

FIO Performance on AWS Storage Optimized Instances

Background

NVMe has already established as storage choice for modern workloads requiring low latency and high performance. On cloud it is available as NVMe ephemeral resources that can be easily attached to computing instances. Though this option is within reach it comes with other challenges

- Storage is ephemeral in nature and cannot be used when persistence is required as the contents do not survive stop
- Compute instances cannot be stopped and it requires removal and recreating of instance if one does. Now your applications need to consider this unnecessarily.
- Only certain storage optimized instances i3,i2 have these direct NVMe SSD
- The minimum capacity is 475Gib even if the application will use only 1/10 of that capacity.
- Hinder any solution that require separation of storage and compute.
- The Elastic Block Storage for a single compute instance will peak at 40K IOPS.

MayaScale provides Elastic NVMe using NVMe Fabric over TCP on AWS Cloud for high-performance shared storage and its performance is compared against running locally using the ephemeral local NVMe SSDs.

Benchmark Setup

MayaScale works better when configured with multiple cores to provide same parallelism seen with NVMe hardware queues and also the compute instance network and storage performance is also tied to the number of provisioned CPU cores. The software has very low memory requirement only. The following i3 instances can be used for MayaScale server.

Instance Type	vCPU	Memory (GiB)	Network	Capacity	Total HW Queues
i3.large	2	15.25	~10 Gbps		
i3.xlarge	4	30.5	~10 Gbps	0.8T	4
i3.2xlarge	8	61	~10 Gbps	1.7T	8
i3.4xl	16	122	~10 Gbps	3.5T	32
i3.8xl	32	244	10 Gbps	6.9T	64
i3.16xl	64	488	25 Gbps	13.8T	248

- server node of instance type from above running MayaScale software
- client node of machine type m5a.2xlarge

FIO Performance Report

The FIO benchmark ran with options numthreads=4, iodepth=32, bs=4k

Instance Type	IO pattern	Block Size	Local IOPS	Client IOPS	Local Latency (us)	Client Latency (us)
i3.large	random write	4K				
i3.large	random read	4K				
i3.xlarge	random write	4K	71.8k	71.3K	1781.89	1793.06
i3.xlarge	random read	4K	210k	175K	609.61	727.94
i3.2xlarge	random write	4K	181K	180k	706.52	706.55
i3.2xlarge	random read	4K	413K	189k	308.56	674.64
i3.4xlarge	random write	4K				
i3.4xlarge	random write	4K				
i3.8xlarge	random write	4K				

The output shows that NVME over Fabrics is a power technology that can match the local NVMe performance with little added latency for read operations only limited by the network bandwidth of 10Gbps that was used in the testing.