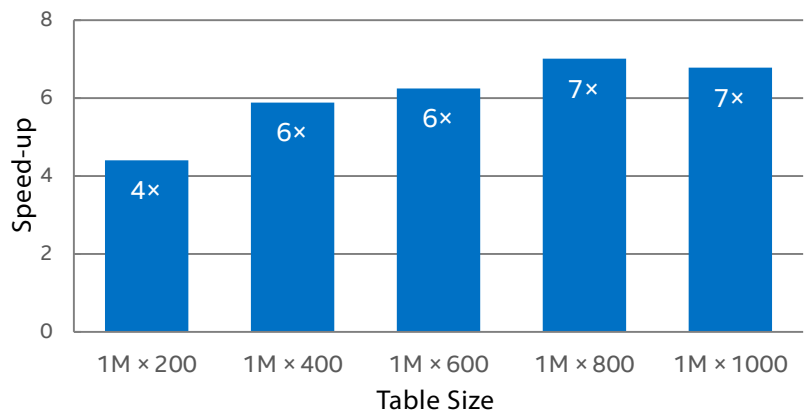intel

Software

Boost Big Data Analytics and Machine Learning Performance

Intel® Data Analytics Acceleration Library
Intel Software Development Tools

Boost Big Data Analytics and Machine Learning Performance with Easy-to-Use Library

- Delivers high application performance across spectrum of Intel®-architecture devices

- Speeds time-to-value through data source and environment integration

- Reduces application development time via wide selection of pre-optimized advanced analytics algorithms

Big data is changing the world of computing by extracting value from the increasing volume, variety and velocity of data generated in many different industries and domains. Genomics, risk, social network and consumer preference analysis are just a few examples of areas where high performance analysis of large data sets is a critical competency in today's compute landscape.

Configuration – Versions: Intel® Data Analytics Acceleration Library 2016, CDH v5.3.1, Apache Spark* v1.2.0; Hardware: Intel® Xeon® Processor E5-2699 v3, 2 Eighteen-core CPUs (45MB LLC, 2.3GHz), 256GB of RAM per node; Operating System: CentOS 6.6 x86_64.

Figure 1: PCA Performance Boost Using Intel DAAL vs. Spark MLLib

For most of these challenges, computational speed is a key ingredient for success. The Intel Data Analytics Acceleration Library (Intel® DAAL) is designed to help

software developers reduce the time it takes to develop their applications and delivery them with maximum performance. Intel DAAL helps you make better predictions — faster, and analyze larger data sets with the available compute resources at hand. Intel DAAL is updated to take advantage of next-generation processors even before they are available. Just link to the newest version, and your code is ready for when those new chips hit the market.
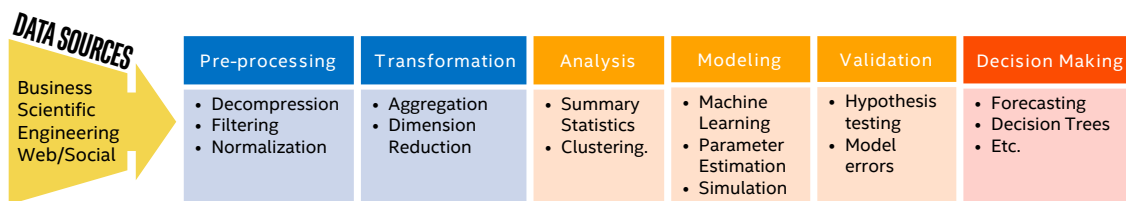

Figure 2: Intel DAAL Addresses All Stages of the Data Analytics Pipeline

Intel DAAL is developed by the same team as the Intel® Math Kernel Library (Intel® MKL), the leading math library in the world.[†] This team works closely with Intel® processor architects to squeeze maximum performance from Intel-based systems. It is available in the Intel® Parallel Studio XE suite or standalone.

Optimized for Your Hardware

For maximum calculation speed, each function is highly tuned to the instruction set, register width and memory architecture of each target processor. Intel DAAL is tuned for a broad range of Intel® processors, including Intel® Atom™, Core™, and Xeon® processors and Intel® Xeon Phi™ coprocessors, targeting Internet of things (IoT) gateways to back-end servers since applications may benefit from splitting analytics processing across several platforms such as IoT gateways and back-end servers.

Algorithms

Data Analysis: Characterization, Summarization, and Transformation

ALGORITHM	DESCRIPTION
Low Order Moments	Computes the basic dataset characteristics such as sums, means, second-order raw moments, variance, standard deviations, etc.
Quantile	Computes quantiles that summarize the distribution of data across equal-sized groups as defined by quantile orders.
Correlation and Variance-Covariance Matrices	Quantifies pairwise statistical relationship between feature vectors.
Cosine Distance Matrix	Measures pairwise similarity between feature vectors using cosine distances.
Correlation Distance Matrix	Measures pairwise similarity between feature vectors using correlation distances.
Cholesky Decomposition	Decomposes a symmetric positive-definite matrix into a product of a lower triangular matrix and its transpose. This decomposition is a basic operation used in solving linear systems, non-linear optimization, Kalman filtration, etc.
QR Decomposition	Decomposes a general matrix into a product of an orthogonal matrix and an upper triangular matrix. This decomposition is used in solving linear-inverse and least-squares problems. It is also a fundamental operation in finding eigenvalues and eigenvectors.
Singular Value Decomposition (SVD)	SVD decomposes a matrix into a product of a left singular vector, singular values, and a right singular vector. It is the basis of principal component analysis, solving linear inverse problems, and data fitting.
Principal Component Analysis (PCA)	PCA reduces the dimensionality of data by transforming input feature vectors into a new set of principal components orthogonal to each other.
K-Means	Partitions a dataset into clusters of similar data points. Each cluster is represented by a centroid, which is the mean of all data points in the cluster.
Expectation-Maximization	Finds maximum-likelihood estimate of the parameters in models. It is used for the Gaussian mixture model as a clustering method. It can also be used in non-linear dimensionality reduction, missing value problems, etc.
Outlier Detection	Identifies observations that are abnormally distant from other observations. An entire feature vector (multivariate), or a single feature value (univariate), can be considered in determining if the corresponding observation is an outlier.
Association Rules	Discovers a relationship between variables with certain level of confidence.

ALGORITHM	DESCRIPTION
Linear and Radial Basis Function Kernel Functions	Map data onto higher-dimensional space.
Quality Metrics	Computes a set of numeric values to characterize quantitative properties of the results returned by analytical algorithms. These metrics include confusion matrix, accuracy, precision, recall, Fscore, etc.

Machine Learning: Regression, Classification, and More

ALGORITHM	DESCRIPTION
Linear Regression	Models relationship between dependent variables and one or more explanatory variables by fitting linear equations to observed data.
Naïve Bayes Classifier	Splits observations into distinct classes by assigning labels. Naïve Bayes is a probabilistic classifier that assumes independence between features. Often used in text classification and medical diagnosis, it works well even when there are some level of dependence between features.
Boosting	Builds a strong classifier from an ensemble of weighted weak classifiers, by iteratively re-weighting according to the accuracy measured for the weak classifiers. A decision stump is provided as a weak classifier. Available boosting algorithms include AdaBoost (a binary classifier), BrownBoost (a binary classifier), and LogitBoost (a multi-class classifier).
Support Vector Machine (SVM)	SVM is a popular binary classifier. It computes a hyperplane that separates observed feature vectors into two classes.
Multi-Class Classifier	Builds a multi-class classifier using a binary classifier such as SVM.

One Year of Product Support and Updates Included

Product purchase provides you access to and support for new updates and releases, as well as older versions. It also entitles you to private, direct and responsive answers to product questions, along with access to decades of product experience from our user community through forums and a library of self-help documents.

Specifications at a Glance

Processors	Supports multiple generations of Intel and compatible processors including, but not limited to, Intel Core processors, Intel Xeon processors, and Intel Xeon Phi coprocessors
Languages	Compatible with compilers from Microsoft, GCC, Intel. C, C++, C#, Fortran, Java*, ASM
Operating Systems	Windows, Linux and OS X* (OS X developers can choose between the C++ or Fortran versions of the Composer Edition).
Development Environment	**Windows:** Integrates into Microsoft Visual Studio **Linux:** Compatible with GNU tools **OS X:** XCode
Additional Details	www.intel.com/software/products/systemrequirements/

For complete information, see **release notes and documentation**.

Included in Intel Parallel Studio XE

Intel Data Analytics Acceleration Library is available standalone or as part of Intel Parallel Studio XE, a comprehensive software development suite for parallel software development.

To Learn More

Intel VTune Amplifier XE

software.intel.com/en-us/intel-vtune-amplifier-xe

For a Free 30-day Evaluation

software.intel.com/en-us/intel-vtune-amplifier-xe/try-buy