SevOne

[WhitePaper]

6 STEPS

TO AN EFFECTIVE PERFORMANCE MONITORING STRATEGY.

Overview

Whether you're looking to improve your application and service delivery, consolidate existing performance monitoring tools and responsibilities or justify the impact of a new technology deployment, the following six steps can help you create the fundamental building blocks of an effective performance monitoring strategy:

1. Collect
2. Baseline
3. Alert
4. Report
5. Analyze
6. Share

Breaking down your strategy into these components will make it easier to understand, articulate and reach consensus on the performance monitoring requirements for your business. This whitepaper also highlights the questions you should ask performance monitoring vendors to ensure they support your strategy.

1 COLLECT.

Any performance monitoring strategy starts with data collection. If you can't monitor it, you can't manage it. It sounds simple enough, but there are many obstacles that prevent proper data collection and leave you with visibility gaps. To prevent this, look for a monitoring platform that supports the following data collection criteria:

Any time series data

Your performance monitoring platform should be data agnostic. It should be able to ingest any time series data, regardless of source. This includes support for collection data via SNMP, NetFlow, IP SLA, WMI, JMX, NBAR, Sylog and more. The industry has begun to move away from supporting standard protocols like SNMP. So you should prepare to collect non-standard performance metrics from third party sources: network probes, proprietary business applications, cloud platforms, or element management systems (EMS) from network equipment vendors. If your data has a time stamp, it should be simple to ingest it into your platform and normalize it with other data sources.

High frequency polling

Traditional five minute polling cycles are insufficient in many instances. A spike in traffic that lasts only a couple seconds represents less than 1% of a five minute polling cycle. The anomaly will be completely flattened and undetectable when averaged out over that time span. Yet this brief spike can disrupt business transactions, VoIP communications, and other latency-sensitive applications. You need a monitoring solution that allows for high frequency polling down to the second. Even one minute polling cycles may be insufficient. Bell Mobility, a wireless provider in Canada, revealed an interesting internal study. It showed average spikes in traffic were as much as 350% higher when viewed at 1.5 second rates compared to 60 second polling intervals.

Raw data retention

Granular data collection is only useful when you can maintain that data for a sufficient time frame. Some monitoring solutions may average and consolidate historical data over time for storage reasons. This results in a lesser understanding of historical events. It also weakens your ability to forecast future capacity needs with accuracy. Many organizations seek a performance monitoring platform that maintains a year of as-polled data. You should not have to invest in extra storage capacity to do this.

Massive scale

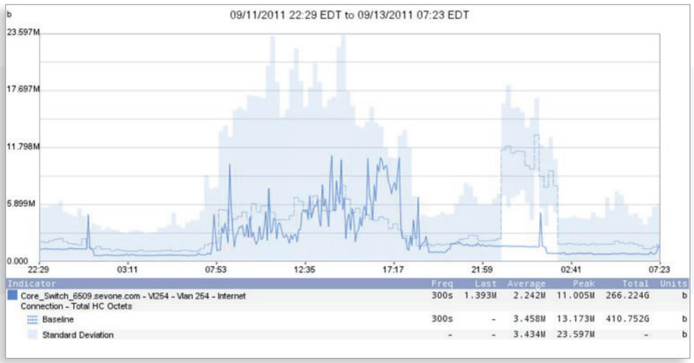
We live in the age of Big Data. Applications, systems and network devices produce massive volumes of machine data. The rise of virtualization and cloud services exacerbates the issue. As you spin-up new resources faster than before, your environment produces exponentially more data. Your performance monitoring platform must scale with your data collection needs. Inability to scale your solution forces you to make tough decisions about what you will and won't monitor. This creates a visibility gap. You never know what might go wrong in your environment. So taking a broad approach to data collection supported by a highly scalable platform is the best strategy.

2 BASELINE.

Once you've collected the broadest set of performance data at the required granularity, it's time to establish a baseline for "normal" performance. Your performance monitoring platform should do this automatically for every metric you collect. This includes baselines for unstructured data, such as logs.

Baselines provide a historical reference point for every 15 minute time frame, every day of the week. With baselines, you can compare real-time infrastructure performance to historical norms. You're also able to view capacity trends and deviations that cause performance-impacting events.

Virtualized data centers with rapid elasticity make baseline technology that much more important. Users may provision and retire resources commensurate with demand and without human intervention. This rapid fluctuation in your environment presents monitoring challenges. It's imperative to understand what normal looks like at any given moment. Baselines then become your basis for a more effective alerting method.



Indicator	Freq	Last	Average	Peak	Total	Units
Core_Switch_6509.sevone.com - Vl254 - Vlan 254 - Internet Connection - Total HC Octets	300s	1.393M	2.242M	11.005M	266.224G	b
Baseline	300s	-	3.458M	13.173M	410.752G	b
Standard Deviation	-	-	3.434M	23.597M	-	b

Performance baselines help you understand how your real-time performance compares to historical norms. They also serve as the foundation for a more effective alerting methodology.

3 ALERT.

You may employ two types of alerts: those based on static thresholds and those based on deviation from baseline performance.

Static thresholds are useful in cases such as wanting to know when a CPU exceeds 95% utilization for a period of 15 minutes or more. Some performance monitoring tools may only allow you to set an upper-level threshold. You should look for a solution that allows you to specify both an upper- and lower-level range. For example, you'd want to know when the voltage on a UPS falls outside of a specified range. Or when the temperature of a device exceeds a high or low manufacturer recommendation.

But you may not always understand what an acceptable range is for performance of a specific device or metric. Often, administrators guess at threshold values. This results in a significant increase in noise from false positive alerts.

In an environment that generates a lot of noise, it's more effective to alert when performance of any metric deviates from historical norms. For example, if your company always runs a backup procedure at 3:00 a.m., you don't want a daily alert about high bandwidth usage. But you would want an alert when an unexpected spike occurs during working hours due to a unique user-initiated action. You should be able to specify how many standard deviations you consider acceptable for any metric. This requires an understanding of baseline historical performance for all metrics monitored.

This method provides a more reliable predictor of service-impacting events. It answers the most critical question: "What is happening in my environment right now that is unique that I need to know about?"

4 REPORT.

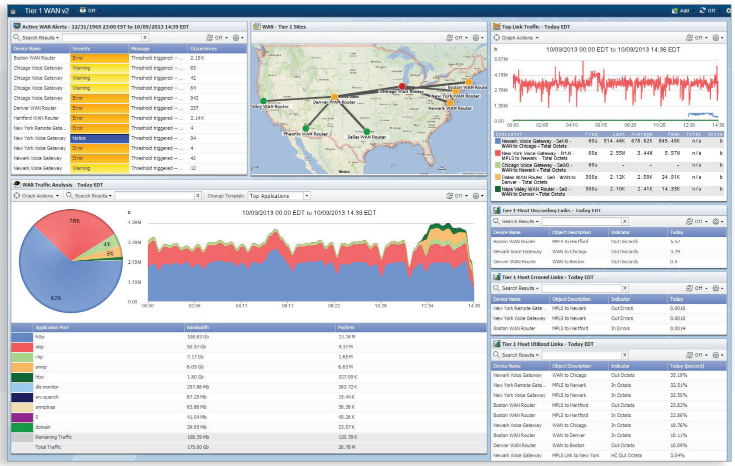
You typically access reports in one of two ways: canned reports provided via a saved template or ad-hoc reports generated to respond to specific business conditions.

Many vendors deliver canned reports that reveal most utilized interfaces, highest packet loss and other key metrics. While these reports are helpful, they may not allow for the level of manipulation required. Troubleshooting the more challenging performance scenarios demands extreme report flexibility

Be sure to ask the following questions of any reporting solution:.

- Can you state which resources, settings, time frames, visualizations and summary information you need to view, on the fly?
- Can you display many disparate data points in a single report or dashboard that allows for visual correlation of events? For example, can you include SNMP data, flow data, device configuration information, fault data from traps and third party metrics from an Element Management System -- all in the same dashboard or report?
- Is there a limit to the number of objects you can present in a single report?
- Can you chain different report graphs together? If you change the time span for one graph, will it update all other graphs in the report at the same time?

Place a high value on report flexibility, as it is the best way to ensure you provide actionable insight to your teams. You should be able to combine any performance metrics in a single dashboard, on the fly. This may include status maps, alert notifications, TopN reports, NetFlow charts, and other data sources.



Active WAN Alerts - 12/31/1969 23:00 EST to 10/09/2013 14:39 EDT

Device Name	Severity	Message	Occurrences
Boston WAN Router	Error	Threshold triggered -- ...	2.15 K
Chicago Voice Gateway	Warning	Threshold triggered -- ...	65
Chicago Voice Gateway	Warning	Threshold triggered -- ...	42
Chicago Voice Gateway	Warning	Threshold triggered -- ...	64
Chicago Voice Gateway	Error	Threshold triggered -- ...	943
Denver WAN Router	Error	Threshold triggered -- ...	257
Hartford WAN Router	Error	Threshold triggered -- ...	2.14 K
New York Remote Gate...	Error	Threshold triggered -- ...	4
New York Voice Gateway	Notice	Threshold triggered -- ...	84
New York Voice Gateway	Error	Threshold triggered -- ...	4
Newark Voice Gateway	Error	Threshold triggered -- ...	42
Newark Voice Gateway	Warning	Threshold triggered -- ...	12

WAN Traffic Analysis - Today EDT

10/09/2013 00:00 EDT to 10/09/2013 14:39 EDT

	Application Port	Bandwidth	Packets
🟦	http	108.83 Gb	13.18 M
🟥	skip	50.57 Gb	4.37 M
🟩	rtip	7.17 Gb	1.65 M
🟧	snmp	6.05 Gb	6.63 M
🟩	hbci	1.80 Gb	327.09 K
🟦	dls-monitor	257.86 Mb	383.72 K
🟦	src-quench	67.25 Mb	15.44 K
🟨	snmptrap	63.86 Mb	36.38 K
🟪	0	41.04 Mb	45.26 K
🟩	domain	39.65 Mb	33.57 K
⬜	Remaining Traffic	108.39 Mb	120.78 K
	Total Traffic	175.00 Gb	26.78 M

Top Link Traffic - Today EDT

10/09/2013 00:00 EDT to 10/09/2013 14:36 EDT

Indicator	Freq	Last	Average	Peak	Total	Units
🟦 Newark Voice Gateway - Se1/0 - WAN to Chicago - Total Octets	60s	514.46K	678.62K	845.45K	n/a	b
🟥 New York Voice Gateway - Et1/0 - MPLS to Newark - Total Octets	60s	2.55M	3.44M	5.57M	n/a	b
🟩 Chicago Voice Gateway - Se0/0 - WAN to Newark - Total Octets	60s	-	-	-	n/a	b
🟧 Dallas WAN Router - Se0 - WAN to Denver - Total Octets	300s	2.12K	2.50K	24.91K	n/a	b
🟩 Napa Valley WAN Router - Se0 - WAN to Denver - Total Octets	300s	2.10K	2.41K	14.33K	n/a	b

Tier 1 Most Discarding Links - Today EDT

Device Name	Object Description	Indicator	Today
Boston WAN Router	MPLS to Hartford	Out Discards	5.52
Newark Voice Gateway	WAN to Chicago	Out Discards	3.19
Denver WAN Router	WAN to Boston	Out Discards	0.5

Tier 1 Most Errored Links - Today EDT

Device Name	Object Description	Indicator	Today
New York Remote Gate...	MPLS to Newark	Out Errors	0.0018
New York Voice Gateway	MPLS to Newark	Out Errors	0.0018
Boston WAN Router	MPLS to Hartford	In Errors	0.0014

Tier 1 Most Utilized Links - Today EDT

Device Name	Object Description	Indicator	Today (percent)
Newark Voice Gateway	WAN to Chicago	Out Octets	35.19%
New York Remote Gate...	MPLS to Newark	In Octets	32.51%
New York Voice Gateway	MPLS to Newark	In Octets	32.50%
Boston WAN Router	MPLS to Hartford	Out Octets	23.83%
Boston WAN Router	MPLS to Hartford	In Octets	22.86%
Newark Voice Gateway	WAN to Chicago	In Octets	10.76%
Boston WAN Router	WAN to Denver	In Octets	10.11%
Denver WAN Router	WAN to Boston	Out Octets	10.09%
Newark Voice Gateway	MPLS Link to New York	HC Out Octets	3.04%

Finally, you need to understand how increases to the number of objects you monitor impacts the speed of your reporting platform. Reports that fail at providing near real-time information are unacceptable. Performance monitoring solutions that rely on a centralized database architecture suffer significant degradation to reporting speed as your monitored domain expands. The centralized reporting database becomes a bottleneck, unable to process report requests with speed. That is why reports sometimes take minutes or hours rather than seconds. It is best to maintain information in a distributed fashion and have the system query the data when needed.

5 ANALYZE.

Data analysis and visualization may take a number of unique forms. At the end of the day, you're looking for actionable insight that allows you to:

- Detect and avoid performance events before they impact end users or customers
- Fine tune your infrastructure to make the most of current resources
- Make more informed decisions about the impact the infrastructure has on your business

Shift to Proactive

Proactive analysis and troubleshooting means you have a better chance of avoiding service-impacting events. For example, in Step 3 we reviewed the value of receiving alerts based on performance deviation (as opposed to static thresholds). But what if you could receive a notification anytime a unique log was generated in your environment? Like an error code on a router reboot or a user accessing an application for the first time. This type of automated analysis helps you stay a step ahead of your end users. Look for ways to make the shift from reactive troubleshooting to proactive analysis that helps you avoid performance events in the first place.

Understand Correlations

Many organizations struggle with performance analysis because their platform fails to provide dashboards and reports that can present disparate data sources in a single view. This makes correlation problematic. You end up referencing multiple screens in an attempt to understand causation. Plan a strategy that allows you to automatically pivot from traditional performance metrics – such as SNMP or IP SLA – to insightful flow or log data.

Forecast Capacity

Capacity teams expect data analysis to reveal which resources are near exhaustion and how much time they have to complete a resource upgrade. How much bandwidth does your business need? How close to maximum utilization are your servers? Which network interfaces will be most utilized 30 days from now?

The ultimate tool for WAN and Data Center capacity teams is a report that reveals the number of days until specific resources reach a user-defined threshold. For example, a capacity planner would like to see all resources that will exceed 80% utilization in 30 days or less. This gives them the ability to plan upgrades based on an understanding of the lapsed time required to complete the upgrade.

Remember, when it comes to capacity planning, volume isn't everything. You must also comprehend the composition of that traffic. Understanding the type of activity taking place can make a big difference in investment plans and monetization strategy. Capacity planning is a numbers game. But the best projection models take into account the value of different types of traffic.

6 SHARE.

You now have a strategy to collect, baseline, alert, report and analyze your performance data. But who can benefit from your insight? Your audience needs may vary:

- Customers through a secure self-service portal
- Co-workers with a .pdf report
- Capacity teams with a .csv export
- Management and executives via real-time dashboards

Providing data for the sake of data is not a productive strategy. The most important lesson to keep in mind when it comes to sharing performance data is ***know your audience***. A CTO does not want to get bogged down in granular metrics about a particular link's capacity or some other device status. They're much more interested in a service-level or even market-level view of performance. For example, "What percent of time are we within SLA compliance?" or "How do my multiple carriers impact performance of my top applications?" Often, a dashboard featuring a simple status map with red/yellow/green indicators for current service health suffices.

Sharing information also means sharing data with other platforms, such as a fault or configuration management solutions. It should be just as easy to export data as it is to ingest it. Does your vendor supply an API for communicating with other platforms as part of your maintenance plan, or is that an additional expense?

SUMMARY

The journey from raw performance monitoring data to actionable insight about the health of your infrastructure is a critical path. The more you can refine the process to eliminate wasted time and energy, the more successful you'll be. We often think of this cycle as Mean Time to Repair (MTTR), but it can be more accurately stated as Mean Time to Action. Ultimately, you're looking to make decisions on the best possible data in the shortest amount of time. This requires a clear understanding of the steps in the process and how you can shave critical moments off each stage.

Breaking down your performance monitoring process into these six fundamental components provides clarity around your strategy. It allows you to answer questions like:

- "How much and what types of data should I prepare to collect?"
- "Do I understand what normal performance looks like in my environment?"
- "How can I see when something is happening right now that I need to know about?"
- "Can I understand exactly what happened at any time in the past year?"
- "Am I able to comprehend the future capacity needs of my organization?"
- "Who needs this information and how can I make it actionable for them?"

By understanding the core requirements of a monitoring strategy, you also arm yourself with the knowledge to make an informed buying decision when evaluating performance monitoring vendors.

About SevOne.

SevOne provides the world's most scalable infrastructure performance monitoring platform to the world's most connected companies. The patented SevOne Cluster™ architecture leverages distributed computing to scale infinitely and collect millions of objects to provide real-time reporting down to the second and help organizations prevent outages. SevOne customers include seven of the world's 13 largest banks, enterprises, CSPs, MSPs and MSOs. SevOne is backed by Bain Capital Ventures. More information can be found at www.sevone.com. Follow SevOne on Twitter at @SevOneInc.