# エネルギー効率の良いデータ・センター・ストレージ：ストレージ製品のエネルギー効率の評価に対する SNIA Emerald™ 分析ツールの有用性


【注意】この日本語版資料は原典の理解を助けるための参考文書です。技術上またはビジネス上の判断をするときには、必ず英語版の原典を参照してください。

# エネルギー効率の良いデータ・センター・ストレージ：ストレージ製品のエネルギー効率の評価に対する SNIA Emerald™分析ツールの有用性

2018 年 9 月(日本語版)


編集者：

Jay Dietrich、IBM

執筆者：

Don Goddard：Net App

Rona Newmark：Dell（retired）

Patrick Stanko：SNIA Green Storage Technical Working Group

Herb Tanzer：SNIA Green Storage Technical Working Group

Gary Verdun：Dell

Chuck Paridon：SNIA Green Storage Technical Working Group

Dave Thiel：SNIA Green Storage Initiative

日本語版編集者：

SNIA 日本支部グリーンストレージ委員会



The Green Grid Association (以下、グリーン・グリッド) Emerald Analysis Working group (以下、Emerald 分析作業部会)は、the Storage Networking Industry Association（以下、SNIA）Green Storage Initiative (以下、グリーン・ストレージ分科会または GSI）および Green Storage Technical Working Group (以下、グリーン・ストレージ技術作業部会または GTWG）の Emerald 分析作業部会との緊密な連携と本ペーパーで提供するデータの収集と分析に対する多大なる貢献に対して感謝の意を表する。

SNIA は、グリーン・グリッドより、本ホワイトペーパー全体の複製、翻訳、および公開に対する権限を与えられている。

## SNIA について

SNIA (Storage Networking Industry Association)は、非営利のグローバル団体で、グローバルなストレージ市場をカバーする会員企業で構成されている。SNIA のミッションは、企業/団体が情報管理を推進するために、標準、技術、教育サービスを発展および促進することで全世界のストレージ市場をリードすることである。この目標に向けて、SNIA は、オープンなストレージ・ネットワーキング・ソリューションをより広範な市場に推進させる標準、教育、サービスを提供することを一意的にコミットしている。詳細については、www.snia.org を参照のこと。

## SNIA グリーン・ストレージ分科会（GSI）について

SNIA グリーン・ストレージ分科会（GSI）は、データ・ストレージ運用の環境影響を最小化することに努めることにより、データセンター・ネットワーク・ストレージ技術のエネルギー効率と維持にフォーカスしている。SNIA のグリーン・ストレージ関連活動は ２つの活動母体、グリーン・ストレージ技術作業部会（GTWG）とグリーン・ストレージ分科会（GSI）で行われている。GTWG は、測定されるエネルギー消費と効率を通じて、エンタープライズ・ストレージ・システムのための再現性があり、公平な測定手法と測定基準を策定することにフォーカスしている。GSI は、業界にエネルギー効率の良いストレージ・ネットワーキングのためのベスト・プラクティスを作成し公開すること、ストレージの設置面積と関連する電力要件を削減するためのストレージ・セントリックなアプリケーションを促進すること、そして、測定手法とベスト・プラクティスを法規制団体と評価を行う組織に教育することにフォーカスしている。

## SNIA Emerald™プログラムについて

SNIA Emerald™プログラムは、ストレージ業界、IT コミュニティ、SNIA GSI と関連する法規制団体コミュニティに対するベンダー中立の公的サービスである。このプログラムは、SNIA Emerald™電力効率測定仕様の利用と進化をサポートする。測定手順とテスト測定基準は SNIA Emerald™電力効率測定仕様に記載されており、これは GSI のガイダンスの元でグリーン・ストレージ技術作業部会（GTWG）によって策定、公開、維持されている。

GSI は、測定者が SNIA Emerald™電力効率測定仕様を一貫性を持ち適正に利用できるような教育プログラム資料を作成している。

EPA の ENEGY STAR データセンター・ストレージ・プログラムは、この仕様で定義されている測定手法が元となっており、仕様に基づいて作成された製品測定結果の公開方法を提供している。いくつかの国の法規制団体はニーズに応じて EPA の ENERGY STAR プログラムを相互参照している。それ以外の国々の法規制団体も SNIA Emerald™電力効率測定仕様を認知しており、将来的にはこの測定手法と測定基準を基とするかもしれない。

## 要旨

グリーン・グリッド（世界中の IT リソースおよびデータ・センター・リソースの効率の改善を目的とした業界団体）では、データ・センター・ストレージに関する ENERGY STAR®プログラム要件に照らしてストレージ製品を認定するために、個々の SNIA 会員企業やグリーン・グリッド会員企業などによって生成された大量の SNIA Emerald™電力効率測定仕様ソフトウェアの実行結果から得られた指標データ(以下、SNIA Emerald™測定データと呼ぶ)とシステム構成情報を分析した。グリーン・グリッドの目的は、ストレージ製品のエネルギー効率を評価するためのツールとしての SNIA Emerald™電力効率測定仕様ソフトウェアの有効性を評価し、ストレージ製品の性能/電力効率しきい値の策定および設定におけるその適用および用途を検討することである。

SNIA®は、2013 年 8 月 12 日に SNIA Emerald™電力効率測定仕様ソフトウェアのバージョン 2.0 をリリースし、続けて、2013 年 8 月 28 日に ENERGY STAR データ・センター・ストレージ V1.0 プログラムを発表した。このタイミングにより、製造業者はストレージ製品のシステム構成とコンポーネントの選択が SNIA Emerald™測定データにどう影響するかについて、最低限の理解を得た。グリーン・グリッド Emerald 分析作業部会（Emerald WG）は、48 のマシン・タイプ/モデルと 155 のシステムから SNIA Emerald™測定データと使用可能なアイドル測定データを収集してブラインド化した。
Emerald WG は、このデータを分析して、テスト対象の 3 つのワークロード・タイプ（容量、シーケンシャル、およびトランザクション）の SNIA Emerald™測定データに対するシステムタイプとコンポーネント選択の影響を把握した。本ホワイトペーパーでは、Emerald 分析作業部会の調査結果の詳細を示し、ストレージ製品のエネルギー効率の評価に対する SNIA Emerald™電力効率測定仕様ソフトウェアの強みと限界に関する情報をストレージ製品の製造業者、規制機関、およびステークホルダーに提供する。最終的に、Emerald 分析作業部会は、ストレージ製品の複雑さと、ストレージ・コントローラの機能とストレージ・デバイスのタイプ、回転数、および容量の重要性のために、 SNIA Emerald™測定データは、ストレージ製品の性能/電力機能の有効な指標になり得るが、ストレージ・システム・エネルギー効率の強制的な枠組みや自主的な枠組みのしきい値を設定するために使用すべきではないと結論付けた。

# 目次

**Table of Figures**

**Table of Tables:**

## 1. はじめに

グリーン・グリッドの Emerald 分析作業部会（Emerald WG）は、データ・センター・ストレージ製品に関する ENERGY STAR プログラム要件に照らして認定するストレージ製品の SNIA Emerald™ 電力効率測定仕様ソフトウェアの実行結果から得られた指標データ(以下、SNIA Emerald™測定データと呼ぶ)を分析した。この SNIA Emerald™測定データは、オンライン 2、3、および 4 の製品カテゴリに及ぶ 48 の個別のマシン・タイプ/モデル番号と 155 のシステムからなる（Table 1 に詳細を示す）。分析では、ストレージ製品のエネルギー効率を評価するツールとして SNIA Emerald™電力効率測定仕様ソフトウェアの有効性を評価した。そして、ストレージ製品の性能/電力効率しきい値の策定および設定における適用および用途を検討している。

Emerald 分析作業部会は、個別のストレージ製品の複雑さが SNIA Emerald™測定データを使用した個別のストレージ製品のエネルギー効率の評価と比較を難しくしていると結論付けた。データの分析は、トランザクション SNIA Emerald™テストとシーケンシャル SNIA Emerald™テストによって測定された性能/W 値が次の要因によって大きく異なることを示している。

- コントローラ・アーキテクチャ、運用ソフトウェア、およびその技術世代
- ドライブのタイプ、フォーム・ファクター、回転数（該当する場合）、および台数
- 冗長なコントローラと電源の使用

このペーパーで提供する分析では、トランザクション・テストとシーケンシャル・テストの性能/W 値と容量 GB/W 値をドライブ・タイプ別、フォーム・ファクター別、およびドライブの回転数別に示すが、コントローラ・アーキテクチャとドライブ数に関してグループ内に大きな違いが存在する。トランザクション・テスト値とシーケンシャル・テスト値を使用してストレージ製品のエネルギー効率しきい値を設定するために有効な方法が存在することを示唆する認識可能なパターンはデータ内に存在しない。特に、アイドル状態値はストレージ製品のエネルギー効率の指標にはならない。指標の特性上、最良のアイドル状態値は、容量が最も大きく回転数が最も低いドライブを使用した時に得られる。この指標は消費エネルギーあたりのストレージ容量が増えるような錯覚を与えるが、この値はストレージ製品がトランザクション・データまたはシーケンシャル・データの読み取りと書き込みを実行する能力と、最終的にストレージ製品の主要な目的になる添付のトランザクション・テスト値とシーケンシャル・テスト値に反比例する。

分析は、データの目的と可用性要件を最も良く満たすドライブ・タイプにデータを割り当てる、ストレージ製品の容量最適化手法（COM）（高度なデータ保護、圧縮、データ重複排除、デルタ・スナップショット、シン・プロビジョニング、およびストレージ階層化）の可用性によってさらに複雑化する。COM とストレージ階層化は、個別の製品のエネルギー消費を高めることで、特定のストレージ製品の設置面積あたりのデータ保存量を増やすが、特定のワークロードまたはデータ・ストレージ・タスクを実現するために必要なストレージ製品とストレージ・デバイスの数が減るため、データ・センター全体のエネルギー消費は下がる。ストレージ製品のエネルギー効率は、SNIA Emerald™テストでの単一製品の視点から単独で分析することができないため、製品の個別の機能と設置面積の両方を評価して、データ・センターで必要なデータ・ストレージ・タスクを実行するために要するエネルギー消費に関連付ける必要がある。容量最適化手法 (COM)の機能とメリットについては、本ホワイトペーパーの 6 章で説明する。

最終的に、ストレージ製品のユーザが運用ニーズを満たすためには、ストレージ製品のシステム構成と機能に多様性が必要である。処理されたワークロードと消費されたエネルギーあたりの保存データによって測定されるエネルギー効率は、個別のストレージ製品のエネルギー消費と機能、真に効果的なデータ・ストレージ容量を増やし、アクセス時間を削減するための容量最適化手法 (COM)の展開、およびワークロードの実行に必要なストレージ製品全体の設置面積に左右される。この評価は、データ・センター内の個別のストレージ・デバイスとストレージ製品の最適化と統合を進めるソフトウェア・デファインド・ストレージ機能とデータ・ストレージを拡張および強化する新しい技術の展開が増えているため、さらに複雑化している。個別のストレージ製品の機能と効率を評価することは有益であるが、Emerald 分析作業部会では、 SNIA Emerald™測定データは製品を評価可能なデータとしては最適であるが、エネルギー効率市場アクセスしきい値の設定には使用しないように結論付けている。

## 2. ストレージ・アーキテクチャの概要

**ストレージ・アーキテクチャの違い**

従来のストレージ・システム・アーキテクチャには基本的な 3 つのタイプがある。直接接続ストレージ（DAS）、ストレージ・エリア・ネットワーク（SAN）、およびネットワーク・アタッチド・ストレージ（NAS）である（Figure 1）。これらのアーキテクチャには重複した要素が含まれており、それらについて説明する。

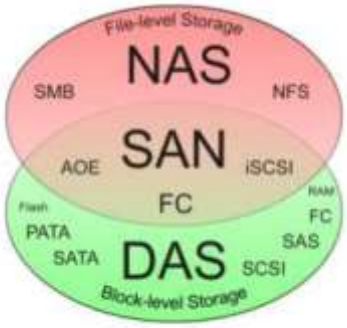



Figure 1: Basic Storage Architectures

DAS は、ブロックレベル・ストレージを提供する最も基本的なストレージ・システムであり、ストレージ・ネットワークを介さずに直接サーバまたはワークステーションに接続される。DAS システムの例を Figure 2 に示す。1 つ目の例では、4 つの Small Computer System Interface（SCSI）ハード・ディスク・ドライブ（HDD）が SCSI ケーブルのデイジー・チェーンを介してホスト・コンピュータに接続されている。2 つ目の例では、Redundant Array of Independent Disks（RAID）/Just a Bunch Of Disks（JBOD）ストレージ・サブシステムにホスト・コンピュータを接続するためにファイバー・チャネル・ケーブルが使用されている。RAID システムは、コントローラを使用してディスクを管理し、ストレージ・システムの信頼性を向上させるのに対して、JBOD システムは、HDD または SSD デバイス・コントローラを使用してワークステーション/サーバと JBOD システム間のデータを管理する。後者のアプローチだけがデータ・センターの実装に関係する。DAS の主な特徴は、個別のコンピュータ/サーバにストレージ・リソースが結びついていることである。これは、安価なストレージ・ソリューションであるが、専用のストレージ・リソースが何らかの制限になる可能性がある。以下に例を示す。

- バスや筐体の設計によってサポートされる HDD の数によってストレージ容量が制限される。
- コンテンツをサーバ間で共有するのではなく複製する必要があるためストレージ・リソースの使用効率が悪く、あるサーバ上の空きストレージ・リソースをディスク領域が不足している別のサーバで使用することができない。
- 接続されたストレージ・リソースにアクセスできなくなるサーバ障害によって可用性が制限される。
- 個別のサーバの処理速度によって性能が制限される。つまり、並列処理によって複数のサーバにワークロードを分散することができない。

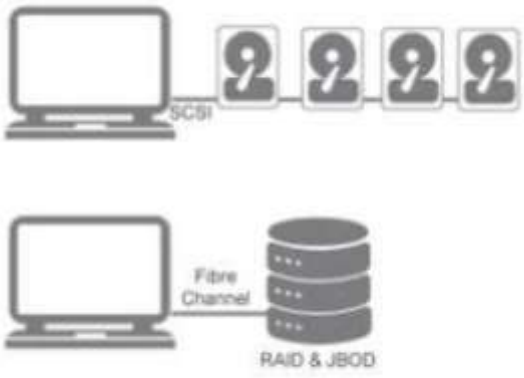



Figure 2: Direct Attach Storage (DAS) Examples

NAS と SAN（Figure 3 と Figure 4 を参照）は、ネットワーク上でストレージを共有するための 2 つの方法であり、DAS に対して本質的メリットを持つ。SAN は、複数のホスト（サーバ）をブロックレベルで単一のストレージ・デバイスに接続することができるため、DAS よりも高度な機能を提供する。通常は、単一ストレージ・ボリュームへの同時接続を許可しないが、あるサーバがボリュームの制御を解放した場合に別のサーバがそのボリュームの制御を引継ぐことは可能である。

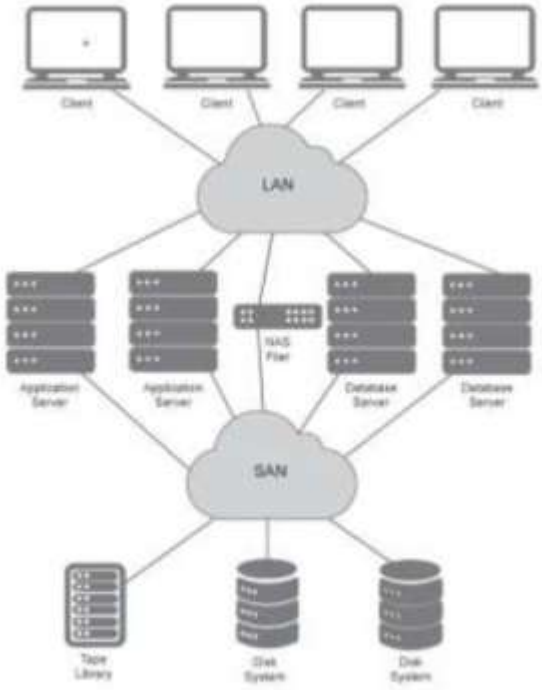



Figure 3: Storage Area Network (SAN)

NAS は、基本的に、SAN または DAS 技術上に構築された「ファイル・サーバ」であり、NFS や CIFS などのネットワーク･ファイル･システム･プロトコルを介してコンピュータ･ネットワーク（LAN）に直接接続する。NAS と SAN の違いは、NAS が「ファイルレベル I/O」であるのに対して、SAN はネットワーク上の「ブロックレベル I/O」である。これは、データがネットワーク経由で受信者に直接ブロックとして転送される（SAN）か、ファイル・データ・ストリームで転送される（NAS）かということであり、NAS の場合、ファイル・アクセス・コマンドが物理ディスク上で一連のブロック・アクセス・コマンドに変換される。ファイル・アクセス・モデルは、より高い抽象化レイヤに構築されるため、追加的な処理レイヤがストレージ・サーバにも、NAS ストレージ・サーバ内でファイル・アクセスとブロック・アクセス間を変換する機能にも必要である。このため、SAN ブロックレベル・アクセスと比較して、多くのアプリケーションで、処理の遅延が大きくなり、I/O スループットに影響が出る。NAS におけるより高いレベルの抽象化がもたらすメリットは使いやすさである。共有ストレージは、NAS システムを使い慣れたエンタープライズ LAN に接続（イーサネット経由）し、ワークステーションとサーバ上の OS を NAS ストレージ・サーバにアクセスするように設定することにより、簡単に実装できる。SAN や DAS と比べた NAS のもう 1 つのメリットは、複数のクライアントで単一のボリュームを共有できることである。SAN ボリュームと DAS ボリュームは、同時に単一のクライアントしかマウントできない。UNIX や LINUX などのオペレーティング・システムには NFS プロトコルのサポートが組み込まれており、Windows OS は CIFS プロトコルをサポートしている。

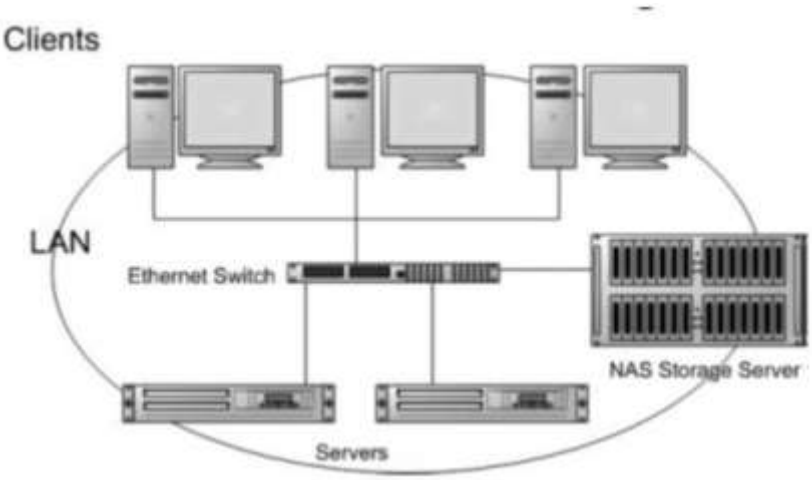



**Figure 4: Network Attached Storage (NAS)**

**ストレージ・システムのスケール方法：スケールアップとスケールアウト**[1]

レガシーまたは従来のシステムは、「スケールアップ」ストレージを使用して容量を増やしている。長年にわたって、標準的なストレージ・アレイは、2 つのストレージ・コントローラ・ヘッド（予備）とメディア装置を収納する複数のディスク・シェルフで構成されてきた。ストレージ・コントローラはストレージ・エリア・ネットワーク（SAN）に接続して、計算サーバにストレージを提供する。すべてのディスク・シェルフがストレージ・コントローラに接続され、すべての計算サーバが 2 つのコン

[1]参照先：http://searchstorage.techtarget.com/blog/Storage-Soup/Scale-out-vs-scale-up-the-basics

トローラを介してこれらのディスクにアクセスする。ストレージ・アレイの容量と性能を高めるには、同じコントローラのペアのディスク・シェルフとディスクを増やす。

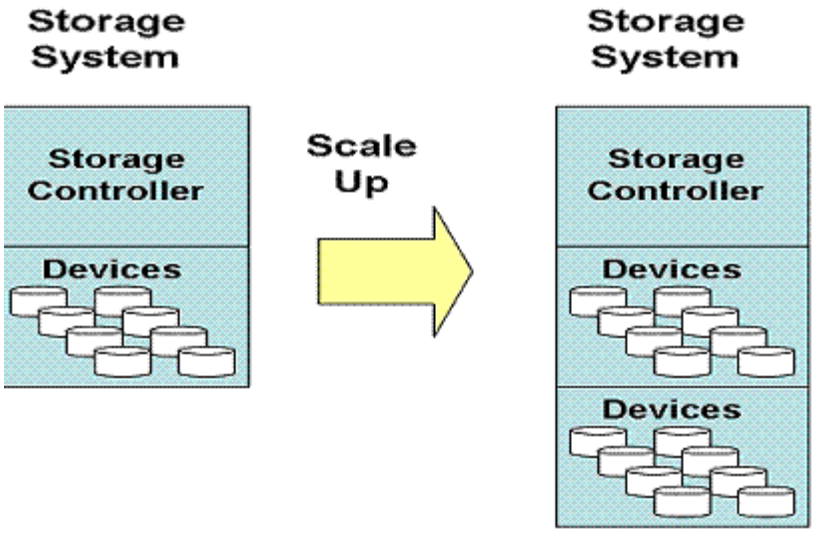



**Figure 5: Scale-up Storage Example**

スケールアップは、ネットワーク接続などのインフラストラクチャー要素を増やすことなく容量問題を解決できる。ただし、ストレージ・デバイスを追加するには、スペース、電力、および冷却を増やす必要がある。スケールアップでは、コントローラ機能が増えないため、追加のホスト活動を処理できない。既存のストレージ・コントローラを使用してストレージ・デバイスが追加されるだけなので容量をスケールするためのコストがかからない。

まず、性能の観点からディスク・スループットが制限要因になることが多いため、ディスクの数を増やせば、通常は、性能が向上する。ただし、ストレージ・アレイに対する負荷が増大（仮想化によってよく引き起こされる状況）し、ディスクの数が増えると、2 つのストレージ・コントローラが互いに RAID 計算とその他のコマンド処理をどんどん CPU に要求し始めるため、それら自体がボトルネックになる。最終的に、コントローラが飽和状態になるまでディスクが追加されると、それ以上性能は上がらない。過負荷状態のコントローラのペアにより速いディスクを追加してもコントローラ上の過負荷状態が進むだけである。このようになる時点を遅らせる 1 つの方法はストレージの計算能力（コントローラの機能）と帯域幅を予め余分に準備しておく方法である。ただし、コストは高くなる可能性がある。性能ピークに到達した場合の次の選択肢は、ストレージ・コントローラの計算機能と帯域幅機能を交換するか、新しいスタンドアロンのストレージ・システムを購入するかである。これらの選択肢は高価であり、管理の負担になることがある。

容量と性能は「スケールアウト」ストレージを使用して拡張できる。通常、スケールアウト・ストレージには、容量と性能の両方を増やすための追加の制御要素とストレージ要素が必要である。スケールアウトと、フロアに追加のストレージ・システムを設置するだけの重要な違いは、スケールアウト・ストレージは単一のシステムとして存在し続けることである。

クラスタ化されたストレージ・システムやグリッド・ストレージ・システムなどのスケールアウトを実現するための方法は複数存在する。Figure 6 に示すスケールアウト・ストレージは、制御機能と

容量の両方が追加されているが、単一システムとしてアクセスできる。このスケーリングには、ストレージをコントローラに接続するためのストレージ・スイッチや、クラスタまたはグリッド内のノード間の接続などの追加のインフラストラクチャーが必要な場合がある。スケーリングアウトでは電力、冷却、およびスペース要件が増え、追加の容量、制御要素、およびインフラストラクチャーのコストが増える。この例のスケールアウト・ソリューションでは、容量が増え、追加の制御機能とともに性能がスケールされる。

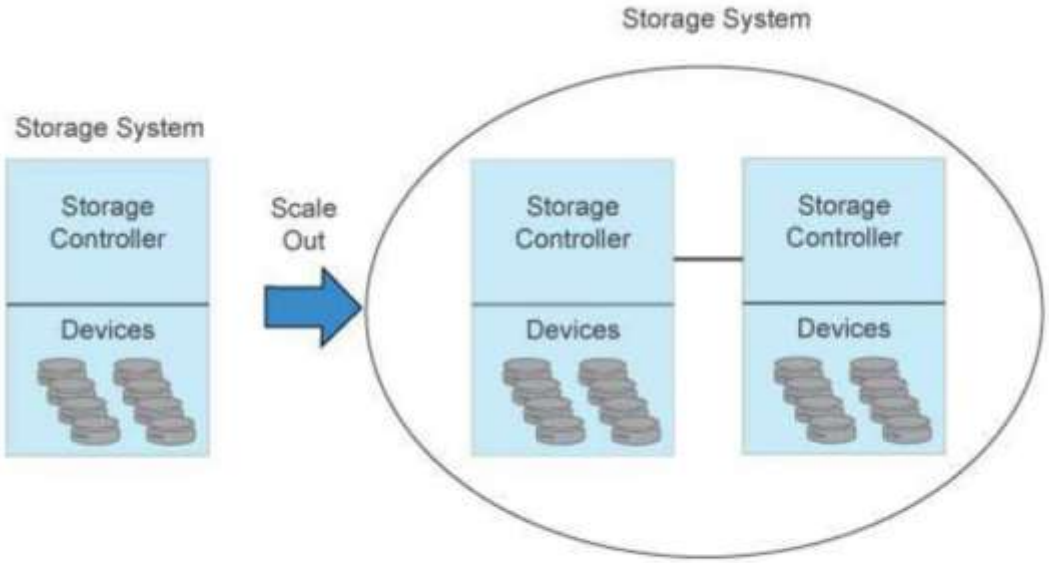



**Figure 6: Scale-out Storage Example**

スケールアウトは必ずしも新しい概念ではないが、ストレージ・ベンダーが旧式のスケールアップ・アーキテクチャの制限を克服するために最新の x86 サーバの能力を活用して新しい構造を作成したため、新しい命が吹き込まれた。このような新しいスケールアウト・ストレージ・システムは、カスタム・コントローラ・プラットフォームではなく、x86 サーバのグループを使用してストレージ・クラスタを形成する。各サーバにはディスクが搭載され、ネットワーク（通常はイーサネット）を使用して、ストレージ・アレイ内の他のサーバと対話する。サーバのグループが協力してクラスタ化されたストレージ・アレイを形成し、従来のアレイと全く同様にネットワーク経由の論理ユニット（LUN）またはファイル共有を提供する。スケールアウト・アレイにディスク・ノードを追加するということは、実際には、新たな x86 サーバをクラスタに追加するということである。それと同時に、ネットワーク・ポート、CPU、および RAM も追加される。スケールアウト・アレイの容量が増えるほど、その性能も増えるため、ノードを追加すると、通常は、性能が線形的に向上する。通常はノードの追加では停止は起きないため、システムを完全に停止することなく、通常のメンテナンス・ウィンドウを使用してアレイの容量と性能を増やすことができる。

**専用（ローカル）ストレージと共有ストレージの比較**

ストレージ・システムは単一のサーバ専用の場合と複数のサーバ間で共有される場合がある。共有ストレージを使用すれば、複数のサーバから同時にまたは別々にストレージにアクセスすることができる。一般的に、共有するように設計されたストレージ・システムは、容量、性能、拡張性、および機能が高く、その代償としてローカル・ストレージ・システムよりコスト高になる。

共有ストレージには、必ず、インターフェイス・カードやネットワーク・スイッチなどのサーバに接続するための余分なハードウェアが必要であり、その分価格が高くなる。共有ストレージは、定期的なチューニングと余分な管理専門知識が必要なほど複雑である。加えて、サーバがストレージを共有する場合は、共有データが正しく更新されるように各サーバのアクションを調整する必要がある。また、サーバのキャッシュ・データに対応するストレージの部分が別のサーバによって変更された時にキャッシュ・データを確実に無効化することにより、サーバのキャッシュを一貫した状態に保つ必要もある。

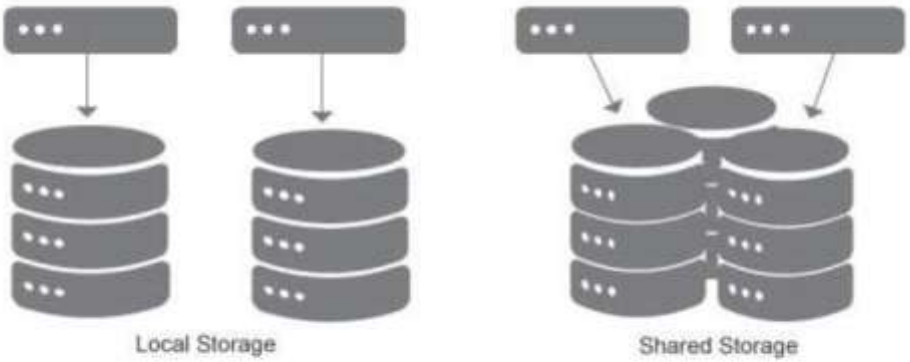



**Figure 7: Local and Shared Storage**

環境やワークロードの要求によっては、ローカル・ストレージで十分な場合と共有ストレージが必要な場合がある。ローカル・ドライブを使用することにより、Input/Output Operations Per second（IOPs）によって測定されるスループットは非常に高くなる可能性があり、ローカル・ストレージのコストは、配備が簡単で管理が単純なため低く抑えられる。ただし、ネットワークに接続された共有ストレージ（SAN/NAS）の拡張性と仮想化の側面は、より高いレベルの冗長性、可用性、および災害復旧性を提供できる。加えて、共有ストレージは、ローカル・ストレージが備えていない様々な追加のソフトウェア機能を提供する。COM[2]（スナップショット、シン・プロビジョニング、圧縮、および重複排除）、自動階層化[3]、データ保護用の追加のフォーム、データのリモート複製などである。

[2] COM（Capacity Optimization Method）は、実際の容量よりも多くのデータを保存できるようにして物理リソースを削減するストレージ・ソフトウェア管理機能である（つまり、使用するディスクと電力を減らす）。

[3]自動ストレージ階層化は、スペース、性能、およびコスト要件を満たすために複数のディスク・タイプと RAID レベル間で情報を動的に移動するストレージ・ソフトウェア管理機能である。

## 3. ストレージの分類(Taxonomy)が必要な理由

データ・ストレージ・システムの基本的な機能は、プライマリ・データ・ストレージ、ミラー・データ・ストレージ、またはバックアップ・データ・ストレージとしてネットワークにストレージ・リソースを提供することである。様々なストレージ・システムが市販されており、異なる性能と機能で様々な目的に使用されている。ストレージ・システムには、JBOD ディスク・アレイ、RAID ディスク・アレイ、テープ・システム、光ストレージ・システムなどがある。このようなデバイス上で提供されるインターフェイスのタイプには、ファイバー・チャネル、SAS、SATA、およびイーサネットが含まれる。

市場では広範囲に及ぶストレージ指向の製品が販売されているため、SNIA Emerald™電力効率測定仕様[4]で分類構造(Taxonomy Structure)が作成された。分類が存在することで、エネルギー効率を評価する場合に、サイズ、容量、性能、高可用性などを考慮した「同様の」製品同士での比較が可能になる。この分類は、下の Figure 8 に示す Online (以下、オンラインと記します)、Near-Online (以下、準オンラインと記します)、Removable Media Library (以下、リムーバブル・メディア・ライブラリ)、Virtual Media Library(以下、仮想メディア・ライブラリ)などのストレージ・カテゴリで構成される。各カテゴリは選択された特徴や機能に基づいて最大 6 つの分類に分割される（Figure 9 にオンライン・カテゴリの中の分類を示す）。

ストレージの分類のもう 1 つの目的は、それぞれ異なるエネルギー効率テスト基準が必要となるストレージ・システムのクラスを明らかにすることである。SNIA Emerald™電力効率測定仕様はオンラインと準オンラインで同じテスト基準を提案しているが、リムーバブル・メディア・ライブラリと仮想メディア・ライブラリではテスト基準が異なる。テスト対象となる製品とそのカテゴリおよび分類を正しく識別し、同等または同様の製品の評価を保証することが有効な測定にとって不可欠である。

| Category / Level | Online (see 5.3) | Near-Online (see 5.4) | Removable Media Library (see 5.5) | Virtual Media Library (see 5.6) | Adjunct Product (see 5.7) | Interconnect Element (see 5.8) |
|---|---|---|---|---|---|---|
| Consumer/ Component [a] | Online 1 | Near-Online 1 | Removable 1 | Virtual 1 | Not defined in this specification | Not defined in this specification |
| Low-end | Online 2 | Near-Online 2 | Removable 2 | Virtual 2 | | |
| Mid-range | Online 3 | Near-Online 3 | Removable 3 | Virtual 3 | | |
| | Online 4 | | | | | |
| High-end | Online 5 | Near-Online 5 | Removable 5 | Virtual 5 | | |
| Mainframe | Online 6 | Near-Online 6 | Removable 6 | Virtual 6 | | |
| [a] Entries in this level of taxonomy include both consumer products and data-center components (e.g., stand-alone tape drives) | | | | | | |

Figure 8: SNIA Emerald Taxonomy Overview

[4] SNIA Emerald™電力効率測定仕様、http://www.sniaemerald.com/download

オンライン・カテゴリは、80 ms 以内にデータ・ブロックの最初のデータを取得可能なストレージ・システムを取り扱う。このようなシステムは、一般的に、ディスク・ベースおよび/またはフラッシュ・ベースである。このカテゴリは、ストレージ・デバイスの台数が少ないコンシューマ/コンポーネント・システムから何百台ものストレージ・デバイスをサポートする大規模システムまでに及び、Figure 9 のように更に分類される。特定のオンライン・カテゴリ内の分類を満たすためには、システムは、分類表に記載された最小ストレージ・デバイス数以上の最大構成をサポートする必要がある。例えば、オンライン 3 に分類されるストレージ製品は少なくとも 12 台のストレージ・デバイスをサポートできる必要がある。ただし、ストレージ製品は、それよりも少ない台数のストレージ・デバイスとともに販売されている場合がある。

準オンライン・カテゴリのストレージ・システムは、最初のデータまで 80 ms という時間要件を満たすことはできないが、それ以外はオンライン・カテゴリと同様である。

リムーバブル・メディア・ライブラリ・カテゴリは、テープ・ライブラリや光ジューク・ボックス向けであり、このようなシステムは、最大 5 分の最初のデータまでの最大時間が必要であり、ストリーミング IO 要求だけをサポートする。仮想メディア・ライブラリ・カテゴリは、最初のデータまで 80 ms という時間要件を満たすカテゴリであり、このようなシステムはシーケンシャル I/O 要求用に設計されたディスク・ベースのシステムであることが多い。

オンライン 2、3、または 4 のストレージの分類カテゴリだけが ENERGY STAR データ・センター・ストレージ V1.0 認定の資格がある。これらが、今日のデータ・センターにおいて、保有ストレージ・システムとエネルギー消費の多くを占めている。ENERGY STAR データ・センター・ストレージ V1.0 認定テストでは、ブロック IO 用に設定できる必要がある。オンライン 1 はコンシューマ市場向けであり、オンライン 5 と 6 は装置製造業者が少なく売上数量も少ないハイエンド超大規模ストレージ・システムに相当する。

| Attribute | Classification | | | | | |
|---|---|---|---|---|---|---|
| | Online 1 | Online 2 | Online 3 | Online 4 | Online 5 | Online 6 |
| Access Pattern | Random/ Sequential | Random/ Sequential | Random/ Sequential | Random/ Sequential | Random/ Sequential | Random/ Sequential |
| MaxTTFD (t) | t < 80 ms | t < 80 ms | t < 80 ms | t < 80 ms | t < 80 ms | t < 80 ms |
| User-Accessible Data | Required | Required | Required | Required | Required | Required |
| Connectivity | Not specified | Connected to single or multiple hosts | Network-connected | Network-connected | Network-connected | Network-connected |
| Consumer/ Component | Yes | No | No | No | No | No |
| Integrated Storage Controller | Optional | Optional | Required | Required | Required | Required |
| Storage Protection | Optional | Optional | Required | Required | Required | Required |
| No SPOF | Optional | Optional | Optional | Required | Required | Required |
| Non-Disruptive Serviceability | Optional | Optional | Optional | Optional | Required | Required |
| FBA/CKD Support | Optional | Optional | Optional | Optional | Optional | Required |
| Maximum Supported Configuration | ≥1 | ≥ 4 | ≥ 12 | > 100 | >400 | >400 |

**Figure 9: SNIA Emerald Online Classifications**

データ・センターの運用では、一般的に、ファイル・アクセス・ベースのネットワーク接続ストレージ（NAS）タイプのオンライン・ストレージが大量に利用される。この関連で、ファイル・アクセス拡張が SNIA Emerald™電力効率測定仕様 V3.0.1 に追加され、2018 年初旬には ENERGY STAR によって採用され、適格なデータ・センター・ストレージ・カテゴリになることが期待されている。

加えて、ストレージ業界は、新しいタイプのストレージ・アーキテクチャと製品を急速に進化させ、市場に投入している。SNIA Emerald™分類の改訂版が策定中であり、将来のストレージ商品を包含するように現在の分類を拡張している。対象となる製品領域には、不揮発性（NV）ソリッド・ステート、オール・フラッシュ・アレイ、オブジェクト・ストア、ビッグ・エンタープライズ仮想化ストレージ・サーバ、およびコンバージド、ハイパーコンバージド、コンポーザブル、およびソフトウェア・デファインド・ストレージが含まれる。ストレージ製品とサーバ製品との境界の定義を正確に記述するのが困難になっており、ソフトウェア・デファインドの実装が注目されていることから、これらを解決するためには、業界の IT パートナー全体での共同の分類の取り組みが必要である。

## 4. ストレージの物理属性

ストレージ・システムは数種類のメディア、コントローラ、および筐体で構成される。メディア装置には、ハード・ディスク・ドライブ（HDD）、ソリッド・ステート・ディスク・ドライブ（SSD）（PCIe や NVMe のバス方式も含む）、非ドライブ形式のソリッド・ステート・メディア、テープ、および光メディアなどがある。光メディアは、エンタープライズ・アプリケーションには適さないし、テープは、適格な ENERGY STAR カテゴリではないためここでは扱われない。ストレージ・アレイ・コントローラまたはプロセッサは、物理ディスクを管理し、ハードウェア RAID を実装し、フロントエンドのコンピュータ HBA（ホスト・バス・アダプタ）とバックエンドで制御されるディスクにインターフェイスを提供する。ドライブ、コントローラ、ミッドプレーン、電源、および冷却ファンが物理筐体に内蔵されている。このような筐体は、単一の統合筐体にすることも、コントローラ・アレイ筐体とドライブ筐体に分けることもできる。単一障害点（SPOF）がないという要件を満たすために、ストレージ・コントローラは冗長なハードウェア・モジュールを備えることになる。ディスクが内蔵された一般的なストレージ・アレイ筐体を Figure 10 に示す。

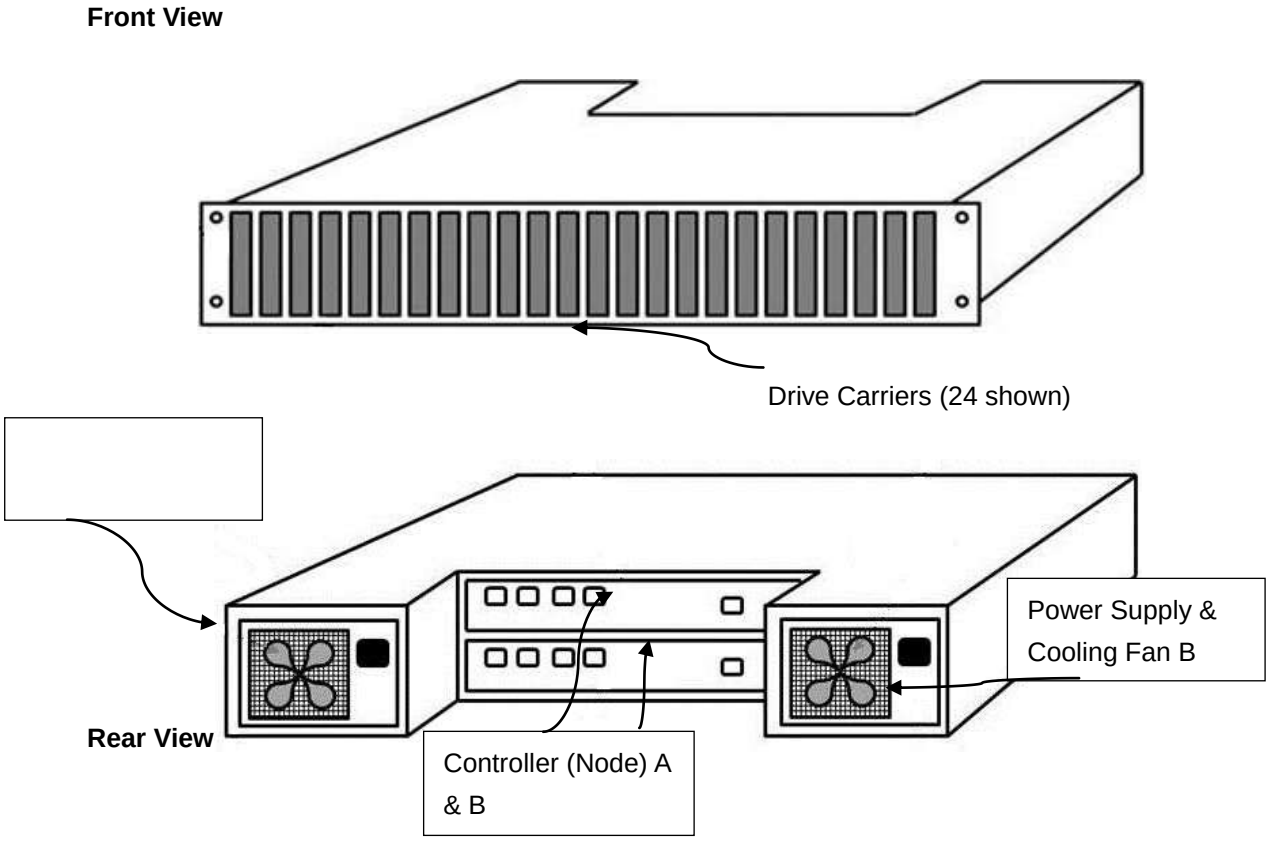



Figure 1: Typical Storage Array Enclosure: Entry Level, Expand Out

ストレージ・アレイ・システムのエネルギー消費は、大まかに、ストレージ・メディアの特性（タイプ（SSD または HDD）、個別のデバイスの容量（GB または TB）と電力容量、およびデバイス数）によって決定される。その他の要因は、RAID レベル、信頼性、可用性、および保守容易性（RAS）機能の選択とアレイ・コントローラの性能である。一般的に、ストレージ・システムの分類が高いほど（ローエンドではなくハイエンド）、追加の機能をサポートするために必要な電力レベルが高くなる。

エンタープライズ・ストレージ製品に使用されているハード・ドライブの中では、7.2Krpm 2.5”ドライブの平均電力需要は約 6W/ドライブで、10Krpm と 15Krpm のより高い回転数ではそれぞれ 7.5W と 8.3W と電力需要が少し高くなる。3.5”ドライブの平均電力需要は 7.2Krpm の回転数で約 10W で、15Krpm の回転数では 14W/ドライブを超える。シングルレベル・セル（SLC）でもマルチレベル・セル（MLC）でも、ソリッド・ステート・ドライブの電力需要は 2.5”ハード・ドライブとほとんど同じである。PCIe ベースのドライブの電力需要は他のすべてのドライブを大きく上回る約 23W である。

ストレージ・ベンダーから電力計算ツールを入手できる。これを使用すれば、すべてのコンポーネントの電力消費を決定し、様々なシステム構成の合計を計算することができる。コントローラ・モジュールの電力は性能によって約 50W～300W の間で異なり、ドライブ筐体 IO モジュールの電力は 20W～30W 程度である。24 台のスモール・フォーム・ファクター（SFF）ハード・ディスク・ドライブ (HDD)を収容する標準的な 2U ドライブ筐体の平均電力は約 250W である。加えて、4U あたり最大で 84 台のラージ・フォーム・ファクター（LFF）ドライブまたは 3U あたり 120 台の SFF ドライブを収容可能な高密度 JBOD 筐体がある。これらの筐体 1 つの電力は 1500W 程度になることがある。標準的なドライブ筐体が 3～4 台を超える大規模なストレージシステムは、ドライブの電力が大部分を占める。

大規模なストレージシステムにおけるドライブのエネルギー消費は、簡単に合計の 60%以上になる。加えて、ストレージ・システム・アイドル電力（アクティブ IO なし、バックグラウンド・アクティビティは存在する場合がある）はフル稼働電力の公称 80%になる。

## 5. ストレージに関する SNIA Emerald™データが示すもの

データ分析に使用された SNIA Emerald™測定データは、2016 年 11 月にまとめられ、2017 年 1 月の Data Center Storage Stakeholders meeting (データ・センター・ストレージ・ステークホルダー会議)で提出された。このデータには 155 システムが含まれ、その内訳は、オンライン 2 システムが 35 システム、オンライン 3 システムが 58 システム、およびオンライン 4 システムが 62 システムである。データはさらに、それぞれのシステムが最適化されたワークロード・タイプ（トランザクションまたはシーケンシャル）に区分できる（Table 1）。

| On-line Category | Number of Families | Number of Configurations | Number of Manufacturers |
|---|---|---|---|
| 2 Transactional | 3 | 11 | 2 |
| 2 Sequential | 5 | 24 | 2 |
| 3 Transactional | 16 | 48 | 5 |
| 3 Sequential | 4 | 10 | 2 |
| 4Transactional | 10 | 35 | 5 |
| 4 Sequential | 10 | 27 | 5 |

**Table 1: Breakout of the ENERGY STAR configuration data presented in this paper**

ストレージ・システムの性能は、コントローラ、コントローラ・キャッシュ、フロントエンドとバックエンドの相互接続、ストレージ・メディアの容量、およびストレージ・メディアの回転数（該当する場合）に依存する。次の節では、トランザクション・ワークロードの指標データと、分類、ドライブのタイプと容量、およびドライブの台数が値に与える影響について見ていく。

**トランザクション・ワークロード用に最適化されたシステムからのデータ**

Figure 11 は、ハード・ディスク・ドライブ（HDD）を使用するシステムのトランザクション・ワークロードである Hot Band 指標(Hot Band テストで測定される IOPS/W 値)と Ready Idle 指標(Ready Idle テストで測定される容量[GB]/W 値)の比較を示している。このグラフは、システムが性能と容量の間で行うトレードオフを示している。7.2K rpm ドライブを使用するシステムは、通常、より高回転のドライブと比較して、Ready Idle 指標の値は高いが、Hot Band 指標の値は低い。一般的に、オンライン 3 システムとオンライン 4 システムは、すべてのディスク・タイプのプロット点が広く分布している。

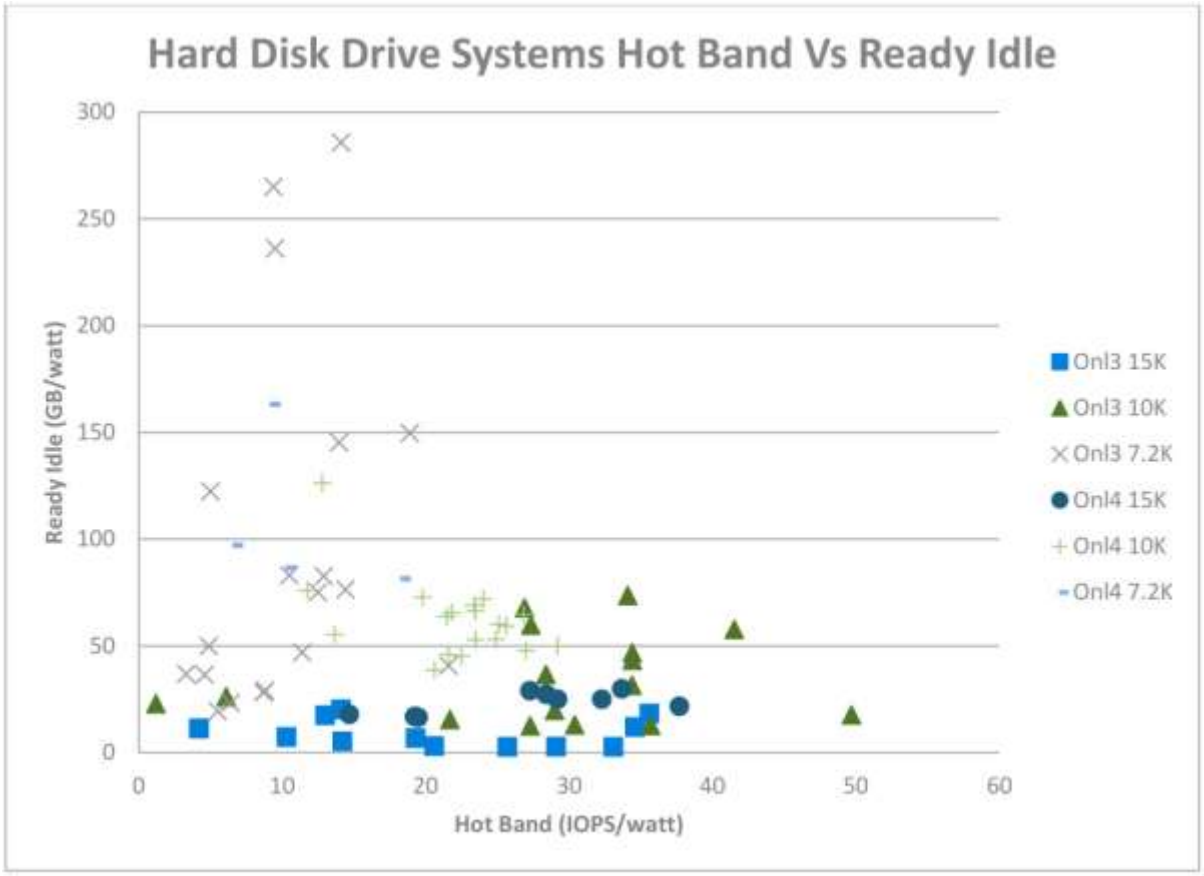



**Figure 11: Hard Drive Systems; Hot Band vs. Ready Idle**

Figure 12 にはオール・フラッシュ・アレイ（AFA）システムが含まれている。一般的に、AFA システムは HDD を搭載したシステムより Hot Band 指標値が高い。AFA システムの Ready Idle 指標値は他のデータセットと同様である。

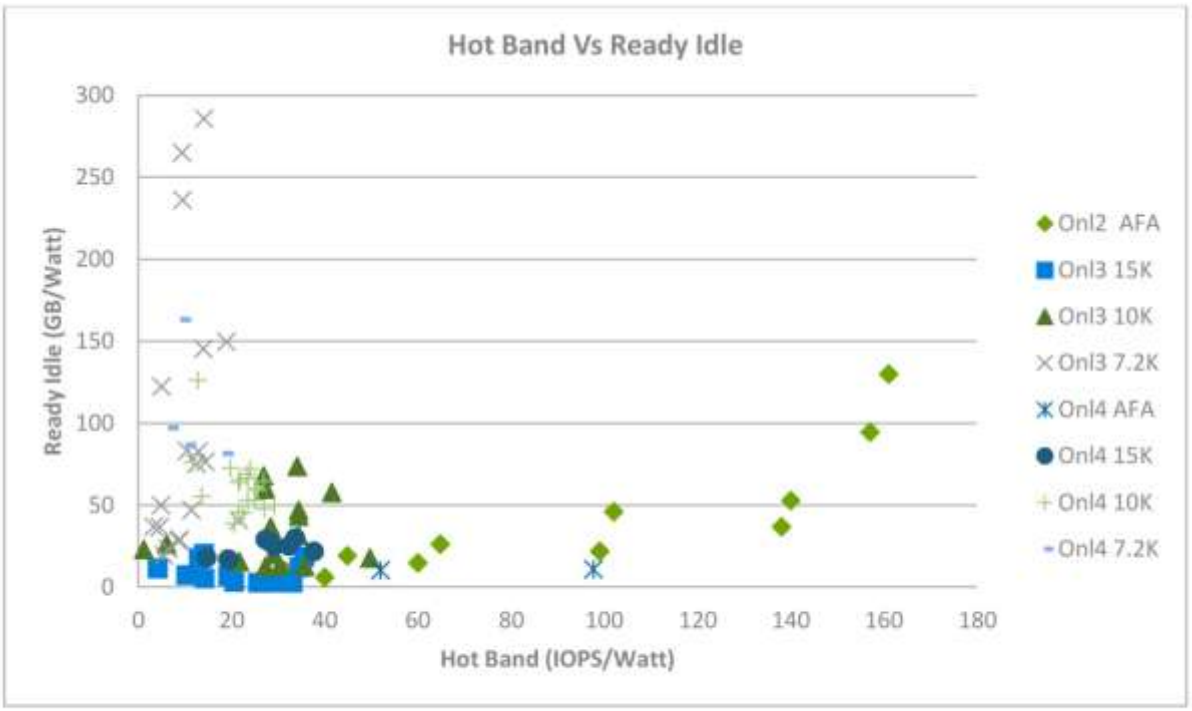



**Figure 12: Hard Drive and All Flash Array Systems; Hot Band vs. Ready Idle**

Figure 13、Figure 14、Figure 15、および Figure 16 は、オンライン・カテゴリ別にシステムの Hot Band 指標、Random Read 指標(Random Read テストで測定される IOPS/W 値)、Random Write 指標(Random Write テストで測定される IOPS/W 値)、Ready Idle 指標とドライブ数を示している。オンライン 2 システムとオンライン 3 システムは、一般的に、コントローラが 1 つの小規模なシステムである。オンライン 2 システムのほとんどがオール・フラッシュ・アレイである。オンライン 4 システムは、デュアル・コントローラと電源を備え、ストレージ・デバイス数が大幅に多い。すべてのオンライン・カテゴリで指標値が広く分布している。

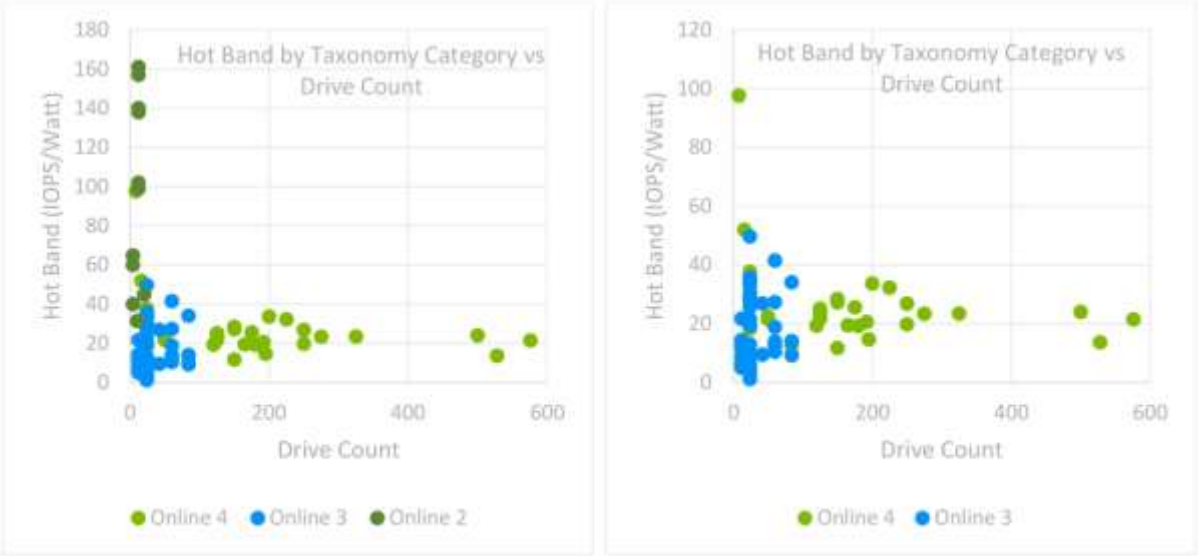



**Figure 13: Hot Band Result vs. Drive Count With and Without Online 2 AFA systems**


Figure 14: Random Read vs. Drive Count with and without Online 2 AFA Systems

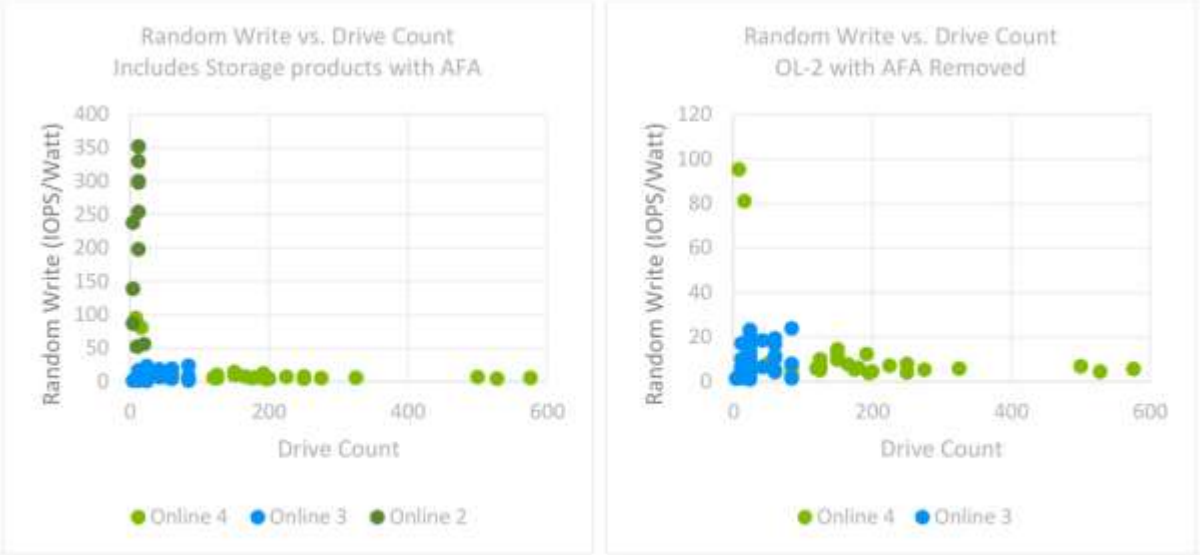



**Figure 15: Random Write vs. Drive Count with and without Online 2 AFA systems**

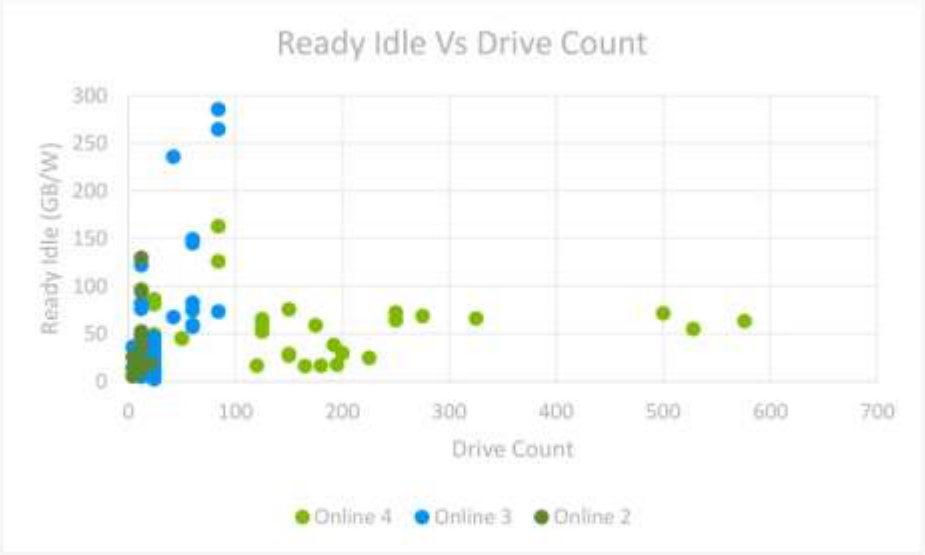



**Figure 16: Transactional Systems Ready Idle vs. Drive Count**

Figure 13、Figure 14、Figure 15、および Figure 16 は、オンライン・カテゴリ別に指標とドライブ数の関係を示している。前述したように、オンライン 2 システムは、一般的に小規模で、ほとんどが AFA である。AFA システムの Hot Band 指標、Random Read 指標、Random Write 指標は、通常、HDD システムのそれより高いが、Ready Idle 指標値は同様である。オンライン 3 システムは、通常、オンライン 4 システムと比較して、ストレージ・デバイス数が大幅に少ない。オンライン 3 システムは、単一のコントローラで構成されることが多いが、オンライン 4 システムは、デュアル・コントローラを備えているため、オンライン 3 システムは電力オーバーヘッドが低く Ready Idle 指標値が優れている。Figure 13、14、15 のオンライン 4 カテゴリの 2 つの高いポイントは、AFA システムである。

Figure 17、Figure 18、および Figure 19 は、オンライン 3 システムとオンライン 4 システムを別々に調査し、結果をドライブ・タイプでさらに分類したものである。

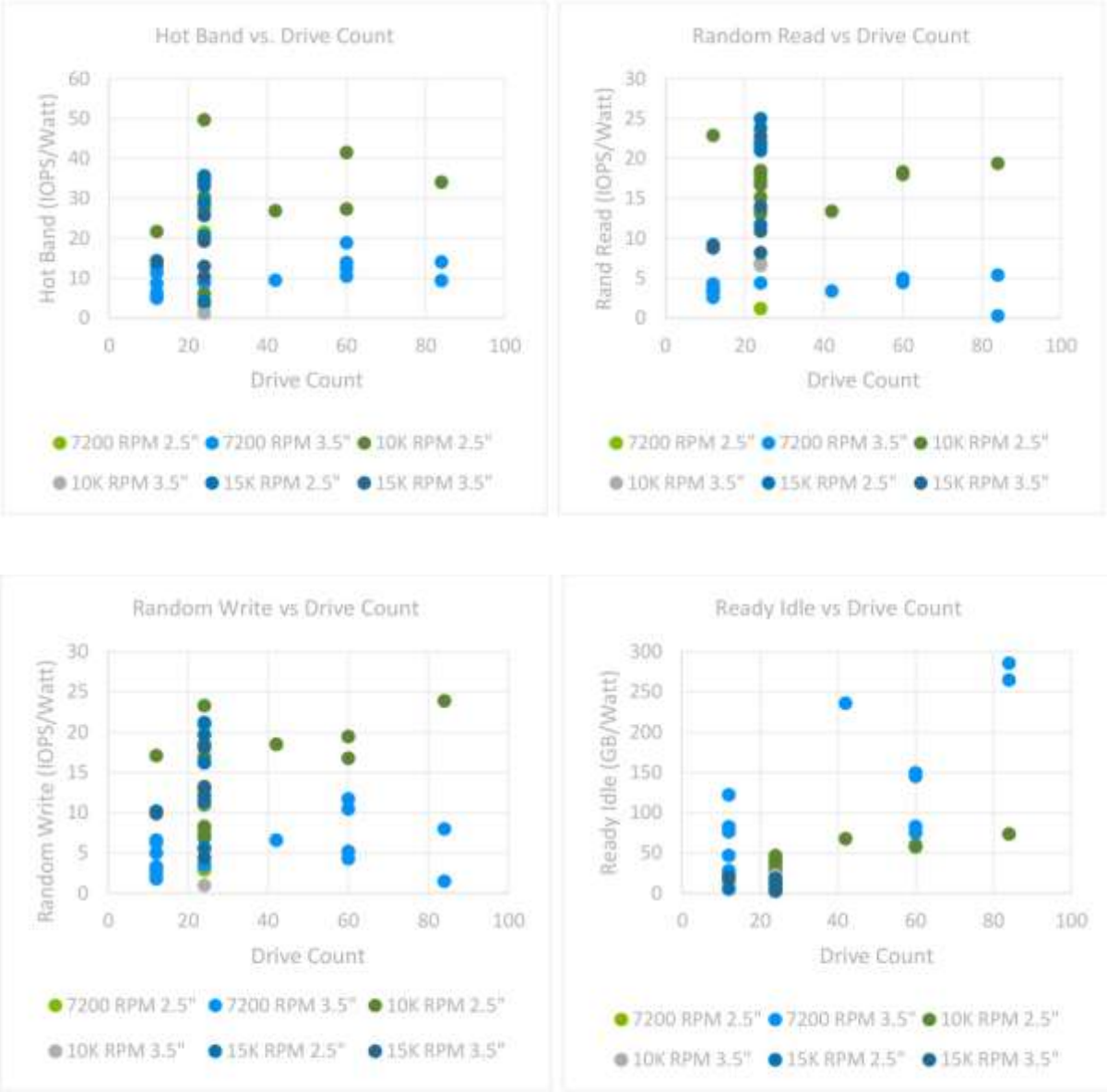



Figure 17: Online 3 Transactional Metrics vs. Drive Count Broken Out by Drive Type

オンライン 3 トランザクションのデータは、ドライブ数とタイプ別で広く分布している。全体として、10Krpm ドライブと 15Krpm ドライブは 7.2Krpm ドライブより良好な性能指標(Hot Band 指標、Random Read 指標、Random Write 指標をまとめてこう呼ぶ。以下同じ)を示し、ドライブ数が多いシステムは小規模なシステムより良好な性能指標を示している。ドライブ数が少ないシステムは、コントローラ電力の 1 ドライブ当たりの負担が大きく、データ経路の一部しか利用されないために指標値は低くなる傾向がある。ドライブの数とタイプが同じシステムでも、コントローラ・アーキテクチャと技術世代の違いや冗長な電源を使用しているか単一の電源を使用しているかによって指標値が異なる。


**Figure 18: Online 4 Transactional Metrics vs. Drive Count including All Flash Arrays**

Figure 18 は、オール・フラッシュ・アレイ・データを含むオンライン 4 の性能指標、Ready Idle 指標とドライブ数の関係を示している。オール・フラッシュ・アレイ製品は、他のほとんどのオンライン 4 システムと比較してストレージ・デバイス数が非常に少ないため、アイドル状態値が大幅に低くなる。コントローラ電力をより少ない数のドライブで負担しなければならないという課題があるにも関わらず、オール・フラッシュ・アレイの Random Write 指標は HDD の値より 8〜10 倍高い。

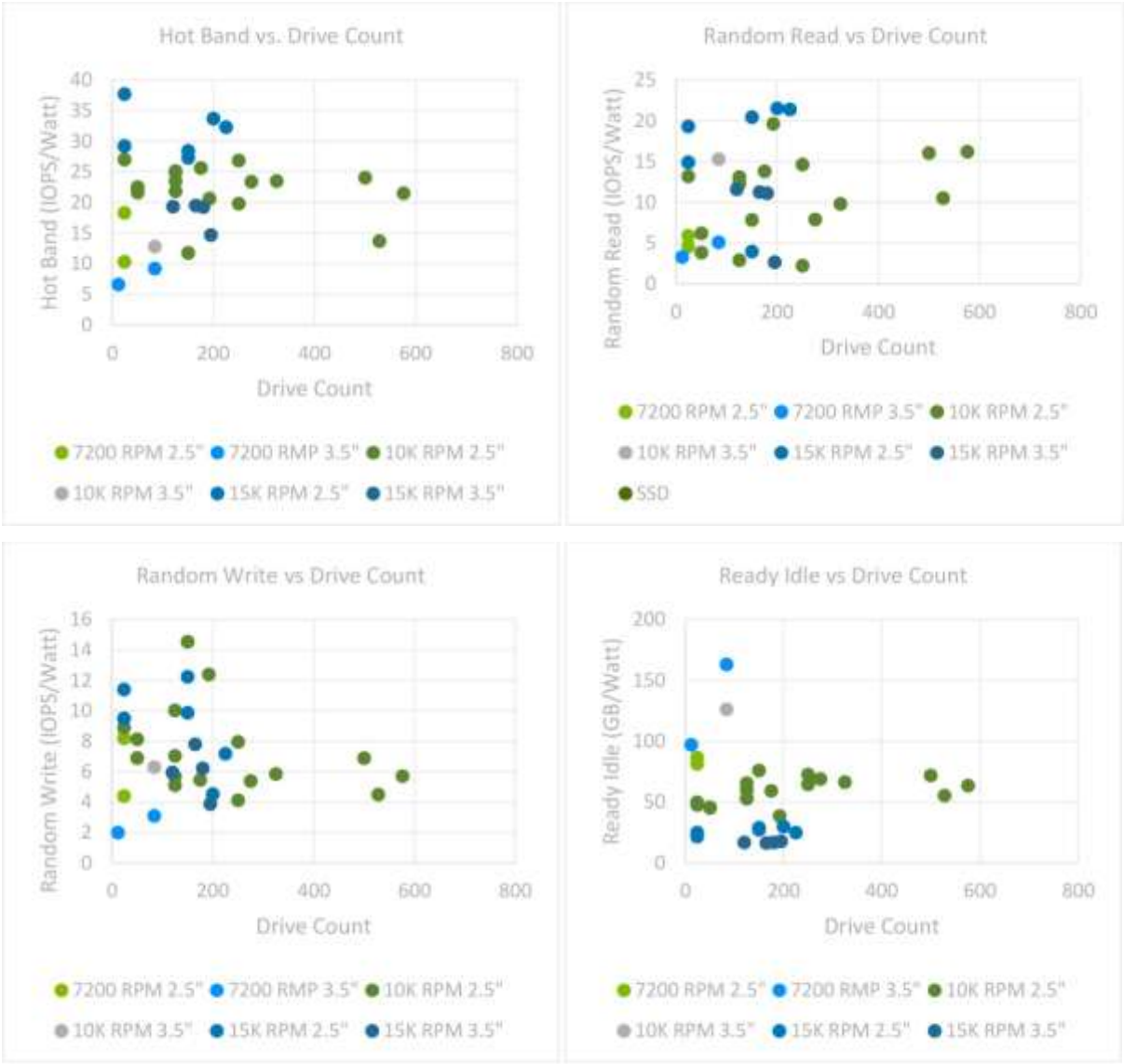



**Figure 19: Online 4 Transactional Metrics vs. Drive Count without All Flash Array Data**

Figure 19 は、ハード・ドライブのみのオンライン 4 の性能指標と Ready Idle 指標とドライブ数の関係を示している。オンライン 4 はオンライン 3 と同じ傾向を示しており、より高回転のドライブはより高い性能/電力指標を示し、より低回転のドライブはより高い容量/電力指標を示す。オンライン 4 性能/電力指標も、同様にドライブのタイプと台数別で広く分布している。Hot Band 指標は、Random Read 指標や Random Write 指標より分散が少ない。これは、オンライン 4 システム・アーキテクチャは、一般的に、キャッシュの容量が大きく、デュアル・コントローラを備えているため Random Read/Write 指標や Sequential Read/Write 指標より Hot Band 指標が良好になりやすいためである。

以下では、同一の製品であるが、搭載ドライブ数などの構成が異なるものを総称して製品ファミリーまたは単にファミリーと呼ぶことにする。

Table 2 は、オンライン 3 での指標を示す。各システムのラベルには、割り当てられたファミリー番号/ドライブ・フォーム・ファクター/ドライブの回転数/ドライブ容量/ドライブ数が記載されている。各ドライブの回転数は色分けされ、各指標の上位 25%が緑色で強調表示されている。

| Label | Hot Band (IOPS/W) | Random Read (IOPS/W) | Random Write (IOPS/W) | Ready Idle (GB/W) |
|---|---|---|---|---|
| FAM15 2.5 7200k 0GB #Dr0 | 21.6 | 6.8 | 7.8 | 40.7 |
| FAM16 2.5 7200k 0GB #Dr0 | 3.3 | 1.2 | 2.9 | 36.8 |
| FAM12 2.5 7200k 1000GB #Dr12 | 5.5 | 3.2 | 2.2 | 19.4 |
| FAM40 3.5 7200k 64GB #Dr4 | 4.6 | 3.97 | 1.4 | 36.49 |
| FAM40 3.5 7200k 384GB #Dr12 | 4.88 | 4.49 | 1.48 | 49.9 |
| FAM11 3.5 7200k 1000GB #Dr12 | 6.4 | 3.5 | 2.8 | 23.3 |
| FAM14 3.5 7200k 1000GB #Dr12 | 8.7 | 4.3 | 3.3 | 28.1 |
| FAM15 3.5 7200k 1000GB #Dr12 | 11.4 | 3.9 | 5 | 46.9 |
| FAM13 3.5 7200k 1000GB #Dr24 | 8.8 | 4.4 | 3.5 | 29.1 |
| FAM15 3.5 7200k 1000GB #Dr60 | 10.5 | 5 | 5.2 | 83 |
| FAM15 3.5 7200k 1000GB #Dr60 | 12.5 | 5 | 4.3 | 75.1 |
| FAM25 3.5 7200k 2000GB #Dr12 | 12.91 | 3.45 | 6.6 | 82.61 |
| FAM28 3.5 7200k 2000GB #Dr12 | 14.42 | 3.14 | 6.3 | 76.38 |
| FAM27 3.5 7200k 2000GB #Dr60 | 18.9 | 4.48 | 10.47 | 149.6 |
| FAM29 3.5 7200k 2000GB #Dr60 | 13.97 | 4.6 | 11.72 | 145.25 |
| FAM17 3.5 7200k 4000GB #Dr12 | 5 | 2.6 | 1.8 | 122.3 |
| FAM33 3.5 7200k 4000GB #Dr42 | 9.5 | 3.4 | 6.6 | 236.1 |
| FAM33 3.5 7200k 4000GB #Dr84 | 14.1 | 5.4 | 8 | 285.7 |
| FAM33 3.5 7200k 800;4000GB #Dr70 | 9.4 | 0.3 | 1.5 | 264.9 |
| FAM15 2.5 10000k 0GB #Dr0 | 6.1 | 16.9 | 11 | 26.2 |
| FAM11 2.5 10000k 600GB #Dr12 | 21.7 | 22.9 | 17.1 | 15.5 |
| FAM12 2.5 10000k 600GB #Dr24 | 27.3 | 17.4 | 21 | 12.5 |
| FAM13 2.5 10000k 600GB #Dr24 | 30.4 | 18.1 | 23.3 | 13 |
| FAM14 2.5 10000k 600GB #Dr24 | 35.7 | 16.7 | 18.5 | 12.7 |
| FAM26 2.5 10000k 600GB #Dr24 | 34.41 | 15.15 | 13.04 | 46.74 |
| FAM30 2.5 10000k 600GB #Dr24 | 34.42 | 13.51 | 13.26 | 43.3 |
| FAM27 2.5 10000k 600GB #Dr60 | 27.34 | 18.32 | 16.77 | 59.71 |
| FAM29 2.5 10000k 600GB #Dr60 | 41.53 | 18.01 | 19.46 | 57.71 |
| FAM33 2.5 10000k 600GB #Dr84 | 34.1 | 19.4 | 23.9 | 73.6 |
| FAM33 2.5 10000k 900GB #Dr42 | 26.9 | 13.4 | 18.5 | 67.8 |
| FAM17 2.5 10000k 1200GB #Dr24 | 28.4 | 13.1 | 16.9 | 36.6 |
| FAM13 2.5 10000k 400;600GB #Dr17 | 29 | 21.8 | 6.9 | 19.8 |
| FAM14 2.5 10000k 400;600GB #Dr17 | 49.7 | 21.1 | 7.2 | 17.7 |
| FAM17 2.5 10000k 400;600GB #Dr17 | 34.4 | 18.5 | 8.3 | 31.5 |
| FAM16 3.5 10000k 0GB #Dr0 | 1.2 | 6.6 | 1 | 22.9 |
| FAM11 2.5 15000k 146GB #Dr24 | 20.6 | 25 | 18.1 | 3.1 |

| Label | Hot Band (IOPS/W) | Random Read (IOPS/W) | Random Write (IOPS/W) | Ready Idle (GB/W) |
|---|---|---|---|---|
| FAM13 2.5 15000k 146GB #Dr24 | 29.1 | 23.8 | 19.7 | 2.8 |
| FAM14 2.5 15000k 146GB #Dr24 | 33.1 | 22 | 16.2 | 2.7 |
| FAM15 2.5 15000k 146GB #Dr24 | 34.6 | 21.7 | 11.5 | 12 |
| FAM16 2.5 15000k 300GB #Dr24 | 4.2 | 11.7 | 5.6 | 11.3 |
| FAM16 2.5 15000k 300GB #Dr24 | 4.2 | 11.7 | 5.6 | 11.3 |
| FAM17 2.5 15000k 300GB #Dr24 | 35.6 | 21 | 21.2 | 18.3 |
| FAM12 2.5 15000k 600GB #Dr12 | 14.2 | 9.2 | 10.2 | 5.3 |
| FAM17 3.5 15000k 600GB #Dr12 | 14.1 | 8.8 | 9.9 | 20.3 |
| FAM13 3.5 15000k 600GB #Dr24 | 10.3 | 14 | 13.1 | 7.3 |
| FAM14 3.5 15000k 600GB #Dr24 | 19.3 | 10.9 | 12.2 | 6.8 |
| FAM12 3.5 15000k 600GB #Dr24 | 25.7 | 22.8 | 18.2 | 2.7 |
| FAM43 3.5 15000k 600GB #Dr24 | 13 | 8.2 | 4.4 | 17.5 |

**Table 2: Online 3 Transactional Data Highlighting the Top 25% for Each Metric/Workload**

Table 2 は、オンライン 3 のグラフと同じ傾向を示している。高回転ドライブを搭載したシステムが性能指標の上位 25%に入る傾向があるのに対して、低速大容量ドライブを搭載したシステムが Ready Idle 指標の上位 25%に入る傾向がある。すべての指標の上位 25%に入っているのは 1 つのシステムだけである。このシステムは、データセット内でドライブ数が最も多いシステムである。

Table 3 は、オンライン 4 指標を示す。ここでもドライブの回転数は色分けされ、各指標の上位 25%が緑色で強調表示されている。

| Label | Hot Band Workload Test (IOPS/W) | Random Read Workload Test (IOPS/W) | Random Write Workload Test (IOPS/W) | Ready Idle Workload Test (GB/W) |
|---|---|---|---|---|
| FAM34 2.5 7200k 1000GB #Dr24 | 10.3 | 4.6 | 4.4 | 86.5 |
| FAM37 2.5 7200k 1000GB #Dr24 | 18.3 | 5.9 | 8.2 | 81.4 |
| FAM34 3.5 7200k 2000GB #Dr12 | 6.6 | 3.3 | 2 | 97.1 |
| FAM35 3.5 7200k 2000GB #Dr84 | 9.2 | 5.1 | 3.1 | 163.1 |
| FAM34 2.5 10000k 1000GB #Dr24 | 29.2 | 14.9 | 9.5 | 49.9 |
| FAM37 2.5 10000k 1000GB #Dr24 | 27 | 13.2 | 8.9 | 47.7 |
| FAM7 2.5 10000k 600GB #Dr50 | 21.6 | 3.85 | 6.9 | 45.8 |
| FAM9 2.5 10000k 600GB #Dr50 | 22.5 | 6.21 | 8.13 | 45.2 |
| FAM2 2.5 10000k 600GB #Dr125 | 23.48 | 12.48 | 5.67 | 52.8 |

| | | | | |
|---|---|---|---|---|
| FAM7 2.5 10000k 600GB #Dr125 | 21.83 | 2.91 | 5.11 | 65.6 |
| FAM9 2.5 10000k 600GB #Dr125 | 25.12 | 13.1 | 7.04 | 60.1 |
| FAM24 2.5 10000k 600GB #Dr125 | 24.91 | 12.3 | 10.03 | 53 |
| FAM19 2.5 10000k 600GB #Dr150 | 11.75 | 7.85 | 14.54 | 75.98 |
| FAM2 2.5 10000k 600GB #Dr175 | 25.62 | 13.81 | 5.46 | 59.3 |
| FAM7 2.5 10000k 600GB #Dr250 | 19.8 | 2.25 | 4.12 | 72.7 |
| FAM24 2.5 10000k 600GB #Dr250 | 26.84 | 14.63 | 7.96 | 64.6 |
| FAM9 2.5 10000k 600GB #Dr275 | 23.38 | 7.9 | 5.39 | 69 |
| FAM2 2.5 10000k 600GB #Dr325 | 23.48 | 9.82 | 5.85 | 66.4 |
| FAM24 2.5 10000k 600GB #Dr500 | 24.03 | 16.04 | 6.89 | 71.9 |
| FAM44 2.5 10000k 600GB #Dr528 | 13.67 | 10.52 | 4.5 | 55.4 |
| FAM42 2.5 10000k 600GB #Dr576 | 21.49 | 16.21 | 5.71 | 63.6 |
| FAM18 2.5 10000k 900GB #Dr192 | 20.63 | 19.63 | 12.39 | 38.6 |
| FAM35 3.5 10000k 0GB #Dr84 | 12.8 | 15.3 | 6.3 | 126.2 |
| FAM34 2.5 15000k 0GB #Dr24 | 29.2 | 14.9 | 9.5 | 25 |
| FAM37 2.5 15000k 0GB #Dr24 | 37.7 | 19.3 | 11.4 | 21.7 |
| FAM7 2.5 15000k 300GB #Dr150 | 27.28 | 3.98 | 9.87 | 29 |
| FAM9 2.5 15000k 300GB #Dr150 | 28.42 | 20.45 | 12.24 | 27.3 |
| FAM2 2.5 15000k 300GB #Dr200 | 33.68 | 21.52 | 4.53 | 29.9 |
| FAM24 2.5 15000k 300GB #Dr225 | 32.29 | 21.39 | 7.18 | 25 |
| FAM9 3.5 15000k 300GB #Dr120 | 19.32 | 11.61 | 5.95 | 16.9 |
| FAM2 3.5 15000k 300GB #Dr165 | 19.48 | 11.26 | 7.8 | 16.5 |
| FAM24 3.5 15000k 300GB #Dr180 | 19.24 | 11.12 | 6.22 | 16.9 |
| FAM7 3.5 15000k 300GB #Dr195 | 14.68 | 2.67 | 3.88 | 17.9 |
| FAM38 2.5 SSD 480GB #Dr0 | 97.64 | 277.97 | 95.3 | 10.8 |
| FAM38 2.5 SSD 480GB #Dr0 | 51.97 | 149.62 | 81.04 | 10.2 |

**Table 3: Online 4 Transactional Data Highlighting the Top 25% Scores for Each Metric/Workload**

Table 3 は、オンライン 4 のグラフおよびオンライン 3 のデータと同じ傾向を示している。高回転ドライブを搭載したシステムが性能指標の上位 25%に入る傾向があるのに対して、低回転大容量ドライブを搭載したシステムが Ready Idle 指標の上位 25%に入る傾向がある。すべての指標の上位 25%に入っている単一のオンライン 4 システムは存在しない。

**シーケンシャル・ワークロード用に最適化されたシステムからのデータ**

次のデータは、シーケンシャル・ワークロード用に最適化されたシステムの ENERGY STAR データセットから抽出されたもので、システム・アーキテクチャ、ドライブの回転数、およびドライブ数の影響を示している。

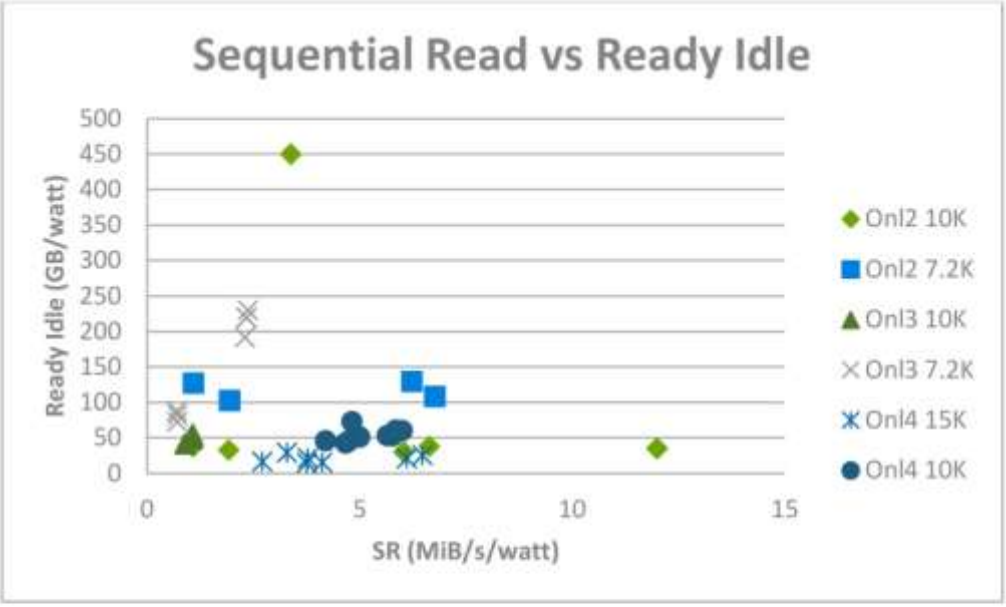



Figure 20: Sequential Read vs. Ready Idle by Taxonomy and Drive Type

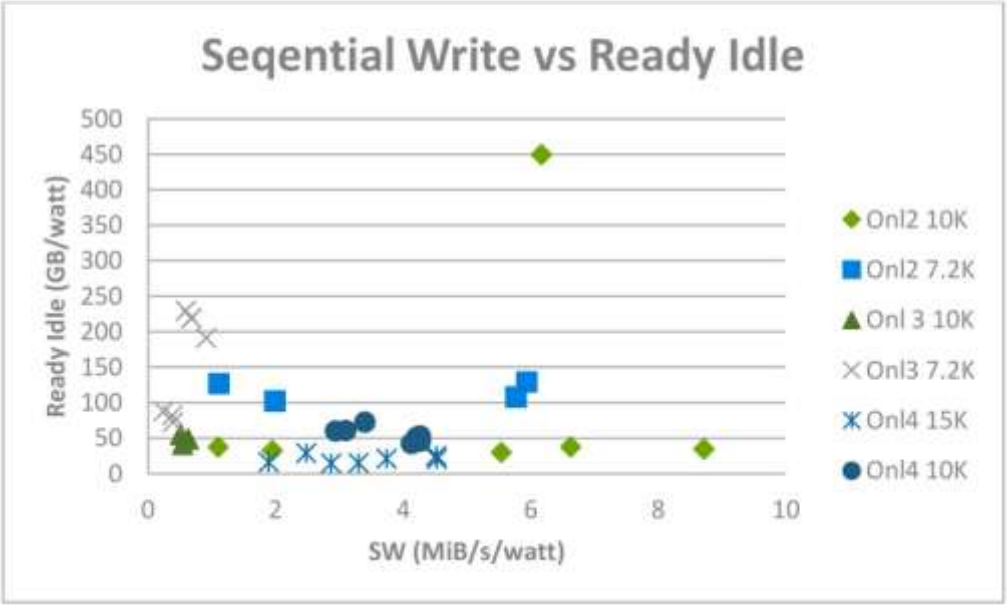
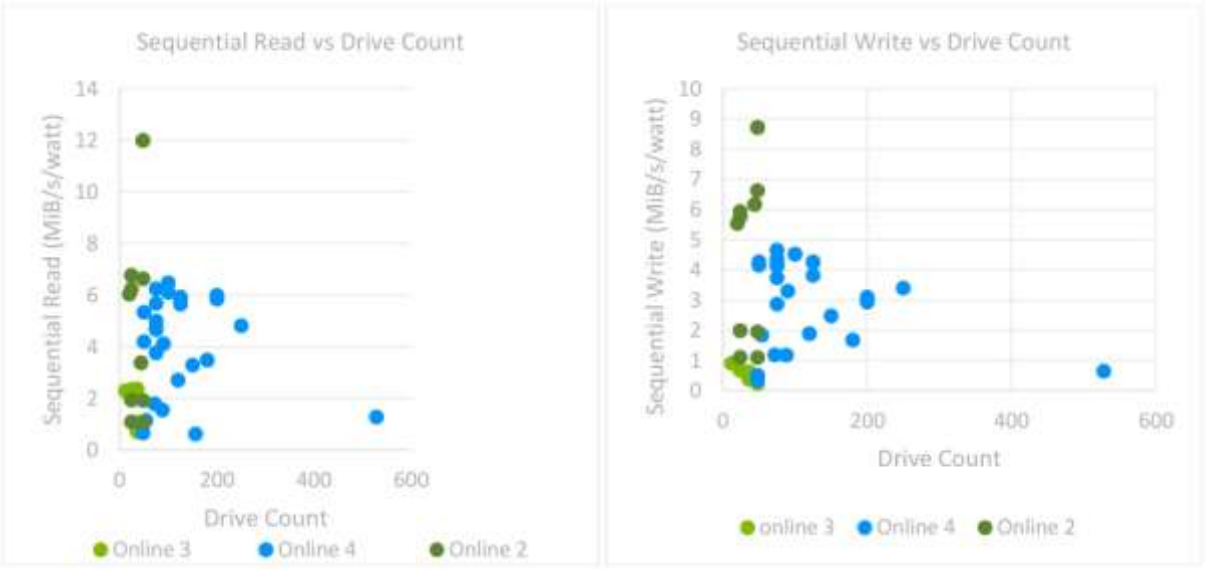



Figure 21: Sequential Write vs. Ready Idle by Taxonomy and Drive Type

Figure 20 と Figure 21 は、前節で観察したのと同様に、ドライブ数別の Sequential Read 指標、Sequential Write 指標、および Ready Idle 指標の関係を示している。分類カテゴリとドライブ・タイプ別で指標が広く分布している。Ready Idle 指標は、7.2Krpm ドライブを搭載したオンライン 3 システムを除いて、分類カテゴリとドライブ・タイプ別でほとんど分散していない。次のグラフはドライブ数別のシーケンシャル指標を示している。



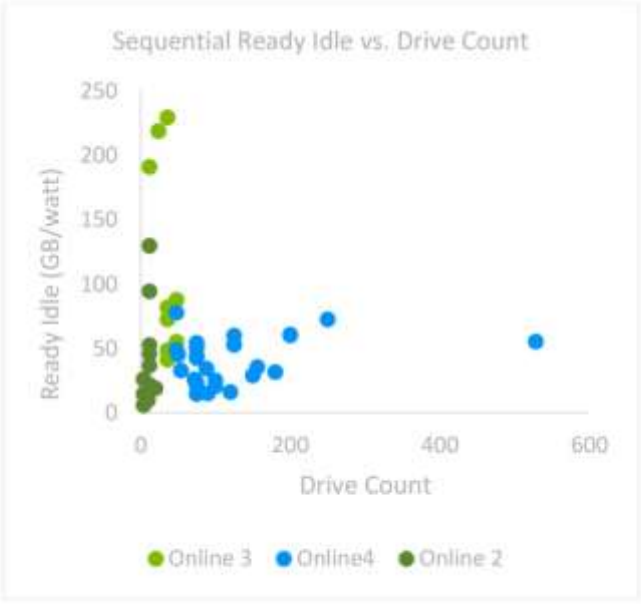




Figure 22: Sequential Metrics vs. Drive Count Broken Out by Taxonomy

Figure 22 は、どの分類カテゴリでも Sequential Read/Write 指標が広く分布していることを示している。「本質的」効率性によってストレージ製品を差別化するために使用可能な構成の明確な傾向や違いは見られない。

Figure 23 と Figure 24 では、データはドライブ・タイプ別に分類され、ドライブの回転数またはフォーム・ファクターに関連付けられた特定の傾向に対して分析されている。

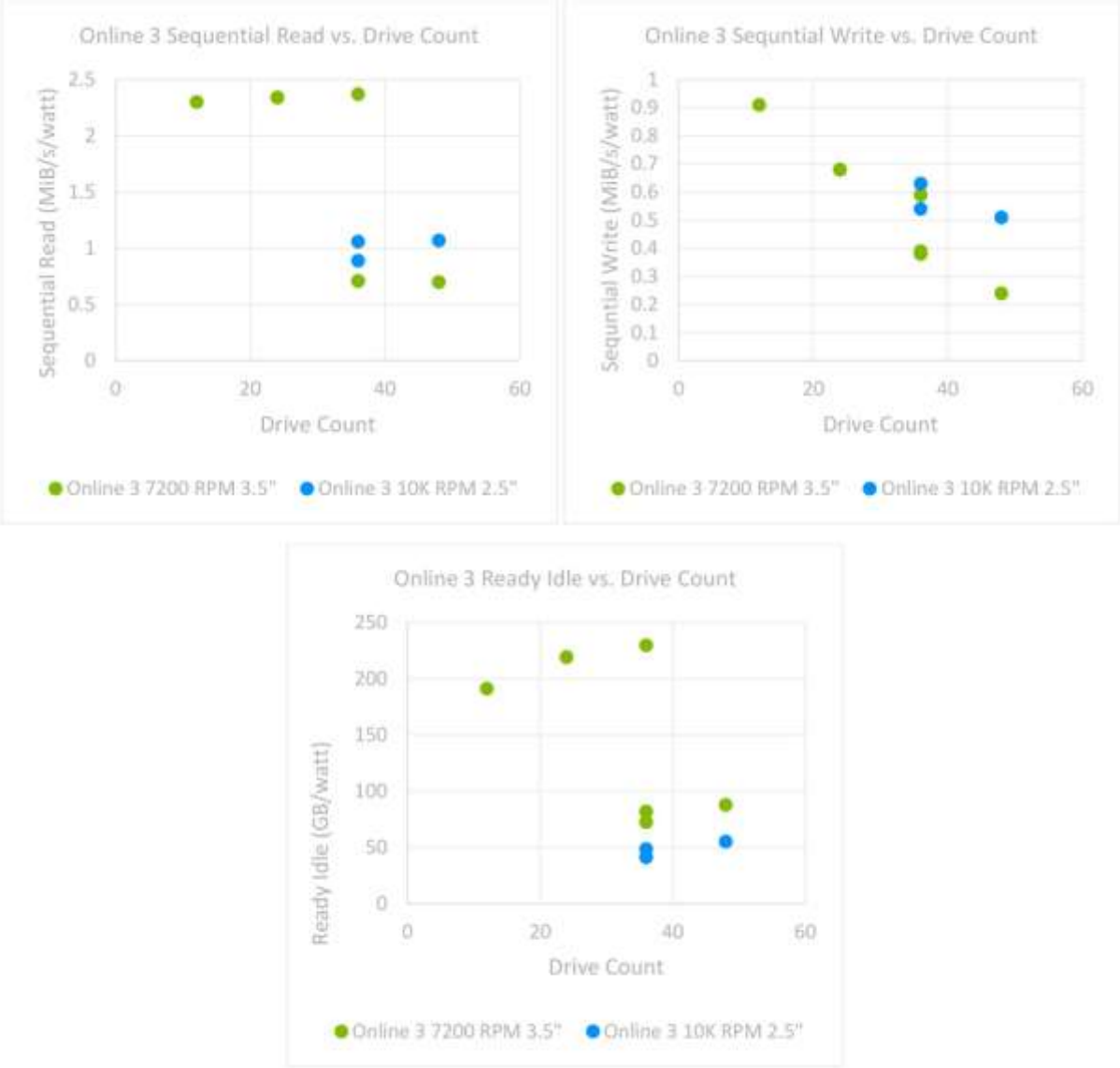



Figure 23: Online 3 Sequential Metrics vs. Drive Count

Figure 23 は、ドライブ・タイプとドライブ数で分類されたオンライン 3 の Sequential Read/Write 指標と Ready Idle 指標を示している。グラフは、概して、前節の結果について得られたものと同じ結論を支持している。より高回転なドライブは性能/電力指標で優れているのに対して、より低回転なドライブは容量/電力指標で優れている。ただし、オンライン 3 の Sequential Read/Write のデータセットは非常に小さく、根拠のある明確な結論を得ることはできない。

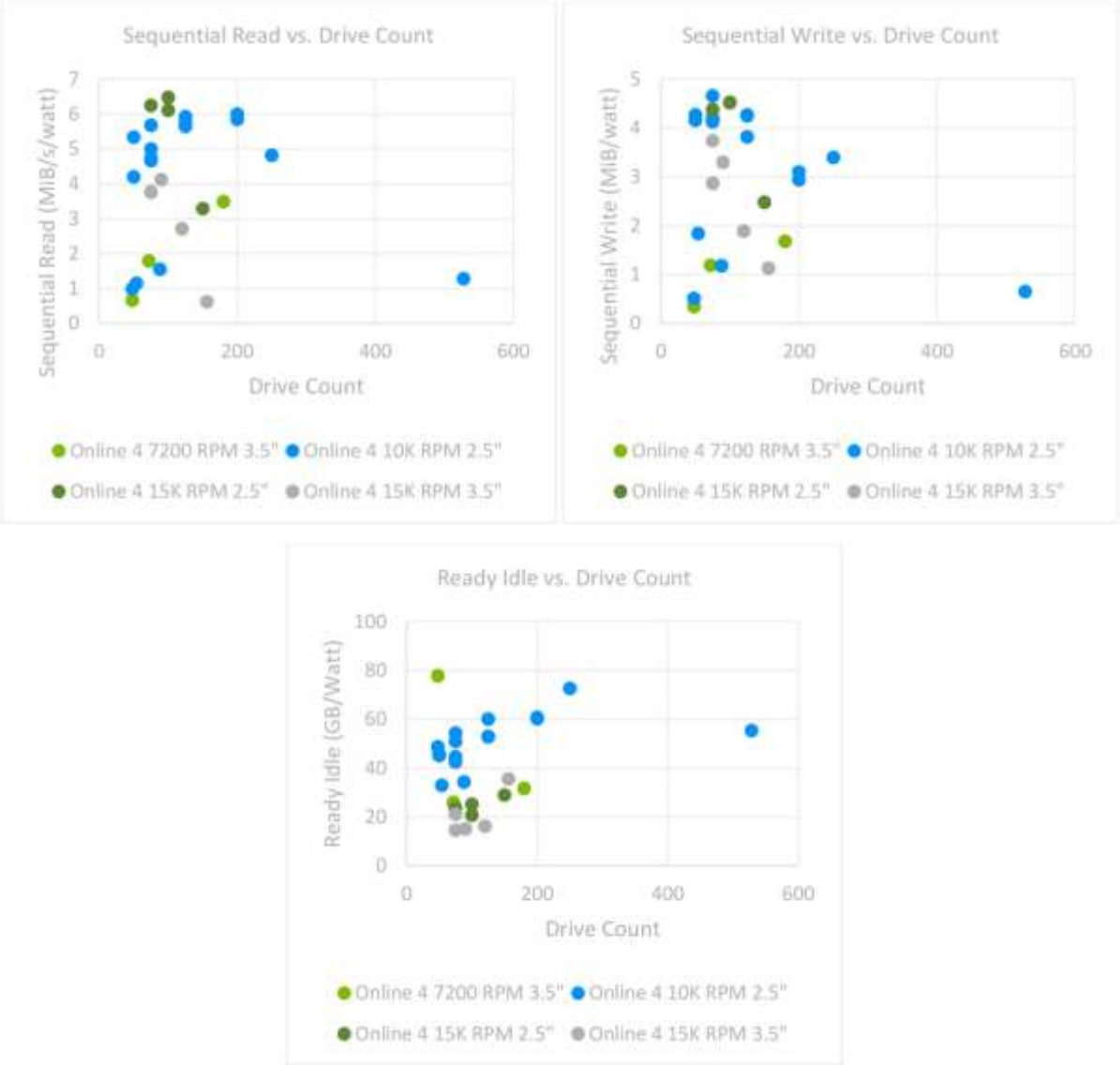



Figure 24: Online 4 Sequential Metrics vs. Drive Count

Figure 24 は、ドライブ・タイプとドライブ数で分類されたオンライン 4 の Sequential Read/Write 指標と Ready Idle 指標を示している。前にドライブ・タイプと台数に関して述べたものと同じ結論に達した。500 台を超えるドライブを搭載した 1 つのシステムは、同様のドライブ数を搭載した他のシステムほど良好な結果を出さなかった。これは、システムのフロントエンドまたはバックエンドでデータ転送が停滞するシステムのボトルネックが原因の可能性が高い。

次の Table 4 と 5 は、オンライン 3 とオンライン 4 の Sequential Read/Write 指標および Ready Idle 指標を示しており、それぞれの指標値で全体の上位 25%が強調表示されている。

オンライン 3 のデータセットは小さいため、特定の結論を導き出すのは困難である。ただし、概して、容量が大きく、回転数が低いドライブほど、Ready Idle 指標が最も高いことが分かっている。こ

のオンライン 3 システムの限定サンプリングでは、ファミリー10 とファミリー32 の Ready Idle 指標と Sequential Read 指標および Sequential Write 指標が最も良い。

| Label | Sequential Read Workload Test | Sequential Write Workload Test | Ready Idle Workload Test (GB/W) |
|---|---|---|---|
| FAM3 3.5 7.2k 1000GB #Dr36 | 0.71 | 0.38 | 81.99 |
| FAM3 3.5 7.2k 1000GB #Dr36 | 0.71 | 0.39 | 72.67 |
| FAM4 3.5 7.2k 1000GB #Dr48 | 0.7 | 0.24 | 87.87 |
| FAM10 3.5 7.2k 4096GB #Dr12 | 2.3 | 0.91 | 191.2 |
| FAM10 3.5 7.2k 4096GB #Dr24 | 2.34 | 0.68 | 219.2 |
| FAM32 3.5 7.2k 4096GB #Dr24 | 2.34 | 0.68 | 219.2 |
| FAM10 3.5 7.2k 4096GB #Dr36 | 2.37 | 0.59 | 229.7 |
| FAM3 2.5 10k 450GB #Dr36 | 1.06 | 0.63 | 48.67 |
| FAM3 2.5 10k 450GB #Dr36 | 0.89 | 0.54 | 41.38 |
| FAM4 2.5 10k 450GB #Dr48 | 1.07 | 0.51 | 55.27 |

**Table 4: Online 3 Sequential Data Highlighting the Top 25% Values for Each Metric/Workload**

| Label | Sequential Read Workload Test | Sequential Write Workload Test | Ready Idle Workload Test (GB/W) |
|---|---|---|---|
| FAM4 3.5 7.2k 1000GB #Dr48 | 0.66 | 0.34 | 77.87 |
| FAM6 3.5 7.2k 2000GB #Dr72 | 1.79 | 1.19 | 26 |
| FAM6 3.5 7.2k 2000GB #Dr180 | 3.49 | 1.68 | 31.7 |
| FAM4 2.5 10k 450GB #Dr48 | 0.99 | 0.51 | 48.75 |
| FAM5 2.5 10k 450GB #Dr54 | 1.15 | 1.84 | 32.95 |
| FAM5 2.5 10k 450GB #Dr88 | 1.55 | 1.18 | 34.32 |
| FAM7 2.5 10k 600GB #Dr50 | 4.2 | 4.27 | 45.8 |
| FAM8 2.5 10k 600GB #Dr50 | 5.34 | 4.16 | 45.2 |
| FAM2 2.5 10k 600GB #Dr75 | 4.73 | 4.18 | 44.7 |
| FAM7 2.5 10k 600GB #Dr75 | 5 | 4.21 | 51 |
| FAM8 2.5 10k 600GB #Dr75 | 5.68 | 4.66 | 54.4 |
| FAM24 2.5 10k 600GB #Dr75 | 4.67 | 4.13 | 42.5 |
| FAM2 2.5 10k 600GB #Dr125 | 5.77 | 4.25 | 52.8 |
| FAM8 2.5 10k 600GB #Dr125 | 5.93 | 3.82 | 60.1 |
| FAM24 2.5 10k 600GB #Dr125 | 5.65 | 4.27 | 53 |
| FAM2 2.5 10k 600GB #Dr200 | 6 | 2.94 | 60.2 |
| FAM24 2.5 10k 600GB #Dr200 | 5.85 | 3.1 | 60.8 |
| FAM7 2.5 10k 600GB #Dr250 | 4.82 | 3.4 | 72.7 |
| FAM44 2.5 10k 600GB #Dr528 | 1.28 | 0.65 | 55.4 |
| FAM8 2.5 15k 300GB #Dr75 | 6.25 | 4.38 | 23.5 |
| FAM2 2.5 15k 300GB #Dr100 | 6.48 | 4.52 | 25.2 |
| FAM24 2.5 15k 300GB #Dr100 | 6.11 | 4.53 | 20.7 |
| FAM7 2.5 15k 300GB #Dr150 | 3.29 | 2.48 | 29 |
| FAM9 3.5 15k 300GB #Dr75 | 3.76 | 2.87 | 14.6 |
| FAM24 3.5 15k 300GB #Dr75 | 3.78 | 3.74 | 21.11 |
| FAM2 3.5 15k 300GB #Dr90 | 4.12 | 3.3 | 15.1 |
| FAM7 3.5 15k 300GB #Dr120 | 2.71 | 1.89 | 16.2 |
| FAM19 3.5 15k 600GB #Dr0 | 0.62 | 1.13 | 35.58 |

Table 5: Online 4 Sequential Data Highlighting the Top 25% of Values for Each Metric/Workload

Table 5 は、オンライン 4 のデータを示し、それぞれの指標の中の上位 25%を強調表示している。3 つすべての指標の上位 25%に入る単一のシステムはない。低回転大容量ドライブを搭載したシステムで最も高いアイドル状態スコアが見られた前節のシステムとは対照的に、アイドル状態の上 25%に入っているのは 7.2Krpm システムの 1 つだけである。この結果の大部分は、10Krpm システム上のド

ライブ数が多く性能指標がよいシステムと比較して 7.2Krpm システムは HDD 数が少ないことが影響している。ドライブ数の多いシステムほど、コントローラ電力を分散して負担できるため、Ready Idle 指標は低くなる。Sequential Read 指標、Sequential Write 指標、Ready Idle 指標全ての上位 25%に入る単一のシステムはない。

**データの分析**

データセットはシステム種と製造業者の数が少ないため、データセット内のコンポーネント・タイプ間の結論を導き出すのは困難である。概して、トランザクション・システムの場合は、高回転ドライブによってより高い性能指標が出力されるが、低回転大容量ドライブはより高い Ready Idle 指標を生成する。この事実は、Ready Idel 指標が運転エネルギー効率の良好な指標にならない理由を示している。Sequential Read/Write 指標はサンプル数が少な過ぎて結論を導き出すことができないが、15 K rpm ドライブを搭載したシステムで Ready Idle 指標が良好なシステムが存在しない事実が、高回転ドライブが競争力のある Ready Idle 指標を達成しないという観察結果を支持している。

ドライブの容量と回転数タイプが同じシステムを比較する場合は、次の属性が指標に影響する。

- コントローラ上のキャッシュ・サイズ（コントローラ上の追加のメモリを通してまたはドライブをキャッシュとして使用して）
- システムに内蔵されたドライブの台数
- システムにデータをプッシュするサーバの台数とストレージ・システムをサーバに接続しているフロントエンド・パイプの数
- サーバ、コントローラ、およびバックエンド・ストレージ・デバイス間の接続タイプ
- 世代、CPU の数、CPU のタイプ、データ移動機能、バックエンド・ストレージ・デバイス・タイプなどのコントローラ・アーキテクチャの特徴

**SSD のエネルギー消費に対する読み取りと書き込みの影響**


**Figure 25: Transactional Metrics for the Online 4 All Flash Array System**

Figure 25 は、オール・フラッシュ・アレイが Random Read 時にメリットを提供するが、Randowm Write 指標値は高性能 HDD ベースのアレイを使用した同様のストレージ・システムの値とあまり変わらないという前述した分析で観察された事実を示している。

**オンライン 4 のデータ**

Figure 26 と Figure 27 はオンライン 4 のデータを示している。図中、同一のファミリーに属するデータを破線で結んである。このようなシステムは、Hot Band や Random Read/Write ワークロード用に最適化されてから、Sequential Read/Write ワークロード用に最適化された。このようなシステムは、異なるドライブ数とドライブ・タイプを使用してテストされ、性能指標値に対するドライブ・タイプとドライブ数の影響の妥当な例を提示している。

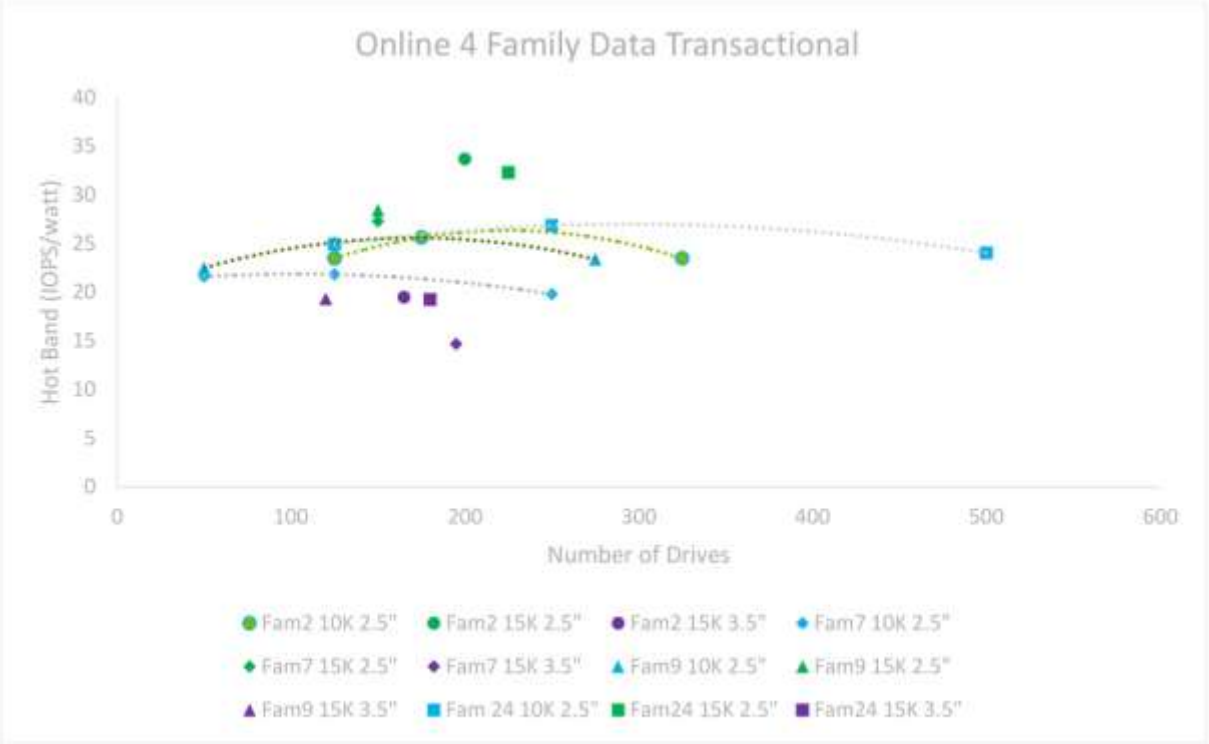



Figure 26: Online 4 Transactional Family Data

Figure 26 は、Hot Band 指標について、異なるドライブ・タイプおよびドライブ数との関係を示している。最も性能の高いシステムは 15Krpm　2.5”フォーム・ファクター・ドライブを搭載したシステムであり、最も性能の低いシステムは 15Krpm　3.5”フォーム・ファクター・ドライブを搭載したシステムである。10Krpm ドライブはその中間で、最小/最適/最大点を示している。最小点と最大点は、最適点の 10%以内で、最小および最大構成点に関する ENERGY STAR V1.0 ストレージ要件で求められている 15%の可変帯以内に適合する。これは、より新しいストレージ・システムでは、所与のストレージ製品内でドライブ数による性能指標の変化が少ないことを意味する。最適ドライブ数の上下のドライブ数で性能指標が低下しにくいいることは、単一の最適点測定が全範囲の構成を適切に代表しているため、データ・センター・ストレージ・バージョン 1 に関する ENERGY STAR プログラム要件で現在要求されているようなトランザクション・テストに対して最小点と最大点をテストする必要がないことを示唆している。

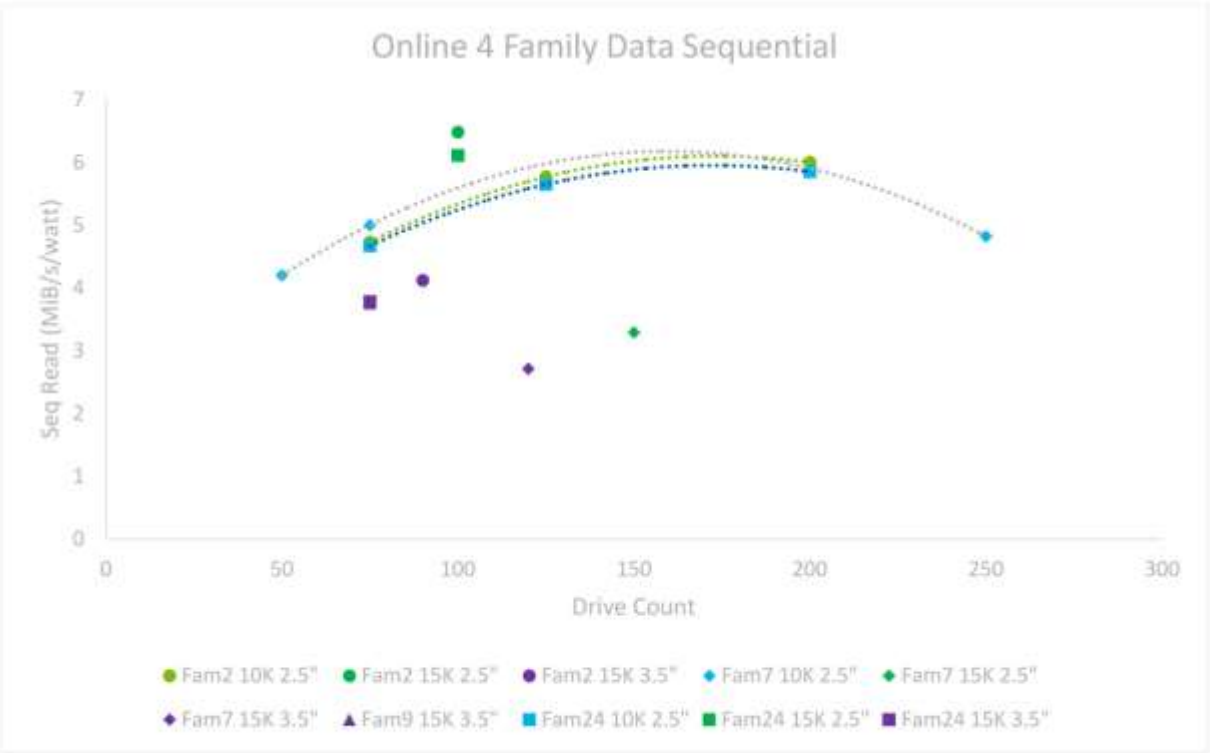



Figure 27: Online 4 Sequential Family Data

Figure 27 は、ドライブ・タイプとドライブ数が異なるオンライン 4 の Sequential Read 指標値を示している。概して、2.5”ドライブの方が 3.5”ドライブより良い指標値を示している。これは、記憶媒体が小さい方がデータ検索が速くなるからである。最小、最適、および最大ドライブ数構成では、10Krpm 2.5”ドライブが使用された。

最適点の上と下では、ドライブ数は異なる目的を果たす。最適点までは、システム性能が注目点である。ストレージ・システムは、最適点を超えるまで成長するように設定されている。最適点を超えると、システムは容量が増大し、Ready Idle 指標が注目点になる。最適点を超えたら、容量、性能、および TCO を評価してストレージ製品を追加するタイミングを決定する。概して、ファミリー内の最小/最適/最大構成セット全体で同様の性能指標値が示されている。

## 6. 容量最適化手法（COM）

容量最適化は、特定のストレージ・デバイスの設置面積に対してより大きなデータ・ストレージ容量を可能にするためにストレージ製品内で使用される一連のソフトウェア・テクニックを意味する一般用語である。保存データ量を減らすことで、必要なドライブ容量を減らすことができ、結果的に特定の操作のエネルギー消費が減る。このようなテクニックのそれぞれが容量最適化手法（COM）と呼ばれている。一般的な容量最適化手法 (COM)の説明を以下に示す。

高度なデータ保護 － パリティ RAID とイレージャ・コーディングを含む：これらは、フル・コピーを使用しないデータ保護の手法である。ミラーリングされたコピーまたはフル・コピーを高度なデータ保護に置き換えた場合の予想スペース節約は 25～50%である。
シン・プロビジョニング：プロビジョニング時にすべての物理容量を割り当てるのではなく、アプリケーションがデータを書き込む時に物理容量を割り当てる技術。
デルタ・スナップショット：複数のフル・コピーを使用せずに、ファイルの複数のバージョンを再構築するタイプのポイントインタイム・コピー。変更が小規模であれば、大きなスペース節約が可能になる。
圧縮：サイズを削減するためのデータ・エンコーディングのプロセス。通常は、2：1 の圧縮率が達成される。
データ重複排除：様々な粒度レベルで、データの重複を共有コピーへの参照に置き換えること。これは大きなスペース節約が得られる。

概して、容量最適化手法 (COM)は、概ね相互に独立しているが、完全に独立しているのではない。容量最適化手法 (COM)は、任意の組み合わせでメリットを提供するが、複合効果は個別のメリットの合計と一致せず､単一のストレージ製品に 3 つ以上の COM を展開した場合は非常に効率が悪くなる。

ENERGY STAR データ・センター・ストレージ V1.0 は、SNIA Emerald™テスト中に適格なオンライン製品カテゴリが高度なデータ保護を有効にするよう求めている。オンライン 3 とオンライン 4 では、上で指定した 4 つの容量最適化手法 (COM)のうち少なくとも 1 つ（高度なデータ保護以外に）を選択可能な機能として使用できるようにする必要がある。テスト検証では、特定の容量最適化手法 (COM)の存在と有効性を確認する検出手法が適用される。テスターは、すべてのアクティブ測定テストやアイドル測定テスト中に無効にすることが可能なすべての容量最適化手法 (COM)を無効にする必要もある。

## 7. ストレージの短期的進化

ストレージ製造業者は、特定の製品の設置面積に対して保存可能なデータの量を増やし、個別の製品レベルとデータ・センター・システム・レベルの両方でストレージ製品全体の設置面積の性能/W 機能を向上させるために、ハードウェア・ベースとソフトウェア・ベースのイノベーションの研究、開発、および提供を継続している。

### ストレージ・メディア

ストレージ・メディアは、ハード・ディスク・ドライブ（HDD）とソリッド・ステート・デバイス（SSD）の両方の継続的技術開発を反映し続けている。この両方のタイプのデバイスは、今後少なくとも 5～10 年間はデータ・センター内に展開され続ける。一部のデータ・センター・ストレージ製品ベンダーは 1 種類のメディアしか使用しないが、他の多くのベンダーは運用要件、容量要件、およびビジネス要件をより良く満たすための選択肢を消費者に提供するようになる。この章では、主な違いと一部の類似点を中心にして、両方のデバイス・タイプの特徴について説明する。違いは、ドライブがストレージに HDD 技術を使用し続ける必要があるか、SSD 技術を使用し続ける必要があるかである。

ドライブ・タイプの一般的な特徴

エンタープライズ・システムの HDD（ハード・ディスク・ドライブ）は高回転と低回転大容量の 2 つのタイプに分けられる。高回転ディスク・ドライブは、SAS とファイバー・チャネルのどちらかの高性能データ・インターフェイスを備えたディスク・ドライブのことである。このドライブの回転数は 10K rpm または 15K rpm と高い。また、比較的密度（データ・ストレージ容量）は低い。低回転大容量ディスク・ドライブは、通常、高回転ディスク・ドライブの 5～10 倍の大きいデータ容量を備えたディスク・ドライブのことである。このドライブは、低性能なデータ・インターフェイス（SATA が一般的）を備えている。また、回転数は 7.2K rpm 以下と低い。

エンタープライズ・システムの SSD（ソリッド・ステート・ディスク・ドライブ）は、ディスク・ドライブと同様に機能するように構成されたフラッシュ・メモリのブロックである。SSD は、ディスク・ドライブのすべての特徴を備えている。また、通常は、高性能なデータ・インターフェイス（SAS が一般的）を使用して構成されている。

以下では、このストレージ・メディア・タイプについてより詳しく説明する。

SSD の特徴

オンライン・ストレージ・メディア（便利なデータ移動手段や不揮発性キャッシュではなく）としての SSD の出現は、シリコン技術の改良から生まれた比較的最近の進歩である。SSD は、次のような特徴の結果として、特定の用途に関して HDD と比べて様々なメリットを提供する。

- 重量：比較的容量が少ない場合、SSD の重量は HDD よりも軽い。このメリットは、多くの場合、非常に大きい容量と密度では維持されないが、家電製品からラップトップまでのポータブル・デバイスに信頼できる軽量ストレージを提供する点において非常に重要であることが証明されている。

- 絶対性能：SSD ストレージは、小規模な I/O 操作の場合に高速（15k rpm）HDD より性能が優れている（例えば、データ転送が 4 KB 単位のトランザクション・ワークロード）。SSD は、高性能なストリーミング・ワークロード（1 回の操作で数百 KB の転送）用には設計されていない。
- 電力ドロー：SSD は従来の高性能 HDD より性能が優れているが、中小サイズの SSD はこれらのドライブより必要な電力が大幅に少なく、通常は、1-2 TB 7.2k RPM HDD の電力ドロー程度である。これは、特定の容量の高性能 HDD を同じ数量の SSD に置き換えることにより、エネルギー消費が下がり、性能が上がることを意味する。

これらのメリットにはマイナス面も伴い、データ・センター・ワークロードの全範囲（SME（中小企業）、エンタープライズ、およびクラウドを含む）にわたって幅広く採用されるまでには時間がかかる。

- 電力/GB が概ね一定である。つまり、2 TB の SSD 容量は 1 TB の電力の約 2 倍を消費する。これは、電力が主に回転数の係数であり、記録密度の高い技術によって容量を倍にしてもエネルギー消費が増えない HDD パラダイムとは異なる。
- コストは大幅に下がっているが、大量の SSD 部品を製造することに課題があり、製造能力が現実的なネックになっている。供給に限度があるため、需要が生産と同程度以上に維持されることで、しばらくは高値が維持される。SSD コンポーネントの製造業者は困難なビジネス・ジレンマに陥っている。生産量を上げるために製造工場に巨額の投資を行っても、供給が過剰になると価格が下がり、投資を回収できないというリスクがある。一方、現状のままでは事実上拡大ペースは抑制されてしまう。各ベンダーは、固有のビジネス状況に基づいて決定を下す必要がある。
- HDD と同様に、SSD も不良セルが生じたり摩耗しやすいため、再配置を繰り返す必要がある。違いは、SSD のセルが処理する書き込み操作の回数に寿命による制限があることである。そのため、SSD の多くには、デバイスの早期老化を避けるための「ウェアレベリング」アルゴリズムが組み込まれている。

**ユース・ケースの比較**

いくつかのデータ・センターのユース・ケースでは、SSD の採用が増え続けている。従来のデータ・センター内のアプリケーションの多くが高い性能を要求する。アプリケーションによっては、一部の機能で高い性能が要求される。このような性能制約にさらされているユーザは HDD の代わりに SSD を採用する可能性がある。この状況では、SSD の使用は、大規模システム内の 1 つのストレージ階層または専用のストレージ・アプライアンスの使用に限定される。他のワークロードは、ストレージ・システム内のキャッシュが増えることで恩恵を受ける可能性があり、そのような用途での SSD は継続的成長も見込まれるが、このカテゴリでは、通常、システム内の HDD は SSD に置き換えられてはいない。このクラスのアプリケーションでは、生産性/経費の観点から SSD への投資はビジネスに有益である。

ほとんどのクラウド・データ・センターは、データがインターネット経由で配信され、そこで大幅な応答の遅延が起きるため、性能が要求されない。何が重要かは、相対的である。前述のアプリケーションは鋭敏性がミリ秒単位で測定されたのに対して、ここではクラウド・データ・センター・ケー

スの最大数秒の遅延を処理できるかどうかの話をしている。このようなデータ・センターは、大量のデータをオンラインで維持しなければならないという別の基本的問題に直面している。また、非常に少ない利益幅で運用している場合も多い。このような状況では、適性な性能レベルでの容量/経費が性能よりも重要なため、高容量 HDD の展開が続く可能性がある。

HDD に依存し続ける可能性が高い SME とエンタープライズ・データ・センターのもう 1 つのユース・ケースは、信頼性を確保するためのデータのバックアップである。ほとんどの企業が、複製から高度なバックアップ・アプライアンスまで、何らかの形態のデジタル・バックアップを使用している。これは、データが事故や装置の故障で破損しないことを保証するためである。これは、大容量が必要な別の状況である。一般的にバックアップ・ソフトウェアは変化した項目のみを保存するように管理するため、性能は大きな問題にはならないが、バックアップは安価に抑える必要があるオーバーヘッドのコストであるため、非常に大きな問題になる。

**メディア開発の方向性**

前述したように、SSD と HDD のどちらも今後展開されるユース・ケースは十分にある。そのため、両方の分野で研究と開発が継続されている。この章では、両方の分野の可能性の高い開発について見ていく。ディスク・ドライブ・サプライヤは、主要顧客である、ドライブをサーバやストレージ製品に統合するシステム・ベンダーによって定義された設計パラメータ内に留まるために各ドライブによって消費される電力を維持または削減する努力を強いられている。この制約の中で、サプライヤは次の点に注目している。

- すべての回転数で容量を増やす。限られた電力予算の中でこれを実現するということは、プラッタをドライブに追加できないことを意味する。この領域の研究は、主に、受け入れ可能な信頼性のレベルで記録密度を上げるための新しいまたは改良された記録技術に焦点が当てられる。
- ドライブの機械部品の変更を試すことによってドライブのエネルギー効率を高める。密封され、ヘリウムで満たされたドライブがこの進行方向の 1 つの例である。

ソリッド・ステート技術の進歩は、以下の点に主に集中している。

- 性能の向上
- パッケージ・サイズ（ボリューム）あたりの容量の増加。これにより、複数の「ディスク相当物」またはシステム・スロットを使用せずにシステム全体の容量が増えるため、より狭い設置面積でより大きなストレージ・システムをサポートできる。SSD の電力/GB は比較的一定のため、パッケージあたりのエネルギー消費は増えるが、必ずしもトランザクションあたりのエネルギー消費は増えないため、ユース・ケースがエネルギー効率を判断する重要な要因になる。
- 需要と採用を増やすためのコスト/GB の削減。ディスク業界と同様に、継続的な開発テーマになり得る。

この章で説明した特徴を組み合わせると、SSD と HDD の両方が今後何年間も多くのデータ・センター環境（レガシー設備と新規購入の両方）で共存し続けるという結論になる。

**最近のデータ・センターの傾向：アプライアンスに取って代わるアプリケーション**

ここ数年の間に、データ・センター業界は新しい"機器"のトレンドを経験した。高度なアプリケーション・ソフトウェア製品は、ストレージ・システムやネットワーク・スイッチなどの従来のデータ・センター・アプライアンスをエミュレートするまでに進化した。文献や論文では、このような製品カテゴリは、ソフトウェア・デファインド・ネットワーキング（以下、SDN）、ソフトウェア・デファインド・ストレージ（以下、SDS）、コンバージド、ハイパーコンバージド、ストレージ・サーバなどという名前で呼ばれている。ここでは、単純に SDN や SDS と呼ぶことにする。このようなアプリケーション・カテゴリの両方が初期採用者の間で大幅な市場シェアの成長を経験しており、両方が時間とともにメインストリーム・データ・センターでの存在感を強めている。

このようなカテゴリの 1 つである SDS を詳しく調べると、運用ソフトウェア製品はインストール先のストレージとサーバのブランドに依存しないように設計されていることが分かった。代わりに、このような製品では、他のアプリケーションと同じように、適切な実行環境についてのガイドラインの類が存在する。つまり、命令セット、オペレーティング・システム、使用可能なメモリ、アプリケーションのインストールに必要なディスク領域などの要件が列挙される。このような SDS アプリケーションは、それが使用するディスク・ドライブとともに、どのようなサーバにも自由にインストールすることができ、そのサーバは、既に別のアプリケーションを実行しているものでもよい。事実、SDS または SDN 製品を使用する主な目的は専用アプライアンスの購入を避けることである。このようなアプリケーションは、通常、SDS を使用してデータを保存するアプリケーションと同じサーバを共有する仮想化されたサーバ環境で動作する。

この傾向の 1 つの重要な結果は、ストレージまたはネットワーク機能がもはやハードウェアの特定のブランドや構成に結びついていないことである。そのエネルギー効率は、ソフトウェアの動作とハードウェアの動作という 2 つの独立した変数によって決められる。これがストレージまたはネットワーク・アプライアンスの表面上の真実であるが、基本的な違いがある。また、例としてストレージを使用すると、特定のストレージ・アプライアンスは実際には特定のコントローラ設計だったり、専用の I/O インフラストラクチャーだったり、特定のストレージ・アプリケーションだったりする。この効率性をテストした場合は、ユーザのデータ・センターに配置された場合でも、同じ結果になることが想定できる。SDS 環境では、性能がストレージ・アプリケーション、特定のサーバ構成（すべてのアプリケーションに必要なメモリや I/O コントローラなどを含む）、およびサーバ上で使用中の他のアプリケーションに左右される。このようなアプリケーションが使われない場合でも、実際のハードウェア構成は、そのソフトウェア・アプリケーションのニーズだけでなく、意図された顧客の使用ニーズを反映することになる。

2 つ目の結果として、このような SDS または SDN アプリケーションのどちらかを実行中のサーバには、データ・センター内の他のサーバとは大きく異なる構成が必要になる。少なくとも、ネットワークまたはディスク・トラフィックの増加に対応するために I/O チャネルを増やす必要がある。同様に、特定のアプライアンスの代用として機能する適切な数の高性能 I/O デバイスをシステムに組み込む必要がある。

**容量最適化手法 (COM)のメリットの定量化**

容量最適化手法 (COM)の簡易存在証明テストを使用するよりも、容量最適化手法 (COM)を有効にしたアイドル指標とアクティブ指標に対する影響を把握するための定量化可能な手法を見出す方が有益かもしれない。アクティブ容量最適化手法 (COM)のメリットの見積もりはいくつか存在するが、アクティブ操作中の効率を正確に考慮する有効な手法を確立するにはもっと多くの特性評価テストが必要である。テストによってアクティブ COM の影響を比較する方法の例を以下にいくつか示す。Table 6 は物理容量削減（つまり、論理容量は一定に保つが、エネルギー消費を削減するために物理ドライブを取り外す）に基づいており、Table 7 は使用可能容量最適化（つまり、物理ストレージの実効仮想容量の増加を定量化する）に基づいている。"Delta (GB/W)"行と"Delta (IO/W)"行は、アイドル状態指標とアクティブ（この例ではランダム・ワークロード）指標の正味変化（プラスまたはマイナス）を示している。すべてのケースにおいて、特に、使用可能容量手法の場合に、アイドル指標の値に正味での向上が見られることに注意されたい。アクティブ指標の結果にはばらつきが多い。物理容量削減（実際のドライブ数が削減されるため）の場合は、アクティブ性能および指標が低下（2 例）または向上（1 例）している。すべての使用可能容量の例において、アクティブ性能指標は同じか向上している。

| COM Type | Thin Provisioning | TP + Data Dedup | Data Compression |
|---|---|---|---|
| Baseline Config | Raid 5 (3+1 Data + Parity) | Full Provisioned No Dedup | Uncompressed |
| Baseline Capacity (GB) | 54,000 (160 – 450 GB disks) | 54,000 (160 – 450 GB disks) | 36,000 (160 – 450 GB disks) |
| Baseline Power (W) | 2360(I) 3134(B) | 2360(I) 3134(B) | 2360(I) 3134(B) |
| Baseline (GB/W) | 22.88 | 22.88 | 15.25 |
| Baseline Perf. (IOPs) | 25,250 | 25,250 | 36,010 |
| Baseline (IO/W) | 8.06 | 8.06 | 11.49 |
| Optimized Config | Thin Provisioning (50% written) | TP+Deduped Patterns (2:1) | Compressed (2:1) |
| Optimized Capacity (GB) | 54,000 (50% written) | 54,000 (50% written-dedup 2:1) | 36,000 (80 – 450 GB disks) |
| Opt. Power (W) | 1464(I) 2030(B) (80 disks) | 1016(I) 1479(B) (40 disks) | 1464(I) 2030(B) |
| Optimized (GB/W) | 36.88 | 53.15 | 24.59 |
| Delta (GB/W) | +61% | +132% | +61% |
| Opt. Perf (IOPs) | 13,560 | 8,580 (2:1 Dedup Ratio) | 38,280 |
| Optimized (IO/W) | 6.68 | 5.81 | 18.86 |
| Delta (IO/W) | (17)% | (28%) | +64% |

Table 6: Active COM Examples (70/30 R/W Random Workload), Based on Physical Capacity Reduction[5]

[5]出典：『Software Defined Storage Energy Efficiency Features』、The Green Grid Forum 2016

| COM Type | Thin Provisioning | TP + Data Dedup | Data Compression |
|---|---|---|---|
| Baseline Config | Raid 5 (3+1 Data + Parity) | Full Provisioned No Dedup | Uncompressed |
| Baseline Capacity (GB) | 54,000 (160 – 450 GB disks) | 54,000 (160 – 450 GB disks) | 36,000 (160 – 450 GB disks) |
| Baseline Power (W) | 2360(I) 3134(B) | 2360(I) 3134(B) | 2360(I) 3134(B) |
| Baseline (GB/W) | 22.88 | 22.88 | 15.25 |
| Baseline Perf. (IOPs) | 25,250 | 25,250 | 36,010 |
| Baseline (IO/W) | 8.06 | 8.06 | 11.49 |
| Optimized Config | Thin Provisioning (50% written) | TP+Deduped Patterns (2:1) | Compressed (2:1) |
| Optimized Capacity (GB) | 108,000 | 216,000 (50% written, 2:1) | 72,000 (2:1 Comp Ratio) |
| Opt. Power (W) | 2360(I) 3134(B) | 2360(I) 3134(B) | 2360(I) 3134(B) |
| Optimized (GB/W) | 45.76 | 91.52 | 30.50 |
| Delta (GB/W) | +100% | +300% | +100% |
| Opt. Perf (IOPs) | 25,250 | 32,825 (2:1 Dedup Ratio) | 72,020 (2:1 Comp Ratio) |
| Optimized (IO/W) | 8.06 | 10.47 | 22.98 |
| Delta (IO/W) | 0% | +30% | +100% |

Table 7: Active COM Examples (70/30 R/W, Random Workload) Based on Usable Capacity Optimizations[6]

物理容量削減手法を使用した場合は、ドライブ数が減ることによる電力節約を容易に特定できる。ただし、存在する物理ドライブが減るため、性能指標（IOPs など）とアクティブ指標が下がる可能性がある。加えて、実際のストレージ・システムでは、運用中ではなく、初期購入時またはセットアップ中にのみ物理ドライブ削減が行われる。実効（または仮想）容量最適化手法を使用した場合は、実際の物理容量が事実上増加する。ただし、現在の SNIA Emerald™仕様では、実効容量の改善ではなく、アイドル状態指標の計算時の分子として物理容量が使用されるため、この容量最適化手法 (COM)のメリットは反映されない。アイドル状態指標は、分子内の物理容量が使用可能または仮想容量に置き換えられることによって大幅に向上する。容量最適化手法 (COM)のメリットを比較するための推奨される定量化可能な手法を導出するにはさらなる調査が必要である。

複数の容量最適化手法 (COM)の同時運用の実現可能性とメリットは、実装と構成に依存する。容量最適化手法 (COM)機能の運用を可能にするアレイ・リソースによっては、最小限の相互作用が生じる場合がある。例えば、コントローラ経由の重複排除、IO スタックの「バックエンド」上にある SSD 内で提供される圧縮機能、およびシン・プロビジョニング機能を有効にする製品を考えてみる。これらのそれぞれが、ストレージ・サブシステムによって提供される全体的な使用可能（論理）容量に貢献できる。理想的なケースでは、圧縮率とシン・オーバープロビジョニング率の両方が 2 : 1 であれば、アプリケーションは論理空間全体を（あたかも物理的に存在しているかのように）利用することができる。重複排除は、圧縮では不可能なより大きなデータの「チャンク」を重複排除することにより圧縮率を高めることができる。最後に、提出者が有効にする容量最適化手法 (COM)の数と選択は提出者次第である。それらのいずれかが性能上の問題や干渉を引き起こす場合は、それらを評価に含めない選択がなされるかもしれない。

連続運転（24 時間無休）が不可欠のエンタープライズ・データ・センター環境では、アクティブ指標がアイドル容量より重視され、容量最適化手法(COM)は最適なメリットを提供するように選択的に適用する必要がある。

[6]出典：『Software Defined Storage Energy Efficiency Features』、The Green Grid Forum 2016

## 8. ENERGY STAR テストの代用となる構成とアプローチ

データ・センター・ストレージは、ENERGY STAR 用に特徴付けるには難しいカテゴリである。規模の範囲は十数台のドライブから数百台のドライブまでに及ぶ。ドライブ自体には様々なフォーム・ファクター、回転数、および容量があり、購入者のワークロードと性能期待に応じて高度に個別の構成に展開できる。技術的変化が急速な進化を続けており、運用環境と実際のドライブの両方に付加価値機能をもたらしている。同時に、システム・ベンダーとコンポーネント製造業者の数が極端に多くないため、複数のストレージ会社の製品ラインで多くのハードウェア・コンポーネントが共通している。

この複雑さがデータ・センター・ストレージ V1.0 環境に次のような結果をもたらしている。

- ENERGY STAR 認証のために多様なシステム構成を認定するには、製品の特性評価と資格証明に多くの労力をかける必要がある。また、この目的のために使用する設備にも多額の投資が必要である。リソース投資は、会社がおそらく製品ラインのテストの ROI を判断する必要が出てくるほど大規模になる。制限付きシステム構成の認定は、ベンダーが個別のシステム構成に認定ラベルを貼ることができない、ENERGY STAR 製品の普及率が下がる、および顧客を失望させるという危険が伴う。
- 仕様に記載されているテスト・プロトコルは現実世界のシステム構成を反映しておらず、現実世界のシステム構成を包含する認定枠を構築するために必要な努力は膨大なものになる。そのような努力をしたとしても、顧客が目的とするシステム構成に合った代表的なシステム構成が QPL 内に見つかる可能性は極めて低い。
  - データ成長の余地のないシステムを購入する顧客はあまりいない。最適点の周りのガードバンドは、（特にシン・プロビジョニングが展開されている場合には）エントリ構成を捕捉する上で不十分である。エントリ・ポイントを下げるには、特別なテストを実施しなければ、システムが ENERGY STAR 認定であると主張することはできない。
  - 1 つの HDD タイプしか搭載していないシステムを購入する顧客はあまりいない。何かの決まった比率と全く同一のタイプ構成で購入する顧客はさらに少ない。混合ドライブ・システムに対応したルールは、簡単ではない。
- テスト参加の壁が高過ぎて一部のベンダーが参加できなかったり、「対象範囲内」の製品のごく一部しかテストされなかったりするだろう。
- 特定の分類内のシステムを比較する時にしきい値を抽出するために使用できるデータが非常に少ない。
- これまでの分析に基づくと、意味のある比較を行うためにはカテゴリ内の区別をさらに細かくする必要があるように見える。
  - 分類カテゴリごとのドライブ・タイプ
  - 分類とドライブ・タイプごとのワークロード区分

ENERGY STAR の目標は、最適に動作する製品を認識し、継続的にエネルギー効率の向上を奨励することである。この目標に沿って、バージョン 2 への 2 本立てのアプローチを提案する。

1. QPL に含めるための基準の厳密さを高める。
2. 既存のベンダーと新しいベンダーの両方が幅広く参加できるようにテスト計画を見直す。

**基準の変更**

実装しきい値ではなく、包含基準を厳密にすることがエネルギー効率の向上を促進すると信じている。考えられる変更を以下に示す。

- 製品は ASHRAE A2 レベル以上の運用に適合する必要がある。
- PSU は、複数出力電源付きの製品の場合は 80 + 「ゴールド」以上に、単一出力電源付きの製品の場合は 80 + 「プラチナ」以上にする必要がある。
- オンライン 3 とオンライン 4 では、提供する COM の最小数を 1 つ増やす必要がある。
- 吸入温度の報告と時間の刻印を必須機能にする必要がある。

**テストの変更**

システムの特性評価プロセスの実行は複雑で膨大なリソースを必要とする。システムが大量のドライブ数（200 ドライブを超える）に対応するためには、十分な材料を組み立てる必要があり、最適点を特定するために必要なワークロードを駆動するのに十分なサーバをテスト装置に含める必要がある。各ドライブ・タイプの特性評価には数週間かかることが多い。その後、認定ラボでの最適点のテストにさらに時間が必要である。一般的な期間は、特性評価の開始から認定テスト結果の完了まで 4〜8 週間である。これでは手間がかかり過ぎて、資格があるのに認定されない製品が増えてしまう。

オンライン 4 のトランザクション・システムとシーケンシャル・システムに対する前記のデータ(Figure 26、Figure 27)において、3 つのテスト構成にわたるトランザクション・テスト点の違いが 5〜10%であり、3 つの構成にわたるシーケンシャル・テスト点のばらつきは少し大きいものの妥当な範囲に留まっていることから、単一の「最適な」構成で特定のワークロード・タイプの望ましいドライブ・タイプをテストする簡易テスト・アプローチの検討を強く推奨する。これにより、会社は、製品の単一構成をテストしてテスト負担を減らしながら、SNIA Emerald™テスト・ワークロードに対するストレージ製品のアクティブ性能/電力指標の評価をデータ・センター・オペレーターに提供できる。アクティブ指標値とアイドル状態指標値は報告する必要があるが、アクティブ効率しきい値とアイドル状態しきい値は設定する必要がない。これは、このホワイトペーパーで提供する分析がサーバの場合のようなしきい値設定を可能にする性能電力値の明確な傾向が見られないことを示しているためである。可能性のあるすべての構成は、対象とする製品ファミリー内のストレージ・デバイス数の範囲にわたって性能指標のばらつきが限定的であることを認識して、単一テスト値に基づいた認定が行われるべきである。

ストレージ製品のエネルギー効率を評価するためには、いかなる形でもアイドル状態を使用しないことが特に重要である。Table 4 と Table 5 にデータの分析の詳細を示したように、GB 容量が大きく、回転数の低いドライブを搭載したシステムで高いアイドル状態値またはしきい値が見られる。アイドル状態は、トランザクション・ワークロードとシーケンシャル・ワークロードのアクティブ指標に対して最適に動作するシステムを示しているわけではない。ストレージ・デバイスに対してすばやくデータを読み書きする能力は、容量の小ささ（読み書き位置をすばやく検索できる）と回転数の高さ（ストレージ・デバイス上の読み書き位置への移動時間を短縮できる）に依存する。アイドル状態は、性能とは全く無関係であり、特定のストレージ・システムが動作するように設計および設定される作業を最も効率良くこなすシステムを特定しない。

ストレージのエネルギー効率基準とテスト・アプローチに対するこれらの提案は、市場参入または調達プログラム用のストレージ製品エネルギー効率要件またはしきい値の設定を検討しているすべての地域に幅広く関係する。

## 9. 結論と提案

### 結論

あらゆる分類にわたって SNIA Emerald™測定データを使用して確実なエネルギー効率しきい値を定義するために十分なデータは、存在しない。特定の分類内では、指標をドライブ・タイプで分ける必要がある。ドライブ・タイプで分けると、しきい値を設定するためのデータが不足する。ただし、これまでのデータが一部の興味深い特性を示している。

Ready Idle 指標：

- Ready Idle 指標は、Hot Band 指標、Random Read/Write 指標、Sequential Read/Write 指標のいずれとも相関関係がない。特に、Ready Idle 指標は Hot Band 指標や Random Read/Write 指標と逆相関を示すことが多い。
- 特定の回転数のハード・ディスク・ドライブ (HDD)の Ready Idle 指標はドライブ容量に直接関係する。Ready Idle 指標を高めるにはより大きい容量のドライブに変更する必要があるだけである。ドライブ容量は定期的に増加するため、次の容量がリリースされるまで待てば容量を増やすことができる。特定のドライブ・ファミリーでは、電力増加が容量増加に対して無視できるほどである。
- SSD のアイドル電力は、容量と「アイドル」中にデバイスによって実行される活動量に左右される。一般的に、アイドル活動はデバイスのメンテナンス・タスクに関連付けられているため、「アイドル」中の活動（ウェアレベリングやハウスキーピングなど）が多いほど、デバイスの信頼性は高くなる。アイドル電力に関して SSD にペナルティを与えれば、デバイスの信頼性の低下につながりかねない。

ドライブ・タイプ：

- 電力が多く容量が小さいドライブは Hot Band 指標、Random Read/Write 指標が高くなる傾向がある。
- SSD は、Hot Band 指標、Random Read/Write 指標または Sequential Read 指標に対して高い性能を示すが、Sequential Write 指標に対してはそうでもない。
- 電力が少なく容量が大きいドライブは Sequential Read/Write 指標が高くなる傾向がある。分析したデータにおいては、10 Krpm ドライブは 7.2 Krpm ドライブより良い Sequential Read/Write 指標を示す。
- 電力が少なく容量が大きいドライブは必ず Ready Idle 指標が高い。
- 一般的に、各ドライブ・タイプのコストとのトレードオフを考慮すべきである。SSD が GB あたり最も高価で、高回転・小容量のドライブが次に高価で、低回転大容量のドライブが最も安価である。

同一製品ファミリーの中の最小、最適、および最大ドライブ数に関するデータのばらつきは、Hot Band 指標、Random Read/Write 指標の 10%未満、Sequential Read/Write 指標の 20%未満である。こ

れに基づくと、最小/最大テストを除外すべきという結論になる。同じワークロードに対して複数のドライブ・タイプを比較すると、必ず、1 つのドライブ・タイプが有利になる。公平に比較するためには、ドライブ・タイプとワークロードを同じにする必要がある。これにより、標準を設定するためのカテゴリの数が増えることになる。

**提案**

上記とこのペーパーの全体で示したデータに基づき、ENREGY STAR の次のバージョンで以下を使用することを提案する。

**基準の変更：**

- 製品は ASHRAE A2 レベル以上の運用に適合する必要がある。
- PSU は「ゴールド」以上にする必要がある。
- オンライン 3 製品とオンライン 4 製品では、COM の最小数を 1 つ増やす必要がある。
- 吸入温度の報告と時間の刻印を必須にする必要がある。

**テスト要件：**

- システムのベスト・フット・フォワード(最適値)をテストし、1 つのワークロードの構成の開示と一緒に結果を投稿する。
  - 製品ファミリー全体に関して報告されたワークロードに対して ENREGY STAR ステータスを付与する。
  - 製品ファミリーはすべてのドライブ・タイプ、容量、およびドライブ数を含むテスト対象コントローラである。

**オプション・テスト：**

- 追加の ENREGY STAR 投稿を取得するために、追加の構成および/またはワークロードをテストすることができる。

これらの提案は、市場参入または調達目的で検討されているすべてのストレージ製品エネルギー効率要件またはプログラムに関係する。