

レノボ System x3650 M5サーバーにおける 2.5インチNVMe SSD評価レポート



レノボ・ジャパン株式会社 | インテル株式会社

検証協力

MKTインターナショナル株式会社

ハセゴイノベーション

2016年8月 作成

Intel Inside® 効果的なクラウドを
インテル® Xeon® プロセッサ搭載



【NVMe SSD評価レポート】

1. はじめに	P.02
2. 本評価の目的	P.02
3. 評価結果の要約	P.02
4. レノボSystem x サーバーにおけるSSD最新事情	P.03
5. 評価対象のSSD一覧	P.04
6. 評価環境	P.04
7. ブロックレベルの性能評価	P.05
7.1. NVME SSD単体のI/O性能評価	P.05
7.2. MDドライバーによるソフトウェアRAID(ストライピング)の検証	P.06
7.3. SSD種類によるCPU負荷の比較	P.07
7.4. ファームウェアにおけるドライブ性能差の評価	P.08
7.5. ブロックI/Oレベル検証のまとめ	P.08
8. データベースを用いた性能評価	P.09
8.1. SATA SSDとの比較	P.09
8.2. 従来型PCIE SSDとの比較	P.09
8.3. レプリケーションを用いた場合の性能比較	P.10
8.4. データベースを用いた性能評価のまとめ	P.11
9. 運用について	P.12
9.1. NVME-CLIパッケージについて	P.12
9.2. NVME ドライブの初期化方法	P.12
9.3. FLASHストレージの障害検出について	P.13
9.4. MDによるRAIDの構成例	P.14
10. まとめ	P.14

【NVMe推奨構成】

1. はじめに	P.16
2. 推奨構成パターン1-3	P.17
3. 参考構成	P.18



NVMe SSD評価レポート

1. はじめに

Lenovoサーバーは、大型コンピュータなどの分野で長年にわたる実績を持つIBM社が手がけてきた高機能のx86サーバーです。IBM社のx86サーバー部門がレノボへと統合された現在、サーバー事業を主導していた米本社（ノースカロライナ州ラレー）をはじめ、当時の製品開発を支えていた研究開発施設やエンジニア、さらには製造・供給体制までもがレノボへとそのまま継承されています。このように、すべての要素がレノボへと引き継がれたx86サーバーは、今もなおIBM社の伝統を受け継ぐ安心・信頼のサーバー製品です。

ITリサーチ大手Technology Business Research社（以下「TBR」）による最新の調査の結果、4四半期連続でお客様満足度No.1を獲得しています。

高い信頼性とパフォーマンスを誇るインテル® Xeon® プロセッサーを搭載した最新世代のレノボ System x M5 シリーズは、最新のNVMeテクノロジーであり、非常に高い転送速度と低いレイテンシー（遅延）を実現した最新のFlashテクノロジーインテル® SSD DC P3700ファミリー（PCI Expressスロット版および2.5インチ版）をラインナップに加え、より高い性能と高い保守性を実現しています。

本ドキュメントは、これらの最新のテクノロジーを活用した検証を行うことにより、サーバー製品の選定のための指針を提供するものです。

2. 本評価の目的

本評価では、実際に安全で高い信頼性とパフォーマンスを誇るインテル® Xeon® プロセッサー E5-2600 v4 製品ファミリーを搭載したレノボ System x M5シリーズおよび安定した高性能を実現するP3700 NVMe SSD（インテル® SSD DC P3700 OEMモデル、以後NVMe SSD）を用いて、最新のFlashテクノロジーによってシステムおよびアプリケーションが得られる性能を計測しました。また、SATA/SASおよびPCIe接続型SSDとの利用感を比較しました。

3. 評価結果の要約

評価結果（要約）は下記のとおりです。

- NVMe SSD によって、SATA SSD 8台に相当するIOPS性能が得られました。
- NVMe SSDを2台～8台までRAID-0構成とした場合にIOPSおよび帯域幅がリニアにスケールすることを確認しました。
- NVMe SSD 1台で従来型PCIe SSD 1台と同等もしくはそれ以上の性能が得られることを確認しました。また、NVMe SSDのRAID-0構成によりさらに性能がスケールしました。



4. レノボSystem x サーバーにおけるSSD最新事情

レノボSystem x サーバーシリーズでは、かねてよりNAND Flashテクノロジーによるソリッドステートドライブ (SSD) の標準オプションとして提供してきました。さらに、最新世代のSystem x M5シリーズでは、最新のNVM Express (NVMe) 準拠した、広帯域幅・低遅延なSSDがオプションに加わりました。

コンピュータシステムでは過去50年以上にわたり、回転型の磁気メディアを用いた記憶装置であるハードディスクドライブ (HDD) が広く利用されてきました。そして現在、ソリッドステートドライブ (SSD) がこれまでのHDDの役割を置き換えつつあります。

身の回りのタブレットデバイスやスマートフォンでは、記憶装置としてNAND Flashメモリを用いられています。また、ノートパソコンでも、NAND Flashメモリを搭載したSSDの利用が一般的となりました。クライアント側へのNAND Flash適用が進む中、サーバーにおいてもクライアントからの大量の要求に対して、安定して良好なレスポンスタイムを実現するためにSSD活用が重要です。

以前のレノボSystem x サーバーでは、SATA/SAS SSDに加えて、NVMe登場以前の独自規格によるPCIe SSDをオプションパーツとして提供し、I/O要求が高いワークロードを実現してきました。System x M5世代で新たにラインナップに加えられたNVMe SSDは、従来製品よりも高いI/O性能や管理性を実現します。

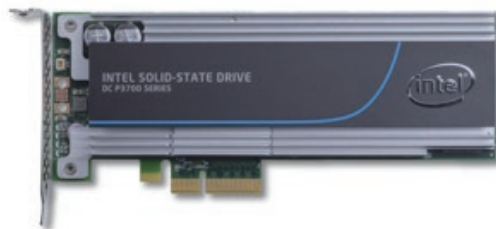


図1 NVMe SSD (PCIeスロット搭載型)

NVM Expressは、不揮発なメモリ技術を用いて広帯域幅・低遅延なSSDを可能にするための標準規格です。ハードディスク時代のインターフェイスではなく、不揮発メモリの性能を引き出すためのインターフェイスとして登場しました。

また、これまでのサーバー搭載型フラッシュストレージでは、サーバー内部のPCIeスロット搭載型SSDが利用されてきました。NVM Expressでは、PCIeスロット搭載型SSDに加え、従来のHDDやSAS/SATA SSDと同様、2.5インチのドライブベイに搭載できるSSDも提供されています。



図2 NVMe SSD (2.5インチ型)



従来のPCIeスロット搭載型 SSDでは、メンテナンスのためにサーバーのカバーを開ける必要がありましたが、NVMeに準拠する2.5インチタイプのドライブであれば、従来のHDDやSSDと同様に、サーバー前面のドライブベイに装着できます。

5. 評価対象のSSD一覧

以下に、今回評価したNVMe SSD の一覧を示します。

評価対象のNVMe SSD (Intel® SSD DC P3700 OEM製品)

P/N	コンポーネント名	容量
00YA818	P3700 400GB NVMe 2.5型 G3HS Enterprise Performance PCIe SSD	400GB
00YA821	P3700 800GB NVMe 2.5型 G3HS Enterprise Performance PCIe SSD	800GB
00YA824	P3700 1.6TB NVMe 2.5型 G3HS Enterprise Performance PCIe SSD	1.6TB
00YA827	P3700 2TB NVMe 2.5型 G3HS Enterprise Performance PCIe SSD	2TB

本評価レポートでは、これまでSystem x 旧モデルで広く利用されてきたSSD製品との性能差を示すため、下記SSD製品のスコアを参考値として掲載しています。

比較対象 (SATA SSD)

P/N	コンポーネント名	容量
90Y4377	1.2TB High IOPS MLC Mono アダプター	1.2TB
00WG620	S3510 120GB Enterprise Entry SATA G3HS 2.5型 SSD	120GB

6. 評価環境

本評価に使用したハードウェア、ファームウェア、そしてソフトウェアのバージョン構成およびバージョンを以下に示します。

評価に使用したハードウェア

P/N	コンポーネント名	数量
8871F4J	System x3650 M5	1
00YJ199	インテル® Xeon® E5-2640 v4 (10) 2.40GHz~2133MHz	2
46W0829	16GB 2R PC4-19200 RDIMM CK	8
00KA498	PCIe ライザー・カード (2x8 FH/FL + 1x8 FH/HL)	2
49Y7960	インテル® X520 デュアルポート 10GbE SFP+ アダプター	



ファームウェアバージョン一覧

コンポーネント	バージョン
UEFI	2.10 (Build: TCE124M)
IMM2	3.00 (Build: TCOO18M)
ServeRAID M5210	24.9.0-0029

ソフトウェアバージョン一覧

ソフトウェア	バージョン
オペレーティングシステム	Red Hat Enterprise Linux 7.2 カーネル 3.10.0-327.22.2.el7.x86_64
Flexible I/O Tester	2.2.8-2.el7.x86_64
MySQL Community Server	5.7.14
tpcc-mysql	コミット faa06df (Feb 29, 2016)

SATA SSDの評価にあたっては、SSD自体の性能を測定するため、システムボードに搭載されたServeRAID M5210(キャッシュ接続なし)に、JBODのドライブとして接続しました。

7. ブロックレベルの性能評価

SATA SSDおよびNVMe SSDに対してブロックI/Oワークロードを実行し、得られるIOPS性能および帯域幅性能を評価しました。評価にあたってはソフトウェアにはFlexible I/O Tester(FIO)を使用しました。FIOにて指定した、おもなパラメータ値は下記のとおりです。

パラメータ名	ブロックサイズ	ジョブ数	IOキュー深度
設定値	4KB	32	32

7.1. NVMe SSD単体のI/O性能評価

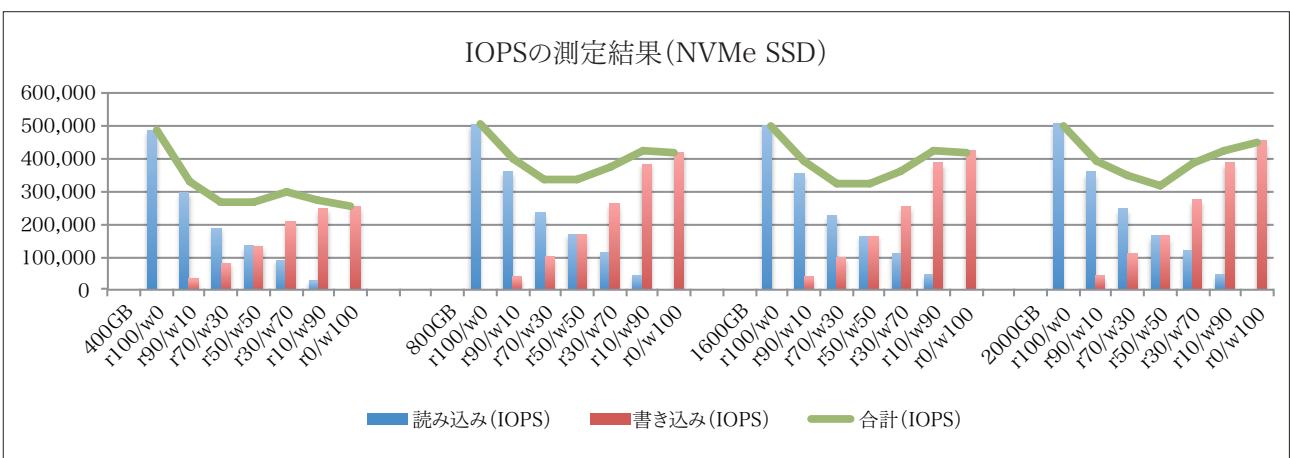


図3 NVMe SSDのIOPS評価(P3700 400GB, 800GB, 1.6TB, 2TBモデル)



Intel Inside® 効果的なクラウドを
インテル® Xeon® プロセッサ搭載

400GBモデルでは、書き込みI/Oが混じるとIOPSの値が低下する傾向があり、読み込みと書き込みを合わせた合計IOPSはL字型のグラフとなりました。800GB以上のモデルでは合計IOPSのプロット結果がV字型となっており、書き込みワークロードにおいても高いIOPS性能が発揮されることが判ります。

続いて、以下のグラフはNVMe SSDにおける帯域幅の測定結果を示します。以下のグラフは、左側から順に、400GB、800GB、1.6TB、2TBモデルにおいてブロックサイズ1MBのランダムI/Oワークロードを実行した結果を示しています。

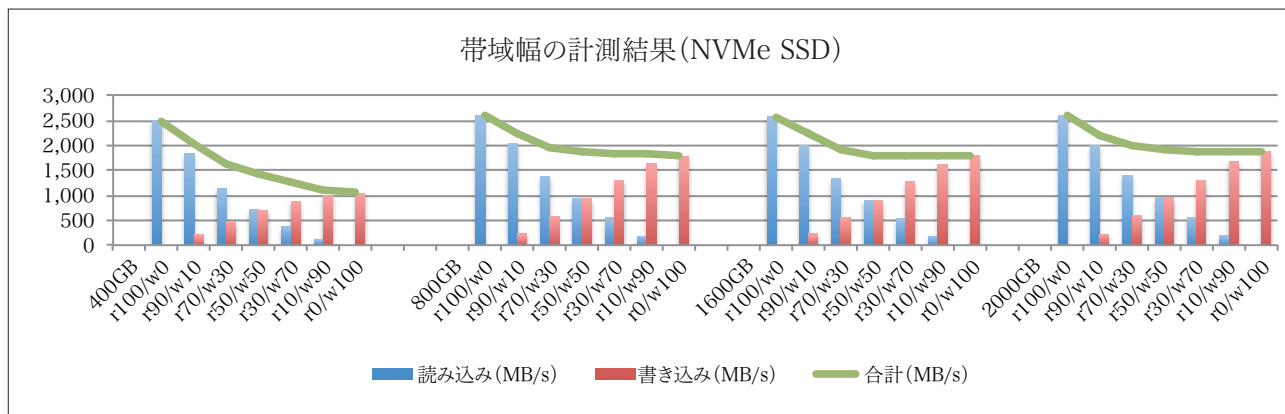


図4 読み書き帯域幅(P3700 400GB, 800GB, 1.6TB, 2TBモデル)

帯域幅のグラフでは、I/Oのバランスが読み込みから書き込みにシフトするにつれて全てL字型となっていますが、特に400GBモデルではその落ち込みの割合が大きくなっています。

SSDは大容量のものほど内部チャンネル数が多く、ピーク性能が高くなる傾向があります。このため、総データ容量の見込みから400GBモデルで十分と考えられる場合でも、性能の観点から800GBモデル、もしくはより大容量なモデルを選定することで、より安定した性能が得られることが示されています。

7.2. mdドライバーによるソフトウェアRAID(ストライピング)の検証

Linuxのmdドライバーを用いて、SATA SSDもしくはNVMe SSDを単体、各2台、4台、6台、8台でストライピングした場合に得られる性能を測定しました。下記にIOPS比較グラフを示します。



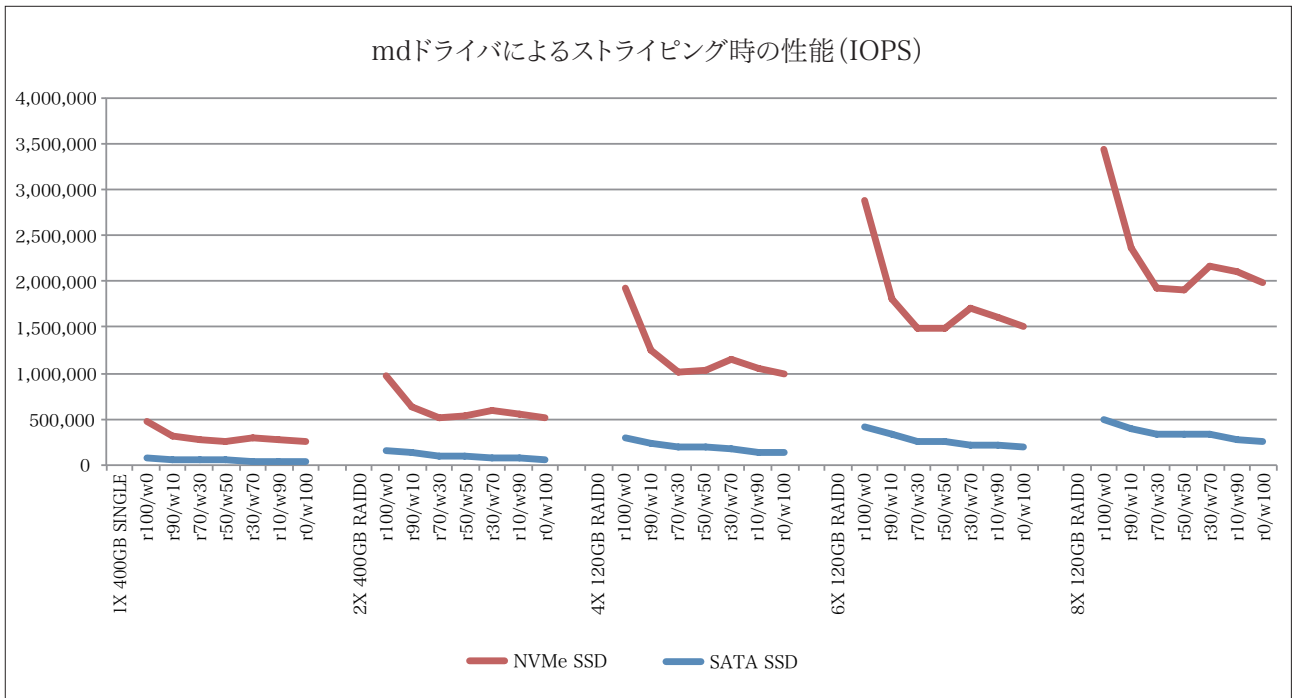


図5 mdドライバーによるストライピング時の性能比較 (SATA/NVMe SSD)

NVMe SSDは、SATA SSDに対して高い性能を示しました。NVMe SSD 1台で、SATA SSDを8台ストライプ相当に匹敵するIOPS性能を示しました。

7.3. SSD種類によるCPU負荷の比較

続いて、NVMe SSDと従来型のPCIe SSD (ioDrive2 OEM)のCPU負荷を測定しました。評価にあたっては、FIOを用いてNVMe SSD、従来型のPCIe SSD各デバイスに対して3倍の容量の4Kランダムライトを実行しました。その後、Flexible I/O Testerにて同ワークロードを1時間実行した間のCPU負荷(ロードアベレージ)をグラフにプロットしました。

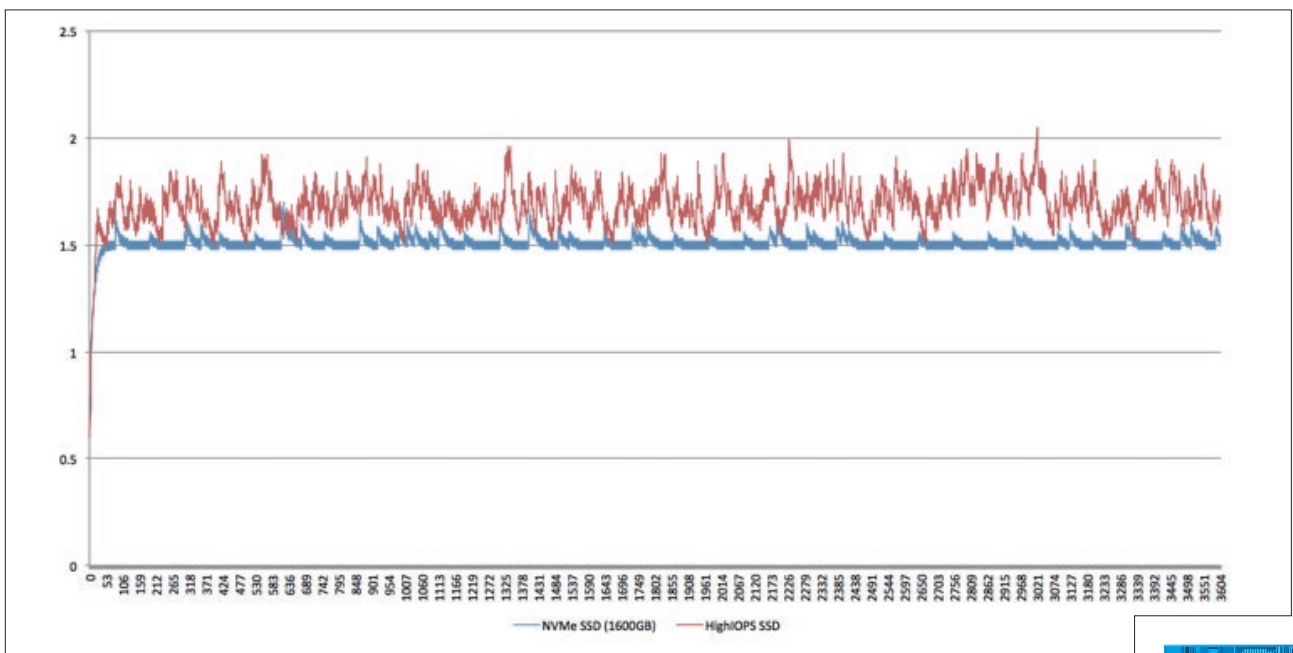


図6 NVMe SSD、従来型PCIe SSD利用時のロードアベレージ比較



Intel Inside® 効果的なクラウドを
インテル® Xeon® プロセッサ搭載

NVMe SSDに対するテスト時のロードアベレージは、およそ1.50でほぼ安定しています。これに対して、従来型のPCIe SSDに対するテスト時のロードアベレージでは、最大でおよそ2.0(従来よりも最大0.5高い値)を示しています。この結果から、従来型のPCIe SSDを利用した場合のほうが、CPU負荷が高くなる傾向にあります。

一般的に、SSDは内蔵コントローラでドライブ内部の制御を行います。今回の比較に使用した従来型のPCIe SSDカードの場合はコントローラが搭載されていないため、ホストCPUがPCIe SSDの制御に利用されており、NVMe SSDよりもCPU負荷が高くなっていることがわかります。このことから、NVMe SSDは性能と同時にホストCPUの負担低減が実現されることが確認できました。

7.4. ファームウェアにおけるドライブ性能差の評価

NVMe SSDのファームウェア 下記2点において、ドライブのI/O性能差の有無を評価しました。

- ・8DV1LP11
- ・8DV1LP10

これら2点のファームウェアにおいて、今回のメトリックにおいて有意な性能差は確認されませんでした。次に、両ファームウェアのIOPS性能グラフを示します

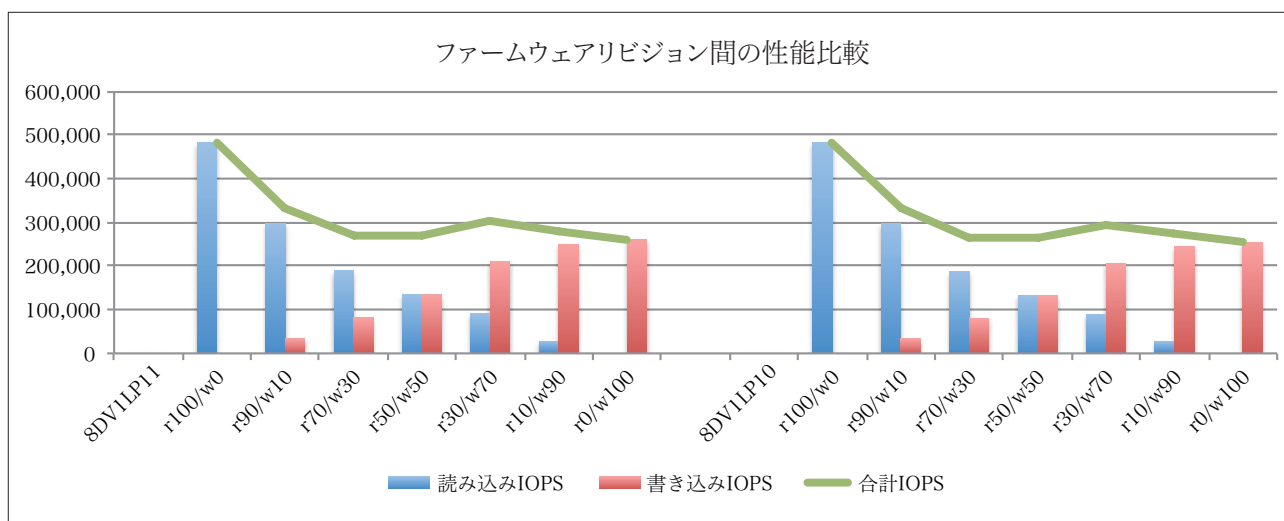


図6 NVMe SSDファームウェアリビジョン間の性能比較 (IOPS)

7.5. ブロックI/Oレベル検証のまとめ

ブロックI/Oレベルの検証結果から、NVMe SSDがSATA SSDや従来型PCIe SSDと比較し、より高い性能とCPU負荷の低減を両立していることを確認できました。

- ・NVMe SSDのうち容量最小の400GBモデルに対し、800GBもしくはそれ以上のモデルがより高い書き込み性能を示しました。
- ・NVMe SSD 1台で、SATA SSD 6台相当、従来型PCIe SSD 1台分以上のIOPS性能を示しました。



8. データベースを用いた性能評価

サーバーにおける高速SSDの活用用途として、リレーショナルデータベース (RDBMS) のデータ領域に対するI/Oの高速化による処理トランザクション数の向上が挙げられます。今回は、実際に各種SSDにおけるMySQLデータベースのトランザクション性能を計測し比較、評価しました。

8.1. SATA SSDとの比較

従来のSATA SSDとNVMe SSDでTPC-Cベンチマークを実行し、時間あたりのトランザクション処理数を計測しました。WH=1000のTPC-Cベンチマークを実行し、時間あたりのトランザクション数を比較しました。

評価にあたってSATA SSDはmdドライバーを用いて2台、4台、6台のドライブでRAID-0(ストライピング)を構成しました。

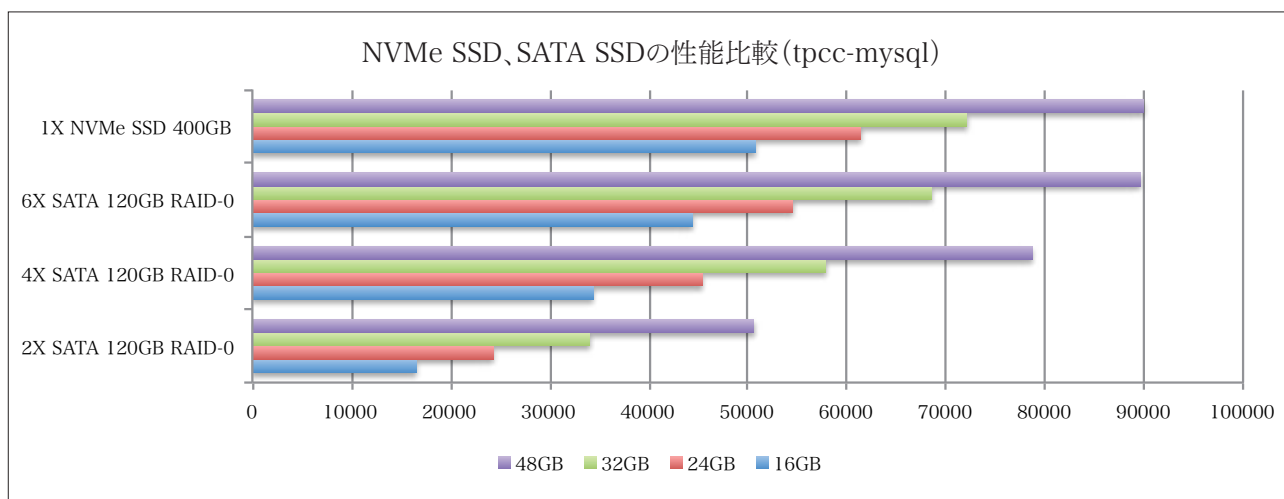


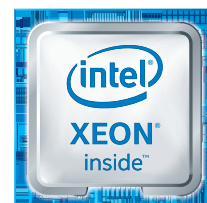
図7 NVMe SSD, SATA SSDの性能比較 (tpcc-mysql)

SATA SSD 2台構成では、NVMe SSD 1台で得られるトランザクション量に対して33%~56%、SATA SSD4台構成では68%~88%の性能を示しました。SATA SSD 6台構成では88%~およそ100%の性能を示し、状況によってはNVMe SSDとほぼ同等の性能となりました。この結果から、NVMe SSD 1台で、SATA SSD 6台相当の性能が得られることがわかります。

全体的な傾向として、MySQLのInnoDBバッファプールサイズが小さいときほど、NVMe SSDとの性能差がひらく傾向にありました。このことから、NVMe SSDはデータ量が大きくなった場合により高い性能を示すことがわかります。

8.2. 従来型PCIe SSDとの比較

5世代までのSystem x サーバーでは、HighIOPS SSD PCIeアダプタ (以後HighIOPS SSD) が提供されてきました。今回は、これまで高速データベース向けに利用されてきたHighIOPS SSDと、NVMe



SSDの性能を相対比較しました。TPC-Cベンチマーク(WH=4000)を実行し、時間あたりのトランザクション数を比較しました。計測に使用したNVMe SSDは800GB、ストライピング状態における性能計測には性能がほぼ同等である1600GB、2000GBを組み合わせて使用しています。

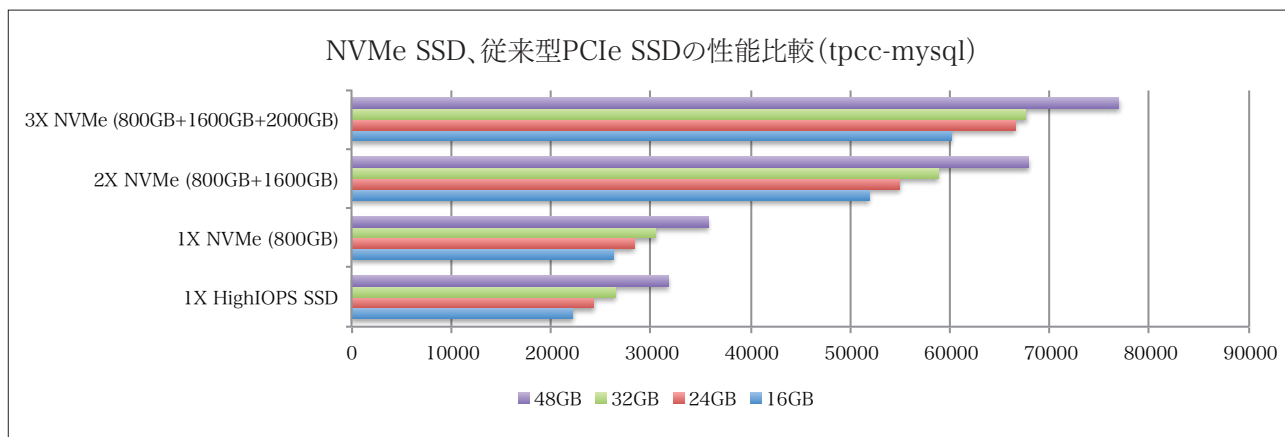


図8 NVMe SSD、従来型PCIe SSDの性能比較 (tpcc-mysql)

NVMe SSD 1台は、従来型PCIe SSDであるHighIOPS SSD 1台よりも高いトランザクション性能を発揮しました。またmdドライバを用いてNVMe SSDを2台するとトランザクション数がほぼ倍増し、3台ストライピングした場合においてもトランザクション数がさらに向上しています。

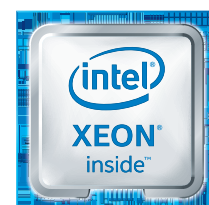
MySQLがシステムメモリ上に確保するバッファ量を指定するInnoDBバッファプールサイズを16GB～48GBの間で変更していくと、各構成において容量が大きいほど時間あたりのトランザクション数が増加する傾向がNVMe SSD、HighIOPS SSDの両方に見られました。

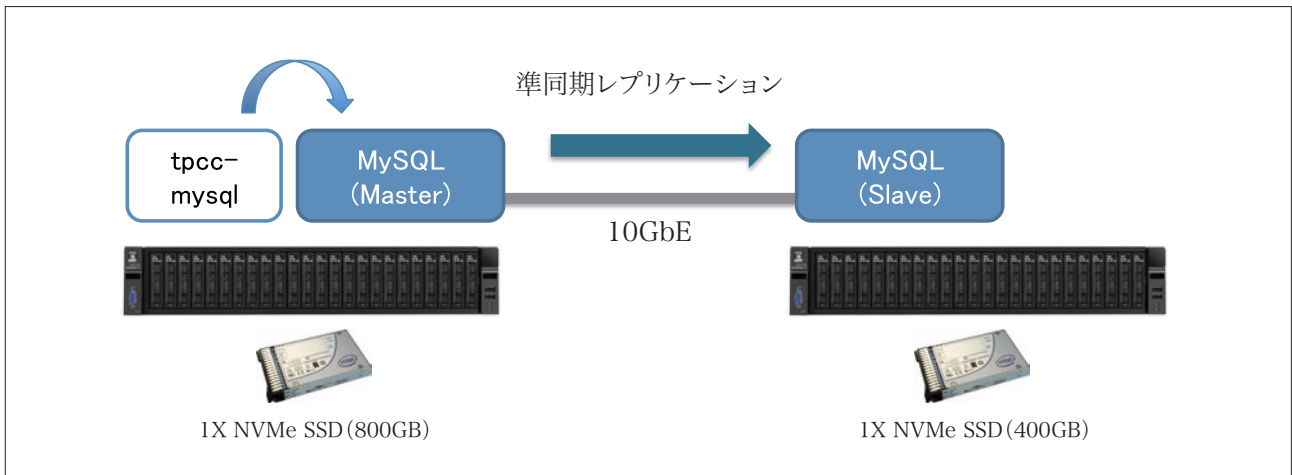
この結果から、従来System x サーバーでHighIOPS SSDを利用していた場合でも、NVMe SSDで同等、もしくはそれ以上のパフォーマンスが得られることがわかります。

8.3. レプリケーションを用いた場合の性能比較

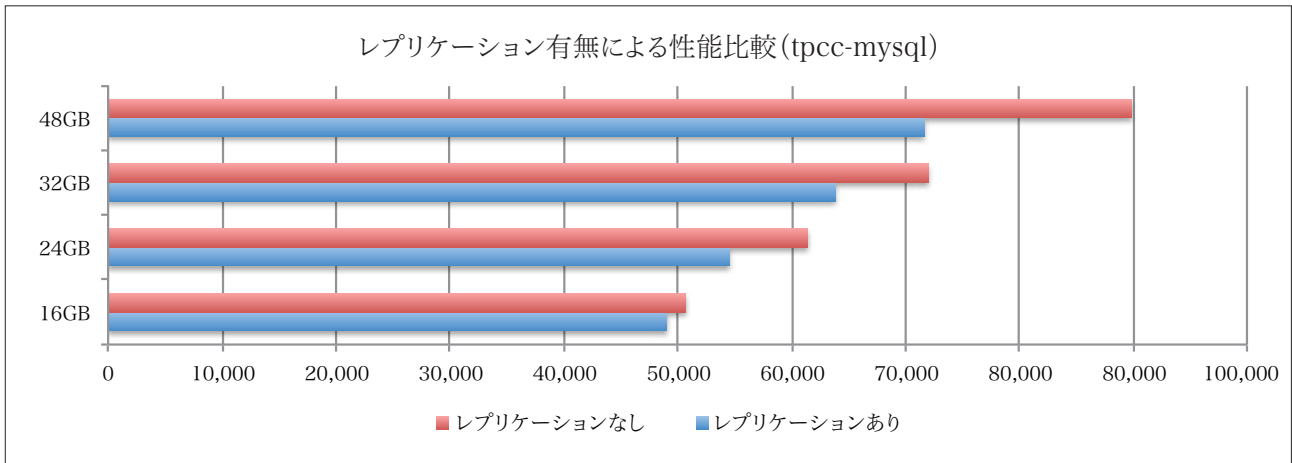
MySQLにはデータベースレプリケーション機能が含まれています。MySQL のレプリケーションでは、マスターサーバで更新された変更がスレーブサーバに伝播、反映されます。レプリケーションを用いることで、複数のMySQLサーバにより多くのワークロードを処理できます。

NVMe SSDを搭載した環境において、レプリケーションあり・なしの場合のMySQLデータベース性能を比較しました。今回の測定では、MySQLの準同期レプリケーションを用いました。準同期レプリケーションは、レプリケーションログが最低でも1台のスレーブサーバのドライブ上への書き込み完了通知を待つレプリケーションモードです。レプリケーションの構成にあたって、ネットワークがボトルネックとなる状況を避けるため、2台のサーバを10Gビットイーサネット接続し、MySQL間の通信経路として使用しました。





性能評価にはTPC-Cベンチマークを実行し、WH=1000(データサイズ 約100GB)のTPC-Cベンチマークを実行し、時間あたりのトランザクション数を比較しました。



評価の結果、準同期レプリケーションを有効化した場合でも、レプリケーションを行わない場合の7～9割のスループットが得られることが確認されました。特に、InnoDBバッファプールに対してデータが多い場合に、レプリケーションなしの場合の性能に近づいています。多くの場合ではデータサイズがInnoDBバッファプールサイズを上回るため、レプリケーションで有効なパフォーマンスが得られると考えられます。NVMe SSDを搭載したSystem x サーバーでMySQLクラスタを構築することにより、NVMe SSDの性能を活かした、高性能なデータベースシステムを構築できることがわかりました。

8.4. データベースを用いた性能評価のまとめ

各種SSDでMySQLを用いたTPC-Cワークロードを実行し、性能を計測した結果、下記のとおりNVMe SSDが優れた性能を示しました。

- NVMe SSD 1台でSATA SSD 6台相当、従来型PCIe SSD 1台以上の性能を発揮しました。
- データベースサイズが大きくなるほど、NVMe SSDの性能が他のSSDよりも高いトランザクション数を示す傾向がありました。



Intel Inside® 効果的なクラウドを
インテル® Xeon® プロセッサ搭載

本結果より、NVMe SSDは従来型のPCIe SSDの後継として利用できることが確認できました。また、NVMe SSDは2.5インチのフォームファクタでこれらの性能を達成でき、ストライピングの構成や運用中のメンテナンスも容易です。より大きなデータセットを、より簡単に扱える点も魅力です。

9. 運用について

サーバー搭載型SSDの魅力のひとつは、セットアップが簡単でありながら、高いI/O性能が得られる点にあります。ただ、SSDはHDDと異なった特性を持つストレージデバイスであり独特の操作があります。また、SSDの導入基準はパフォーマンスだけではありません。システム運用にあたっては、信頼性に加え運用監視の方法が明確であることが重要です。

ここでは、System x に装着されたNVMe SSDの基本的操作の例、およびLinux環境におけるソフトウェアRAIDの利用方法など、はじめてのNVMe SSDの導入・運用にあたってまずは覚えておきたい幾つかの事項について紹介します。

9.1. nvme-cliパッケージについて

Red Hat Enterprise Linux 7で利用する場合、オープンソースのNVMe 管理ツール nvme-cliパッケージに含まれるnvmeコマンドを用いることによってNVMe SSDのメンテナンスに必要な機能を利用できます。

9.2. NVMe ドライブの初期化方法

SSDのNAND Flashメモリには、それまでに書き込まれたユーザーデータ、SSD自体のメタデータなどが保存されます。SSDの性能は累積した書き込みデータの状況に影響を受けます。このため、従来型PCIe SSDでは論理フォーマットを行い初期状態に戻すコマンドなどが提供されている場合があります。また、SATAドライブの場合にはTRIMもしくはDISCARDコマンドによって書き込み済みのデータをクリアする手法がとられることもあります。

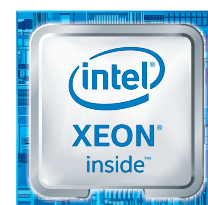
NVMe SSDの場合にもフォーマット処理によって内部のデータ構造をリセットすることが可能です。nvmeコマンドで下記の要領でNVMe SSDをフォーマットできます。

```
# nvme format /dev/nvme0
```

※1台目のNVMe SSDを初期化する場合

```
[root@localhost ~]# nvme list
Node                               SN                               Model                               Version  Namespace  Usage                               Format                               FW Rev
-----
/dev/nvme0n1                       CVFT5434007P400GGN             INTEL SSDPE2MD400G4L              1.0     1          400.09 GB / 400.09 GB             512 B + 0 B                         80V1LP10
/dev/nvme1n1                       CVFT5402007X400GGN             INTEL SSDPE2MD400G4L              1.0     1          400.09 GB / 400.09 GB             512 B + 0 B                         80V1LP10
/dev/nvme2n1                       CVFT5434008P400GGN             INTEL SSDPE2MD400G4L              1.0     1          400.09 GB / 400.09 GB             512 B + 0 B                         80V1LP10
/dev/nvme3n1                       CVFT5434008F400GGN             INTEL SSDPE2MD400G4L              1.0     1          400.09 GB / 400.09 GB             512 B + 0 B                         80V1LP10
[root@localhost ~]# nvme format /dev/nvme3
Success formatting namespace: ffffffff
[root@localhost ~]#
```

図6 nvme formatコマンドの実行例



Intel Inside® 効果的なクラウドを
インテル® Xeon® プロセッサ搭載

9.3. Flashストレージの障害検出について

他のコンポーネントと同様、SSDが健全に稼働しているか監視する手法は、システム管理者にとっての関心事のひとつです。また、Flashメモリは書き換え回数が有限な記憶メディアとして知られています。このため、各SSDへの累計書き込み量も運用監視の対象として関心が寄せられます。

SSDは、Flashメモリの特性上、書き込みの繰り返しにより、メディアの信頼性が低下していきます。繰り返しの書き込みやNAND Flashメディアの不具合により一部領域が利用できなくなった場合でも、通常はSSDの内蔵コントローラによりスペア領域の割り当てが自動的に行われます。しかし、メディア上に不良ブロックが大量に発生し、ドライブに搭載されているスペア領域を使い切ってしまうと、二次記憶装置としての機能に支障をきたす可能性があります。このため、SSDには、ドライブへ書き込み可能な総書き込み量がPBW(Petabytes Written)もしくはTBW(Terabytes Written)として設定されています。この挙動を考慮し、SSDの運用監視にあたってはスペア消費状況やメディアへの累計書き込み量がモニタリング項目として注目されます。

従来のHDDやSATA/SAS SSDではSMARTにより健全性を示すデータを取得することが可能ですが、NVMe SSDにおいてもコマンドによりデバイスの健全性の確認が可能です。

次にnvme smart-logコマンドの出力例を示します。

```
[root@node1 dest]# nvme smart-log /dev/nvme0n1
Smart Log for NVME device:nvme0n1 namespace-id:ffffffff
critical_warning          : 0
temperature              : 19 C
available_spare           : 100%
available_spare_threshold : 10%
percentage_used           : 0%
data_units_read          : 20421220
data_units_written       : 22838914
host_read_commands       : 713650779
host_write_commands      : 377662231
controller_busy_time     : 0
power_cycles              : 101
power_on_hours            : 1675
unsafe_shutdowns         : 0
media_errors              : 0
num_err_log_entries      : 0
Warning Temperature Time : 0
Critical Composite Temperature Time : 0
Temperature Sensor 1     : 0 C
Temperature Sensor 2     : 0 C
Temperature Sensor 3     : 0 C
Temperature Sensor 4     : 0 C
Temperature Sensor 5     : 0 C
Temperature Sensor 6     : 0 C
Temperature Sensor 7     : 0 C
Temperature Sensor 8     : 0 C
[root@node1 dest]#
```

図7 nvme smart-logコマンドの実行例

percentage_usedは、これまでホスト側から要求された書き込み操作が、書き込み可能な総書き込み量の何パーセントにあたるかを示します。この値を定期的にレビューし、運用期間中にドライブの交換が必要か確認できるほか、その傾向からいつごろ必要となるか予測できます。

available_spareは、ドライブが持つスペア領域のうち何パーセントがまだ残っているかを示しています。SSDを長期間にわたって運用し記憶メディアの不良ブロックが発生し、ブロックの代替処理が進行するにつれて、この値が低下していきます。残りスペア領域の減少状況を監視することで、ドライブのスペア領域が不足する自体になる前に対処できます。



9.4. mdによるRAIDの構成例

mdはLinuxオペレーティングシステムに含まれるソフトウェアRAIDドライバです。複数のSSDでRAID-0(ストライピング)を構成すると、大容量かつ高速な仮想デバイスとしてSSDを利用できます。RAID-1(ミラーリング)を構成すると、ひとつのデータが複数のSSDに同時書き込みされ、仮にいずれかのSSDの動作が停止した場合でもデータ喪失を避けられます。

```
mdadm --create /dev/md0 --level=0 --raid-devices=2 ¥
/dev/nvm[01]n1
```

システム再起動時、アレイを初期化するには/etc/mdadm.confが参照されます。/etc/mdadm.confファイルは下記の要領で生成できます。アレイの情報が本ファイルに保存されていない場合、システム再起動後にRAIDアレイを示すブロックデバイス名が変化する場合があります。

```
mdadm --detail --scan >> /etc/mdadm.conf
```

RAIDアレイの稼働状況は/proc/mdstatファイルを読み取ることで確認できます。下記は3つのNVMeドライブでRAID-0(ストライピング)を構成し、正常に動作している場合の例です。

```
# cat /proc/mdstat
Personalities : [raid0]
```

```
md0 : active raid0 nvme2n1[2] nvme1n1[1] nvme0n1[0]
1171740672 blocks super 1.2 512k chunks
```

Red Hat Enterprise Linux 7のmdドライバの詳細については、同オペレーティングシステムのドキュメントを参照してください。

10. まとめ

レノボ System x 3650 M5とP3700 NVMe SSDを組み合わせた場合の性能について評価した結果、以下の結果が得られました。

- NVMe SSD 1台で、SATA SSD 8台相当のIOPS性能、6台相当のデータベース性能を実現
- 旧 System x サーバー向けの従来型PCIe SSDを超える性能を発揮
- NVMe SSDが、大きなデータセットに対してより高い性能を発揮

性能面の優位性に加え、レノボ System x M5では、2.5インチタイプ/PCIeスロットタイプのP3700 NVMe SSD (インテル® SSD DC P3700 OEM品)が用意されています。旧世代のSystem x シリーズよりも柔軟な構成が可能であるほか、2.5インチタイプの場合はサーバー前面のドライブベイにて装着/脱着が可能のため、保守性に優れるなど、様々なメリットがあります。



本レポートはレノボ・ジャパン株式会社ならびにインテル株式会社の依頼により、MKTインターナショナル株式会社、ハセゴイーノベーション、日本仮想化技術株式会社(監修)の三社が第三者検証協力としてNVMeの評価を実施し、Lenovo製品の推奨構成を付与したものです。

従来型SSDとNVMeディスクとの比較についてIO性能、アプリケーション性能、運用上のポイントの観点で記載しております。サーバー検討、ディスク構成の検討に是非お役立てください。

【MKTインターナショナル株式会社について】

MKTインターナショナル株式会社は、「知識をオープンに共有し、価値を創造する使命を持って、お客様、社会から信頼される企業を目指す」ことをミッションに掲げ、2011年4月の設立以降、日本で随一の法人向けマーケティング活動を支援してきました。

会 社 名：MKTインターナショナル株式会社

所 在 地：神奈川県川崎市

設立年月日：2011年4月28日

代 表 者：代表取締役社長 赤井 誠

主な事業内容：事業企画及びマーケティングコンサルティング、マーケティング委託業務等

ホームページ：<http://www.mkt-i.jp/>

【ハセゴイーノベーションについて】

屋 号：ハセゴイーノベーション

所 在 地：神奈川県川崎市

代 表 者：長谷川 猛

主な事業内容：各種コンサルティングおよびコンピュータソフトウェア開発等

【日本仮想化技術株式会社について】

会 社 名：日本仮想化技術株式会社

所 在 地：東京都渋谷区

設 立：2006年12月

取 締 役：代表取締役社長兼CEO 宮原 徹

取締役CTO 伊藤 宏通

ホームページ：<http://virtualtech.jp/>



NVMe推奨構成

1. はじめに

NVMeは様々なサーバー製品でサポートされておりますが、3種類の用途の推奨構成と1U/2Uでの最大構成を用意させていただきました。2.5inchのSSDにて構成しています。

パターン1:一般的な1U Serverで高速I/Oを体験! 普通の使い方ならこれで十分!?

- 高速なI/Oを高信頼のサーバーでご用意いたしました。VDIや仮想化、クラウドインフラに最適。
- シンプルに1UサーバーにNVMeのディスクを構成、2.5inch HDDも8本構成可能です。ネットワーク帯域も10Gb以上の構成も可能です。
- x3550M5とNVMeの構成にすることで、400GBから4TBまでのNVMe構成をサポートしています。

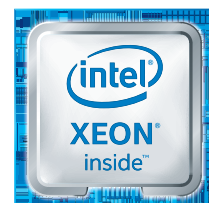
パターン2:DBサーバーやBigData、仮想化インフラ、SDSで大容量のNVMe領域を!

- 大容量のNVMeが必要な方はこちら!DBサーバーや、BigData、高速性の求められる仮想化インフラ、Lenovoが注力するSDS、とにかく高速で大容量のディスクが必要な方にお勧めです。
- x3650M5とNVMe構成で、400TBから16TBまでのNVMe構成をサポートしています。

パターン3:冗長構成をMySQL レプリカで取りましょう。

- 高信頼のDBを中規模のサーバーでレプリケーション構成で構築しましょう。二重書きをすることでデータの保全もばっちり。

参考:1U/2Uサーバーの最大容量



パターン1 一般的なIU Serverで高速I/Oを体験! 普通の使い方ならこれで十分!?

- 高速なI/Oを高信頼のサーバーでご用意いたしました。VDIや仮想化、クラウドインフラに最適。
- シンプルにIUサーバーにNVMeのディスクを構成、2.5inch HDDも8本構成可能です。ネットワーク帯域も10Gb以上の構成も可能です。
- x3550M5とNVMeの構成にすることで、400GBから4TBまでのNVMe構成をサポートしています。



CPU	XeonE5-2650v4(12) 2.20GHz x 2	Network	1Gb x 4 10Gb x 2	NVMe SSD	P3700 400GB x 2
Memory	128GB	OS用HDD	120GB SSD x 2	HDD	2TB 7.2k 12Gbps x 6

8869G2J	System x3550 M5(HS 2.5)/XeonE5-2650v4(12) 2.20GHz-2400MHzx1/PC4-19200 16.0GB(16x1) (Chipkill)/RAID-M5210/RAID-5200-1GF/POW(900Wx1)/OSなし/3年保証24x7(CRU)/SS90	1	
00YE898	XeonE5-2640v4(10) 2.40GHz-2133MHz	1	
46W0829	16GB 2R PC4-19200 RDIMM CK	7	
00NA496	2TB 7.2K 12Gbps NL SAS 2.5型 Gen3 HS HDD	6	
00YC385	120GB SATA 2.5型 MLC HS Gen3 Enterprise Entry SSD	2	
00YA818	P3700 2TB NVMe 2.5型 Gen3 eMLC HS Enterprise Performance PCIe SSD	2	
90Y3901	(FoD) Integrated Management Module(IMM) Advanced UPG	1	
00KA066	PCIe ライザー・カード (2x8 FH/FL + 1x8 FH/HL)	1	
00YL497	追加8HDDキット with エキスパンダー	1	
49Y7960	NVMe対応 PCIe SSDx4 接続キット	1	
00KA098	インテル X520 デュアルポート 10GbE SFP+ アダプター	1	
46M2593	高効率 750W リダンダント電源機構	2	
00FF244	NEMA5-15P to IEC C13 電源ケーブル (2.8m)	1	

パターン2 DBサーバーやBigData、仮想化インフラ、SDSで大容量のNVMe領域を!

- 大容量のNVMeが必要な方はこちら! DBサーバーや、BigData、高速性の求められる仮想化インフラ、Lenovoが注力するSDS、とにかく高速で大容量のディスクが必要な方にお勧めです。
- x3650M5とNVMe構成で、400GBから16TBまでのNVMe構成をサポートしています。



CPU	XeonE5-2650v4(12) 2.20GHz x 2	Network	1Gb x 4 10Gb x 4	NVMe SSD	P3700 800GB x 2
Memory	256GB	OS用HDD	120GB SSD x 2	HDD	2TB 7.2k 12Gbps x 6

8871G2J	System x3650 M5(HS 2.5)/XeonE5-2650v4(12) 2.20GHz-2400MHzx1/PC4-19200 16.0GB(16x1) (Chipkill)/RAID-M5210/RAID-5200-1GF/POW(750Wx1)/OSなし/3年保証24x7(CRU)/SS90	1	
00YJ197	XeonE5-2650v4(12) 2.20GHz-2400MHz	1	
46W0829	16GB 2R PC4-19200 RDIMM CK	15	
00NA496	2TB 7.2K 12Gbps NL SAS 2.5型 Gen3 HS HDD	6	
00YC385	120GB SATA 2.5型 MLC HS Gen3 Enterprise Entry SSD	2	
00YA821	P3700 800GB NVMe 2.5型 Gen3 eMLC HS Enterprise Performance PCIe SSD	2	
90Y3901	(FoD) Integrated Management Module(IMM) Advanced UPG	1	
00KA498	PCIe ライザー・カード (2x8 FH/FL + 1x8 FH/HL)	2	
00FK677	NVMe対応 PCIe SSDx4 接続キット	1	
49Y7960	インテル X520 デュアルポート 10GbE SFP+ アダプター	2	
00FK932	高効率 750W リダンダント電源機構	1	
46M2593	NEMA5-15P to IEC C13 電源ケーブル (2.8m)	2	
00FF244	Windows Server 2012 R2 Standard レノボ版 (1-2CPU/2VM)	1	

パターン3 冗長構成をMySQL レプリカで取りましょう。

- 高信頼のDBを中規模のサーバーでレプリケーション構成で構築しましょう。二重書きをすることでデータの保全もばっちり。



サーバースペック(1台)

CPU	XeonE5-2650v4(12) 2.20GHz x 2	Network	1Gb x 4 10Gb x 4	NVMe SSD	P3700 800GB x 2
Memory	256GB	OS用HDD	120GB SSD x 2	HDD	2TB 7.2k 12Gbps x 6

8871G2J	System x3650 M5(HS 2.5)/XeonE5-2650v4(12) 2.20GHz-2400MHzx1/PC4-19200 16.0GB(16x1) (Chipkill)/RAID-M5210/RAID-5200-1GF/POW(750Wx1)/OSなし/3年保証24x7(CRU)/SS90	2	
00YJ197	XeonE5-2650v4(12) 2.20GHz-2400MHz	2	
46W0829	16GB 2R PC4-19200 RDIMM CK	30	
00NA496	2TB 7.2K 12Gbps NL SAS 2.5型 Gen3 HS HDD	12	
00YC385	120GB SATA 2.5型 MLC HS Gen3 Enterprise Entry SSD	4	
00YA821	P3700 800GB NVMe 2.5型 Gen3 eMLC HS Enterprise Performance PCIe SSD	4	
90Y3901	(FoD) Integrated Management Module(IMM) Advanced UPG	2	
00KA498	PCIe ライザー・カード (2x8 FH/FL + 1x8 FH/HL)	4	
00FK677	NVMe対応 PCIe SSDx4 接続キット	2	
49Y7960	インテル X520 デュアルポート 10GbE SFP+ アダプター	4	
00FK932	高効率 750W リダンダント電源機構	2	
46M2593	NEMA5-15P to IEC C13 電源ケーブル (2.8m)	4	
00FF244	Windows Server 2012 R2 Standard レノボ版 (1-2CPU/2VM)	2	



Intel Inside® 効果的なクラウドを
インテル® Xeon® プロセッサ搭載

参考 NVMe 2本構成1U

- NVMeの最大容量をx3550M5で構成した場合
- ・HDD 8本(16TB)、NVMe 2本(4TB)の構成が可能
- ・10Gb Ethernetを2本用意



CPU	XeonE5-2650v4(12) 2.20GHz x 2	Network	1Gb x 4 10Gb x 4	NVMe SSD	P3700 800GB x 2
Memory	256GB	OS用HDD	120GB SSD x 2	HDD	2TB 7.2k 12Gbps x 6

8871G2J	System x3650 M5(HS 2.5)/XeonE5-2650v4(12) 2.20GHz-2400MHzx1/PC4-19200 16.0GB(16x1) (Chipkill)/RAID-M5210/RAID-5200-1GF/POW(750Wx1)/OSなし/3年保証24x7(CRU)/SS90	1
00YJ197	XeonE5-2650v4(12) 2.20GHz-2400MHz	1
46W0829	16GB 2R PC4-19200 RDIMM CK	15
00NA496	2TB 7.2K 12Gbps NL SAS 2.5型 Gen3 HS HDD	6
00YC385	120GB SATA 2.5型 MLC HS Gen3 Enterprise Entry SSD	2
00YA821	P3700 800GB NVMe 2.5型 Gen3 eMLC HS Enterprise Performance PCIe SSD	2
90Y3901	(FoD) Integrated Management Module(IMM) Advanced UPG	1
00KA498	PCIe ライザー・カード (2x8 FH/FL + 1x8 FH/HL)	2
00FK677	NVMe対応 PCIe SSDx4 接続キット	1
49Y7960	インテル X520 デュアルポート 10GbE SFP+ アダプター	2
00FK932	高効率 750W リダンダント電源機構	1
46M2593	NEMA5-15P to IEC C13 電源ケーブル (2.8m)	2
00FF244	Windows Server 2012 R2 Standard レノバ版 (1-2CPU/2VM)	1

参考 HDD16本搭載/NVMe 4本構成2U

- HDDを最大本数搭載した場合のNVMe構成はこちらです。
- ・OS用SSD 2本、HDD 14本(28TB)、NVMe 4本(8TB)の構成が可能
- ・10Gb Ethernetを4本用意



CPU	XeonE5-2640v4(10) 2.40GHz x2	Network	1Gb x 4 10Gb x 4	NVMe SSD	P3700 2TB x4
Memory	256GB	OS用HDD	120GB SSD x 2	HDD	2TB 7.2k 12Gbps x 14

8871F4J	System x3650 M5(HS 2.5)/XeonE5-2640v4(10) 2.40GHz-2133MHzx1/PC4-19200 16.0GB(16x1) (Chipkill)/RAID-M5210/RAID-5200-1GF/POW(750Wx1)/OSなし/3年保証24x7(CRU)/SS90	1
00YJ199	XeonE5-2640v4(10) 2.40GHz-2133MHz	1
46W0829	16GB 2R PC4-19200 RDIMM CK	15
00NA496	2TB 7.2K 12Gbps NL SAS 2.5型 Gen3 HS HDD	14
00YC385	120GB SATA 2.5型 MLC HS Gen3 Enterprise Entry SSD	2
00YA827	P3700 2TB NVMe 2.5型 Gen3 eMLC HS Enterprise Performance PCIe SSD	4
90Y3901	(FoD) Integrated Management Module(IMM) Advanced UPG	1
00KA498	PCIe ライザー・カード (2x8 FH/FL + 1x8 FH/HL)	2
00FK661	追加8HDDキット with エキスパンダー	1
00FK677	NVMe対応 PCIe SSDx4 接続キット	1
49Y7960	インテル X520 デュアルポート 10GbE SFP+ アダプター	2
00FK932	高効率 750W リダンダント電源機構	1
46M2593	NEMA5-15P to IEC C13 電源ケーブル (2.8m)	2

参考 HDD8本搭載/NVMe 8本構成2U

- NVMeを最大本数搭載した場合の構成はこちらです。
- ・OS用SSD 2本、HDD 6本(12TB)、NVMe 8本(16TB)の構成が可能
- ・10Gb Ethernetを4本用意



CPU	XeonE5-2640v4(10) 2.40GHz x2	Network	1Gb x 4 10Gb x 4	NVMe SSD	P3700 2TB x8
Memory	256GB	OS用HDD	120GB SSD x 2	HDD	2TB 7.2k 12Gbps x 6

8871F4J	System x3650 M5(HS 2.5)/XeonE5-2640v4(10) 2.40GHz-2133MHzx1/PC4-19200 16.0GB(16x1) (Chipkill)/RAID-M5210/RAID-5200-1GF/POW(750Wx1)/OSなし/3年保証24x7(CRU)/SS90	1
00YJ199	XeonE5-2640v4(10) 2.40GHz-2133MHz	1
46W0829	16GB 2R PC4-19200 RDIMM CK	15
00NA496	2TB 7.2K 12Gbps NL SAS 2.5型 Gen3 HS HDD	6
00YC385	120GB SATA 2.5型 MLC HS Gen3 Enterprise Entry SSD	2
00YA827	P3700 2TB NVMe 2.5型 Gen3 eMLC HS Enterprise Performance PCIe SSD	8
90Y3901	(FoD) Integrated Management Module(IMM) Advanced UPG	1
00KA498	PCIe ライザー・カード (2x8 FH/FL + 1x8 FH/HL)	2
00FK677	NVMe対応 PCIe SSDx4 接続キット	2
49Y7960	インテル X520 デュアルポート 10GbE SFP+ アダプター	2
00FK932	高効率 750W リダンダント電源機構	1
46M2593	NEMA5-15P to IEC C13 電源ケーブル (2.8m)	2



Intel Inside® 効果的なクラウドを
インテル® Xeon® プロセッサ搭載

“夢にみたスループットを実現”

レノボの飽くなきチャレンジとイノベーションが、
時代のスタンダードに。



Intel Inside® 効果的なクラウドを
インテル® Xeon® プロセッサ搭載

サーバー・ストレージ関連の最新情報はこちら

▶ <http://www.lenovo.jp.com/server/>

エンタープライズ・ソリューションに驚異的なパフォーマンスと高い汎用性を、
最高レベルの信頼性とセキュリティと共にご提供するラック型2Uサーバー

System x3650 M5



System x3650 M5は、最大2個のインテル® Xeon® プロセッサ E5-2600 v4製品ファミリーと、2ソケット・サーバーでは屈指の大容量ストレージ(最大116TB)のサポート、およびストレージの選択肢により、仮想デスクトップ環境はもとより、クラウドからビッグデータまで、あらゆるワークロードを最適化し、System x Trusted Platform Assurance による最高レベルのセキュリティと信頼性と共に、ビジネスを加速させます。

お電話やメールでのお問い合わせはこちら！

法人のお客様向け見積依頼
・ご購入相談窓口

▶ **0120-68-6200**

✉ hojin_jp@lenovo.com

受付時間：月曜日から金曜日9時から17時30分
(祝日および年末年始休業日を除く)

Lenovo™

レノボ・ジャパン株式会社

〒101-0021

東京都千代田区外神田四丁目14番1号 秋葉原UDX



<http://www.lenovo.jp.com/business/>