

# ビッグデータ解析および機械学習の パフォーマンスを大幅に向上

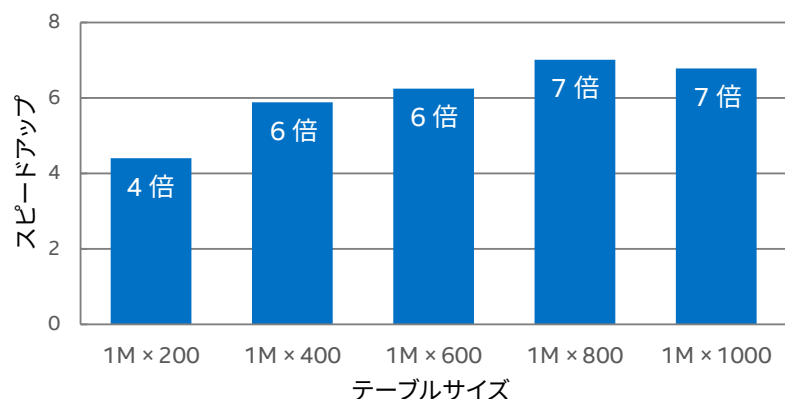
## インテル® DAAL

インテル® ソフトウェア開発ツール

### 使いやすいライブラリーによりビッグデータ解析と機械学習の パフォーマンスを大幅に向上

- インテル® アーキテクチャー・ベースのデバイス全体にわたって高いアプリケーション・パフォーマンスを実現
- データソースと環境の統合により、成果が得られるまでの時間を短縮
- あらかじめ最適化された幅広い高度な解析アルゴリズムによりアプリケーション開発時間を短縮

ビッグデータにより、コンピューティングの世界では、多種多様な業界 / 分野において目覚ましい速度で生成される膨大な量と種類のデータから値を抽出して処理する必要性が増えています。ゲノム解析、リスク解析、ソーシャル・ネットワーク / 消費者選好解析は、大規模なデータセットのハイパフォーマンスな解析が不可欠とされるいくつかの例です。



システム構成: バージョン: インテル® DAAL 2016、CDH v5.3.1、Apache Spark® v1.2.0。ハードウェア: インテル® Xeon® プロセッサ E5-2699 v3、2 x 18 コア CPU (45MB LLC、2.3GHz)、ノードあたり 256GB RAM。オペレーティング・システム: CentOS® 6.6 x86\_64。

図 1: PCA パフォーマンスの向上 — インテル® DAAL vs Spark MLlib

ほとんどの場合、計算速度が成功の鍵となります。インテル® Data Analytics Acceleration Library (インテル® DAAL) は、ソフトウェア開発者がアプリケーションの開発期間を短縮し、パフォーマンスを

最大限に引き出せるように支援します。インテル® DAAL を利用することで、より正確に素早く予測し、利用可能な計算リソースで大きなデータセットを解析することができます。インテル® DAAL のアップデートは、次世代のプロセッサを利用できるように、それらのプロセッサが出荷される前に提供されます。最新バージョンにリンクするだけで、コードを最新のプロセッサに対応させることができます。



図 2: インテル® DAAL はデータ解析パイプラインのすべてのステージに対応

インテル® DAAL は、業界トップの算術ライブラリー<sup>†</sup>である、インテル® マス・カーネル・ライブラリー (インテル® MKL) と同じチームによって開発されています。インテル® DAAL 開発チームは、インテル® プロセッサのアーキテクトと緊密に連携して、インテル® プロセッサ・ベースのシステムでパフォーマンスを最大限に引き出せるように取り組んでいます。インテル® DAAL は、インテル® Parallel Studio XE スイートまたはスタンドアロンで利用できます。

## ハードウェア向けに最適化

最大限の計算速度を達成するには、各ターゲット・プロセッサの命令セット、レジスター幅、メモリー・アーキテクチャー向けに各関数が高度にチューニングされていなければなりません。IoT (Internet of Things) ゲートウェイとバックエンド・サーバー間のように複数のプラットフォームで解析処理を分割するとアプリケーションが利点を得られる場合があるため、インテル® DAAL は、IoT ゲートウェイからバックエンド・サーバーまでをターゲットとする、インテル® Atom™ プロセッサ、インテル® Core™ プロセッサ、インテル® Xeon® プロセッサ、インテル® Xeon Phi™ コプロセッサを含む幅広いインテル® プロセッサ向けにチューニングされています。

## アルゴリズム

### データ解析：特性評価、集計、変換

アルゴリズム	説明
低次モーメント	データセットの基本特性 (合計、平均、2 次モーメント、差異、標準偏差など) を計算します。
クォンタイル	クォンタイルの位数で等分されたデータ分布を集計するクォンタイルを計算します。
相関行列と分散共分散行列	特徴ベクトルのペアの統計的関係を定量化します。
コサイン距離行列	コサイン距離を使用して、特徴ベクトルのペアの類似性を評価します。
相関距離行列	相関距離を使用して、特徴ベクトルのペアの類似性を評価します。
コレスキー分解	対称正定値行列を下位三角行列とその転置の積に分解します。線形システム、非線形最適化、カルマンフィルターなどを解くための基本操作として使用されます。
QR 分解	一般行列を直交行列と上三角行列の積に分解します。線形逆問題と最小二乗問題を解くのに使用されます。また、固有値と固有ベクトルの計算にも使用されます。
特異値分解 (SVD)	行列を左特異ベクトル、特異値、右特異ベクトルの積に分解します。主成分分析、線形逆問題の計算、データ・フィッティングに使用されます。
主成分分析 (PCA)	入力特異ベクトルを互いに直交する主成分の新しいセットに変換することでデータの次元を減らします。
K 平均法	データセットをほぼ同じデータポイント数のクラスターに分割します。各クラスターは、そのクラスターのすべてのデータポイントの平均であるセントロイド (中心) により表されます。
期待値最大化	モデルのパラメーターの最大尤度を推定します。クラスタリング手法としてガウス混合モデルで使用されます。非線形次元縮退、欠測値問題にも使用できます。
外れ値検出	異常に離れた観測点とほかの観測点を識別します。特徴ベクトル全体 (多変量) または 1 つの特徴値 (単変量) を考慮して、外れ値かどうか決定できます。
相関ルール	特定の信頼レベルで変数の関係を検出します。
線形および放射基底関数カーネル	データをより高い次元の空間にマップします。
品質メトリック	数値のセットを計算し、解析アルゴリズムによって返される結果の定量的特性を評価します。メトリックには、混同行列、正解率、適合率、再現率、F 値などが含まれます。

## 機械学習：回帰、分類、その他

アルゴリズム	説明
線形回帰	線形方程式を観測データに当てはめることで、従属変数と1つ以上の説明変数の関係をモデル化します。
ナイーブベイズ分類器	ラベルを割り当てることで、観測値を別々のクラスに分割します。ナイーブベイズは、特徴間の独立性を仮定する確率的分類器です。テキスト分類と医療診断によく使用され、特徴間にある程度の従属関係がある場合でもうまく動作します。
ブースティング	弱分類器の正解率に応じて繰り返し重み付けを見直し、一連の重み付けされた弱分類器をまとめて強分類器を作成します。弱分類器が決定木となります。AdaBoost (二項分類器)、BrownBoost (二項分類器)、LogitBoost (多クラス分類器) を含むブースティング・アルゴリズムを利用できます。
サポート・ベクトル・マシン (SVM)	SVM はポピュラーな二項分類器です。観測された特徴ベクトルを2つのクラスに分割する超平面を計算します。
多クラス分類器	SVM などの二項分類器を使用して多クラス分類器を作成します。

## 1年間の製品サポートとアップデート

製品を購入すると、製品のサポート、アップデート、新しいリリースに加えて、古いバージョンも利用することができます。また、製品に関する質問を直接問い合わせたり、ユーザー・コミュニティ (フォーラム) やセルフヘルプ・ドキュメントから製品に関するさまざまな情報を得ることができます。

## 動作環境

プロセッサ	複数の世代のインテル® プロセッサと互換プロセッサをサポートしています。例：インテル® Core™ プロセッサ、インテル® Xeon® プロセッサ、インテル® Xeon Phi™ コプロセッサ。
言語	Microsoft* コンパイラ、GCC、インテル® コンパイラと互換。C、C++、C#、Fortran、Java*、ASM
オペレーティング・システム	Windows*、Linux*、および OS X* (OS X* 開発者は Composer Edition の C++ または Fortran バージョンを選択可能)。
開発環境	<b>Windows*:</b> Microsoft* Visual Studio* に統合 <b>Linux*:</b> GNU* ツール互換 <b>OS X*:</b> XCode*
関連情報	<a href="http://www.intel.com/software/products/systemrequirements/">www.intel.com/software/products/systemrequirements/</a> (英語)

詳細は、[リリースノートおよびドキュメント \(英語\)](#) を参照してください。

## インテル® Parallel Studio XE スイートに含まれるコンポーネント

インテル® DAAL は、並列ソフトウェア開発向けの統合ソフトウェア開発スイートであるインテル® Parallel Studio XE またはスタンドアロンで使用できます。

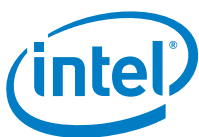
### 関連情報

#### インテル® DAAL

[software.intel.com/en-us/intel-daal](http://software.intel.com/en-us/intel-daal)

#### 30 日間の評価版

[software.intel.com/en-us/intel-daal/try-buy](http://software.intel.com/en-us/intel-daal/try-buy)



† 出典：Evans Data Software Developer surveys 2011-2015

本資料に掲載されている情報は、インテル製品の概要説明を目的としたものです。本資料は、明示されているか否かにかかわらず、また禁反言によるとらえずにかかわらず、いかなる知的財産権のライセンスも許諾するものではありません。製品に付属の売買契約書『Intel's Terms and Conditions of Sale』に規定されている場合を除き、インテルはいかなる責任を負うものではなく、またインテル製品の販売や使用に関する明示または黙示の保証 (特定目的への適合性、商品適格性、あらゆる特許権、著作権、その他知的財産権の非侵害性への保証を含む) に関していかなる責任も負いません。インテルによる書面での合意がない限り、インテル製品は、その欠陥や故障によって人身事故が発生するようなアプリケーションでの使用を想定した設計は行われていません。

性能に関するテストに使用されるソフトウェアとワークロードは、性能がインテル® マイクロプロセッサ用に最適化されていることがあります。SYSmark\* や MobileMark\* などの性能テストは、特定のコンピュータ・システム、コンポーネント、ソフトウェア、操作、機能に基づいて行なったものです。結果はこれらの要因によって異なります。製品の購入を検討される場合は、他の製品と組み合わせた場合の本製品の性能など、ほかの情報や性能テストも参考にして、パフォーマンスを総合的に評価することをお勧めします。ベンチマークの出典：インテル コーポレーション

最適化に関する注意事項：インテル® コンパイラは、互換マイクロプロセッサ向けには、インテル製マイクロプロセッサ向けと同等レベルの最適化が行われられない可能性があります。これには、インテル® ストリーミング SIMD 拡張命令 2 (インテル® SSE2)、インテル® ストリーミング SIMD 拡張命令 3 (インテル® SSE3)、ストリーミング SIMD 拡張命令 3 補足命令 (SSSE3) 命令セットに関連する最適化およびその他の最適化が含まれます。インテルでは、インテル製ではないマイクロプロセッサに対して、最適化の提供、機能、効果を保証していません。本製品のマイクロプロセッサ固有の最適化は、インテル製マイクロプロセッサでの使用を目的としています。インテル® マイクロアーキテクチャに非固有の特定の最適化は、インテル製マイクロプロセッサ向けに予約されています。この注意事項の適用対象である特定の命令セットに関する詳細は、該当する製品のユーザー・リファレンス・ガイドを参照してください。改訂 #20110804

インテル製品は、予告なく仕様や説明が変更されることがあります。機能または命令の一覧で「留保」または「未定義」と記されているものがありますが、その「機能が存在しない」あるいは「性質が留保付である」という状態を設計の前提にしないでください。これらの項目は、インテルが将来のために留保しているものです。インテルが将来これらの項目を定義したことにより、衝突が生じたり互換性が失われたりしても、インテルは一切責任を負いません。この情報は予告なく変更されることがあります。この情報だけに基つて設計を最終的なものとししないでください。

本資料で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、公表されている仕様とは異なる動作をする場合があります。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。最新の仕様をご希望の場合や製品をご注文の場合は、お近くのインテルの営業所または販売代理店にお問い合わせください。本資料で紹介されている資料番号付きのドキュメントや、インテルのその他の資料を入手するには、1-800-548-4725 (アメリカ合衆国) までご連絡いただくか、インテルの Web サイト ([www.intel.com](http://www.intel.com)) を参照してください。

© 2015 Intel Corporation. 無断での引用、転載を禁じます。Intel、インテル、Intel ロゴ、Intel Atom、Intel Core、Intel Xeon Phi、Xeon は、アメリカ合衆国および / またはその他の国における Intel Corporation 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。