

製品ソリューション概要

第4世代インテル® Xeon® スケーラブル・プロセッサ
ハイパフォーマンス・コンピューティング

intel®

ハイパフォーマンス・コンピューティングに 必要な高性能テクノロジー

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、
コアあたりの高いパフォーマンス、コア数の増加、I/Oとメモリーのサブシステム進化、
幅広い内蔵ハードウェア・アクセラレーターなどのメリットをHPCワークロードにもたらします。

AIとHPCのワークロードの統合が進むに伴い、従来のパフォーマンスの課題に新たな側面が加わりました。企業規模で行われる推論やトレーニングはどちらもシステムのリソースに大きな負荷がかかりますが、好ましいユーザー体験を提供するためにはこの負荷に効率的に対応できなければなりません。これらの要件は、ライフサイエンスや材料科学、製造、シミュレーション/モデリング、ファイナンスなどの分野で使われるHPCワークロードに対する従来のニーズや現在増大しつつあるニーズに追加されるものです。その結果、世界のHPC市場は年平均成長率7.7%で拡大し、2026年までには592億米ドルに達すると見込まれています。¹

2026年までの年平均成長率

7.7% 世界の
HPC市場の成長¹

2026年までに

592億米ドル 世界の
HPC支出¹

HPCシステムのバランスを調整

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、HPCワークロードで迅速な価値創出を実現する画期的なパフォーマンスを提供します。このプラットフォームは、1コアあたりのパフォーマンスがより高く、ソケットあたり最大60コアとシステムあたり2、4または8ソケットを備えた最新のアーキテクチャーとなっています。これは最大120スレッドのコアあたり密度に相当し、前世代と比較して50%の増加となります。

これらのコア数の増加とバランスをとるため、このプラットフォームではメモリーと入出力 (I/O) サブシステムも同時に高度化されています。DDR5メモリーは最大1.5倍の帯域幅とDDR4の速度を提供し、4800MT/秒で動作します。プラットフォームはソケットあたり80レーンのPCIe Gen 5に対応し、従来のプラットフォームと比較してI/Oが劇的に向上しました。高いファブリック帯域幅と取り付けられたアクセラレーターの効率性をサポートするCXL (Compute Express Link 1.1) も提供されています。

急速に成長する各種ワークロードのパフォーマンスを加速します。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、市場で販売されているあらゆるCPUの中で最も多くの内蔵アクセラレーターを搭載し、AI、データ分析、ネットワーキング、ストレージ、HPCのパフォーマンスを向上させます。例えば以下のワークロードに活用できます。

- ・ **オプション価格の設定:** 時間の制約が厳しい中での意思決定、多様な要件の複雑なアプリケーション、AIの使用増加に伴う市場の需要変化に対応。
- ・ **ライフサイエンスのアプリケーション:** シミュレーション精度を上げるためのモデルの改善や大規模な演算の実行により、発見の迅速化とリサーチの有効性向上を実現。
- ・ **コンピューターを使ったエンジニアリング:** コンピューターを使ったエンジニアリングのアプリケーションで結果を迅速に引き出し、コスト低減、製品の安全性や設計の向上、市場投入までの時間短縮を実現。

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、HPC向けのインテル® HPC エンジンなどの内蔵のハードウェア・アクセラレーターによりパフォーマンス実現の新たなパラダイムを生み出します。

intel

XEON

パフォーマンス・プルーフポイント

最大 **1.56倍** 28の一般的なHPCワークロードで前世代と比べて
GEOMEAN パフォーマンスが向上^{2,3}

インテル® HPC エンジン



パフォーマンス・プルーフポイント

最大 **1.68倍** LAMMPS ワークロードで前世代と比べて
 GEOMEAN パフォーマンスが向上^{2,3}

内蔵アクセラレーターによる高度な機能

ワークロードが複雑化し、コンピューティング・リソースへの負荷も増大していますが、特定の機能をCPUのコアからオフロードし、これらの実行リソースをビジネスに不可欠なワークロードのタスクに向けることも可能です。これら特定の機能には、AI、セキュリティ、共有ストレージ、ネットワークなどの機能が含まれます。

第4世代Intel® Xeon® スケーラブル・プロセッサのシリコンに直接内蔵されたハードウェア・アクセラレーターは、プラットフォーム内のデータの移動や処理を向上させる機能を提供しています。プロセッサに内蔵されているため、外付けのソリューションやコアで実行されるソフトウェアベースのソリューションと異なり、PCIeバスへ送られる時のレイテンシーが発生せず、その分だけ電力消費も抑えられます。これらの内蔵アクセラレーターを活用したユースケースでは、設備投資と運用コストの削減とともに、パフォーマンスの向上も実現されています。

- **パフォーマンス:** 専用のアクセラレーターが対象のワークロード向けにスループットを大幅に向上。
- **機器コスト:** 第4世代Intel® Xeon® スケーラブル・プロセッサにアクセラレーターが内蔵されているため、追加機器への投資が不要。
- **運用コスト:** 内蔵のアクセラレーターにより、機器のラックにコアを追加する必要性が低減するため、大幅な節電が実現可能。

Intel® Advanced Matrix Extensions (Intel® AMX) : ディープラーニングを加速

マシンラーニングはHPCワークロードの効率性と有効性を高めることが実証されています。Intel® AMXは、ディープラーニングのアルゴリズムの中心でテンソル形式の処理を高速化することで推論やトレーニングのパフォーマンスを飛躍的に向上させる内蔵のハードウェア・アクセラレーターです。このテクノロジーには、従来のもより大きなかたまりのデータを格納するコアあたり最大8の拡張可能な2Dレジスター・タイルのセットであるTILEと、TILEの最初の演算子となる行列積の命令セットであるTMUL (Tile Matrix Multiply) が含まれます。Intel® AMXは、ディープラーニングのソフトウェアが、一定時間内により多くの推論を完了、また、より迅速にソリューションを導き出せるようにすることで、価値実現までの時間を短縮します。

Intel® Advanced Vector Extensions 512 (Intel® AVX-512) : 最新のX86ベクトル命令セット

テクノロジーが何世代も経るごとに、ベクトル化はますます高度になり、より大きなデータセットをより迅速に計算できるようになりました。最新のX86ベクトル命令セットであるIntel® AVX-512は、前世代のベクトル処理能力を基盤として構築され、データ負荷の高いワークロードの処理を高速化します。HPCアプリケーションは、最も演算負荷の高いワークロード向けに、2つの512ビット融合積和 (FMA) ユニットを使用して、クロックサイクルごとに1秒当たり32個の倍精度浮動小数点演算と64個の単精度浮動小数点演算を512ビットベクトルにパックすることができます。また、8個の64ビット整数と16個の32ビット整数もパックでき、ビジネス・インテリジェンスを促進します。このテクノロジーにより、Intel® Advanced Vector Extensions 2 (Intel® AVX2) に比べ、レジスターの数とFMAユニットの幅が倍増します。

インテル® データ・ストリーミング・アクセラレーター (インテル® DSA) : ストリーミング・データの移動を最適化

データの移動と変換処理は、ストレージ、ネットワーキング、HPCの分析など、データ負荷の高いワークロードに極めて重要です。インテル® DSAは、大規模なデプロイでオーバーヘッドの原因となる最も一般的なデータ移動のタスクをオフロードすることで、これらの機能向けのパフォーマンスを向上させます。チェックサム、メモリの比較、チェックポイントリングなど、ほぼすべてのデータ移動処理を肩代わりすることで、インテルは、メモリ、ストレージ、ネットワーキング・サブシステムのデータの出し入れに係るオーバーヘッドからCPUコアを開放します。インテル® DSAは、接続されたすべてのメモリ、ストレージ、ネットワーク・デバイス、さらにCPU、メモリ、キャッシュを跨いでストリーミング・データの処理を最適化します。

インテル® クイックアシスト・テクノロジー (インテル® QAT) : 暗号化と圧縮の高速化

暗号化やデータ圧縮に係るオーバーヘッドの削減はクラスター全体のパフォーマンスの向上に重要な役割を果たせます。インテル® QATは、第4世代インテル® Xeon® スケーラブル・プロセッサにハードウェア・アクセラレーターとして内蔵され、移動中のデータの暗号化や復号を高速化し、データ圧縮を効率化します。この最新版のテクノロジーは、従来の各世代と比べ、暗号サイファ、セキュアハッシュ、パブリックキー暗号化、圧縮/展開のパフォーマンスを高速化します。インテル® QATは、プロセッサ・コアからこれらのタスクをオフロードすることで、ほかの作業のためにリソースを使えるようにして、全体のスループットを増加させます。インテル® QATは、いかなるインフラストラクチャーにおいても、また保存時、転送時、使用時のどの段階においても、重要なワークロード向けのパフォーマンスを犠牲にすることなくデータを保護するゼロトラスト・セキュリティー戦略を支えます。

主なテクノロジー

<p>最大60コア ソケットあたり 最大50%増加² システムあたり2、4または8ソケット</p>	<p>最大8チャンネル DDR5、最大4800MT/s メモリ帯域幅およびスピードが 最大50%向上²</p>	<p>最大80レーン ソケットあたりPCIe 5.0 I/O容量の増加²</p>
---	---	--

第4世代インテル® Xeon® スケーラブル・プロセッサは、コアあたりの浮動小数点パフォーマンスの向上と幅広い内蔵ハードウェア・アクセラレーターを提供

開発者のイネーブルメントおよびサポート

インテル® oneAPI ツールキットは、HPCソフトウェア・エコシステムへのインテルの長年の取り組みから生まれたもので、インテルのアーキテクチャー向けに最適化された高品質のソフトウェアの開発過程を効率化するコンパイラー、ライブラリー、パフォーマンス・ツールを提供しています。開発者が第4世代インテル® Xeon® スケーラブル・プロセッサに内蔵されたアクセラレーターを利用しようとする場合、オープンな標準ベースのソフトウェア開発スタックを備えたツールキットは導入の近道となります。開発者はインテル® oneAPI ツールキットを使用して、アクセラレーター内蔵のCPU、GPU、FPGAなど、インテルのあらゆるアーキテクチャーでパフォーマンスを高速化するコードを作成できます。



インテル® oneAPI ベース・ツールキット

高性能なデータ中心のアプリケーションを開発するためのコア・コンパイラー、ライブラリー (インテル® oneAPI マス・カーネル・ライブラリーを含む)、その他のツール



インテル® oneAPI HPC ツールキット

インテル® Fortran コンパイラー、OpenMP GPU オフロード、MPIによる拡張性



インテル® AIアナリティクス・ ツールキット

最適化されたフレームワークおよびPythonライブラリーでデータサイエンスやアナリティクスのパイプラインを高速化



インテル® oneAPI レンダリング・ツールキット

レンダリングやレイトレーシングのライブラリーにより高性能で再現度が高いビジュアル体験を実現

oneAPIに基づいたオープン・スタンダードのコード開発を行うことで、オープンソースのツール、API、ドライバーを含む大規模でオープンなエコシステムの恩恵を受けられます。その柔軟性により、新しいアーキテクチャーの導入が効率化され、エンジニアやプログラマーもコードの維持ではなく新しいコードの導入が可能になり、企業も複雑さ、コスト、時間の要件を低減し、新しいサービスやソリューションを市場に投入することが可能になります。

既存のインフラストラクチャーとの容易な統合

企業は、インテルの技術を熟知した実績のあるパートナーからなる膨大なエコシステムの活用により導入の時間を短縮できます。世界中のハードウェア・ベンダー、ソフトウェア・ベンダー、ソリューション・インテグレーターがインテル® Xeon® スケーラブル・プロセッサを搭載した製品を構築し、数多くの導入実績に裏付けされた豊富な選択肢と相互運用性を提供します。

パフォーマンス・プルーフポイント

VASP のパフォーマンスの向上^{2,3}

最大
1.61倍
GEOMEAN
第4世代インテル® Xeon®
スケーラブル・プロセッサと
前世代の比較

最大
2.01倍
GEOMEAN
インテル® Xeon® プロセッサ・マックス・
シリーズと2S 第3世代インテル® Xeon®
スケーラブル・プロセッサの比較

すべての分野のHPCユースケース向けに設計

第4世代インテル® Xeon® スケーラブル・プロセッサは、高パフォーマンス、DDR5を使用したメモリー帯域幅の向上、PCIe Gen 5およびCXL 1.1を使用したI/Oの高度化により、実環境での幅広いユースケースを加速します。開発者は、HPCアプリケーションが箱から出してすぐに高いパフォーマンスを出せるよう最適化されたインテルのトップレベルのソフトウェア・ライブラリーやコンパイラーで、より迅速にコードを作成できます。コードやモデルは、演算負荷の最も高いワークロード向けに、コアあたり2つのFMAユニットを備えたパワフルなインテル® AVX-512テクノロジーを活用できます。インテル® MPIライブラリーにより、ワークロードを複数のHPCクラスターに拡張することができます。インテル® Optane™ パーシステント・メモリーを追加すれば、メモリー容量が大きいほどメリットのある大規模な計算をサポートできます。

HPCワークロードをサポートするインテルのテクノロジーで可能性を広げる



帯域幅を最大化: 最新のインテル® Xeon® プロセッサ・マックス・シリーズは、モデリング、AI、HPC、データ分析など、メモリー中心のワークロード向けにボトルネックを解消し、DDR5のみを備えたプラットフォームと比較してパフォーマンスが4倍になります。これは高帯域幅メモリーとアクセラレーターをプロセッサ・パッケージに統合した最初のx86 CPUで、最大64GBのHBM2eを備えています。DDRへの依存の低減、最新のソフトウェア・ツール、コードの優れた再利用性により、総所有コスト(TCO)を改善します。



インパクトを最大化: フラッグシップ製品となるインテル® データセンター GPU マックス・シリーズは、エクサスケールの時代のAI、HPC、高度な分析ワークロードを高速化する、インテルの最も高度なIPとパッケージングのテクノロジーを使用します。インテル® X® HPCアーキテクチャーを基盤とし、GPU内で最も大容量かつ高帯域幅のキャッシュを提供します。SIMT/SIMDの柔軟性を備えたoneAPIオープン・エコシステムのサポートにより、GPUは高帯域幅メモリーを含む複数のIPイノベーションをパッケージに統合します。



マイクロ秒でデータにアクセス: DAOS(Distributed Asynchronous Object Storage)は、オープンソースのソフトウェア定義によるスケールアウト型オブジェクト・ストアで、高帯域幅、低レイテンシー、1秒ごとのI/O(IOPS)が高いストレージコンテナを単一のストレージ層でコスト効率よくHPCやAIのアプリケーションに提供します。DAOSは、構造化データセット、半構造化データセット、非構造化データセットをサポートすると同時に、従来の分散ストレージの制約を取り除きます。

詳細情報

www.intel.co.jp/xeon/scalable/

www.intel.co.jp/hpc/



- Intersect360 Research 「Total HPC Market Revenue Grew 5.2% to \$41.0 Billion in 2021, Says Intersect360 Research」 2022年5月20日。
<https://www.hpcwire.com/off-the-wire/total-hpc-market-revenue-grew-5-2-to-41-0-billion-in-2021-says-intersect360-research/> (英語)
- 2S 第3世代インテル® Xeon® スケーラブル・プロセッサーとの比較。
- [intel.com/processorclaims/](https://www.intel.com/processorclaims/) (英語): [4th Gen Intel® Xeon® Scalable Processors](https://www.intel.com/processorclaims/4th-gen-intel-xeon-scalable-processors/) をご覧ください。実際の結果は異なる場合があります。

通知と免責事項

アクセラレーターの有無は、SKUによって異なります。製品に関する追加情報に関しては、[インテル製品の仕様情報ページ](#)にアクセスしてください。

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、<https://www.intel.com/PerformanceIndex/>を参照してください。

性能の測定結果は構成情報に記載された日付時点のテストに基づくものです。また、公開中のすべてのアップデートが適用されているとは限りません。構成の詳細については、公開されている構成情報を参照してください。絶対的なセキュリティを提供できる製品またはコンポーネントはありません。

インテルは、サードパーティーのデータについて管理や監査を行っていません。ほかの情報を参考にして、正確さを評価してください。

実際のコストと結果は異なる場合があります。

インテルのテクノロジーを使用するには、対応したハードウェア、ソフトウェア、またはサービスの有効化が必要となる場合があります。

本資料に記載されているインテル製品に関する侵害行為または法的調査に関連して、本資料を使用または使用を促すことはできません。本資料を使用することにより、ユーザーは、インテルに対し、本資料で開示された内容を含む特許クレームで、その後作成したものについて、非独占的かつロイヤルティ無料の実施権を許諾することに同意することになります。

説明されている製品には、エラッタと呼ばれる設計上の不具合が含まれている可能性があり、製品が公開されている仕様とは異なる場合があります。現在確認済みのエラッタについては、インテルまでお問い合わせください。

© Intel Corporation. Intel, インテル, Intel ロゴ, その他のインテルの名称やロゴは、Intel Corporation またはその子会社の商標です。その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。