Infiniflash 功能验证与性能测试

测试项目名称	Infiniflash+Nexenta 测试	项目实施地点	上海

项目名称：				
文件名：测试报告模板 V1.0				生成日期：2016-05-26
版本号	日期	作者	修正章节	变更记录
1.0	2016-05-26	Louis Liu		

测试背景

本次项目会进行 NexentaStor 和 SanDisk InfiniFlash 测试。测试的目的是证明 Infiniflash 在大规模数据仓库应用方面的可行性。

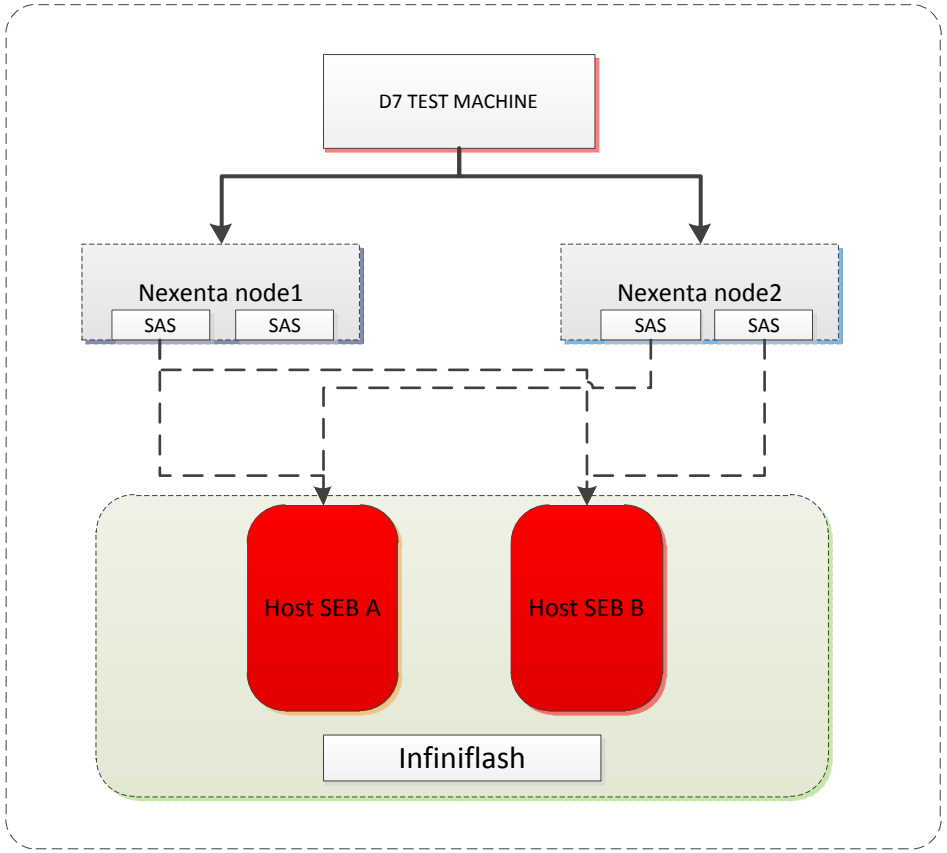
此测试分为两个主要部分：

1. 安装和配置 - 涵盖包括整个系统的硬件和软件组件的安装和配置。
2. 性能和功能测试 - 包括配置、特定功能测试，以及基于现有配置的性能验证。

测试环境

测试拓扑图

本次项目将一台 InfiniFlash 作为 Nexenta 的后端 全闪存存储，通过四条 6Gb/s SAS 进行连接。两台服务器安装 NexentaStor 软件和 HA 插件，实现 HA 环境下的相关功能测试。拓扑图如下：



硬件环境

SuperMicro 服务器配置 ：

机器	DELL PE R730 * 2
CPU	Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz 32cores
Memory	256GB (32GB*16)
OS 版本	CentOS Linux release 7.0.1406 (Core)
文件系统	ext4
SAS HBA	9266-8i MegaRAID SAS HBA

软件环境

1. NexentaStor 版本 ： Ver.40-0-61
2. 客户端操作系统版本：Linux CentOS 7.0

用途 ：

1) 压力测试
2) 功能验证测试

3. 测试工具 ：

1）FIO
2）TPCC

性能测试

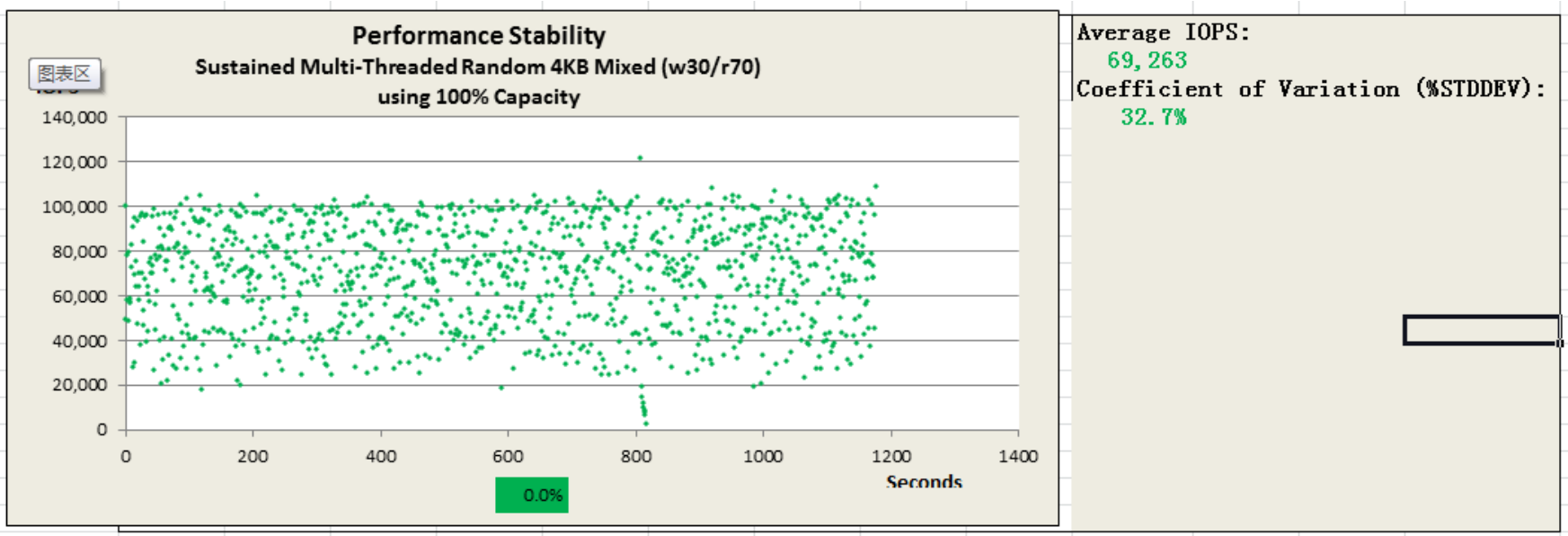
对一块 200GB 的 lun 做性能测试 ，首先用不同的块大小擦写两遍，使其完全碎片化。

Configurable Primary Properties

Size	200G The logical size of the zvol.
Block Size	16K The block size of the zvol.
Refreservation	203G The minimum amount of space guaranteed to a dataset, not including its descendents. When the amount of space used is below this value, the dataset is treated as if it were taking up the amount of space specified by refreservation. If you do not want to use this property - keep it "none".
Reservation	none The minimum amount of space guaranteed to a dataset and its descendents. When the amount of space used is below this value, the dataset is treated as if it were taking up the amount of space specified by its reservation. Reservations are accounted for in the parent datasets space used, and count against the parent datasets quotas and reservations. If you do not want to use this property - keep it "none".
Description	fiotest Human-readable description for this zvol.
Primary Cache	all Controls what is cached in the primary cache (ARC). If this property is set to all, then both user data and metadata is cached. If this property is set to none, then neither user data nor metadata is cached. If this property is set to metadata, then only metadata is cached. The default value is all.
Secondary Cache	all Controls what is cached in the secondary cache (L2ARC). If this property is set to all, then both user data and metadata is cached. If this property is set to none, then neither user data nor metadata is cached. If this property is set to metadata, then only metadata is cached. The default value is all.
Read Only	off Controls whether this dataset can be modified. Default is "off".
Compression	LZ4 Controls the compression algorithm used for this dataset. Default is "on".
Checksum	on Controls the checksum used to verify data integrity. The default value is on, which automatically selects an appropriate algorithm (currently, fletcher2, but this may change in future releases). The value off disables integrity checking on user data. Disabling checksums is NOT a recommended practice.
Deduplication	Off Controls the deduplication option for the volume. If enabled, it will optimize use of duplicate copies of data. Default is off.
Log Bias	latency Provide a hint to ZFS about handling of synchronous requests in this dataset. If logbias is set to latency (the default), ZFS will use pool log devices (if configured) to handle the requests at low latency. If logbias is set to throughput, ZFS will not use configured pool log devices. ZFS will instead optimize synchronous operations for global pool throughput and efficient use of resources.
Number of copies	1 Controls the number of copies of data stored for this dataset. Default is 1.
Sync	standard Controls synchronous requests (standard - ensure all synchronous requests are written to stable storage; always - every file system transaction will be written and flushed to stable storage by system call return; disabled - synchronous requests are disabled). Default is standard.

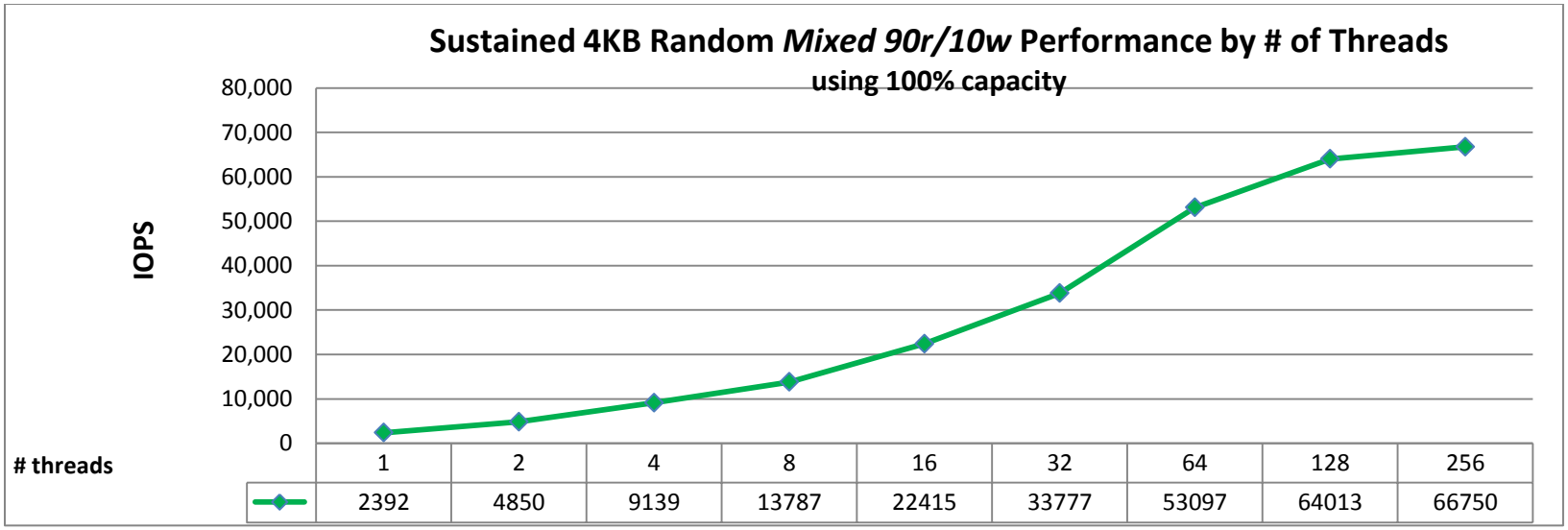
4K 随机读写 (w30,r70)

点的分布比较分散，主要由于读已经达到网卡的上限，另外随机写的性能相对较弱，平均 IOPS 还是可以接受。



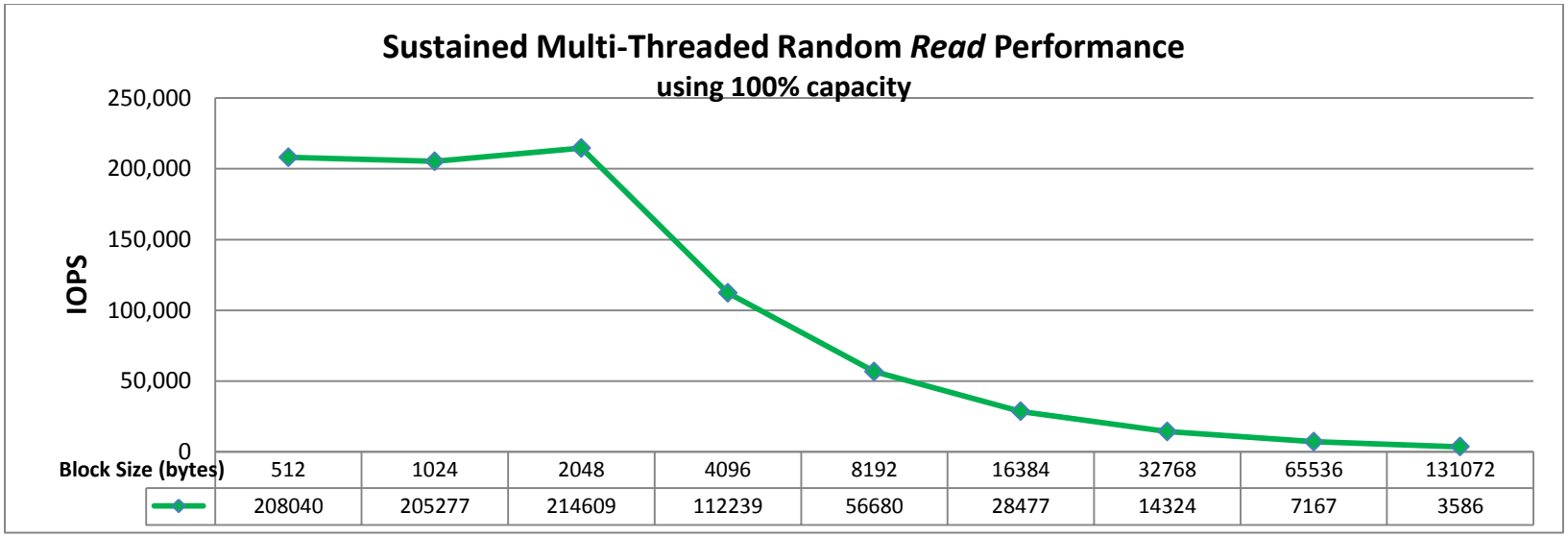
4K 随机读写 (w90 r10 ,多 thread)

当达到 128 threads 的 时候 IOPS 趋于平稳



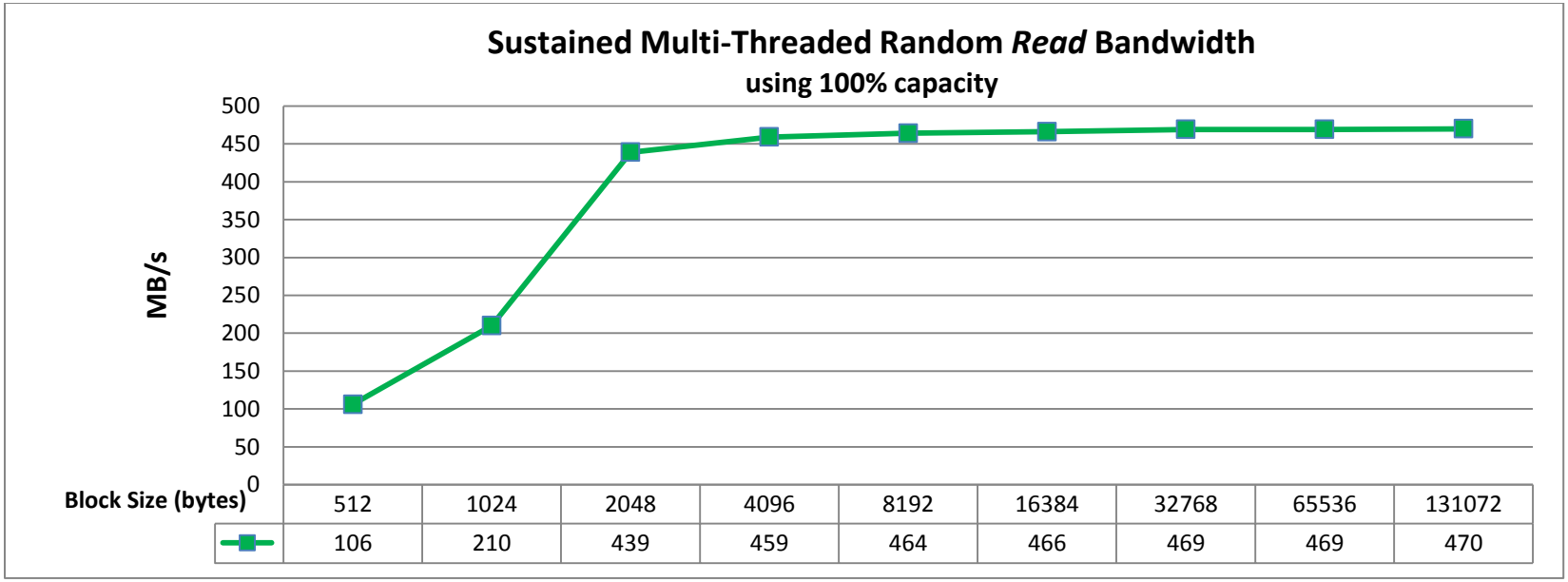
不同 block size 的随机读性能

在 512byte - 2K 的时候维持在较高的水准 后续因为网卡流量达到上限而下降



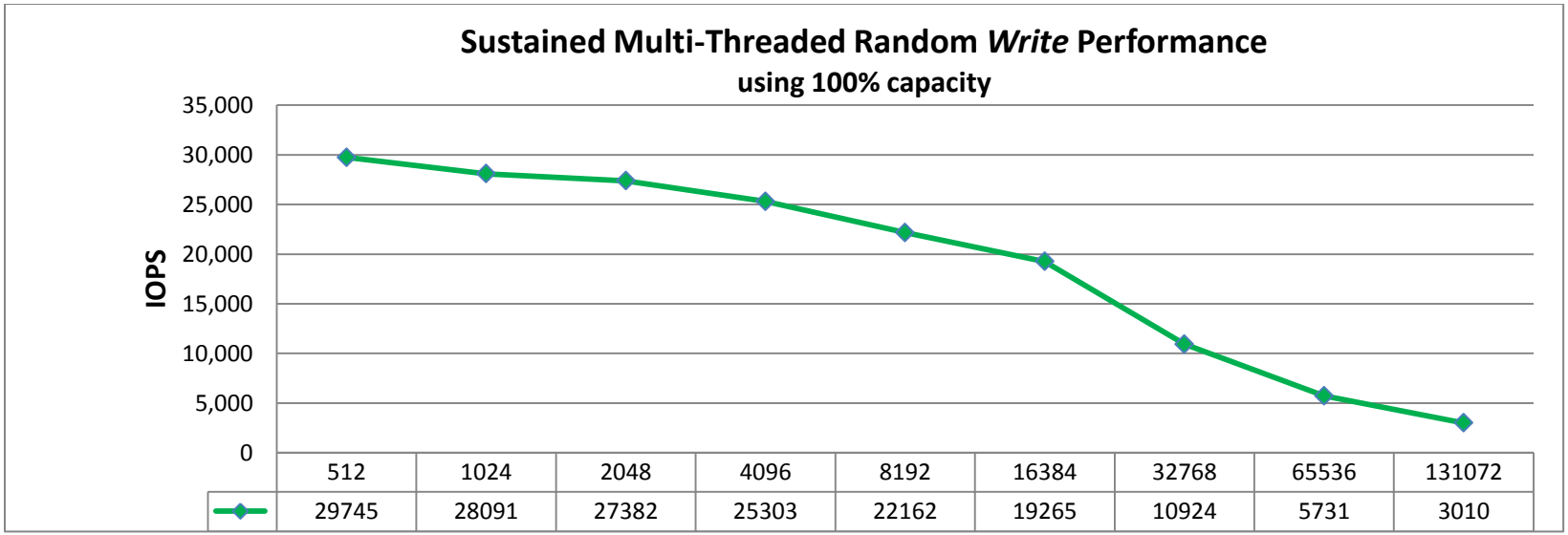
不同 block size 下的读吞吐量

最后将 4 张网卡打满，从 2048byte 开始即达到网卡的上限



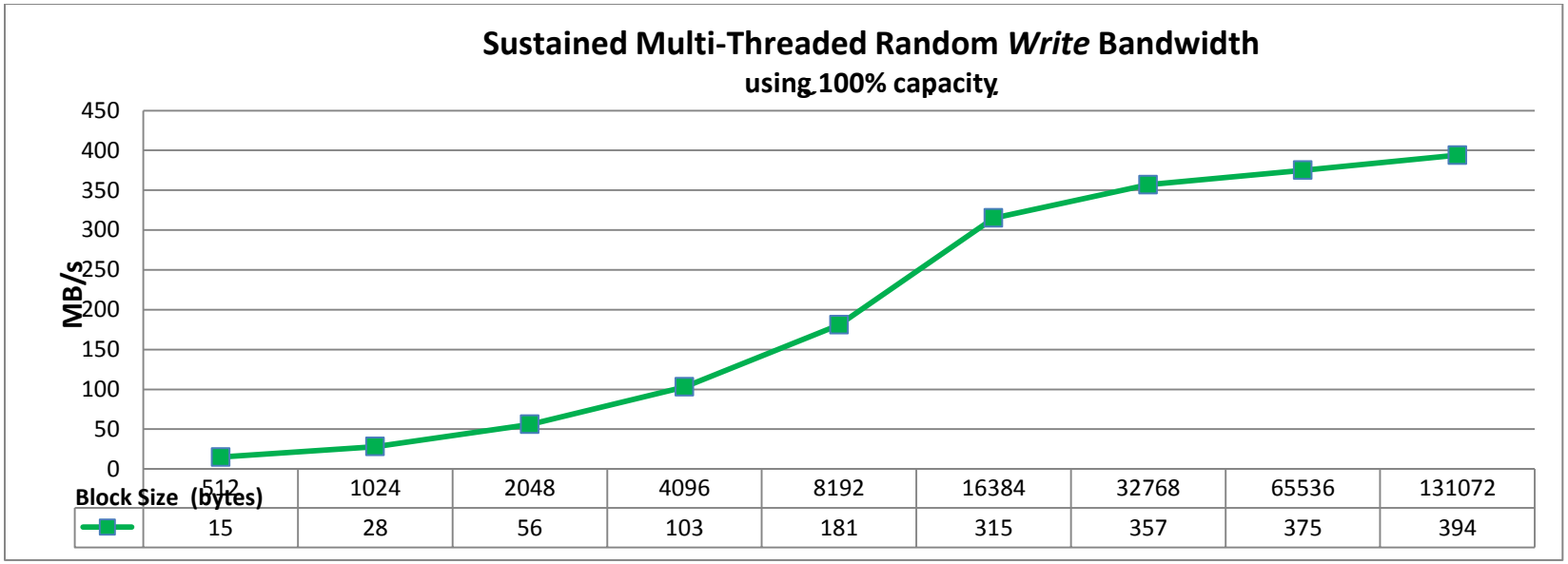
不同 block size 下的随机写性能

随机写的 IOPS 较随机读的性能所有下降



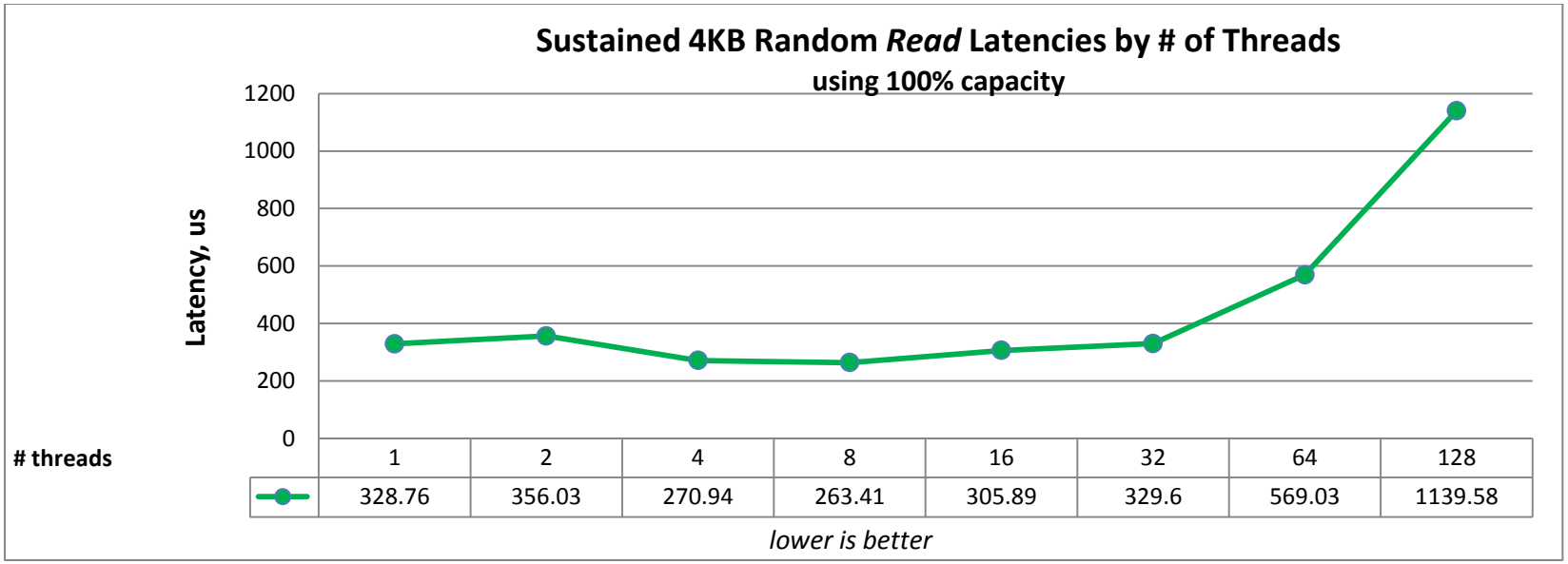
不同 Block size 下的随机写吞吐量

由于写的性能限制，并没有能将网卡打满



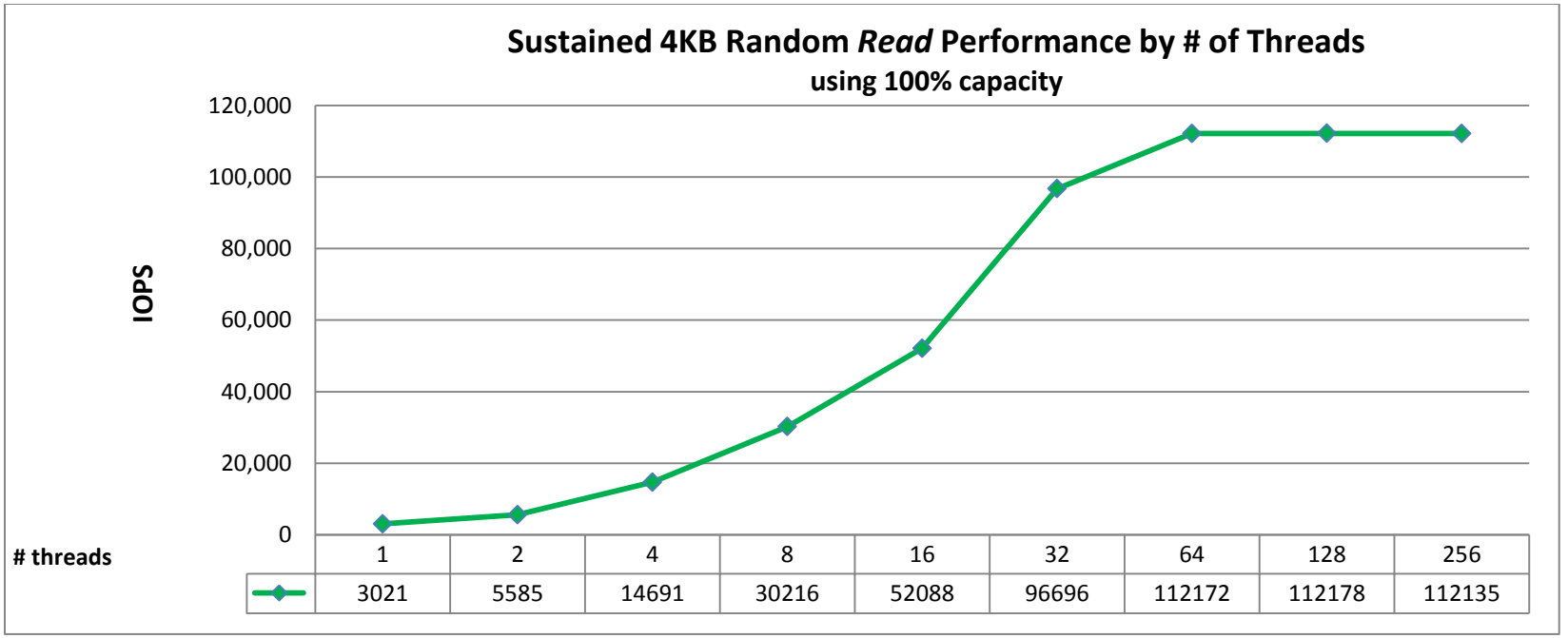
Block size 4K 情况下的读延迟

读延迟可以接受，在 128 threads 情况下为 1ms



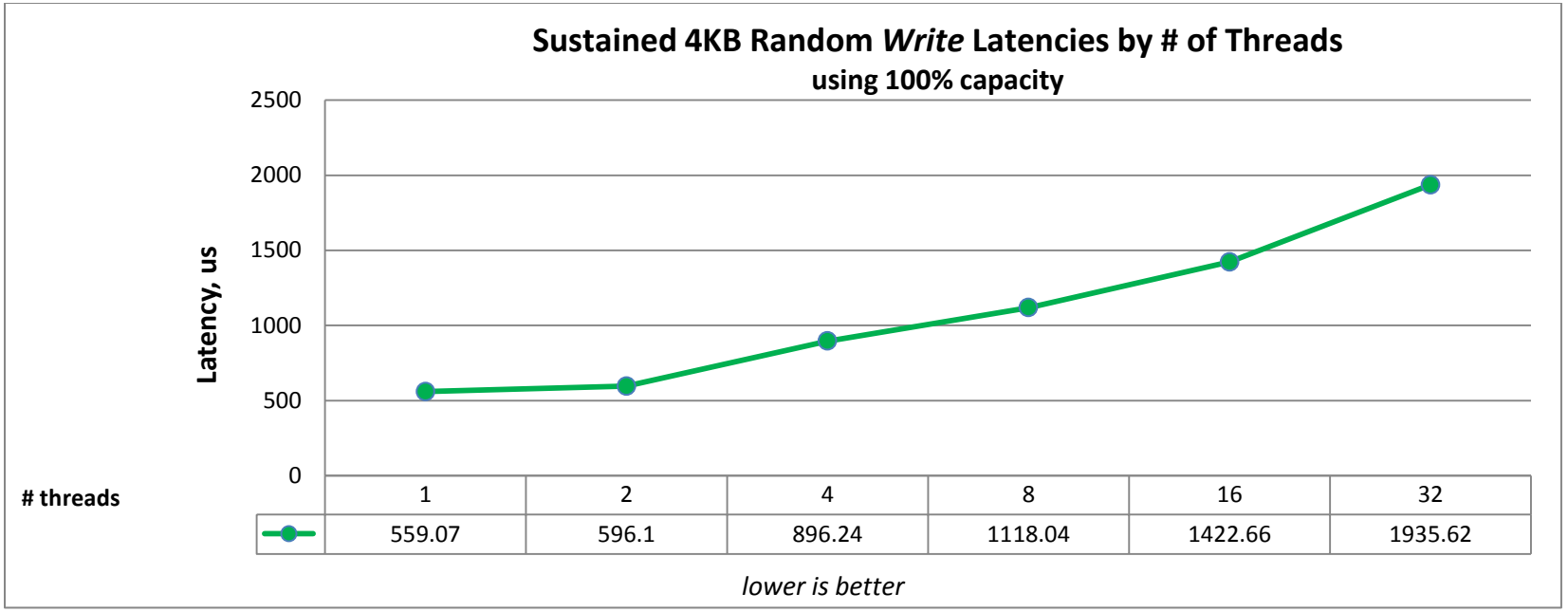
Block size 4K 情况下的 IOPS

在 64-256 达到上限 （由于网卡被打满 可能更高）



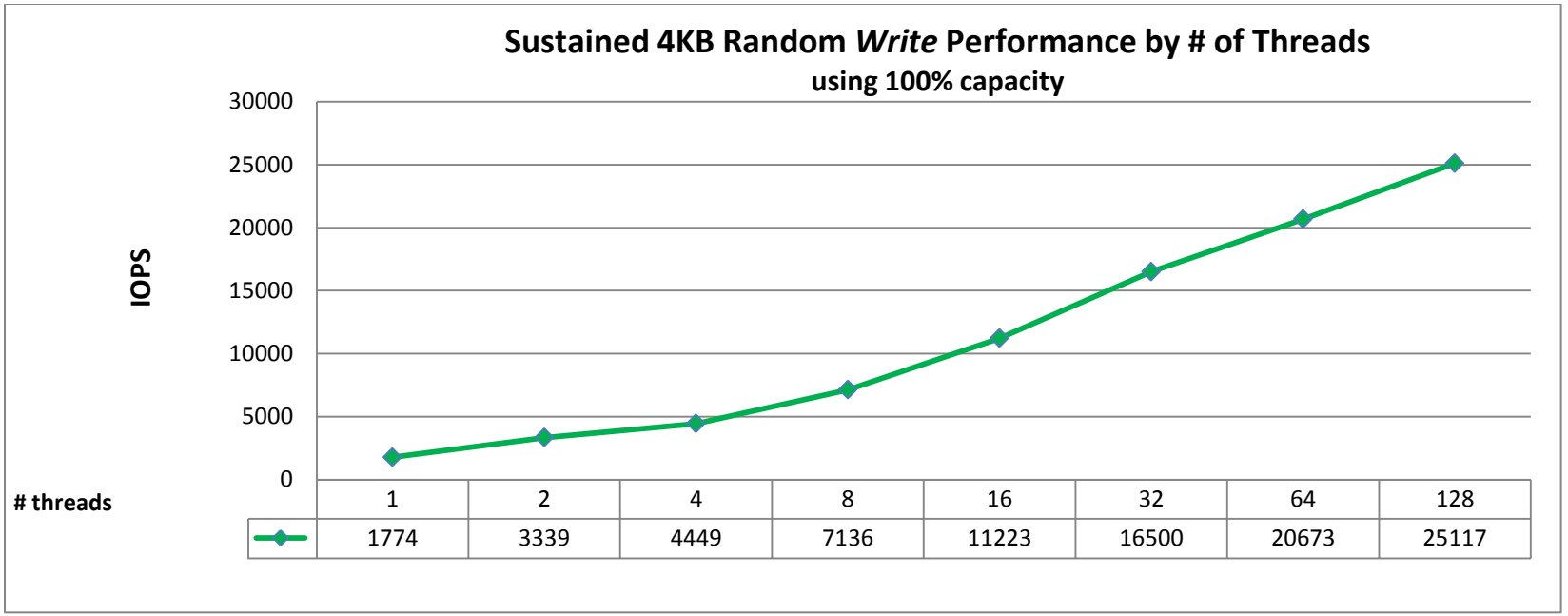
Block size 4K 的随机写延迟

可以看到随机写的延迟明显比读高了很多

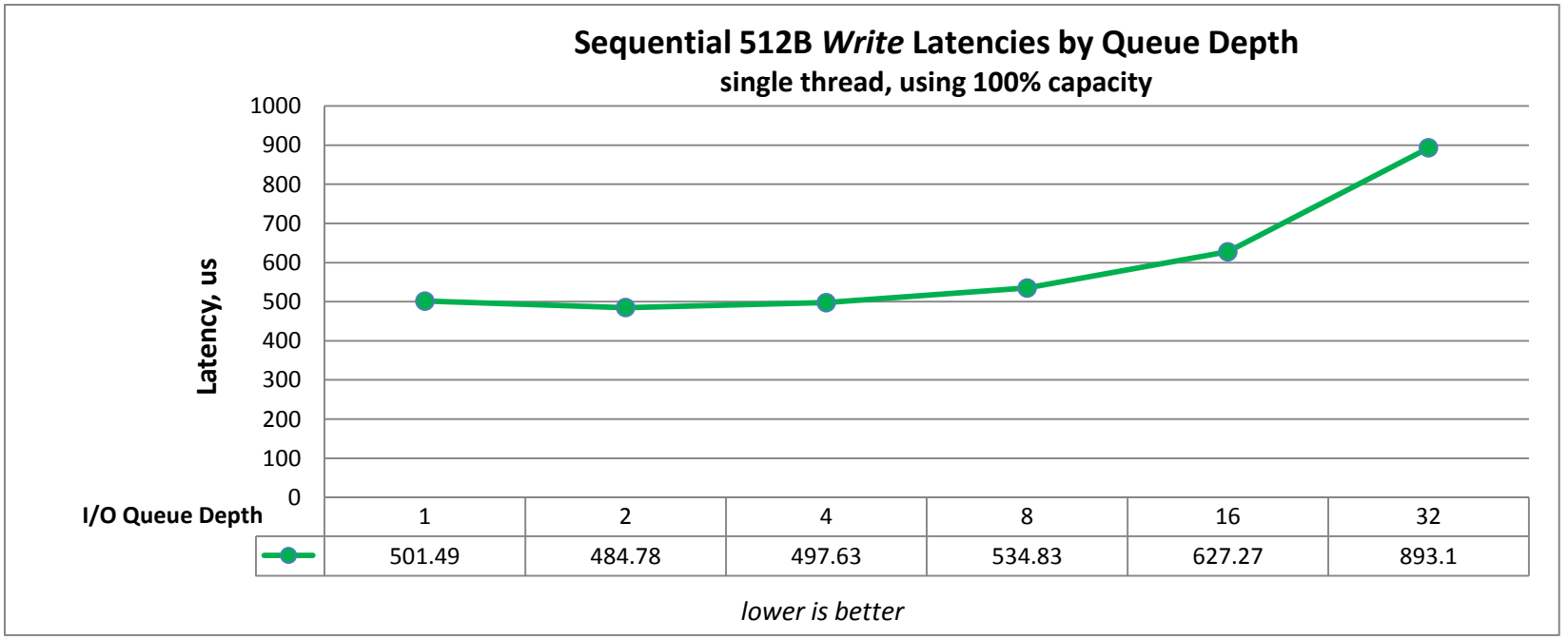


Block size 4K 随机写的 IOPS

无法将网卡打满

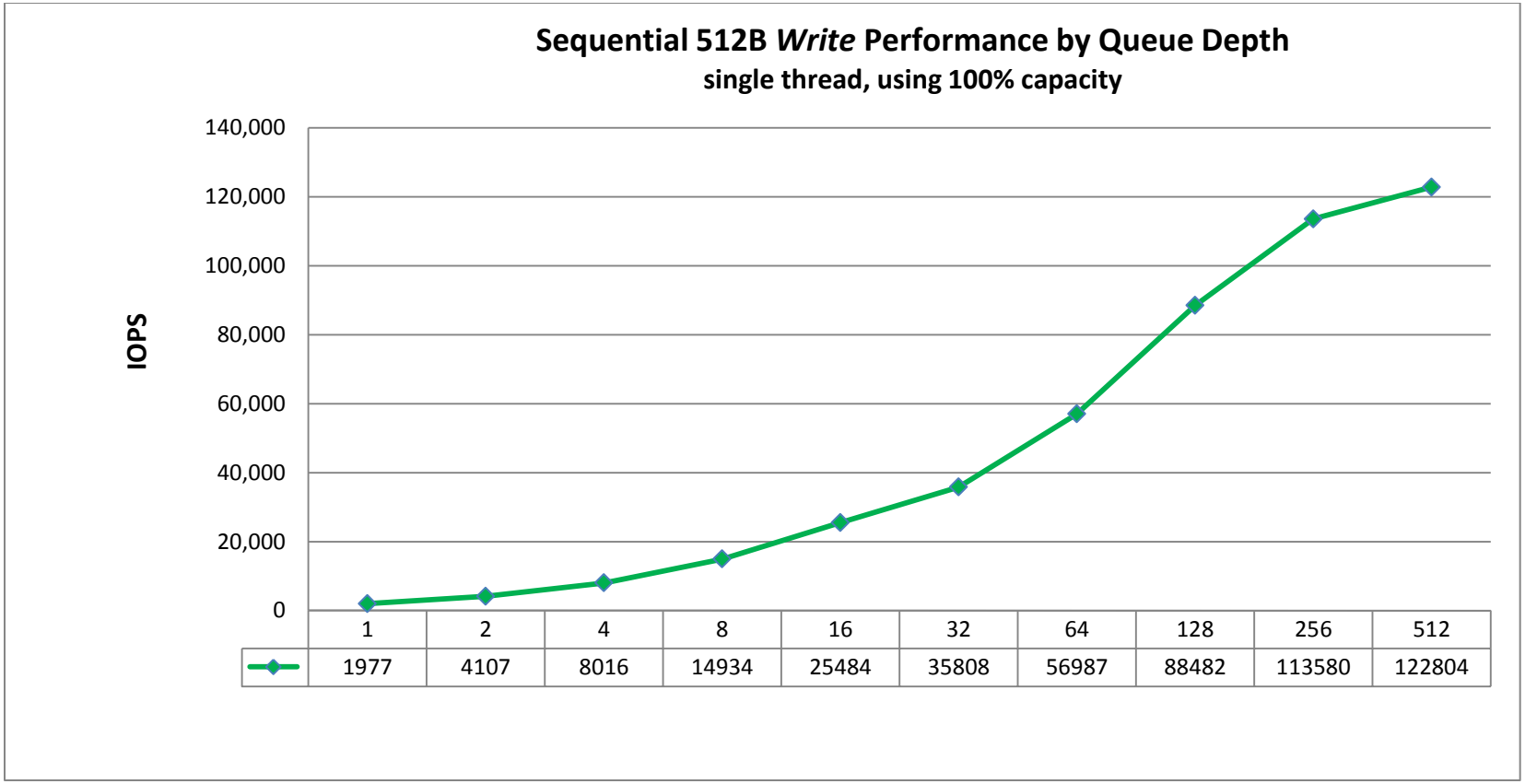


不同 IOdepth 情况下的 小块写延迟



不同 IO depth 情况下的 IOPS 表现

Depth=512 的时候 单线程的 IOPS 为 10 万



TPCC Infiniflash vs FusionIO

Infiniflash

```
****************************************
*** ###easy### TPC-C Load Generator ***
****************************************
option h with value '10.128.181.216'
option d with value 'tpcc100'
option u with value 'dbtest'
option p with value 'dbtest'
option P with value '3307'
option w with value '100'
option c with value '32'
option r with value '60'
option l with value '120'
option f with value 'tpcc_32_parallel.log'
<Parameters>
     [server]: 10.128.181.216
     [port]: 3307
     [DBname]: tpcc100
       [user]: dbtest
       [pass]: dbtest
  [warehouse]: 100
 [connection]: 32
     [rampup]: 60 (sec.)
    [measure]: 120 (sec.)

RAMP-UP TIME.(60 sec.)

<Raw Results>
  [0] sc:84197  lt:0  rt:0  fl:0
  [1] sc:84091  lt:0  rt:0  fl:0
```

FusionIO

```
****************************************
*** ###easy### TPC-C Load Generator ***
****************************************
option h with value '10.128.181.216'
option d with value 'tpcc100'
option u with value 'dbtest'
option p with value 'dbtest'
option P with value '3307'
option w with value '100'
option c with value '32'
option r with value '60'
option l with value '120'
option f with value 'tpcc_32_parallel.log'
<Parameters>
     [server]: 10.128.181.216
     [port]: 3307
     [DBname]: tpcc100
       [user]: dbtest
       [pass]: dbtest
  [warehouse]: 100
 [connection]: 32
     [rampup]: 60 (sec.)
    [measure]: 120 (sec.)

RAMP-UP TIME.(60 sec.)

<Raw Results>
  [0] sc:214993  lt:0  rt:0  fl:0
  [1] sc:214899  lt:0  rt:0  fl:0
```

```
  [2] sc:8421  lt:0  rt:0  fl:0
  [3] sc:8420  lt:0  rt:0  fl:0
  [4] sc:8421  lt:0  rt:0  fl:0
 in 120 sec.

<Raw Results2(sum ver.)>
  [0] sc:84208  lt:0  rt:0  fl:0
  [1] sc:84207  lt:0  rt:0  fl:0
  [2] sc:8421  lt:0  rt:0  fl:0
  [3] sc:8422  lt:0  rt:0  fl:0
  [4] sc:8421  lt:0  rt:0  fl:0

<Constraint Check> (all must be [OK])
 [transaction percentage]
        Payment: 43.45% (>=43.0%) [OK]
   Order-Status: 4.35% (>= 4.0%) [OK]
       Delivery: 4.35% (>= 4.0%) [OK]
    Stock-Level: 4.35% (>= 4.0%) [OK]
 [response time (at least 90% passed)]
      New-Order: 100.00%  [OK]
        Payment: 100.00%  [OK]
   Order-Status: 100.00%  [OK]
       Delivery: 100.00%  [OK]
    Stock-Level: 100.00%  [OK]

<TpmC>
                 42098.500 TpmC
```

```
  [2] sc:21499  lt:0  rt:0  fl:0
  [3] sc:21498  lt:0  rt:0  fl:0
  [4] sc:21500  lt:0  rt:0  fl:0
 in 120 sec.

<Raw Results2(sum ver.)>
  [0] sc:214995  lt:0  rt:0  fl:0
  [1] sc:214997  lt:0  rt:0  fl:0
  [2] sc:21499  lt:0  rt:0  fl:0
  [3] sc:21498  lt:0  rt:0  fl:0
  [4] sc:21500  lt:0  rt:0  fl:0

<Constraint Check> (all must be [OK])
 [transaction percentage]
        Payment: 43.47% (>=43.0%) [OK]
   Order-Status: 4.35% (>= 4.0%) [OK]
       Delivery: 4.35% (>= 4.0%) [OK]
    Stock-Level: 4.35% (>= 4.0%) [OK]
 [response time (at least 90% passed)]
      New-Order: 100.00%  [OK]
        Payment: 100.00%  [OK]
   Order-Status: 100.00%  [OK]
       Delivery: 100.00%  [OK]
    Stock-Level: 100.00%  [OK]

<TpmC>
                 107496.500 TpmC
```

基本功能测试

第一阶段测试内容涵盖软件的安装配置，基本功能验证和性能测试。测试结果中包括测试方法、测试过程、结果记录以及完成情况说明。

本次测试概括为六个方面：

1. **安装与基本配置**
2. **高可用性测试**
3. **管理功能与安全性测试**
4. **其他基本功能测试**
5. **特殊功能测试**

测试用例一：安装与基本配置测试

安装与基本配置测试

测试用例	测试用例名称	NexentaStor 安装与基本配置	
测试过程记录	测试步骤	是否完成	完成情况说明
	1. NexentaStor 安装	完成	
	2. 冷启动与热启动验证	完成	
	3. 网络端口与主机名配置	完成	
	4. 创建磁盘 Raid 组	完成	配置不同保护级别：Mirror、RaidZ1(5)、Z2(6)、Z3(7)配置

	5. 创建存储池	完成	
	6. 存储池在线扩容	完成	
	7. 在线添加新磁盘组，系统自动识别	完成	
	8. 移除正在访问的磁盘，验证是否会有 Raid 保护机制以及重建数据/时间	完成	配置 Mirror 为例
	9. 在线断电，验证存储重启后是否有数据损坏	完成	
测试结果	**考核指标**	**结果**	**说明**
	1. NexentaStor 与兼容性验证	兼容	
	2. 软件 Raid 以及存储池功能	支持	1. 支持 3 块校验盘 2. 同一存储池支持不同 Raid 级别，支持读写加速、支持数据块和文件访问
	3. 在线扩容	支持	
	4. 磁盘拔除保护以及断电保护	支持	

测试用例二：高可用性测试

高可用性测试根据 iSCSI 场景。

高可用性测试

测试用例	2.1	测试用例名称	NexentaStor 高可用性测试（iSCSI 连接后端磁盘）	
测试过程记录	测试步骤		是否完成	完成情况说明
	1. NexentaStor HA 环境搭建		完成	
	2. 手动发起 failover 操作，确认切换是否成功以及切换时间		完成	验证切换时间主要通过观察客户端拷贝文件服务停止到继续的时间和 NexentaStor 管理界面状态改变时间
	3. 断电主节点（模拟节点故障），确认切换是否成功以及切换时间		完成	
	4. 手动发起 failback 操作，确认切回是否成功以及切回时间		完成	
	5. 切换后，手工断电 secondary 节点，确认是否自动切回，以及切回时间		完成	
	6. 手工发起 failover 切换，切换		完成	

	成功后，关掉该节点		
	7. 重启主节点，验证该节点在重启后是否可以认到后端共享磁盘	完成	
测试结果	**考核指标**	**结果**	**说明**
	1. NexentaStor HA 功能（手动/自动）	支持	
	2. NexentaStor HA 切换时间	支持，客户端应用无报错	1. 手动 Failover ：18-20s 2. 自动 Failover ： 30s 左右 3. 手动 Failback ：18-20s 4. 自动 Failback ：30s 左右
	3. 在线切换节点后，该节点下电，重启	支持	可实现 HA 环境下，在线节点的更换。
测试方法	1. 验证切换时间主要通过观察客户端拷贝文件服务停止到继续的时间和 NexentaStor 管理界面状态改变的时间 2. 每项测试验证 5 次，取平均值		

3.3 测试用例三：管理功能与安全性测试

管理功能与安全性测试

测试用例	**3**	**测试用例名称**	**NexentaStor 文件系统相关测试**	
测试过程记录	**测试步骤**		**是否完成**	**完成情况说明**
	1. 通过 SSH,远程访问 NexentaStor		完成	

	2. 通过浏览器（端口 8457），远程访问 NexentaStor	完成	
	3. SSL 功能	完成	
	4. 通过非 root/admin 账户登录，来创建存储池	完成	
测试结果	**考核指标**	**结果**	**说明**
	1. NexentaStor 远程管理方式	支持 SSH、Web 多种方式	
	2. NexentaStor 用户管理	支持用户管理	

3.4 测试用例四：其他基本功能测试

其他基本功能测试

测试用例	4	**测试用例名称**	NexentaStor 其他基本功能测试
测试过程记录	**测试步骤**		**是否完成**	**完成情况说明**
	1. NexentaStor 产品 License Key 注册		完成	
	2. 网络接口配置		完成	
	3. DNS 配置		完成	
	4. 默认网关配置		完成	
	5. NMV 配置向导		完成	
	6. NTP 服务配置		完成	
	7. SMTP 服务配置		完成	

	8. checkpoint 创建	完成	
	9. 插件管理与安装配置	完成	
	10.SSH 绑定功能配置	完成	
	11.LDAP 功能配置	完成	
	12.SSL 认证	完成	
	13.TLS 认证	完成	
	14.加入 AD 域	完成	
	15.加入 CIFS 工作组	完成	
	16.iSCSI initiator 配置	完成	
	17.iSCSI target 配置	完成	
	18.NDMP 配置	完成	
	19.SNMP 配置	完成	
	20.Syslog 配置	完成	
	21.NMV 用户管理	完成	
	22.UNIX 用户、用户组管理	完成	
	23.创建存储池	完成	
	24.创建目录	完成	
	25.创建 Zvol	完成	
	26.存储池导入导出	完成	
	27.目录查询功能	完成	
	28.快照创建、删除	完成	
	29.CIFS 共享	完成	
	30.NFS 共享	完成	

	31.webDAV 功能	完成	
	32.FTP 功能	完成	
	33.Rsync 功能配置	完成	
	34.创建 iSCSI 目标、映射	完成	
	35.iSCSI 目标组管理	完成	
	36.autoSnap 功能配置	完成	
	37.autotier 功能配置	完成	
	38.autosync 功能配置	完成	
	39.autoscrub 功能配置	完成	
	40.管理界面实现重启功能	完成	
	41.管理见面实现关机功能	完成	
测试结果	**考核指标**	**结果**	**说明**
	NexentaStor 软件基础功能验证	支持	

3.5 测试用例五：特性功能测试特性功能测试

特殊功能测试

测试用例	5	**测试用例名称**	NexentaStor 高级功能测试
测试过程记	**测试步骤**	**是否完成**	**完成情况说明**

录	1. 数据复制功能（异步）	完成	Autosyn 可基于文件系统、LUN
	2. 快照功能	完成	
	3. 在线扩容	完成	
	4. 数据重建	完成	3.2TB 实际数据写入，RaidZ1(5)
测试结果	考核指标	结果	说明
	Nexenta 特性功能验证		