INTELIGÊNCIA ARTIFICIAL NA ARQUITETURA INTEL

Igor Freitas

Intel

Legal Notices and Disclaimers

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL® PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. INTEL PRODUCTS ARE NOT INTENDED FOR USE IN MEDICAL, LIFE SAVING, OR LIFE SUSTAINING APPLICATIONS.

Intel may make changes to specifications and product descriptions at any time, without notice.

All products, dates, and figures specified are preliminary based on current expectations, and are subject to change without notice.

Intel, processors, chipsets, and desktop boards may contain design defects or errors known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Any code names featured are used internally within Intel to identify products that are in development and not yet publicly announced for release. Customers, licensees and other third parties are not authorized by Intel to use code names in advertising, promotion or marketing of any product or services and any such use of Intel's internal code names is at the sole risk of the user.

Intel product plans in this presentation do not constitute Intel plan of record product roadmaps. Please contact your Intel representative to obtain Intel's current plan of record product roadmaps.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to http://www.intel.com/performance

Intel, the Intel logo, Xeon and Xeon logo , Xeon Phi and Xeon Phi logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries

Other names and brands may be claimed as the property of others.

All products, dates, and figures are preliminary and are subject to change without any notice.

This document contains information on products in the design phase of development.

Optimization Notice

Optimization Notice

Intel® compilers, associated libraries and associated development tools may include or utilize options that optimize for instruction sets that are available in both Intel® and non-Intel microprocessors (for example SIMD instruction sets), but do not optimize equally for non-Intel microprocessors. In addition, certain compiler options for Intel compilers, including some that are not specific to Intel micro-architecture, are reserved for Intel microprocessors. For a detailed description of Intel compiler options, including the instruction sets and specific microprocessors they implicate, please refer to the “Intel® Compiler User and Reference Guides” under “Compiler Options." Many library routines that are part of Intel® compiler products are more highly optimized for Intel microprocessors than for other microprocessors. While the compilers and libraries in Intel® compiler products offer optimizations for both Intel and Intel-compatible microprocessors, depending on the options you select, your code and other factors, you likely will get extra performance on Intel microprocessors.

Intel® compilers, associated libraries and associated development tools may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include Intel® Streaming SIMD Extensions 2 (Intel® SSE2), Intel® Streaming SIMD Extensions 3 (Intel® SSE3), and Supplemental Streaming SIMD Extensions 3 (Intel® SSSE3) instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors.

While Intel believes our compilers and libraries are excellent choices to assist in obtaining the best performance on Intel® and non-Intel microprocessors, Intel recommends that you evaluate other compilers and libraries to determine which best meet your requirements. We hope to win your business by striving to offer the best performance of any compiler or library; please let us know if you find we do not.

Notice revision #20101101

Agenda

- Inteligência Artificial (IA) na Arquitetura Intel
- Big Data, IA e Computação de Alta Performance
- Casos de sucesso nacionais
- Tutoriais, treinamentos e informações sobre IA na Arquitetura Intel

INTELIGÊNCIA ARTIFICIAL NA ARQUITETURA INTEL

intel AI PORTFOLIO

SOLUTIONS

Data Scientists

Technical Services

Reference Solutions

PLATFORMS

Intel® AI DevCloud

Intel® Deep Learning System‡

intel Saffron REASONING

TOOLS

Intel® Deep Learning Studio‡

Intel® Deep Learning Deployment Toolkit†

Intel® Computer Vision SDK†

Intel® Movidius™ Software Development Kit (SDK)

FRAMEWORKS

TensorFlow* Caffe* mxnet* BigDL ON Spark* neon Caffe2*‡ PYTORCH*‡ Microsoft CNTK*‡ PaddlePaddle*‡

LIBRARIES

Intel® MKL/MKL-DNN, clDNN, DAAL, Intel Python Distribution, etc.

DIRECT OPTIMIZATION

Intel® nGraph™ Library†

CPU Transformer† | NNP Transformer‡ | Other

TECHNOLOGY

intel XEON inside | intel NERVANA inside‡ | intel ARRIA 10 inside | intel ATOM inside | intel CORE i7 inside | intel Iris Graphics | intel Movidius

END-TO-END COMPUTE

SYSTEMS & COMPONENTS

†Beta available
‡ Future
*Other names and brands may be claimed as the property of others.

INTEL AI PLATFORMS

INTEL® AI DEVCLOUD

Developer Zone

INTEL® AI ACADEMY

Tools

Intel® AI DevCloud

Now Available

Free cloud compute is now available for Intel® AI Academy members. Use Intel® AI DevCloud powered by Intel® Xeon® Scalable processors for your machine learning and deep learning training and inference compute needs.

Request Access

Benefits

- Four weeks of access
- 200 GB of file storage
- Access to a remote cluster of Intel® Xeon® Scalable processors
- Get started without making any investment

Available Frameworks and Tools

- neon™ framework
- Intel® Optimization for Theano*
- Intel® Optimization for TensorFlow*
- Intel® Optimization for Caffe*
- Intel® Distribution for Python* (including NumPy, SciPy, and

Use your existing Intel® Xeon® processor-based cluster

–OR–

Get 4-weeks access to our cluster for FREE including 200GB storage, pre-configured libraries & frameworks

INTEL® DEEP LEARNING SYSTEM*

BETA

DL Studio
+
Frameworks
+
Libraries
+
Processors

Enterprise-centric “turnkey” deep learning stack available via rackable on-premise system

INTEL® SAFFRON COGNITIVE SOLUTION

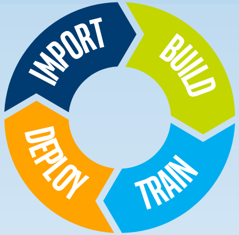
Collaborative AI decision system for fraud detection, prescriptive maintenance, churn analysis, root cause analysis & more

*Available from select OEMs in 2018

INTEL AI TOOLS

FUTURE

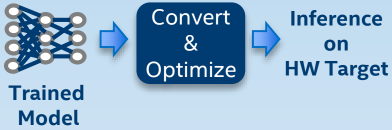
INTEL® DEEP LEARNING STUDIO



Enterprise customer tool to compress full DL development cycle; coming to the Intel® Deep Learning System

BETA

INTEL® DEEP LEARNING DEPLOYMENT TOOLKIT

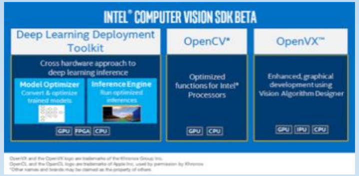


Trained Model → Convert & Optimize → Inference on HW Target

Facilitates optimized inference deployment models trained using the following frameworks: TensorFlow, Caffe, or MXNet

BETA

INTEL© COMPUTER VISION SDK



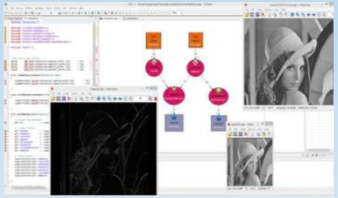
INTEL® COMPUTER VISION SDK BETA

Deep Learning Deployment Toolkit	OpenCV*	OpenVX™
Cross hardware approach to deep learning inference Model Optimizer: Convert & optimize trained models Inference Engine: Run optimized inferences GPU FPGA CPU	Optimized functions for Intel* Processors GPU CPU	Enhanced, graphical development using Vision Algorithm Designer GPU IPU CPU

OpenVX and the OpenVX logo are trademarks of the Khronos Group Inc.
OpenCL and the OpenCL logo are trademarks of Apple Inc. used by permission by Khronos
*Other names and brands may be claimed as the property of others.

DL deployment kit
-PLUS-
OpenCV* and OpenVX* support for deep learning-based computer vision on CPU, IPU, GPU & FPGA

INTEL® MOVIDIUS™ SDK



Intel® Movidius™ Vision Processing Units (VPU) software development kit for inference deployment

Other names and brands may be claimed as the property of others.

INTEL AI FRAMEWORKS

Popular DL Frameworks are now optimized for CPU!

CHOOSE YOUR FAVORITE FRAMEWORK

TensorFlow™ *

Caffe *

mxnet *

BigDL FOR APACHE Spark™ *

neon™

See installation guides at ai.intel.com/framework-optimizations/

More under optimization: Caffe2 * PYTORCH * Microsoft CNTK * PaddlePaddle * *and others to be enabled via Intel® nGraph™ Library*

SEE ALSO: Machine Learning Libraries for Python (Scikit-learn, Pandas, NumPy), R (Cart, randomForest, e1071), Distributed (MlLib on Spark, Mahout)
**Limited availability today*
Other names and brands may be claimed as the property of others.

INTEL AI LIBRARIES

DIRECT OPTIMIZATION

MKL-DNN Open-source optimized deep neural network functions for new frameworks

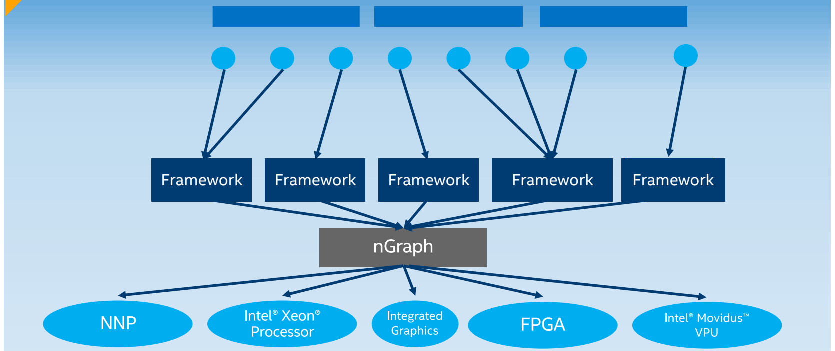
clDNN Open-source optimized deep neural network functions for Intel GPUs

DAAL Data Analytics Acceleration Library for analytics and machine learning

Intel Python Distribution Optimized distribution of most popular & fastest growing language for machine learning

BETA

INTEL® NGRAPH™ LIBRARY



Translates participating deep learning framework compute graphs into hardware-optimized executables for many different targets
(CPU, GPU, NNP, FPGA, VPU, etc.)

INTEL AI COMPUTE

GENERAL AI

intel ATOM inside / intel CORE inside / intel XEON inside

Mainstream AI

+

intel STRATIX 10 inside

Flexible Acceleration

DEEP LEARNING

TRAINING	DATA CENTER/ WORKSTATION	intel XEON inside Mainstream Training	+	intel NERVANA inside ‡ Intensive Training				
INFERENCE	DATA CENTER/ WORKSTATION	intel XEON inside Mainstream Inference	+	intel STRATIX 10 inside Real-time Inference				
	EDGE/GATEWAY	intel ATOM inside / intel CORE inside / intel XEON inside Mainstream Inference	+	intel Iris Graphics Higher Inference Throughput	intel Movidius™ Vision 1-20W	Intel GNA (IP) Speech/Audio 1-100+mW	MOBILEYE An Intel Company Autonomous driving	intel STRATIX 10 inside Custom Inference

‡ Future

All products, computer systems, dates, and figures are preliminary based on current expectations, and are subject to change without notice.

INTEL® XEON® PROCESSOR PLATFORM PERFORMANCE

Hardware plus optimized software

INFERENCE THROUGHPUT

Up to

198x

Intel® Xeon® Platinum 8180 Processor
higher Intel optimized Caffe GoogleNet v1 with Intel® MKL
inference throughput compared to
Intel® Xeon® Processor E5-2699 v3 with BVLC-Caffe

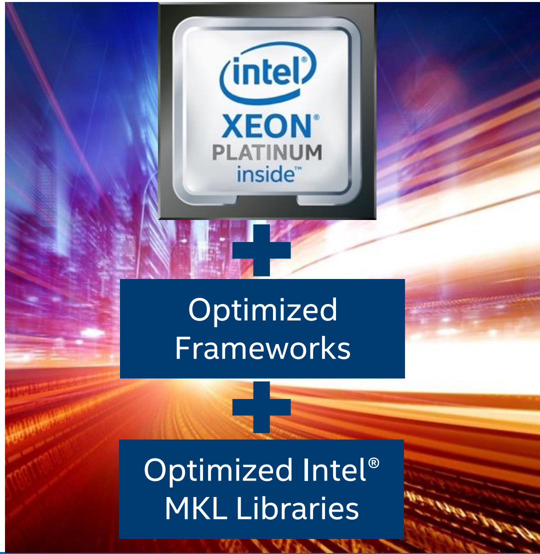
TRAINING THROUGHPUT

Up to

127x

Intel® Xeon® Platinum 8180 Processor
higher Intel Optimized Caffe AlexNet with Intel® MKL
training throughput compared to
Intel® Xeon® Processor E5-2699 v3 with BVLC-Caffe

Inference and training throughput uses FP32 instructions



Deliver significant AI performance with hardware and software optimizations on Intel® Xeon® Scalable Family

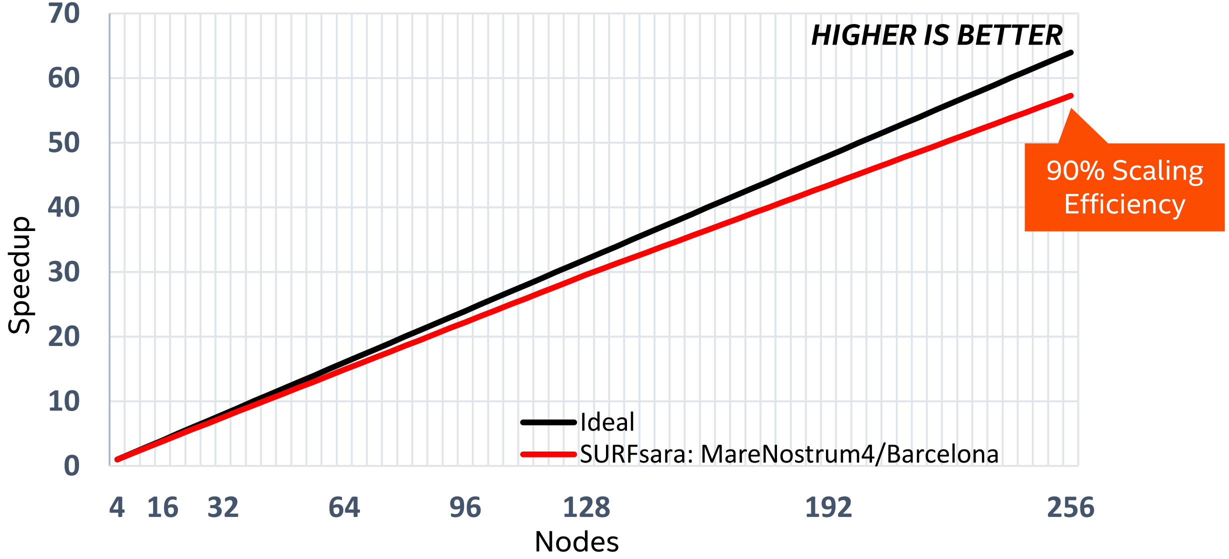
Up to 191X Intel® Xeon® Platinum 8180 Processor higher Intel optimized Caffe Resnet50 with Intel® MKL inference throughput compared to Intel® Xeon® Processor E5-2699 v3 with BVLC-Caffe
Up to 93X Intel® Xeon® Platinum 8180 Processor higher Intel optimized Caffe Resnet50 with Intel® MKL training throughput compared to Intel® Xeon® Processor E5-2699 v3 with BVLC-Caffe

Performance estimates were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown." Implementation of these updates may make these results inapplicable to your device or system.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit: http://www.intel.com/performance Source: Intel measured as of June 2017. Configurations: See the last slide in this presentation. *Other names and brands may be claimed as the property of others.

FAST & EFFICIENT DL SCALING ON CPU

Intel® - SURFsara* Research Collaboration - Multi-Node Intel® Caffe ResNet-50 Scaling Efficiency on 2S Intel® Xeon® Platinum 8160 Processor Cluster



- MareNostrum4 Barcelona Supercomputing Center
- ImageNet-1K
- 256 nodes
- 90% scaling efficiency
- Top-1/Top-5 > 74%/92%
- Batch size of 32 per node
- Global BS=8192
- Throughput: 15170 Images/sec

Time-To-Train: 70 minutes (50 Epochs)

90% scaling efficiency with up to 74% Top-1 accuracy on 256 nodes

Configuration Details: see end of presentation

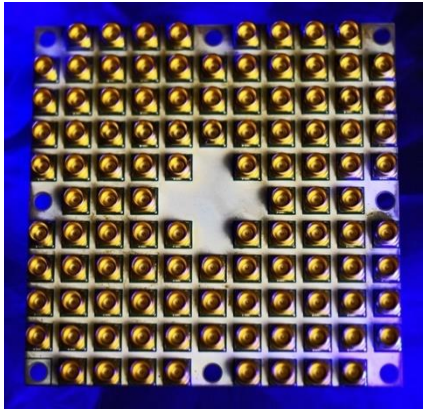
Performance estimates were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown." Implementation of these updates may make these results inapplicable to your device or system. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit: http://www.intel.com/performance Source: Intel measured as of June 2017

Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

LEADING AI RESEARCH

Choose a partner on the cutting-edge of AI breakthroughs

Neuromorphic Computing Test Chip Codenamed “Loihi”



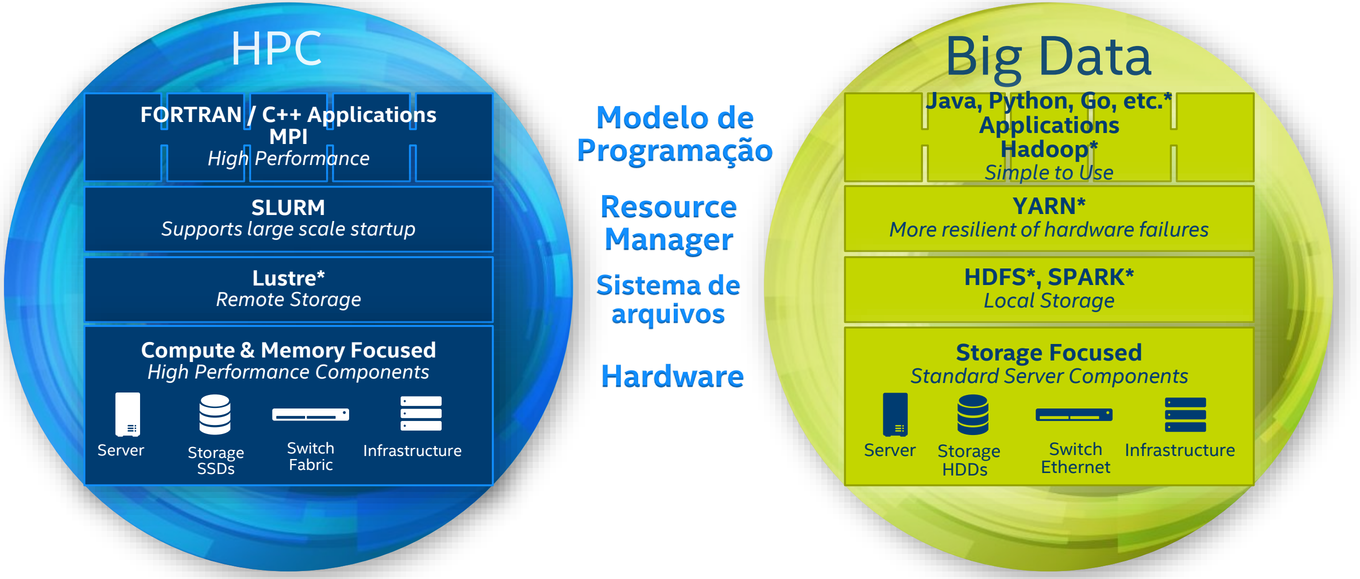
Quantum Computing 49-Qubit Test Chip Codenamed “Tangle-Lake”

NEW AI Technologies @Intel Labs

BIG DATA, INTELIGÊNCIA ARTIFICIAL E COMPUTAÇÃO DE ALTO DESEMPENHO

Big Data Analytics

HPC != Big Data Analytics != Inteligência Artificial ?



HPC

FORTRAN / C++ Applications
MPI
High Performance

SLURM
Supports large scale startup

Lustre*
Remote Storage

Compute & Memory Focused
High Performance Components

Modelo de Programação

Resource Manager

Sistema de arquivos

Hardware

Big Data

Java, Python, Go, etc.*
Applications
Hadoop*
Simple to Use

YARN*
More resilient of hardware failures

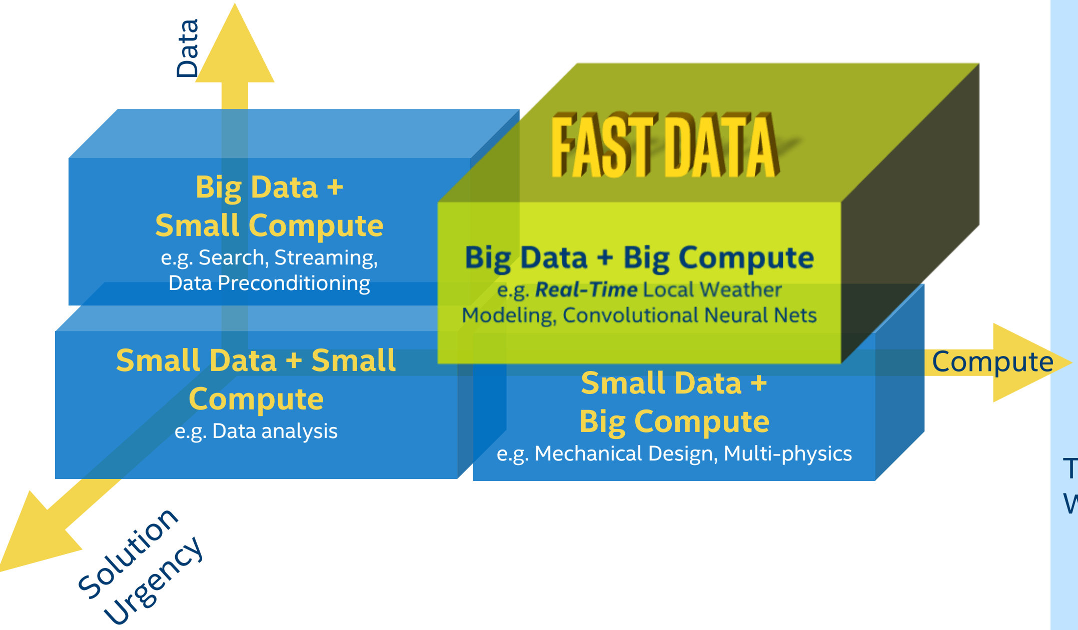
HDFS*, SPARK*
Local Storage

Storage Focused
Standard Server Components

*Other brands and names are the property of their respective owners.

Big Data Analytics

HPC em tempo real



Data

FAST DATA

Big Data + Small Compute
e.g. Search, Streaming, Data Preconditioning

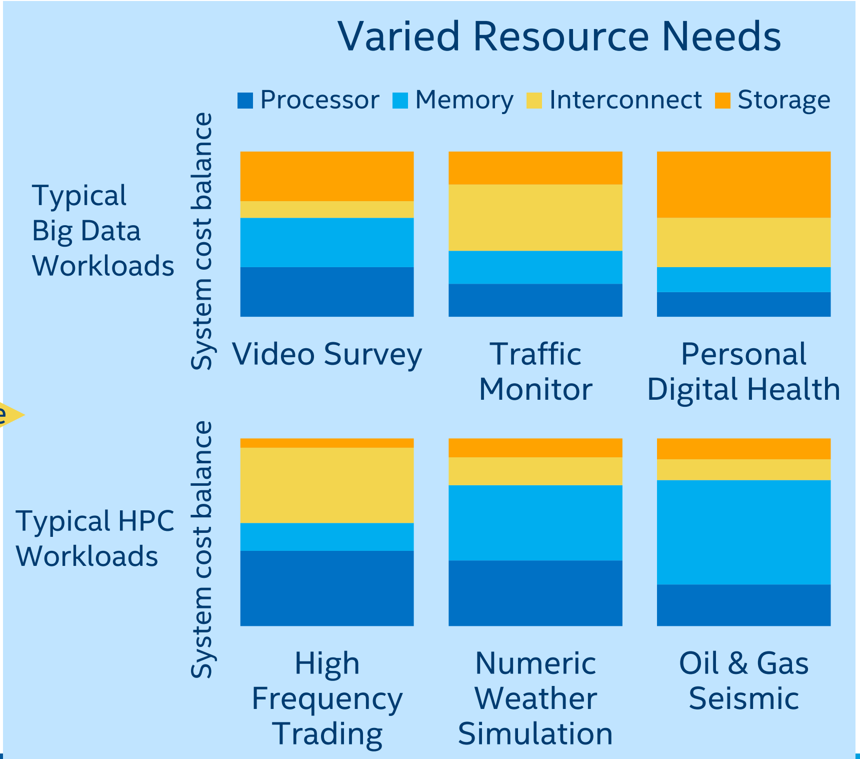
Big Data + Big Compute
e.g. ***Real-Time*** Local Weather Modeling, Convolutional Neural Nets

Small Data + Small Compute
e.g. Data analysis

Small Data + Big Compute
e.g. Mechanical Design, Multi-physics

Compute

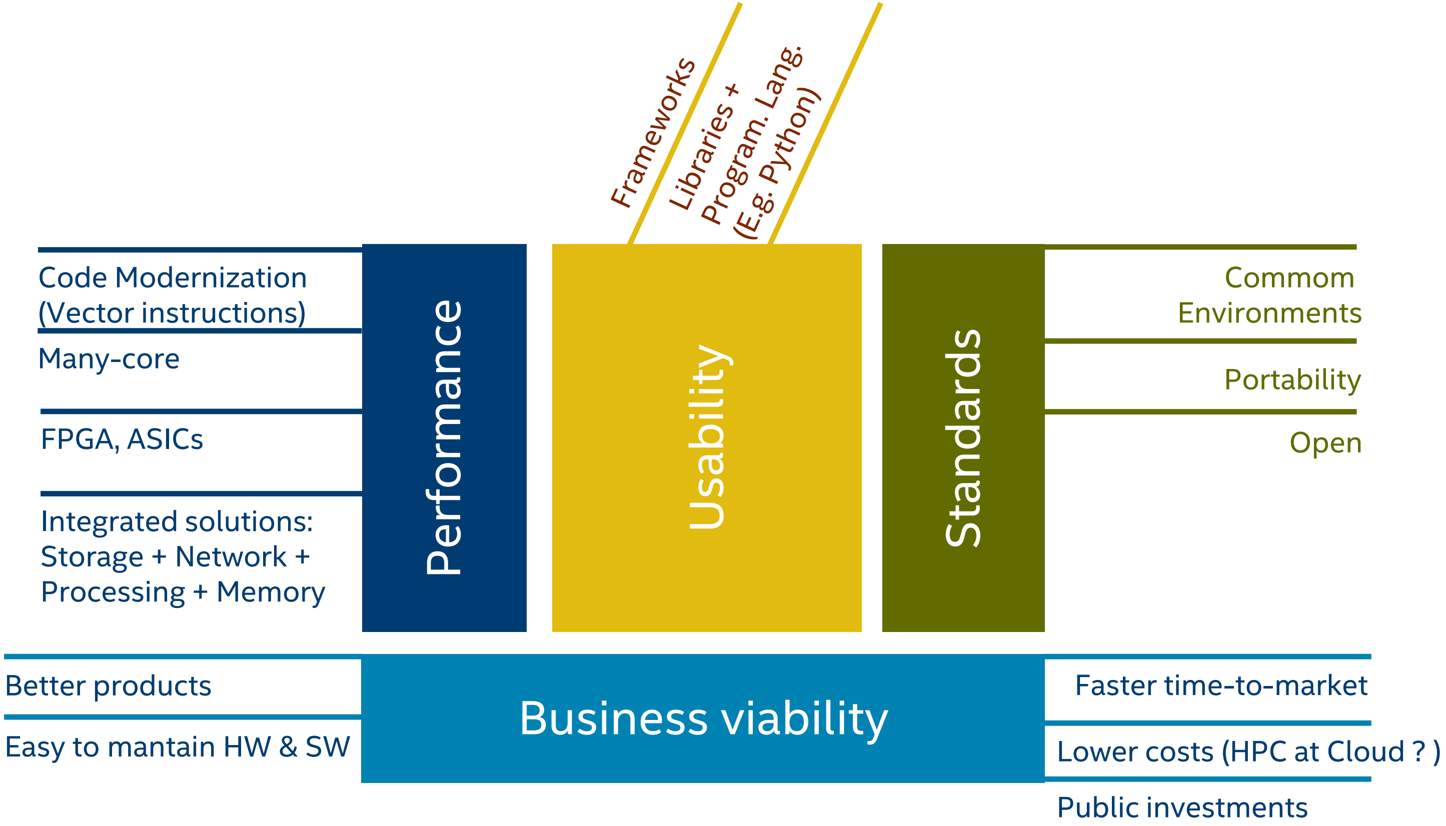
Solution Urgency



Varied Resource Needs

https://www.intelnervana.com/framework-optimizations/

Trends in HPC + Big Data Analytics



Frameworks

Libraries + Program. Lang. (E.g. Python)

Code Modernization (Vector instructions)

Many-core

FPGA, ASICs

Integrated solutions: Storage + Network + Processing + Memory

Performance

Usability

Standards

Commom Environments

Portability

Open

Better products

Easy to mantain HW & SW

Business viability

Faster time-to-market

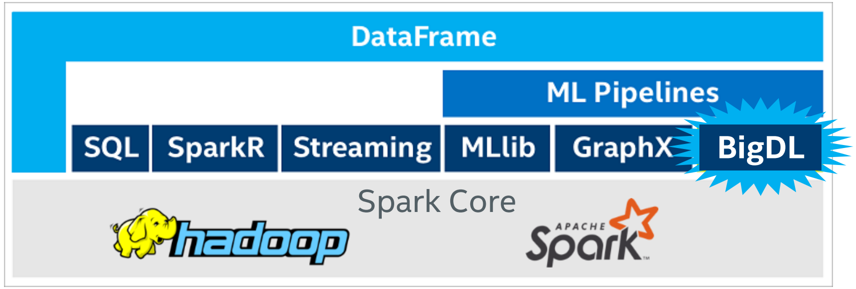
Lower costs (HPC at Cloud ?)

Public investments

https://www.intelnervana.com/framework-optimizations/

BigDL

High Performance Deep Learning for **FREE** on CPU Infrastructure[1]



BigDL is a distributed deep learning library for Apache Spark* that can run directly on top of existing Spark or Apache Hadoop* clusters with direct access to stored data and tool/workflow consistency!

No need to deploy costly accelerators, duplicate data, or suffer through scaling headaches!

Feature Parity with Caffe* and Torch*

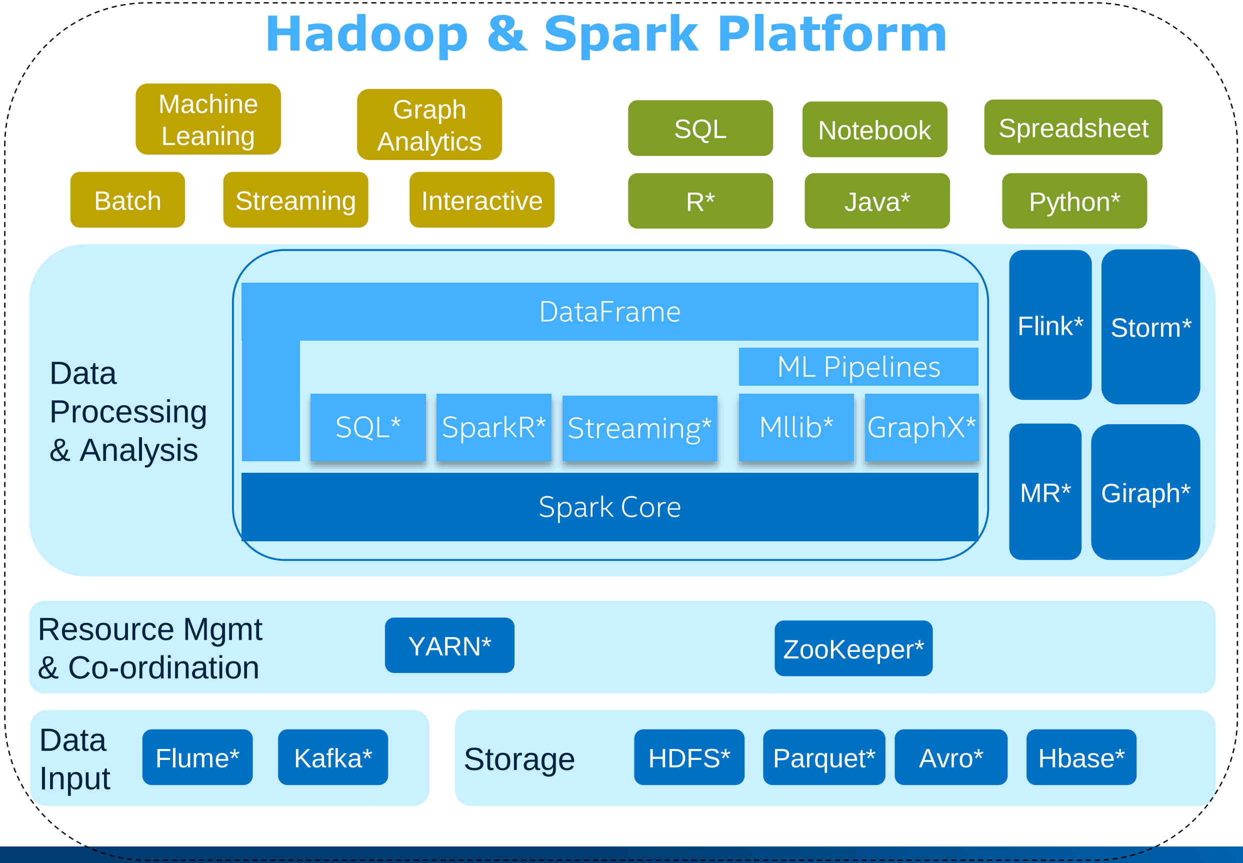
Lower TCO and improved ease of use with existing infrastructure

Deep Learning on Big Data Platform, Enabling **Efficient Scale-Out**

software.intel.com/bigdl

[1] Open-source software is available for download at no cost; 'free' is also contingent upon running on existing idle CPU infrastructure where the operating cost is treated as a 'sunk' cost

Unified Big Data Analytics Platform



How to Run Deep Leaning Workloads Directly on Big Data Platform?

- Integrated with Big Data ecosystem
- Massively distributed, shared-nothing
- Scale-out
- Send compute to data
- Fault tolerance
- Elasticity
- Incremental scaling
- Dynamic resource sharing
- …

intel
experience what's inside™

CASOS DE SUCESSO NACIONAIS EM INTELIGÊNCIA ARTIFICIAL

Parcerias entre Intel, Academia e Indústria no Brasil

Centros de Excelência Intel em Inteligência Artificial c/ Startups

Apoio de P&D no desenvolvimento de soluções inovadoras em IA

Centros de Excelência em Inteligência Artificial em parceria com a Intel

AI2BIZ

AXONDATA

- Transferência de conhecimento em IA e HPC (High Performance Computing)
- Apoio no desenvolvimento de protótipos utilizando software livre
- Workshop em IA e HPC
- Acesso a servidores Intel de alto desempenho

Entendimento do problema

- Definição do problema
- Apresentação de casos de sucesso
- Como a Intel pode ajudar

Workshop

- Treinamento técnico em:
  - IA na prática
  - Melhoria de performance
  - Hands-on

Prova de Conceito / MVP

- Baseado em código livre
- 30 à 90 dias
- Repositório de código e docs
- Protótipo otimizado p/ produção

Fase Piloto

- Suporte p/ testes
- Otimização de performance

Fase da solução

- Suporte p/ deploy
- Otimização de performance
- Casos de sucesso
- Press release
- Eventos

Centros de Excelência em Inteligência Artificial - Intel

Casos de sucesso

SERPRO
Serviço Federal de Processamento de Dados

Thiago Oliveira, superintendente de Engenharia de Infraestrutura do SERPRO

“Validador Cognitivo de Infrações de Trânsito”

✓ Performance 22.5x mais rápida em “Xeon Scalable Processors”

“...um processamento de multas que antes levava 45 horas agora poderá ser realizado em menos de 2 horas.”

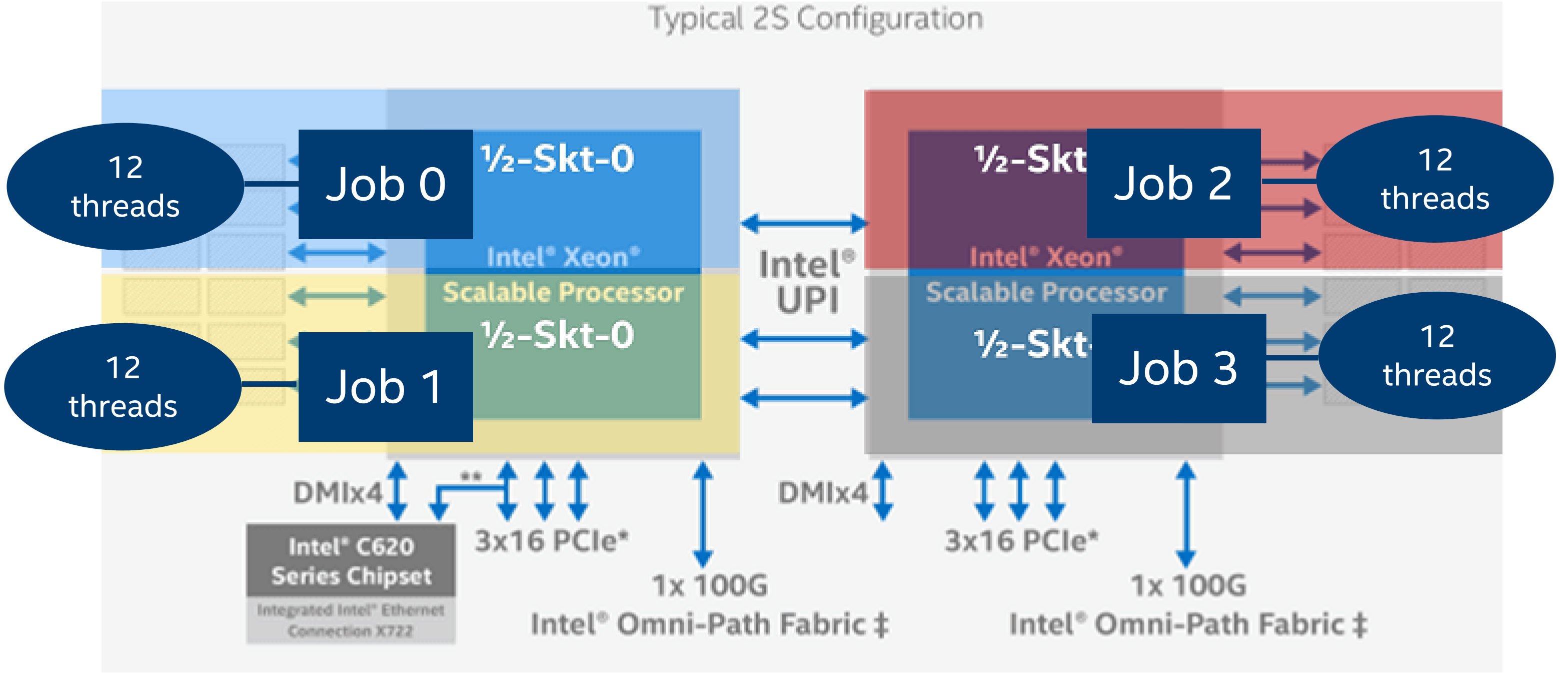
✓ Desenvolvimento do modelo matemático

“Com isso, tivemos uma acurácia de 90% no sistema, além da automação de todo o projeto”,
disse Gustavo Rocha, chefe de divisão do SERPRO,“

Intel Python Distribution + Caffe / TensorFlow otimizados + MKL + Técnicas de HPC

Centros de Excelência em Inteligência Artificial - Intel

Técnicas de HPC aplicadas para IA



`libnumactl`   `kmp_affinity`

https://software.intel.com/en-us/articles/boosting-deep-learning-training-inference-performance-on-xeon-and-xeon-phi

Centros de Excelência em Inteligência Artificial - Intel

Provas de Conceito em andamento

Automação da Análise de Processos Jurídicos

- Classificação e agrupamento automático de processos
- Identificação de processos válidos, seguindo critérios do cliente (falta de carimbo, data errada, nome das partes, etc.)
- OCR de documentos mal digitalizados

Otimização de Deep Learning em Ambiente de Produção
Área Financeira – Banco Público

- Melhoria de performance e acurácia
- Uso da BigDL em cluster Hadoop + Spark
- Do Protótipo ao Produto

Predição de falhas em Ponto de Vendas (POS)
Área Financeira - Setor de Logística e Pagamentos

- Diminuir custo de manutenção
- Análise histórica das falhas
- Entendimento da eficiência das máquinas
- Eventos externos e internos
- Predição de quando ocorrerão novas falhas
- Análise de 100k terminais

Predição de falhas em Caixas Eletrônicos (ATM)
Área Financeira – Setor de TI

- Análise histórica das falhas
- Predição de quando ocorrerão novas falhas

Um pouco do histórico da Intel em HPC no Brasil

Casos de Sucesso

Oil & Gas - Reservoir Simulator at PETROBRAS

- Up to 10.5x performance gains in their Reservoir Simulator software[1]

LNCC - National Laboratory for Scientific Computing
Largest HPC cluster in Latin America

- Up to 30x performance gain in Oil & Gas applications[2]
- Up to 3.4x speedup via AVX (vector instructions)
- Link white-paper

Health & Life Sciences

- Up to 11x speedup in Molecular Dynamics – NCC/UNESP & LNCC – white-paper link
- Xeon only:
  - Original code vs Modernized code: up to 11x speedup
- Xeon + 1 Xeon Phi (same optimized code)
  - 1.14x speedup

Article link

INPE/CPTEC
Code Modernization of BRAMS

- Initial results – white-paper link

Authors:
[1] CENPES team and Gilvan Vieira - gilvandsv@gmail.com
[2] LNCC - Frederico Cabral - fredluiscabral@gmail.com
[3] NCC/UNESP - Silvio Stanzani silvio.stanzani@gmail.com

TUTORIAIS, TREINAMENTOS E INFORMAÇÕES SOBRE IA NA ARQUITETURA INTEL

INTEL® AI ACADEMY

For developers, students, instructors and startups

LEARN

- Online tutorials
- Webinars
- Student kits
- Support forums

DEVELOP

- Intel Optimized Frameworks
- Exclusive access to Intel® AI DevCloud

TEACH

- Comprehensive courseware
- Hands-on labs
- Cloud compute
- Technical Support

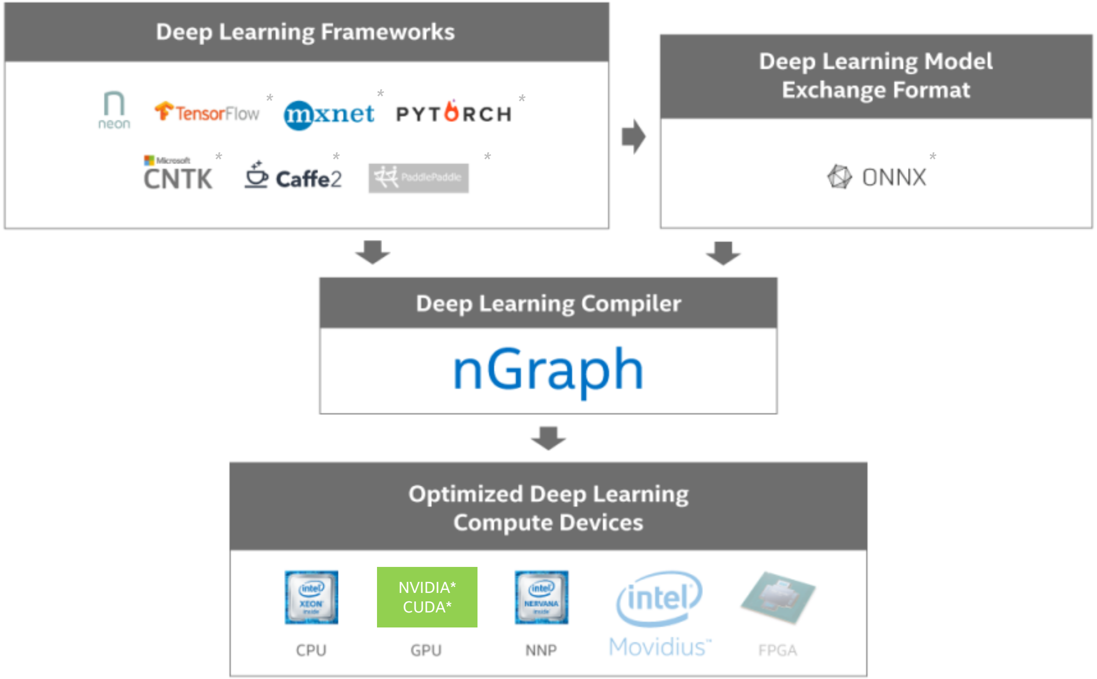
SHARE

- Project showcase opportunities at
- Intel Developer Mesh
- Industry & Academic events

software.intel.com/ai

INTEL® NGRAPH™ COMPILER

Optimized Compute Devices for Neural Networks



*Other names and brands may be claimed as the property of others.
All products, computer systems, dates, and figures are preliminary based on current expectations, and are subject to change without notice.

INTEL DISTRIBUTION FOR PYTHON

Advancing Python Performance Closer to Native Speeds

For developers using the most popular and fastest growing programming language for AI

Easy, Out-of-the-box Access to High Performance Python

- Prebuilt, optimized for numerical computing, data analytics, HPC
- Drop in replacement for your existing Python (no code changes required)

Drive Performance with Multiple Optimization Techniques

- Accelerated NumPy/SciPy/Scikit-Learn with Intel® MKL
- Data analytics with pyDAAL, enhanced thread scheduling with TBB, Jupyter* Notebook interface, Numba, Cython
- Scale easily with optimized MPI4Py and Jupyter notebooks

Faster Access to Latest Optimizations for Intel Architecture

- Distribution and individual optimized packages available through conda and Anaconda Cloud
- Optimizations upstreamed back to main Python trunk

software.intel.com/intel-distribution-for-python

INTEL® MKL-DNN

Math Kernel Library for Deep Neural Networks

For developers of deep learning frameworks featuring optimized performance on Intel hardware

Distribution Details

- Open Source
- Apache 2.0 License
- Common DNN APIs across all Intel hardware.
- Rapid release cycles, iterated with the DL community, to best support industry framework integration.
- Highly vectorized & threaded for maximal performance, based on the popular Intel® MKL library.

github.com/01org/mkl-dnn

Examples:

Direct 2D Convolution

Local response normalization (LRN)

Rectified linear unit neuron activation (ReLU)

Maximum pooling

Inner product

INTEL® DEEP LEARNING DEPLOYMENT TOOLKIT

BETA Now Available!

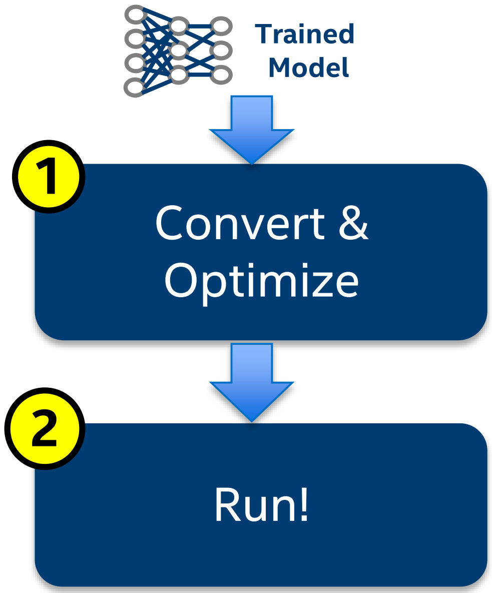
For developers looking to run deep learning models on the edge

1

- Imports trained models from popular DL framework regardless of training HW
- Enhances model for improved execution, storage & transmission

2

- Optimizes Inference execution for target hardware (computational graph analysis, scheduling, model compression, quantization)
- Enables seamless integration with application logic
- Delivers embedded friendly Inference solution

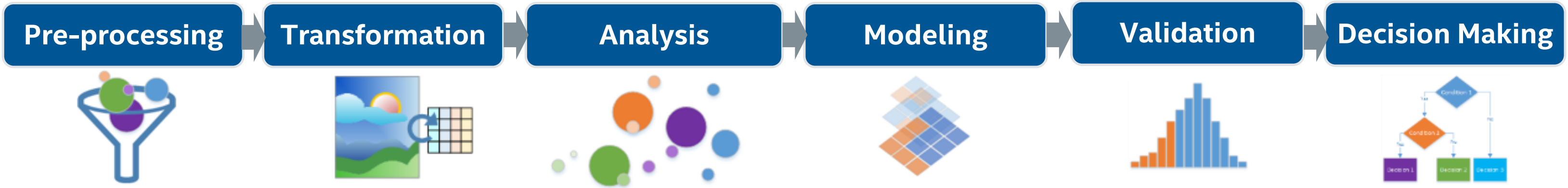


Ease of use + Embedded friendly + Extra performance boost

INTEL® DATA ANALYTICS ACCELERATION LIBRARY (INTEL® DAAL)

High Performance ML and Data Analytics library

Building blocks for all data analytics stages, including data preparation, data mining & machine learning



Open Source • Apache 2.0 License

Common Python, Java and C++ APIs across all Intel hardware

Optimized for large data sets including streaming and distributed processing

Flexible interfaces to leading big data platforms including Spark and range of data formats (CSV, SQL, etc.)

FIND OUT MORE

LEARN

More information at ai.intel.com

EXPLORE

Use Intel's performance-optimized libraries & frameworks

ENGAGE

Contact your Intel representative for help and POC opportunities