

SOLUTION BRIEF

Infinidat InfiniBox[®] and SAS[®] Analytics

SAS is one of the leading analytic platforms for AI, Machine Learning and business intelligence. The resources required to support the SAS platform, such as CPU, memory and I/O, are very predictable and isolated in such a way that each can be optimized to improve performance, reliability, and availability for all phases of SAS workloads.



Optimizing SAS Performance

Of all resources that affect performance, the most influential in any SAS environment is I/O. You can add memory or CPUs to a server farm, but if the I/O sub-system is not capable of keeping up with the workflow, the other resources could be under-utilized waiting for I/O to complete. It is important when choosing an enterprise storage platform to look for a provider that will perform well supporting every phase of the SAS processing cycle.



The SAS Processing Cycle

DATA ACCESS: Getting the data from a source to the SAS environment requires large block writes to the SAS servers.

ETL: In this phase, data is then mapped, transformed and loaded into SAS data structures.

USAGE: This is where the user community begins to analyze, synthesize and distill the data to provide the answers SAS was designed to deliver.

Each of these phases have characteristics that have vastly different I/O requirements, and each utilizes the storage system in different ways.

I/O Testing Tool: To give the end user a sense for what a particular I/O system can support, SAS provides an I/O testing tool that simulates these three phases. The [SAS support page](#) describes the tool, how to use it, and how to interpret the results.

Throughput Rates: Performance of your SAS Analytics deployment is determined by the throughput rates for each of the 2 SAS file systems: **/saswork** and **/sasdata**. This document presents test results for running these two file systems on an InfiniBox enterprise storage system. At a minimum, SAS recommends an average throughput rate of 125MB/s per physical core in the system.

SCALE TO WIN

INFINIDAT

TEST 1: RESULTS FOR SINGLE SERVER RUNNING SAS

This test was conducted using a single, large bare-metal Linux system with 2 volumes configured: /sasdata and /saswork. The server had 20 cores and 128GB of RAM. There were 4 x 16GB Fiber Channel HBA's connecting to the InfiniBox F63xx. The rhel_iotest.sh script was run against each file system.

SAS I/O Testing Tool Results:

READ THROUGHPUT RATE: 308.12 megabytes/second per physical core

WRITE THROUGHPUT RATE: 256.91 megabytes/second per physical core

InfiniMetrics®, the InfiniBox performance monitoring tool, shows the I/O performance through the entire SAS Processing cycle.



Performance Notes:

- ▶ The green line shows peak read throughput rate is over 6GB/s, with very low latency
- ▶ The red line shows peak write throughput rate is 5.3GB/s

The graph shows 3 phases of processing: Data copy, ETL, and read data. Data copy and ETL are write intensive, with data copy using 1MB write block size. The Reads are small block sequential reads.

TEST 2: RESULTS FOR 3 SERVERS RUNNING SAS

The second test was to show how much SAS workload can be placed on a single InfiniBox F63xx. Each server is a large bare-metal Linux system, as described in Test 1, used to drive the the SAS workload, run simultaneously on all 3 servers. The results are as follows:

Server 1: SAS I/O Testing Tool Results

READ THROUGHPUT RATE: 262.23 megabytes/second per physical core

WRITE THROUGHPUT RATE: 225.86 megabytes/second per physical core

Server 2: SAS I/O Testing Tool Results

READ THROUGHPUT RATE: 293.67 megabytes/second per physical core

WRITE THROUGHPUT RATE: 230.52 megabytes/second per physical core

Server 3: SAS I/O Testing Tool Results:

READ THROUGHPUT RATE: 252.74 megabytes/second per physical core

WRITE THROUGHPUT RATE: 254.76 megabytes/second per physical core

InfiniMetrics shows the I/O performance through the entire SAS Processing cycle.



Performance Notes:

- ▶ Read throughput shows peak read throughput rate is over 16GB/s
- ▶ Write throughput exceeded 14.8GB/s

INFINIBOX ADVANTAGES FOR SAS

SAS requires high-performance, high-resiliency storage at scale. That system should also be easy to manage so that administrators can focus on the needs of the business instead of the needs of the infrastructure. InfiniBox excels at meeting these requirements while also delivering a low total cost of ownership (TCO) with a rapid return on investment (ROI).

InfiniBox is built on 5 fundamental principles that are a perfect match for SAS solutions:

1. High Performance

A truly innovative cache management algorithm combined with an ultra-efficient data layout delivers maximum performance while simplifying deployment and overall systems management. High throughput, at sub-millisecond latency, is the key to high performance operation as well as powering synchronous and asynchronous replication for block and file.

2. High Availability and Reliability

InfiniBox's self-healing architecture, designed to deliver unmatched availability, combined with our patented InfiniRaid® and predictive analytics, delivers seven nines uptime and non-disruptive upgrades. The InfiniBox "component group redundancy" design enables rapid recovery from any component failure without impacting performance.

3. Multi-Petabyte Scale

Maximum system capacity utilization is possible due to extremely efficient thin provisioning, continuous space reclamation, and inline data compression. Packaged in a single 42U rack and scaling to well over 8PB or more effective capacity, multiple system consolidations are easy to accomplish and remarkably cost-effective.

4. Simple and Powerful Management

An intuitive HTML5 GUI simplifies the most complex storage management operations. A comprehensive RESTful API and a powerful CLI help automate complex tasks, including policy management for quality of service. Easily facilitate service level coordination across tenants, workloads, and volumes. Monitor and measure all feature performance elements using InfiniMetrics.

5. Low Total Cost of Ownership (TCO) and High Return on Investment (ROI)

Infinidat offers a TCO of 30-50% less on a price per TB basis compared the competition. According to Forrester's Total Economic Impact (TEI) assessment⁴, InfiniBox delivers an average ROI of 125% on a payback period of less than six months.

⁴ <https://www.infinidat.com/news/press-releases/independent-study-finds-infinidat-storage-can-provide-customers-125-roi-over-three-years-and-payback-in-less-than-six-months/>

INFINIDAT INFINIBOX FOR SAS

The InfiniBox enterprise storage solution delivers faster than all-flash performance, high availability, and capacity density at multi-petabyte scale for a multitude of mixed workloads. Zero-impact snapshots, synchronous and asynchronous replication, and data-at-rest encryption assure maximum data security and reliability.

With InfiniBox, SAS implementations are simpler to implement, operate, and protect while reducing costs and delivering the results that help companies discover, innovate, and drive their success.

Key Features

- ▶ 100% availability guarantee
- ▶ Architected for triple redundancy and continuous data integrity checking
- ▶ High performance delivering over 2M IOPS & 25GB/s throughput
- ▶ Multi-protocol-FC, iSCSI, and NFSv3 all supported
- ▶ Multi-petabyte Scale-up 8PB in one rack
- ▶ Simple management-HTML5 GUI, RESTful API
- ▶ Cost effective-all features included at no additional cost

InfiniBox for SAS in Action

Configuring Infinidat storage for SAS is extremely simple. SAS requires 2 main file systems, /saswork and /sasdata. Deploying on InfiniBox requires a single volume for each. There is no reason to create multiple volumes for each which complicates the configuration. When a single volume is created on InfiniBox, using InfiniRaid, the data objects are spread across every spindle in the frame.

InfiniBox F63xx



Multiple SAS Nodes Scale-out configuration



Mount points for each

- ▶ /saswork
- ▶ /sasdata

CONCLUSION

Analytics platforms such as SAS require rapid access to critical datasets with predictable performance regardless of the workload. Since the data is vital to your business, it should also be highly reliable, easy to deploy, and scale with your business growth. The testing with SAS publicly available test tools shows that a single InfiniBox can deliver performance advantages whether running in a single SAS node or in a scale-out SAS model. SAS running on InfiniBox enables customers to focus on the needs of their business instead of worrying about the cost and complexity of the storage infrastructure by providing:

- ▶ **FASTER INGEST**-increase the ingest speed and overall amount of data ingested WHILE supporting multiple user reports.
- ▶ **HIGHEST AVAILABILITY**-capture all critical data with ultra-low latency and near instant write acknowledgement. InfiniBox availability means your SAS solution is always open for business.
- ▶ **SPEED/CAPACITY/LOWEST TCO**-with faster-than-flash speed AND petabyte-scale capacity, InfiniBox addresses the speed and capacity in a single solution, at the lowest total cost of ownership.

Coupling SAS Analytics with InfiniBox, we enable your competitive business advantage at any scale.