

Storage Workload Analysis – The WorkloadWisdom Way

By Logan Harbaugh, senior analyst
Deni Connor, founding analyst
SSG-NOW

December 2, 2015

Workload performance analysis has been around for quite a while on data networks. It involves capturing network traffic, analyzing the traffic, particularly in the context of network applications, and can help determine causes of a spike in traffic, increases in latency, or other issues, as well as whether it's an isolated anomaly or a regular or irregular event. The tools that exist are from specific vendors and only work in networks involving their switches, network adapters or routers.

The same tools for analyzing storage workloads until now have largely been vendor-specific too. They are proprietary and work with FC SANs and IP SAN networks that also involve a particular switch, storage system or host bus adapter, and are unable to characterize traffic over a heterogeneous storage network implemented with gear from many different storage vendors.

There are further complications with the workload analysis tools. In data networks, the tools that are available for Ethernet may not work well with iSCSI or Fibre Channel over Ethernet. The same can be said for the tools for storage networks – some work well with Fibre Channel networks, but lack iSCSI capability. Other tools handle block protocols well, but lack an understanding of objects and file data (SMB and NFS). And, further, many tools are not able to correlate performance problems with the behavior of application workloads, such as Oracle or SAP.

Lack of visibility into workload behavior

Storage traffic is also more difficult to characterize as belonging to a particular workload or application than network traffic. Since they are focused on a specific type of hardware (switch or HBA), existing tools could characterize traffic between a host and a particular storage device or LUN, but not identify the workload behavior that could be causing a transient or persistent problem. For example, most tools are unable to capture information on the block-level, file-level or object level data, let alone find hot spots associated with particular data patterns.

Typical changes to workload behavior (Upgrades, etc.)

In addition to diagnosing problems, workload analysis is also critical as a predictive tool for predicting changes in storage infrastructure performance due to new applications or increased use of existing applications – the upgrades that might be needed to support new traffic loads, the increases in latency that adding more applications might cause or the trends in latency and throughput as applications are utilized to an increasing degree. The ideal scenario would be to not only capture the storage traffic associated with a particular workload mix, but to generate the same type of traffic at increased levels, to simulate

an increased numbers of users of a particular application, the addition of more databases to an existing server, or the addition of a new application to an existing storage system.

With proprietary tools, the storage administrator is also strapped at discovering and troubleshooting storage infrastructure problems. For instance, they can tell that storage traffic is spiking at 2:00PM on Thursday, but they then have to discover for themselves what application, infrastructure, or workload change is creating the spike.

There haven't been many solutions that look at workload acquisition and workload analysis from a holistic standpoint. There haven't been solutions that support both iSCSI and Fibre Channel protocols, file and object protocols and provide visibility into switches, hosts and storage arrays, collect and replay data over extended periods of time, and be able to characterize traffic as belonging to specific application workloads. Until now, that is.

How Virtual Instruments implements workload analysis

Virtual Instruments has been shipping a Workload Generation Appliance and its WorkloadWisdom software (formerly known as Load DynamiX Enterprise) application since 2013. The solution includes workload modeling, load generation and performance validation, enabling storage architects and storage engineers to optimize the performance, availability and cost of their storage infrastructure. The company has just added several new appliances and software capabilities to its family of workload modeling and performance validation tools. They are expanding their focus to the IT operations teams in addition to the storage engineering and architecture teams.

The Workload Data Importer, part of WorkloadWisdom, collects workload and performance data from storage device log files or vendor performance monitoring tools. It collects historical data for any period of time as determined by the storage manager. This data is then sent to the Workload Analyzer that is also included with WorkloadWisdom. The Workload Analyzer statistically analyzes the workload data and automatically creates one or more workload models that can be used as a basis for storage performance planning.

iSCSI Log Data 27-Jul-2015

Importing Log Data Started on 07/27/2015 at 1:07 PM by Samantha Carter
 using analysis policy iSCSI Analysis

Elapsed Approx. Remaining

21:30

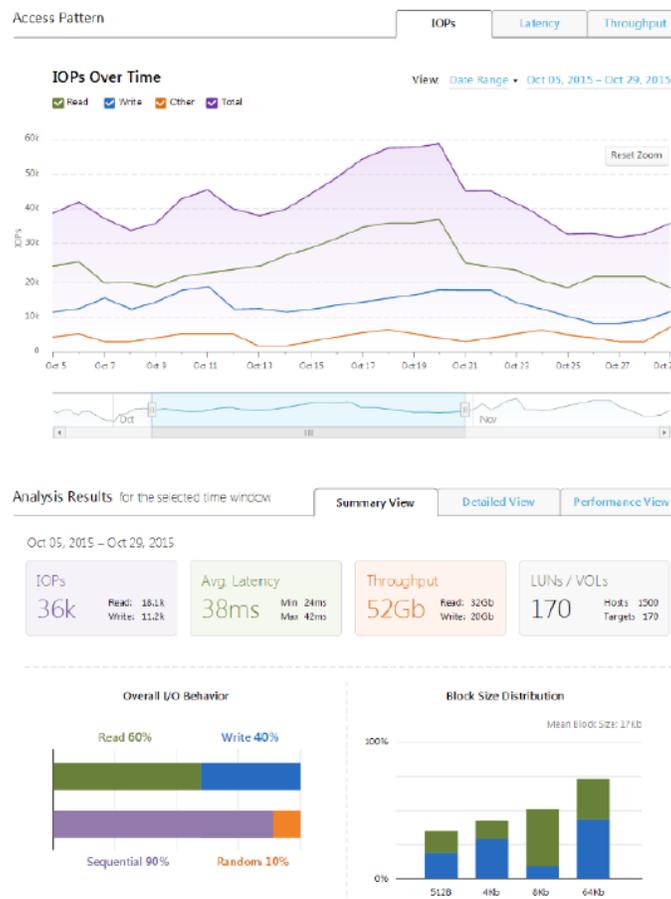
5:40

Last Log Record: 07/27/2015 1:07:30 PM | Success | Project start successful [View Log](#)

Analysis Results (7)

Date/Time	Target Range	Clustered by	Algorithm	# of clusters	# of workloads	<input type="checkbox"/>	<input type="checkbox"/>
Not started yet	Whole Test	Time	dbscan	<input type="text" value="waiting for import completion"/>		<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Not started yet	Cluster with Max Throughput	Device	dbscan	<input type="text" value="waiting for initial analysis results"/>		<input type="checkbox"/>	<input type="checkbox"/>
Not started yet	Whole Test	Time	k-means	<input type="text" value="waiting for dbscan results"/>		<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Not started yet	Cluster with Max Throughput	Device	k-means	<input type="text" value="waiting for initial analysis results"/>		<input type="checkbox"/>	<input type="checkbox"/>

The output from the Workload Data Importer is used by the Workload Analyzer software, shown below, which automatically creates highly accurate workload models that can be replayed in a lab environment. The combination of these three new capabilities, paired with the Workload Generation Appliance, is extremely powerful, as it allows the capture of both historical and real-time data, allowing for evaluation and troubleshooting of existing storage systems and networks, predictive analysis to model the performance impact of changes and upgrades in applications, and optimization of storage infrastructure to avoid over-provisioning of storage.



Since performance results are correlated with workload data, administrators can build multiple profiles, including a wide variety of parameters such as block size, command mix, queue depths, compression and dedupe ratios, and so forth, and test with multiple profiles, either simultaneously or in sequence, building up simulations of real world traffic under multiple scenarios. It becomes easy to test for issues such as spikes in traffic during holidays or other events, simulate the effects of adding additional applications to an existing server, or modeling the efficacy of adding flash storage to reduce latency.



SSG-NOW Take

Organizations with large storage estates can use the Workload Data Importer to troubleshoot existing problems and to perform predictive analysis to support new applications or increases in traffic on existing applications, including testing with increased traffic using workload profile data from existing apps. The devices typically pay for themselves by allowing companies to buy only what is needed rather than over-provisioning storage systems in the hopes of avoiding bottlenecks.

By determining the type of upgrades that will actually produce the desired results, administrators could provide hard data on the need for additional storage, or for adding flash memory (or determining why a hybrid system would work better than all-flash system). The ability to provide tabulated hard data to justify storage expenses would well make storage architects lives much easier.

Service organizations could also install the devices in client networks, quickly find problems or capture traffic data, and then move the devices to new client, offering a level of service not available otherwise. The data collected and the ability to simulate increased traffic levels would provide the organization with real justifications to add capacity and improve planning, as well as troubleshooting transient problems with a degree of accuracy not available with simple log tools, as well as the ability to correlate particular storage activity with specific applications. While extremely experienced storage troubleshooters might be able to associate specific kinds of data traffic with a particular application in some cases, the WorkloadWisdom system can do this for all types of applications and traffic. ●

About Storage Strategies NOW™

Storage Strategies NOW™ (SSG-NOW) is an industry analyst firm focused on storage, server, cloud and virtualization technologies. Our goal is to convey the business value of adopting these technologies to corporate stakeholders in a concise and easy-to-understand manner.

Note: The information and recommendations made by Storage Strategies NOW are based upon public information and sources and may also include personal opinions both of Storage Strategies NOW and others, all of which we believe to be accurate and reliable. As market conditions change however, and not within our control, the information and recommendations are made without warranty of any kind. All product names used and mentioned herein are the trademarks of their respective owners. Storage Strategies NOW, Inc. assumes no responsibility or liability for any damages whatsoever (including incidental, consequential or otherwise), caused by your use of, or reliance upon, the information and recommendations presented herein, nor for any inadvertent errors which may appear in this document.