, hypervisors, storage systems, networks and more. Splunk's proprietary implementation of MapReduce is designed for massive horizontal scalability—to terabytes of indexing per day. With no dependence on a backend database, Splunk lets you create a schema on the fly to allow unlimited (and different) views of the same data and to run queries on data without needing to understand its structure.

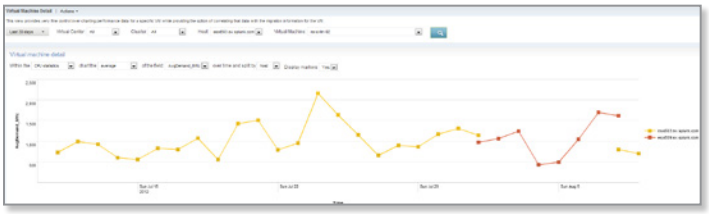
Splunk's powerful search and reporting language allows you to visualize, analyze, report, trend and correlate on large and diverse datasets and create powerful, information-rich reports. Using Splunk's easy-to-create dashboard views and flexible reports that summarize data or display information in real time, Splunk users can harness machine data for customized reporting, trending and big data analytics and to gain new levels of operational visibility and intelligence about their environment.

Splunk App for VMware: A Reality Check for Your Virtual World

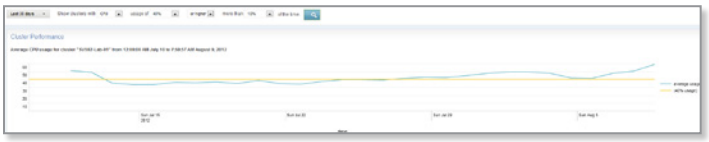
A natural extension to the Splunk platform is the Splunk App for VMware. This App extends the data collection, indexing and analytic power of the Splunk platform to VMware environments. The Splunk App for VMware provides a mechanism to collect granular data about the virtualization layer directly from ESX/i hosts and Virtual Centers and store it in full fidelity within Splunk. It can then be analyzed in the context of data from all the other technology tiers such as operating systems, storage, networks and applications. The App allows administrators to proactively plan for capacity, monitor changes, track assets, report on security and troubleshoot issues.

In addition to performance data, the Splunk App for VMware also collects log data which contains important clues about hypervisors, storage, networking and other failures. With topology information and task & event data from VCs, this solution provides a means to audit usage and changes across your virtual deployment.

With over 30 reports providing details on inventory, performance, tasks, events and security, and out-of-the-box troubleshooting views with direct access to the ESX/i hosts and VC logs, the App provides a unified view of your virtual assets.



Trend Over Time of Virtual Machine Performance by Host



Over Time Capacity report on Cluster Performance

Solving Your Virtualization Woes with Splunk

Data illuminates your view of any issue or problem in virtualized environments. With Splunk software, you can go back in time and recreate the state of the environment, drilling down to root cause and issue resolution. In addition, Splunk gives you the flexibility to retain as much data as you need at appropriate levels of detail for as long as you need (in full fidelity or aggregated), without any performance impact to vCenter Servers or ESX/i hosts.

Getting the data into Splunk and visualizing it alongside data from other technology tiers helps to accurately pinpoint the root cause of any problem—whether an application failed, or a virtual machine resides next to a noisy neighbor, or if a hypervisor is sharing resources inefficiently, or any other infrastructure issue.

The same data can also provide you with a holistic view of your security posture and help you address compliance reporting requirements. Any search can be turned into an alert with Splunk. Once you discover an issue, you can set up proactive alerts to help keep reoccurring issues from causing havoc in your environment. Thus, you can use Splunk to help prevent downtime, outages, security threats or just resolve tickets.

Deploying a big data analytics solution like Splunk helps use the virtualization data for more than just issue resolution; with Splunk's powerful analytics, reporting and dashboarding capabilities, this data can be used for operational insights such as capacity planning, resource utilization reporting on SLA compliance, customer usage and more.

Conclusion

Although virtualization offers extensive flexibility and efficiency in workload management, it is crucial to have visibility and monitoring of this environment in order to realize its full potential. To proactively manage virtualized environments, it is critical to go beyond just the virtualization layer. A big data approach is needed to provide visibility, reporting and analytics across these environments. Splunk, a big data analytics engine, allows you to comprehensively gain top to bottom visibility of your physical and virtual infrastructure, make informed decisions and mitigate risk.

Free Download

Download Splunk for free. You'll get a Splunk Enterprise license for 60 days and you can index up to 500 megabytes of data per day. After 60 days, or anytime before then, you can convert to a perpetual Free license or purchase an Enterprise license by contacting sales@splunk.com.

250 Brannan St, San Francisco, CA, 94107 info@splunk.com | sales@splunk.com 866-438-7758 | 415-848-8400 www.splunkbase.com