

# What are Workloads and why you should care



*by George Crump, Lead Analyst*

Data Centers are under increasing pressure to build infrastructures that are responsive to the application workloads of the organization. Understanding what a workload is and what its characteristics are is critical to creating a responsive infrastructure that can satisfy the variety of demands being placed on IT by an organization's customers, users and application owners.

## What is a Workload?

It is a common misconception that workloads are a synonym for virtual machines. They can be, but in most cases a workload is a set of I/O characteristics running through a group of virtual machines that interface with network and storage infrastructures. For example, an application workload may interact with a web-server, one or several database servers as well as other application servers. The combination of all of these servers and the associated networked storage makes up that application's workload. Another example is a Virtual Desktop Infrastructure (VDI), which could be comprised of several physical hosts and hundreds, if not thousands, of virtual machines.

## Understanding Workload Characteristics

Each workload will have unique characteristics, and each of these characteristics impacts storage latency, IOPS and throughput. These characteristics include:

- I/O Mix
  - is the workload read heavy, write heavy or balanced?
- I/O type
  - does the workload write or read data sequentially or randomly?
- Data/metadata mix
  - does the workload read or manipulate metadata more so than actual data?
- Block or file size distribution
  - does the workload write in small or large blocks?
- Data efficiency appropriateness
  - does the workload have highly redundant or compressible data so that functions like deduplication and compression work effectively?
- Is the workload prone to specific hot spots?
- How do all of the above characteristics change over your relevant time period?

## Understanding How Workloads Change

The only constant in modern business is change. These changes impact workloads and IT needs to understand how far their current infrastructures can be extended in order to support workloads whose demands are increasing. They also need to understand the capabilities of their next potential storage system.

Businesses can expand by adding more of the same customers, and growing the number of users through acquisition or contract thanks to consolidation. There can be spikes in demand, such as commonly seen in e-commerce sites. They can also branch into new business opportunities that change the workload's characteristics. When looking at storage performance, changes at the virtual machine and server level, such as OS updates or HBA queue depth settings, can have a significant impact on the workload characteristics.

## The Importance of Workload Knowledge

Each of these workload characteristics will demand something different from the storage system. For example, while a data set with highly redundant data may benefit from deduplication and compression to lower its capacity requirements, it may also make the storage system controllers work harder because it has to manage all the links to redundant data, which in turn could impact performance.

The biggest challenge however, is when the workloads share the same storage system. The more varied these I/O characteristics are, the more difficult it is for a single storage system to be able to support them all. In contrast, if workload characteristics are understood and only a specific type is going to be used on the storage system, then the storage system can be fine-tuned to that specific use case. A classic example might be video streaming applications which benefit from large block sizes.

The reality is most data centers have a variety of workloads they need to support. The challenge is IT professionals don't have tools available to them that provide insight into the I/O characteristics of each workload in their environment. The result is data centers either buy multiple storage systems to distribute workloads across or they overbuy on a single storage system so it can support all of their workloads simultaneously.

## Using Workload Knowledge to Build Better Infrastructures

It is critical that IT professionals have the ability to capture, analyze and forecast workload I/O profiles. Armed with workload data, IT professionals can better optimize their current environment, plan their storage refresh cycles based on actual storage demands instead of arbitrary four year storage refresh cycles and have a baseline from which to evaluate new storage systems.

## Acquire

The first step in understanding workload I/O characteristics is to capture the I/O profiles of each workload. The temptation here is to use monitoring tools provided by the application.



The problem with this approach is that they provide a very myopic view of the workload and only report the I/O characteristics of the particular server that the monitoring tool is running on. For the same reason, storage system I/O reporting tools should be avoided because they only report the result.

Remember, workload I/O profiles are the result of multiple physical or virtual machines generating I/O simultaneously, not just one. IT professionals should look for a tool that collects I/O characteristics from every component of the workload.

## Analyze

The next step is to analyze the I/O profiles captured for each workload by presenting the results into a single graphical interface. The interface should allow the overlaying of workload profile data and even artificially combine them to determine what the impact of mixing the workloads are. Armed with the profiles of each workload IT professionals are ready to take action to resolve performance issues, optimize their current storage infrastructure and to make better storage system choices in the future.

## Act

The final step is to act on the workload information now available to the data center. Workload I/O profiling should be a continuous capture of data, not just a one-time informational event. The continuous capture of workload I/O characteristics enables using the information to troubleshoot workload performance problems as they occur.

For example, if a workload owner blames the storage infrastructure of causing a performance problem, the storage manager can now play back the workload profiles on that storage system at the time in question. They can see if there was I/O generated from that or other workloads that caused an I/O problem. They could also show there was no significant I/O and point the problem back at the application. If the problem was I/O related the storage manager would have the information available to them that would allow them to move a conflicting workload to another storage system.

Another action that IT professionals armed with workload profiles should take is optimizing the existing storage environment. Again, with continuous analysis and modeling of I/O data, the workload information can be played back to look for predictable moments in time. This playback capability provides insight into where specific workloads could be moved to a higher performance tier of storage (flash for example) or look for workloads that never stress the storage system that they are on and downgrade those to a lower performing, less expensive, tier.

Finally, a clear understanding of the I/O characteristics of the data center's workloads allows an IT planner to test properly systems being considered for a storage system refresh. As Storage Switzerland discussed in its article, "[The Value of an Independent Storage Performance Testing Platform](#)", the biggest challenge facing IT planners during a storage system refresh is how to test the various systems under consideration with their production workloads.

IT planners today are often too busy or don't have the lab resources to test each storage system properly that they are considering. As a result, they count on vendor supplied data to make their choice. Even if they do have the time and some resources available, as we discussed in our article, "[You can do better than ioMeter and Vdbench](#)", the rudimentary freeware testing tools have little relationship with the data center's actual workloads. Without real world simulation of workloads, the testing of storage systems provides virtually no insight into how the storage system will actually perform in your data center.

With the workload I/O characteristics captured, the I/O profile data can be modeled to be used with an I/O load generator. The I/O load generator is typically a single appliance that can simulate the I/O profiles of multiple workloads. Instead of playing back the I/O to a GUI, as is the case in the diagnostics described above, the workload generator will play the I/O profiles onto the evaluated storage systems.

The workloads can be tested individually, in combination or all at once, to see how the storage system will handle the load. The I/O generator also can modify the workload to simulate growth in I/O demand and worst case scenarios, providing insight to how far in the future the given storage system will meet the organization's SLAs. Solutions such as those from companies like Load DynamiX provide an integrated workload analysis, modeling and load generation platform that solves these issues.

## Conclusion

A workload is the combined I/O of a distributed application often serviced by multiple servers. Without the proper tools, understanding the I/O requirements of these multi-tier workloads is difficult, and comparing the impact of multiple, frequently changing workloads is almost impossible. The ability to capture workload I/O characteristics, analyze that data and regenerate it is a critical capability for data centers to master. Workload profiling enables organizations to troubleshoot and optimize their current environment as well as plan, for the future.

*Sponsored by: Load DynamiX*

## About Load DynamiX

Load DynamiX is a Storage Performance Analytics innovator providing unique insight into application workload performance to empower storage professionals to optimize costs and assure performance by more intelligently deploying and troubleshooting storage infrastructure. The combination of advanced workload analysis and modeling software with extreme workload generation appliances give IT organizations the ability to cost-effectively validate and stress today's most complex networked storage infrastructure to its limits. See more at: <http://www.loadynamix.com/company/#sthash.gR93dz2w.dpuf>