TOSHIBA

Informe de Laboratorio

JBOD PROMISE VTrak J5960

Rainer Kaese, director senior Desarrollo
de Negocio de Productos Almacenamiento
de Toshiba Electronics Europe GmbH

Introducción

La lucha por la neutralidad climática y la sostenibilidad es clave en los centros de datos modernos. El objetivo: mantener bajos el consumo de energía y la huella de carbono total, al mismo tiempo que se procesa un volumen enorme de datos 24/7, datos que se han almacenado y se almacenarán principalmente en unidades de disco duro.

Las iniciativas ecológicas afectan a todas las áreas del centro de datos, desde medidas técnicas como la reutilización del calor residual o la refrigeración con frío, hasta el uso de un inventario producido de forma sostenible y de tecnologías de disco duro (HDD) de última generación.

Aquí es donde empiezan nuestros informes de laboratorio:

El nuevo JBOD de carga superior VTrak J5960 4U 60-Bahías de PROMISE se promociona como un “JBOD con ADN verde”, incluyendo un compromiso con la protección del medio ambiente y la producción sostenible. Toshiba Electronics Europe GmbH (“Toshiba”) ha probado y evaluado el JBOD con 60 HDD Toshiba Enterprise con una capacidad de 18TB, lo que suma una capacidad total de 1080TB.

Toshiba ha probado el funcionamiento, el rendimiento, la acústica y el consumo de energía, con el foco puesto en la evaluación de los atributos ecológicos que destaca PROMISE.

Imagen 1: El JBOD VTrak J5960 de PROMISE en el laboratorio Toshiba.

TOSHIBA

Dimensiones & características mecánicas:

Con una altura de 4U y una longitud de chasis de solo 666mm, el J5960 es el JBOD de alta densidad más corto que jamás hemos visto en el laboratorio de HDD de Toshiba. Es tan largo como el chasis de un servidor típico de 2U de 66cm y se adapta convenientemente a cualquier rack existente. Esto es una gran ventaja sobre otros JBOD: muchos superan los 1000mm de longitud y requieren racks más largos o tienden a crear problemas de cableado.

Los módulos IO (IOM) intercambiables en caliente del J5960 se extraen por la parte delantera del JBOD – el cableado conectado permanece en la parte posterior de la unidad. Esto facilita el reemplazo de un IOM, mientras que los intercambios traseros exigen acceder a través de un montón de cables de alimentación y señal.

Una característica mecánica interesante es la tapa del JBOD. Se mantiene unida y permanece en el rack cuando se extrae el JBOD para algún servicio. Lo hemos probado y funciona realmente bien. Al no ser necesario levantar las tapas, solo hay que extraer el JBOD hasta donde lo requiera la unidad defectuosa.
También puede instalarse en zonas altas del rack.

Los HDDs se fijan en bandejas metálicas con 4 tornillos.

Los LEDs de estado están normalmente apagados y solo se activan cuando se retira la tapa (es decir, cuando el JBOD se extrae del rack). Esto ahorra Watios de energía adicional.

Configuración en el Lab de Toshiba

Modelo:	PROMISE VTrak J5960 4U-SAS-60-D BP
Firmware:	1023
SO Host:	Linux (Centos 7.9)
SO Host:	Windows (Windows Server 2019 Standard)
Adaptador de bus HOST (HBA):	Broadcom Avago HBA 9500-16e (Host IF: 8x PCIe-Gen4)
Adaptador RAID:	Microchip Adaptec® SmartRAID Ultra 3254-16e/e (16x PCIe-Gen4)

Pruebas con unidades SAS de capacidad empresarial (nearline):

Nombre modelo:	ToshibaMG09SCA18TE
Tamaño de bloque	512B emulado
Firmware	0104

Imagen 2: el J5960 con la tapa abierta.

Diámetro exterior ratio datos: 282MB/s

Consumo de energía

Inactivo_B:	3.36W
Escritura secuencial:	
	7.62
W Lectura secuencial:	
	8.71
W	
Escritura aleatoria:	6.64W
Lectura aleatoria:	9.47W

Funciones básicas del JBOD:

Función básica:	ok
IOM SAS detectado:	ok
Conexión en caliente /reinserción:	ok
Lectura inteligente:	ok
Gestión de la caja:	ok, probada con un conector serial RJ11

Imagen 3: Toshiba HDD MG09SCA18TE en bandeja PROMISE.

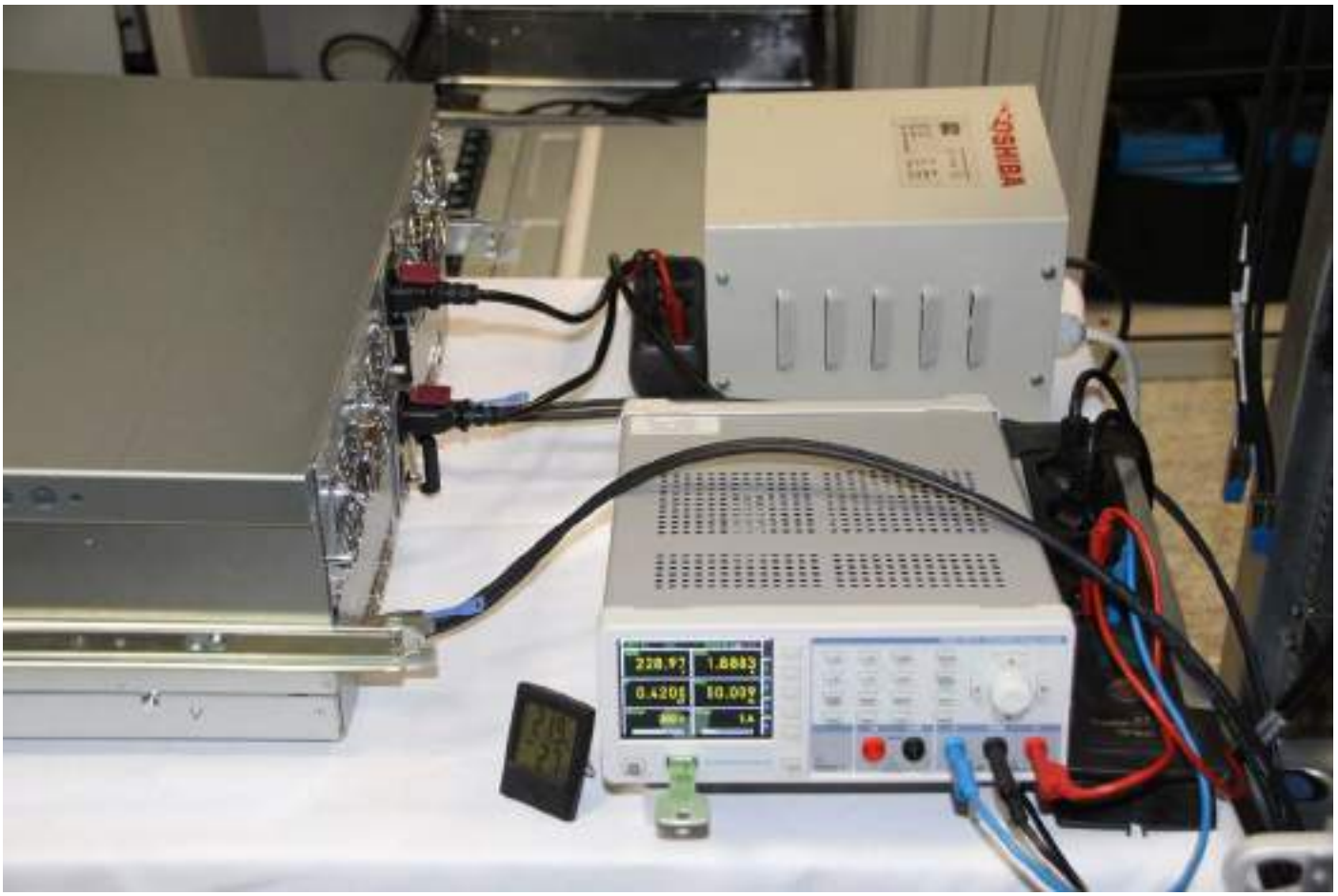

Imagen 4: Configuración de medición de potencia en el Lab de Toshiba.

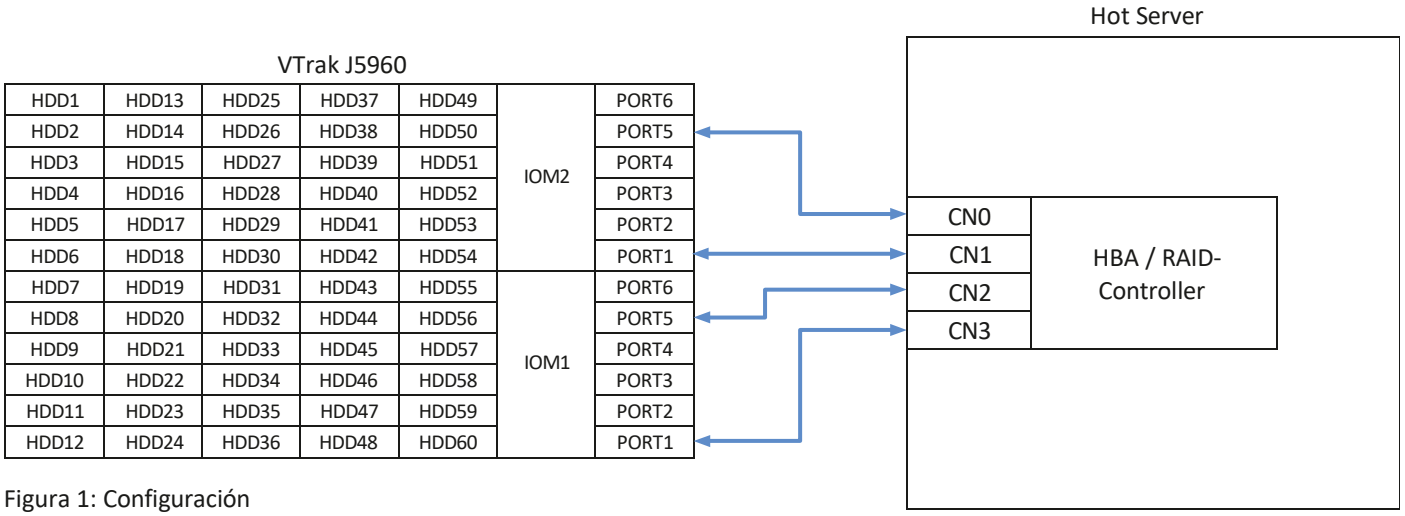

Figura 1: Configuración

Para la medición precisa del consumo de energía hemos utilizado un analizador profesional de potencia de alta precisión (R&S HMC8015).

JBOD encendido, sin discos, SAS conexión al host activa: 100W
JBOD con discos, potencia máxima arranque más de 500ms. 850W
JBOD con discos raw en HBA Inactivo_B: 305W
Lambda (ratio de energía activa y reactiva) 0.96
Ruido a 1 m distancia 80dB
Temperatura ambiente.....23°C

El VTrak J5960 tiene una configuración predeterminada que pone todos los discos (SAS) en estado Inactivo cuando no se accede a ellos durante aproximadamente 2 minutos.

Unos 100W for un JBOD con IOM dual sin discos es un valor de potencia excelente. En JBOD con IOM individuales dedicados puede ser tan bajo como 80W en iguales condiciones, pero el rango típico en JBODs con IOM dual es 200-300W.
300W en "inactivo con discos" también es un sorprendente valor bajo, otros JBOD de 60 bahías parten de 400W o más.
Un factor lambda alto de 0.96 significa que la ratio de energía reactiva generada por las fuentes de alimentación es muy baja (equivale al 4% de la energía total). Cuanto mayor (mejor) es el factor lambda, menor será la energía reactiva.
La energía reactiva no cuesta y no genera calor, pero los rieles de suministro de energía deben dimensionarse para la energía activa y reactiva – en los centros de datos, un lambda alto es un factor importante.

Mediciones de rendimiento en el Lab de Toshiba

Para las pruebas con discos SAS, hemos conectado ambos IOM del JBOD con dos cables mini-SAS-HD a los 4 puertos mini-SAS-HD del HBA 6e y al controlador RAID.

Esta configuración proporciona un ancho de banda de acceso al JBOD/HDD teórico de 4x 4.8GB/s = 19.2GB/s, pero requiere agregar rutas con la funcionalidad de instalación multiruta.
En cuanto a la configuración con HBA, la multiruta debe habilitarse manualmente en Linux/Windows. Un controlador RAID (como el Microchip Adaptec® modelo Ultra-3254) detectará las configuraciones automáticamente e invocará una configuración multiruta correcta. La multiruta y la agregación de enlaces SAS manual solo funciona en discos SAS. La configuración con SATA se beneficiará de la agregación IOPS de los 60 HDDs, pero el ancho de banda secuencial suele limitarse a un enlace mini-SAS-HD (4.8GB/s).

Hemos probado varias configuraciones de disco con "fio" – un software de pruebas I/O flexible) – midiendo el rendimiento secuencial, aleatorio y con cargas de trabajo mixtas, así como el consumo de energía relacionado.
Se han hecho pruebas en discos individuales conectados vía HBA y en configuraciones RAID como unidades físicas y lógicas. En las unidades lógicas también medimos el rendimiento y consumo de una copia (es decir, lectura y escritura) de un archivo grande.

Comentarios sobre la configuración del JBOD

La configuración predeterminada del VTrak J5960 es "HDD inactivo tras 2 min de inactividad". Se necesitan unos 1200ms para que el HDD inactivo pase de estado Inactivo a Activo.
En RAID se recomienda deshabilitar esta función – en un RAID grande es posible que algunos HDDs no estén activos por más de 2 mins, incluso con carga de datos completa. Si algunos HDDs del RAID se configurarán en modo inactivo se producirían largas latencias de 1200ms si se accediera a ellos.
Desabilitar el cambio a modo inactivo: vía cable serial al puerto de gestión de IOM (115200/8/N/1), el comando CLI es "enclosure –m –idlep 0".

Las pruebas de Toshiba se realizaron en “modo zoning 0”, (=configuración predeterminada). Esto significa “sin zoning”, por lo que se puede acceder a todos los HDD desde ambos IOM.

Algunos controladores RAID o HBAs pueden rechazar la conexión de 4 cables en dos IOM. Si es así, el “modo zoning 1” puede ser una alternativa. Aquí 30 unidades están conectadas desde un puerto SAS a IOM1 y otras 30 están conectadas desde un puerto SAS a IOM2. La configuración aparece como dos JBOD de 30 discos.
El comando CLI para cambiar el zoning mode es “enclosure –m –z 1”.

Todos los discos en paralelo con dispositivos físicos individuales (multiruta):

SO:	Linux (Centos 7.9)
HBA/Controlador:	Broadcom HBA9500-16e
HDD:	60x Toshiba MG09SCA18TE
Configuración:	Dual IOM 2x 2 cables Mini-SAS HD (3m longitud) Configuración multiruta en discos

Carga de trabajo	Energía (W)	IOPS	Ancho de banda (MB/s)
Escritura secuencial 1024K	610		13300
Lectura secuencial 1024K	640		14500
Escritura aleatoria 4K	510	24100	
Lectura aleatoria 4K	540	33900	
Mix 4K/64K/256K/2M	540	22600	2350
Temperatura ambiente	23°C		
Temperatura min HDD	27°C		
Temperatura max HDD	36°C		

El ancho de banda máximo teórico es 282MB/s (disco único) x 60 = 16.2GB/s. Con 14.5GB/s y un IOPS superior a 20K, esta configuración está cerca del límite teórico.

Las cifras de potencia máxima de 640W prueban las credenciales verdes del JBOD. La diferencia de temperatura de menos de 10°C entre la unidad más caliente y la más fría con una temperatura máxima de menos de 14°C sobre la temperatura ambiente, respaldan la fiabilidad y larga vida útil de los HDDs.

Todos los discos en RAID10, unidad física de Windows:

SO:	Windows Server 2019
RAID/Adaptador:	Microchip Adaptec® SmartRAID Ultra 3254-16e/e (16x PCIe-Gen4)
HDD:	60x Toshiba MG09SCA18TE
Configuración:	Dual IOM 2x 2 cables Mini-SAS HD (3m longitud)

Carga de trabajo	Energía (W)	IOPS	Ancho de banda (MB/s)
Escritura secuencial 1024K	510		8600
Lectura secuencial 1024K	570		15300
Escritura aleatoria 4K	600	12800	
Lectura aleatoria 4K	730	9900	
Mix 4K/64K/256K/2M	680	6100	1800
Inactivo (fondo raid)	470		
Temperatura ambiente	25°C		
Temperatura min HDD	29°C		
Temperatura max HDD	38°C		

Script 1 – Todos los discos en paralelo como dispositivos físicos individuales (multiruta):

```
fio --direct=1 --bs=1m –-iodepth=16 --size=32g --ioengine=libaio --group_reporting --rw=write --output=se-
qwrite.log --name=/dev/mapper/mpath{a..z} -- name=/dev/mapper/mpatha{a..z} –-name=/dev/mapper/mpathb{a..h}

fio --direct=1 --bs=1m –-iodepth=16 --size=32g --ioengine=libaio --group_reporting --rw=read --output=se-
qread.log --name=/dev/mapper/mpath{a..z} -- name=/dev/mapper/mpatha{a..z} –-name=/dev/mapper/mpathb{a..h}

fio --direct=1 --bs=4k –-iodepth=16 --size=512m --ioengine=libaio --group_reporting --rw=randwrite --out-
put=randwrite.log --name=/dev/mapper/mpath{a..z} -- name=/dev/mapper/mpatha{a..z} –-name=/dev/mapper/mpath-
b{a..h}

fio --direct=1 --bs=4k –-iodepth=16 --size=512m --ioengine=libaio --group_reporting --rw=randread --out-
put=randread.log --name=/dev/mapper/mpath{a..z} -- name=/dev/mapper/mpatha{a..z} –-name=/dev/mapper/mpath-
b{a..h}

fio --direct=1 --bssplit=4k/20:64k/50:256k/20:2M/10 –-iodepth=16 --size=8g –ioengine=libaio --group_report-
ing --rw=randrw --output=mixed.log --name=/dev/mapper/mpath{a..z} -- name=/dev/mapper/mpatha{a..z} –-name=/
dev/mapper/mpathb{a..h}
```

TOSHIBA

Script 2 – Todos los discos en RAID10, unidad física de Windows:

```
fio --filename=\\.\Physicaldrive1 --direct=1 --rw=write --bs=1m --iodepth=16 --time_based --runtime=300
--group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=64 --norandommap --randrepeat=0
--output=seqwritephysical.log

fio --filename=\\.\Physicaldrive1 --direct=1 --rw=read --bs=1m --iodepth=16 --time_based --runtime=300
--group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=64 --norandommap --randrepeat=0
--output=seqreadphysical.log

fio --filename=\\.\Physicaldrive1 --direct=1 --rw=randwrite --bs=4k --iodepth=16 --time_based --runtime=300
--group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=64 --norandommap --randrepeat=0
--output=randwritephysical.log

fio --filename=\\.\Physicaldrive1 --direct=1 --rw=randread --bs=4k --iodepth=16 --time_based --runtime=300
--group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=64 --norandommap --randrepeat=0
--output=randreadphysical.log

fio --filename=\\.\Physicaldrive1 --direct=1 --rw=randrw --bssplit=4k/20:64k/50:256k/20:2M/10 --iodepth=16
--time_based --runtime=300 --group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=64
--norandommap --randrepeat=0 --output=mixedphysical.log
```

Como la escritura en RAID10 siempre va a dos dispositivos espejo en paralelo, la velocidad de escritura se reduce a la mitad del valor. La energía inactiva en configuraciones RAID está casi al mismo nivel de la energía activa ya que el controlador RAID siempre accede a los discos para realizar comprobaciones de consistencia en segundo plano.

Todos los discos en RAID10, volumen lógico Windows:

SO:	Windows Server 2019
RAID/Adaptador:	Microchip Adaptec® SmartRAID Ultra 3254-16e/e (16x PCIe-Gen4)
HDD:	60x Toshiba MG09SCA18TE
Configuración:	Dual IOM 2x 2 cables Mini-SAS HD (3m de longitud)

Carga de trabajo	Energía (W)	IOPS	Ancho de banda (MB/s)
Escritura secuencial 1024K	520		6900
Lectura secuencial 1024K	550		15000
Escritura aleatoria 4K	520	11100	
Lectura aleatoria 4K	540	29500	
Mix 4K/64K/256K/1M	550	8100	2400
Copia de Windows	500		550
Inactivo (fondo raid.)	470		
Temperatura ambiente	25°C		
Temperatura min HDD	29°C		
Temperatura max HDD	39°C		

Script 3 – Todas las unidades en RAID10, volumen lógico de Windows:

```
fio --filename=test --size=1T --direct=1 --rw=write --bs=1m --iodepth=16 --time_based --runtime=300
--group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=64 --norandommap --randrepeat=0
--output=seqwritelogical.log

fio --filename=test --size=1T --direct=1 --rw=read --bs=1m --iodepth=16 --time_based --runtime=300 --group_
reporting --name=job1 --ioengine=windowsaio --thread --numjobs=64 --norandommap --randrepeat=0 --out-
put=seqreadlogical.log

fio --filename=test --size=1T --direct=1 --rw=randwrite --bs=4k --iodepth=16 --time_based --runtime=300
--group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=64 --norandommap --randrepeat=0
--output=randwritelogical.log

fio --filename=test --size=1T --direct=1 --rw=randread --bs=4k --iodepth=16 --time_based --runtime=300
--group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=64 --norandommap --randrepeat=0
--output=randreadlogical.log

fio --filename=test --size=1T --direct=1 --rw=randrw --bssplit=4k/20:64k/50:256k/20:2M/10 --iodepth=16
--time_based --runtime=300 --group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=64
--norandommap --randrepeat=0 --output=mixedlogical.log
```

Este benchmark para un volumen lógico en Windows no usa el tamaño completo de la unidad, sino las operaciones en un archivo de prueba de 1TB. Es, por tanto, más realista para los casos de uso reales. El IOPS de más de 32k se debe al hecho de que la operación de búsqueda no cubre el rango de capacidad completo. Eso también resulta en un consumo de energía significativamente menor en las operaciones de escritura en comparación con las búsquedas completas de 500TB en las unidades físicas anteriores.

Configuración con unidades SATA

También probamos el JBOD VTrak J5960 JBOD con unidades SATA (Modelo MG09ACA18TE). Como la interfaz solo tiene una ruta de señal, usamos una configuración de IOM única (se desconectó IOM2 y los cuatro cables se conectaron a IOM1).

El rendimiento secuencial suele limitarse al nivel de un cable mini-SAS HD (unos 4.3GB/s lectura y escritura secuencial). Los valores IOPS encajan bien con los resultados con SAS al no tener límite de ancho de banda (10k IOPS en capacidad total, 30k en unidades lógicas con rango datos 1TB).

El consumo de energía con unidades SATA y una operación con un único IOM es en torno a 70~80W inferior a la configuración equivalente con unidades SAS e IOM dual. Esto se debe a que la propia unidad SATA tiene un consumo de energía entre 0.4~0.8W inferior (dependiendo de la carga) que la misma unidad con una interfaz SAS y el segundo IOM faltante.

Consideraciones del sistema

Un ancho de banda de red de 100Gbit/s y un ancho de banda de almacenamiento de 12.5GB/s combinan bastante bien. Para sistemas que requieren tan alto rendimiento secuencial recomendamos utilizar los HDD de capacidad empresarial nearline SAS de la serie MG de Toshiba en configuraciones de IOM dual.

Si el ancho de banda de la red es de 25Gbit/s o menos y el principal objetivo es la capacidad más alta, también pueden usarse unidades nearline SATA en configuraciones de un solo IOM ya que habitualmente el ancho de banda de almacenamiento se limita a 4GB/s, lo que nuevamente coincide con una velocidad de enlace de red de 25Gbit/s.

Conclusión

El VTrak J5960 de PROMISE Vtrak es un JBOD de carga superior de 60 bahías energéticamente eficiente y fácil de mantener. Es el modelo más compacto de su clase con una longitud de solo 666mm. Completamente equipado con discos Toshiba de 18TB ofrece una capacidad total de más de un PB con solo alrededor de 500W de consumo de energía.

En la evaluación de las configuraciones de almacenamiento basadas en este JBOD, Toshiba ha mostrado con 60 HDDs rendimientos agregados de hasta 15GB/s de rendimiento secuencial y más de 30k de IOPS aleatorio, al tiempo que la gestión eficiente del flujo de aire y la refrigeración del JBOD apoyan la larga vida útil y alta fiabilidad de los discos duros al mantener su temperatura en pleno funcionamiento siempre menos de 14°C sobre la temperatura ambiente

Nota de agradecimiento a nuestros partners

La colaboración ha sido clave en el éxito de este informe de laboratorio. “Agradezco a todos nuestros socios el apoyo al proyecto. PROMISE nos proporcionó el JBOD ecológico Vtrak J5960, Microchip lo apoyó con controlador raid Adaptec SmartRAID Ultra 3254-16e /e y Broadcom contribuyó con el Adaptor de Bus Host HBA 9500-16e. Junto con nuestras unidades de disco duro Toshiba, pudimos construir en nuestro laboratorio una configuración de centro de datos, mostrado un nuevo e impresionante nivel de resultados de rendimiento”.

Rainer Kaese, director senior Desarrollo de Negocio, División Productos Almacenamiento, Toshiba Electronics Europe GmbH

TOSHIBA

Toshiba Electronics Europe GmbH

Hansaallee 181
40549
Düsseldorf
Alemania

info@toshiba-storage.com
toshiba-storage.com

Copyright © 2022 Toshiba Electronics Europe GmbH. Todos los derechos reservados. Especificaciones de producto, configuraciones, precios y componentes / opciones de disponibilidad están sujetos a cambios sin aviso. El diseño de los productos, las especificaciones y colores están sujetos a cambio sin aviso y pueden diferir de los que se muestran. Errores y omisiones excluidos.