

Seeweb S.r.l.
Corso Lazio 9a
Frosinone
Tel. 800 112825
www.seeweb.com



SAS, SSD E XIV A CONFRONTO

GIUGNO 2014

*Benchmark comparativi dei sistemi di storage più diffusi nell'infrastruttura Cloud.
Test di scrittura, lettura e latenza effettuati in ambito enterprise su lame IBM
BladeCenter HS23*

SOMMARIO

Sommario

Scenario	1
Il test	5
Risultati	8
Conclusioni	15
Riferimenti:	17
Informazioni di contatto	18
Informazioni sulla società	18

Scenario

AMBIENTE DI TEST

I test sono stati effettuati su tre differenti server dello stesso IBM Blade Center HS23, dotati ciascuno di 2 processori da 6 core e 48 GB di RAM. Su ogni server è stato installato o configurato uno dei sistemi di storage da testare.

Le CPU sono Intel(R) Xeon(R) CPU E5-2620 da 2.00GHz, con un valore di cache size di 15360 KB.

Sui server è stata installata la medesima versione del sistema operativo Linux CentOS 6.5 (Final) con kernel 2.6.32-431.17.1.el6.x86_64; sul sistema sono stati installati esclusivamente i software utilizzati per i benchmark in modo che i risultati non fossero influenzati da altri fattori.



IBM BladeCenter HS23 con dischi SAS

GLI STORAGE TESTATI

Di seguito i tre sistemi di storage testati. Tutti i dischi erano nuovi e sono stati utilizzati per la prima volta per l'esecuzione delle nostre prove.

SCENARIO

DISCHI SAS

I dischi SAS testati sono IBM / SEAGATE 900GB 10K RPM SAS 2.5 SFF SYSTEM X HARD DRIVE HDD ST9900805SS-IBM.

Driver: Serial Attached SCSI controller: LSI Logic / Symbios Logic SAS2004 PCI-Express Fusion-MPT SAS-2 [Spitfire] (rev 03).

I dischi sono stati montati in RAID 1.



DISCHI SSD

I dischi SSD su cui sono stati effettuati i test sono dei Samsung SSD 840 BB6Q.

Driver: Serial Attached SCSI controller: LSI Logic / Symbios Logic SAS2004 PCI-Express Fusion-MPT SAS-2 [Spitfire] (rev 03).

I dischi sono stati montati in RAID 1.



SCENARIO

STORAGE XIV

Infine ulteriori benchmark sono stati eseguiti sul sistema di Storage Area Network IBM XIV Storage Gen3.

Driver: Fibre Channel: Emulex Corporation OneConnect 10Gb FCoE Initiator (be3) (rev 03).



SCENARIO

I TOOL UTILIZZATI

I software utilizzati per i test sono stati installati dai repository del sistema operativo o compilati direttamente sul server seguendo le specifiche degli sviluppatori.

HDPARM

Hdparm fornisce una interfaccia a riga di comando per i vari sistemi di storage ioctls supportati dai driver Linux ATA / IDE. In particolare per questi test, questo tool, è stato ricompilato per supportare l'ultima release del kernel.

DD

Dd copia un file, formattandolo e convertendolo a seconda degli operandi passati in input.

IOPING

Questo strumento consente di monitorare la latenza di I/O in tempo reale.

DBENCH

Questo tool è una emulazione della funzione di caricamento file system del benchmark NetBench. Questo benchmark è una buona simulazione di una configurazione server vera e propria in cui una grande quantità di file con dimensioni e posizioni diverse devono essere creati, scritti, letti, ed eliminati. Permette di eseguire dei test con processi in concorrenza.

Il test

Lo scopo dei test era quello di verificare le reali prestazioni in scrittura e lettura, sia per singoli processi che con processi paralleli, dei tre sistemi di storage esaminati.

I test sono stati eseguiti molteplici volte, e i risultati ottenuti dalle diverse esecuzioni sono sempre stati sufficientemente simili da renderli attendibili.

I benchmark sono stati eseguiti utilizzando le medesime impostazioni e opzioni per ciascun programma utilizzato.

Di seguito, i test eseguiti nel dettaglio. In tutti gli script, con la variabile **\$DEV** indichiamo il device.

HDPARM

Come primo test, è stato eseguito il software `hdparm`, e eseguito con le seguenti opzioni:

```
hdparm -tT $DEV
```

Dove le opzioni riportate indicano rispettivamente:

- T

Esegue un test per avere i tempi di cache ai fini di un confronto di benchmark. Questa opzione visualizza la velocità di lettura direttamente dalla cache buffer Linux senza accesso al disco. Misura essenzialmente un'indicazione della velocità del processore, cache, e la memoria del sistema in esame.

- t

Esegue un test di lettura di un dispositivo ai fini di un confronto di benchmark. Visualizza la velocità di lettura attraverso la cache del disco senza alcun caching preventivo dei dati. Dà un'indicazione di quanto velocemente il disco può sostenere letture di dati sequenziali sotto Linux, senza alcun sovraccarico filesystem.

IL TEST

DD

Il secondo test è stato eseguito usando il comando dd. Il test è composto da 2 fasi, la prima scrive un file di 100GB, mentre la seconda lo legge.

```
dd if=/dev/zero of=tempfile bs=1M count=102400 conv=fdatasync
```

L'opzione fdatasync passata a dd serve a richiedere un "sync" completo una volta, proprio prima di uscire. Così impegna l'interamente i 100 GB di dati creati (count=102400), indicando al sistema operativo di assicurarsi di copiarlo completamente su disco. Solo allora misura il tempo totale impiegato per fare tutto questo e calcola il risultato benchmark.

Dopo la scrittura viene eseguito il seguente comando per svuotare la cache in modo da non alterare eventuali test successivi.

```
echo 3 > /proc/sys/vm/drop_caches
```

La seconda fase del test legge il file precedentemente creato.

```
dd if=tempfile of=/dev/null bs=1M
```

Tutti test sono stati eseguiti senza nessun dato nella cache del disco.

IOPING

Il terzo benchmark è stato eseguito utilizzando il tool ioping, un progetto sviluppato da Google per misurare la latenza sul disco:

```
ioping -R $DEV  
ioping -RL $DEV
```

Con tali opzioni, il test ha come risultato la latenza minima, massima, media della ricerca all'interno del disco. L'opzione -L fa un test con operazioni sequenziali piuttosto che random.

DBENCH

L'ultimo test, è stato eseguito con dbench, sia in maniera sincrona che asincrona, con differenti numeri di processi concorrenti:

```
##### dbench #####
dbench -D $DIR -t 120 1
dbench -D $DIR -t 120 2
dbench -D $DIR -t 120 4
dbench -D $DIR -t 120 8
dbench -D $DIR -t 120 16
dbench -D $DIR -t 120 32
dbench -D $DIR -t 120 64
dbench -D $DIR -t 120 128

##### dbench sync #####
dbench -D $DIR -s -t 120 1
dbench -D $DIR -s -t 120 2
dbench -D $DIR -s -t 120 4
dbench -D $DIR -s -t 120 8
dbench -D $DIR -s -t 120 16
dbench -D $DIR -s -t 120 32
dbench -D $DIR -s -t 120 64
dbench -D $DIR -s -t 120 128
```

Il carico di lavoro per ogni processo dbench è specificato da un file di configurazione client.txt nella directory (testing) di lavoro. Si compone di una miscela di operazioni di file system eseguiti da ogni processo dbench. Esegue dei processi paralleli e fornisce un solo valore di conseguenza. Il valore risultante è un erogato medio delle operazioni di file system descritte in client.txt e misurata in megabyte al secondo.

Con una grande quantità di memoria come nel nostro caso, le misurazioni dbench possono essere utilizzati per rilevare gli effetti di scalabilità della memoria.

Per il carico di lavoro dbench utilizzato per le nostre misurazioni il numero di clienti è stato scalato come mostrato nella seguente sequenza: 1, 4, 8, 16, 32, 64, 128, sempre con una durata di 120 secondi (-t 120).

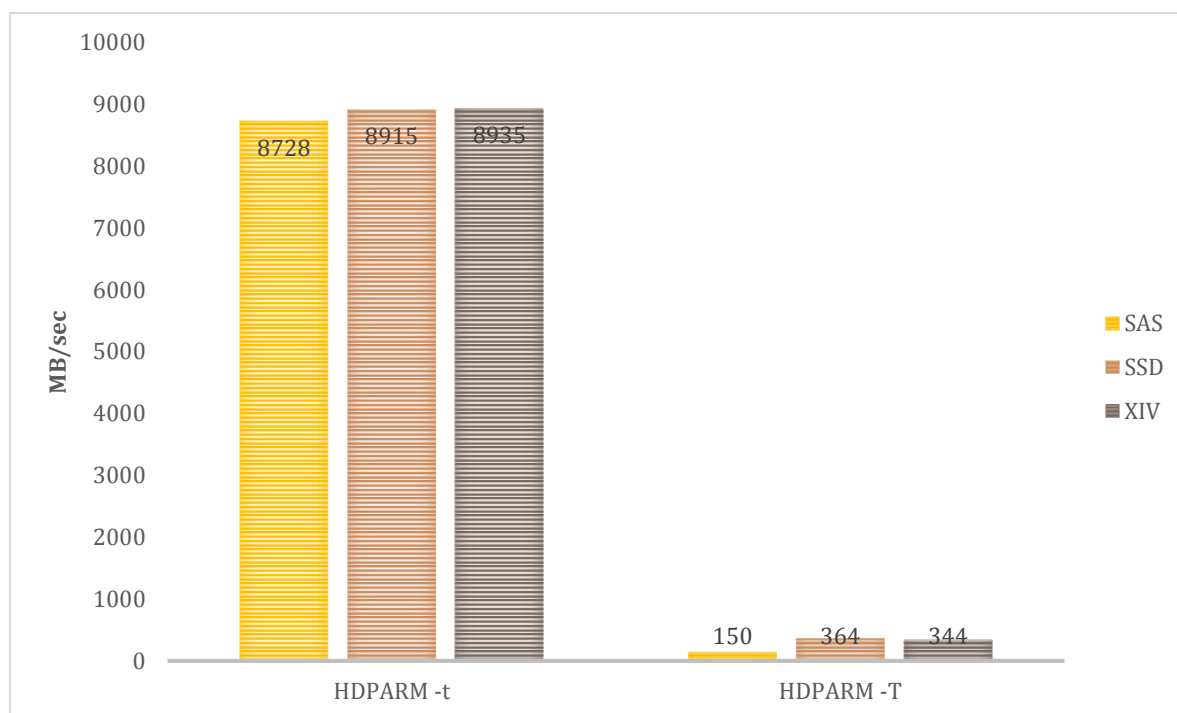
RISULTATI

Risultati

Di seguito l'esito dei risultati di ciascun test:

HDPARM

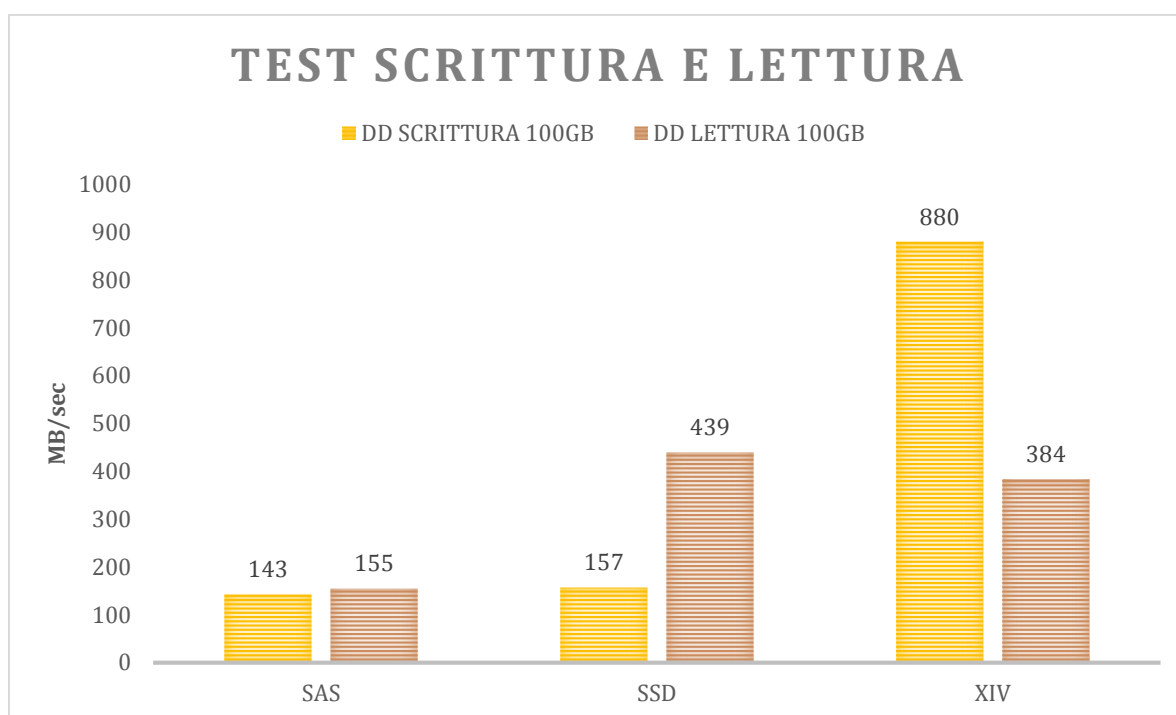
	SAS	SSD	XIV
HDPARM -t	8728 MB/sec	8915 MB/sec	8935 MB/sec
HDPARM -T	150.32 MB/sec	364.36 MB/sec	344.20 MB/sec



RISULTATI

DD

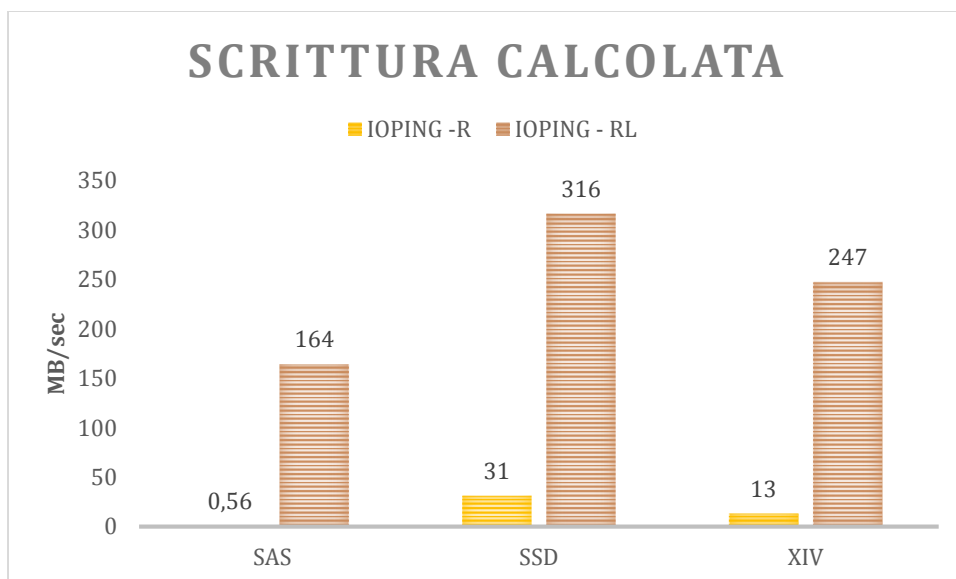
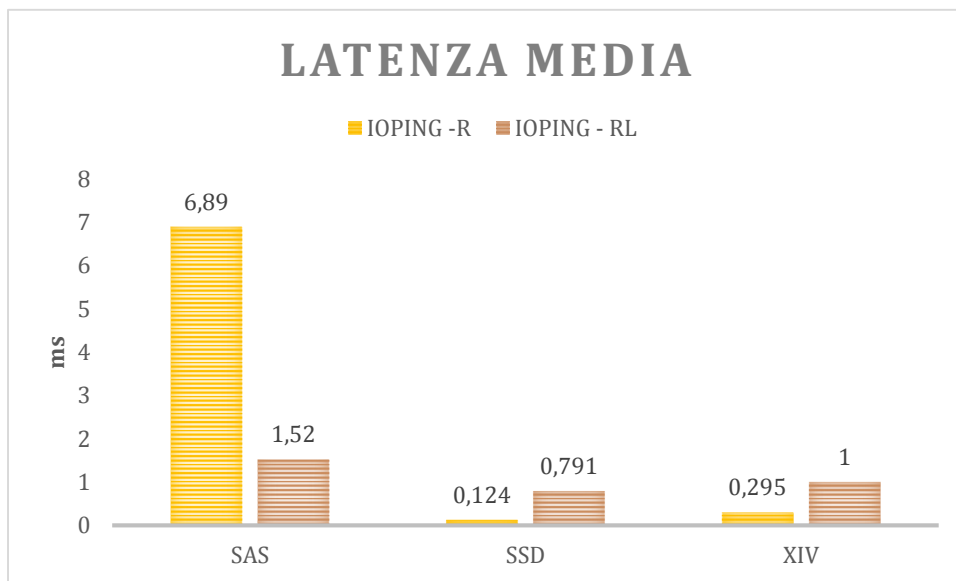
	SAS	SSD	XIV
DD SCRITTURA 100GB	143 MB/s	157 MB/s	880 MB/s
DD LETTURA 100GB	155 MB/s	439 MB/s	384 MB/s



RISULTATI

IOPING

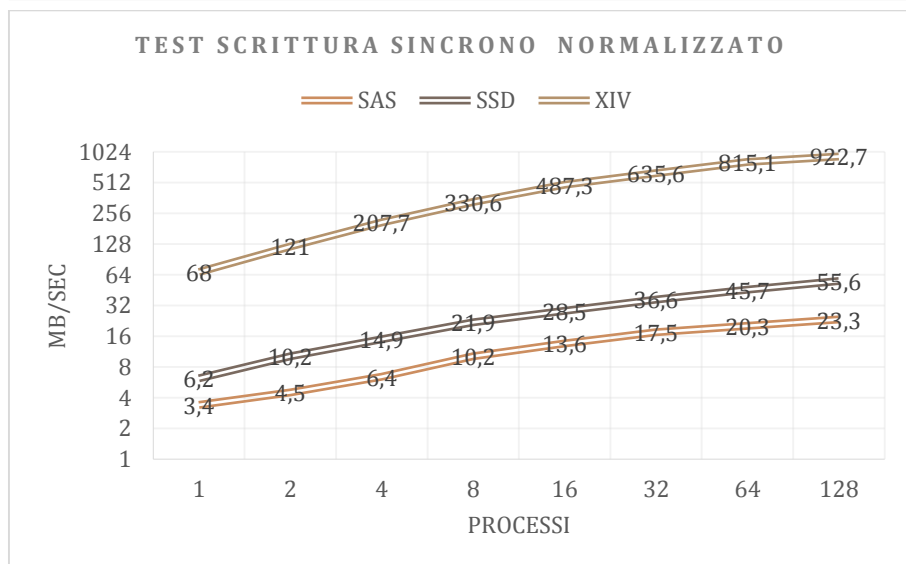
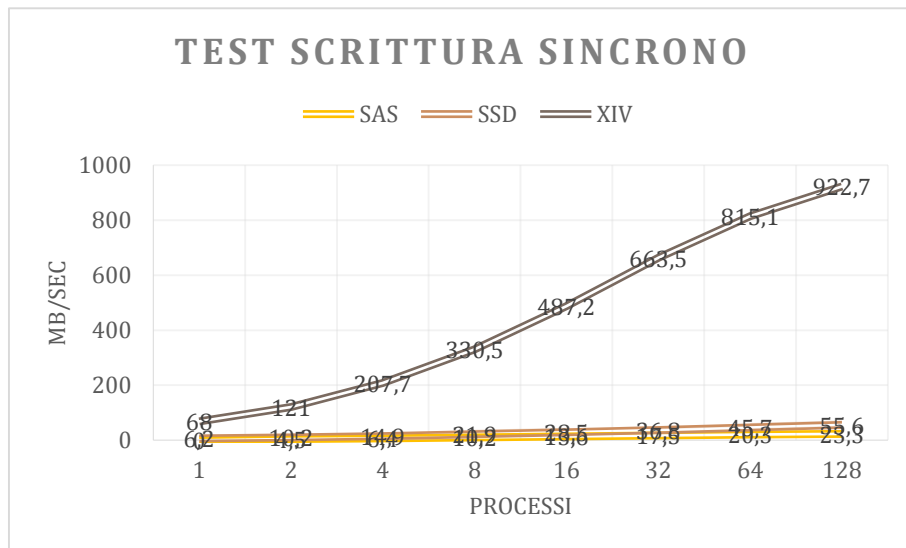
	SAS	SSD	XIV
IOPING -R	145 iops, 580.6 KiB/s min/avg/max/mdev (ms) 1.75 ms / 6.89 ms / 12.8 ms / 2.24 ms	8.0 k iops, 31.3 MiB/s min/avg/max/mdev 62 us / 124 us / 7.6 ms / 88 us	3.4 k iops, 13.2 MiB/s min/avg/max/mdev 162 us / 295 us / 17.0 ms / 557 us
IOPING -RL	657 iops, 164.4 MiB/s min/avg/max/mdev 1.30 ms / 1.52 ms / 13.5 ms / 405 us	1.3 k iops, 316.0 MiB/s min/avg/max/mdev 740 us / 791 us / 2.9 ms / 108 us	987 iops, 247.0 MiB/s min/avg/max/mdev 818 us / 1.0 ms / 4.7 ms / 254 us



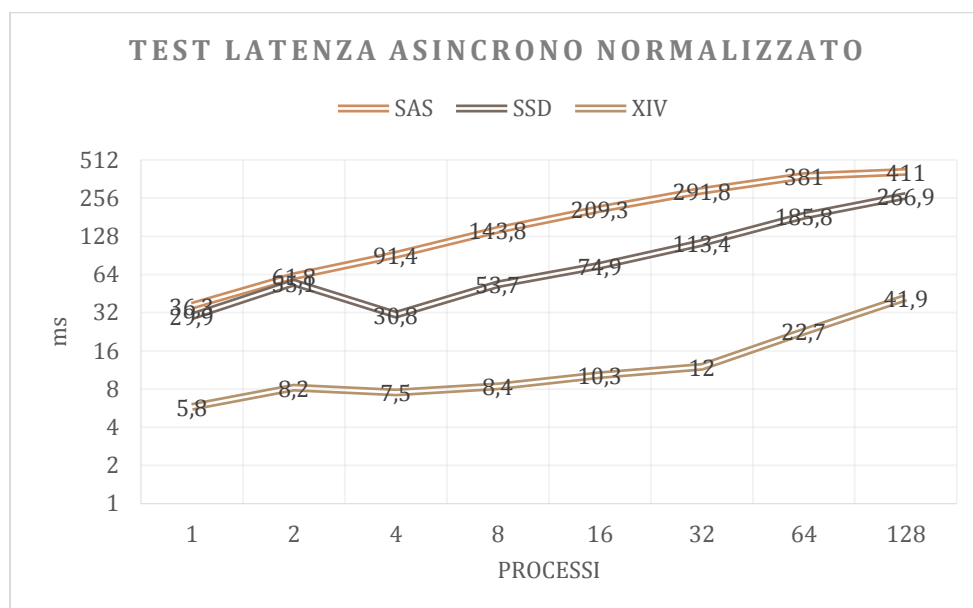
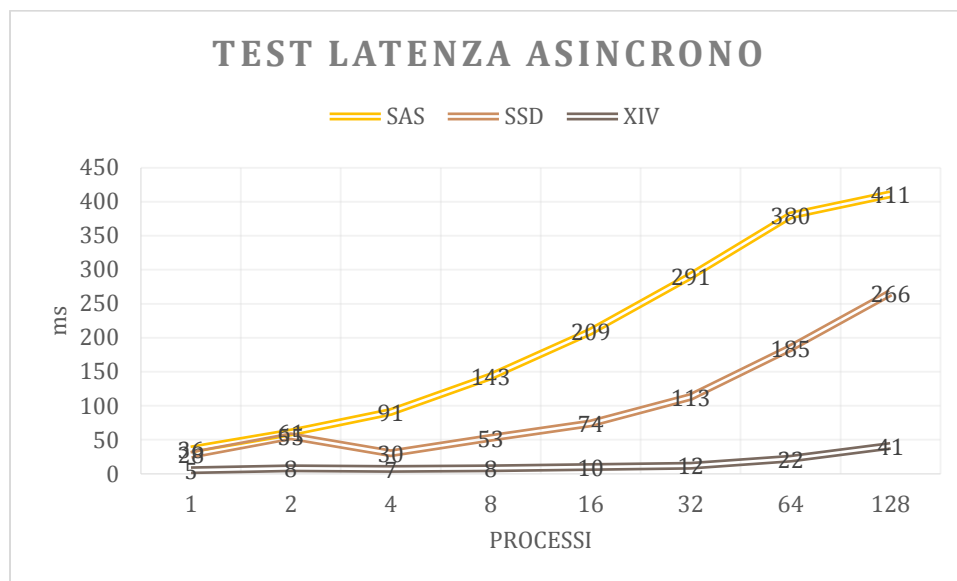
RISULTATI

DBENCH SINCRONO

	SAS	SSD	XIV
DBENCH 1 PROC	3.47987 MB/sec latency 36.325 ms	6.28501 MB/sec latency 28.933 ms	68.0084 MB/sec latency 5.771 ms
DBENCH 2 PROC	4.59014 MB/sec latency 61.770 ms	10.2345 MB/sec latency 55.078 ms	121.057 MB/sec latency 8.155 ms
DBENCH 4 PROC	6.43253 MB/sec latency 91.437 ms	14.9811 MB/sec latency 30.819 ms	207.732 MB/sec latency 7.579 ms
DBENCH 8 PROC	10.2296 MB/sec latency 143.847 ms	21.9186 MB/sec latency 53.665 ms	330.555 MB/sec latency 8.419 ms
DBENCH 16 PROC	13.6704 MB/sec latency 209.320 ms	28.5849 MB/sec latency 74.942 ms	487.267 MB/sec latency 10.269 ms
DBENCH 32 PROC	17.5596 MB/sec latency 291.777 ms	36.8647 MB/sec latency 113.413 ms	653.597 MB/sec latency 12.052 ms
DBENCH 64 PROC	20.3587 MB/sec latency 380.990 ms	45.7354 MB/sec latency 185.764 ms	815.146 MB/sec latency 22.757 ms
DBENCH 128 PROC	23.3151 MB/sec latency 411.056 ms	55.6715 MB/sec latency 266.860 ms	922.71 MB/sec latency 41.857 ms



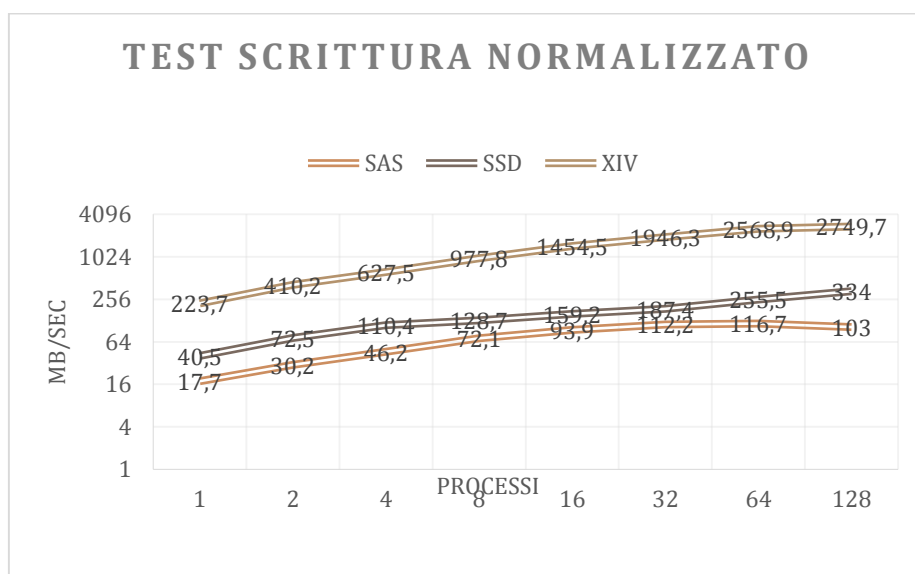
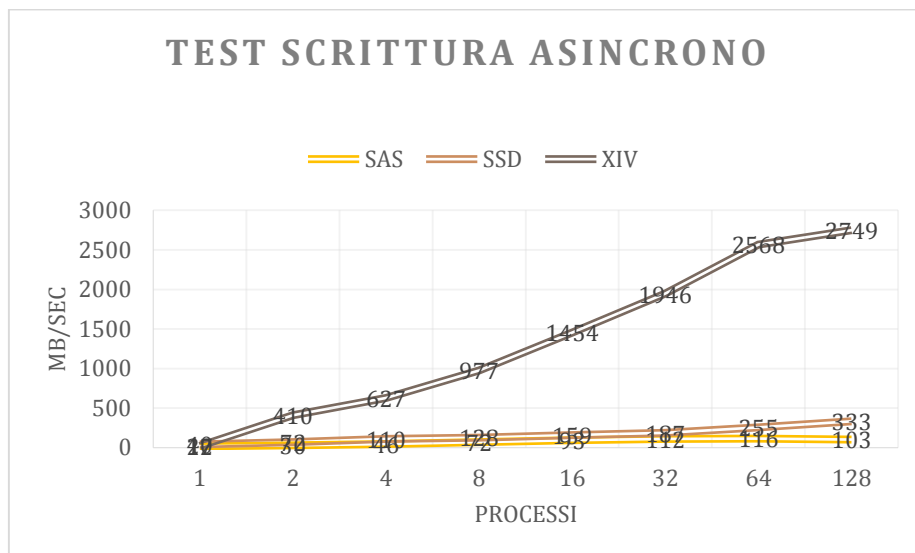
RISULTATI



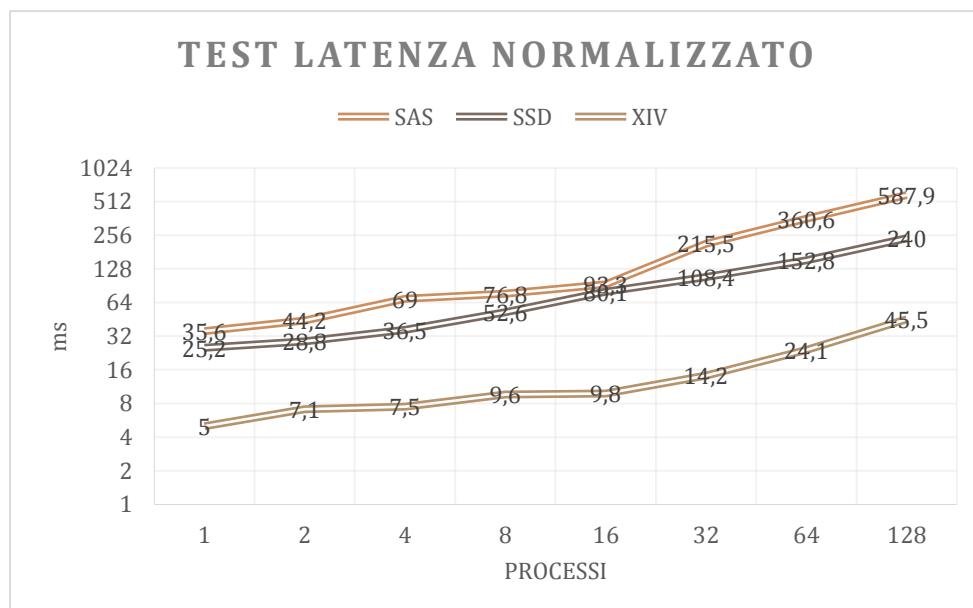
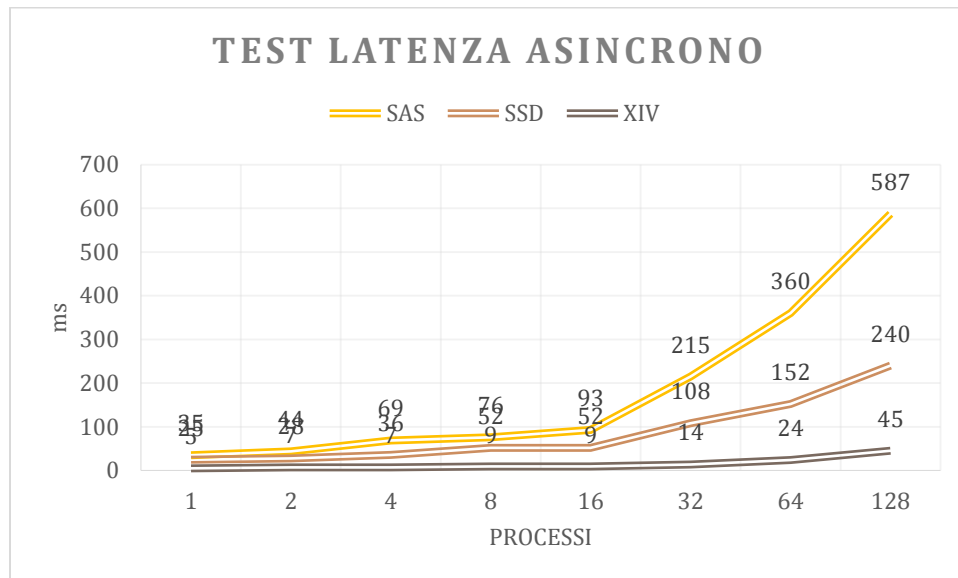
RISULTATI

DBENCH ASINCRONO

	SAS	SSD	XIV
DBENCH 1 PROC	17.7284 MB/sec latency 35.618 ms	40.4829 MB/sec latency 25.218 ms	223.688 MB/sec latency 5.030 ms
DBENCH 2 PROC	30.1915 MB/sec latency 44.229 ms	72.488 MB/sec latency 28.764 ms	410.228 MB/sec latency 7.182 ms
DBENCH 4 PROC	46.2279 MB/sec latency 69.034 ms	110.394 MB/sec latency 36.497 ms	627.514 MB/sec latency 7.522 ms
DBENCH 8 PROC	72.1488 MB/sec latency 76.873 ms	128.743 MB/sec latency 52.582 ms	977.844 MB/sec latency 9.558 ms
DBENCH 16 PROC	93.9349 MB/sec latency 93.267 ms	159.246 MB/sec latency 80.103 ms	1454.52 MB/sec latency 9.786 ms
DBENCH 32 PROC	112.19 MB/sec latency 215.464 ms	187.384 MB/sec latency 108.374 ms	1946.36 MB/sec latency 14.247 ms
DBENCH 64 PROC	116.683 MB/sec latency 360.578 ms	255.488 MB/sec latency 152.761 ms	2568.89 MB/sec latency 24.176 ms
DBENCH 128 PROC	103.067 MB/sec latency 587.912 ms	333.975 MB/sec latency 240.033 ms	2749.69 MB/sec latency 45.548 ms



RISULTATI



Conclusioni

Dall'analisi dei vari risultati dei test, possiamo fare alcune osservazioni interessanti sul comportamento dei dispositivi in esame, sia nei singoli test e iterazioni, che nell'andamento delle loro performance. Ecco alcune delle considerazioni sull'esito dei nostri test:

LETTURA

Il sistema più performante in lettura risulta essere il disco SSD con 439 Mb/sec, subito dopo troviamo XIV con 384 MB/s, pessime performance invece per il disco SAS con solo 155 MB/s MB/sec.

SCRITTURA

Si evidenzia una grandissima differenza di performance in scrittura tra gli storage. In particolare, l'XIV dimostra una capacità di scrittura quasi 6 volte superiore agli altri due dischi con una media di 880 MB/s. Il disco SAS e quello SSD in scrittura nel test eseguito con dd rimangono attestati su valori simili tra loro (circa 150 MB/sec ciascuno).

TEST PARALLELI

Nei test paralleli, si riesce ad evidenziare in che modo questi storage scalano all'aumentare dei processi paralleli. Il grafico mostra chiaramente come la crescita di velocità in scrittura per l'XIV sia quasi lineare all'aumentare dei processi concorrenti nei processi asincroni, e ancora di più in quelli sincroni.

La scalabilità invece dello storage SSD e ancora di più di quello SAS ha un andamento sub-lineare.

LATENZA

I test sulla latenza mettono in evidenza come, all'aumentare dei processi paralleli, lo storage SAS vada in difficoltà, con valori quasi proporzionali all'aumentare dei processi.

Discorso analogo per i dischi SSD, ma con meno criticità all'aumentare dei processi.

Sorprendente invece il risultato per il test di latenza sia sincrono che asincrono dell'XIV, dove la latenza rimane quasi invariata e sempre in un range di valori trascurabili.

I valori medi della latenza nei test di Dbench sono stati calcolati facendo la media dei processi di scrittura per ogni iterazione, scartando il 5% dei valori massimi e minimi di ciascuna ripetizione.

CONCLUSIONI

DRIVER

Dai risultati ottenuti, si evidenzia anche la differenza dei driver utilizzati dai tre dispositivi. In particolare, dalle nostre osservazioni, il driver Emulex utilizzato da XIV risulta essere più performante del driver LSI LOGIC utilizzati sia dai dischi SAS che SSD.

CONSIDERAZIONI

I test eseguiti mettono in evidenza come la scrittura sullo storage XIV sia notevolmente superiore rispetto agli altri due sistemi di storage esaminati, con valori superiori di quasi 6 volte rispetto all'SSD sulla singola scrittura, e più di 8 volte superiore con 128 processi paralleli, raggiungendo picchi di scrittura di 2800 MB/sec nel caso del test asincrono.

I risultati dei test riportano chiaramente che l'IBM XIV essendo uno dei sistemi più evoluti e complessi presenti sul mercato è decisamente più performante in scrittura rispetto a sistemi più semplici come i dischi SSD e i dischi SAS.

RIFERIMENTI:

Riferimenti:

1. IBM XIV Gen 3: <http://www-03.ibm.com/systems/it/storage/disk/xiv/>
2. SAN: http://it.wikipedia.org/wiki/Storage_Area_Network
3. IBM BladeCenter HS23:
<http://www-03.ibm.com/systems/it/bladecenter/hardware/servers/hs23/>
4. CentOS: <http://www.centos.org/>
5. Hdparm: <http://linux.die.net/man/8/hdparm>
6. Dd: <http://linux.die.net/man/1/dd>
7. Ioping: <https://code.google.com/p/ioping/wiki/man>
8. Dbench: <http://linux.die.net/man/1/dbench>

INFORMAZIONI DI CONTATTO

Informazioni di contatto

Tutti i test sono stati eseguiti nel datacenter di Frosinone 2 di Seeweb. Benchmark e articolo sono stati realizzati dai tecnici di Seeweb:

MARCO
CRISTOFANILLI



marco.c@seeweb.it
@m4oc

STEFANO
DI PEDE



stefano.dp@seeweb.it
@stefanodipede

Informazioni sulla società

Seeweb S.r.l.
Corso Lazio 9a
Frosinone
Tel. 800 112825
www.seeweb.com

