

第 4 世代 Intel® Xeon® スケーラブル・プロセッサ 上でのメディア処理の基本チューニング・ガイド

この記事は、Intel® デベロッパー・ゾーンに公開されている「[Media Processing Basics Tuning Guide on 4th Gen Intel® Xeon® Scalable Processors](#)」の日本語参考訳です。原文は更新される可能性があります。原文と翻訳文の内容が異なる場合は原文を優先してください。

はじめに

本ガイドは、メディア処理に精通しているユーザー向けです。一般的なメディア処理において、合理的なベースライン・パフォーマンスを達成するハードウェアとソフトウェアの設定に関する推奨事項を示します。メディア処理は複雑であり、最適なパフォーマンスを達成するには、本ガイドで紹介する推奨事項以外にも検討が必要なことがあります。

第 4 世代 Intel® Xeon® スケーラブル・プロセッサは、AI、暗号化、HPC、ストレージ、データベース・システム、およびネットワーク向けに改善されたアーキテクチャーと内蔵アクセラレーションにより、ワークロードに最適化されたパフォーマンスを提供します。また、独自のセキュリティ技術により、オンプレミスやクラウド上のデータ保護を支援します。

基本的なメディア処理に直接恩恵をもたらす改善として、コア数の増加、DDR5 によるメモリー・パフォーマンス、大容量キャッシュなどがあります。

アプリケーションにおいて、基本的なメディア処理が、人工知能、コンテンツ配信、リアルタイム・ストリーミングなど、さまざまな機能と統合されることは珍しくありません。第 4 世代 Intel® Xeon® スケーラブル・プロセッサの主な機能は次のとおりです。

- AI、HPC、ネットワーク、セキュリティ、ストレージ、アナリティクス向けの新しい内蔵アクセラレーター
- Intel® ウルトラ・パス・インターコネクト (Intel® UPI)
- Intel® スピード・セレクト・テクノロジー (Intel® SST)
- ハードウェアで強化されたセキュリティ
- 新しいフレックスバス I/O インターフェイス (PCIe* 5.0 + CXL)
- 新しいフレキシブル I/O インターフェイス—最大 20 HSIO レーン (PCI 3.0)
- Intel® UPI 2.0 によるマルチソケット帯域幅の増加 (最大 16GT/秒)
- Intel® データ・ストリーミング・アクセラレーター

本ガイドは、ハードウェア、ファームウェア、ソフトウェアをカバーしています。メモリー構成のように、調整方法が単純なものもあれば、スケーリング・ガバナナーのように、BOIS 設定やオペレーティング・システム API など、さまざまな方法で調整できるものもあります。ソリューション・スタックの上位のレベルが、スタックの下位レベルで行われた設定を変更することは珍しくありません。意図したチューニング設定が実行中に有効であることを確認するため、インストルメンテーションを使用して、実行時に重要なパラメーターを取得することを推奨します。実行時の設定が指定したチューニング設定と異なる場合、下流のファームウェアやソフトウェアがチューニング・パラメーターを変更している可能性があります。

リファレンス・ワークロード

インテルは、一般的なメディア処理のリファレンスとして、FFmpeg ベースのメディア・トランスコード・ベンチマークを採用しています。このガイドで推奨するチューニングは、ベンチマークの集計結果全体にわたって良好なパフォーマンスを提供することを目的としています。インテルはベンチマークのライセンスを付与していませんが、以下のガイダンスに従って再現できます。

ユースケース

このベンチマークは、以下の表に示すように、4 つのコーデックの 24 のユースケースに対して、ファイルベースの 1 つの入カストリームと 1 つの出カストリームを実装します。

コーデック	入力解像度	Xcode* の種類	プリセット	ISA	GOP サイズ (秒)	エンコードするフレーム	出力解像度 (入力に一致)	出力 FPS (入力に一致)	出力ビットレート (Mb/秒)	最大ビットレート (Mb/秒)	バッファサイズ (Mb)	プロファイル	プレビデオスイッチ	その他のスイッチ	エンコーダーパラメーター
svt av1	FHD	1:1	5	AVX2	2	すべて	FHD	60	4	12	16	n/a		-rc 1 -g 119-sc_detection 0	
svt av1	FHD	1:1	5	AVX3	2	すべて	FHD	60	4	12	16	n/a		-rc 1 -g 119-sc_detection 0	
svt av1	FHD	1:1	8	AVX2	2	すべて	UHD4Kc	60	4	24	48	n/a		-rc 1 -g 119-sc_detection 0	
svt av1	FHD	1:1	8	AVX3	2	すべて	UHD4Kc	60	4	24	20	n/a		-rc 1 -g 119-sc_detection 0	
svt av1	FHD	1:1	12	AVX2	2	すべて	UHD4Kc	60	4	24	20	n/a		-rc 1 -g 119-sc_detection 0	
svt av1	FHD	1:1	12	AVX3	2	すべて	UHD4Kc	60	4	24	20	n/a		-rc 1 -g 119-sc_detection 0	
svt av1	UHD4Kc	1:1	8	AVX2	2	すべて	UHD4Kc	60	9	18	36	n/a		-rc 1 -g 119-sc_detection 0	
svt av1	UHD4Kc	1:1	8	AVX3	2	すべて	UHD4Kc	60	9	18	36	n/a		-rc 1 -g 119-sc_detection 0	
svt av1	UHD4Kc	1:1	12	AVX2	2	すべて	FHD	60	9	18	16	n/a		-rc 1 -g 119-sc_detection 0	
svt av1	UHD4Kc	1:1	12	AVX3	2	すべて	FHD	60	9	10	16	n/a		-rc 1 -g 119-sc_detection 0	
svt hevc	FHD	1:1	1	AVX2	2	すべて	UHD4Kc	60	5	18	20	Main		-rc 1 -g 119-sc_detection 0	
svt hevc	FHD	1:1	5	AVX2	2	すべて	UHD4Kc	60	5	18	36	Main		-rc 1 -g 119-sc_detection 0	
svt hevc	FHD	1:1	5	AVX3	2	すべて	UHD4Kc	60	5	18	36	Main		-rc 1 -g 119-sc_detection 0	
svt hevc	FHD	1:1	9	AVX2	2	すべて	UHD4Kc	60	5	18	36	Main		-rc 1 -g 119-sc_detection 0	
svt hevc	UHD4Kc	1:1	1	AVX3	2	すべて	FHD	60	12	10	36	Main10		-rc 1 -g 119-sc_detection 0	
svt hevc	UHD4Kc	1:1	5	AVX3	2	すべて	FHD	60	12	10	36	Main10		-rc 1 -g 119-sc_detection 0	

コーデック	入力解像度	Xcode* の種類	プリセット	ISA	GOP サイズ (秒)	エンコードするフレーム	出力解像度 (入力に一致)	出力 FPS (入力に一致)	出力ビットレート (Mb/秒)	最大ビットレート (Mb/秒)	バッファサイズ (Mb)	プロファイル	プレビデオスイッチ	その他のスイッチ	エンコーダー・パラメーター
svt hevc	UHD4Kc	1:1	9	AVX2	2	すべて	FHD	60	12	24	36	Main10		-rc 1 -g 119-sc_detection 0	
svt hevc	UHD4Kc	1:1	9	AVX3	2	すべて	FHD	60	12	10	36	Main10		-rc 1 -g 119-sc_detection 0	
x264	FHD	1:1	高速	AVX2	2	すべて	FHD	60	6	8	24	High		-tune=psnr	keyint=120;min-keyint=120:sliced-threads=0;scene-cut=0;threads=4
x264	FHD	1:1	普通	AVX2	2	すべて	FHD	60	6	8	24	High		-tune=psnr	keyint=120;min-keyint=120:sliced-threads=0;scene-cut=0;threads=4
x264	FHD	1:1	低速	AVX2	2	すべて	FHD	60	5	8	48	High		-tune=psnr	keyint=240;-min-keyint=240:sliced-threads=0;scene-cut=0;threads=8
x265	FHD	1:1	普通	AVX3	2	すべて	FHD	60	5	8	48	Main		-tune=psnr	keyint=120,min-keyint=120;pools=4
x265	FHD	1:1	普通	AVX2	2	すべて	FHD	60	5	8	48	Main		-tune=psnr	keyint=120,min-keyint=120;pools=4
x265	UHD4Kc	1:1	低速	AVX2	2	すべて	FHD	60	12	8	48	Main10		-tune=psnr	keyint=240;min-keyint=240;pools=8

入力ファイル

入力ファイルには、アニメーション、ライブ、ゲームなどさまざまなタイプのビデオコンテンツから 10 シーンが集約されています。各シーンには先頭にキーフレームがあり、長さは 240 フレーム、つまり 60fps で 4 秒です。入力ビデオ全体では 2400 フレーム、つまり 60fps で 40 秒です。

入力ファイルは FHD とコンシューマー向け 4K 解像度でレンダリングされ、パフォーマンス測定のため、ユースケースの出力解像度とマッチングされます。

CPU コアのロード手法

正確な結果を得るには、スケジューラーをスラッシングさせることなく、プロセッサ利用率を高める (90% 以上に) するように、コアをロードすることが重要です。インテルでは、コア数に応じて FFmpeg インスタンスをディスパッチしています。以下の表にまとめます。

コーデック	解像度	MSO	プリセット	FFMPEG インスタンス	スレッド/エンコード
x.264	FHD	1:1	低速	Ceiling(Logical Cores/4)	8†
x.264	FHD	1:1	普通	Ceiling(Logical Cores Div/2)	8†
x.264	FHD	1:1	高速	Ceiling(Logical Cores Div/2)	8†
x.265	UHD4Kc	1:1	低速	Ceiling(Logical Cores/16)	32†
x.265	FHD	1:1	普通	Ceiling(Logical Cores Div 4)	16†
svt hevc	FHD	1:1	1	Ceiling(Logical Cores/8)	‡
svt hevc	FHD	1:1	5	Ceiling(Logical Cores/8)	‡
svt hevc	FHD	1:1	9	Ceiling(Logical Cores/4)	‡
svt hevc	UHD4Kc	1:1	1	Ceiling(Logical Cores/12)	‡
svt hevc	UHD4Kc	1:1	5	Ceiling(Logical Cores/12)	‡
svt hevc	UHD4Kc	1:1	9	Ceiling(Logical Cores/12)	‡
svt av1	FHD	1:1	12	Ceiling(Logical Cores/8)	‡
svt av1	FHD	1:1	8	Ceiling(Logical Cores/8)	‡
svt av1	FHD	1:1	5	Ceiling(Logical Cores/8)	‡
svt av1	UHD4K	1:1	12	Ceiling(Logical Cores/12)	‡
svt av1	UHD4K	1:1	8	Ceiling(Logical Cores/12)	‡

† FFmpeg コマンドラインのエンコーダー・パラメーターで設定
‡ コーデックにより自動的に設定される

チューニング

プラットフォームの選択に関する考察

最大メモリー速度

第 4 世代インテル® Xeon® スケーラブル・プロセッサ・ファミリーは、DDR5 メモリーに対応していますが、最大速度は特定のプロセッサ SKU により異なります。最大速度 4000MT/s、4400MT/s、4800MT/s をサポートする SKU が用意されています。メディア・トランスコードは一般に、高速メモリーの恩恵を受けるため、最高のパフォーマンスを得るためには、4800MT/s をサポートする SKU を選択することを推奨します。

システム設定

以下のセクションでは、BIOS やオペレーティング・システムで設定可能なパラメーターについて説明します。推奨設定は、一般的なメディア処理のユースケースで良好なパフォーマンスを発揮します。適切な設定は、アプリケーションにより異なります。

重要: パフォーマンスの設定方法は、マシンにより異なります。メーカー、モデル、アーキテクチャ、および BIOS によって、大きく異なる可能性があります。SUT に精通していない場合は、十分な知識を持つパフォーマンス・エンジニアから指導を受けることを強く推奨します。

安全な既知のデフォルト

安全な出発点を確保するため、BIOS とホスト・オペレーティング・システムをデフォルト設定にリセットします。BIOS のリセットは、通常、BIOS サブシステムから行うことができます。BIOS の供給元に確認してください。オペレーティング・システムのリセットは、通常、オペレーティング・システムの配布物の一部として文書化されています。オペレーティング・システムの開発元に確認してください。

一般的なパラメーター

以下の表は、一般に広く認識されている 6 つのパラメーターと、その推奨設定および簡単な説明を示します。

チューニング・パラメーター	場所*	推奨設定	説明
電力とポリシー	BIOS	パフォーマンス	パフォーマンス向けにシステムを最適化します。
CPU 周波数ガバナナー	OS	パフォーマンス	CPU を利用可能な最高周波数に設定します。
ターボブースト	BIOS	有効	CPU が最大ターボ周波数を維持できるようにします。
C ステート	BIOS	無効	CPU が低電力状態へ移行するのを防ぎます。
アンコア周波数	BIOS	最小	コアへの電力供給を優先します。
ハイパースレディング	BIOS	有効	1 つの物理マイクロプロセッサが 2 つの論理マイクロプロセッサのように動作できるようにします。

*チューニング・パラメーターの設定メカニズムは、マシンにより異なります。場所は、一般的な場所であり、実際の場所は異なる可能性があります。詳細は、ハードウェアとソフトウェアの参考資料を確認してください。ワークロードをインストールして、実行時に実際のパフォーマンス・パラメーターを取得することを強く推奨します。

ホームレス・プリフェッチャー

ホームレス・プリフェッチャーを使用すると、デマンドミスを実際の中間レベルのキャッシュに早期にフェッチできます。ホームレス・プリフェッチャーは、エクストリーム・コア・カウント (XCC) と呼ばれる 4 タイルのダイでは無効にし、モノリシック・メディア・コア・カウント (MCC) ダイでは有効にする必要があります。

プロフェッショナルなメディア処理アプリケーションは、多くの場合、中程度のコア数の SKU でパフォーマンス電力比を向上します。

Sub-NUMA Cluster (SNC)

(必須ではありませんが、良い例です。)

SNC は、以前のプロセッサ・ファミリーに搭載されていたクラスターオンダイ (Cluster-On-Die) と同様の局所性のメリットを、COD のデメリットなしに提供する機能です。SNC は、ラストレベル・キャッシュ (LLC) をアドレス範囲に基づいて不連続のクラスターに分割し、各クラスターをシステム内のメモリー・コントローラーのサブセットにバインドします。SNC は LLC の平均レイテンシーを改善する、以前のプロセッサ・ファミリーにあった COD 機能の代替となるものです。

メモリー構成

メディア・トランスコード・ワークロードは、メモリーの速度と構成の影響を受けます。アーキテクチャーでサポートされている最速のメモリーを選択し、CPU コアと間のデータ移動距離を最小限に抑えるように各メモリーチャンネルを構成します。メモリーサイズは、エンコーダー、解像度、および品質のバッファー要件に対応できるサイズにする必要があります。詳細は、以降のセクションで説明します。

メモリー速度

前述のとおり、最高のパフォーマンスを達成するには、最大スピード 4800MT/s をサポートする第 4 世代 Intel® Xeon® スケーラブル・プロセッサ SKU を選択します。DDR5 DIMM が 4800MT/s 以上であることを確認します。4800MT/s よりも高速なメモリーでは、CPU がそれ以上高速化できないため、パフォーマンスは向上しません。4800MT/s よりも遅いメモリーでは、メモリーの速度に合わせてシステムの速度が低下します。

DIMM 構成

第 4 世代 Intel® Xeon® スケーラブル・プロセッサ (開発コード名 Sapphire Rapids) は、最大 4800MT/s の DDR5 をサポートする 8 チャンネルのメモリー・アーキテクチャーです。メモリー・コントローラーは、チャンネルあたり最大 2 スロットをサポートします。そのため、一般的なメインボードには、各 CPU に対して 16 個のメモリー DIMM スロット (CPU ごとに 8 チャンネル×チャンネルごとに 2 スロット=CPU ごとに 16 スロット) が搭載されています。FFmpeg メディア・ベンチマークのようなトランスコード・アプリケーションでは、コンピューターの各チャンネル (スロットではなく) にメモリーを搭載することが重要です。したがって、第 4 世代 Intel® Xeon® スケーラブル・プロセッサでは、CPU ごとに (少なくとも) 8 つの DIMM を装着する必要があります (つまり、1 つおきのスロットは空でも良い)。シングル CPU マシンでは、(少なくとも) 8 つの DDR5 DIMM が必要です。デュアルソケットのマシンでは、(少なくとも) 16 個の DDR5 DIMM が必要です。

メモリーサイズ

DIMM のサイズはもっと複雑です。一般的なガイダンスは、論理コアあたり少なくとも 2GB の空きメモリーを確保することです。Intel® ハイパースレッディング・テクノロジー (Intel® HT テクノロジー) を有効にすると、第 4 世代 Intel® Xeon® スケーラブル・プロセッサでは、各物理コアに対して 2 つの論理コアを利用できます。例えば、Intel® Xeon® Platinum 8468 プロセッサは、2 ソケットのアプリケーションを想定した 48 コアの CPU です。2S サーバーは 192 個の論理コアを提供します (2 ソケット * 48 物理コア/ソケット * 2 論理コア/物理コア)。論理コアあたり 2GB の空きメモリーを割り当てれば、4K トランスコードまでのほとんどのユースケースに対応できます。この場合、推奨する空きメモリー容量は、384GB (2GB/論理コア * 192 論理コア) です。

メモリー サイズに関する補足説明: 多くのアプリケーションは、論理コアあたり 2GB よりかなり少ないメモリーで十分です。ただし、アプリケーションによっては、より多くのメモリーを必要とする場合があります。ファームウェアとオペレーティング・システムでは、メモリーのオーバーヘッドが生じるため、メモリーは重要なコストドライバーです。最大のパフォーマンスを確保しながらコストを最小化するには、エンドツーエンド (E2E) プラットフォームの特定のインスタンスをプロファイルする必要があります。

ストレージ、ディスク構成、および設定

ファイルベースのトランスコードの場合、サーバークラスの SSD は適切な I/O パフォーマンスを提供します。ほかのアプリケーションでも RAM ディスクから恩恵を得られる可能性があります。

ネットワークの構成と設定

一般に、オフラインメディア処理アプリケーションには、ネットワーク帯域幅による制限がありません。ライブ・アプリケーションやビデオ制作などのユースケースには、さまざまな要件があります。

関連ツールと情報

チューニング・パラメーターの設定と変更には、さまざまな仕組みがあります。オペレーティング・システム、ツール、またはアプリケーションによって、システム起動時に設定されたパラメーターが変更されることがあります。ワークロード実行時の設定が意図したものと一致するように、インテル® システム・ヘルス・インスペクター (svr-info)¹ やインテル® Power Thermal Utility² を使用して、システム構成を照会することを推奨します。

フィードバック

皆様からのフィードバックをお待ちしております。本ガイドに関するご意見や、本ガイドで取り上げていないトピックについて関心がある場合は、[お知らせください](#) (英語)。

参考資料

CNDA を締結されている方のみご覧いただけます。

¹ <https://www.intel.com/content/www/us/en/secure/design/confidential/products-and-solutions/software-and-services/cloud-performance-optimization-tools.html?s=Newest> (英語)

² <https://www.intel.com/content/www/us/en/secure/content-details/615177/intel-power-thermal-utility-intel-ptu-user-s-guide.html?wapkw=ptu&DocID=615177> (英語)

製品および性能に関する情報

¹ 性能は、使用状況、構成、その他の要因によって異なります。詳細については、<http://www.intel.com/PerformanceIndex/> (英語) を参照してください。