oneAPI

縛りのない、生産性とパフォーマンスに優れた
高速コンピューティング

oneAPI および インテル® ソフトウェア 開発ツール

インテル コーポレーション
ソフトウェア&先端技術部門 (SATG)
ソフトウェア製品 & エコシステム

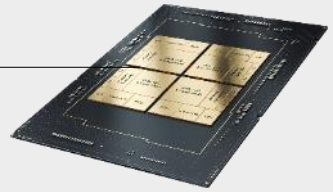
2023年11月

intel.

ここに記載されているすべての情報は、予告なく変更されることがあります。
インテルの最新の製品仕様およびロードマップをご希望の方は、インテルの担当者までお問い合わせください。

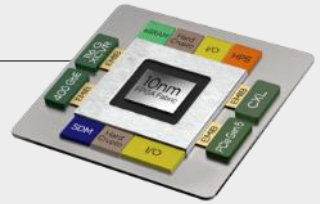
多様なハードウェア・アクセラレーションの課題を解決

CPU



GPU

FPGA



その他のアクセラレーター

開発者の 48% が 2 種類以上のプロセッサーやコアを使用する
ヘテロジニアス・システムをターゲットにしている[1]

開発者の課題: 複数のアーキテクチャー、ベンダー、プログラミング・モデル

oneAPI

オープンな標準ベースのマルチアーキテクチャー・プログラミング

ビデオ: oneAPI とは? 概要と利点 (英語)

1. 2022年6月に Evans Data が実施した「Global Development Survey Report 22.1」

oneAPI
仕様とオープンソース

最良の選択を行う

- ベンダー固有や独自のアーキテクチャーに代わる、ターゲット・アーキテクチャーの変更が容易な、オープンな選択肢を提供します。
- オープンな標準ベースのプログラミング (C++ with SYCL*) であり、ソフトウェアへの投資は将来の世代のハードウェアでも引き続き価値を保ちます。

パフォーマンス - ハードウェアの価値を最大限に引き出す

- CPU、GPU、FPGA、その他のアクセラレーターの最先端の機能をすべて活用し、パフォーマンスを最大限に引き出します。
- ドメイン固有の関数を高速化する強力なライブラリー。

生産性 - パフォーマンスの高いコードを素早く開発

- 1 つのプログラミング・モデル。CUDA* コードの SYCL* への移行など、既存のコードとの統合を容易にします。
- 馴染みのある C++ ベース。新しい言語を学ぶ必要はありません。
- Fortran、C/C++、OpenMP*、MPI、Python* (最適化された Python* ライブラリーの豊富なセットを含む) などの既存の HPC 標準規格と相互運用可能です。

詳細は、oneapi.io (英語) または https://uxlfoundation.org/ (英語) を参照してください。

oneAPI

マルチアーキテクチャーの高速コンピューティングの実現に向けた、ベンダーに依存しないソフトウェア・エコシステムを推進するオープンな業界イニシアチブ

現在は Linux Foundation が管理

UXL FOUNDATION
Unified Acceleration

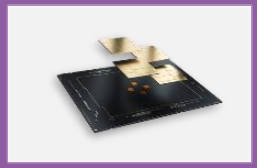
ミドルウェア/フレームワーク

PyTorch | TensorFlow | scikit learn | OpenVINO | NumPy | XGBOOST

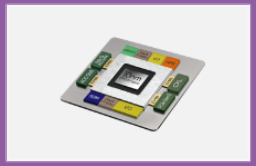
oneAPI 業界仕様

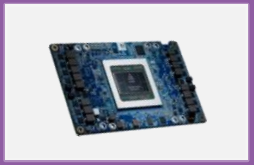
ダイレクト・プログラミング	API ベースのプログラミング		
SYCL* (C++)	演算 oneMKL	スレッド化 oneTBB	並列 STL oneDPL
	アナリティクス/ML oneDAL	DNN oneDNN	ML 通信 oneCCL

低水準ハードウェア・インターフェイス (oneAPI レベルゼロ)


CPU

GPU


FPGA


その他の
アクセラレーター

オープンでスケーラブルなアクセラレーションへの取り組み

開発者を独占的なエコシステムから解放

UXLFOUNDATION
Unified Acceleration

Linux Foundation が管理するマルチアーキテクチャーの高速コンピューティングの実現に向けた、ベンダーに依存しないソフトウェア・エコシステムを推進するオープンな業界イニシアチブ

創設メンバー: ARM、Fujitsu、Google Cloud、Imagination Tech、Intel、Qualcomm、Samsung、VMware

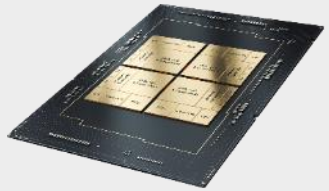
oneAPI 仕様 (oneAPI.io) が出発点

oneAPI

オープンな標準ベースのマルチアーキテクチャー・プログラミング

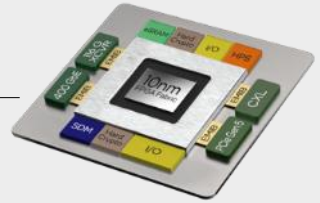
パフォーマンス | 生産性 | ベンダー依存からの解放

CPU



GPU

FPGA



その他のアクセラレーター

oneAPI

マルチアーキテクチャー、マルチベンダーのアクセラレーター・ソフトウェア・エコシステムを構築するオープンなイニシアチブとコミュニティー

現在は Linux Foundation が管理

UXL FOUNDATION
Unified Acceleration

定期的な仕様の更新

イニシアチブとテクノロジーの進歩

2020

oneAPI SYCL* 実装

インテルの CPU、GPU、FPGA をサポート

NVIDIA* GPU 用の SYCL*

PyTorch* が oneDNN によりモデルを高速化

2021

富岳が Arm に oneDNN を導入

ハイデルベルク大学が AMD の CPU と GPU に SYCL* を導入

NERSC とアルゴンヌが NVIDIA の GPU に SYCL* を導入

アルゴンヌとオークリッジ国立研究所が AMD の GPU に SYCL* を導入

Huawei CCE & Ascend が oneAPI を採用

2022

NVIDIA の GPU、AMD の CPU と GPU、Arm の CPU

インテルの CPU と GPU、NVIDIA と AMD の GPU 向けの **GROMACS** SYCL* コード

oneAPI コミュニティー・フォーラムとオープンガバナンスの設立

Julia* の oneAPI インターフェイス

TensorFlow* が oneDNN によりモデルを高速化

SYCL* のパフォーマンスが NVIDIA/AMD のネイティブシステム言語に匹敵

2023

NVIDIA および AMD 向け **Codeplay の oneAPI プラグイン**

Linux Foundation が oneAPI 仕様を定義する Unified Acceleration Foundation (英語) を発表

創設メンバー: ARM、Fujitsu、Google Cloud、Imagination Tech、Intel、Qualcomm、Samsung、VMware

Unified Acceleration (UXL) Foundation

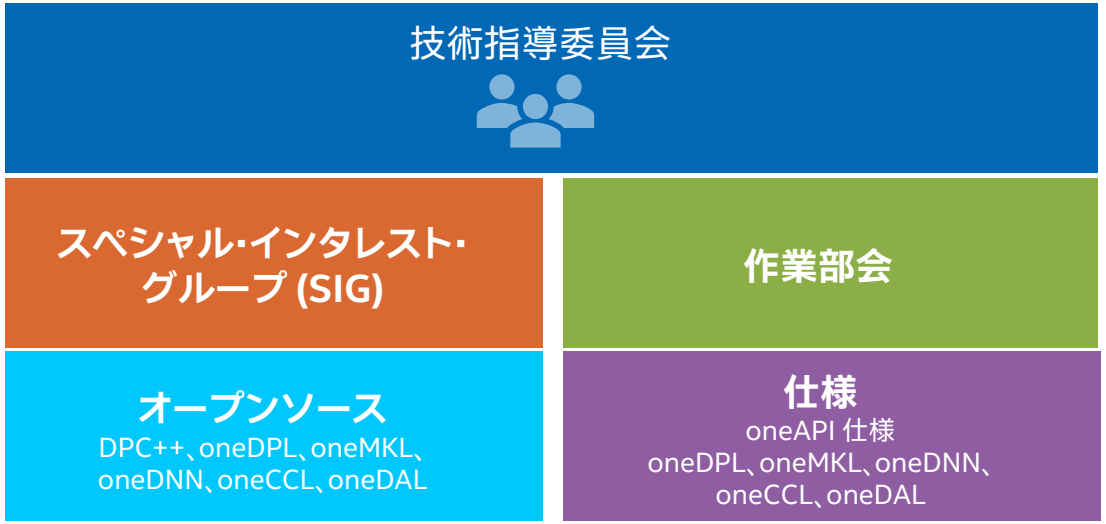
マルチアーキテクチャー、マルチベンダーのアクセラレーター・ソフトウェア・エコシステムを構築

- 管理: Linux Foundation の Joint Development Foundation
- 目的: ヘテロジニアス計算エコシステムをオープンスタンダードで統合する
- 出発点: oneAPI 仕様 (oneAPI.io)
- 目標: 幅広い業界の参加と貢献
- SIG: AI、ハードウェア、言語、数学
- 参加方法: SIG に参加
  - www.UXLFoundation.org (英語)

Steering Members

arm, FUJITSU, Google Cloud, Imagination, intel, Qualcomm, SAMSUNG, vmware

UXL Foundation の構造



Unified Acceleration (UXL) Foundation

2023年9月19日発表

THENEWSTACK

OneAPI Computing Aims to Ease Multi-Architecture Computing

OneAPI is all about using universal APIs for parallel application development whenever possible.

Sep 19th, 2023 10:11am by Steven J. Vaughan-Nichols

ITOpsTimes

Linux Foundation forms new group to drive common model for developing cross-platform applications

Latest News Published: September 19th, 2023 - Jenna Barron

VentureBeat

Intel CTO highlights open and secure advances for AI deployment

TECHZINE

Unified Acceleration Foundation (UXL) gives oneAPI a boost, and more governance

TC

The Unified Acceleration Foundation wants to create an open standard for accelerator programming

Frederic Lardinois

embedded

oneAPI motivates new foundation for open standard accelerated compute

September 19, 2023 Nitin Dahad

Unified Acceleration (UXL) Foundation wants to build largest open ecosystem for accelerated computing, using existing open standards, to foster a multi-architecture and multi-vendor programming platform for all accelerators.

THENEXTPLATFORM

Bring the benefits of the cloud into you network. Napatech IPUs powered by Intel®

THE NEW UXL FOUNDATION'S HAS A BOLD BLUEPRINT FOR OPEN ACCELERATION

phoronix

MAKE A POWERFUL CHOICE.

Intel oneAPI Initiative Evolves Into The Unified Acceleration "UXL" Foundation

Written by Michael Larabel in Software on 19 September 2023. Page 1 of 1. 2 Comments

Here's a surprise announcement I was briefed on last week and now made public by the Linux Foundation and Intel... The Linux Foundation is forming the Unified Acceleration (UXL) Foundation that is

siliconANGLE [the voice of enterprise and emerging tech]

Intel, Samsung and other chipmakers back new accelerator programming initiative

BY MARIA DEUTSCHER

Open Source Insider

Cross-architecture unification: Linux Foundation forms Unified Acceleration (UXL) Foundation

Adrian Bridgwater

Published: 20 September 2023 12:17

As part of the Linux Foundation's Open Source Summit 2023, the foundation has announced the formation of the Unified Acceleration (UXL) Foundation.

Forbes

Unified Acceleration Foundation To Broaden OneAPI's Use For Heterogeneous Programming

Steven Leibson Contributor

Tirias Research Contributor Group

Follow

insideHPC

Linux Foundation: Unified Acceleration Foundation Formed for Open Accelerated Compute and Cross-Platform Performance

September 19, 2023 by staff | Leave a Comment

Runtime

The race for an Nvidia alternative

Tom Krazit

hackster.io

The Unified Acceleration Foundation Aims to Boost ML, AI Workloads No Matter the Architecture

Cross-industry group including Arm, Intel, Qualcomm, and more looks to evolve oneAPI into an architecture-agnostic accelerator standard.

Gareth Halfacree

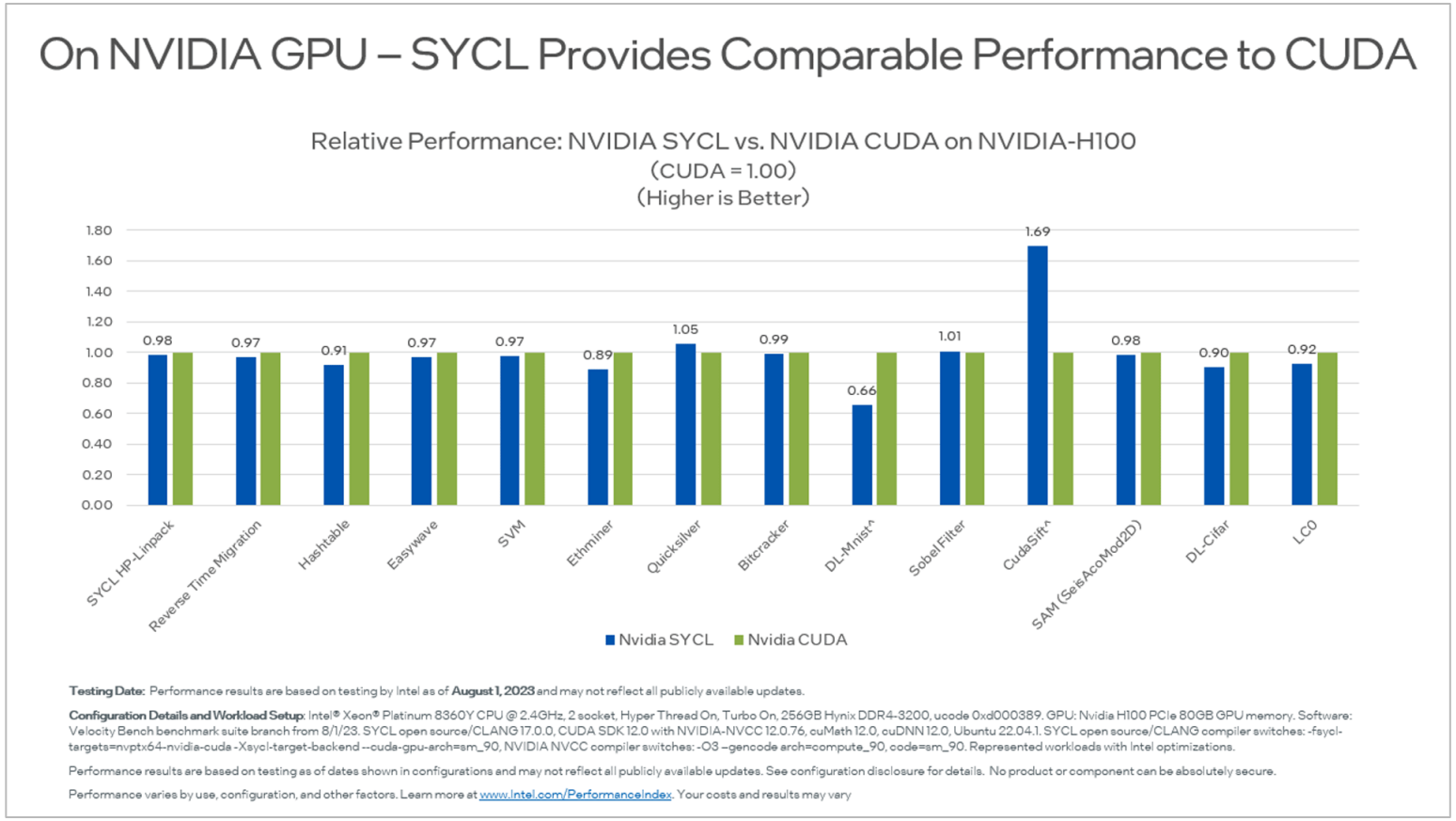
Linux-Foundation will einheitlichen Beschleuniger-Zugriff

記事のリンクは補足資料を参照してください。

SYCL* で高速化する

Khronos Group の標準仕様

- オープンで標準ベース
- マルチアーキテクチャー・パフォーマンス
- ベンダー依存からの解放
- NVIDIA* GPU 上のネイティブ CUDA* に匹敵するパフォーマンス
- 広く使用されている C++ 言語の拡張
- インテル® DPC++ 互換性ツール (英語) またはオープンソースの SYCLomatic (英語) による迅速なコードの移行



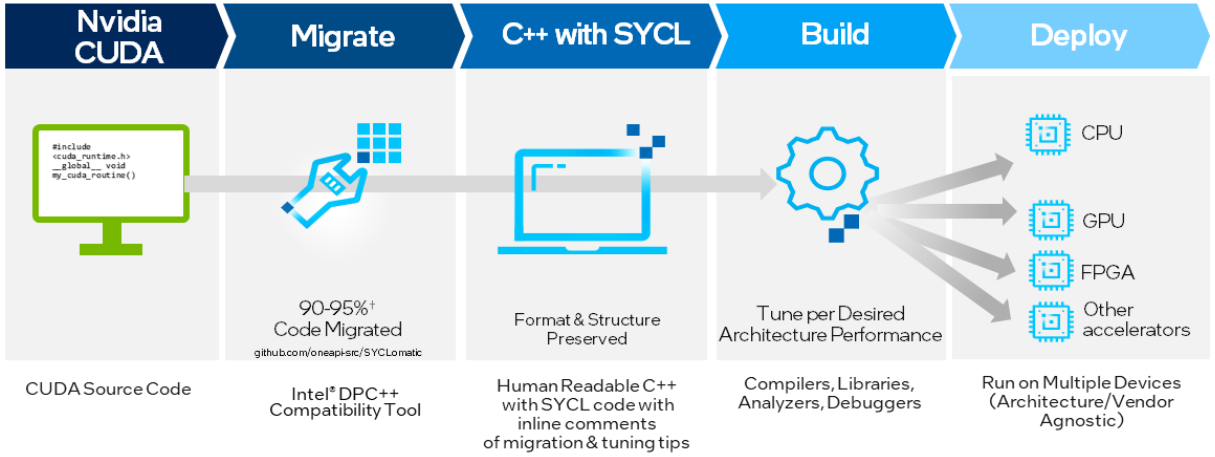
アーキテクチャー	インテル | NVIDIA | AMD CPU & GPU | RISC-V | ARM Mali* | PowerVR* | Xilinx

CUDA* から C++ with SYCL* への移行

アーキテクチャーごとに異なるコードベースを作成/保守する必要はない

CUDA* から
C++ with SYCL* への
コード移行ポータル

- 高速コンピューティング・ハードウェアを選択して、パフォーマンスの移植性を備えたコードを再利用
- 単一の C++ with SYCL* コードベースでマルチベンダーの複数のアーキテクチャーのアクセラレーター上で実行可能
- インテル® DPC++ 互換性ツールとオープンソースの SYCLomatic は、典型的な CUDA* アプリケーションの約 90～95%* を自動で SYCL* に移行
- 移行を完了し、パフォーマンスをチューニングするのに役立つコメントを生成
- チュートリアル、ベスト・プラクティス、サンプルコード、アプリケーション・カタログ、コミュニティー・サポートについては、CUDA* から C++ with SYCL* へのコード移行ポータルを参照



移行の成功事例

Argonne NATIONAL LABORATORY
THE UNIVERSITY OF TENNESSEE KNOXVILLE
Arcvideo
blender
inesc id lisboa
TÉCNICO LISBOA
FAST. FLEXIBLE. FREE. GROMACS
BittWare a molex company
THE UNIVERSITY OF UTAH
OLD DOMINION UNIVERSITY
SAMSUNG MEDISON
ZUSE INSTITUTE BERLIN
Stockholm University
KTH

[1]出典: 2023年3月現在のインテルによる推定。Rodinia、SHOC、PENNANT など、85 種類の HPC ベンチマークとサンプルの測定結果に基づいています。結果は異なることがあります。

NVIDIA* および AMD* 向け oneAPI プラグイン

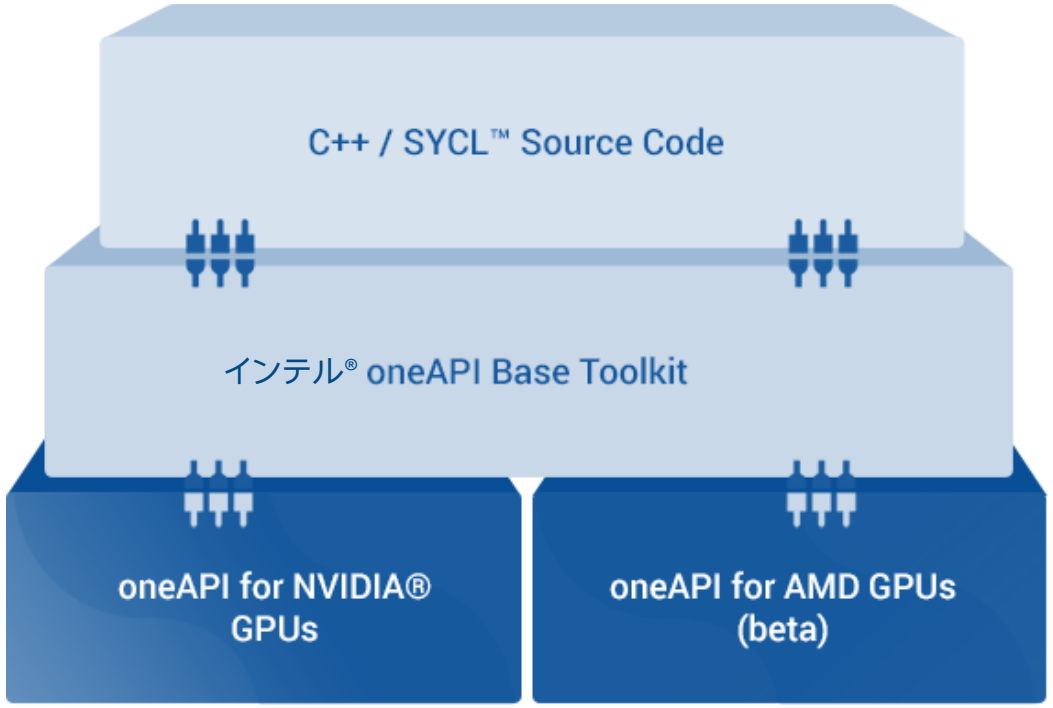
インテル® oneAPI ベース・ツールキットに Codeplay の NVIDIA* GPU と AMD* GPU サポートを追加

NVIDIA* GPU および AMD* GPU 向け oneAPI

- インテル® oneAPI DPC++/C++ コンパイラーへの
  バイナリープラグインを無料でダウンロード可能
- NVIDIA* GPU
- AMD* GPU (ベータ)
- ソースからのビルドは不要
- SYCL* 2020 の適合性とパフォーマンスに合わせて
  四半期ごとに更新される

優先サポート

- インテル、Codeplay、およびチャネルから入手可能
- インテル® DPC++/C++ コンパイラーに対する有効な
  インテルの優先サポートが必要
- インテルが受け付け、Codeplay がサポートを提供
- 古いバージョンのプラグインへアクセス可能



画像の出典: Codeplay Software Ltd.

NVIDIA* GPU プラグイン (英語)

AMD* GPU プラグイン (英語)

Codeplay のブログ (英語)

Codeplay のプレスリリース (英語)

oneAPI の採用実績

エンドユーザー

archanan
FOUNDRY MODO
BioDataAnalysis GmbH GeoEast
BENTLEY
CINESITE
LAIKA
Accrad Accelerated Radiology
kt
AUTODESK ARNOLD
中国石油集团东方地球物理勘探有限责任公司 BGP INC.,CHINA NATIONAL PETROLEUM CORPORATION
iOmniscient Autonomous
EURECOM
Brightskies
PHILIPS
大势智慧 DASPATIAL
WeBank
SAMSUNG MEDISON
TANGENT ANIMATION
allegro.ai
ILLUMINATION MACGUFF
Verizon*
MediaKind

国立研究所

Argonne NATIONAL LABORATORY
UT-BATTELLE Oak Ridge National Laboratory
CERN openlab
CINECA
Peraton Labs
Laboratório Nacional de Computação Científica
Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities
SANKHYA SUTRA LABS

ISV & OSV

codeplay
CANONICAL
CERTIFACE
ANACONDA
Ansys
MAX PLANCK COMPUTING & DATA FACILITY
MPCDF
AI SINGAPORE
sas
Red Hat OpenShift Data Science
KATANA GRAPH
CGG
AIBLE
CHAOZGROUP
MAXON
Ospirent
FOUNDRY
mercenaries engineering Guerilla
Hisense Medical 海信医疗
SENAI CIMATEC FIEB
SAP
E4 COMPUTER ENGINEERING
YUAN
GIGASPACES innovate with confidence
MEGH COMPUTING
KFBIO
Tech Mahindra
AsiaInfo 亚信科技
SUSE
vmware
UNITED IMAGING

OEM & SI

BittWare a molex company
Hewlett Packard Enterprise
DELL Technologies
Atos
Lenovo
HCL
rENIAC
MEGWARE

大学 & 研究機関

LOBACHEVSKY UNIVERSITY
TECHNION Israel Institute of Technology
UNIVERSITY OF CAMBRIDGE
中国科学院计算技术研究所 INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES
Ben-Gurion University of the Negev
Berkeley UNIVERSITY OF CALIFORNIA
UC DAVIS UNIVERSITY OF CALIFORNIA
THE UNIVERSITY OF TENNESSEE KNOXVILLE
UNIVERSIDAD COMPLUTENSE MADRID
OLD DOMINION UNIVERSITY
CtG
北京大学软件与微电子学院 School of Software & Microelectronics
THE UNIVERSITY OF UTAH
ILLINOIS
UNIVERSITY OF OREGON
TÉCNICO LISBOA
inesc id lisboa
ZIB
PURDUE UNIVERSITY Elmore Family School of Electrical and Computer Engineering
FACULTY OF MATHEMATICS AND PHYSICS Charles University
University College London
Durham University
University of Stuttgart Germany
IIT GOA
Indian Institutes of Technology Delhi / Kharagpur / Roorkee
SDSC SAN DIEGO SUPERCOMPUTER CENTER
सी डैक CDAC
UKRI Science and Technology Facilities Council Scientific Computing
ATLAS EXPERIMENT
TACC
UNIVERSIDAD DE MÁLAGA
HEIDELBERG UNIVERSITY COMPUTING CENTRE
Northern Illinois University
University of BRISTOL
Stockholm University
KTH
Indian Institute of Science Bangalore
Indian Institute of Science Education & Research Pune
IISER PUNE

CSP & フレームワーク

Microsoft Azure
Alibaba Cloud
Google Cloud
TensorFlow
Taboola
Tencent 腾讯
DataRobot
Baidu 百度
飞桨 PaddlePaddle
QCT
NAVER CLOVA
PyTorch

これらの組織は、クロスアーキテクチャー開発を実現する単一の統合プログラミング・モデルである oneAPI イニシアチブをサポートしています。
インテルの製品の購入または使用に関する契約を示すものではありません。

インテル® ソフトウェア開発ツール

柔軟で、オープンな、統合ソフトウェア・スタック – powered by oneAPI

intel. HPC TOOLKIT

- モデリングとシミュレーション
- 科学計算
- アナリティクス

インテルにより最適化された AI ソフトウェア・ツールとフレームワーク

- 大規模なデータ・アナリティクス: MODIN, pandas, NumPy, SciPy
- DL 推論とトレーニング: TensorFlow, PyTorch, OpenVINO, インテル® ニューラル・コンプレッサー
- 古典的な ML: scikit learn, dmlc XGBoost, python

intel. RENDERING TOOLKIT

- 高忠実度グラフィックス
- ビジュアル・コンピューティング
- レイトレーシング

intel. 1 oneAPI BASE TOOLKIT

ツール:	インテル® DPC++ 互換性ツール	インテル® VTune™ プロファイラー	インテル® Advisor	インテル® ディストリビューションの GDB	インテル® ディストリビューションの Python*		
パフォーマンス・ライブラリー:	インテル® oneMKL	インテル® oneDNN	インテル® oneDAL	インテル® oneCCL	インテル® oneTBB	インテル® oneDPL	インテル® MPI
ダイレクト・プログラミング:	C++ with SYCL*	C++	Fortran	OpenMP*			
コンパイラー:	インテル® Fortran コンパイラー	インテル® DPC++/C++ コンパイラー					

ハードウェア・インターフェイス – oneAPI レベルゼロ

CPU GPU FPGA

ダウンロード: www.xlsoft.com/jp/products/intel/oneapi
または インテル® デベロッパー・クラウドで実行

インテル® ソフトウェア開発ツール 2024.0

インテルの CPU と GPU 上でマルチアーキテクチャー・コンピューティングを高速化 – 2023年11月リリース

ハードウェアの選択/将来に対応したプログラミング

- インテル® ディストリビューションの Python*、Modin 向けインテル® ディストリビューション (英語)、XGBoost 向けインテル® オプティマイゼーション (英語)、インテル® ニューラル・コンプレッサー (英語)、インテル® レンダリング・ツールキットの新しい GPU サポートにより、AI、HPC、レンダリング・アクセラレーションを拡大
- インテル® DPC++/C++ コンパイラーに SYCL* 2020 をほぼ完全に実装することでコードのオフロードが容易に
- インテル® DPC++ 互換性ツール (英語) (オープンソースの SYCLomatic プロジェクト・ベース) で CUDA* ライブラリー API を移行することにより、CUDA* コードを同等のパフォーマンスの SYCL* コードに移行する際の時間を節約

アプリケーション・パフォーマンスの向上

- すべてのツールでサポートされる第 5 世代インテル® Xeon® スケーラブル・プロセッサーとインテル® Core™ Ultra プロセッサー上でパフォーマンスを最大化
- インテルにより最適化された PyTorch* (英語) と TensorFlow* (英語) フレームワークにより、インテルの CPU と GPU 上でネイティブに実行される AI および マシンラーニング・アプリケーションをさらに高速化し、数値計算ワークロードに標準の Python* を使用する際のデータ並列拡張を改善
- CPU および GPU 上の乱数ジェネレーターに対する インテル® MKL の新しいベクトル演算の最適化により、時間を節約してパフォーマンスを向上
- インテル® レンダリング・ツールキットのインテル® GPU サポートにより、スケーラブルなリアルタイム・レンダリングを高速化

開発者の生産性を高める

- インテル® DPC++/C++ コンパイラーとインテル® Fortran コンパイラーの LLVM サニタイザーにより、インテルの CPU 上でメモリーリーク、初期化されていないメモリー、スレッドデータ競合、デッドロック、未定義の動作を素早く排除
- 個々のツールのダウンロードや、独自のニーズに合わせてカスタマイズ可能なプリセットのバンドルなど、柔軟で合理化されたプロセスにより、AI フレームワーク、ツール、ライブラリーを迅速に利用可能
- 高度なプレビュー機能の評価が容易に。GPU オフロードを容易にする C++ Parallel STL、計算ノードのリソース使用量を最適化する動的デバイス選択、GPU オフロードのオーバーヘッドを軽減する SYCL* グラフ、OpenMP* のスレッド・オーバーサブスクリプションを防ぐスレッド・コンポーザビリティー、NPU (ニューラル・プロセッサー・ユニット) へのオフロードコードのプロファイルを含む。

詳細は、補足資料または最新情報を参照してください。

HPC 向けインテル® ソフトウェア開発ツールの新機能

第 4 世代インテル® Xeon® プロセッサー、インテル® Xeon® CPU マックス・シリーズ、インテル® データセンター GPU マックス・シリーズの価値を最大化

インテル® MPI ライブラリー 4.0

開発者は非常に大きな MPI メッセージを簡単に渡し、MPI セッションを使用して単一の MPI_Init/MPI_COMM_WORLD よりも詳細なプロセスでグループ化できます。

- 効率的なメッセージパッシングと集合操作インフラストラクチャーにより、インテル® データセンター GPU マックス・シリーズを含むノードで構成されるシステム上で MPI アプリケーションのパフォーマンスを向上します。
- 開発者は、PMIx 標準に基づくソフトウェア管理スタックを使用するシステムをターゲットにする場合、インテル® MPI ライブラリーを使用できるようになりました (エクサスケール・システムの新たな管理ニーズに対応しています)。
- ILP64 のネイティブサポートによる 8 バイト整数のシームレスなサポートにより、Fortran コードでより大きなデータセットを使用できます。

インテル® ディストリビューションの Python*

Data Parallel Extensions for Python* により、低レベルの独自プログラミング API を使用することなく、インテルのすべての CPU および GPU で標準ベースのアクセラレーションを可能にします。

インテル® oneCCL

ハードウェア・リソースの有効活用により、分散 AI ワークロードのパフォーマンスを向上します。

インテル® Fortran コンパイラー

初期 Fortran 2023 標準規格のサポートが追加され、OpenMP* 5.0 および 5.1 の対応が強化されました。CPU 上の Fortran、OpenMP*、OpenMP* オフロードアドレス、メモリーリーク、初期化されていないメモリー、スレッドデータ競合、デッドロック、未定義の動作を検出できる LLVM サニタイザーの追加により、開発者の生産性を大幅に向上します。

インテル® oneAPI DPC++/C++ コンパイラー

SYCL* 2020 をほぼ完全に実装し、開発者の生産性および CPU と GPU のオフロード・パフォーマンスを向上し、OpenMP* 5.0 と 5.1 への準拠が強化されました。

- CPU 上の C++、SYCL*、OpenMP* アドレス、メモリーリーク、初期化されていないメモリー、スレッドデータ競合、デッドロック、未定義の動作を容易に検出できる LLVM サニタイザーが追加されました。
- 仮想関数を使用する C++ コードを SYCL* デバイスオフロードで実行する簡単な方法が追加され、SYCL* および OpenMP* コードのエラーメッセージ処理とエラー処理が改善されました。

インテルの AI ツールの新機能

GPU と CPU 向けのインテルの最新の最適化により主要な AI ライブラリーとフレームワークを高速化

新しい AI Tools Selector
必要なフレームワーク、ツール、ライブラリーを個別に、またはプリセットのバンドルで、正確に入手するための柔軟で合理的なプロセスを提供します。

TensorFlow* 向けインテル® エクステンションと PyTorch* 向けインテル® エクステンション
インテルの GPU と CPU のネイティブサポートにより、大幅なパフォーマンス向上を実現します。高度な最適化と機能により、大規模で複雑な AI ワークロードの処理がさらに強化されます。

インテル® ディストリビューションの Python*
インテルの GPU をサポートし、標準 Python* を使用して数値計算ワークロードと最適化されたデータ並列拡張機能 (Python*、NumPy*、Numba) を CPU および GPU システムにデプロイして、パフォーマンスと開発者の生産性を向上します。

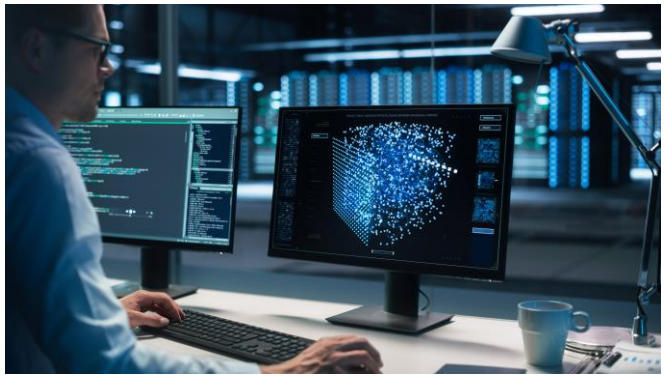
XGBoost 向けインテル® オプティマイゼーション
GPU 向けの新たな最適化により、インテルのすべての CPU と GPU で実行できるようになりました。

インテル® ニューラル・プロセッサー
インテルの GPU をサポートし、統合 Neural Network Compression Framework により、インテルの AI 実行プロバイダー間での最適化を簡素化することで、シームレスなユーザー体験とリソースの節約を実現します。TensorFlow* 2.14、PyTorch* 2.2、ONNX* RT 1.15 をサポートしました。

scikit-learn* 向けインテル® エクステンション
新しい K 平均法と低次アルゴリズムにより、疎データの取り扱いを改善し、アナリティクスとマシンラーニングのパフォーマンスを向上します。これにより、Python* 開発者は、パフォーマンスを向上し、ユースケースを高速化する選択肢が 2 つ増えました。

インテル® oneDNN
ストレージ効率の合理化、インテル® Xeon® プロセッサー上でのパフォーマンスの最適化、グラフ・コンパイラー機能との互換性の強化、コンパイラーの Xbyak バックエンドによる高度なコード生成、TensorFlow* 2.5 を使用したインテル® Xeon® プロセッサーでの sparse_tensor_dense_matmul() パフォーマンスの向上により、開発の生産性を高め、アプリケーションを高速化します。

Modin 向けインテル® ディストリビューション
AI パフォーマンス、効率とイノベーション、インテルの GPU サポートを向上するため、データテスクを高速化します。インタラクティブなジョブのパフォーマンス、ベンチマークの競争力、効率的なデータ取り込みに重点を置いて、データ管理を強化しています。



レンダリングとレイトレーシングの新機能

インテルの GPU 上でレイトレース・ハードウェア・アクセラレーションによりリアルタイムのレンダリングを実現

インテル® Embree

受賞歴のある CPU でのレイトレースのサポートに加えて、インテル® Arc™ グラフィックス、インテル® データセンター GPU フレックス・シリーズおよび新しく追加されたインテル® データセンター GPU マックス・シリーズでプロダクションレベルのレイトレース・ハードウェア・アクセラレーションを実現します。

インテル® オープン・イメージ・デノイズ

CPU でのサポートに加えて、インテル® Arc™ グラフィックス、インテル® データセンター GPU フレックス・シリーズおよびインテル® データセンター GPU マックス・シリーズで最先端のノイズ除去をサポートします。

インテル® オープン・ボリューム・カーネル・ライブラリー

CPU のサポートに加えて、インテル® Arc™ グラフィックス、インテル® データセンター GPU フレックス・シリーズおよびインテル® データセンター GPU マックス・シリーズをサポートし、CPU と GPU 上のレンダリングに共通のボリューム・ライブラリーを提供します。

インテル® OSPRay

CPU に加えて、インテル® Arc™ グラフィックス、インテル® データセンター GPU フレックス・シリーズおよびインテル® データセンター GPU マックス・シリーズ上で使用できるレンダリング API とレイトレーシング・エンジンを提供します。

インテル® OSPRay Studio

CPU に加えて、インテル® Arc™ グラフィックス、インテル® データセンター GPU フレックス・シリーズおよびインテル® データセンター GPU マックス・シリーズ上で、インテル® OSPRay レイトレーシング・エンジンの GUI を提供します。

4004 Moore's Lane モデルは、インテルの GPU 上でインテル® Embree ベースのパストレーサーを使用してレンダリングされました。
(インテルにより作成されたイメージ)

第 4 世代および第 5 世代
インテル® Xeon® スケーラブル・プロセッサー向け開発ツール

内蔵アクセラレーターを活用するツールでパフォーマンスを活性化

AI とアナリティクスを高速化

インテル® oneAPI ベース・ツールキットとインテルの AI ツールを使用して、マシンラーニングとデータサイエンスのパイプラインを高速化

インテル® AMX とインテル® AVX-512 を活用

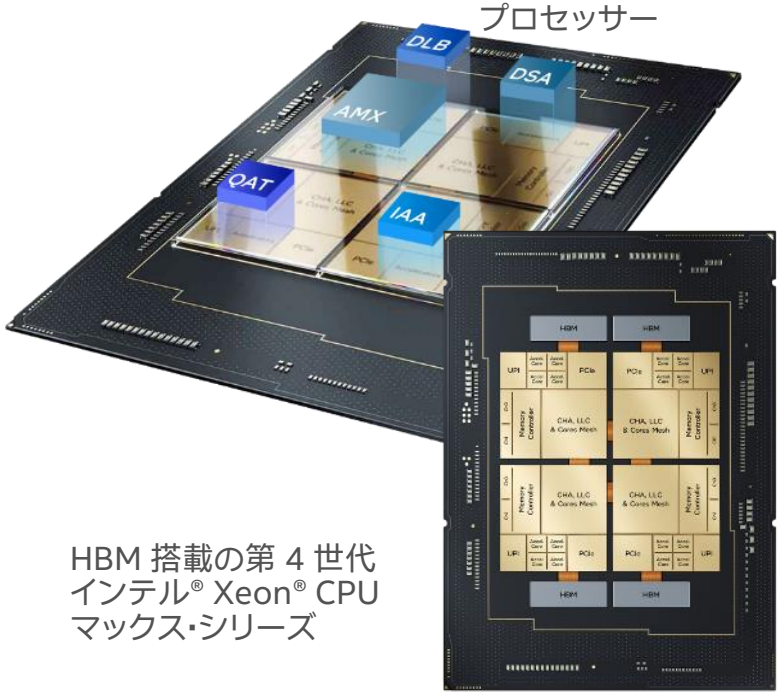
- インテル® oneDNN
- インテル® oneDAL

TensorFlow* と PyTorch* ディープラーニング AI フレームワークのトレーニングと推論を大幅に最適化

マルチアーキテクチャーのパフォーマンスと移植性

ベンダー固有や独自のアーキテクチャーに代わって、インテル® DPC++ 互換性ツールまたはオープンソースの SYCLomatc により CUDA* コードを SYCL* に移行して、インテル® Xeon® GPU マックス・シリーズを含むマルチベンダーのアーキテクチャーに容易に移植できるように単一のソースコード・ベースを作成

第 4 世代インテル® Xeon® スケーラブル・プロセッサー



HBM 搭載の第 4 世代インテル® Xeon® CPU マックス・シリーズ

スケーラブルで高速な HPC アプリケーションを実現

HPC とテクニカル・コンピューティングの生産性を向上

- インテル® Fortran コンパイラー
- インテル® oneMKL

第 4 世代インテル® Xeon® CPU マックス・シリーズの高帯域幅メモリーによりメモリー依存のワークロードのボトルネックを解消

- インテル® VTune™ プロファイラーはインテル® Xeon® プロセッサー・マックス・シリーズの高帯域幅メモリー (HBM) の情報を提供

データベース管理システムを高速化

インテル® IAA を有効にすることで、データベース分析ワークロードをより高速に実行し、メモリー帯域幅を節約し、より高いクエリー・スループットを実現

- インテル® クエリー・プロセシング・ライブラリー (インテル® QPL) (英語)

インテル® DSA を有効にすることで、ストリーミング・データの移動と変換処理を最適化し、レイテンシーを短縮し、メモリー転送のパフォーマンスを向上

- インテル® Data Mover Library (インテル® DML) (英語)

関連情報: 第 4 世代インテル® Xeon® プロセッサーと第 4 世代インテル® Xeon® プロセッサー・マックス・シリーズ向けのソフトウェア (英語)

インテル® データセンター GPU フレックス・シリーズ向け開発ツール

メディア、視覚推論、およびクラウド・ワークロードをマルチアーキテクチャーで高速化

インテル® oneVPL

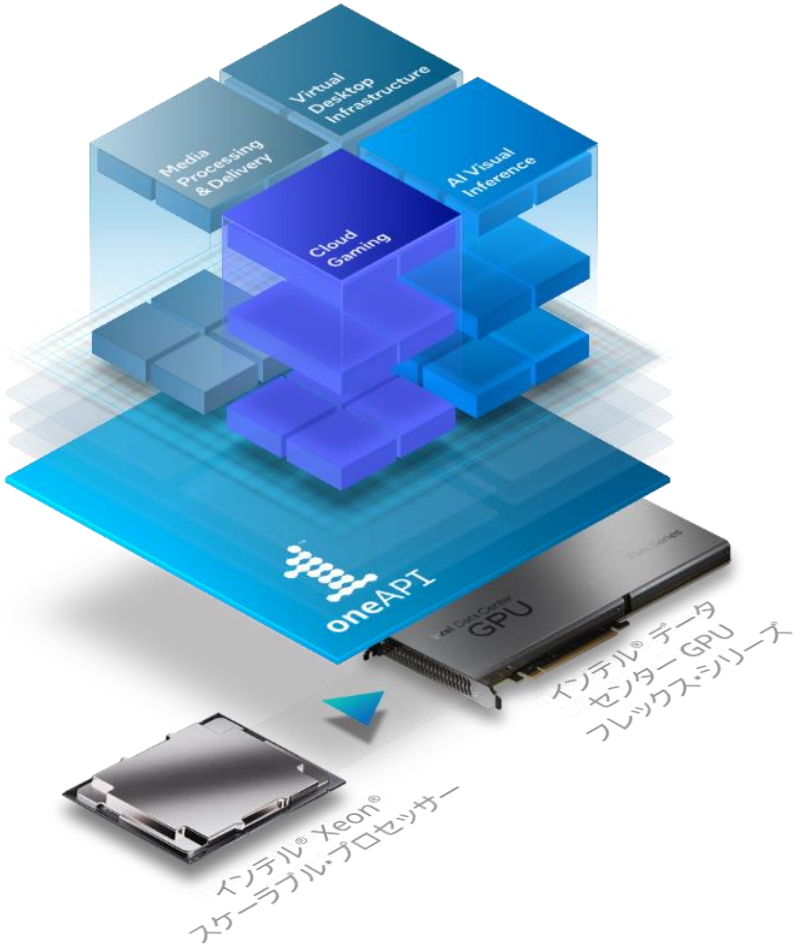
GPU に使用。CPU 速度のメディア処理、配信、クラウド・ゲーム・ストリーミングには GStreamer + FFmpeg を使用

インテル® oneAPI DPC++/C++ コンパイラー

同じコードで AI 視覚推論ワークロードを CPU と GPU 間で最大 150TOPS INT8 にスケールアップ

インテル® VTune™ プロファイラー

アプリケーションのパフォーマンスを解析し、最も時間のかかる GPU コード領域特定し、SYCL*、OpenCL* コード、Microsoft* DirectX*、OpenMP* オフロードコードの GPU オフロードスキーマとデータ転送を最適化して、計算集約型のタスクを高速化



インテル® グラフィックス・パフォーマンス・アナライザー (英語)

CPU と GPU 上でビジュアル・コンピューティングとグラフィックスのワークロードを最適化

インテル® ディストリビューションの OpenVINO™ ツールキット

TensorFlow* や PyTorch* などの一般的なディープラーニング・フレームワークをサポートしており、AI 視覚推論の効率を高め、迅速なデプロイを実現

インテルの AI ツール

最適化された DL & ML フレームワークとハイパフォーマンスの Python* ライブラリーでエンドツーエンドのマシンラーニングとデータサイエンスのパイプラインを高速化

関連情報: インテル® データセンター GPU フレックス・シリーズ向けのソフトウェア (英語)

インテル® データセンター GPU マックス・シリーズ向けの開発ツール

HPC と AI で優れたパフォーマンスを発揮するマルチアーキテクチャー・アプリケーションを開発

HPC を高速化

インテル® oneAPI DPC++/C++ コンパイラー、インテル® Fortran コンパイラー、パフォーマンス・ライブラリー (インテル® oneMKL、インテル® oneDNN、インテル® MPI ライブラリー) は、インテル® Xe マトリクス・エクステンション (インテル® XMX) によるアクセラレーションとインテル® Xe リンクによる GPU 間の直接通信を利用して、複数の GPU で実行するアプリケーションを高速化

解析とデバッグツールは、最も時間のかかる GPU コード領域を特定することで、開発者が計算集約型のタスクを解析して高速化できるようにし、最適に動作するコードを設計するための実用的なアドバイスを提供

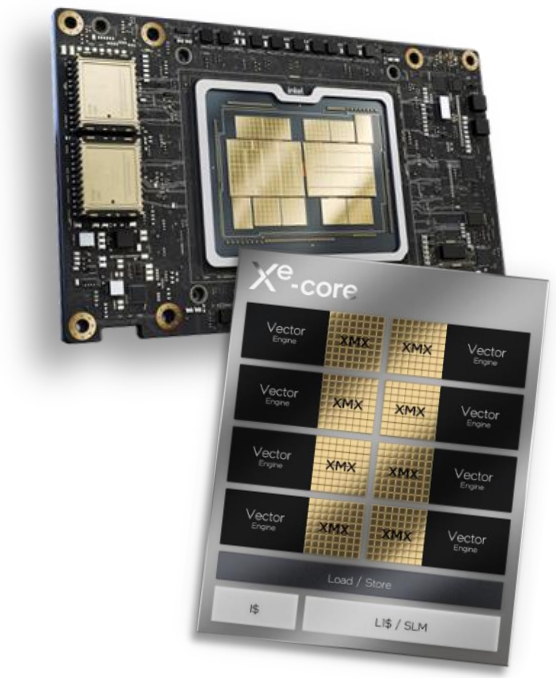
インテル® ディストリビューションの Python* で利用可能な Parallel Extensions for Python* により、低レベルの独自プログラミング API を使用することなく、標準ベースの CPU および GPU アクセラレーションが可能

AI と DL 推論を高速化

インテル® oneAPI ベース・ツールキットに含まれるインテル® oneDNN はインテル® XMX を活用してディープラーニング・フレームワークの AI トレーニング/推論を高速化

インテル® ディストリビューションの OpenVINO™ ツールキット powered by oneAPI は AI 視覚推論を高速化し、デプロイを合理化

どちらのツールも、TensorFlow* や PyTorch* などのディープラーニング・フレームワークをサポート



コード移行ツールでマルチアーキテクチャー・コードを効率良く作成

CUDA* コードを SYCL* に移行して、マルチベンダーのアーキテクチャーに容易に移植できるように単一のソースコード・ベースを作成

オープンソースの SYCLomatic (英語) をベースとするインテル® DPC++ 互換性ツール (英語) がほとんどのプロセスを自動化し、継続的なコードの保守にかかる時間を大幅に短縮

ハイパフォーマンスかつ高忠実度のレイトレーシングとレンダリング

インテル® レンダリング・ツールキットのライブラリーは、CPU のサポートに加えて、インテル® データセンター GPU フレックス・シリーズやインテル® データセンター GPU マックス・シリーズを含む、インテル® Xe アーキテクチャー・ベースの GPU をサポート

インテル® Embree (英語)
ハードウェア・レイトレーシング・ユニットと拡張メモリー容量を活用して、大規模なデータセットのレンダリング・パフォーマンスを向上

インテル®オープン・イメージ・デノイズ (英語)
インテル® XMX を利用して高品質でハイパフォーマンスなノイズ除去機能と AI 最適化を提供

インテル® オープン・ボリューム・カーネル・ライブラリー
CPU と GPU 上のレンダリングに共通のボリューム・ライブラリーを提供

インテル® OSPRay
スケーラブルで、移植可能な、分散レンダリング API

HPC、SciVis、ProRender パストレーサー

関連情報: インテル® データセンター GPU マックス・シリーズ向けの開発ツール (英語)

インテル® Core™ Ultra プロセッサー向けの開発ツール

AI PC 向けに最適化および高速化された移植性の高いアプリケーションの作成

インテル®ディストリビューションの OpenVINO™ ツールキット

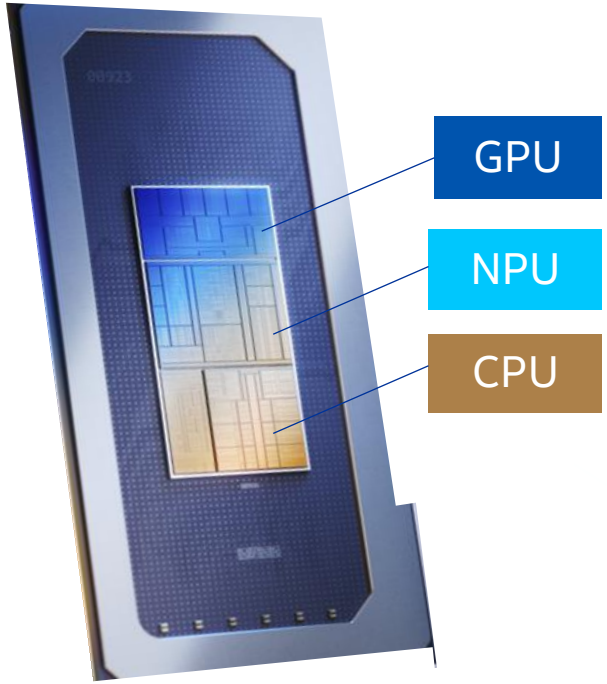
TensorFlow* や PyTorch* などの一般的なディープラーニング・フレームワークをサポートしており、AI 視覚推論の効率を高め、迅速なデプロイを実現

インテルの AI ツール

最適化された DL & ML フレームワークとハイパフォーマンスの Python* ライブラリーでエンドツーエンドのマシンラーニングとデータサイエンスのパイプラインを高速化

インテル® oneAPI DPC++/C++ コンパイラー

同じコードで AI 視覚推論ワークロードを CPU と GPU 間で最大 150TOPS INT8 にスケールアップ



ニューラル・プロセシング・ユニット搭載の
インテル® Core™ Ultra プロセッサー

インテル® グラフィックス・パフォーマンス・アナライザー (英語)

CPU と GPU 上でビジュアル・コンピューティングとグラフィックスのワークロードを最適化

インテル® oneVPL

GPU に使用。CPU 速度のメディア処理、配信、クラウド・ゲーム・ストリーミングには GStreamer + FFmpeg を使用

インテル® VTune™ プロファイラー

アプリケーションのパフォーマンスを解析し、最も時間のかかる GPU コード領域特定し、SYCL*、OpenCL* コード、Microsoft* DirectX*、OpenMP* オフロードコードの GPU オフロードスキーマとデータ転送を最適化して、計算集約型のタスクを高速化

インテル® ソフトウェア開発ツールの使用例

マルチアーキテクチャーのパフォーマンスと生産性を顧客にもたらす

HPC

アルゴンヌが公開したインテル® データセンター GPU マックス・シリーズを使用した Aurora のパフォーマンス (英語)

TACC の Frontera スーパーコンピューター (英語) は oneAPI を使用してエクサスケールの科学計算を加速

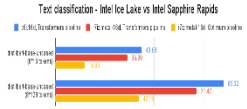
ケンブリッジ大学 (英語) による oneAPI を使用したゼタスケール実現への取り組み

HPC 向けに Google Cloud (英語) を高速化
ビデオ (英語) | ポッドキャスト (英語)

AI/ML/DL

Red Hat

Red Hat* はデータサイエンスのワークフローを最適化 (英語)



インテルの AI による HuggingFace Transformer のスケーリングとパフォーマンスの最適化 (英語)

IBM

NLP と NLU による IBM Watson のパフォーマンスの最適化 (英語)

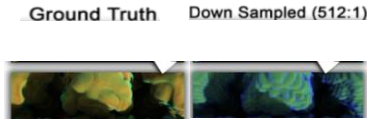
PyTorch Foundation

インテルの最適化による PyTorch* の進化 (英語)

レンダリング

THE UNIVERSITY OF TENNESSEE KNOXVILLE

テネシー大学 (英語) は oneAPI を使用してクラウドベースの RaaS (Rendering-as-a-Service) 環境を実現

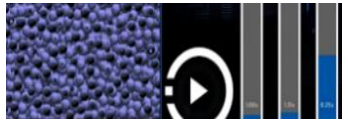


カリフォルニア大学デイビス校 (英語) はパフォーマンスを 3 倍向上し、サイエンティフィック・レンダリングで 100 倍のデータ圧縮を達成

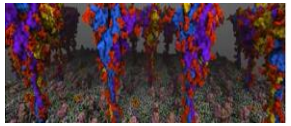
スティーブン・ホーキング宇宙論研究所 (英語) が宇宙物理学を視覚化

Cinesite は『アダムス・ファミリー 2』(英語) でレンダリング効率を 10～25% 向上

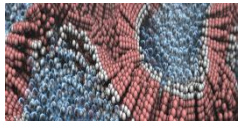
CUDA* コードを SYCL* へ移行



Aurora スーパーコンピューター向けの NAMD 分子動力学 (英語) の準備



LAMMPS: エクサスケールの材料探索を高速化



ストックホルム大学 GROMACS 2022 (英語)

Samsung Medison (英語) は oneAPI を利用して産科超音波診断システムを構築

THE UNIVERSITY OF UTAH®

ユタ大学 (英語) は大規模な画像データセットを使用 SYCL* を使用したバインド

1. 詳細は、「法務上の注意書き」を参照してください。インテル® ソフトウェア製品のパフォーマンスおよび最適化に関する注意事項については、software.intel.com/articles/optimization-notice (英語) を参照してください。詳細は、http://www.intel.com/PerformanceIndex/ (英語) を参照してください。結果は異なることがあります。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。ほかの情報も参考にして、正確かどうかを評価してください。

ダウンロードまたはクラウドで実行可能

ローカルで実行

www.xlsoft.com/jp/products/intel/oneapi にアクセス

- ダウンロード
- リポジトリー
- コンテナー

サンプルコード、
クイック・スタート・ガイド、
ウェビナー、トレーニング

インテル® デベロッパー・クラウドで実行

ウェブサイトにアクセス

- ハードウェアの取得不要
- ダウンロード、インストール、設定不要
- サンプルコードとドキュメントを利用可能
- すぐに使えるデプロイ環境と開発環境
- 最新の学習リソースにアクセス

エキスパートとコミュニティーによるサポートを利用可能

- 無料でダウンロード、またはクラウド上でツールを実行できます。
- インテル® oneAPI ベース・ツールキット、インテル® HPC ツールキット、インテル® レンダリング・ツールキットの有償ライセンスには優先サポートが含まれます。
- インテル® デベロッパー・クラウドには、Standard (無料)、Premium (従量課金)、Enterprise (チームアクセス) があります。

*バージョン 2024.0 のツールは、インテル® デベロッパー・クラウドで利用可能です。

インテル® デベロッパー・クラウド

最新のインテルのハードウェアとソフトウェアでマルチアーキテクチャー・アプリケーションを開発してワークロードをテスト

開発者

最新のインテルの CPU および GPU でサポートされるインテルにより最適化された AI ソフトウェアに**簡単にアクセスして使用**可能

エンタープライズ

インテルの製品とテクノロジーの**採用とデプロイを加速**し、**新しいソフトウェアとサービス**を作成可能

パートナー

パフォーマンスと**コストが最適化されたインテルの AI コンピューティング・サービス**を顧客に提供可能

無料のクラウドクレジットが利用可能

intel Developer Cloud

利用可能なソフトウェア・ツールと最適化

intel 1 oneAPI BASE TOOLKIT

intel HPC TOOLKIT

intel RENDERING TOOLKIT

AI ツール

MODIN, pandas, NumPy, SciPy, TensorFlow, PyTorch, OpenVINO, インテル® ニューラル・コンプレッサー, scikit learn, dmlc XGBoost, python

利用可能なプラットフォーム

intel Xeon, intel Xeon MAX SERIES, intel DATA CENTER GPU MAX SERIES, intel DATA CENTER GPU FLEX SERIES, intel Gaudi

インテル® デベロッパー・クラウドにアクセス

*バージョン 2024.0 のツールは、インテル® デベロッパー・クラウドで利用可能です。

インテル® デベロッパー・クラウド

さまざまなインテル® アーキテクチャーでマルチアーキテクチャー・アプリケーションを開発してワークロードをテスト

	開発者 インテルの AI を利用する	エンタープライズ AI の大規模なデプロイ	パートナー SaaS プロバイダー向け AI インフラストラクチャー・サービス
	評価 学習 開発 研究	認証とベンチマーク AI トレーニングと推論 プロダクション・ワークロード	サードパーティー AI Saas 向けの インテルのコンピューティング・ サービス
クラウド・インターフェイス	インテルのクラウドポータルと AI		
クラウドサービス	AI インフラストラクチャーとプラットフォーム・サービス		
クラウド・インフラストラクチャー	intel. XEON / intel. XEON MAX SERIES / intel. DATA CENTER GPU MAX SERIES / intel. DATA CENTER GPU FLEX SERIES / intel. GAUDI		

インテル® デベロッパー・クラウドのサービスの種類

Standard

無料

インテルの AI 製品を評価

以下が含まれます。

- ✓ 最新のインテルの製品の評価
- ✓ AI スキルの習得
- ✓ 最先端の学習リソースへのアクセス
- ✓ インテル・コミュニティー・サポート

Premium

従量課金

1 人のユーザーが最新のインテルの製品にアクセス

以下が含まれます。

- ✓ リリース前のハードウェアへのアクセス
- ✓ AI/ML ソフトウェア・ツールキット
- ✓ インテルの優先サポート

Enterprise

確約利用割引

チームで最新のインテルの製品にアクセス

以下が含まれます。

- ✓ チーム向けの有償サブスクリプション
- ✓ 独自のインフラストラクチャーに推論をデプロイ
- ✓ CPU、GPU、AI アクセラレーターの使用 ***
- ✓ インテルによる 24時間 365日の優先サポート

エキスパートとコミュニティーによるサポートを利用可能

インテルのツールキットの優先サポート

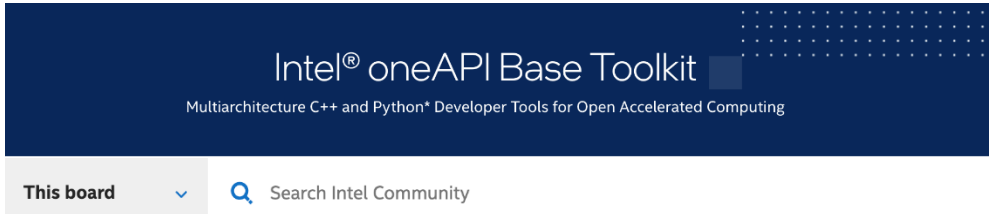
インテル® oneAPI ベース・ツールキット、インテル® HPC ツールキット、インテル® レンダリング・ツールキットの有償ライセンスには優先サポートが含まれる

- **インテルのエンジニアに直接問い合わせたり**、サポート要求を送信することが可能
- ツールキットに関する技術的な質問や製品に関するその他のニーズに対して**迅速な対応**が得られる
- 製品の新しいアップデートと以前のバージョンを**無料でダウンロード**可能
- 製品の機能や品質についての**リクエストを送る**ことが可能
- エスカレーションされた不具合について優先的なサポートが得られる
- 過去数十年のハイパフォーマンス・コード作成の経験を基に構築された**ドキュメント・ライブラリー**を利用可能
- インテルのテクニカル・コンサルティング・エンジニアによるオンサイト/オンライン・トレーニングおよびコンサルテーションを含む**オプションのサービス**を追加料金で利用可能

無料のコミュニティー・サポート

パブリックの**デベロッパー・ソフトウェア・フォーラム**でインテル・コミュニティーと繋がることができる

- コミュニティーのテクニカル・エキスパートによりサポートが提供され、インテルのエンジニアがモニタリング
- よくある質問と回答
- オンライン・チュートリアルとセルフヘルプ・フォーラムにアクセス可能
- 開発者仲間からのトラブルシューティングに関するガイダンス



intel®

バージョン 2024 の新機能

コンポーネントの詳細

インテル® ソフトウェア開発ツール 2024

インテルの CPU と GPU 上でマルチアーキテクチャー・コンピューティングを高速化

今後登場する第 5 世代インテル® Xeon® スケーラブル CPU とインテル® Core™ Ultra CPU を含む、インテルの CPU と GPU でパフォーマンスを最大化します。

Codeplay の oneAPI プラグインにより NVIDIA と AMD の GPU をサポートします。また、インテル® Arria® FPGA、インテル® Cyclone® 10 GX FPGA、インテル® Stratix® FPGA、インテル® Agilex® FPGA をサポートします。

コンパイラー

- インテル® oneAPI DPC++/C++ コンパイラーは、SYCL* 2020 をほぼ完全に実装し、開発者の生産性および CPU と GPU のオフロード・パフォーマンスを向上し、OpenMP* 5.0 と 5.1 への準拠が強化されました。
- インテル® Fortran コンパイラーは、初期 Fortran 2023 標準規格をサポートし、メモリーリーク、初期化されていないメモリー、スレッドデータ競合、デッドロックなど、CPU 上のさまざまな動作を開発者が検出できる LLVM サニタイザーが追加されました。

Python* プログラミング

- Data Parallel Extensions for Python* により、低レベルの独自プログラミング API を使用することなく、標準ベースの CPU および GPU アクセラレーションが可能です。
- NumPy* API のサブセットの代わりに Data Parallel Extension for NumPy* パッケージを使用して CPU と GPU の両方を活用することで、パフォーマンスを向上します。
- Data Parallel Extension for Numba を使用して、Numba で CPU をプログラムするのと同じ方法で、GPU 向けの開発を生産的に行うことができます。

パフォーマンス・ライブラリー

- インテル® MPI ライブラリーは、GPU 上での大きな MPI メッセージパッシングを簡素化し、Fortran コードで大きなデータセットを使用できるようにし、開発者が PMIx 標準に基づくソフトウェア管理スタックを使用するシステムをターゲットにできるようにします。
- インテル® oneMKL は、マルチベンダーの CPU および GPU 上で、HPC シミュレーション、統計サンプリングなどの RNG にベクトル演算の最適化を統合します。

解析ツール

- インテル® Advisor は、第 4 世代インテル® Xeon® スケーラブル・プロセッサーで FP16 および BF16 拡張機能とインテル® AMX のプロファイルをサポートし、さらに GPU およびマルチタイル・アーキテクチャーのアプリケーション・パフォーマンス特性 (帯域幅依存、命令ミックス、キャッシュラインの利用状況) の検出を可能にします。
- インテル® VTune™ プロファイラーは、開発者が各ノードのインテル® Xe リンクを介した GPU 間のトラフィックと帯域幅に加えて、暗黙的な USM データの移動がパフォーマンスの低下を引き起こしているかどうか、およびその理由を理解できるようにします。

レンダリングとレイトレーシング

- インテル® レンダリング・ツールキットがインテル® Xe アーキテクチャー・ベースの GPU (インテル® Arc™ GPU、インテル® データセンター GPU フレックス・シリーズ、インテル® データセンター GPU マックス・シリーズ) をサポートしました。

詳細: バージョン 2024 の機能

インテル® ソフトウェア開発ツール 2024 - 特長

コンパイラー

インテル® oneAPI DPC++/C++ コンパイラー

- SYCL* 2020 をほぼ完全に実装し、開発者の生産性および CPU と GPU のオフロード・パフォーマンスを向上します。
- OpenMP* 5.0 と 5.1 への準拠が強化されました。
- C++、SYCL*、OpenMP* アドレス、メモリーリーク、初期化されていないメモリー、スレッドデータ競合、デッドロック、CPU 上の未定義の動作を容易に検出できる LLVM サニタイザーが追加されました。
- 仮想関数を使用する C++ コードを SYCL* デバイスオフロードで実行する簡単な方法の追加と、SYCL* および OpenMP* コードのエラーメッセージ処理とエラー処理の改善により、開発者の生産性を高めます。

インテル® Fortran コンパイラー

- CPU 上の Fortran、OpenMP*、OpenMP* オフロードアドレス、メモリーリーク、初期化されていないメモリー、スレッドデータ競合、デッドロック、未定義の動作を検出できる LLVM サニタイザーが追加されました。
- 初期 Fortran 2023 標準規格のサポートが追加され、OpenMP* 5.0 および 5.1 の対応が強化されました。

インテル® ディストリビューションの Python*

- Data Parallel Extensions for Python* により、低レベルの独自プログラミング API を使用することなく、インテルのすべての CPU および GPU で標準ベースのアクセラレーションを実現します。
- Data Parallel Extension for NumPy* パッケージにより、インテルの CPU と GPU で NumPy* を実行することを可能にします。
- Data Parallel Extension for Numba を使用して、Numba で CPU をプログラムするのと同じ方法で GPU 向けに開発できます。

インテル® ソフトウェア開発ツール 2024 - 特長

パフォーマンス・ライブラリー

インテル® oneDNN

- インテル® Xeon® プロセッサー上でストレージ効率を合理化し、パフォーマンスを最適化します。
- グラフ・コンパイラー機能との互換性が強化されました。
- コンパイラーの Xbyak バックエンドによるコード生成が向上し、TensorFlow* 2.5 と oneDNN により sparse_tensor_dense_matmul() が高速化しています。

インテル® oneMKL

- マルチベンダーの CPU および GPU でのハイパフォーマンス・コンピューター・シミュレーション、統計サンプリング、その他の分野向けに、ベクトル演算の最適化を乱数ジェネレーターに統合します。
- インテル® oneMKL SYCL* ライブラリーは、バイナリー・フットプリントとダウンロード・サイズを小さくするために分割されました。
- インテルの GPU で FP16 データ型のベクトル演算をサポートします。
- インテル® Xeon® CPU マックス・シリーズおよびインテル® データセンター GPU マックス・シリーズ向けに最適化されたハイパフォーマンス・ベンチマーク HPL および HPL-AI を提供します。

インテル® oneCCL

- ハードウェア・リソースの有効活用により、分散 AI ワークロードのパフォーマンスを向上します。

インテル® oneDAL

- マシンラーニング・モデルのビルドおよび配布のために Microsoft* のオープンソース ML.NET マシンラーニング・フレームワークと統合されました。

インテル® MPI ライブラリー

- インテル® MPI ライブラリー 4.0 の機能により、開発者は非常に大きな MPI メッセージを簡単に渡し、MPI セッションを使用して単一の MPI_Init/MPI_COMM_WORLD よりも詳細なプロセスでグループ化できます。
- 効率的なメッセージパッシングと集合操作インフラストラクチャーにより、インテル® データセンター GPU マックス・シリーズを含むノードで構成されるシステム上で MPI アプリケーションのパフォーマンスを向上します。
- ILP64 のネイティブサポートによる 8 バイト整数のシームレスなサポートにより、Fortran コードでより大きなデータセットを使用できます。
- 開発者は、PMIx 標準に基づくソフトウェア管理スタックを使用するシステムをターゲットにする場合、インテル® MPI ライブラリーを使用できるようになりました (エクサスケール・システムの新たな管理ニーズに対応しています)。

インテル® IPP

- インテル® IPP Cryptography に AES-GCM アルゴリズム向けの VAES インテル® AVX2 最適化と RSA アルゴリズム向けのインテル® AVX-512 最適化が導入され、ユーザーがデータを安全かつ高速に送信できるようになりました。
- RGB から XYZ のカラー変換向けのインテル® AVX512-VNNI ベースの最適化により、画像ドメインのパフォーマンスが向上
- インテル® AVX-512 最適化により信号処理ドメイン統計関数 L2 Norm のパフォーマンスが向上

インテル® oneTBB

- Emscripten を使用した WebAssembly (WASM) でのコンパイルが可能になり、ウェブブラウザー上で実行されるアプリケーションでの使用が容易になります。

詳細: バージョン 2024 の機能

インテル® ソフトウェア開発ツール 2024 - 特長

レンダリング

インテル® レンダリング・ツールキットのライブラリーは、CPU のサポートに加えて、インテル® Arc™ グラフィックス、インテル® データセンター GPU フレックス・シリーズ、インテル® データセンター GPU マックス・シリーズを含む、インテル® Xe アーキテクチャー・ベースの GPU をサポートしました。

インテル® Embree

受賞歴のある CPU でのレイトレースのサポートに加えて、インテル® Arc™ グラフィックス、インテル® データセンター GPU フレックス・シリーズおよび新しく追加されたインテル® データセンター GPU マックス・シリーズでプロダクションレベルのレイトレース・ハードウェア・アクセラレーションにより、スケーラブルなリアルタイムのレンダリングを実現します。

インテル® オープン・イメージ・デノイズ

CPU でのサポートに加えて、インテル® Arc™ グラフィックス、インテル® データセンター GPU フレックス・シリーズおよびインテル® データセンター GPU マックス・シリーズで最先端のノイズ除去をサポートします。

インテル® オープン・ボリューム・カーネル・ライブラリー

CPU のサポートに加えて、インテル® Arc™ グラフィックス、インテル® データセンター GPU フレックス・シリーズおよびインテル® データセンター GPU マックス・シリーズをサポートし、CPU と GPU 上のレンダリングに共通のボリューム・ライブラリーを提供します。

インテル® OSPRay

- CPU に加えて、インテル® Arc™ グラフィックス、インテル® データセンター GPU フレックス・シリーズおよびインテル® データセンター GPU マックス・シリーズ上で使用できるレンダリング API とレイトレーシング・エンジンを提供します。
- インテル® OSPRay Studio は、CPU に加えて、インテル® Arc™ グラフィックス、インテル® データセンター GPU フレックス・シリーズおよびインテル® データセンター GPU マックス・シリーズ上で、インテル® OSPRay レイトレーシング・エンジンの GUI を提供します。

コードの移行、解析、およびデバッグツール

インテル® DPC++ 互換性ツール

CUDA* から同等のパフォーマンスの SYCL* コードへ生産的に移行します。CUDA* ライブラリー API と、AI、ディープラーニング、暗号化、科学シミュレーション、画像処理などの分野の 20 の一般的なアプリケーションが追加されています。

インテル® VTune™ プロファイラー

- 開発者が各ノードのインテル® Xe リンクを介した GPU 間のトラフィックと帯域幅を理解できるようにします。
- 統合共有メモリー (USM) ワークロードを使用する SYCL* および DPC++ アプリケーションをプロファイルして、パフォーマンスの低下を招く暗黙の USM データ移動の理由を特定します。
- NPU (ニューラル・プロセシング・ユニット) にオフロードされたコードをプロファイルして、NPU から DDR メモリーに転送されるデータ量を理解し、NPU で実行に最も時間を費やしているタスクを特定します。[1]

インテル® Advisor

- 自動ルーフライン解析でハードウェアの制限に対するパフォーマンスのヘッドルームを理解するため、Python* コードをプロファイルします。
- 第 4 世代インテル® Xeon® スケーラブル・プロセッサーで FP16 および BF16 拡張機能とインテル® AMX のプロファイルをサポートします。
- インテルの GPU、マルチタイル・アーキテクチャー、VNNI と ISA で、帯域幅依存、命令ミックス、キャッシュラインの使用など、アプリケーションのパフォーマンス特性の評価をサポートします。

インテル® ディストリビューションの GDB

- デバッガーのパフォーマンスを向上し、コマンドライン、Visual Studio*、Visual Studio* Code の UI を改善します。さらに、インテルの CPU と GPU 向けのアプリケーションを効率良くデバッグするため、ロック制御を微調整する高度なスケジューラー・ロックを提供します。

1. この機能は、現在テクニカルプレビュー段階です。

詳細: バージョン 2024 の機能

インテル® ソフトウェア開発ツール 2024 - 特長

AI ツール

新しい AI Tools Selector は、必要なフレームワーク、ツール、ライブラリーを正確に入手するための柔軟で合理的なプロセスを提供します。

インテル® ディストリビューションの Python*

インテルの GPU をサポートし、標準 Python* を使用して数値計算ワークロードと最適化されたデータ並列拡張機能 (Python*、NumPy*、Numba) を CPU および GPU システムにデプロイすることで、開発者がより多くの作業を、より高速に行えるようにします。

TensorFlow* 向けインテル® エクステンションと PyTorch* 向けインテル® エクステンション

インテルの GPU と CPU をネイティブにサポートし、パフォーマンスを大幅に向上します。高度な最適化と機能により、大規模で複雑な AI ワークロードの処理がさらに強化されます。

XGBoost 向けインテル® オプティマイゼーション

GPU 向けの新たな最適化により、インテルのすべての CPU と GPU で実行できるようになりました。

scikit-learn* 向けインテル® エクステンション

新しい K 平均法と低次アルゴリズムにより、疎データの取り扱いを改善し、アナリティクスとマシンラーニングのパフォーマンスを向上します。これにより、Python* 開発者は、パフォーマンスを向上し、ユースケースを高速化する選択肢が 2 つ増えました。

Modin 向けインテル® ディストリビューション

AI パフォーマンス、効率とイノベーション、インテルの GPU サポートを向上するため、データタスクを高速化します。インタラクティブなジョブのパフォーマンス、ベンチマークの競争力、効率的なデータ取り込みに重点を置いて、データ管理を強化しています。

インテル® ニューラル・コンプレッサー

インテルの GPU をサポートし、統合 Neural Network Compression Framework により、インテルの AI 実行プロバイダー間での最適化を簡素化することで、シームレスなユーザー体験とリソースの節約を実現します。TensorFlow* 2.14、PyTorch* 2.2、ONNX* RT 1.15 をサポートしました。

インテル® oneDNN

ストレージ効率の合理化、インテル® Xeon® プロセッサー上でのパフォーマンスの最適化、グラフ・コンパイラー機能との互換性の強化、コンパイラーの Xbyak バックエンドによる高度なコード生成、TensorFlow* 2.5 を使用したインテル® Xeon® プロセッサーでの sparse_tensor_dense_matmul() パフォーマンスの向上により、開発の生産性を高め、アプリケーションを高速化します。

詳細: バージョン 2024 の機能

バージョン 2024 のプレビュー機能

- **インテル® oneAPI DPC++/C++ コンパイラー**のプレビュー機能である**インテル® oneDPL** を使用した PSTL オフロードにより、CPU と GPU の両方で C++ 標準の並列アルゴリズム (PSTL) を容易に実行できます。
- **インテル® oneDPL** の動的デバイス選択テクノロジーのプレビュー版は、ラウンドロビン、ロードベース、および自動調整ポリシーのいずれかを選択して、利用可能な計算デバイスでの作業をスケジュールすることで、計算ノードのリソースを最適化してアプリケーションを高速に実行できるようにします。
- **インテル® oneAPI DPC++/C++ コンパイラー**に SYCL* Graph がプレビュー機能として追加されました。開発者は、オーバーヘッドを軽減するため、単一の CPU 操作を通じて複数の GPU 操作を起動するメカニズムを提供することで、単一の操作ではなくグラフとして作業を定義できます。
- **インテル® oneTBB** のスレッド・コンポーザビリティー・マネージャーにより、インテル® oneTBB スレッドと OpenMP* スレッドを入れ子にするときの柔軟性とワークロード・パフォーマンスが向上します。
- **インテル® MPI ライブラリー**は、CPU および GPU からの通信で MPI RMA (一方向通信) を利用して、より効率的なメッセージパッシングを提供します。
- **SYCL* イメージ API 拡張機能**により、画像処理およびメディア処理エンジニアは、インテル LLVM オープンソース・プロジェクトの DPC++ コンパイラーを使用して、NVIDIA* GPU でバインドレス・イメージを高速化できます。

詳細: バージョン 2024 の機能

インテル® ソフトウェア開発ツールの詳細

インテル® oneAPI ベース・ツールキット

インテル® oneAPI ベース・ツールキット

データセントリックなワークロードを高速化

インテルの CPU、GPU、FPGA でハイパフォーマンスなアプリケーションを開発するための基本ツールとライブラリーの基本セット

対象ユーザー

- さまざまな業界の広範な開発者
- アドオン・ツールキットのユーザー (これはすべてのツールキットの基本となるため)

主な特徴/利点

- データ並列 C++ コンパイラー、ライブラリー、解析ツール
- CUDA* コードから C++ with SYCL* への移行を支援する SYCLomatic/インテル® DPC++ 互換性ツール
- 高速化された scikit-learn*、NumPy*、SciPy* ライブラリーを含む Python* ディストリビューション
- スレッド化、数学、データ・アナリティクス、ディープラーニング、ビデオ/画像/信号処理向けに最適化されたパフォーマンス・ライブラリー

詳細とダウンロード

インテル® oneAPI ベース・ツールキット

ダイレクト・プログラミング

- インテル® oneAPI DPC++/C++ コンパイラー
- インテル® DPC++ 互換性ツール
- インテル® ディストリビューションの Python*
- oneAPI ベース・ツールキット用インテル® FPGA アドオン

API ベースのプログラミング

- インテル® oneAPI DPC++ ライブラリー (インテル® oneDPL)
- インテル® oneAPI マス・カーネル・ライブラリー (インテル® oneMKL)
- インテル® oneAPI データ・アナリティクス・ライブラリー (インテル® oneDAL)
- インテル® oneAPI スレッディング・ビルディング・ブロック (インテル® oneTBB)
- インテル® oneAPI コレクティブ・コミュニケーション・ライブラリー (インテル® oneCCL)
- インテル® oneAPI ディープ・ニューラル・ネットワーク・ライブラリー (インテル® oneDNN)
- インテル® インテグレーテッド・パフォーマンス・プリミティブ (インテル® IPP)

解析/デバッグツール

- インテル® VTune™ プロファイラー
- インテル® Advisor
- インテル® ディストリビューションの GDB

intel. 1 oneAPI BASE TOOLKIT

インテル® oneAPI DPC++/C++ コンパイラー

生産性とパフォーマンスに優れた SYCL* コンパイラー

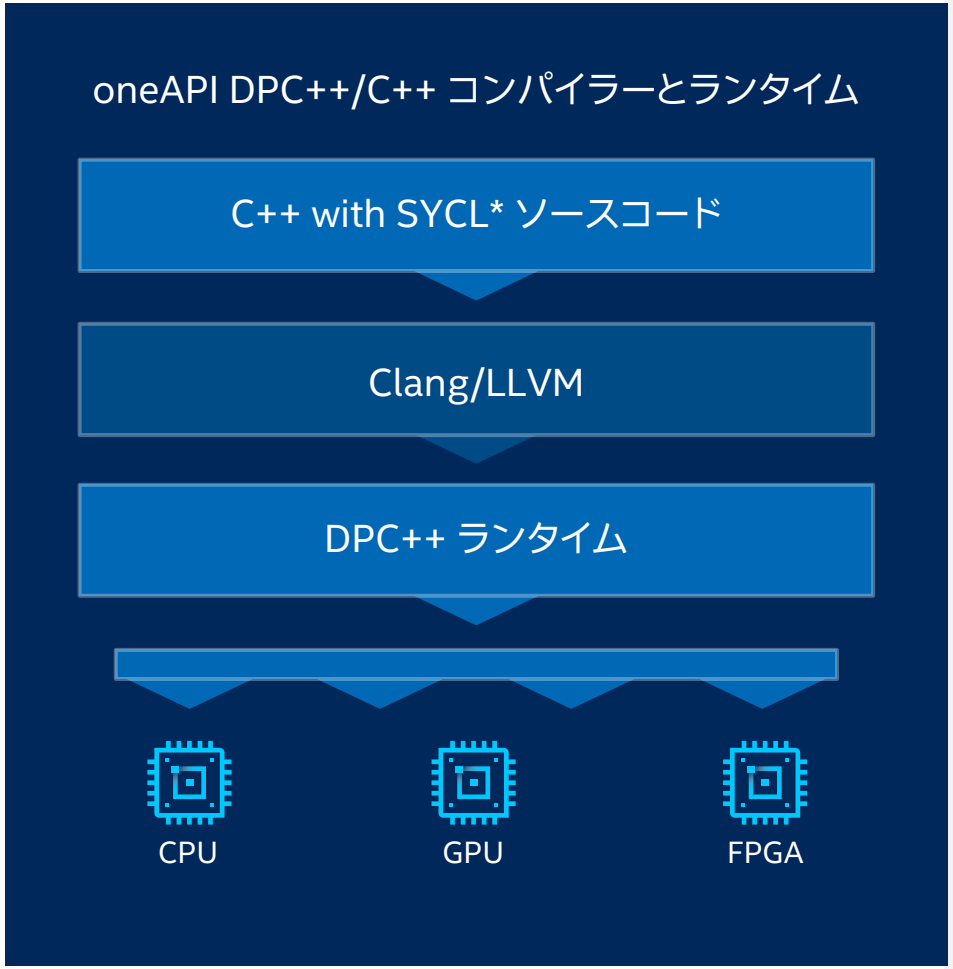
CPU とアクセラレーターに妥協のない並列プログラミングの生産性とパフォーマンスを提供

- ターゲット・ハードウェア間でコードの再利用が可能、特定のアクセラレーター向けのカスタム・チューニングを行うことが可能
- 単一アーキテクチャー専用の言語に代わる、オープンな業界全体の代替手段

Khronos SYCL* 標準

- C++ の生産性の利点を提供、一般的で使い慣れた C および C++ 構造を使用
- データ並列処理とヘテロジニアス・プログラミングをサポートするため Khronos Group が作成

インテルの数十年にわたるアーキテクチャーとハイパフォーマンス・コンパイラーの経験を基に構築



詳細とダウンロード

現時点では各アーキテクチャー向けにチューニングする必要があります。

インテル® DPC++ 互換性ツール

コードの移行時間を最小化

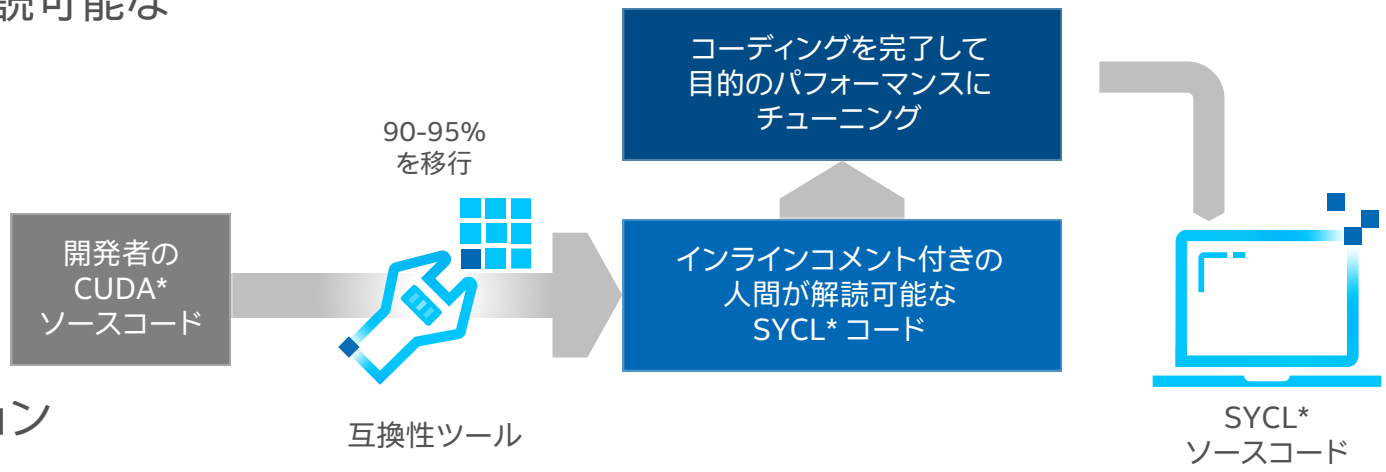
CUDA* で記述されているコードを C++ with SYCL* に移行する開発者を支援、可能な場合は人間が解読可能なコードを生成

通常はコードの 90-95% を自動的に移行[1]

開発者がアプリケーションの移行を完了できるように支援するインラインコメントを提供

SYCLomatic ツールはオープンソース・バージョン

インテル® DPC++ 互換性ツールの使用フロー



ベンダー依存からの解放と SYCL* への容易な移行

詳細とダウンロード

[1]出典: 2023年3月現在のインテルによる推定。Rodinia、SHOC、PENNANT など、85 種類の HPC ベンチマークとサンプルの測定結果に基づいています。結果は異なることがあります。

解析/デバッグツール

ハードウェアを最大限に活用

設計

インテル® Advisor

- コードを GPU へ効率良くオフロード
- CPU/GPU コードの計算とメモリーを最適化
- ベクトル並列処理をさらに有効にして効率を向上
- スレッド化されていないアプリケーションに効率的なスレッド化を追加

デバッグ

インテル® ディストリビューションの GDB

- CPU、GPU、FPGA を含むさまざまなアクセラレーターをサポート
- SYCL*、C、C++、OpenMP*、Fortran クロスアーキテクチャー・アプリケーションの詳細なシステム全体のデバッグが可能
- Microsoft* Visual Studio*、VS Code、Eclipse* との IDE 統合

チューニング

インテル® VTune™ プロファイラー

- GPU、CPU、FPGA、NPU† 向けにチューニング
- オフロードのパフォーマンスを最適化
- SYCL*、C、C++、Fortran、Python*、Go*、Java*、混在言語をサポート

† テクニカルプレビュー機能

強力なパフォーマンス･ライブラリー

豊富な機能

あらゆるユースケースに対応する最適化されたパフォーマンス･ライブラリー

スレッド、オフロード、数学、データ･アナリティクス、データ処理レンダリング、レイトレーシング、DNN、通信、暗号、など

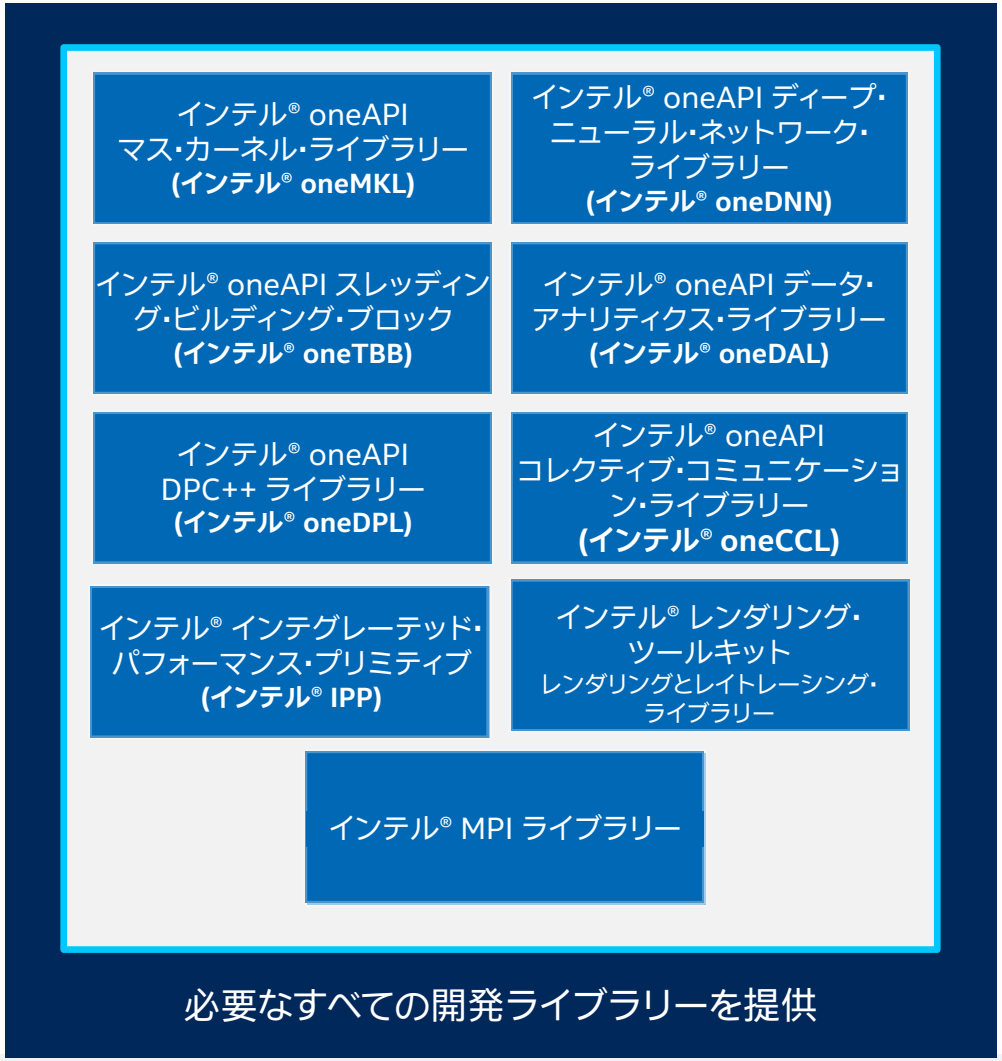
ハードウェアの価値を最大限に引き出す

主要なドメイン固有の関数を高速化するように設計

選択の自由

最高のパフォーマンスが得られるように各プラットフォーム向けに事前に最適化

ライブラリーの詳細



必要なすべての開発ライブラリーを提供

インテル® oneAPI DPC++ ライブラリー (インテル® oneDPL)

インテルの CPU、GPU、FPGA 上で SYCL* C++ カーネルを高速化

最適化された C++ 標準アルゴリズム

さまざまなハードウェア上での効率的なアプリケーション開発とデプロイメントのために 75 の並列化された C++17 アルゴリズムとユーティリティーを含む

C++ 開発者がすでに利用している並列ライブラリーがベース

簡単に開発者が採用できるように一般的なライブラリー Parallel STL および Boost.Compute を含む

インテル® DPC++ 互換性ツールと統合

すべての oneAPI DPC++ コンポーネントを補完して開発者の CUDA* コードから DPC++ コードへの移行を単純化

C++ でマルチアーキテクチャー・コンピューティングを実現

詳細とダウンロード

インテル® oneAPI ディープ・ニューラル・ネットワーク・ライブラリー (インテル® oneDNN)

ハイパフォーマンスなディープラーニングを実現

開発者がハイパフォーマンスなディープラーニング・フレームワークを作成できるように支援

命令セットとパフォーマンス最適化の複雑さを抽象化

インテルの CPU と GPU で同じ API を利用して、ジョブに最適なテクノロジーを使用

Linux*、Windows* をサポート

オープンソース化によりコミュニティーに貢献

詳細とダウンロード

インテル® oneAPI コレクティブ・コミュニケーション・ライブラリー (インテル® oneCCL)

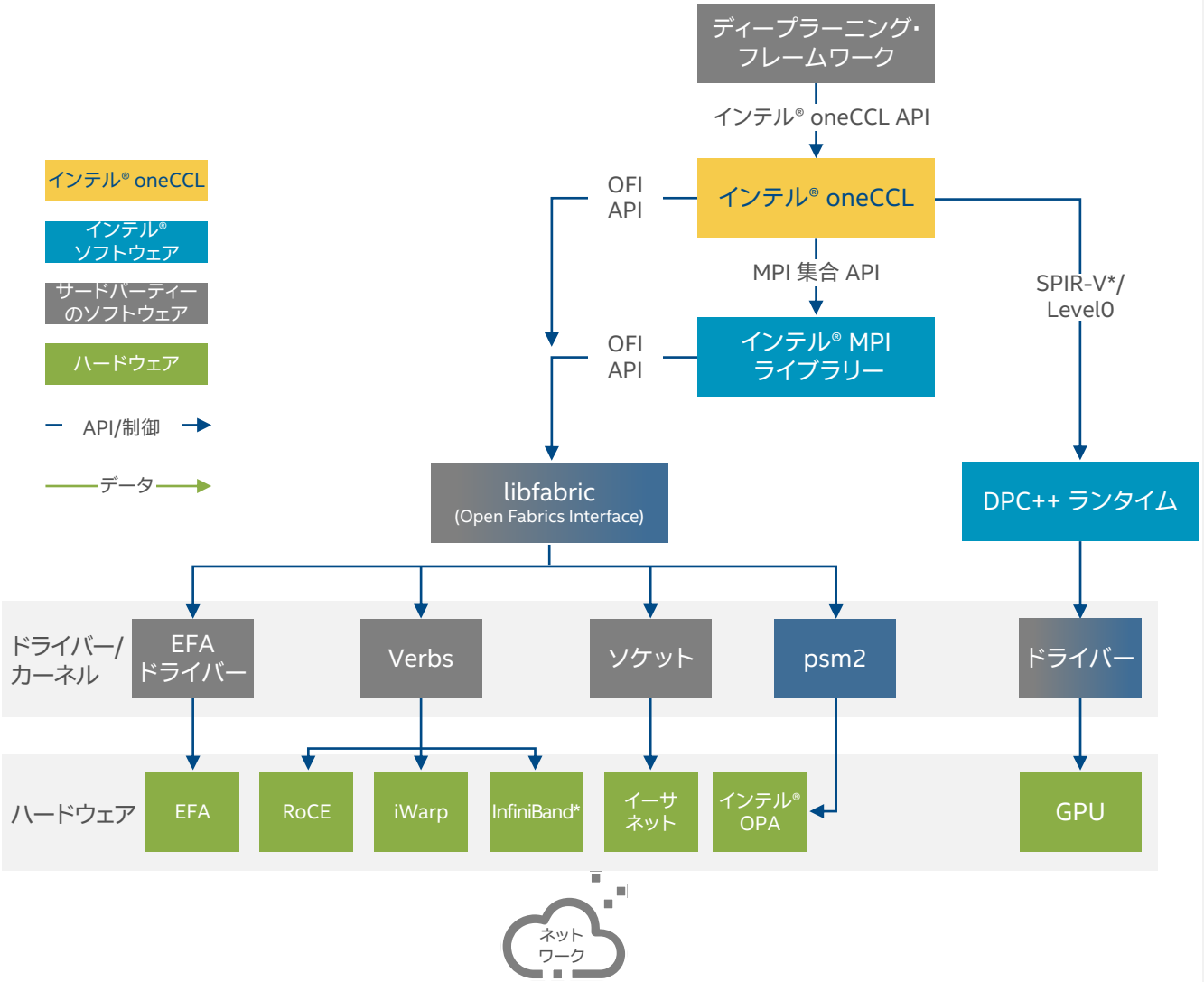
通信パターンを最適化

複数のノードでの分散モデル・トレーニングのためにインテルの CPU と GPU 向けにハイパフォーマンスな最適化された通信パターンを提供

インテル® Omni-Path アーキテクチャー (インテル® OPA)、InfiniBand*、イーサネットなど、多くのインターコネクトを透過的にサポート

低レベルの通信ミドルウェア上に構築された MPI と libfabrics

ディープラーニングのトレーニングに使用される処理 (all-gather、all-reduce、reduce-scatter) の効率的な実装が可能



詳細とダウンロード

インテル® oneAPI マス・カーネル・ライブラリー (インテル® oneMKL)

インテル® アーキテクチャー・ベースのシステムで最も高速で、最も使用されている数学ライブラリー[1]

- 高度に最適化、スレッド化、ベクトル化された数学関数により、科学、工学、金融アプリケーションを高速化
- 密/疎線形代数 (BLAS、LAPACK、直接法スパースソルバー)、FFT、ベクトル演算、サマリー統計、補間、行列乗算などの主な機能を提供
- 第 4 世代インテル® Xeon® スケーラブル・プロセッサーとインテル® データセンター GPU マックス・シリーズ向けの最新の最適化を含む
- 現在および将来の世代のインテルの CPU と GPU で一貫して優れたパフォーマンスを発揮
- コードを分岐せずに各プロセッサー向けに最適化されたコードを自動ディスパッチ
- SYCL*、C、Fortran API の言語サポート
- 無料およびロイヤルティー・フリーで利用可能

intel oneMKL

- 密線形代数
- 疎線形代数
- 高速フーリエ変換
- ベクトル演算
- ベクトル RNG
- サマリー統計

Evans Data のサイト (英語)

詳細とダウンロード

[1] データの出典: Evans Data が実施した「Global Development Survey Report 22.1」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。

インテル® インテグレーテッド・パフォーマンス・プリミティブ (インテル® IPP)

画像、信号、データ処理アプリケーション向けのビルディング・ブロック

画像/信号処理、データ圧縮/暗号化計算タスクを高速化する、すぐに使えるドメイン固有の関数

画像処理

- 医療用画像
- コンピューター・ビジョン
- デジタル監視
- ADAS
- 自動ソート
- 生体認証
- 視覚探索

信号処理

- ゲーム
- エコー・キャンセレーション
- 通信
- エネルギー

データ圧縮と暗号化

- データセンター
- ID の検証
- 電子署名
- エンタープライズ・データ管理
- スマートカード/スマートウォレット
- 情報セキュリティー/サイバーセキュリティー

詳細とダウンロード

インテル® oneAPI スレッディング・ビルディング・ブロック (インテル® oneTBB)

高速なアプリケーションのための高度なスケーリング

並列処理向けの柔軟な C++ ライブラリー

開発者に深いハードウェアの知識がなくてもアプリケーションで並列処理を表現できる簡単な方法

将来もアプリケーションのパフォーマンスをスケーリング

現在および将来のプラットフォームで計算負荷の高いワークロードのパフォーマンスを効率良く並列化およびスケーリング

ほかのスレッド化パッケージとの互換性

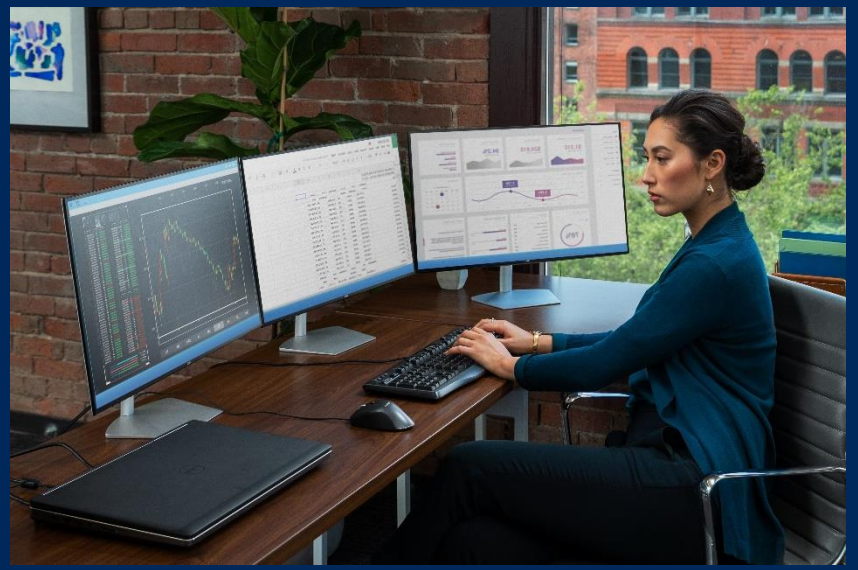
レガシーコードを保ちながら新しい実装にはインテル® oneTBB を使用、ほかのスレッド化パッケージとシームレスに共存

アプリケーションの構成可能性を簡素化/拡張

モジュール式のタスク並列処理と既存のプログラミング・モデルとの相互運用性により、現在および将来のヘテロジニアス・プラットフォームで利用可能

intel oneTBB

並列プログラミングで最も広く利用されている C++ ライブラリーの 1 つ



詳細とダウンロード

インテル® MPI ライブラリー

柔軟で、効率良い、スケーラブルなクラスターメッセージ処理を実現

最適化された MPI アプリケーション・パフォーマンス

- アプリケーション固有のチューニング
- 自動チューニング
- 最新のインテル® Xeon® スケーラブル・プロセッサーをサポート

低レイテンシーおよび複数のベンダーとの相互運用性

- 業界トップレベルのレイテンシー
- OpenFabrics* インターフェイス (OFI) により、ファブリック向けに最適化されたパフォーマンスをサポート

高速な MPI 通信

- 最適化された集合操作

持続性のあるスケーラビリティー

- ネイティブ InfiniBand* インターフェイス・サポートにより、低レイテンシー、高帯域幅、メモリー使用量の軽減を実現

主要な新機能

- インテル® GPU ピニングをサポート
- 分散型非同期オブジェクト・ストレージ (DAOS) をサポート
- インテル® Xeon® Platinum 92XX プロセッサーの最適化
- Mellanox ConnectX*: 3/4/5/6 (FDR/EDR/HDR) サポートを拡張

インテル® oneAPI データ・アナリティクス・ライブラリー (インテル® oneDAL)

CPU と GPU にハイパフォーマンスなデータサイエンスをデプロイ

ML とデータ・アナリティクスのパフォーマンスを向上

- アプリケーションがより速く、より優れた予測を行えるように支援
- 最高のパフォーマンスを引き出せるようにデータの取り込みとアルゴリズムの計算を最適化
- 広範なアプリケーションのニーズを満たす、オフライン、ストリーミング、分散型使用モデルをサポート
- エッジデバイスとクラウドサービス間で解析ワークロードを分割してアプリケーション全体のスループットを最適化

インテル® oneDAL による GPU サポート

以下のアルゴリズムがサポートされています。

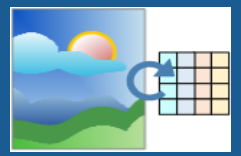
- **統計:** 相関、提示モーメント†
- **分類**: 線形回帰†、ロジスティック回帰†、KNN、SVM
- **教師なし学習**: K 平均法クラスタリング、DBSCAN
- **分類と回帰**: ランダムフォレスト
- **次元縮小**: PCA

NEW: scikit-learn* 1.2 を完全サポート

†

前処理 → 変換 → 解析 → モデリング → 検証 → 意思決定

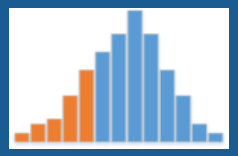
展開、フィルタリング、正規化



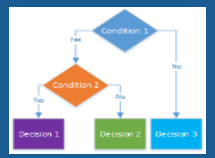
集約、次元縮小

サマリー統計ライブラリー、クラスタリングなど

ML (トレーニング) パラメーター推定、シミュレーション



仮説検証、モデルエラー



予測、決定木など

詳細とダウンロード

† GPU 実装と既存の oneAPI 仕様の oneDAL (英語) では CPU 実装のサブセットが増えています。

インテル® VTune™ プロファイラー

CPU、GPU、FPGA 向けにチューニング

SYCL* コードの解析

最も多くの時間を費やしている SYCL* コード行を確認

インテルの GPU、CPU、FPGA 向けのチューニング

サポートしているハードウェア・アクセラレーター向けに最適化

オフロードの最適化

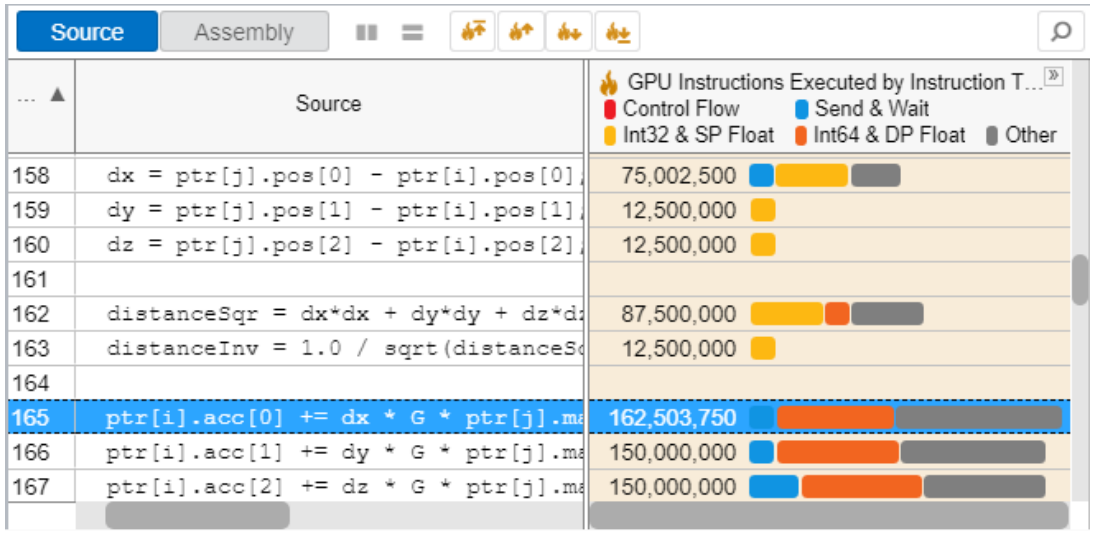
OpenMP* オフロードのパフォーマンスをチューニング

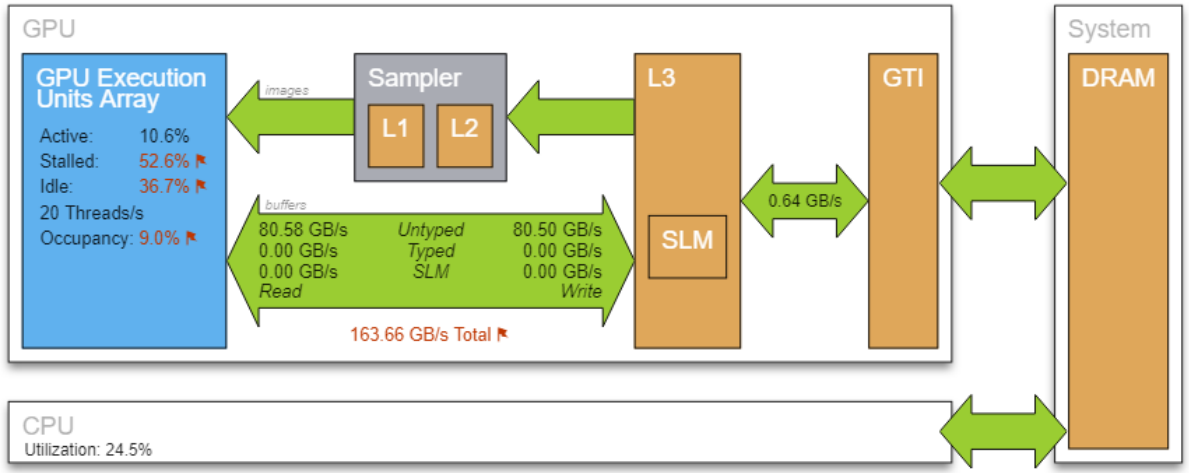
広範なパフォーマンス・プロファイル

CPU、GPU、FPGA、スレッド、メモリー、キャッシュ、ストレージなど
フレームグラフ表示により hotspot の視認性を向上

一般的な言語をサポート

SYCL*、C、C++、Fortran、Python*、Go*、Java*、またはこれらの混在





SYCL* コードの解析と GPU オフロード・プロファイル

詳細とダウンロード

現時点では各アーキテクチャー向けにチューニングする必要があります。

インテル® Advisor
高速な計算ソリューションを構成

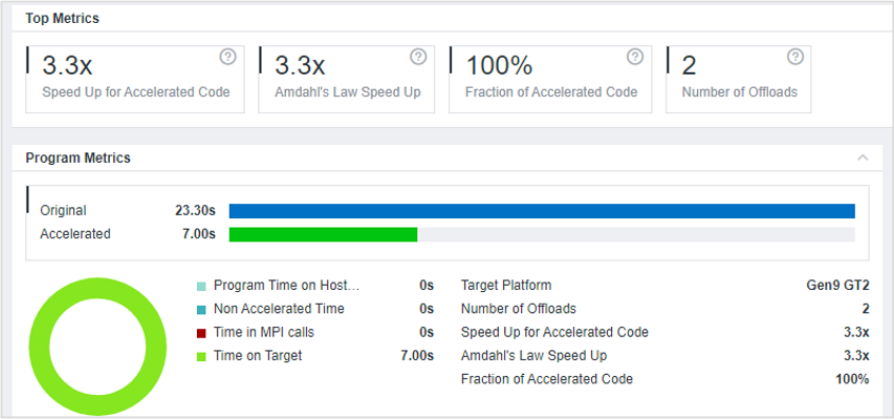
オフロードのモデル化
アクセラレーターにオフロードした際のパフォーマンスを推定

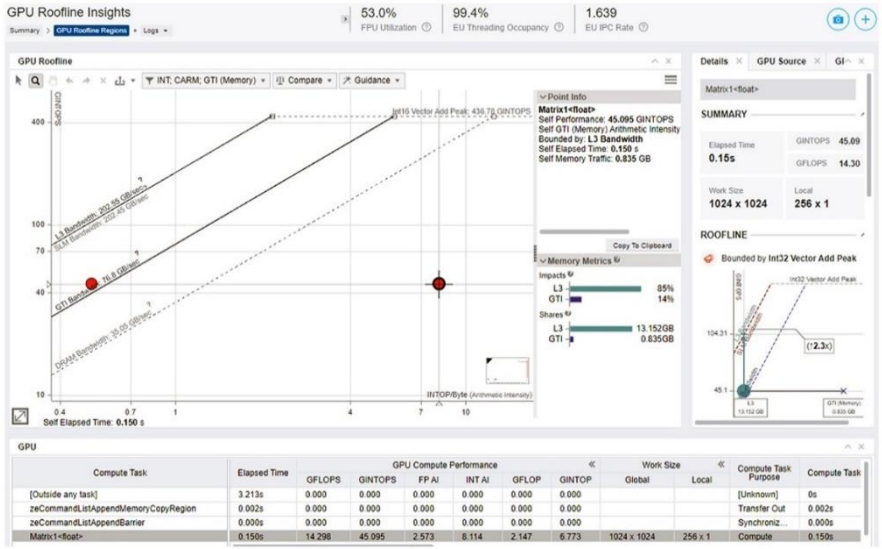
ルーフライン解析
CPU/GPU コードの計算とメモリーを最適化

ベクトル化アドバイザー
ベクトル化とその最適化

スレッド化アドバイザー
スレッド化されていないアプリケーションに効率的なスレッド化を追加

フローグラフ・アナライザー
効率的なフローグラフを作成して解析





オフロードする前に確認

詳細とダウンロード

現時点では各アーキテクチャー向けにチューニングする必要があります。

インテル® ディストリビューションの GDB

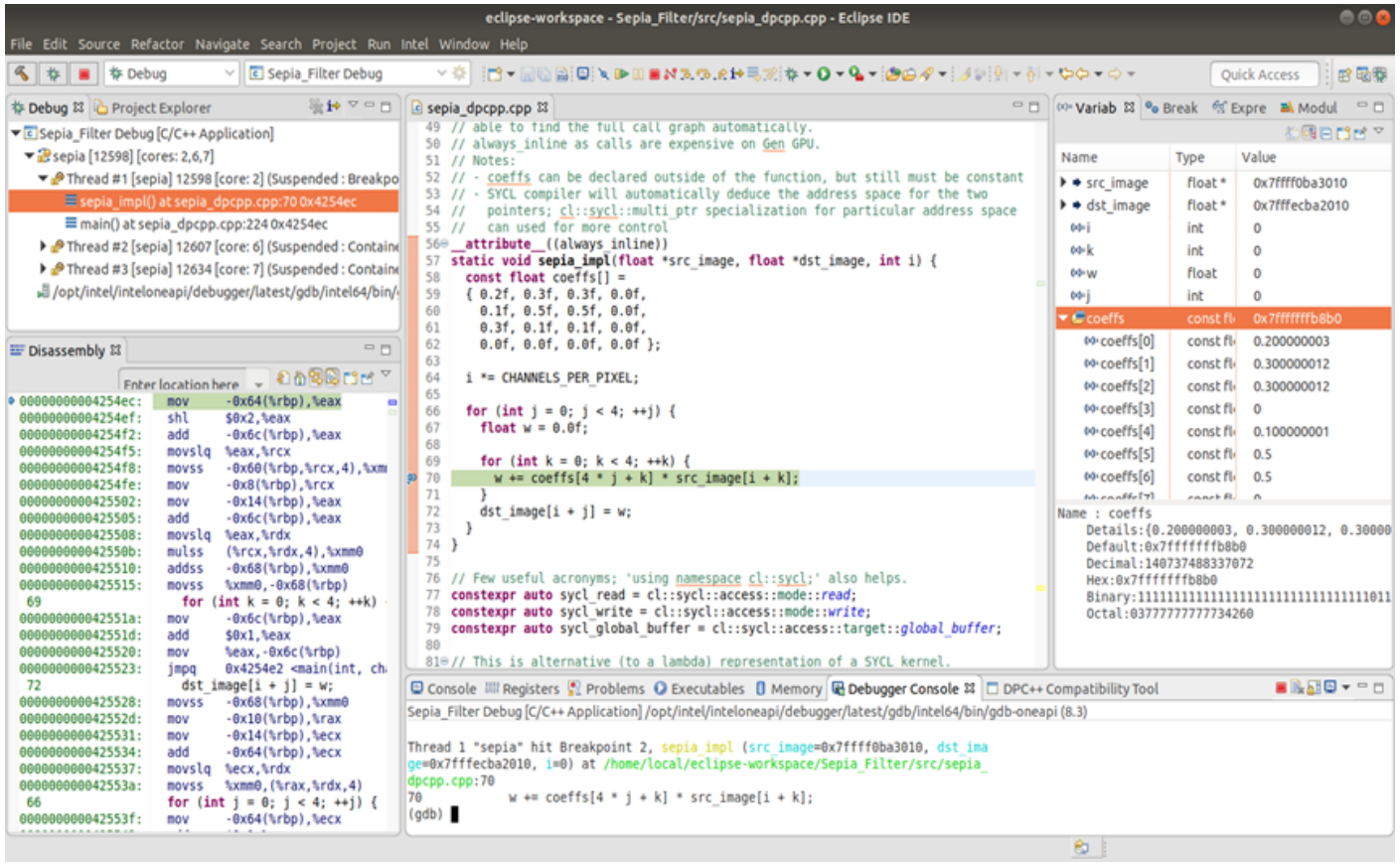
ヘテロジニアス・アプリケーションのデバッグ

高水準言語のデバッグをサポート

さまざまなアクセラレーターをサポート: インテルの CPU、GPU、FPGA のエミュレーション

アプリケーションの実行中にアクセラレーターのアーキテクチャーを自動検出

GDB ベースの独自仕様でないオープンソースのソリューション



詳細とダウンロード

インテル® ディストリビューションの Python*

数値計算と科学計算向けの CPU および GPU アクセラレーション

基本パッケージ群によりネイティブコードに近いパフォーマンスを達成

- コアとなる数値計算パッケージおよびマシンラーニング・パッケージを高速化
- 最新の命令をサポートし、ラップトップ、デスクトップ、強力なサーバーで利用可能なすべての CPU コアを活用
- Python* コードを最適化された命令にコンパイルする生産性向上ツール
- Python* プロジェクトとインテルのネイティブツールの統合を容易にするために不可欠な Python* バインディング

Data Parallel Extensions for Python*

- 低レベルの独自プログラミング API を使用することなく、インテルのすべての CPU および GPU で標準ベースのアクセラレーションを実現
- Data Parallel Extension for NumPy* パッケージにより、インテルの CPU と GPU 向けに NumPy* を使用して開発
- Data Parallel Extension for Numba を使用して、Numba で CPU をプログラムするのと同じ方法で GPU 向けに開発

詳細とダウンロード

インテル® ディストリビューションの Python*	
パッケージと環境マネージャー	パッケージと環境を管理するための必須ツール
データ処理とモデリングパッケージ	数値計算とデータサイエンス・ワークフロー向けのパッケージ
マシンラーニング・パッケージ	マシンがデータから自動的に学習するためのパッケージ
Python* インタープリターとコンパイラー	さまざまなインタラクティブ体験とスケールアップしたパフォーマンスを実現するツール
高度なプログラミング・パッケージ	データとデバイスの管理、同時実行、並列処理を細かく制御
開発パッケージとランタイム	インテルにより最適化された Python* パッケージ全体でパフォーマンスを発揮するためのランタイムパッケージ
優先サポート	インテル® oneAPI ベース・ツールキットで利用可能

サポートされるハードウェア・アーキテクチャー

CPU　GPU

oneAPI ベース・ツールキット用 インテル® FPGA アドオン

空間アーキテクチャーの SYCL* コーディング

経験豊富な FPGA 開発者向け

使いやすさ

同じヘテロジニアス SYCL* コード経験豊富な FPGA ユーザーはデータ並列 C++ を使用して合理化されたプログラミング・モデルを活用できる

ランタイム解析のサポート

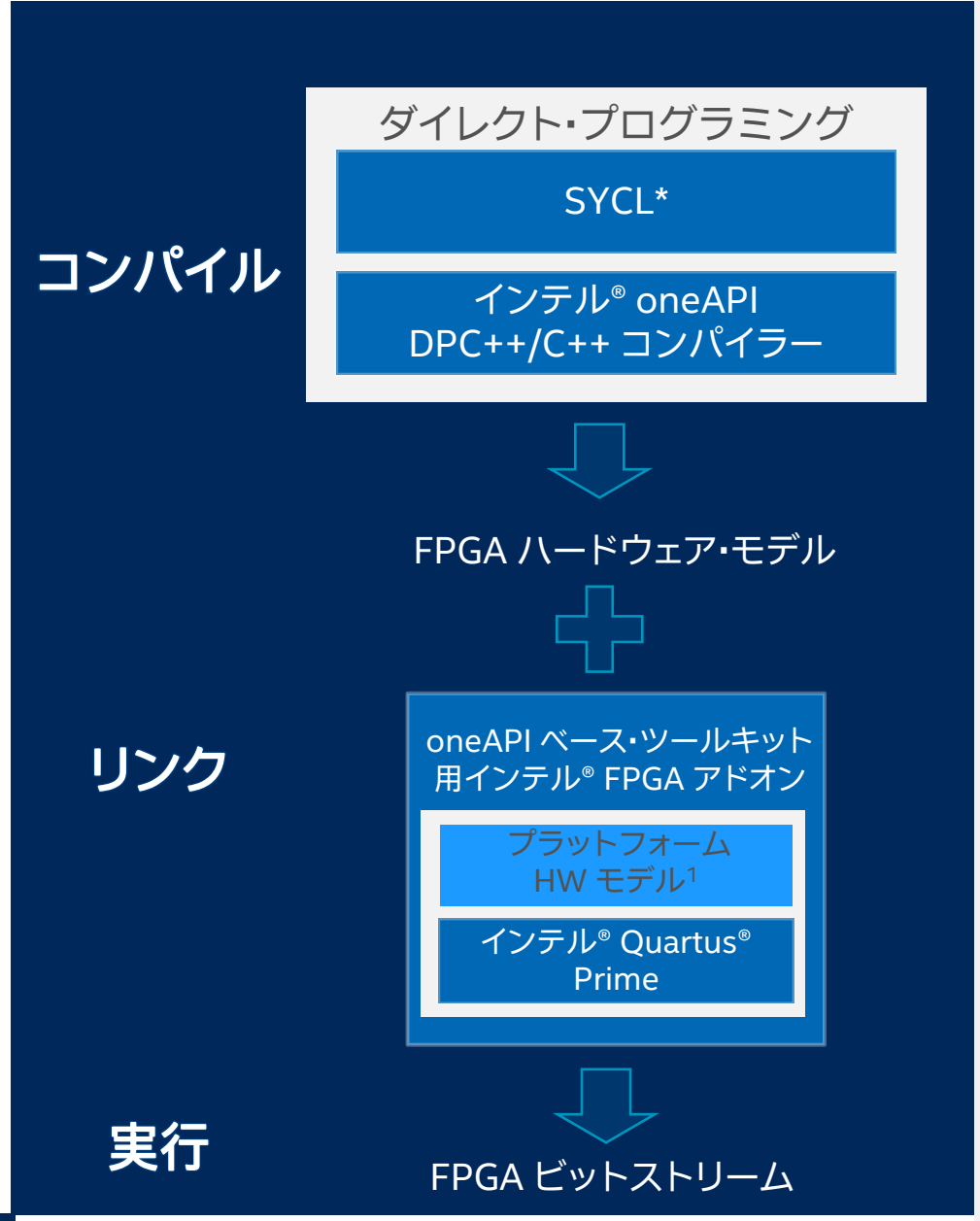
インテル® VTune™ プロファイラーを使用してランタイムにプロファイル・データを収集し、CPU と FPGA のインタラクションを解析

リアルタイム処理

決定性のある低レイテンシー、低消費電力、高スループットでデータを高速に処理

デバイス固有の最適化

経験豊富な FPGA 開発者向けに、FPGA 向けに oneAPI コードの最適化を開始する 1 日のトレーニング・クラスを提供



1. サードパーティーのプラットフォーム・ベンダーから入手可能

開始方法

	1 ソースコードを最適化	2 プログラムを新規作成または既存の CUDA* コードを移行	3 ハードウェア構成にアクセラレーターを追加
コンパイラーとコードの移行	インテル® DPC++ 互換性ツール インテル® oneAPI DPC++/C++ コンパイラー インテル® Fortran コンパイラー インテル® ディストリビューションの Python*	インテル® DPC++ 互換性ツール インテル® oneAPI DPC++/C++ コンパイラー インテル® Fortran コンパイラー インテル® ディストリビューションの Python*	インテル® DPC++ 互換性ツール インテル® oneAPI DPC++/C++ コンパイラー インテル® Fortran コンパイラー インテル® ディストリビューションの Python*
ライブラリー	インテル® oneAPI ライブラリー	インテル® oneAPI ライブラリー	インテル® oneAPI ライブラリー
アナライザー	インテル® VTune™ プロファイラー インテル® Advisor	インテル® VTune™ プロファイラー インテル® Advisor	インテル® Advisor インテル® VTune™ プロファイラー
	インテル® VTune™ アナライザーでアプリケーション・コードのパフォーマンス解析から始める **古いコンパイラーを使用している場合**最新のインテル® コンパイラーで再コンパイルすることでスピードアップが期待できる	新しいプログラムの場合ツールキットに含まれる**最先端のコンパイラーとライブラリー**で最高のパフォーマンスを発揮 移行する場合インテル® DPC++ 互換性ツールを使用して CUDA* コードを SYCL* に移行し、マルチアーキテクチャーに対応	最適化する前に**インテル® Advisor** で最適なデバイスの組み合わせを定義し、適切なアクセラレーターをターゲットにする さらに**インテル® VTune™ プロファイラー**でアプリケーション・コードのパフォーマンスを解析してチューニングする

	4 新しいハードウェアを使用	
ブリングアップ	インテル® システム・ブリングアップ・ツールキット	**インテル® システム・ブリングアップ・ツールキット**で新しいハードウェアのシステム・ブリングアップと検証をスピードアップ

関連情報

oneAPI 関連情報

使用する

- intel.com/oneapi (英語)
- ドキュメント (英語) + デベロッパー・ガイド
- サンプルコード (英語)
- インテル® デベロッパー・クラウド

oneAPI 仕様

- UXL Foundation (英語)
- oneAPI.io (英語)
- oneAPI オープン業界仕様 (英語)
- オープンソース実装 (英語)

学ぶ

- トレーニング (英語) とウェビナー (英語)
- awesome-oneAPI リポジトリー (英語)
- サミットとワークショップ (英語): ライブおよびオンデマンドの仮想ワークショップ、コミュニティー主導のセッション
- Intel Software の YouTube ページ (英語) の動画

エコシステム

- コミュニティー・フォーラム (英語)
- インテル® DevMesh イノベーター・プロジェクト (英語)
- oneAPI センター・オブ・エクセレンス (英語): 研究、コードの有効化、カリキュラム、教育

特殊なワークロード向けの
ドメイン固有のアドオン・ツールキット

- インテル® HPC ツールキット
- インテルの AI ツール
- インテル® レンダリング・ツールキット

インテルの HPC 向けツール

インテル® HPC ツールキット

スケーラブルで高速なアプリケーションを実現

ツールキットの概要

C++、Fortran、SYCL*、OpenMP* および MPI でエンタープライズ、クラウド、HPC、AI 向けのハイパフォーマンスでスケーラブルな並列アプリケーションを開発するためインテル® oneAPI ベース・ツールキットに追加するツールキット

対象ユーザー

- OEM/ISV
- C++、Fortran、OpenMP*、MPI 開発者

ツールキットの重要性

- インテル® Xeon® プロセッサー、インテル® Core™ プロセッサー、およびインテルのアクセラレーター上でパフォーマンスを向上
- 少ない労力で、高速かつ安定した、スケーラブルな業界標準の並列コードを作成

詳細とダウンロード

インテル® oneAPI ベース & HPC ツールキット

ダイレクト・プログラミング

- インテル® Fortran コンパイラー・クラシック
- インテル® Fortran コンパイラー
- インテル® oneAPI DPC++/C++ コンパイラー
- インテル® DPC++ 互換性ツール
- インテル® ディストリビューションの Python*
- oneAPI ベース・ツールキット用インテル® FPGA アドオン

API ベースのプログラミング

- インテル® MPI ライブラリー
- インテル® oneAPI DPC++ ライブラリー (インテル® oneDPL)
- インテル® oneAPI マス・カーネル・ライブラリー (インテル® oneMKL)
- インテル® oneAPI データ・アナリティクス・ライブラリー (インテル® oneDAL)
- インテル® oneAPI スレッディング・ビルディング・ブロック (インテル® oneTBB)
- インテル® oneAPI コレクティブ・コミュニケーション・ライブラリー (インテル® oneCCL)
- インテル® oneAPI ディープ・ニューラル・ネットワーク・ライブラリー (インテル® oneDNN)
- インテル® インテグレーテッド・パフォーマンス・プリミティブ (インテル® IPP)

解析/デバッグツール

- インテル® Inspector
- インテル® Trace Analyzer & Collector
- インテル® VTune™ プロファイラー
- インテル® Advisor
- インテル® ディストリビューションの GDB

インテル® **HPC** ツールキット +
インテル® oneAPI **ベース**・ツールキット

intel. HPC TOOLKIT 1 oneAPI

スケーラブルで高速な HPC アプリケーション

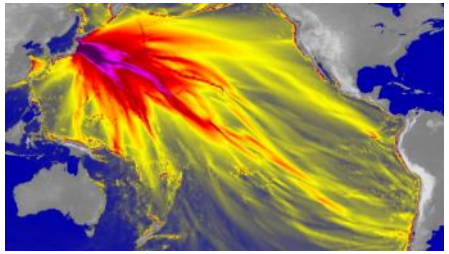
ユースケース – インテル® oneAPI ベース/HPC/レンダリング・ツールキット

intel 1 oneAPI BASE TOOLKIT | intel HPC TOOLKIT | intel RENDERING TOOLKIT

ユースケースほか

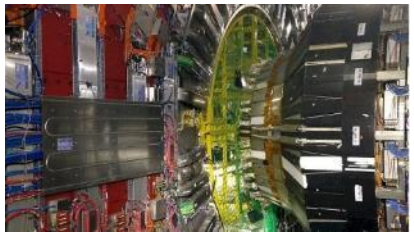
アルゴンヌ国立研究所のスーパーコンピューター Aurora (英語)
米国内初のエクサスケール・システムの 1 つである Argonne Leadership Computing Facility のインテル-HPE*/Cray* スーパーコンピューター Aurora 用のコードにインテル® ソフトウェア開発ツールを使用

計算密度と HBM により LAMMPS のエクサスケールの材料探索を高速化
インテルの CPU と GPU 上で動作する LAMMPS ワークロードがインテルの前世代と AMD を上回るパフォーマンスを実現

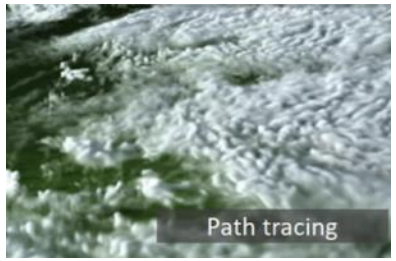


Zuse Institute Berlin (ZIB) の *easyWave* 津波シミュレーション・アプリケーション
CUDA* から oneAPI へ移行して**インテルの CPU、GPU、FPGA** と NVIDIA* P100 で優れたパフォーマンスを実現

HPC 向けに Google Cloud* を高速化
C2 は HPC ワークロードに優れたパフォーマンスを提供: インテル® Xeon® プロセッサーと AMD で**コアあたりのパフォーマンスが 40% 向上**し、インテル® oneAPI **ベース** & **HPC** ツールキットにより最適化
ビデオ (英語) | ポッドキャスト (英語)



HPC と AI 推論の高速化 (英語)
CERN、SURFsara、およびインテルは共同で、科学、エンジニアリング、金融アプリケーションで使用される**シミュレーションのパフォーマンスを飛躍的に向上させる**アプローチを研究



Texas Advanced Computing Center (TACC) (英語) の Frontera スーパーコンピューターのビジュアライゼーションおよびファイルシステムはインテル® Xeon® プロセッサーの大容量メモリー・ファット・ノードとインテル® Optane™ パーシステント・メモリーを使用†

ストックホルム大学/KTH (英語) は新薬の設計に用いられるシミュレーション・アプリケーション GROMACS を oneAPI で最適化し、CUDA* コードを SYCL* に移行してインテルの CPU と複数のアクセラレーターをターゲットにするマルチアーキテクチャー・コードを作成

HPC リーダーとの oneAPI ディスカッションのビデオ (英語) | HPC リーダーによる問題解決のアプローチの紹介ビデオ (英語)

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。ほかの情報も参考にして、正確かどうかを評価してください。
詳細は、http://www.intel.com/PerformanceIndex/ (英語) を参照してください。結果は異なることがあります。

インテルの AI ソフトウェア・ポートフォリオ

データ処理 → モデル作成 → 最適化とデプロイ

インテルにより最適化されたオープンソースの AI ツールとフレームワーク

- MODIN
- SciPy
- pandas
- NumPy
- Spark
- Numba

大規模なデータ・アナリティクス

- dmlc XGBoost
- PyTorch
- ONNX RUNTIME
- scikit learn
- TensorFlow
- DirectML
- SigOpt AutoML
- intel ニューラル・コンプレッサー

マシンラーニングとディープラーニングのフレームワーク、最適化、デプロイツール

インテルのディストリビューション

- OpenVINO
  1 つのコードをどこでもデプロイ

oneAPI

- インテル® oneDNN ライブラリー
- インテル® oneCCL
- インテル® oneMKL
- インテル® oneDAL

CPU、GPU、その他のアクセラレーター向けのオープンなクロスアーキテクチャー・プログラミング・モデル

クラウドとエンタープライズ

クライアントとワークステーション

エッジ

インテル® デベロッパー・クラウドとインテル® Developer Catalog
インテルの最新のツールとハードウェアを試し、最適化された AI モデルを利用可能

cnvrg.io
フルスタックの ML オペレーティング・システム

Hugging Face
インテルの最適化とファイン・チューニング、最適化された推論モデル、モデルの供給

注: スタックの各レイヤーにあるコンポーネントは、予想される AI 利用モデルに基づいて、ほかのレイヤーにある対象コンポーネント向けに最適化されており、すべてのコンポーネントが右端の列のソリューションで利用される訳ではありません。

インテルの AI ツール

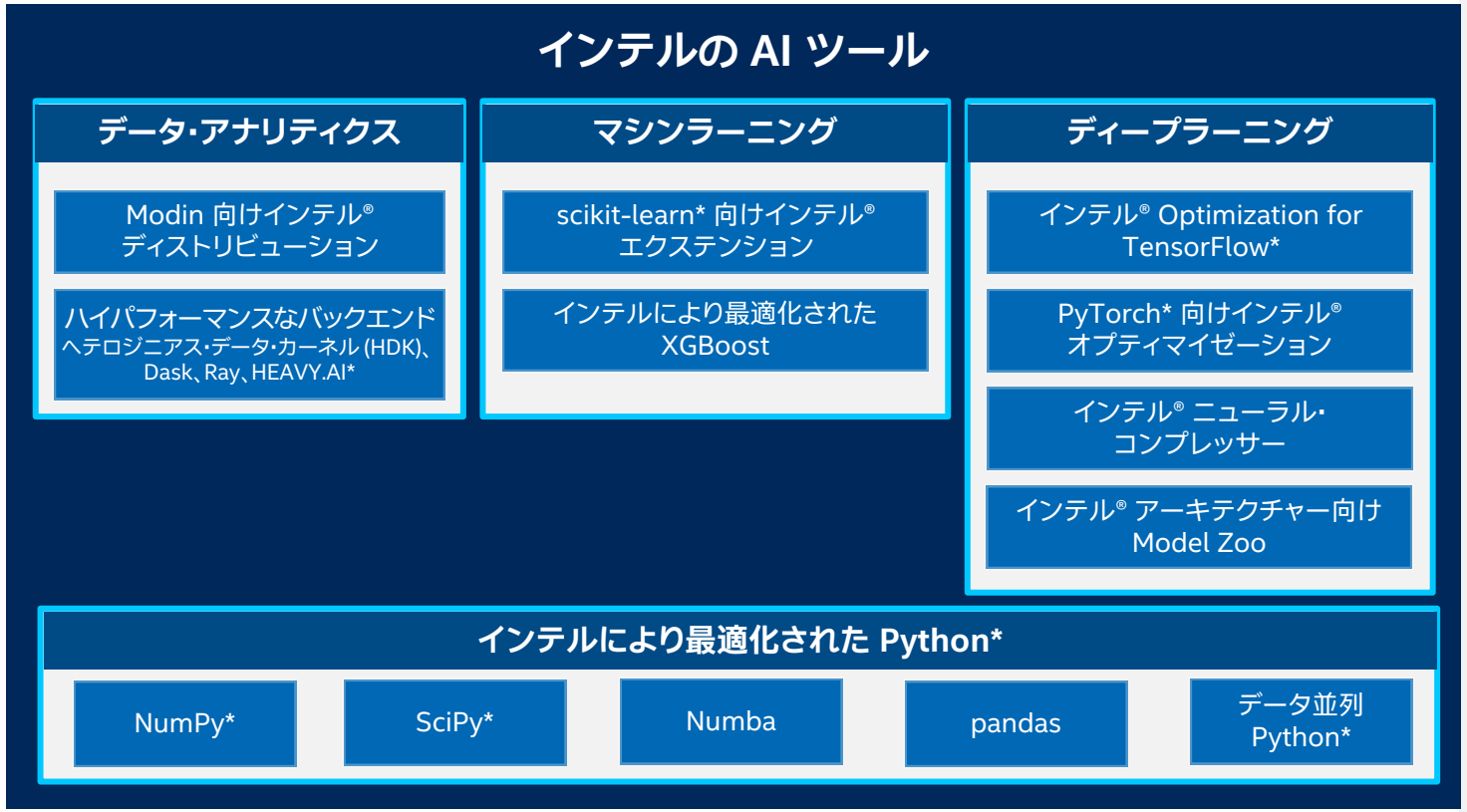
インテルの最新の最適化により、主要な AI ライブラリーとフレームワークを高速化し、そのすべてを新しく柔軟なウェブベースのインストール・ツールで利用できるようになりました。

対象ユーザー

データ・サイエンティスト、AI 研究者、マシンラーニングおよびディープラーニング開発者、AI アプリケーション開発者

主な特徴/利点

- 幅広いインテルの CPU および GPU ハードウェアから、AI ターゲット・プラットフォームを選択できます。
- ディープラーニング・フレームワーク、マシンラーニング・ライブラリー、インテルの AI ツールのパフォーマンスを大幅に向上します。
- 新しいウェブベースの AI Tools Selector を使用して AI ツールを素早く取得できます。AI Tools Selector は、必要なフレームワーク、ツール、ライブラリーを個別に、またはプリセットのバンドルで、正確に入手するための柔軟で合理的なプロセスを提供します。



インテル® ディストリビューションの OpenVINO™ ツールキット

powered by oneAPI

ハイパフォーマンスな DL 推論を提供

エッジからクラウドのビジョン/AI アプリケーションで使用されるハイパフォーマンスなディープラーニング推論とコンピューター・ビジョンの開発を高速化するツールキットで、ハードウェア・アクセラレーターでのディープラーニングやインテルの CPU、GPU、FPGA、VPU へのデプロイメントが可能です。

対象ユーザー

- コンピューター・ビジョン、ディープラーニング・ソフトウェア開発者
- データ・サイエンティスト
- OEM、ISV、システム・インテグレーター

使用法

セキュリティー監視、ロボット工学、小売、医療、AI、オフィス・オートメーション、交通、非ビジョン・ユースケース (音声、NLP、オーディオ、テキスト) など

詳細とダウンロード

エッジ AI とビジョン・アライアンス 2019 VISION PRODUCT OF THE YEAR

インテル® ディストリビューションの OpenVINO™ ツールキット

ディープラーニング

インテル® ディープラーニング・デプロイメント・ツールキット

モデル・オプティマイザー 変換および最適化 → IR → 推論エンジン 最適化された推論

IR = 中間表現ファイル

Open Model Zoo

- 40 以上の事前トレーニング済みモデル
- サンプル
- モデル・ダウンローダー

ディープラーニング・ワークベンチ

- キャリブレーション・ツール
- モデル・アナライザー
- ベンチマーク・アプリケーション
- 精度チェッカー
- AUX機能

従来のコンピューター・ビジョン

最適化されたライブラリーとコードサンプル

- OpenCV*
- OpenVX*
- サンプル

インテルの CPU と GPU/インテル® プロセッサー・グラフィックス

ツールとライブラリー

メディア/ビデオ/グラフィックス・パフォーマンスを向上

- インテル® メディア SDK オープンソース・バージョン
- OpenCL* ドライバーとランタイム

GPU/インテル® プロセッサー・グラフィックス

インテル® FPGA の最適化 (Linux* のみ)

- FPGA ランタイム環境 (インテル® FPGA SDK for OpenCL*)
- ビットストリーム

oneAPI を利用して作成された AI リファレンス・キット

一般的な AI ユースケースにおける導入の課題に対応

QR コードをスキャン

金融/保険

- 請求書類の自動化
- クレジットカード取引の不正検知
- 債務不履行リスク予測
- 災害査定プロセス

プロセス自動化

- 視覚処理検出
- インテリジェント・ドキュメント・インデックス
- 請求から入金までの自動化
- 歴史的資産の文書処理

製造業/公益事業

- ドローン・ナビゲーション・セグメンテーション
- 送電線障害検出
- 予測資産分析
- ビジュアル品質検査
- デジタルツイン
- エンジニアリング設計
- 交通カメラの物体検出
- 空気力学/流体流動プロファイル

ヘルスサイエンス/ライフサイエンス

- 医用画像処理
- 病気の予測
- セラピスト向け AI 文字起こし

カスタマーケア

- 購入予測
- 顧客セグメンテーション
- 顧客解約予測
- カスタマーケア・チャットボット
- 需要予測
- 製品推奨
- 受注から納品までの予測

テクノロジー/セキュリティー

- 垂直検索エンジン
- ネットワーク侵入検知
- データ保護
- IoT (データ・ストリーミング異常検出)

合成データ

- AI 合成データ (構造化)
- AI 合成データ (非構造化 – テキスト)
- AI 合成データ (非構造化 – 画像)
- AI 合成データ (非構造化 – 音声)

- 詳細: https://www.intel.com/aireferencekit (英語)
- GitHub* からダウンロード可能: https://github.com/oneapi-src (英語)

サーバー・プラットフォーム別 インテルの AI ソフトウェア (2023年4月)

■ インテル® Xeon® スケーラブル・プロセッサー
■ インテル® データセンター GPU
■ DL 向けインテル® Gaudi® AI アクセラレーター

カテゴリー	ソフトウェア	オープンソース	最適化のアップストリーム†	インテルの拡張††	インテルのディストリビューション	インテルのツール/キット
オーケストレーション	Cnvrg.io	×				■ ■ ■
最適化	BigDL	○				■
	OpenVINO™ ツールキット	○			■ ■	■ ■
	ニューラル・コンプレッサー	○				■ ■
	SigOpt	○				■ ■ ■
DL フレームワーク	TensorFlow*	○	■	■ ■	■ ■ ■	
	PyTorch*	○	■	■ ■	■	
	ONNX*	○	■			
	PDPD	○	■			
ML フレームワーク	XGBoost	○	■			
	scikit-learn*	○		■ ■		
	CatBoost	○	■			
	LightGBM	○	■			
データ前処理	Modin* (pandas 向け)	○	■		■	
	インテル® ディストリビューションの Python*	○			■ ■	
	Spark*	○	■	■		

† インテルは、できるだけ早く、できるだけ多くのハードウェア・ターゲット向けの最適化をアップストリームするよう努めています。
†† API 準拠の拡張機能を介して、インテルによる最適化とターゲット・ハードウェア・サポートを利用できます。

最適なツールの選択

ツールは相互に補完的 - AI の導入状況に応じて選択

- **データを調査して分析**しています。**モデルを開発**しています。
- **パフォーマンス**と、使用しているフレームワークとライブラリーとの**互換性**を求めています。
- コードをほとんど変更することなく**ドロップインで高速化**したいです。
- **新しいツールや言語は学びたくありません。**

データ・サイエンティスト/ML 開発者
インテルの AI ツール

- **モデルをデプロイ**しています。
- 複数のターゲット HW で**最先端のパフォーマンスと効率**を求めています。
- デプロイに不可欠な**メモリー・フットプリントを小さくする**ことに関心があります。
- **新しいツールや API を学んだり、採用することに抵抗がありません。**

アプリケーション開発者
インテル® ディストリビューションの OpenVINO™ ツールキット

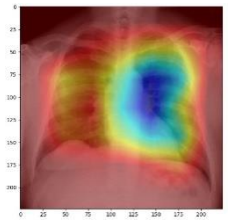
プリミティブを使用したり、oneAPI ライブラリー (oneDNN、oneCCL、oneDAL) を使用して
直接カーネルやアルゴリズムを最適化する場合は、インテル® oneAPI ベース・ツールキットを使用

AI ワークロードでエンドツーエンドのパフォーマンスを達成 – ユースケース

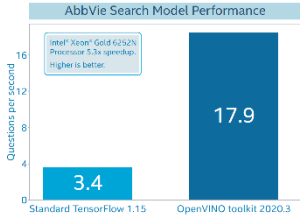
トレーニングと推論を高速化 - インテルの AI ツールで最適化

必要に応じてインテル® ディストリビューションの OpenVINO™ ツールキット、インテルの convrg.io、インテル® Gaudi®、インテル® oneAPI ベース・ツールキット、インテル® レンダリング・ツールキットも使用

ユースケースほか > その他のケーススタディー (英語)

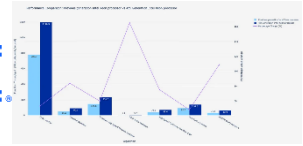


Accrad の AI ベースのソリューションが迅速な肺疾患診断を支援 (英語)
トレーニング + 推論の高速化

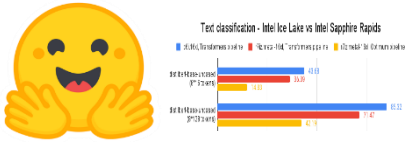


AbbVie 機械翻訳ソリューション (英語)
プロセッサーに最適化された機能により自然言語処理の推論モデルを高速化

RedHat はエンタープライズ規模でのサプライチェーン最適化
Red Hat* OpenShift* Data Science とインテル® アーキテクチャーを使用してオープンソースの AI テクノロジーを活用



自然言語処理: インテルのツールにより IBM が Watson NLP ライブラリーを高速化 (英語)



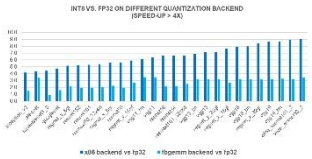
CPU 上で自然言語処理を大衆化 (英語)
Hugging Face とインテルのテクノロジーにより Falcon LLM 7B パラメーター・モデルのパフォーマンスを向上

インテル® AI アナリティクス・ツールキットと
インテル® ディストリビューションの OpenVINO™ ツールキットで最適化

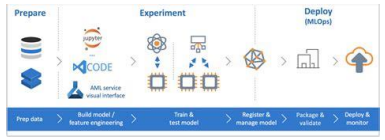
delphai AI ソリューションが B2B 企業検索エンジンを高速化 (英語)
インテル® Xeon® プロセッサー上でインテル® ニューラル・コンプレッサー、PyTorch* 向けインテル® エクステンション、インテル® AI Engine を使用して NLP モデルを高速化

AI がエッジに拡大 (英語)
インテル、Red Hat、Guise AI、OnLogic がエッジにインテリジェンスをもたらす

HippoScreen は oneAPI ツールにより AI パフォーマンスを 2.4 倍向上 (英語)
より幅広い精神疾患や病気にシステムの応用範囲を拡大



PyTorch* での x86 CPU 向け INT8 量子化 (英語)
INT8 量子化は x86 CPU プラットフォームで DL 推論を高速化する強力な手法



AZURE* ML - インテルの AI フレームワークにより Azure* Machine Learning モデルを最大化およびスケーリング (英語)
インテルと Microsoft は協力して scikit-learn* 向けインテル® エクステンションにより Azure* MaaS に AI の最適化をもたらす

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。ほかの情報も参考にして、正確かどうかを評価してください。

インテル® レンダリング・ツールキット

グラフィックス AI、レイトレーシング、ビジュアライゼーション、ほか

インテル® レンダリング・ツールキット (レンダリング・キット) は、オープンソースのレンダリング、レイトレーシング、ノイズ除去、パス・ガイディング・ライブラリーの強力なセットです。AI 合成データ生成、デジタルツイン、高忠実度のハイパフォーマンスなビジュアライゼーション、没入型コンテンツ作成など、さまざまな用途に使用できます。これらのライブラリーとインテルのハードウェア (CPU および GPU) を使用して、スケーラブルなソリューション・スタックを構成する、最適化されたレンダリング・パフォーマンスを提供します。

対象ユーザー

- シミュレーションと予測に合成データ生成を使用する **AI、ロボット工学、自律走行車の開発者**
- あらゆる状況を表現し、数学モデルを検証し、ML モデルをトレーニングするため、合成データを使用する **ML および DL アプリケーション**
- マルチテラバイトのデータセットを処理できる HPC 分散レンダリング・ソリューションで、in-situation シミュレーションや最高品質の画像を必要とする**研究科学者**
- ワークステーションからレンダリング・ファームまでスケーリング可能な統合ワークフローの恩恵を受ける、没入型体験を創造する必要のある**クリエーター、開発者、アーティスト**
- 製品のライフサイクル管理で正確なデジタルツインを使用する**製品デザイナーやエンジニア**

intel. RENDERING TOOLKIT 1 oneAPI

作品を より鮮やかに

インテルのレンダリング・ライブラリーと レイトレーシング・ライブラリー

インテル® Embree

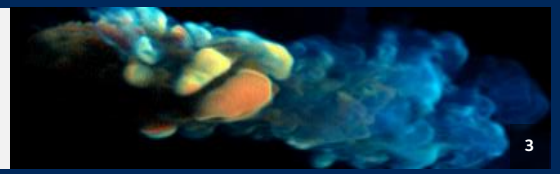
受賞歴を誇るハイパフォーマンスで機能豊富な レイトレーシングと写実的なレンダリング

インテル® オープン・イメージ・デノイズ

CPU、インテルおよびサードパーティーの GPU で優れたビジュアル品質を実現する AI で高速化されたデノイザー

インテル® オープン・ボリューム・カーネル・ライブラリー

3D 空間データ処理のレンダリングとシミュレーション



インテル® オープン PGL

重点サンプリングのパス・トレーシング、同じまたは より低い計算コストでより高い忠実度

インテル® OSPRay

スケーラブルで、移植可能な、分散レンダリング API HPC、SciVis、ProRender パストレーサー

インテル® OSPRay Studio

この新しいシーン・グラフ・アプリケーションを使用した GUI によるリアルタイム・レンダリング

インテル® OSPRay for Hydra

プラグインによりレンダリング・ツールキットのライブラリーから Universal Scene Description Hydra レンダリング・サブシステムに接続

詳細: xlsoft.com/jp/products/intel/oneapi/rendering/

イメージの出典 [1]『アダムス・ファミリー 2』- デジタル・ドメイン、マーベル・スタジオ、カオスグループ V-Ray*
[2] Scene © Blender Foundation | cloud.blender.org/spring
[3] 煙のシミュレーション、データ提供: OpenVDB サンプル・リポジトリー
[4] Blender Institute: Nishita Demo (CC-BY-ND)
[5] Copyright Animal Logic Pty Limited 2021. 無断での引用、転載を禁じます。

インテル® レンダリング・ツールキット

最高の忠実度でビジョンをレンダリング: 画期的なレイトレーシングへのオープンなパス

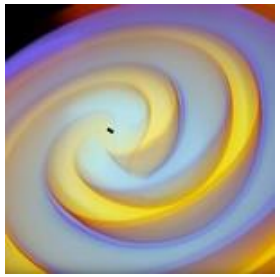
ユースケースほか > その他のケーススタディー (英語)

Blender は、インテルの CPU と GPU をサポートし、インテル® レンダリング・ライブラリーを使用することで、アーティストや 3D デザイナーに高度なレンダリング機能へのアクセスを提供
詳細: Blender 3.6LTS で GPU HW を解き放つ (英語)

映画『アダムス・ファミリー 2』で Cinesite はレンダリング効率を 10 ~ 25% 向上
詳細: ケーススタディー (英語)

Maxon の Cinema 4D は AI でトレーニングしたノイズ除去により CG ワークフローを強化 (英語)
ケーススタディー (英語) | CG ワークフローのセッション [7:03] (英語) | シーン制作/3D スキャンのチュートリアル [16:33] (英語)

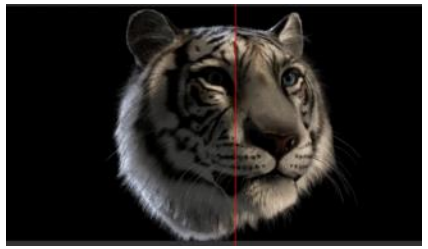
Bentley Motors Limited のカー・コンフィグレーター (英語)
ビジュアライゼーションと AI により 170 万以上のイメージをレンダリングして顧客にオプションを提供
PR (英語) | ビデオ [2:31] (英語) | デモ [6:58] (英語) | セッション [10:37] (英語)



スティーブン・ホーキング宇宙論研究所が宇宙物理学を視覚化 (英語)

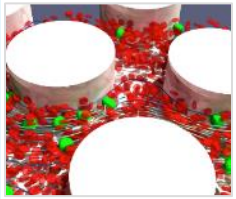
Tangent Studios (英語)
インテル® Embree と AI ベースのインテル® オープン・イメージ・デノイズを使用して、レンダリングを **1/5 ~ 1/6に削減**
ビデオ [2:35] (英語)

Texas Advanced Computing Center (TACC) (英語)
Frontera スーパーコンピューターのビジュアライゼーションおよびファイルシステムはインテル® Xeon® プロセッサーの大容量メモリー・ファット・ノードとインテル® Optane™ パーシステント・メモリーを使用



Mercenaries Engineering (英語)
プロダクション対応のツールを利用して AI ベースのアニメーションと VFX を提供し、レンダリング時間を最大 138% 向上

Chaos Group ケーススタディー
ビデオ [2:22] (英語)
セッション [17:42] (英語)



HPC セクション: インテル® ソフトウェア開発ツールが Aurora スーパーコンピューター向けのコードの準備を支援 (英語)

導入事例 [ビデオ 2:34] (英語)

[1]出展: ブレンダーとインテル® Embree を使用した Tangent Animation と Baozou Production の共同制作。メディアの出典: Netflix, Inc.、現在 Netflix で配信中。視聴するには Netflix への加入が必要です。
*構成の詳細は、スライドのノートを参照してください。
インテル® ソフトウェア製品のパフォーマンスおよび最適化に関する注意事項については、software.intel.com/articles/optimization-notice#opt-jp を参照してください。
詳細は、http://www.intel.com/PerformanceIndex/ (英語) を参照してください。結果は異なることがあります。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。ほかの情報も参考にして、正確かどうかを評価してください。.

法務上の注意書き (1/2)

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。結果は異なることがあります。

性能の測定結果はシステム構成の日付時点のテストに基づいています。また、現在公開中のすべてのセキュリティー・アップデートが適用されているとは限りません。構成の詳細は、補足資料を参照してください。
絶対的なセキュリティーを提供できる製品またはコンポーネントはありません。

Texas Advanced Computing Center (TACC) の Frontera に関する参考文献
記事: *HPCWire: Visualization & Filesystem Use Cases Show Value of Large Memory Fat Notes on Frontera* (英語)
www.intel.com/content/dam/support/us/en/documents/memory-and-storage/data-center-persistent-mem/Intel-Optane-DC-Persistent-Memory-Quick-Start-Guide.pdf (英語)
software.intel.com/content/www/us/en/develop/articles/introduction-to-programming-with-persistent-memory-from-intel.html (英語)
wreda.github.io/papers/assise-osdi20.pdf (英語)

KFBIO
KFBIO m. tuberculosis screening detectron2 model throughput performance on 2nd Intel® Xeon® Gold 6252 processor: NEW: Test 1 (single instance with PyTorch 1.6: Tested by Intel as of 5/22/2020. 2-socket 2nd Gen Intel® Xeon® Gold 6252 Processor, 24 cores, HT On, Turbo ON, Total Memory 192 GB (12 slots/16 GB/2666 MHz), BIOS: SSE5C620.86B.02.01.0008.031920191559 (ucode: 0x500002c), Ubuntu 18.04.4 LTS, kernel 5.3.0-51-generic, mitigated Test 2 (24 instances with PyTorch 1.6: Tested by Intel as of 5/22/2020. 2-socket 2nd Gen Intel Xeon Gold 6252 Processor, 24 cores, HT On, Turbo ON, Total Memory 192 GB (12 slots/16 GB/2666 MHz), BIOS: SSE5C620.86B.02.01.0008.031920191559 (ucode: 0x500002c), Ubuntu 18.04.4 LTS, kernel 5.3.0-51-generic, mitigated BASELINE: (single instance with PyTorch 1.4): Tested by Intel as of 5/22/2020. 2-socket 2nd Gen Intel Xeon Gold 6252 Processor, 24 cores, HT On, Turbo ON, Total Memory 192 GB (12 slots/16 GB/2666 MHz), BIOS: SSE5C620.86B.02.01.0008.031920191559 (ucode: 0x500002c), Ubuntu 18.04.4 LTS, kernel 5.3.0-51-generic, mitigated.

Tangent Studios
Configurations for Render Times with Intel® Embree, testing conducted by Tangent Animation Labs. Render farm: 8x Intel® Core™ processors +hyperthread*2 + 128gig. In-office workstations: Intel® Xeon® processors HP blade c7000 chassis, with HP460 gen8 blades - 2x Intel Xeon E5-2650 V2, Eight Core 2.6GHz-128GB. Software: Blender 2.78 with custom build using Intel® Embree. For more information on Tangent's work with Embree, watch this video:
www.youtube.com/watch?time_continue=251&v=_2Ia4h8q3xs&feature=emb_logo (英語)
Recreation of the performance numbers can be recreated using Agent327, Blender and Embree.

Chaos Group - Up to 90% Memory Reduction for Displacement
Testing conducted by Chaos Group with Intel® Embree 2020. Software Corona Renderer 5 with Intel Embree. Up to 90% memory reduction calculated using Corona Renderer 5 with regular displacement grids per triangle of 154 bytes versus Corona Renderer 5 with Intel Embree, which has a displacement capability grid of 12 bytes per grid triangle. (12/154 = 7.8% usage or >90% memory reduction.) Recreation of the performance numbers can be accomplished using Corona Renderer 5 and Embree. For more information, visit the Corona Renderer Blog: blog.corona-renderer.com/corona-renderer-5-for-3ds-max-released/ (英語)

The Addams Family 2 - Gained a 10% to 20%—and sometimes 25%—efficiency in rendering, saving thousands of hours in rendering production time.
Testing Date: Results are based on data conducted by Cinesite 2020-21. 10% to up to 25% rendering efficiency/thousands of hours saved in rendering production time/15 hrs per frame per shot to 12-13 hrs.
Cinesite Configuration: 18-core Intel® Xeon® Scalable processors (W-2295) used in render farm, 2nd gen Intel Xeon processor-based workstations (W-2135 and -2195) used. Rendering tools: Gaffer, Arnold, along with optimizations by Intel® Open Image Denoise.

実際の費用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。

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。ほかの情報も参考にして、正確かどうかを評価してください。

© 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。
SYCL は Khronos Group 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。

法務上の注意書き (2/2)

HippoScreen increased AI performance by 2.4x – Test configurations Intel® Xeon® Gold 6330 CPU @ 2.00 GHz, 2 sockets, 28 cores per socket, note: www.intel.com /performanceindex

Univ. of Calif. at Davis achieved 3x performance increase & >100x data compression for scientific rendering – VIDI researchers achieved compression rates of several hundred times by implementing a new compression mechanism that uses a combination of positional encoding (Figure 2) and multilayer perceptrons (Figure 3) to learn a mapping from sampling positions to volume densities. Figure 1. A comparison of training results between the proposed method and two state-of-the-art techniques: fV-SRN, which was adjusted to match the models' compression ratios, and tthresh, which was adjusted to match our models' PSNRs after 20k steps of training.Configurations: Models were trained on Windows* machine with RTX 3090, while fV-SRN models were trained on a Linux* machine with a faster RTX 3090Ti and Intel® Xeon® Scalable processor (E5-2699) due to operating system compatibility issues. The tthresh experiments were run on an 88-core Linux server with 256 GB of memory because tthresh is a CPU-based algorithm. The table indicates experiments that ran out of memory (as OOM). The method outperforms state-of-the-art techniques, with the best and worst results within each category highlighted in red and gray, respectively. Figure 13. Relative speed-up times when compared to the baseline ray-marching renderer. An Intel® Xeon® Scalable processor (E5-2699) with 256 gigabytes of memory were the 88-core/176-thread workhorses used to render datasets and train the machine learning models that powered these projects. More details are in *Instant Neural Representation for Interactive Volume Rendering*

The Addams Family 2/Cinesite achieving up to 25% efficiency in rendering - Testing Date: Results are based on data conducted by Cinesite 2020-21. 10% to up to 25% rendering efficiency/thousands of hours saved in rendering production time/15 hrs per frame per shot to 12-13 hrs. Cinesite Configuration: 18-core Intel® Xeon® Scalable processors (W-2295) used in render farm, 2nd gen Intel Xeon processor-based workstations (W-2135 and -2195) used. Rendering tools: Gaffer, Arnold, along with optimizations by Intel® Open Image Denoise.

University of Utah -: the performance for growing problem sizes (from 10'000x10'000 pixels to 40'000x40'000 mosaic) using Intel CPU (Intel® Xeon® Platinum 8480+) and GPU (Intel® Data Center GPU Max 1100) available on the Intel® Developer Cloud and using the original code on an NVIDIA A6000 on a workstation.

Ali- DP **Testing Date:** Performance results are based on **testing by Alibaba as of August 21, 2022**. Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Configuration Details and Workload Setup:. 3rd Gen Intel® Xeon® Scalable processors 8369B CPU @ 2.70GHz, 32v CPU(s), 64G memory, 40G ESSD. Lammps configuration file: Lammps default configuration file. Release: 23 Jun 2022, Iteration Count: 2M, Number of test processes and threads: 32P1T, 16P2T, 8P4T. Comparing compilers: GCC-10.2, Intel(R) oneAPI DPC++/C++ Compiler 2022.0.0 (2022.0.0.20211123). Performance evaluation indicators: the exectuion time. Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See configuration disclosure for details. No product or component can be absolutely secure. Performance varies by use, configuration, and other factors. Learn more at www.Intel.com/PerformanceIndex. Your costs and results may vary

性能は、使用状況、構成、その他の要因によって異なります。
実際の費用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。
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。ほかの情報も参考にして、正確かどうかを評価してください。

© 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。
SYCL は Khronos Group 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。

UXL の関連情報

- Unified Acceleration Foundation プレスリリース: https://www.linuxfoundation.org/press/announcing-unified-acceleration-foundation-uxl (英語)
- Codeplay の Rod Burns によるブログ: https://www.oneapi.io/blog/announcing-the-unified-acceleration-uxl-foundation/ (英語)
- インテルの Sanjiv Shah によるブログ: https://www.oneapi.io/blog/our-kids-graduating-from-college/ (英語)

- https://uxlfoundation.org/ (英語)
- https://www.oneapi.io/ (英語)

- embedded:「oneAPI motivates new foundation for open standard accelerated compute」(英語)
- Phoronix:「Intel oneAPI Initiative Evolves Into The Unified Acceleration "UXL" Foundation」(英語)
- Techzine Europe:「Unified Acceleration Foundation (UXL) gives oneAPI a boost, and more governance」(英語)
- TechCrunch:「The Unified Acceleration Foundation Wants to Create an Open Standard for Accelerator Programming」(英語)
- Inside HPC:「Linux Foundation: Unified Acceleration Foundation Formed for Open Accelerated Compute and Cross-Platform Performance」(英語)
- ITOps Times:「Linux Foundation forms new group to drive common model for developing cross-platform applications」(英語)
- The New Stack:「OneAPI Computing Aims to Ease Multi-Architecture Computing」(英語)
- The Next Platform:「The New UXL Foundation Has a Bold Blueprint for Open Acceleration」(英語)
- Forbes:「Unified Acceleration Foundation To Broaden OneAPI's Use For Heterogeneous Programming」(英語)
- VentureBeat:「Intel CTO highlights open and secure advances for AI deployment」(英語)
- ComputerWeekly:「Cross-architecture unification: Linux Foundation forms Unified Acceleration (UXL) Foundation」(英語)

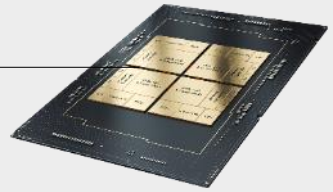
- EENews Europe:「UXL looks to standardise heterogeneous compute」(英語)
- Golem.de:「Linux-Foundation will einheitlichen Beschleuniger-Zugriff」(英語)
- Heise:「Mittwoch: Helfersuche für Google Maps, Social Engineering für Cyberangriffe」(英語)
- Runtime News:「The Race for a Nvidia Alternative」(英語)
- SiliconAngle:「Intel, Samsung and other chipmakers back new accelerator programming initiative」(英語)

Intel Confidential. Share under NDA only.

intel®

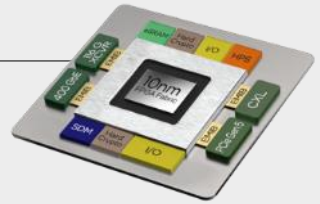
多様なハードウェア・アクセラレーションの課題を解決

CPU



GPU

FPGA



その他のアクセラレーター

開発者の 48% が 2 種類以上のプロセッサーやコアを使用する
ヘテロジニアス・システムをターゲットにしている[1]

開発者の課題: 複数のアーキテクチャー、ベンダー、プログラミング・モデル

oneAPI

今業界に必要なベンダーに依存しないプログラミング

ビデオ: oneAPI とは? 概要と利点 (英語)

1. 2022年6月に Evans Data が実施した「Global Development Survey Report 22.1」

インテル® oneAPI ツールキット – 実証済みのパフォーマンス

ユースケース

HPC マルチアーキテクチャー・システム

- アルゴンヌ国立研究所 (英語) は現在のインテルの CPU と GPU 上でエクサスケール規模のアプリケーションを効率良く実行
- Zuse Institute Berlin (ZIB) (英語) は津波をシミュレーション・アプリケーション「easyWave」を CUDA* から SYCL* に移行し、マルチベンダーの複数のアーキテクチャーでパフォーマンスを発揮
- GROMACS (英語) は oneAPI とインテルのツールにより高速化し、CUDA* コードを SYCL* に移行して、インテル® GPU 上で実行
- LAMMPS (英語) ワークロードは、インテル® Xeon® CPU マックス・シリーズ上で 6 基のインテル® データセンター GPU マックス・シリーズにカーネルをオフロードし、最大 **12.8 倍**のパフォーマンス向上を達成して AMD を凌駕

HPC & AI – CERN はインテル® DL ブーストと oneAPI を利用して (英語)、精度を落とさずに推論を約 **2 倍**高速化してシミュレーションをスピードアップ†

クラウド/CSP – QCT (英語) はインテル® VTune™ プロファイラーを使用して、SYCL* 対応 PolyBench-ACC ベンチマークのパフォーマンスを、OpenMP* ベースライン実装と比較して 38% 向上

コンテンツ作成 – Blender はインテルの CPU と GPU で oneAPI をサポートし、インテルの複数のレンダリング・ライブラリー (英語) を使用してレンダリングの忠実度とパフォーマンスを向上

高度なレイトレーシングを使用した超現実的なビジュアライゼーションと AI – Bentley Motors Limited の AI ベースのカー・コンフィグレーター (英語) は **170 万枚以上**の画像を処理し、1 モデルあたり最大 **10 億**通りの設定が可能。Cinesite は『アダムス・ファミリー 2』(英語) でインテルのデノイザーによりレンダリング効率を **25%** 向上

IoT – Samsung Medison (英語) は複数のインテル® アーキテクチャー上でエッジの超音波診断を高速化して精度の向上と迅速な診断を実現

主要な CSP とフレームワークが oneAPI を支持 (英語) – Microsoft* Azure*、Google Cloud*、TensorFlow*

FPGA – oneAPI を使用することで **Bittware** (英語) は Verilog* や VHDL* では数週間かかるアプリケーションを**数日**で実行†

ビデオ [3:45] (英語)

†各事例の詳細は、「oneAPI Customer Use Cases」デックに記載されています。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。ほかの情報も参考にして、正確かどうかを評価してください。詳細は、「法務上の注意書き」を参照してください。

データ並列 C++

oneAPI の SYCL* 実装

DPC++ = ISO C++ と Khronos SYCL* およびコミュニティーによる拡張

選択の自由: 将来に対応したプログラミング・モデル

- 複数のハードウェア・ターゲットでコードを再利用可能
- 特定のアクセラレーター向けのカスタム・チューニングが可能
- 専用の言語に代わる、オープンな業界全体の代替手段

DPC++ = ISO C++ と Khronos SYCL* およびコミュニティーによる拡張

- データ並列プログラミングの生産性を向上
- 標準 C++ と同等のネイティブな高水準言語パフォーマンスを提供
- Khronos Group の SYCL* を継承し、データ並列処理とヘテロジニアス・プログラミングをサポート

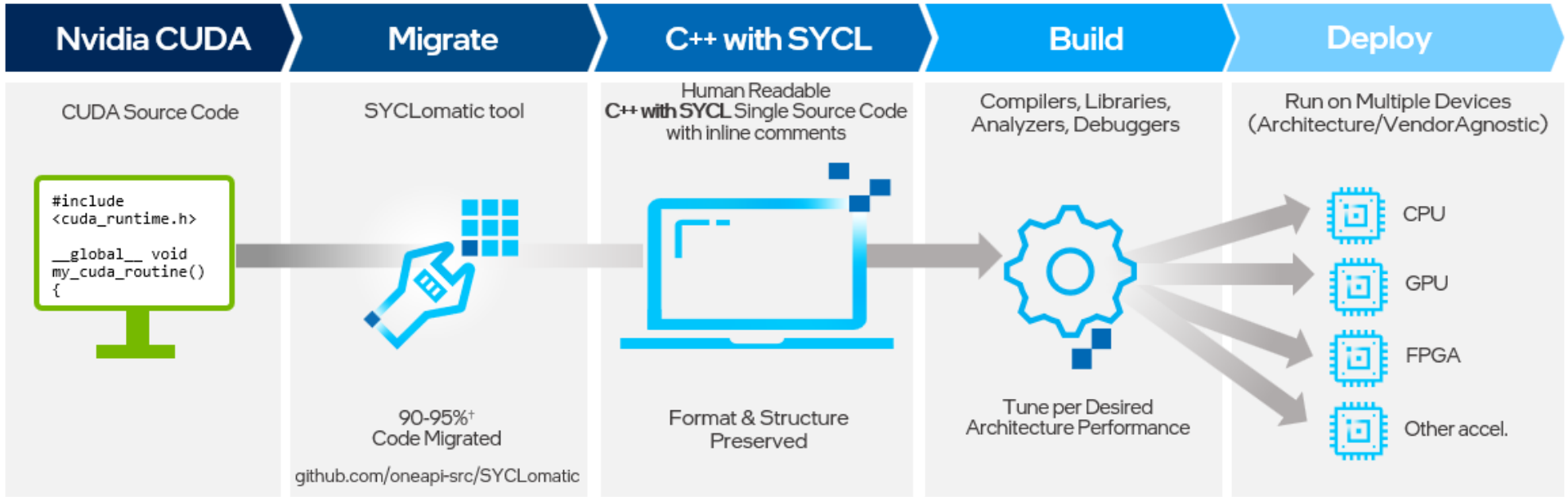
コミュニティー・プロジェクトを通じて言語の拡張を推進

- データ並列プログラミングを簡素化する拡張機能を提供
- オープンな共同開発により継続的に進化

オープンソースおよびインテルの DPC++/C++ コンパイラーはインテルの CPU、GPU、FPGA をサポートしています。
Codeplay* は、NVIDIA* GPU をターゲットとする DPC++ コンパイラー (英語) を発表しています。

SYCLomatic: CUDA* から SYCL* へ簡単に移行

開発者がソフトウェアを実行する場所を選択



オープンソースの SYCLomatic ツールは、すでに CUDA* で記述されているコードを C++ with SYCL* に移行する開発者を支援し、可能な場合は**人間が解読可能な**コードを生成

通常はコードの 90-95% を自動的に移行[1]

開発者がアプリケーションの移行を完了できるように支援するインラインコメントを提供

インテル® DPC++ 互換性ツールはインテル® oneAPI ベース・ツールキットに含まれるインテルの実装

github.com/oneapi-src/SYCLomatic
(英語)

[1]出典: 2023年3月現在のインテルによる推定。Rodinia、SHOC、PENNANT など、85 種類の HPC ベンチマークとサンプルの測定結果に基づいています。結果は異なることがあります。