



**AI ON
INTEL**

**AI IN THE ENTERPRISE:
THE INTEL® AI BUILDERS SHOWCASE**

**BRIGITTE ALEXANDER, MANAGING DIRECTOR AI PARTNER PROGRAMS
INTEL AI PRODUCTS GROUP
SEPTEMBER 10, 2019**

TODAY'S LINE UP

- Intel® AI Builders Program
- Why Intel AI?
- Intel AI Builders Partner Showcase
- How to Deploy on Intel Architecture
- After party!
- Match-making w/ Partners



INTEL AI BUILDERS PROGRAM

INTEL® AI BUILDERS IS A THRIVING ENTERPRISE ECOSYSTEM

250+ Enterprise Partners with 100+ Optimized AI Solutions on Intel Technologies

Vertical Partners

Retail	Healthcare	Finance & Insurance	Transportation	Art & Entertainment	Government	Software	Prof. Services	Agriculture	Communications

Horizontal Partners

OEM					SI				
<div><div> CISCO</div><div> COLFAX Controlled Solutions</div><div> DELL EMC</div><div> hp</div><div> inspur</div><div> Hewlett Packard Enterprise</div><div> SUPERMICRO</div><div> Lenovo</div><div> IBM</div></div>					<div><div> HCL</div><div> Tech Mahindra</div><div> QUEST Solve to Enrich</div><div> accenture High performance. Delivered.</div><div> L&T Technology Services</div><div> ALTOROS</div><div> wipro</div><div> valtech.</div><div> Computacenter</div><div> TATA CONSULTANCY SERVICES</div></div>				
Data Analytics	Video Surveillance & Analytics	Image/Object Recognition	Conv. Bots and Voice Agents	Data Prep & Mgmt	Anomaly Detection	Facial Detection/Recognition	Robotic Process Automation		
<div><div> skymind</div><div> C3.ai</div><div> DataRobot</div><div> bluedata</div><div> cloudera</div><div> PROPHESTOR</div><div> sas</div><div> H2O</div></div>	<div><div> DEEPCLINT 精英深瞳</div><div> vedax</div><div> anyVISION.</div><div> 慧眼达 Sight 1 A</div></div>	<div><div> Matroid</div><div> altaia SYSTEMS</div><div> wrnch</div><div> gesteo</div><div> viso.ai</div><div> CrowdAI</div><div> allegro</div><div> DEEViA Deep Vision ANALYSIS</div><div> InstaDeep</div><div> clarifai</div></div>	<div><div> avamo</div><div> AVAYA</div><div> gamalon</div><div> VERINT</div><div> iJ verbio</div><div> aivo</div><div> [24]7.ai</div><div> gnani.ai</div><div> Jacada</div><div> Clova</div><div> SET SAIL</div><div> voicell</div><div> 云知声 CloudVoice</div></div>	<div><div> CTACCEL</div><div> MoBagel</div><div> Petuum</div><div> iMerit</div><div> FORMCEPT</div><div> DOMINO</div><div> ALEGION</div><div> Paperspace</div><div> clusterone</div><div> JQUANT</div><div> CORE SCIENTIFIC</div><div> NIMBAX</div><div> sparkcognition</div></div>	<div><div> QuikFynd</div><div> TIVIT</div><div> axondata</div><div> DATAValue</div><div> iMerit</div><div> ALEGION</div><div> minds.ai</div><div> deepsense.ai</div></div>	<div><div> KIBERNETIKA AI</div><div> ENTROPIK TECH</div><div> JETWARE</div><div> VISIONGENII Being Different</div><div> LEAPMIND</div><div> znv 力维</div></div>	<div><div> blueprism</div><div> UiPath</div></div>		

INTEL® AI BUILDERS BENEFITS FOR PARTNERS

Partner Activation

Available to all partners



TECH ENABLEMENT

- Solution definition
- Technical support
- Intel® AI Dev Cloud for Builders
- Neural Compute Stick 2
- Learning resources

Partners demonstrating solutions on Intel AI



CO-MARKETING

- Event demos & speakerships
- Intel AI Builders Document Library
- Digital content/ social channels

Partners optimized on Intel AI



MATCH-MAKING

- Match-making of optimized partners with Intel enterprise customers

Select start ups



INVESTMENT

- Intel Capital consideration by our dedicated AI investment team

INTEL® AI BUILDERS CO-MARKETING

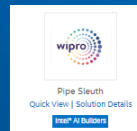
Events demos,
sessions, & keynote
walk-ons



Global PR stories
including TV
coverage



Blogs, white papers & solution briefs
highlighting AI optimizations on Intel



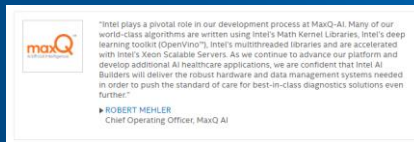
Weekly podcasts



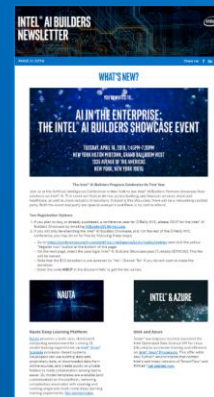
Partner videos



Partner advocates for Intel® AI and
the Intel® AI Builders Program



Intel® AI Builders
newsletters



Social amplification



Partner microsites



Analyst days,
press briefings and
Intel Capital
summits

INTEL® AI BUILDERS BENEFITS FOR ENTERPRISE END USERS

Road Map
Alignment

Engagement and Support from
Planning to Deployment

Faster and Safer
Deployments

CONNECT

with preferred solution, providers, get access to state of the art solutions and benefit from best practices



TEST

your roadmap of interoperable, scalable, and flexible solutions by utilizing the resources available only to Intel® AI Builders members



DEPLOY

optimized, tested, and reliable solutions, and continue to drive the transformation of your system

SOLUTIONS CATALOG

Find optimized and market-ready partner solutions to support your AI deployment needs.

REFINE RESULTS

Reset

Search



Ecosystem Partners

All

Intel® AI Builders

Solution Geographic Availability

Compute

Deployment Channel

Framework Optimizations

Industry

Model Training

Offering Type

Operating System

Software Libraries

Toolkits

Topology

Use Case

Application Type



Retail Shelvesight Solution
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



AI-driven Crowd Counting Solution
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



Healthcare AI Platform
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



Skymind Intelligence Layer (SKIL)
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



Cloudera Data Platform (CDP)
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



Vision Processing Accelerator for
DL Inference
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



Intelligent Computing
Orchestration Platform
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



Stratifyd Platform
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



Dell Ready Solution for AI Deep
Learning w Intel
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



FPGA Inference Acceleration Card
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



OneClick.ai
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



SubtlePET
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



ignio AIOps™
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



AI based container usage
optimization tool
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



QuEST ThirdEye - Vision Analytics
Platform
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



Wipro Pipe Sleuth
[Quick View](#) | [Solution Details](#)

Intel® AI Builders



WE INVITE YOU TO LEARN MORE ABOUT INTEL AI BUILDERS...

- Go to builders.intel.com/ai
- See more Intel® AI Builders in the Intel booth tomorrow & Thurs:

ACCELERATE SUSPECTED INTRACRANIAL HEMORRHAGE
DETECTION WITH AI



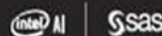
Intel® AI Builders Member

AUTOMATED ML: SOLVING CRITICAL BUSINESS PROBLEMS IN LESS TIME



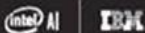
Intel® AI Builders Member

PREDICTIVE MAINTENANCE USING STREAMING ANALYTICS



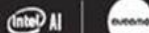
Intel® AI Builders Member

END-TO-END DATA ANALYTICS: COLLECT, ORGANIZE, ANALYZE, & INFUSE



Intel® AI Builders Member

VIRTUAL ASSISTANTS IN HEALTHCARE: HOW HUMAN CAN THEY BE?



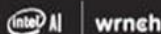
Intel® AI Builders Member

INTELLIGENT DATA EXTRACTION FROM ENGINEERING DOCUMENTS



Intel® AI Builders Member

AI-DRIVEN PERSONALIZED TRAINING & FITNESS



Intel® AI Builders Member

WHY INTEL AI?

PARTNER WITH INTEL® TO ACCELERATE YOUR AI JOURNEY

Tame your data deluge
with our data layer expertise



Speed up development
with open AI software



Choose any approach
from analytics to deep learning



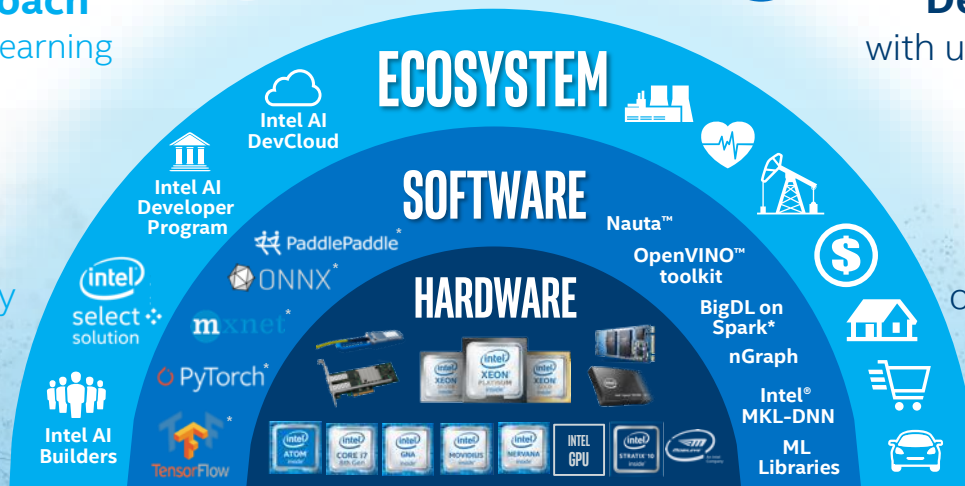
Deploy AI anywhere
with unprecedented HW choice



Simplify AI
via our robust community



Scale with confidence
on the platform for IT & cloud



www.intel.ai

HARDWARE

Multi-purpose to purpose-built
AI compute from cloud to device



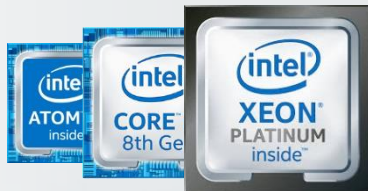
MAINSTREAM

INTENSIVE

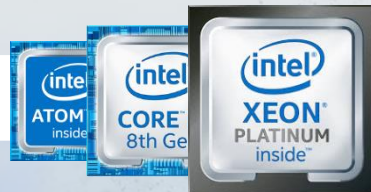
DEEP
LEARNING

TRAINING

INFERENCE



MOST
OTHER AI



ONE SIZE DOES NOT FIT ALL

INTEL® XEON® PROCESSORS

Now Optimized For Deep Learning

INFERENCE THROUGHPUT



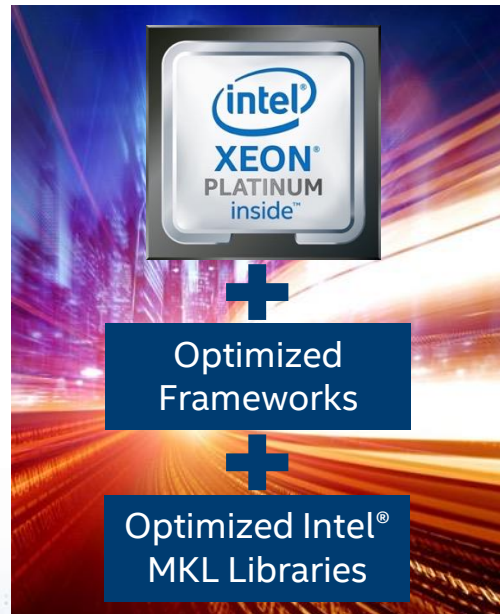
Intel® Xeon® Platinum 8180 Processor
higher Intel optimized Caffe GoogleNet v1 with Intel® MKL
inference throughput compared to
Intel® Xeon® Processor E5-2699 v3 with BVLC-Caffe

TRAINING THROUGHPUT



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



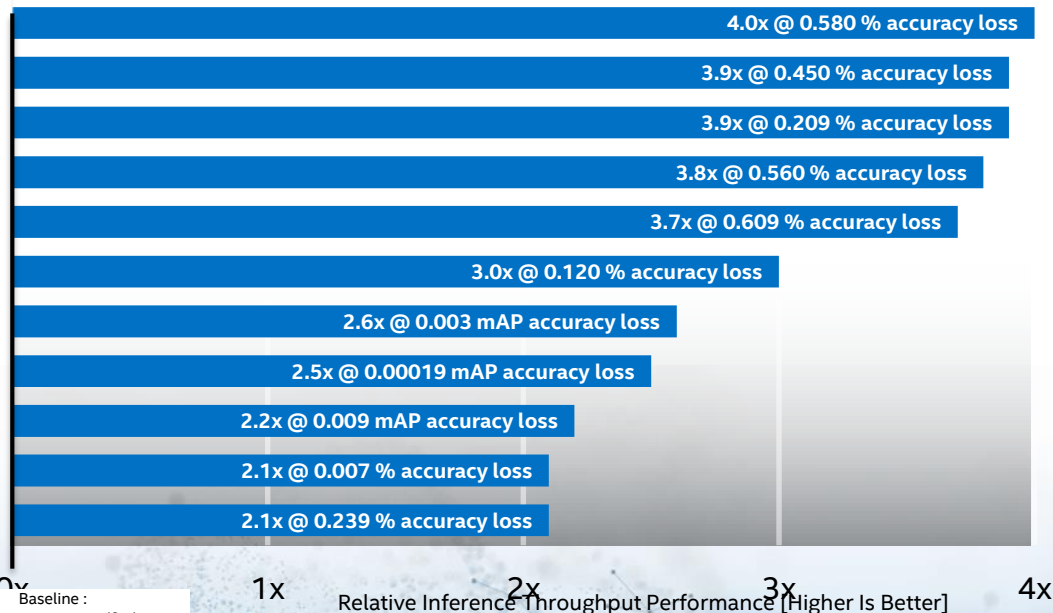
Deliver significant AI performance with hardware and software optimizations on Intel® Xeon® Scalable Family

¹ Performance results are based on testing as of June 2018 and may not reflect all publicly available security updates. Configurations: See slide 21&22. For more complete information about performance and benchmark results, visit www.intel.com/benchmarks.

ENABLING DEEP LEARNING USE CASES WITH INTEL® DEEP LEARNING BOOST

2S Intel® Xeon® Platinum 8280 Processor vs 2S Intel® Xeon® Platinum 8180 Processor

INT8 w/ INTEL® DL BOOST VS FP32 PERFORMANCE GAINS



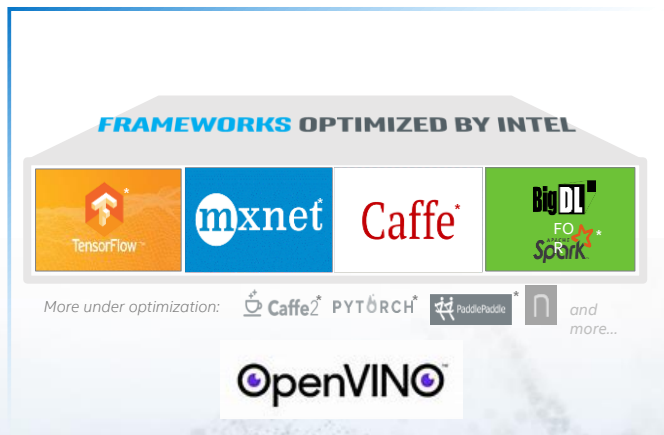
TensorFlow	ResNet-101	Image Recognition
TensorFlow	ResNet-50	
OpenVINO	ResNet-50	
mxnet	ResNet-101	
PyTorch	ResNet-50	
mxnet	ResNet-50	
PyTorch	RetinaNet	Object Detection
mxnet	SSD-VGG16	
Caffe	SSD-MobileNet	
TensorFlow	Wide and Deep	Rec. Systems
mxnet	Wide and Deep	

SIGNIFICANT PERFORMANCE GAINS USING INTEL® DL BOOST ACROSS POPULAR FRAMEWORKS AND DIFFERENT CUSTOMER USECASES

Performance results are based on testing as of 3/26/2019 and may not reflect all publicly available security updates. Configurations: See slide 21&22. For more complete information about performance and benchmark results, visit www.intel.com/benchmarks. 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.

SOFTWARE: THE BEST “OUT-OF-THE-BOX” AI EXPERIENCE

SOFTWARE FRAMEWORKS



CONTAINERS



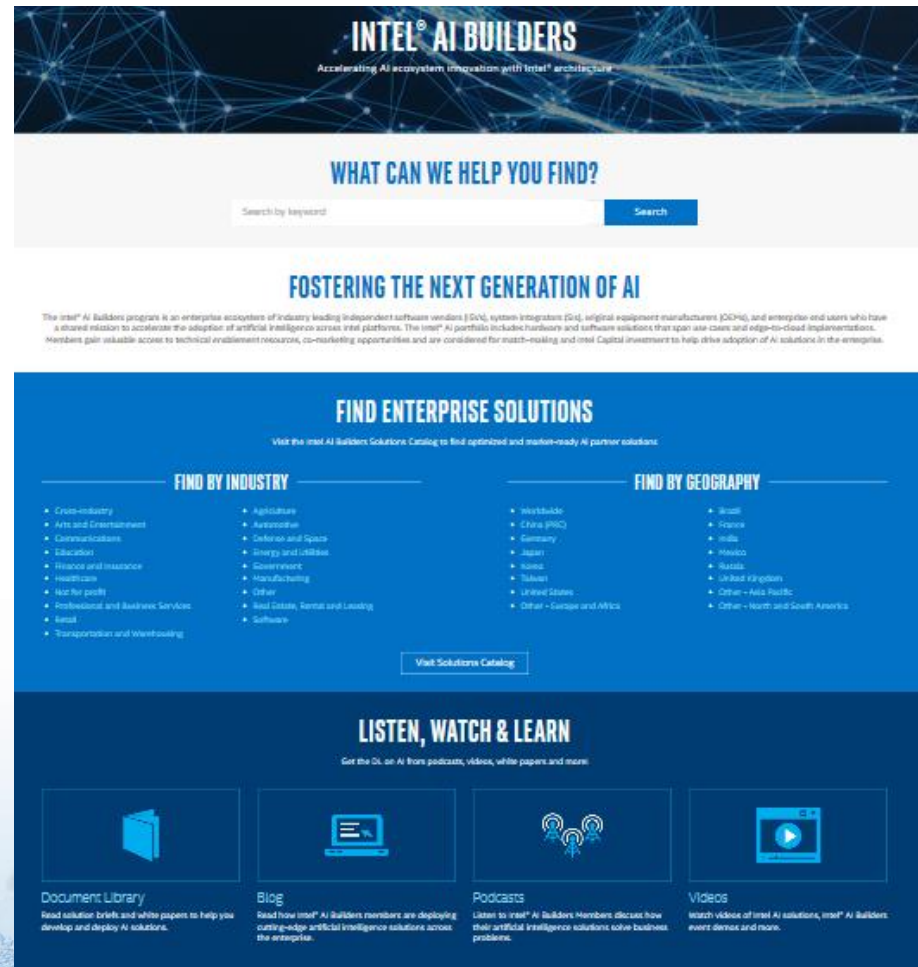
CLOUD ENVIRONMENTS



1. Frameworks: All Intel optimizations up-streamed to <https://github.com/tensorflow/tensorflow>
2. Anaconda: `conda install -c anaconda tensorflow`; `conda install -c intel tensorflow`
3. Docker: `$ docker pull intelai/g/intel-optimized-tensorflow:1.10.0`
4. Amazon: <https://www.intel.ai/amazon-web-services-works-with-intel-to-enable-optimized-deep-learning-frameworks-on-amazon-ec2-cpu-instances>
5. GCP:
6. Microsoft Azure:

ECOSYSTEM

- Global
- Cross-industry
- All AI use cases
- Intel optimized
- Market-ready solutions



INTEL® AI BUILDERS
Accelerating AI ecosystem innovation with Intel® architecture

WHAT CAN WE HELP YOU FIND?

Search by keyword

FOSTERING THE NEXT GENERATION OF AI

The Intel® AI Builders program is an enterprise ecosystem of industry leading independent software vendors (ISVs), system integrators (SIs), original equipment manufacturers (OEMs), and emerging and users who have a shared mission to accelerate the adoption of artificial intelligence across Intel platforms. The Intel® AI portfolio includes hardware and software solutions that span use cases and edge-to-cloud implementations. Members gain valuable access to technical enablement resources, co-marketing opportunities and are considered for match-making and Intel Capital investments to help drive adoption of AI solutions in the enterprise.

FIND ENTERPRISE SOLUTIONS

Visit the Intel AI Builders Solutions Catalog to find optimized and market-ready AI partner solutions.

FIND BY INDUSTRY

- Cross-industry
- Arts and Entertainment
- Communications
- Education
- Finance and Insurance
- Healthcare
- Not for profit
- Professional and Business Services
- Retail
- Transportation and Warehousing
- Agriculture
- Automotive
- Defense and Space
- Energy and Utilities
- Government
- Manufacturing
- Other
- Real Estate, Rental and Leasing
- Software

FIND BY GEOGRAPHY

- Worldwide
- China (PRC)
- Germany
- Japan
- Korea
- Taiwan
- United States
- Other - Europe and Africa
- Brazil
- France
- India
- Mexico
- Russia
- United Kingdom
- Other - Asia Pacific
- Other - North and South America

LISTEN, WATCH & LEARN

Get the full on AI from podcasts, videos, white papers and more!

Document Library
Read solution briefs and white papers to help you develop and deploy AI solutions.

Blog
Read how Intel® AI Builders members are deploying cutting-edge artificial intelligence solutions across the enterprise.

Podcasts
Listen to Intel® AI Builders Members discuss how their artificial intelligence solutions solve business problems.

Videos
Watch videos of Intel AI solutions, Intel® AI Builders event demos and more.

Configuration for Intel® DL Boost Performance Gains over FP32 on Xeon®

4.0x performance boost with TensorFlow ResNet101: Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8280L Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0348.011820191451 (ucode:0x5000017), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: TensorFlow: <https://hub.docker.com/r/intelai/gp/intel-optimized-tensorflow:PR25765-devel-mkl> (<https://github.com/tensorflow/tensorflow.git> commit: 6f2eaa3b99c241a9c09c345e1029513bc4cd470a + Pull Request PR 25765, PR submitted for upstreaming), Compiler: gcc 6.3.0, MKL DNN version: v0.17, ResNet101: https://github.com/IntelAI/models/tree/master/models/image_recognition/tensorflow/resnet101 commit: 87261e70a902513f934413f009364c4f2eed6642, Synthetic data, Batch Size=128, 2 instance/2 socket, Datatype: INT8 vs Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8180 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2666 MHz), BIOS: SE5C620.86B.0D.01.0286.121520181757 (ucode:0x2000057), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: TensorFlow: <https://hub.docker.com/r/intelai/gp/intel-optimized-tensorflow:PR25765-devel-mkl> (<https://github.com/tensorflow/tensorflow.git> commit: 6f2eaa3b99c241a9c09c345e1029513bc4cd470a + Pull Request PR 25765, PR submitted for upstreaming), Compiler: gcc 6.3.0, MKL DNN version: v0.17, ResNet101: https://github.com/IntelAI/models/tree/master/models/image_recognition/tensorflow/resnet101 commit: 87261e70a902513f934413f009364c4f2eed6642, Synthetic data, Batch Size=128, 2 instance/2 socket, Datatype: FP32

3.9x performance boost with TensorFlow ResNet50: Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8280L Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0348.011820191451 (ucode:0x5000017), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: TensorFlow: <https://hub.docker.com/r/intelai/gp/intel-optimized-tensorflow:PR25765-devel-mkl> (<https://github.com/tensorflow/tensorflow.git> commit: 6f2eaa3b99c241a9c09c345e1029513bc4cd470a + Pull Request PR 25765, PR submitted for upstreaming), Compiler: gcc 6.3.0, MKL DNN version: v0.17, ResNet50: https://github.com/IntelAI/models/tree/master/models/image_recognition/tensorflow/resnet50 commit: 87261e70a902513f934413f009364c4f2eed6642, Synthetic data, Batch Size=128, 2 instance/2 socket, Datatype: INT8 vs Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8180 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2666 MHz), BIOS: SE5C620.86B.0D.01.0286.121520181757 (ucode:0x2000057), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: TensorFlow: <https://hub.docker.com/r/intelai/gp/intel-optimized-tensorflow:PR25765-devel-mkl> (<https://github.com/tensorflow/tensorflow.git> commit: 6f2eaa3b99c241a9c09c345e1029513bc4cd470a + Pull Request PR 25765, PR submitted for upstreaming), Compiler: gcc 6.3.0, MKL DNN version: v0.17, ResNet50: https://github.com/IntelAI/models/tree/master/models/image_recognition/tensorflow/resnet50 commit: 87261e70a902513f934413f009364c4f2eed6642, Synthetic data, Batch Size=128, 2 instance/2 socket, Datatype: FP32

3.9x performance boost with OpenVino™ ResNet50: Tested by Intel as of 1/30/2019. 2 socket Intel® Xeon® Platinum 8280 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0271.120720180605 (ucode:0x4000013), Linux 4.15.0-43-generic-x86_64-with-debian-buster-sid, Compiler: gcc (Ubuntu 7.3.0-27ubuntu1~18.04) 7.3.0, Deep Learning Toolkit: OpenVINO R5 (DLDTK Version:1.0.19154, AIXPRT CP (Community Preview) benchmark (<https://www.principledtechnologies.com/benchmarkxpri/aixpri/>) BS=64, Imagenet images, 1 instance/2 socket, Datatype: INT8 vs Tested by Intel as of 1/30/2019. 2 socket Intel® Xeon® Platinum 8180 Processor, 28 cores HT On Turbo ON Total Memory 192 GB (12 slots/ 16GB/ 2633 MHz), BIOS: SE5C620.86B.0D.01.0271.120720180605, Linux 4.15.0-29-generic-x86_64-with-Ubuntu-18.04-bionic, Compiler: gcc (Ubuntu 7.3.0-27ubuntu1~18.04) 7.3.0, Deep Learning Toolkit: OpenVINO R5 (DLDTK Version:1.0.19154, AIXPRT CP (Community Preview) benchmark (<https://www.principledtechnologies.com/benchmarkxpri/aixpri/>) BS=64, Imagenet images, 1 instance/2 socket, Datatype: FP32

3.8x performance boost with MXNet ResNet101: Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8280L Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0348.011820191451 (ucode:0x5000017), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: MXNet <https://github.com/apache/incubator-mxnet.git> -b master da5242b732de39ad47d8eece582f261ba5935fa9, Compiler: gcc 6.3.1, MKL DNN version: v0.17, ResNet101: https://github.com/apache/incubator-mxnet/blob/master/python/MXNet/gluon/model_zoo/vision/resnet.py, Synthetic Data, Batch Size=64, 2 instance/2 socket, Datatype: INT8 vs Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8180 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2666 MHz), BIOS: SE5C620.86B.0D.01.0286.121520181757 (ucode:0x2000057), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: MXNet <https://github.com/apache/incubator-mxnet.git> -b master da5242b732de39ad47d8eece582f261ba5935fa9, Compiler: gcc 6.3.1, MKL DNN version: v0.17, ResNet101: https://github.com/apache/incubator-mxnet/blob/master/python/MXNet/gluon/model_zoo/vision/resnet.py, Synthetic Data, Batch Size=64, 2 instance/2 socket, Datatype: FP32

3.7x performance boost with PyTorch ResNet50: Tested by Intel as of 2/25/2019. 2 socket Intel® Xeon® Platinum 8280 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0271.120720180605 (ucode: 0x4000013), Ubuntu 18.04.1 LTS, kernel 4.15.0-45-generic, SSD 1x sda INTEL SSDSC2BA80 SSD 745.2GB, 3X INTEL SSDPE2KX040T7 SSD 3.7TB, Deep Learning Framework: Pytorch with ONNX/Caffe2 backend: <https://github.com/pytorch/pytorch.git> (commit: 4ac91b2d64eeea5ca21083831db5950dc08441d6) and Pull Request link: <https://github.com/pytorch/pytorch/pull/17464> (submitted for upstreaming), gcc (Ubuntu 7.3.0-27ubuntu1~18.04) 7.3.0, MKL DNN version: v0.17.3 (commit hash: 0c3cb94999919d33e4875177def662bd9413dd4), ResNet-50: <https://github.com/intel/optimized-models/tree/master/pytorch>, Synthetic Data, Batch Size=512, 2 instance/2 socket, Datatype: INT8 vs Tested by Intel as of 2/25/2019. 2 socket Intel® Xeon® Platinum 8180 Processor, 28 cores HT On Turbo ON Total Memory 192 GB (12 slots/ 16GB/ 2666 MHz), BIOS: SE5C620.86B.00.01.0015.110720180833 (ucode: 0x200004d), CentOS 7.5, 3.10.0-693.el7.x86_64, Intel® SSD DC S4500 SERIES SSDSC2KB480G7 2.5" 6Gbps SATA SSD 480GB, Deep Learning Framework: Pytorch with ONNX/Caffe2 backend: <https://github.com/pytorch/pytorch.git> (commit: 4ac91b2d64eeea5ca21083831db5950dc08441d6) and Pull Request link: <https://github.com/pytorch/pytorch/pull/17464> (submitted for upstreaming), gcc (Ubuntu 7.3.0-27ubuntu1~18.04) 7.3.0, MKL DNN version: v0.17.3 (commit hash: 0c3cb94999919d33e4875177def662bd9413dd4), ResNet-50: <https://github.com/intel/optimized-models/tree/master/pytorch>, Synthetic Data, Batch Size=512, 2 instance/2 socket, Datatype: FP32

Configuration for Intel® DL Boost Performance Gains over FP32 on Xeon® (cont.)

performance boost with MXNet ResNet50: Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8280L Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0348.011820191451 (ucode:0x5000017), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: MXNet <https://github.com/apache/incubator-mxnet> -b master da5242b732de39ad47d8eece582f261ba5935fa9, Compiler: gcc 6.3.1, MKL DNN version: v0.17, ResNet50: https://github.com/apache/incubator-mxnet/blob/master/python/MXNet/gluon/model_zoo/vision/resnet.py, Synthetic Data, Batch Size=64, 2 instance/2 socket, Datatype: INT8 vs Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8180 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2666 MHz), BIOS: SE5C620.86B.0D.01.0286.121520181757 (ucode:0x2000057), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: MXNet <https://github.com/apache/incubator-mxnet> -b master da5242b732de39ad47d8eece582f261ba5935fa9, Compiler: gcc 6.3.1, MKL DNN version: v0.17, ResNet50: https://github.com/apache/incubator-mxnet/blob/master/python/MXNet/gluon/model_zoo/vision/resnet.py, Synthetic Data, Batch Size=64, 2 instance/2 socket, Datatype: FP32

2.5x Performance boost with MXNet SSD-VGG16 Inference: Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8280L Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0348.011820191451 (ucode:0x5000017), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: MXNet <https://github.com/apache/incubator-mxnet> -b master da5242b732de39ad47d8eece582f261ba5935fa9, Compiler: gcc 6.3.1, MKL DNN version: v0.17, SSD-VGG16: https://github.com/apache/incubator-mxnet/blob/master/example/ssd/symbol/vgg16_reduced.py, Synthetic Data, Batch Size=224, 2 instance/2 socket, Datatype: INT8 vs Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8180 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2666 MHz), BIOS: SE5C620.86B.0D.01.0286.121520181757 (ucode:0x2000057), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: MXNet <https://github.com/apache/incubator-mxnet> -b master da5242b732de39ad47d8eece582f261ba5935fa9, Compiler: gcc 6.3.1, MKL DNN version: v0.17, SSD-VGG16: https://github.com/apache/incubator-mxnet/blob/master/example/ssd/symbol/vgg16_reduced.py, Synthetic Data, Batch Size=224, 2 instance/2 socket, Datatype: FP32

2.2x performance boost with Intel® Optimized Caffe SSD-MobileNet v1: Tested by Intel as of 2/20/2019. 2 socket Intel® Xeon® Platinum 8280 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0271.120720180605 (ucode: 0x4000013), Ubuntu 18.04.1 LTS, kernel 4.15.0-45-generic, SSD 1x sda INTEL SSDSC2BA80 SSD 745.2GB, Deep Learning Framework: Intel® Optimization for Caffe version: 1.1.3 (commit hash: 7010334f159da247db3fe3a9d96a3116ca06b09a), ICC version 18.0.1, MKL DNN version: v0.17 (commit hash: 830a10059a018cd2634d94195140c2d8790a75a), model: https://github.com/intel/caffe/blob/master/models/intel_optimized_models/int8/ssd_mobilenet_int8_prototxt, Synthetic Data, Batch Size=64, 2 instance/2 socket, Datatype: INT8 vs Tested by Intel as of 2/21/2019. 2 socket Intel® Xeon® Platinum 8180 Processor, 28 cores HT On Turbo ON Total Memory 192 GB (12 slots/ 16GB/ 2666 MHz), BIOS: SE5C620.86B.0D.01.0015.110720180833 (ucode: 0x200004d), CentOS 7.5, 3.10.0-693.el7.x86_64, Intel® SSD DC S4500 SERIES SSDSC2KB480G7 2.5" 6Gb/s SATA SSD 480GB, Deep Learning Framework: Intel® Optimization for Caffe version: 1.1.3 (commit hash: 7010334f159da247db3fe3a9d96a3116ca06b09a), ICC version 18.0.1, MKL DNN version: v0.17 (commit hash: 830a10059a018cd2634d94195140c2d8790a75a), model: https://github.com/intel/caffe/blob/master/models/intel_optimized_models/int8/ssd_mobilenet_int8_prototxt, Synthetic Data, Batch Size=64, 2 instance/2 socket, Datatype: FP32

2.6x performance boost with PyTorch RetinaNet: Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8280 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0271.120720180605 (ucode: 0x4000013), Ubuntu 18.04.1 LTS, kernel 4.15.0-45-generic, SSD 1x sda INTEL SSDSC2BA80 SSD 745.2GB, 3X INTEL SSDPE2KX040T7 SSD 3.7TB, Deep Learning Framework: Pytorch with ONNX/Caffe2 backend: <https://github.com/pytorch/pytorch> (commit: 4ac91b2d64eeea5ca21083831db595dc08441d6) and Pull Request link: <https://github.com/pytorch/pytorch/pull/17464> (submitted for upstreaming), gcc (Ubuntu 7.3.0-27ubuntu1~18.04) 7.3.0, MKL DNN version: v0.17.3 (commit hash: 0c3cb94999919d33e4875177fde662bd9413dd4), RetinaNet: https://github.com/intel/Detector/blob/master/configs/12_2017_baselines/retinanet_R-101-FPN_1x.yaml, BS=1, synthetic data, 2 instance/2 socket, Datatype: INT8 vs Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8180 Processor, 28 cores HT On Turbo ON Total Memory 192 GB (12 slots/ 16GB/ 2666 MHz), BIOS: SE5C620.86B.0D.01.0015.110720180833 (ucode: 0x200004d), CentOS 7.5, 3.10.0-693.el7.x86_64, Intel® SSD DC S4500 SERIES SSDSC2KB480G7 2.5" 6Gb/s SATA SSD 480GB, Deep Learning Framework: Pytorch with ONNX/Caffe2 backend: <https://github.com/pytorch/pytorch> (commit: 4ac91b2d64eeea5ca21083831db595dc08441d6) and Pull Request link: <https://github.com/pytorch/pytorch/pull/17464> (submitted for upstreaming), gcc (Ubuntu 7.3.0-27ubuntu1~18.04) 7.3.0, MKL DNN version: v0.17.3 (commit hash: 0c3cb94999919d33e4875177fde662bd9413dd4), RetinaNet: https://github.com/intel/Detector/blob/master/configs/12_2017_baselines/retinanet_R-101-FPN_1x.yaml, BS=1, synthetic data, 2 instance/2 socket, Datatype: FP32

2.1x performance boost with TensorFlow Wide & Deep: Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8280L Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0348.011820191451 (ucode:0x5000017), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: TensorFlow <https://github.com/tensorflow/tensorflow> A3262818d9d8f9630f04df23033032d39a7413 + Pull Request PR26169 + Pull Request PR26261 + Pull Request PR26271, PR submitted for upstreaming, Compiler: gcc 6.3.1, MKL DNN version: v0.18, Wide & Deep: https://github.com/IntelAI/models/tree/master/benchmarks/recommendation/tensorflow/wide_deep_large_ds commit: a044cb3e7d2b082aebae2edbe6435e7a2cc18f, Model: https://storage.googleapis.com/intel-optimized-tensorflow/models/wide_deep_int8_pretrained_model.pb, Dataset: Criteo Display Advertisement Challenge, Batch Size=512, 1 instance/1 socket, Datatype: INT8 vs Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8180 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2666 MHz), BIOS: SE5C620.86B.0D.01.0286.121520181757 (ucode:0x2000057), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: TensorFlow <https://github.com/tensorflow/tensorflow> A3262818d9d8f9630f04df23033032d39a7413 + Pull Request PR26169 + Pull Request PR26261 + Pull Request PR26271, PR submitted for upstreaming, Compiler: gcc 6.3.1, MKL DNN version: v0.18, Wide & Deep: https://github.com/IntelAI/models/tree/master/benchmarks/recommendation/tensorflow/wide_deep_large_ds commit: a044cb3e7d2b082aebae2edbe6435e7a2cc18f, Model: https://storage.googleapis.com/intel-optimized-tensorflow/models/wide_deep_int8_pretrained_model.pb, Dataset: Criteo Display Advertisement Challenge, Batch Size=512, 1 instance/1 socket, Datatype: FP32

2.1x performance boost with MXNet Wide & Deep: Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8280L Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0348.011820191451 (ucode:0x5000017), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: MXNet <https://github.com/apache/incubator-mxnet> commit f1de8e51999ce3acaa95538d21a91fe43a0286ec applying https://github.com/intel/optimized-models/blob/v1.0.2/mxnet/wide_deep_criteo/patch.diff, Compiler: gcc 6.3.1, MKL DNN version: commit: 08bd90cca77683d5d1c98068cea8b92ed05784, Wide & Deep: https://github.com/intel/optimized-models/tree/v1.0.2/mxnet/wide_deep_criteo commit: c3e7cbde4209c3657ecb6c9a14271c3672654a5, Dataset: Criteo Display Advertisement Challenge, Batch Size=1024, 2 instance/2 socket, Datatype: INT8 vs Tested by Intel as of 3/26/2019. 2 socket Intel® Xeon® Platinum 8180 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2666 MHz), BIOS: SE5C620.86B.0D.01.0286.121520181757 (ucode:0x2000057), CentOS 7.6, Kernel 4.19.5-1.el7.elrepo.x86_64, SSD 1x INTEL SSDSC2K96 960GB, Deep Learning Framework: MXNet <https://github.com/apache/incubator-mxnet> commit f1de8e51999ce3acaa95538d21a91fe43a0286ec applying https://github.com/intel/optimized-models/blob/v1.0.2/mxnet/wide_deep_criteo/patch.diff, Compiler: gcc 6.3.1, MKL DNN version: commit: 08bd90cca77683d5d1c98068cea8b92ed05784, Wide & Deep: https://github.com/intel/optimized-models/tree/v1.0.2/mxnet/wide_deep_criteo commit: c3e7cbde4209c3657ecb6c9a14271c3672654a5, Dataset: Criteo Display Advertisement Challenge, Batch Size=1024, 2 instance/2 socket, Datatype: FP32

INTEL AI BUILDERS PARTNER SHOWCASE

SHOWCASE SCHEDULE

Meeting rooms are available to meet with presenting partners throughout the event.

Time	Session	Speaker
1:30pm	Welcome and AIB overview and growth: Impact on AI ecosystem	Brigitte Alexander (Intel)
1:40pm	Driving business impact with the Intel AI technology portfolio	Ananth Sankaranarayanan (Intel)
2:00pm	Clinical deployment of radiology AI powered by OpenVINO	Liren Zhu (Subtle Medical)
2:10pm	Industrialize AI with Cloudera	Jessie Lin (Cloudera)
2:20pm	Automated time series forecasting and optimization for enterprises	Yuan Shen (OneClick.ai)
2:30pm	QuEST vision analytics solution with OpenVINO and Intel AI	Rubayat Mahmud (QuEST Global)
2:40pm	Saving Antarctic penguins with deep learning	Ganes Kesari (Gramener)
2:50pm	Pipe Sleuth: AI-based pipeline assessment	Sundar Varadarajan (Wipro)

Time	Session	Speaker
3:00-3:30pm	Afternoon break Outside meeting rooms	
3:30pm	AI-based container usage optimization tool	Amine Kerkeni (InstaDeep)
3:40pm	The turnkey high-compliance AI platform	David Talby (John Snow Labs)
3:50pm	Accelerating AI from research to production in the enterprise	Ari Kamlani (Skymind)
4:00pm	Running enterprise IT more efficiently, improving customer experience, and increasing the agility and stability of IT	Anjali Gajendragadkar (Digitate)
4:10pm	Using AI to accelerate time to customer	Derek Wang (Stratifyd)
4:20pm	Data analytics at the retail edge	Han Yang (Cisco)
4:30pm	Accelerate innovation with DevOps-like agility for machine learning pipelines	Nanda Vijaydev (Hewlett Packard Enterprise)
4:40pm	Accelerating deep learning workloads in the cloud and data centers	Ravi Panchumathy (Intel)
5:00 PM	How to leverage powerful Intel-based instance types to create new solutions	Carlos Escapa (Amazon Web Services)
5:10 PM	Intel-based AI solutions from cloud to edge	Bob Anderson (Inspur)
5:20 PM	Lenovo intelligent computing orchestration	Matt Ziegler (Lenovo)
5:30 PM	Dell Ready Solutions	Philip Hummel (Dell EMC)
5:40 PM	Closing Remarks	Brigitte Alexander (Intel)





**AI ON
INTEL**

**AI BUILDERS SHOWCASE
SUBTLE MEDICAL**

AGENDA

- Company overview
- Business problem they solve by prioritized vertical
- Use of Intel® AI technology
- Results
- Contact



Founder & CEO

Enhao Gong, PhD

*EE @ Stanford University
BME @ Tsinghua University
Forbes 30under30 China & Asia*



Founder

Greg Zaharchuk, MD PhD

*Radiologist @ Stanford Hospital
Professor @ Stanford Medicine
Director, Center of Advanced Functional Neuroimaging
Chair, ISMRM Machine Learning Workshop*



- Subtle Medical Inc was founded in July 2017 from Stanford University
- First AI-powered product SubtlePET received FDA clearance and CE Mark in Nov 2018

Key Team Members



Tao Zhang, PhD

Head of Research & Development



GE Healthcare



Liren Zhu, PhD

Head of Engineering



Caltech



Lusi Chien, MBA

Head of Commercialization



Praveen Gulaka, PhD

Head of Clinical & Regulatory Affairs

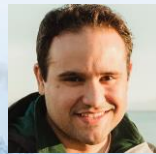


SAMSUNG



Anna Menyhart

Head of Marketing



Jorge Guzman

Head of Platform



GE Healthcare



Ajit Shankaranarayanan, PhD

Head of Partnership



GE Healthcare

BUSINESS PROBLEM SOLVED

The Radiology Workflow



- More frequent usage
- Immediate and bigger financial value
- Fundamental for downstream application

90% of the cost of imaging

BUSINESS PROBLEM SOLVED

Better Economics



More patients on existing machines without additional capital purchase



+1 patient scanned per day
= approx. \$500k a year

Better Care



Competitive advantage: Latest technology for better quality and service



Increased patient comfort
= Higher patient satisfaction

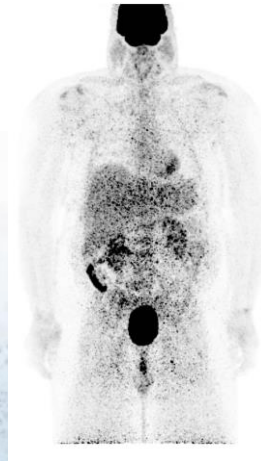
BUSINESS PROBLEM SOLVED

SubtlePET™ enhances up to 4x faster PET scans



Original Scan Time 18 min

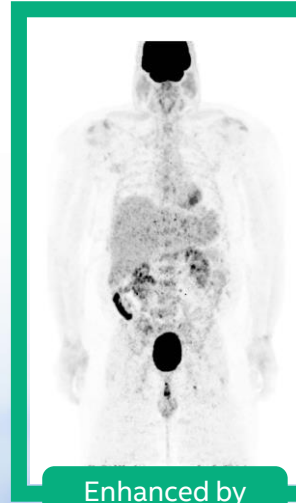
4.5 min scan



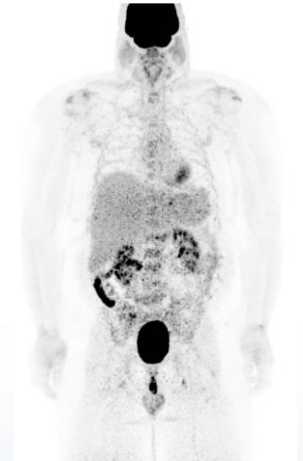
Fast scan - Noisy

Powered by
Intel HW and SW

4.5 min scan



Enhanced by
SubtlePET™



USE OF INTEL® AI TECHNOLOGY

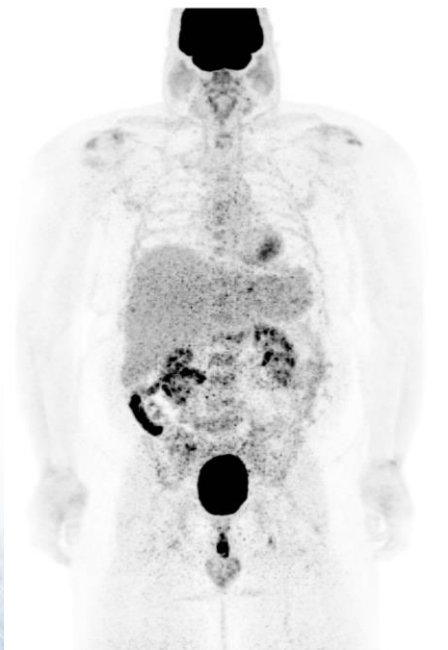
HW: Intel® Xeon® CPU E5 family, Intel® Core™ i5 and i7 series

SW: Intel® Distribution of OpenVINO™ Toolkit 2019 R1.1

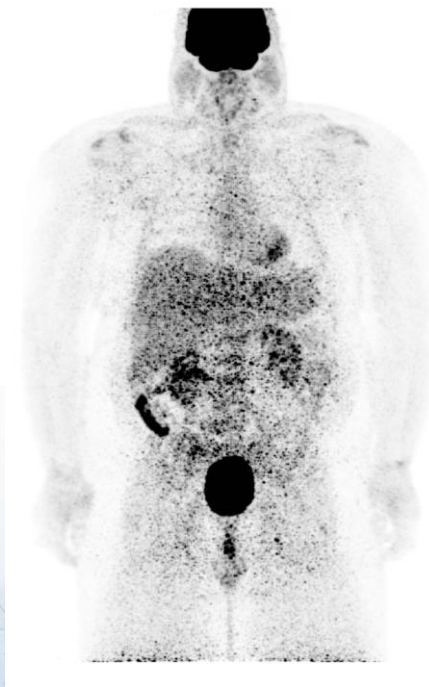


RESULTS

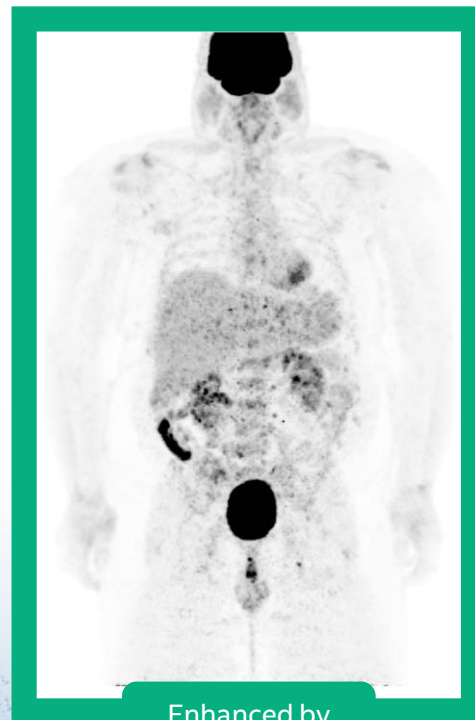
24 min scan



6 min scan



6 min scan

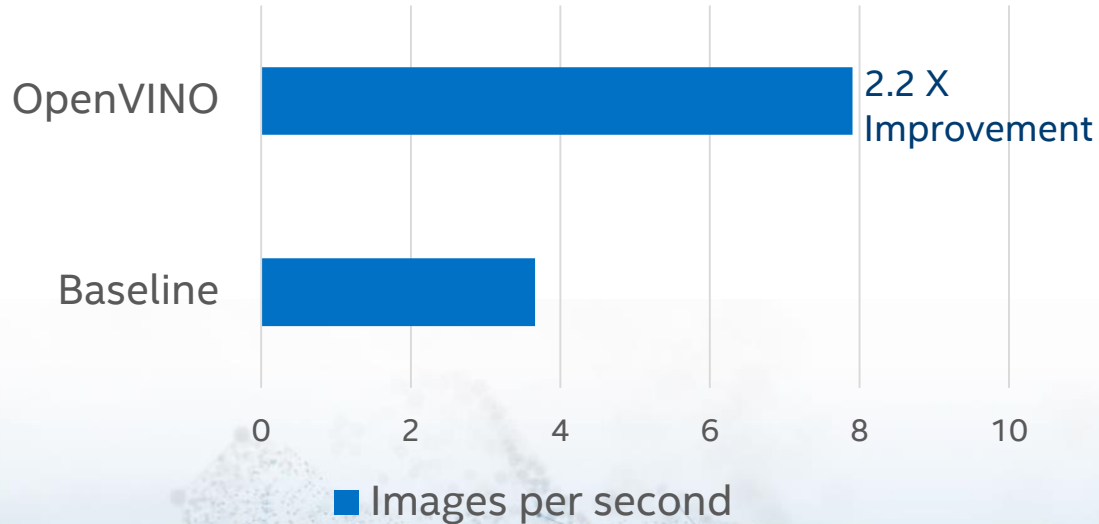


Enhanced by
SubtlePET™

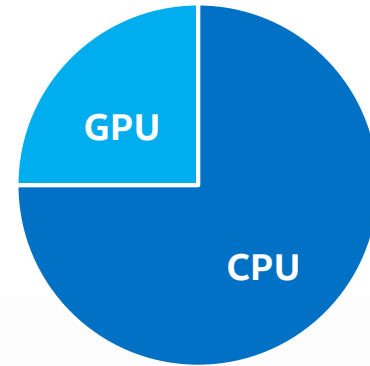
*Note: Images are Coronal Maximum Intensity Projection

RESULTS

2.2X Inference Speed Improvement



75% Deployments Use CPU



CONFIGURATION SPECIFICATIONS

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

Configuration: AWS p3.2xlarge instance, Intel® Xeon® E5-2686 v4 @ 2.30 GHz, 61 GB RAM, Intel® OpenVINO™ Toolkit. Testing done by Subtle Medical, Jan 2019

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

Performance results are based on testing as of dates shown in configuration and may not reflect all publicly available security updates. No product can be absolutely secure. See configuration disclosure for details.

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. 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>

CONTACT

Liren Zhu

Head of Engineering

liren@subtlemedical.com



Visit our table with your questions, or stop by the Intel® AI Builders matchmaking table to set up a private meeting.





**AI ON
INTEL**

**AI BUILDERS SHOWCASE
CLOUDERA**

AGENDA

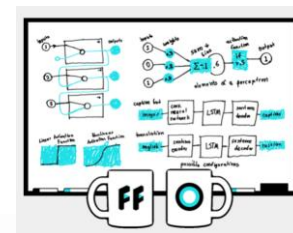
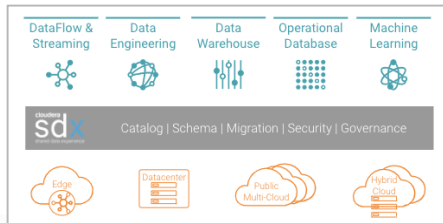
CLUSTERA

- Company overview
- Predictive Maintenance Use Case
- Use of Intel® AI technology
- Contact



MACHINE LEARNING AT CLUSTERA

Our approach

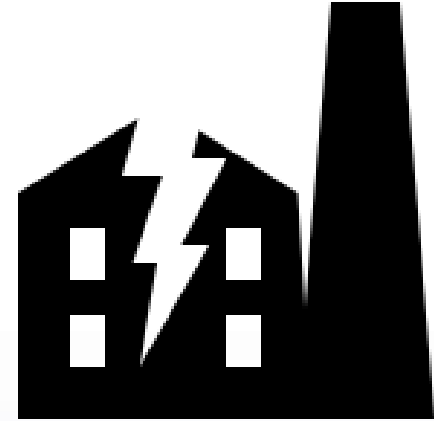
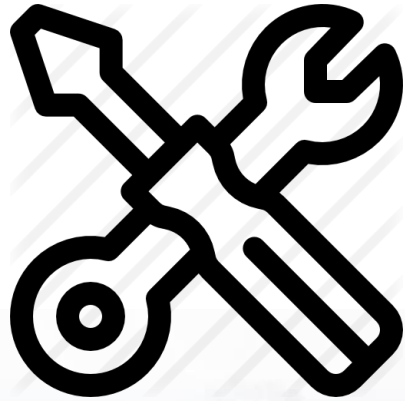


Open **platform** to build, train, and deploy many scalable ML applications

Comprehensive data science **tools** to accelerate team productivity

Expert guidance & services to fast track value & scale

PREDICTIVE MAINTENANCE



SOLUTION ARCHITECTURE

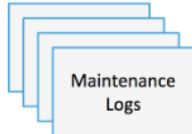
Real-time Sensor Ingest
Using Kafka Messaging System



Batch Database Ingest
Using Sqoop DB Transfer



Batch Text Ingest
Using HUE File Transfer



Data Enrichment & Scoring
Using Spark Streaming

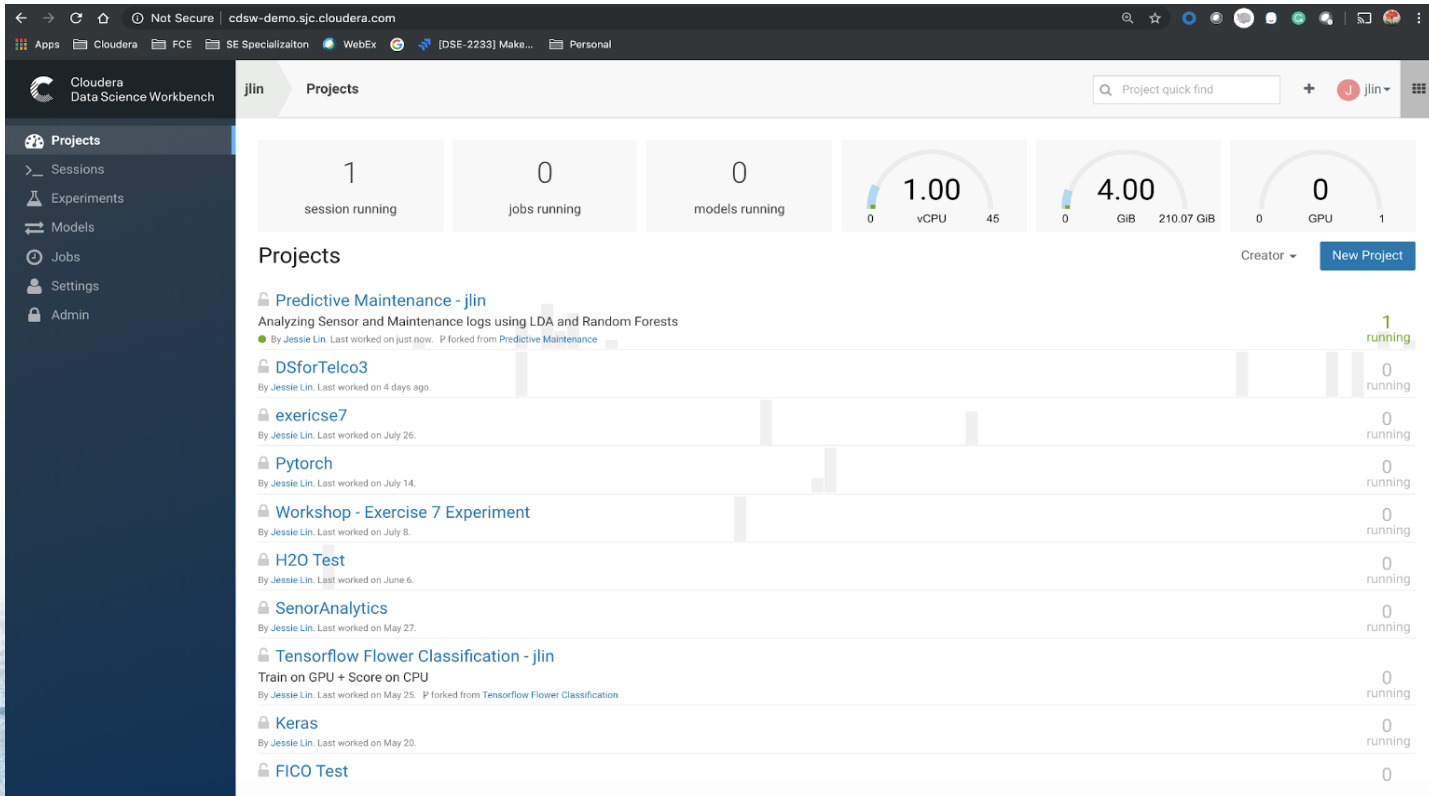


Data Modelling & Machine Learning
Using Spark



Reporting & Data Visualization
Using Impala and Cloudera Search through HUE

CLOUDERA DATA SCIENCE WORKBENCH



WORKBENCH EDITOR

SensorAnalytics_stream.r
datagenerator_local.py
interactive_sensorAnalys
SensorAnalytics_kudu.py
setup.sh
config.ini
datagenerator.py

Predictive Maintenance -
jlin
▼ cdsw
batchScoring_sensorAr
experiment_sensorAna
flatten_all_spark_scher
interactive_sensorAnal
predict_SensorAnalysis
SensorAnalytics_hdfs.r
SensorAnalytics_kudu.j
SensorAnalytics_strear
streamScoring_sensor
cdsw-build.sh
config.ini
► datagen
► img
► maintenance
► models
README.md
► sampledata
► seaborn-data
setup.sh
► slides
spark-defaults.conf
spark_rf.tar
► sql

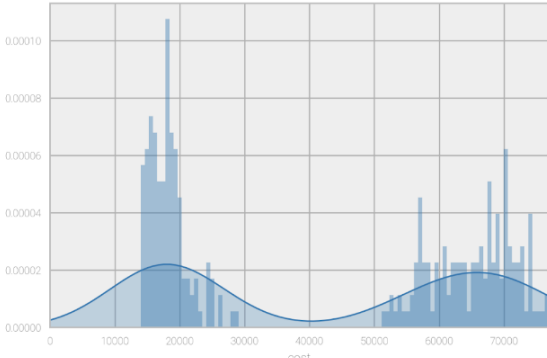
File Edit View Navigate Run cdsw/interactive_sensor...
93 import pandas as pd
94
95
96 # ### Create a Spark Session
97 spark = SparkSession.builder.appName("Sensor Anal
98 sc = spark.sparkContext
99 sqc = SQLContext(sc)
100
101 # ### Analyze Maintenance Costs
102 # We start our analysis with visualizing the dist
103 # ### Histogram of Maintenance Costs
104 rawMaintCosts = sc.textFile("sampledata/maint_cos
105 schemaString = "date cost"
106 schema = StructType([StructField(field_name, Stri
107 maintCosts = spark.createDataFrame(rawMaintCosts,
108 maintCosts = maintCosts.select(maintCosts.date.ca
109 maintCostsPD = maintCosts.toPandas()
110 maintCostsPD.describe()
111 maintCosts.groupBy(F.date_format('date', 'yyyyMM')
112 .agg(F.round(F.sum('cost')).alias('cost'))\
113 .orderBy('month').toPandas().plot(kind='line',
114 sb.distplot(maintCostsPD['cost'], bins=100, hist=
115
116 # ## Text Analytics of Maintenance Logs
117 # Spark provides rich text analytics capabilities
118 # stop words removal, vectorization, and more tha
119 # models based on textual data.
120 # ### Sample of Maintenance Logs
121 maintenance = spark.read.format("csv").option("de
122 .load("sampledata/maint_notes.txt")\
123 .withColumnRenamed('_c0', 'date')\
124 .withColumnRenamed('_c1', 'note')\
125 .withColumnRenamed('_c2', 'duration')\
126 .withColumn('note', F.lower(F.regexp_replace('n
127 .select(F.col('date').cast('date'), 'note', F.c
128 maintenance.show(5, truncate=False)
129
130 # ### Sample of 2-word nGrams on Maintenance Note
131 tk = Tokenizer(inputCol="note", outputCol="words"
132 maintTokenized = tk.transform(maintenance)
133 swr = StopWordsRemover(inputCol="words", outputCo
134 maintFiltered = swr.transform(maintTokenized)
135 ngram = NGram(n=2, inputCol="filtered", outputCol
136 maintNGrams = ngram.transform(maintFiltered)
137 maintNGrams.select('ngrams').show(5, truncate=Fa

Untitled Session Timeout
By Jessie Lin - Python 3 Session - 1 vCPU / 4 GiB Memory - just now

Session Logs Spark UI Collapse Share

0 month

> sb.distplot(maintCostsPD['cost'], bins=100, hist=True, kde_kws={"shade": True}).set(xlim=(0, maxi
[(0, 76872)]



Text Analytics of Maintenance Logs

1.6.0.1294376 (46715e4)

MAINTENANCE NOTES

Sample of Maintenance Logs

```
> maintenance = spark.read.format("csv").option("delimiter", "|")\
  .load("sampledata/maint_notes.txt")\
  .withColumnRenamed('_c0', 'date')\
  .withColumnRenamed('_c1', 'note')\
  .withColumnRenamed('_c2', 'duration')\
  .withColumn('note', F.lower(F.regexp_replace('note', '[!?-]', ' ')))\
  .select(F.col('date').cast('date'), 'note', F.col('duration').cast('int'))
> maintenance.show(5, truncate=False)
```

date	note	duration
2014-04-08	asset failure due to high sensor_6 and sensor_5 asset shutdown corrective maintenance required	22
2014-04-10	program maintenance tests all passed asset healthy sensor readings normal	2
2014-04-16	asset failure due to high sensor_8 and sensor_5 asset shutdown corrective maintenance required	22
2014-04-19	asset failure due to high sensor_2 and sensor_5 asset shutdown corrective maintenance required	18
2014-04-20	program maintenance tests all passed asset healthy sensor readings normal	3

only showing top 5 rows

Topic Clustering using Latent Dirichlet Allocation (LDA)

LDA is a form of un-supervised machine learning that identifies clusters, or topics, in the data

```
> cv = CountVectorizer(inputCol="ngrams", outputCol="features", vocabSize=50)\
  .fit(maintNGrams) # CountVectorizer converts nGram array into a vector of counts
> maintVectors = cv.transform(maintNGrams)
> vocabArray = cv.vocabulary
> lda = LDA(k=3, maxIter=10)
> ldaModel = lda.fit(maintVectors)
> ldaModel.write().overwrite().save('lda.mdl')
> maint_topics=[None] * 3
> topics = ldaModel.describeTopics(5)
```

We see below that each maintenance log can be clustered based on its text into 1 of 3 topics below. The nGrams in each cluster show clearly 3 types of maintenance activities

1. Preventive maintenance occurs when we have 'abnormal readings' or a 'component replacement'
2. Corrective maintenance occurs when we have a 'asset shutdown' event or 'asset failure'
3. The rest of the logs indicate that no downtime is required (ie. 'maintenance tests passed', 'asset healthy')

We see below that each maintenance log can be clustered based on its text into 1 of 3 topics below. The nGrams in each cluster show clearly 3 types of maintenance activities

1. Preventive maintenance occurs when we have 'abnormal readings' or a 'component replacement'
2. Corrective maintenance occurs when we have a 'asset shutdown' event or 'asset failure'
3. The rest of the logs indicate that no downtime is required (ie. 'maintenance tests passed', 'asset healthy')

```
> for topic in topics.collect():
  print('Topic %d Top 5 Weighted nGrams' % (topic[0]+1))
  for termIndex in topic[1]:
    print(' %s' % vocabArray[termIndex])
```

Topic 1 Top 5 Weighted nGrams

sensor readings
tests passed
asset healthy

program maintenance
readings normal

Topic 2 Top 5 Weighted nGrams

sensor_5 asset
asset shutdown
corrective maintenance
due high
failure due

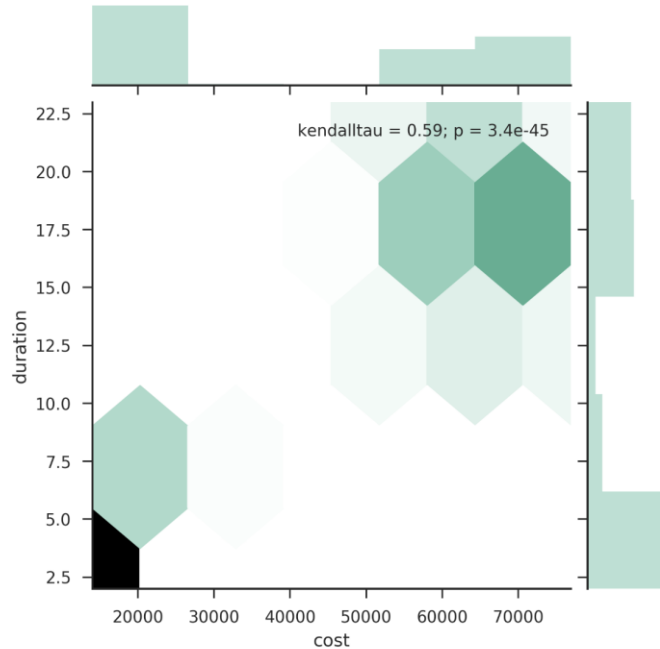
Topic 3 Top 5 Weighted nGrams

showing abnormal
abnormal readings
component replacement
readings preventive
required scheduling

MAINTENANCE COST AND DURATION

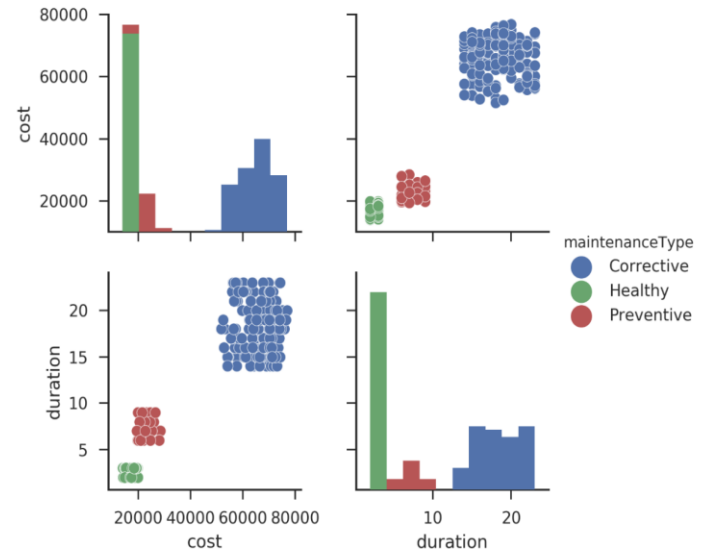
```
> sb.jointplot(maintClusters['cost'], maintClusters['duration'],  
              kind="hex", stat_func=kendalltau, color="#4CB391")
```

```
<seaborn.axisgrid.JointGrid at 0x7f2aecc92550>
```



```
> sb.pairplot(maintClusters, hue="maintenanceType", vars=['cost', 'duration'])
```

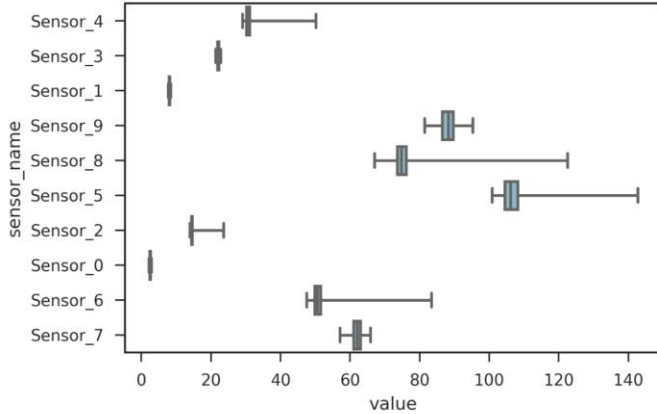
```
<seaborn.axisgrid.PairGrid at 0x7f2aec9817f0>
```



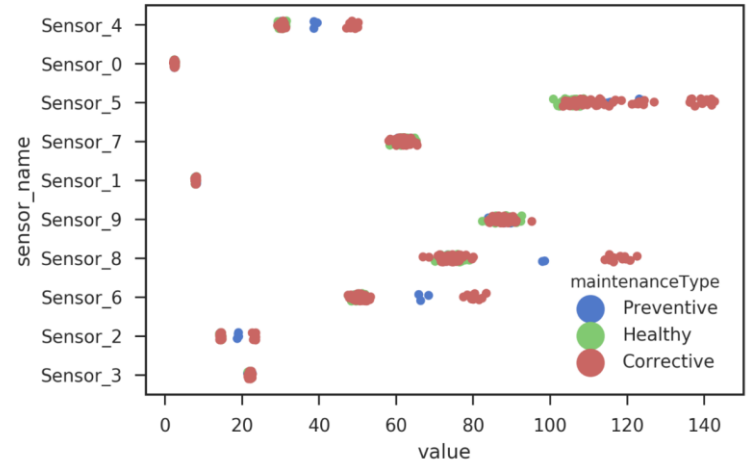
SENSOR READINGS

```
> sb.set(style="ticks", palette="muted", color_codes=True)
> ax = sb.boxplot(x="value", y="sensor_name",
    data=dailyRawMeasurements.filter('value!=0 and year(day)>=2016').select('
    whis=np.inf, color="c")

/usr/local/lib/python3.6/site-packages/seaborn/categorical.py:462: FutureWarning: remove
_na is deprecated and is a private function. Do not use.
    box_data = remove_na(group_data)
```



```
rawSensorsByMaint = progPrevMaint.union(corrMaint)
> sb.stripplot(y="sensor_name", x="value", hue="maintenanceType", jitter=True,
    data=rawSensorsByMaint.filter('year(date)>=2016').select('sensor_name', 'value', 'maintenanceTyp
<matplotlib.axes._subplots.AxesSubplot at 0x7f2aec3e6eb8>
```



MODEL TRAINING AND EVALUATION

Model Training

We split the data into 2 subsets - one to train the model, and one to test/evaluate it. We then build a pipeline of all steps involved in running the model on some data. The pipeline will have the following steps:

1. VectorAssembler - put all features into a single column
2. StringIndexer - convert the maintenance type strings to a numeric index
3. RandomForestClassifier - classify the data into one of the different indexes
4. IndexToString - convert the maintenance type indexes back to strings

```
> (trainingData, testData) = modelData.randomSplit([0.7, 0.3])
```

Now we need to convert our feature columns (sensor names) into a vector for each row

```
> va = VectorAssembler(inputCols=sensorNames, outputCol="features")
```

Index the labels (maintenance type)

```
> li = StringIndexer(inputCol='maintenanceType', outputCol='label')\
    .fit(modelData)
> rf = RandomForestClassifier(labelCol="label", featuresCol="features", num1
> i2s = IndexToString(inputCol="prediction", outputCol="predictedLabel",
    labels=li.labels)
> pipeline = Pipeline(stages=[va, li, rf, i2s])
> model = pipeline.fit(trainingData)
```

Model Evaluation

The training data was used to fit the model (ie. train it), now we can test the model using the test subset, and calculate the accuracy (ie. false prediction rate)

```
> predictions = model.transform(testData)
> evaluator = MulticlassClassificationEvaluator(
    labelCol="label", predictionCol="prediction", metricName="accuracy")
> accuracy = evaluator.evaluate(predictions)
> print("Test Error = %g" % (1.0 - accuracy))

Test Error = 0.038961
```

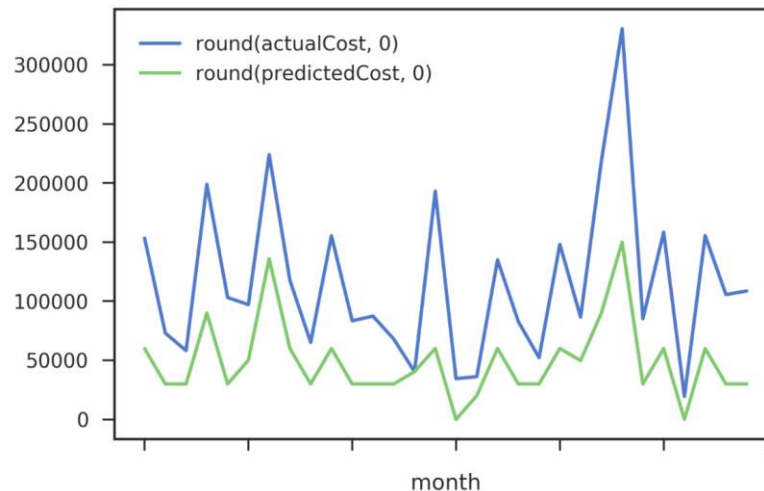
```
> predictions.groupBy(predictions.predictedLabel.alias('Prediction'),
    predictions.maintenanceType.alias('Actual'))\
    .count().orderBy('Actual', 'Prediction')\
    .toPandas()
```

	Prediction	Actual	count
0	Corrective	Corrective	44
1	Healthy	Healthy	26
2	Corrective	Preventive	2
3	Healthy	Preventive	1
4	Preventive	Preventive	4

PROJECTED SAVING

```
> def f(actual, predicted, cost):
    if actual==predicted:
        if actual=='Corrective':
            return 30000
        elif actual=='Preventive':
            return cost
        elif actual=='Healthy':
            return 0
    else:
        return cost
```

```
> csPD = costSavings.select('month', F.round('actualCost'), F.round('predictedCost')).toPandas()
> csPD.plot(kind='line', x='month')
<matplotlib.axes._subplots.AxesSubplot at 0x7f2aec200048>
```



```
> costSavings.agg(F.sum('actualCost').alias('TotalCost'),
                  F.sum('predictedSavings').alias('TotalSavings($')))\
.withColumn('TotalSavings(%)', F.col('TotalSavings($')/F.col('TotalCost')*100)\
.select('TotalSavings($)', 'TotalSavings(%)')\
.toPandas()
```

	TotalSavings(\$)	TotalSavings(%)
0	2008901	57.807273



BATCH SCORING

cdsw-demo.sjc.cloudera.com/jlin/predictive-maintenance-jlin/jobs/954/settings

Apps Cloudera FCE SE Specialization WebEx [DSE-2233] Make... Personal

Cloudera Data Science Workbench

All Projects

Overview

Sessions

Experiments

Models

Jobs

Files

Team

Settings

jlin Predictive Maintenance - jlin Jobs **DailyBatchScoring** Settings

DailyBatchScoring

Overview History Dependencies Settings

General

Name

DailyBatchScoring

Script

cdsw/batchScoring_sensorAnalysis.py

Engine Kernel

☐ Python 2

☒ Python 3

☐ R

☐ Scala

Schedule

Manual

Engine Profile

1 vCPU / 4 GiB Memory

GPUs

0 GPUs

Timeout in Minutes (optional) 30 ☐ Kill on Timeout

Jobs exceeding timeout send warning email if notifications enabled.

DailyBatchScoring

Success Pause Run

Overview History Dependencies Settings

Script: dsfortelco_interactive.py

Schedule: Every 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 and 55th minute past every hour

Engine Profile:

Created By: Jessie Lin

Latest Run: 3 minutes ago

Duration: 00:41

Runs: 275

Failures: 1



DEPLOY THE MODEL AS A REST API

RFModel

Deployed Stop Restart Deploy New

Overview Deployments Builds Monitoring Settings

Description test

Sample Code

Shell Python R

```
curl -H "Content-Type: application/json" -X POST http://cdsw-demo.sjc.cloudera.com/api/altus-ds-1/models/call-model -d '{"accessToken":"md1sze9zz86wyetmadjdmvh12jugayo j","request":{"features":"75.0, 3.0, 14.0, 107.0, 31.0, 22.0, 8.0, 53.0, 64.0, 88.0"}}'
```

Test Model

Input

```
{
  "features": "75.0, 3.0, 14.0, 107.0, 31.0, 22.0, 8.0, 53.0, 64.0, 88.0"
}
```

Test Reset

Result

Status	● success
Response	{ "result": "Healthy" }

Model Details

Model Id	519
Deployment	1381
Build	716
Deployed By	jlin
Comment	
Kernel	python3
Engine Image	Base Image v7
File	predict_SensorAnalysis.py
Function	predict

Model Resources

Replicas	1
Total CPU	1 vCPUs
Total Memory	4.00 GiB

RFModel

Overview Deployments Builds Monitoring Settings

Id	Build	Status	Deployed At	Stopped At	Deployed By
1381	2	Deployed	Aug 29, 2019, 03:22 PM		jlin
1190	2	Stopped	Jun 21, 2019, 12:23 PM	Jun 21, 2019, 12:24 PM	jlin
1093	2	Stopped	May 23, 2019, 04:58 PM	May 23, 2019, 08:04 PM	jlin
1076	2	Stopped	May 20, 2019, 10:05 PM	May 23, 2019, 12:13 PM	jlin
1074	2	Stopped	May 20, 2019, 01:28 PM	May 20, 2019, 09:57 PM	jlin
1072	2	Stopped	May 19, 2019, 02:46 PM	May 19, 2019, 09:43 PM	jlin
1071	1	Stopped	Never	May 19, 2019, 01:55 PM	jlin

USE OF INTEL® AI TECHNOLOGY

HW: Intel® Xeon® Processors

SW: Intel® Math Kernel Library (Intel® MKL)

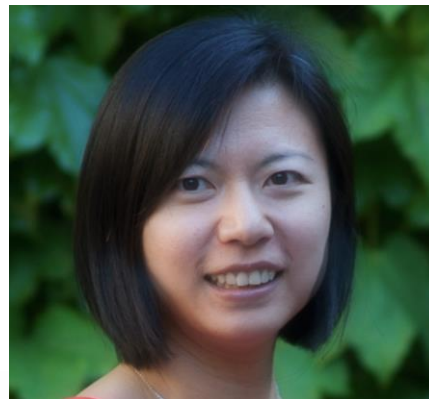
- a. <https://blog.cloudera.com/using-native-math-libraries-to-accelerate-spark-machine-learning-applications/>



CONTACT

CLUSTERA

Jessie Lin
Solution Engineer
jlin@cloudera.com



Visit our table with your questions, or stop by the Intel® AI Builders
matchmaking table to set up a private meeting.



**AI ON
INTEL**

**AI BUILDERS SHOWCASE
ONECLICK.AI**

AGENDA

- Company overview
- Business problem they solve by prioritized vertical
- Use of Intel® AI technology
- Results
- Contact

- OneClick.ai is an SaaS forecasting and optimization Platform powered by Automated Deep Learning(AutoDL) technology
- OneClick.ai empowers business analysts in Financial, CPG, Retail and Manufacturing industries to solve their top challenges in day-to-day operations with self-serving AI software to advance to AI-driven decision-making.

BUSINESS PROBLEM SOLVED

OneClick.ai can help enterprises to automate most of the customization of AI solutions

Financial
Forecast

Smart
Inventory

Dynamic
Pricing

Product
Recommendation

CHALLENGES

1. New territory of big data: new format, new sources

Temporal
factors

Operational
factors

Economical
factors

Product
information

Social media

2. Traditional methods can not keep up what advanced AI technologies like Deep Learning can achieve

TECHNOLOGY

Structured
+
Unstructured



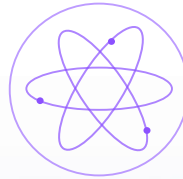
Images

Product Pictures



Time Series

Sales Record



Numbers

Transactions,
Product Category



Text

Product Description
User Reviews



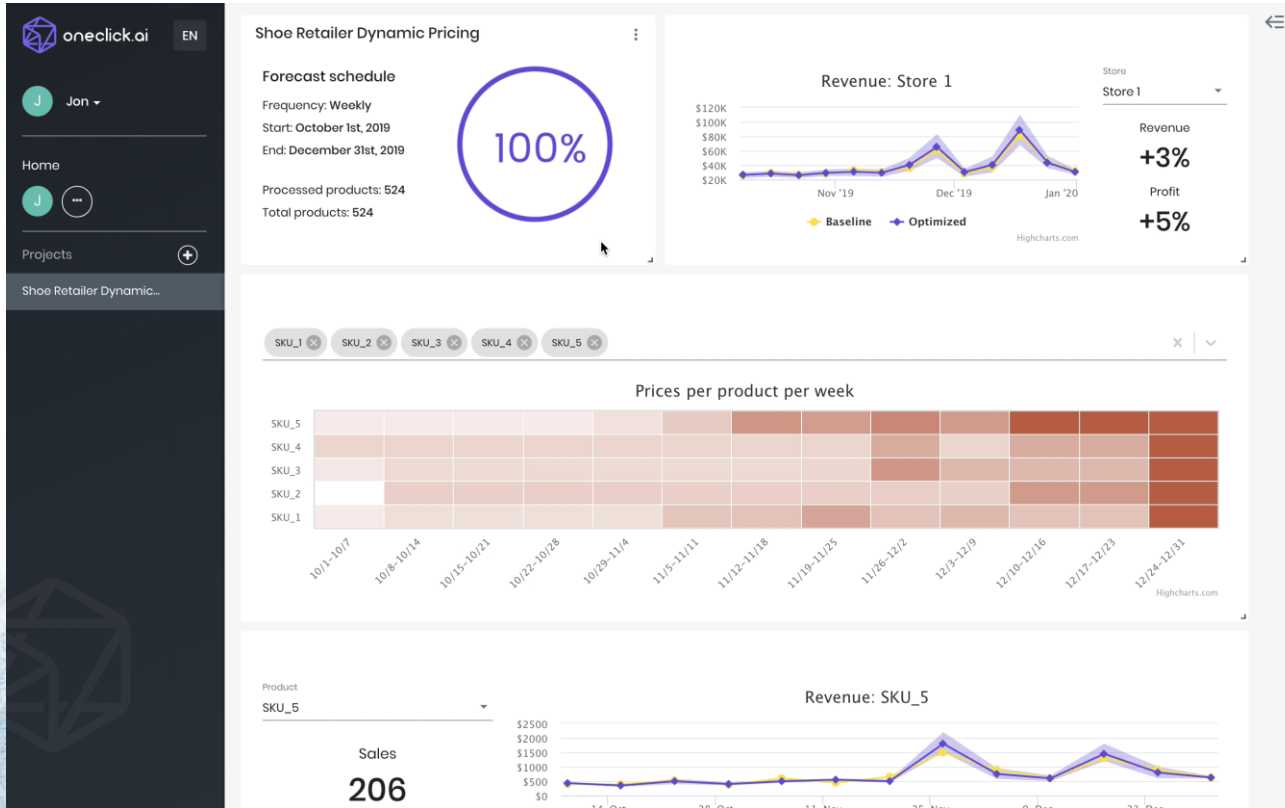
Mixed

Any combination

TECHNOLOGY



PRODUCT DEMO



USE OF INTEL® AI TECHNOLOGY

HW: Intel® Xeon® Scalable Processors

SW: Intel® Distribution for TensorFlow* 1.12 with Intel® MKL-DNN, Intel® Distribution for Python* 2.7

Data: Retail historical sales records with product attributes

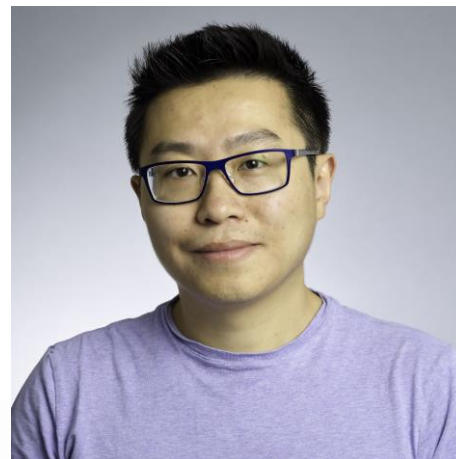
Time Series Forecasting Model: Custom hybrid deep neural network models with near 450,000 trainable parameters

Improvement: Model training and inference time

CONTACT

Yuan Shen
Founder and CEO of OneClick.ai

yuans@oneclick.ai



Visit our table with your questions, or stop by the Intel® AI Builders
matchmaking table to set up a private meeting.





AI ON INTEL

**AI BUILDERS SHOWCASE
QUEST GLOBAL INC.**

AGENDA

1

Company Overview

2

QuEST - Intel® AI Solutions

3

Use Of Intel® AI Technology

4

Financial Sector and Industrial Use case

4

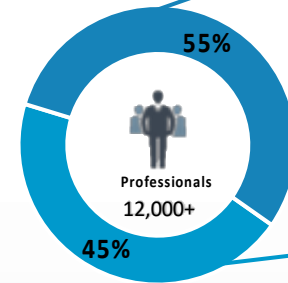
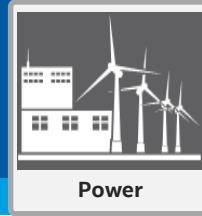
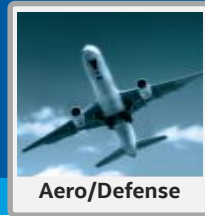
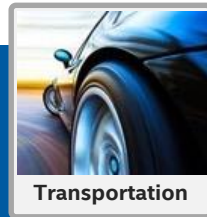
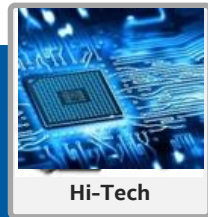
Results

5

Contact

COMPANY OVERVIEW: QUEST GLOBAL INC.

Engineering-focused solutions provider to Fortune-500 Technology companies. We deliver next-gen technologies like AI/DL, IOT, AR/VR to accelerate digital transformation for the world's leading organizations.




- **Embedded & Software**
Product Engineering
- **Enterprise Solutions**
- **Technology Solutions:**
AI/DL, AR/VR, Big Data
Analytics, IoT, Security

Engineering
(Mechanical Engineering,
Design Engineering, Process
and Manufacturing etc.)

 **1997**
Founded

 **14**
Countries

 **65+**
Locations

 **12,000+**
Employees

QUEST - INTEL® AI SOLUTIONS

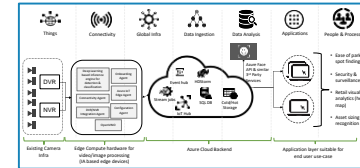


Healthcare &
Medical Imaging

Retail, Banking,
Hospitality & Education

QuEST AI Solutions in Intel® AI Builder Solution Catalogue (Intel® OpenVINO™/AI)

QuEST ThirdEye - Vision Analytics Platform



Intel AI Builders Solution Catalogue: <https://builders.intel.com/ai/solutionscatalog/qu-est-thirdeye-vision-analytics-platform-554>



QuEST Intel® AI Lab

Develop & Scale IA based AI solutions globally.

Automotive &
Transportation

Industrial &
Manufacturing

Intel® AI Builders

System Integration (SI) Partner

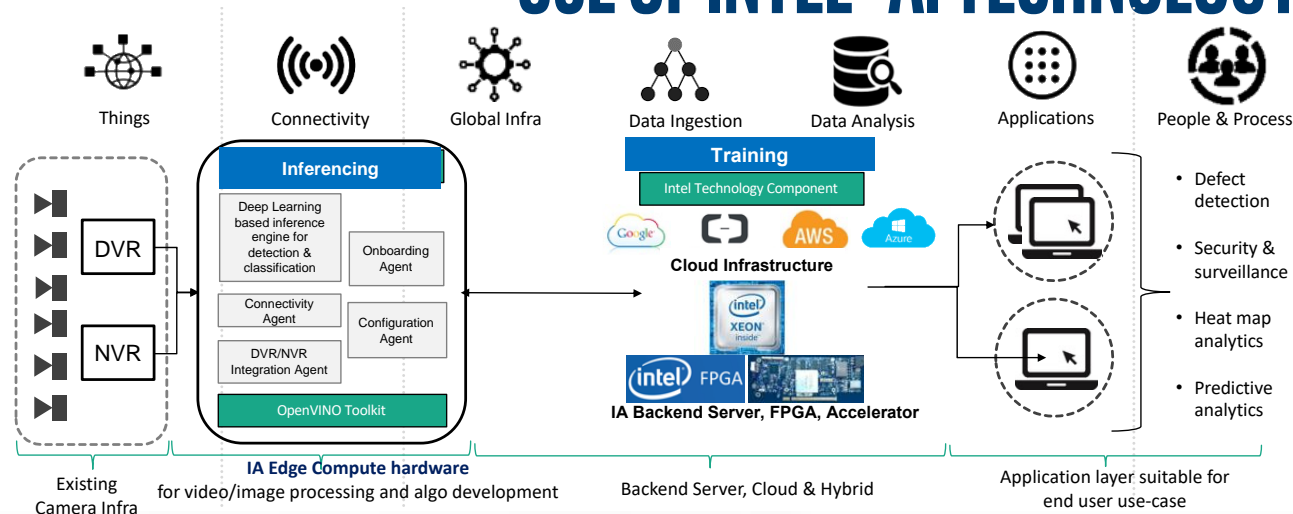
<https://builders.intel.com/ai/membership/quest>

QuEST Lung Nodule Detection (Early detection of Lung Cancer)



Intel® AI Builders Solution Catalogue: <https://builders.intel.com/ai/solutionscatalog/lung-nodule-detection-in-ct-scans-549>

USE OF INTEL® AI TECHNOLOGY



Intel® HW

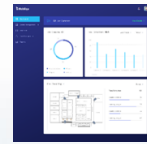
Intel® Core™ i5/ i7
Intel® Xeon® (Skylake)

Intel® SW

Intel® Distribution of OpenVINO™ Toolkit



```
{
  "cameraId": 1,
  "zoneId": 32,
  "personDetected": 1,
  "bbox": {57, 89, 250, 310}
}
```



DVR/NVR Aggregator

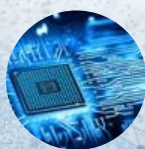
Edge Compute

On-Prem/Cloud Backend

Presentation/App Layer



Automotive



Hi-Tech



Medical Devices



Industrial



Oil & Gas



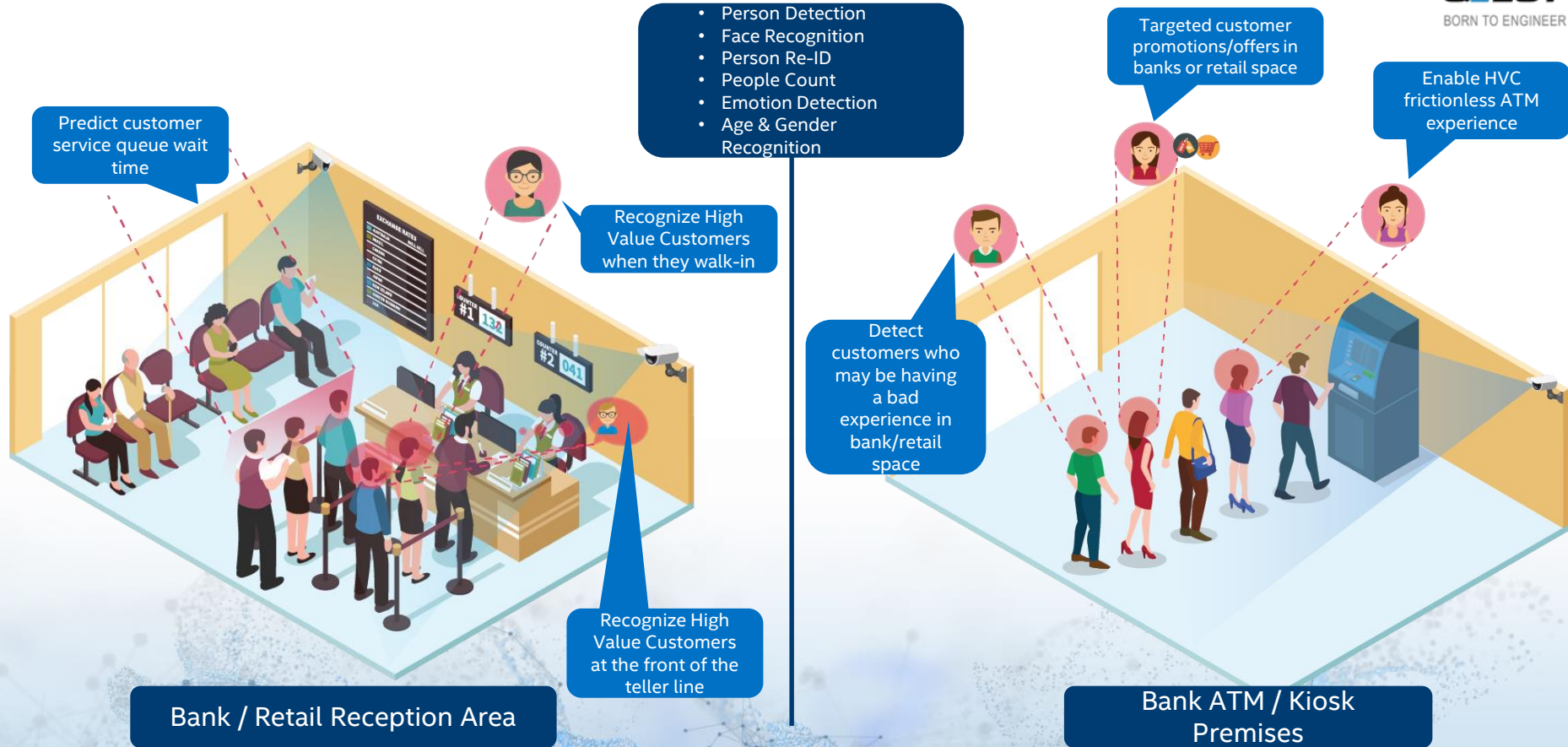
Power



Rail

Showcasing vision application in Financial & Industrial verticals

RETAIL AND BANKING SECTOR APPLICATION



RESULTS- INTEL TECHNOLOGY VALUE PROPOSITION FOR AI

Intel Technology Impact and value Proposition:

- Intel technology with Intel® Distribution of OpenVINO™ Toolkit gives the advantage of utilizing the existing CPU infrastructure to its fullest potential for creating accelerated AI inferencing applications.
- We are able to achieve this without compromising speed or quality.
- Customers are really interested in such solutions on edge devices.

CONTACT

Dr. Rubayat Mahmud
Director of Sales and Business Development
Rubayat.Mahmud@QuEST-Global.com



Visit our table with your questions, or stop by the Intel® AI Builders matchmaking table to set up a private meeting.





**AI ON
INTEL**

**AI BUILDERS SHOWCASE
GRAMENER INC**

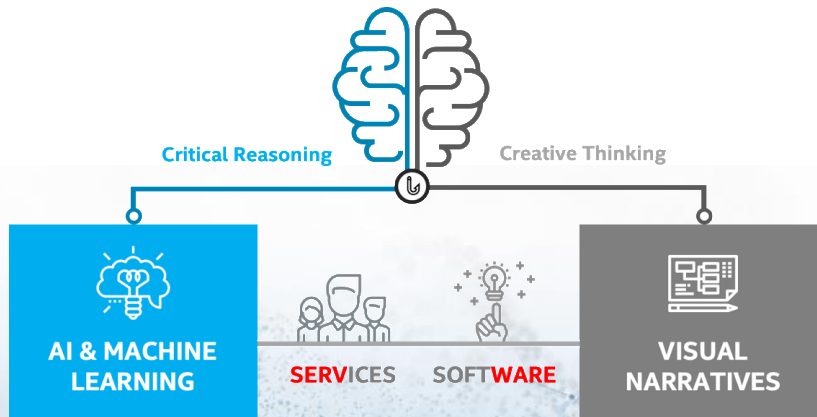
AGENDA

- Gramener overview
- Saving Antarctic Penguins with Deep Learning
- Approach and Intel® AI technology used
- Results
- Contact



GRAMENER

- Solve Business Problems through consultative data science applications
- “Insights as Stories”



- 100+ clients, 200+ employees, 5 global locations
- Won Microsoft-CNBC award for AI



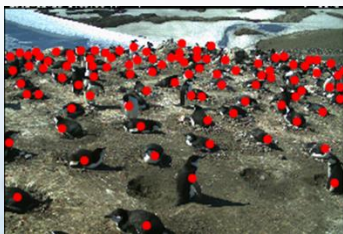
SAVING ANTARCTIC PENGUINS WITH DEEP LEARNING



- Antarctica faces major environmental threats
- Penguin populations suffer the biggest impact

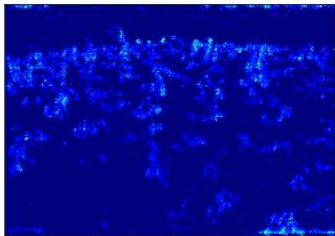


- Penguin watch is a citizen science initiative
- 100 cameras setup with 10+ years of time-lapse pictures

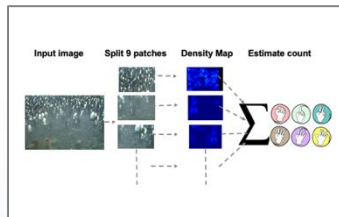


- Data labeling done by crowd-sourced annotations
- Solution built by Gramener and Microsoft AI for Earth

APPROACH & INTEL® AI TECHNOLOGY



- Challenges such as occlusion, density difference, perspective distortion and camera angles
- Crowd counting approach using density estimations



- Cascaded multi-column CNN architecture adopted
- High-level prior stage, followed by density estimation



- Intel® Xeon® processors
- Intel® Optimization for PyTorch*

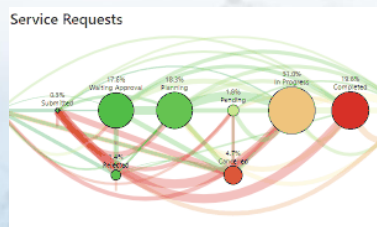
RESULTS



- Robust solution with good accuracy (Mean Abs. Error < 10)
- Intel architecture benchmarked for high performance with cost savings



- Solution architecture repurposed for drug discovery
- Applied to estimate store footfalls & count crowds



- AI in Text analytics, Image recognition, Crowd counting
- Solutions such as CX - VoC Analytics, Legal Contract Risk
- Data storytelling to solve specific business problems

CONTACT



Naveen Gattu

Co-founder and COO

naveen.gattu@gramener.com



Ganes Kesari

Co-founder and Head of Analytics

ganes.kesari@gramener.com

Visit our table with your questions, or stop by the Intel® AI Builders matchmaking table to set up a private meeting.





**AI ON
INTEL**

**AI BUILDERS SHOWCASE
WIPRO**

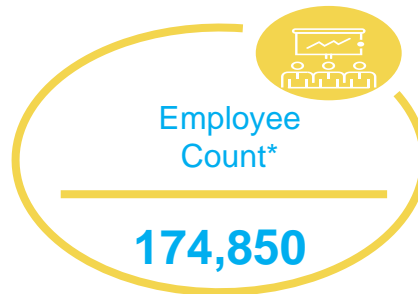
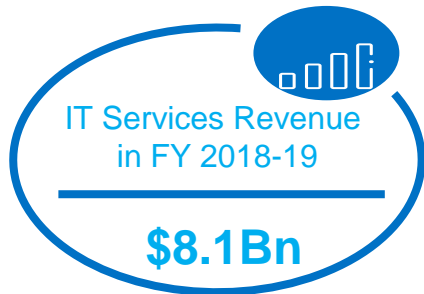
AGENDA



- Company Overview
- Business Problem Solved
- Use of Intel® AI Technology
- Results



WIPRO OVERVIEW



Wipro's AI/ML Service Offerings

AI Consulting

- Help in AI strategy
- AI CoE- Operating Model
- AI maturity assessment
- AI Technology Recommendations
- AI Readiness
- Business Case creation

AI Services

- Productionizing POCs/Pilots
- Deployment
- Technology Identification
- Build & Support AI Data Platforms
- Custom AI Models
- Co-Create Solutions

ML Ops

- Build, Maintain ML Data Pipeline
- Data Ingestion, Data Exploration, Data Wrangling
- Machine Learning Lifecycle Mgmt
- AI Validation as a Service

AI Apps

- Industry Specific Solutions built on AI services like Vision, Speech, Video Intelligence, Chatbots, Language Understanding
- Custom Insights, Reports

*Figures based on Q1 2019-20 for Global IT Services business

INTEL® AI BUILDERS: BUSINESS PROBLEMS SOLVED



Pipe Sleuth



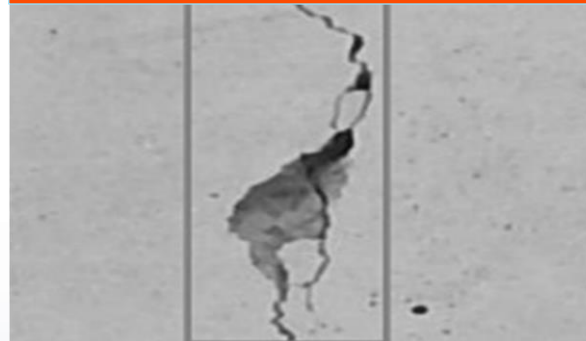
Pipeline Assessment

Medical Imaging



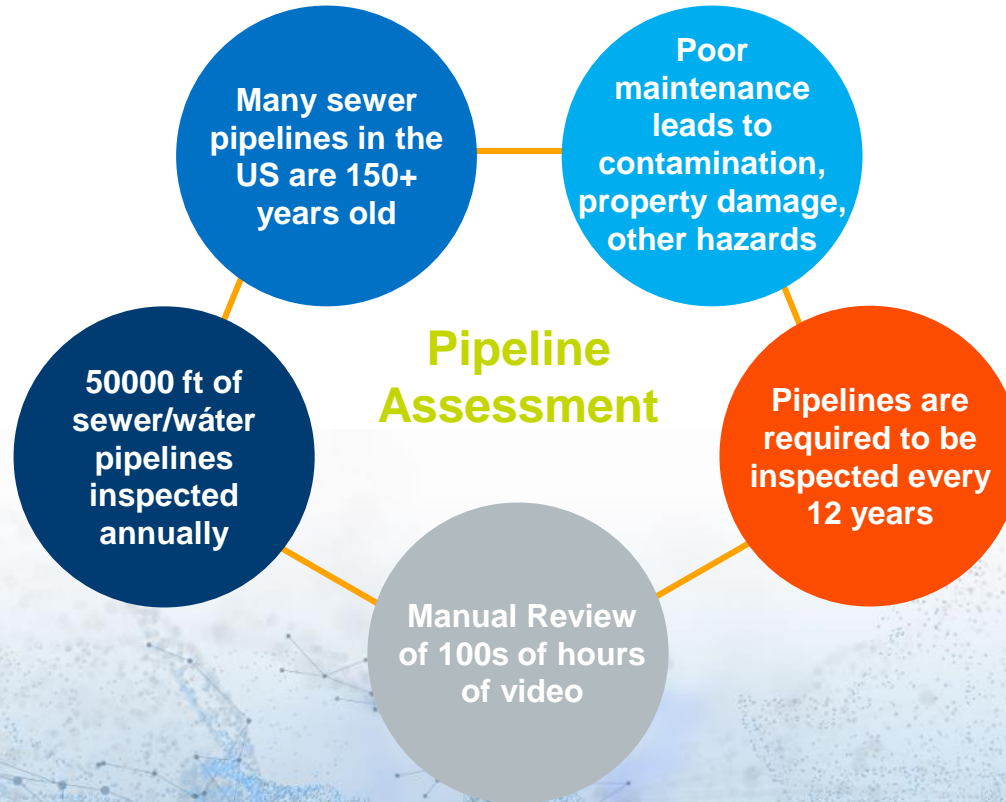
Lung Segmentation, Lung Disease Diagnosis

Surface Crack Detection



External Structural Cracks Inspection

PIPE SLEUTH - PROBLEM CONTEXT



PIPELINE VIDEO CAPTURE & ASSESSMENT PROCESS



Pipeline Video Capture

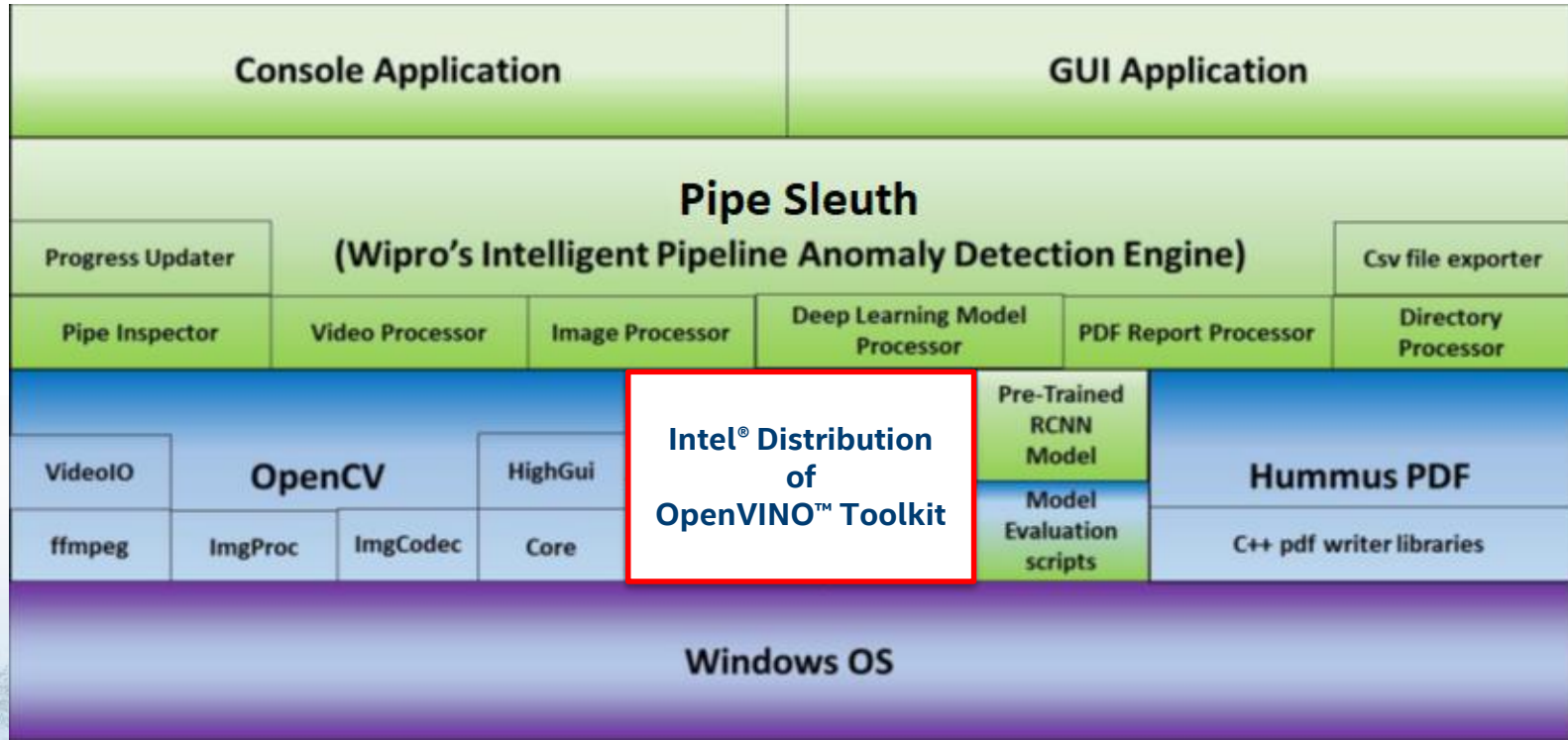


Pipeline Assessment

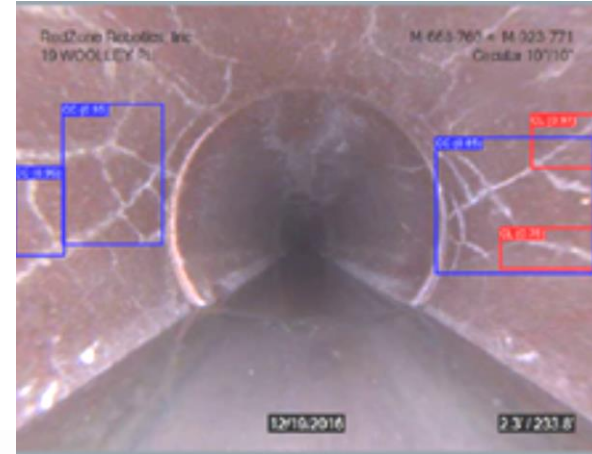
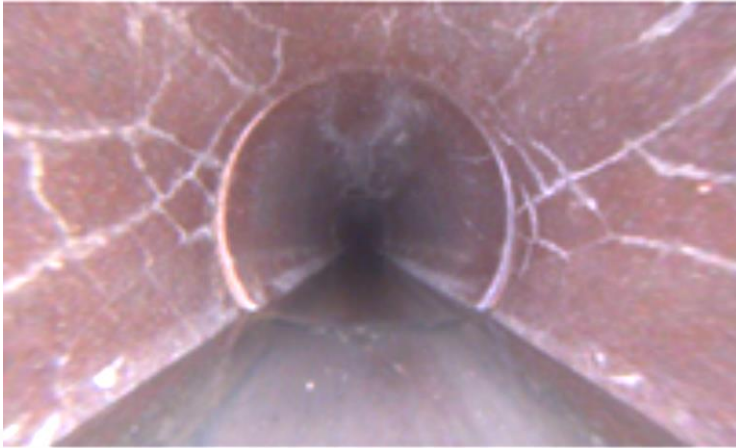


- Expert time: **1 hour** -> **10-mins** (for 1-hour of video)
- From Tedious -> **Not Tedious**
- **Scalable, Reliable**
- **ROI: 3.5-4x over 5 years**

PIPE SLEUTH SOLUTION



PIPELINE ANOMALIES



~20 Defect Types

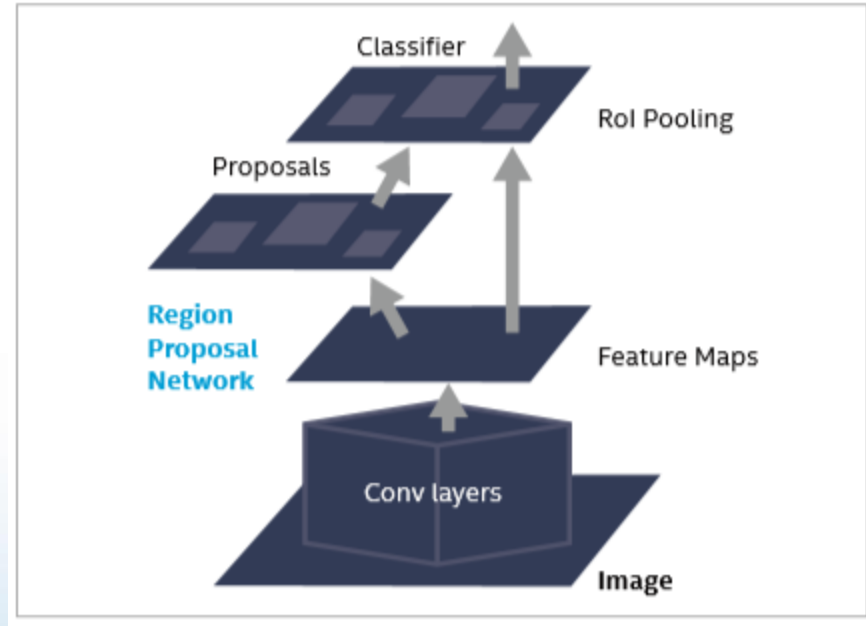
- Cracks (Longitudinal, Circumferential, Hinge, Spiral, Multiple)
- Deposits (Attached – Encrustation, Grease, Others; Settled – Others)
- Roots (Ball, Medium, Fine)
- Collapse
- Others

DEEP LEARNING MODEL

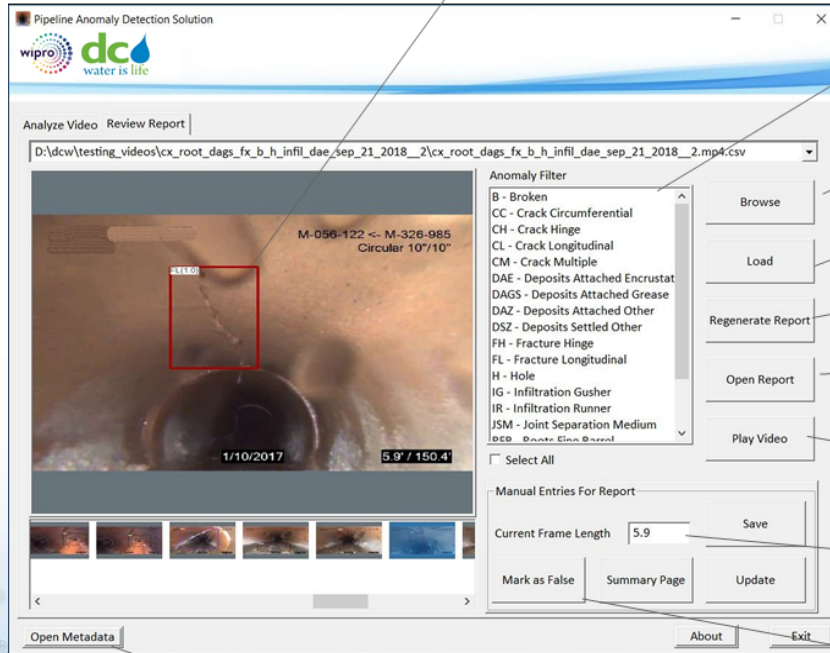
ResNet-101 (Feature Extractor)

layer name	output size	101-layer
conv1	112×112	7×7, 64, stride 2
conv2_x	56×56	3×3 max pool, stride 2 $\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
conv4_x	14×14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$
conv5_x	7×7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax

Faster R-CNN (Object Detector)



APPLICATION



Anomaly location and type are identified with confidence score

Lists all anomaly types found. User can select one or more types to review only those

User can select pre-evaluated video to review

Load anomaly images from selected video

Regenerate report with user updates

Open the original report and its folder location

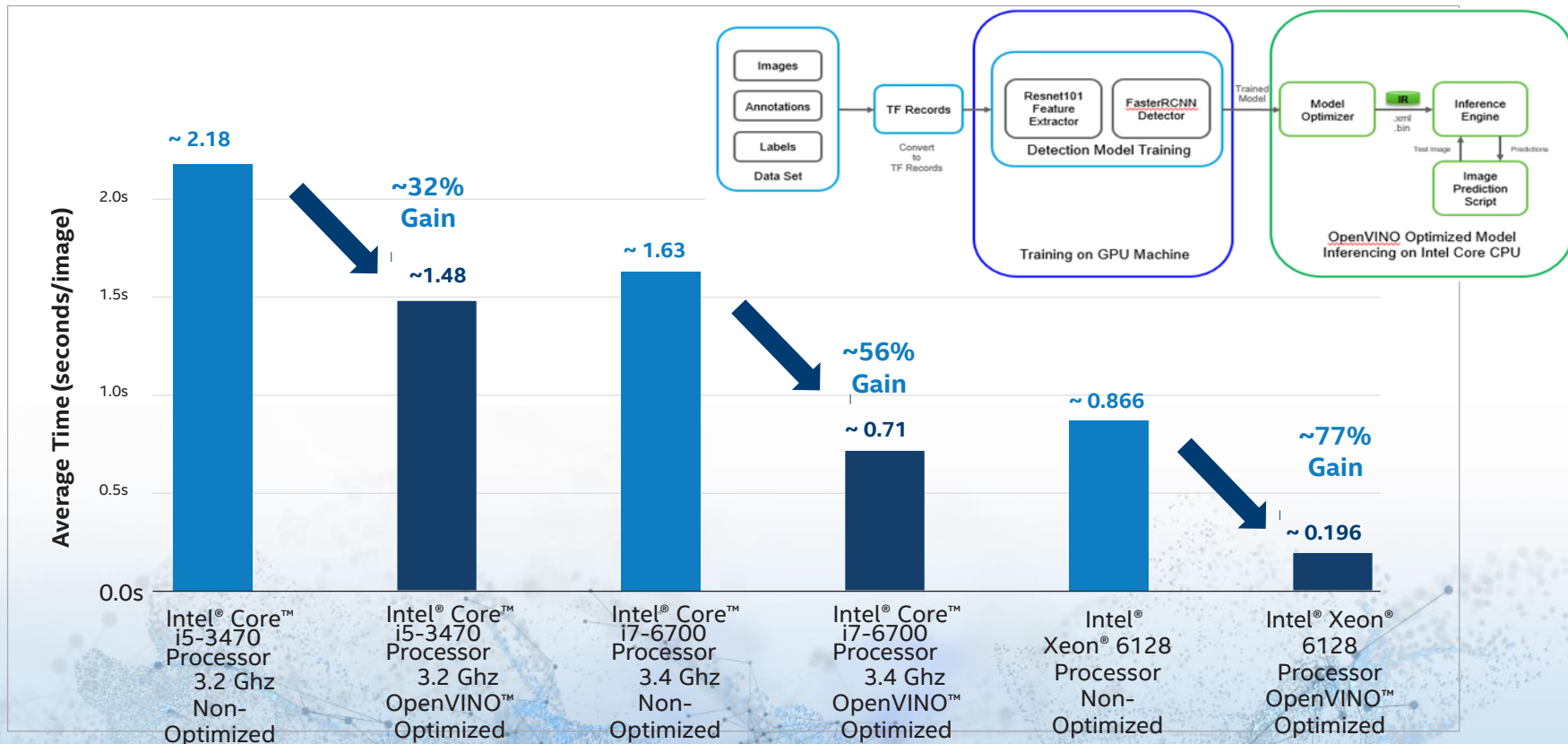
Launches video replay application

Extracted distance information from the frame. User can also update the same

User can mark individual frames as false and exclude from report

Opens the excel with the accumulated defect details for each video assessed

OPTIMIZATION ON INTEL® - PERFORMANCE



CONFIGURATION DETAILS

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

Configuration: Intel® Xeon® Platinum 8153 CPU @ 2GHz with Intel® Distribution of OpenVINO™ toolkit. Testing done by Wipro Q3 2019 Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

Performance results are based on testing as of dates shown in configuration and may not reflect all publicly available security updates. No product can be absolutely secure. See configuration disclosure for details.

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. 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>

CONTACT



Sundar Varadarajan
Consulting Partner – AI & ML
sundar.varadarajan@wipro.com



Sunil Baliga
Director of Sales
sunil.baliga@wipro.com





**AI ON
INTEL**

**AI BUILDERS SHOWCASE
INSTADEEP**

AGENDA

- Company overview
- Business Problem
- Use of Intel® AI technology
- Results
- Contact



FOUNDED

2015

TEAM

76

R&D AI Engineers,
Software,
Hardware,
Visualization, ...

OFFICES

**LONDON, TUNIS,
PARIS, NAIROBI,
LAGOS**

WHAT WE DO

- Solve a wide range of prediction, classification and optimization problems for our global industrial partners
- Allow clients to unlock data insights, realize value, and increase efficiency and speed across their organization

OUR APPROACH

1. RESEARCH

2. ENGINEERING

3. DEPLOYMENT, VISUALIZATION

BUSINESS PROBLEM SOLVED

Problem

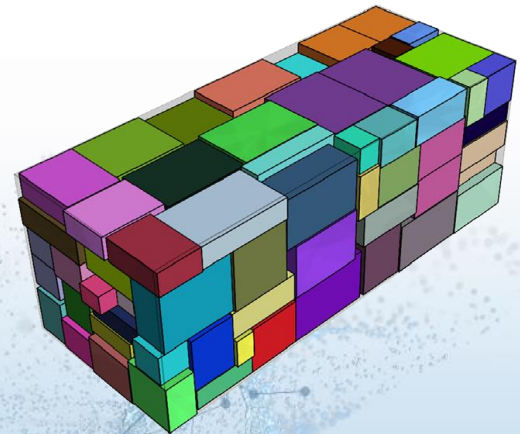
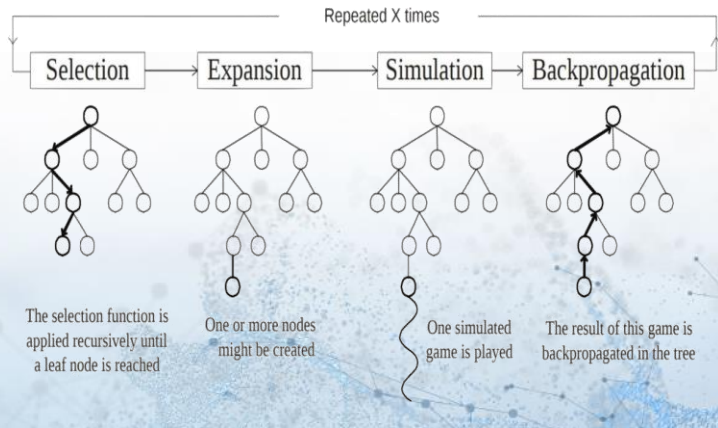
Packing a set of items into fixed-size bins while minimising a cost function e.g. number of bins required, surface of the packing

Methodology

- Formulation as a Markov Decision Process
- AlphaZero-like Algorithm
 - Policy-Value Neural Network
 - Planning Algorithm
 - Adversarial Learning Process

Results

- Outperform a sophisticated heuristic, i.e. lego
- Approach applicable to many NP-Hard problems



USE OF INTEL® AI TECHNOLOGY

HW: Second Generation Intel® Xeon® Scalable Processors

Contrary to traditional deep learning, reinforcement learning relies heavily on general purpose computing both during training and inference. The workload can be highly parallelized and the training time is highly correlated to the number of cores used.



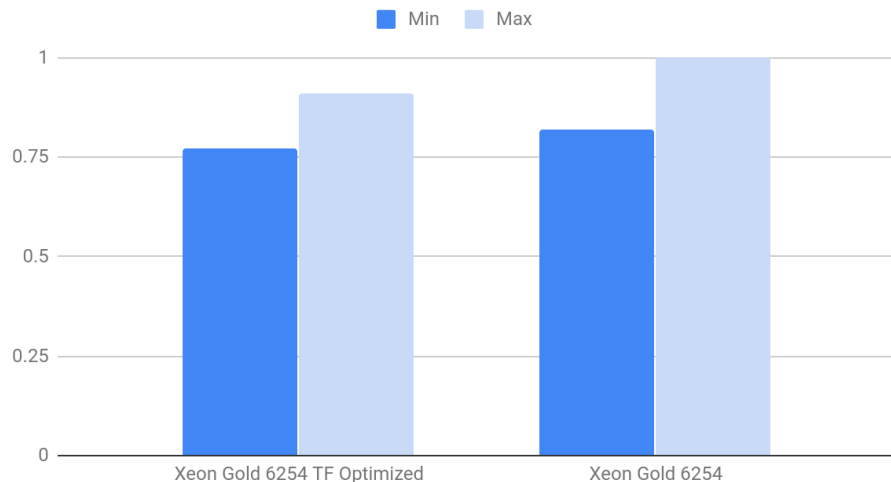
RESULTS

The benchmark consists in measuring the inference time which represents the major bottleneck in the **learning speed**.

We measured the average inference time from **10K** inferences initiated from a **Golang** based agent dispatching the requests on a thread pool. Using **2nd Gen Intel® Xeon® Scalable Processors** we see **~10%** improvement in inference performance **Intel® Optimization for TensorFlow***

Demo

Xeon Gold 6254 TF Optimized vs Non Optimized



CONFIGURATION SPECIFICATIONS

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

Configuration: Intel® Xeon® Gold 6254 @ 3.10GHz with Intel optimized TensorFlow. Testing done by InstaDeep August 2019

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

Performance results are based on testing as of dates shown in configuration and may not reflect all publicly available security updates. No product can be absolutely secure. See configuration disclosure for details.

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. 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>

CONTACT

Amine Kerkeni
Head Of Engineering
ak@instadeep.com



Visit our table with your questions, or stop by the Intel® AI Builders matchmaking table to set up a private meeting.





**AI ON
INTEL**

**AI BUILDERS SHOWCASE
JOHN SNOW LABS**

AGENDA

- Accelerating AI in Healthcare
- Getting AI from concept to production in a high-compliance industry
- The Turnkey Enterprise Platform for Healthcare AI
- Optimized for Intel® AI Software & Hardware
- Customer Success
- Contact

ACCELERATING AI IN HEALTHCARE

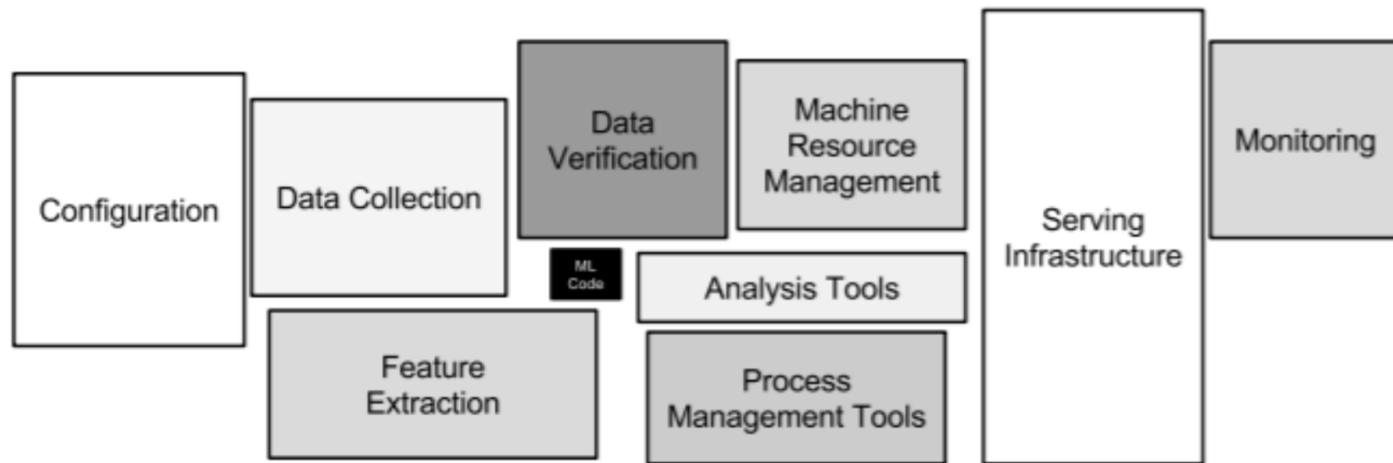
CIO ARTIFICIAL
INTELLIGENCE
Review SOLUTION PROVIDER
OF THE YEAR - 2018



Best BI or Analytics Solution

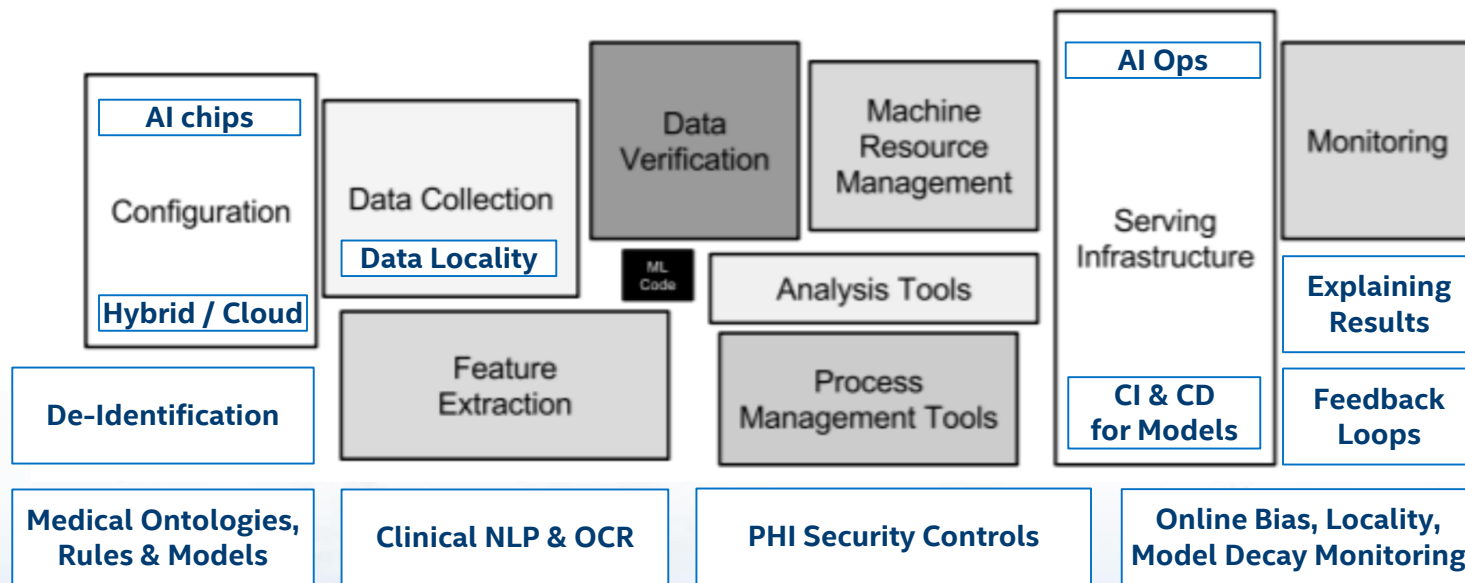


GETTING AI FROM CONCEPT TO PRODUCTION



“Hidden Technical Debt in Machine Learning Systems”, Google, NIPS 2015

GETTING AI FROM CONCEPT TO PRODUCTION



Building Real-World AI Systems in Healthcare & Life Science, 2019

THE TURNKEY ENTERPRISE PLATFORM FOR HEALTHCARE AI



Data Engineer

Data Integration

Self-serve data ingestion & wrangling



Data Analyst

Data Exploration

Interactive data analysis without coding



Data Scientist

Data Science

Interactive notebooks in Python & R



App Developer

Model Deployment

Scalable & Secure model API's

Dataflows

Batch & streaming data ingestion

Connectors

Input & output to standard protocols

Curated Medical Data

Current, Cleaned & Enriched

Data Quality

Monitor feed health in real time

Governance

Search metadata, lineage & stats

Search

Full-text, faceted & geospatial search

Visualization

Real-time, drag & drop dashboards

Time Series

Interactive time series analysis

Metadata

Dataset & schema search

SQL

Live SQL editor & dashboards

Notebooks

JupyterLab and JupyterHub

Machine Learning

Certified ML & data mining libraries

Deep Learning

Train & debug on CPU's and GPU's

Natural Language

Spark NLP with TensorFlow

Vision & OCR

Pre-trained deep learning models

Model Server

Deploy models as micro-services

CI & CD for Models

Auto-test & deploy machine learning

Model Repository

Reusable & versioned model store

Pipelines

Track experiments, jobs, and runs

API Gateway

Security, logging & rate limiting

Hub

Portal & Single Sign on

Orchestration

Hardened Kubernetes

Operations



DataOps

Monitoring

Pre-configured agents & alerts

Identity

Authentication, Authorization, Audit

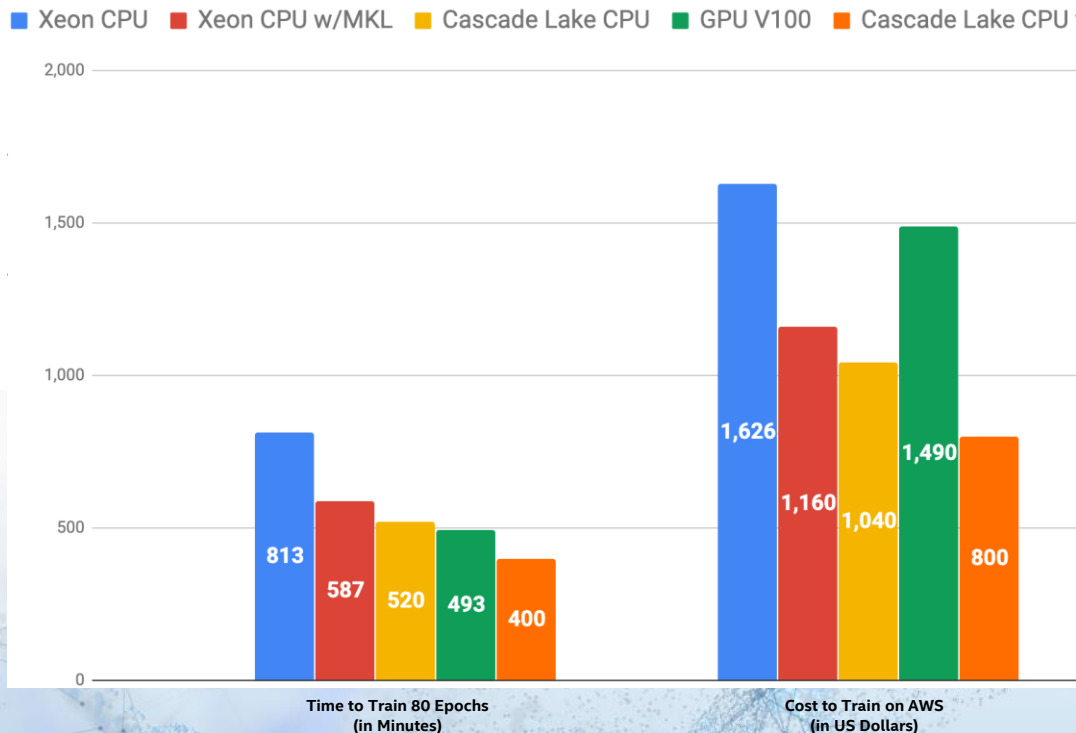
USE OF INTEL® AI TECHNOLOGY

Goal: Get the most from your Intel hardware, out-of-the-box

Software: Intel® MKL, MKL-DL, Intel® Optimization for TensorFlow*

Hardware: 2nd Gen Intel® Xeon® Scalable Processors

Results: Training Deep-Learning NLP on Intel® Xeon® Scalable Processors 2nd gen is **19% faster** and **46% cheaper** than GPU-optimized. Training on Tesla v100. Tested on AWS.



CONFIGURATION SPECIFICATIONS

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

Configuration: Intel® Xeon® Platinum 8175M @ 2.5GHz with Intel optimized TensorFlow*. Testing done by John Snow Labs September 2019

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

Performance results are based on testing as of dates shown in configuration and may not reflect all publicly available security updates. No product can be absolutely secure. See configuration disclosure for details.

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. 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>

CUSTOMER SUCCESS



KAISER PERMANENTE®

**Improving Patient Flow Forecasting
at Kaiser Permanente**



usermind

**Usermind built its data science platform
from scratch to production in 3 months**



**How Roche Automates Knowledge
Extraction from Pathology Reports**



the SHM
foundation

**Using Advances Analytics to Improve
Mental Health for HIV-Positive Adolescents**

SelectData™

**How SelectData Uses AI to Better
Understand Home Health Patients**

DEEP 6 AI

**Feature Engineering with Spark NLP
to Accelerate Clinical Trial Recruiting**

CONTACT

Dr. David Talby
CTO

david@johnsnowlabs.com



Visit our table with your questions or stop by the Intel® AI Builders
matchmaking table to set up a private meeting.





**AI ON
INTEL**

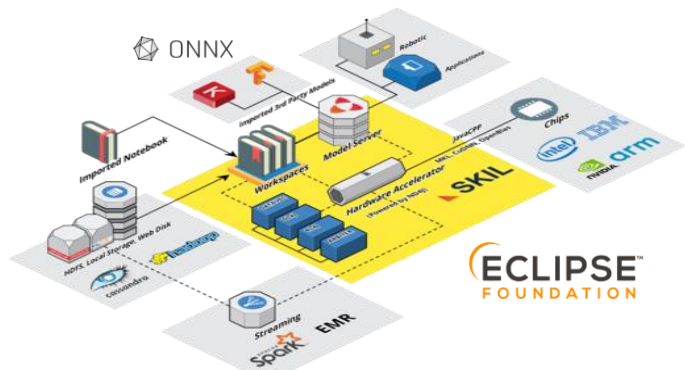
**AI BUILDERS SHOWCASE
SKYMIND**

AGENDA

- Skymind overview
- Business problem solved by prioritized vertical
- Use of Intel® AI technology
- Results
- Contact



SKYMIND



ECLIPSE
FOUNDATION

Apache Zeppelin



Enabling the AI-Driven Enterprise

Accelerating Experimentation to Production Lifecycle

Founded

2014

Funding

\$17.5M Funding, Series A Startup

Open Source

DeepLearning4J Suite > 200k downloads/mo

Enterprise

Skymind Intelligence Layer (SKIL) AI Infrastructure

Services

Professional Services, Customer POCs

Investors



BUSINESS PROBLEM SOLVED

Vision

Our mission is to bridge the gap between Deep Learning Workflows and delivering scalable production-ready AI Deployment

Modeling
Algorithms



Infrastructure
Compute



Device
Performance



AI
Platforms



Vendor
Partnerships



Business Engagements

Many of our large enterprise clients engage to advance their machine intelligence, accelerate & scale their operating cycle initiatives, delivering value faster along their AI Maturity Journey.



servicenow

relative sampling of select Industries

Business Challenges

Many **State-of-the-Art (SOTA)** models require high memory, distributed compute power, and a proper strategy to minimize latency to gain insights from their data.



Computer Vision

Image Classification
High Dimensionality

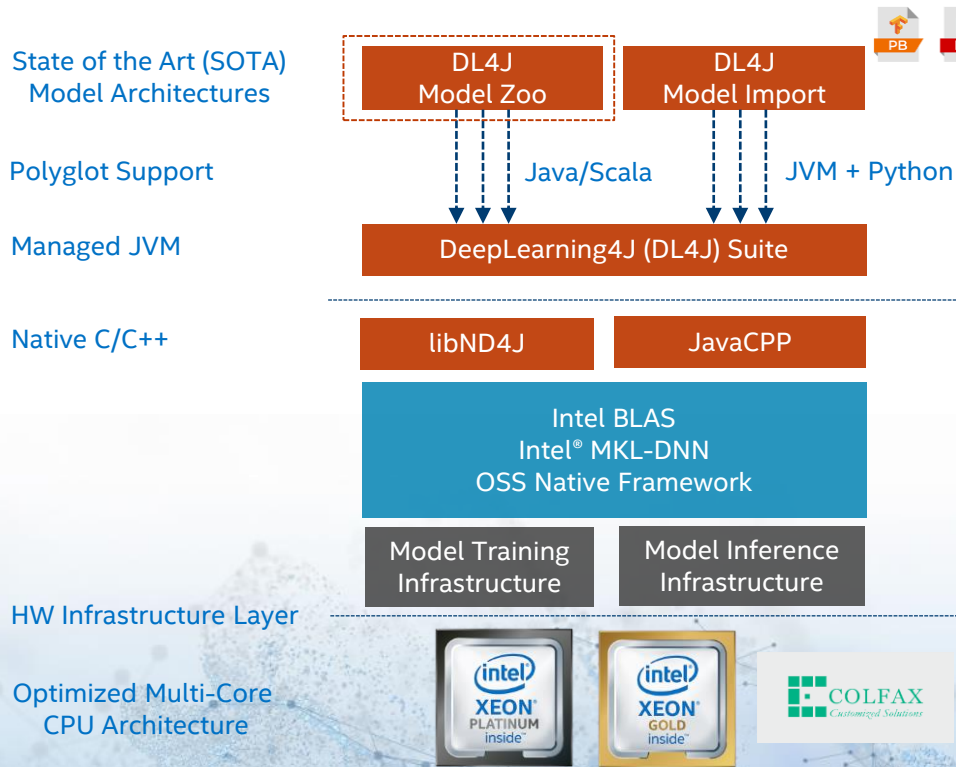


NLP Text

Agents and Bots
High Cardinality & Context

USE OF INTEL® AI TECHNOLOGY

■ Skymind
 ■ Intel OSS
 ■ Intel Hardware
 ■ Resource Management



Native JVM Model Zoo VGG16 CNN Architecture
2-Class 4600+ Image Dataset, Batch Size=128

DL4J Suite Beta Release 1.0.0-BETA4
CONV Layers/Operations support Intel® Math Kernel Library (Intel® MKL)

Adaptive backend switch across chip architectures

MKL (Math Kernel Library) Performance Benchmarking
Native DL4J Models on Intel® MKL-DNN

2nd Generation Intel® Xeon® Scalable processors
1 x 6248 (Intel® Xeon® Gold) Multi-Core CPU @2.5 GHz
2 x 8256 (Intel® Xeon® Platinum) Multi-Core CPU @3.8 GHz

RESULTS

Our partnership with Intel® has enabled **Intel® MKL-DNN** as the default execution on **2nd Gen Intel® Xeon® Scalable** processors, replacing OpenBLAS* per optimized Linear Algebra

Facilitating Faster Time to Market, Productivity, and Operational Costs

6X	42 min	Training Time	<i>Possibly Reduced Operational Cost</i>
5.9X	6 records/sec	Throughput (Record)	<i>Leading to Reduced Inference Latency Time</i>

Reference the following for upcoming published DL4J Benchmarks
software.intel.com/en-us/frameworks
github.com/IntelAI/models/benchmarks

CONFIGURATION SPECIFICATIONS

SkyminD DL4J Benchmark on Custom Code using VGG16 - Throughput Performance on Intel® Xeon® Platinum 8256 Processor:

NEW: Tested by Intel as of 09/03/2019. 2 socket Intel® Xeon® Platinum 8256 Processor, 4 cores per socket, Ucode 0x500001c, HT On, Turbo On, OS Ubuntu 18.04.2 LTS, Kernel 4.15.0-48-generic, Total Memory 384 GB (12 slots/ 32GB/ 2666 MTs), Deep Learning Framework: DL4J (from Intel_benchmark branch of <https://github.com/deeplearning4j/dl4j-benchmark>), custom train data of 2 classes with a total 4608 samples from Distracted Driver Dataset from kaggle with batch size of 128 for VGG16. Trained with MKLDNN & KMP_BLOCKTIME as zero & KMP_AFFINITY as fine, balanced

BASELINE: Tested by Intel as of 09/03/2019. 2 socket Intel® Xeon® Platinum 8256 Processor, 4 cores per socket, Ucode 0x500001c, HT On, Turbo On, OS Ubuntu 18.04.2 LTS, Kernel 4.15.0-48-generic, Total Memory 384 GB (12 slots/ 32GB/ 2666 MTs), Deep Learning Framework: DL4J (from Intel_benchmark branch of <https://github.com/deeplearning4j/dl4j-benchmark>), custom train data of 2 classes with a total 4608 samples from Distracted Driver Dataset from kaggle with batch size of 128 for VGG16. Trained without MKLDNN with maxbyte as 376GB, maxphysicalbytes as 373GB and Xmx as 6GB

CONTACT

Ari Kamlani

Principal Deep Learning Technologist

ari@skymind.io



Visit our table with your questions, or stop by the Intel® AI Builders matchmaking table to set up a private meeting.

Learn more: builders.intel.com/ai/membership/skymind





**AI ON
INTEL**

**AI BUILDERS SHOWCASE
DIGITATE**



AGENDA

:digitate

- Digitate overview
- Business problem solved by prioritized vertical
- Use of Intel® AI technology
- Results
- Contact

Launched in June 2015

100+ Customers globally

*Managing millions of
Applications, Batch, SAP,
& Infra components*

600+ Employees worldwide

70+ Patents globally



Self-Awareness
Category



Best Overall AI Company
of the Year



Tata Steel for
Operational
Efficiency



Most Innovative Tech
Company of the Year



Software Company
of the Year



Fastest Growing Software
Company of the Year

:digitate

The Business
Transformation &
Operational
Excellence
Awards 2019
Finalists: Loblaw
& Credit Suisse



Software Defined
Infrastructure

➤ Digital transformation for an enterprise

- Powering IT operations with ignio, AI/ML based Cognitive Platform
 - Data-driven context aware insights, predictions and actions
 - Superior experience with integration of context-rich collaboration channels & varied types of data

➤ Challenges

- Scale: 100K+ technology components
- Complexity: Deep and dynamic interdependencies, diversity of technology vendors & versions
- Accelerating rate of change: Technology footprint, everything on-demand,
Expectations: SLAs → instant gratification, cyber security threats and regulatory requirements

Case Study: Face Recognition for AIOps

RESULTS

Public



from



UP TO 5.8X INCREASE

in Inference workload performance over baseline using Intel® Distribution of OpenVINO™ toolkit



Partner: Digitate is a pioneer in bringing artificial intelligence to IT and Business Automation. Digitate's AI-enabled software, ignio™, helps organizations run IT infrastructures far more efficiently, improves customer experience and increases the agility and stability of IT operations

Challenge: Reduce image processing/ interpreting time for face detection, in a deep learning workload for workflow automation. Additionally, there was a need to adhere to UX design guidelines for facial login response time.

Solution: The solution took advantage of efficient multi-core processing on Intel Xeon® Scalable processors along with Facenet* and witnessed significant improvement in Inference performance, thereby bringing down the time to detect the face for auto-login. This resulted in quicker login times by performing image inferencing in milliseconds now (vs. seconds earlier), while adhering to UX requirements.

Intel® OpenVINO™ improves FaceNet inference times by up to 1051X, resulting in an end-to-end 5.8X improvement in Digitate's auto-login process, with Intel® Xeon® Platinum 8256 processors.

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

Configuration: Intel® Xeon® Platinum 8256CPU @ 3.80GHz/ 4 cores per socket / 2 sockets / 384GB / Tensorflow* 1.13.1/ OpenVINO™ 2019 R1.1.133

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>. Performance results are based on testing as of August 2018 and may not reflect all publicly available security updates.

See configuration disclosure for details. No product can be absolutely secure.

Whitepaper/Solution
Brief: WIP



126



RESULTS

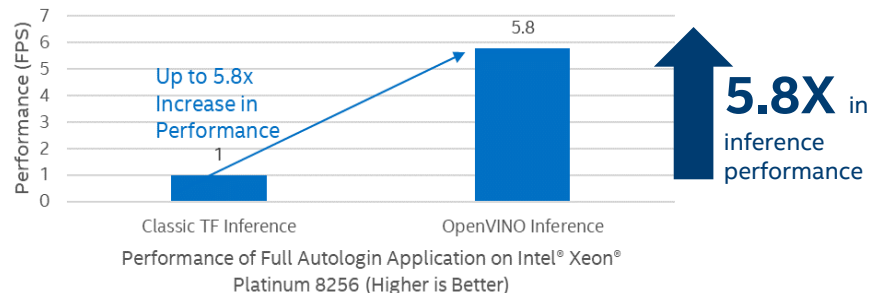


➤ Enhanced Cognitive Automation Performance with Intel® technology

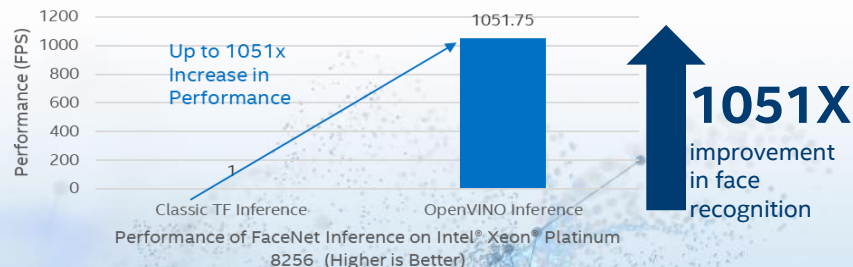
- Use of Intel® Distribution of OpenVINO™ toolkit delivered significant, tangible performance improvements compared to Tensorflow
- Both the deep learning-based face recognition workload (FaceNet) and ignio's overall auto-login workload benefited from Intel hardware and software

ignio, along with Intel® AI hardware and software, delivers the capabilities and performance required for future-ready, agile and proactive IT operations and infrastructure.

Relative Digitate Autologin Performance (FPS)



Relative FaceNet Inference Performance (FPS)





Hardware: 2nd Generation Intel® Xeon® Scalable processor

Choosing the right hardware helps drive IT efficiency, agility and gain a competitive edge

An infrastructure to seamlessly scale and support demanding AI Ops workloads

Robust & reliable platforms for the data-fueled enterprise



Software: Intel® Distribution of OpenVINO™ toolkit

Incredible efficiency and hardware acceleration achieved through Intel® Distribution of OpenVINO™ toolkit

The toolkit supports heterogeneous execution across various hardware through a common API

Includes easy to use libraries and pre-optimized kernels

CONFIGURATION SPECIFICATIONS



Digitate Custom Face Detector:

NEW: Tested by Intel as of 05/19/2019. 2 socket Intel® Xeon® Platinum 8256 Processor, 4 cores per socket, HT On, Turbo On, Total Memory 374 GB (12 slots/ 32GB/ 2666 MHz), BIOS: SE5C620.86B.02.01.0008.031920191559, Deep Learning Framework: Intel® OpenVINO® 1.1.133 using Intel® MKL 2019.3. FaceNet topology customized by Digitate*, custom test data, tested using batches of 1 (optimized for latency).

BASELINE: Tested by Intel as of 05/19/2019. 2 socket Intel® Xeon® Platinum 8256 Processor, 4 cores per socket, HT On, Turbo On, Total Memory 374 GB (12 slots/ 32GB/ 2666 MHz), BIOS: SE5C620.86B.02.01.0008.031920191559, Deep Learning Framework: TensorFlow 1.13.1 (Anaconda repo 'tensorflow') using Intel® MKL 2019.3. FaceNet topology customized by Digitate*, custom test data, tested using batches of 1 (optimized for latency).

Name: Anjali Gajendragadkar

Title: ignio™ Product Manager

Email address: anjali.sg@digitate.com



Visit our table with your questions and to see a demo, or stop by the Intel® AI Builders matchmaking table to set up a private meeting.

We look forward to meeting you!

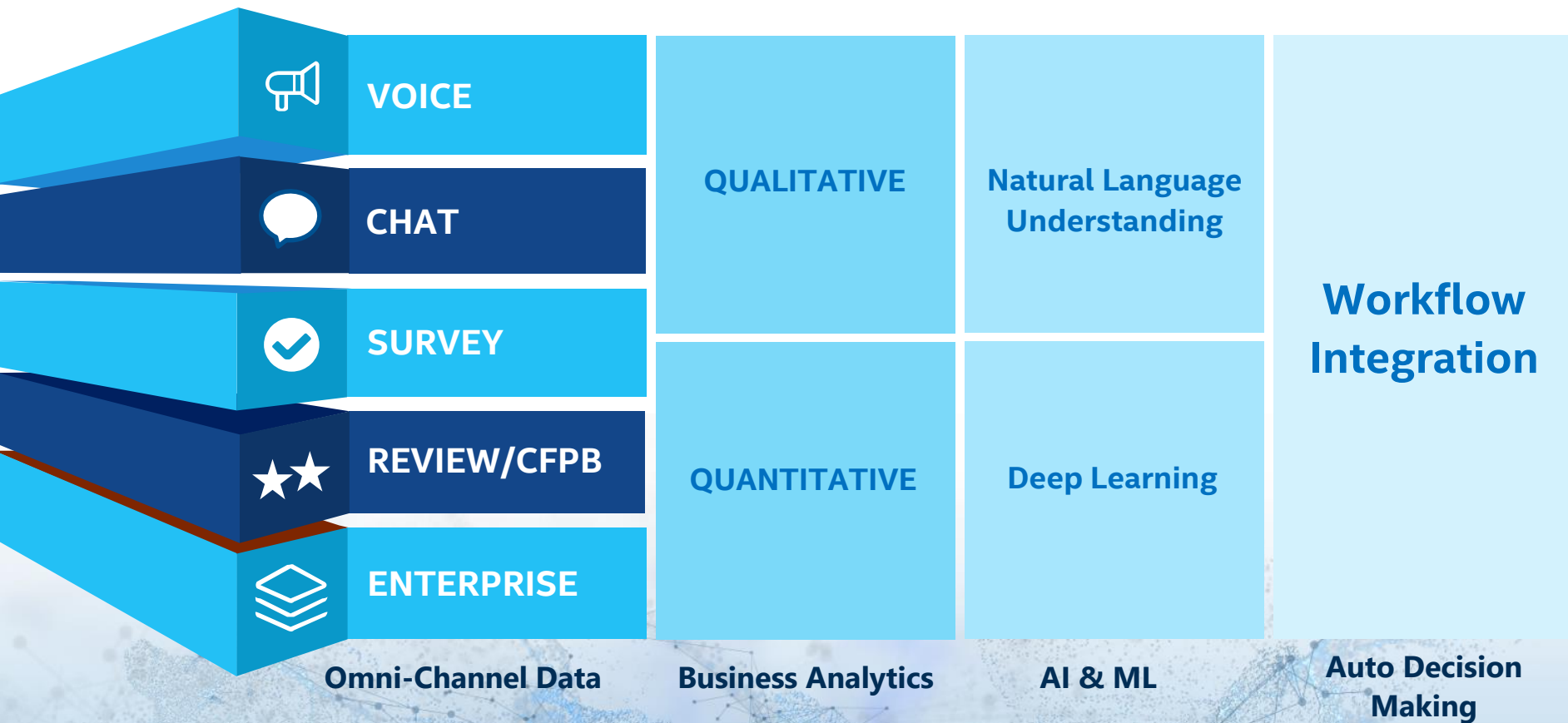




**AI ON
INTEL**

**AI BUILDERS SHOWCASE
STRATIFYD**

END-TO-END CUSTOMER ANALYTICS PLATFORM POWERED BY AI



USE OF INTEL AI TECHNOLOGY



ARTIFICIAL INTELLIGENCE

AI Solutions Catalog
(Public & Internal)



DEEP LEARNING DEPLOYMENT

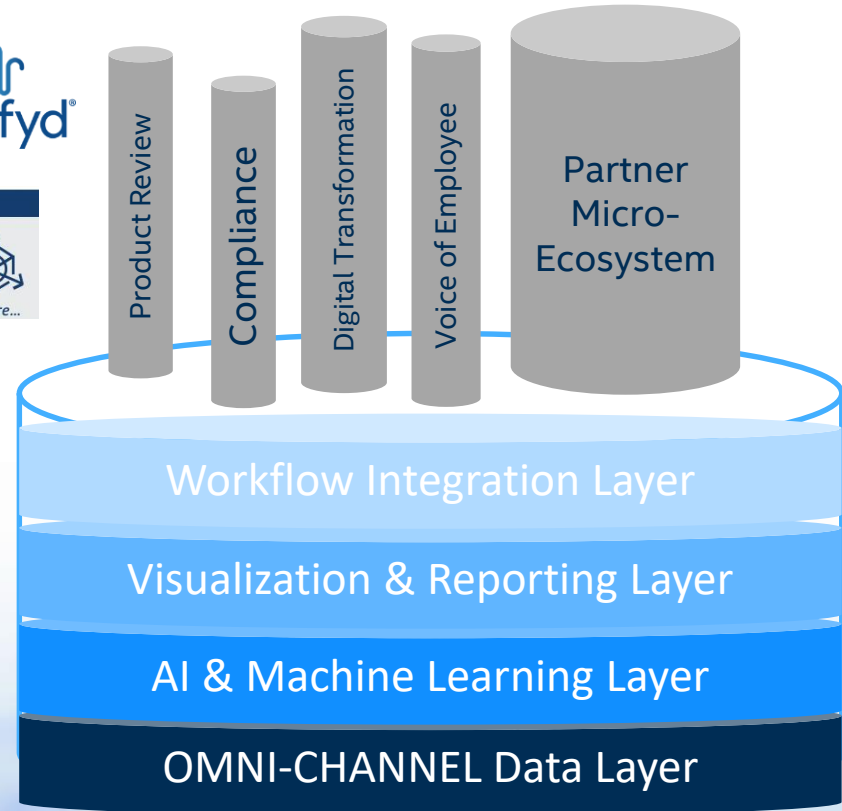
Intel® Distribution of OpenVINO™ Toolkit

Open Visual Inference & Neural Network
Optimization toolkit for inference
deployment on CPU/GPU/FPGA for TF,
Caffe* & MXNet*

Intel® Movidius™ SDK

Optimized inference deployment
on Intel VPUs for
TensorFlow* & Caffe*

AI FOUNDATION



RESULT IN STRATIFYD AI MARKETPLACE



- **Voice of the Customer**
- **Voice of the Employee**
- **App Store Feedback**
- **Complaint vs. Feedback**
- **Compliance**
- **Revenue**
- **Operational Efficiencies**
- **Many more...**

mastercard.

intuit.



Prudential

Microsoft

GlaxoSmithKline

Lilly

ally



charles SCHWAB

Gartner

MASCO



citi

Capital One

Etsy

Lenovo

Mercedes-Benz

Kimberly-Clark



PHILIP MORRIS
INTERNATIONAL

LIVEPERSON

arvato

BERTELSMANN

CONTACT



Derek Wang
Founder and CEO

Derek.Wang@stratifyd.com



Visit our table with your questions, or stop by the Intel® AI Builders
matchmaking table to set up a private meeting.





**AI ON
INTEL**

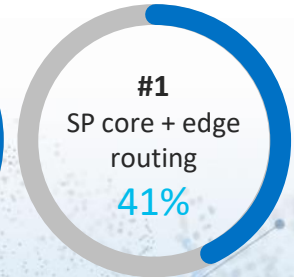
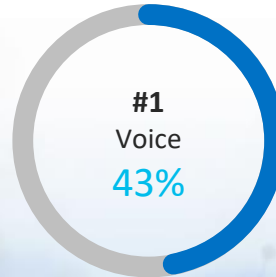
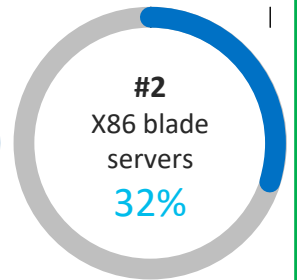
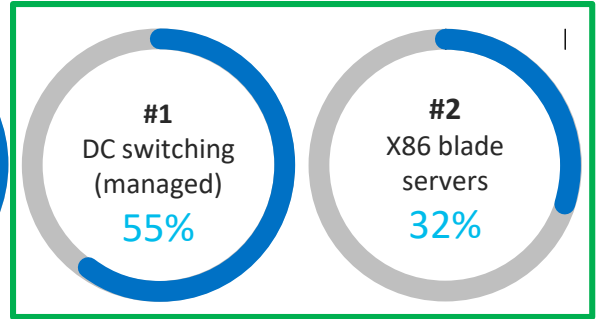
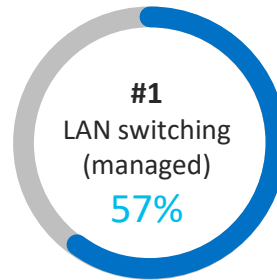
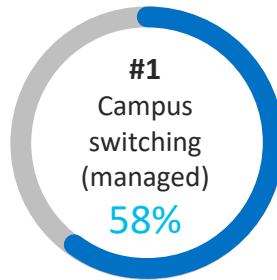
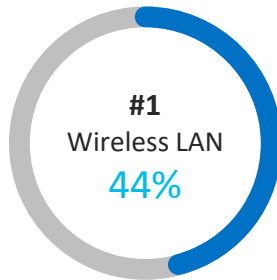
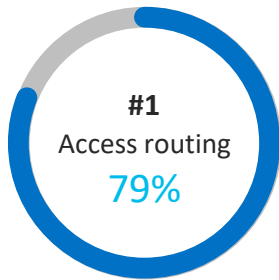
**AI BUILDERS SHOWCASE
CISCO**

AGENDA



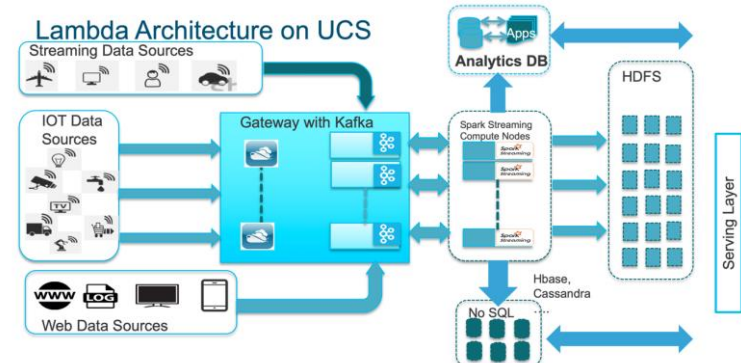
- Cisco Overview
- How Cisco is creating a bridge on AI solutions
- Retail Shelvesight Solution
- Use of Intel® AI technology
- Contact

CISCO SYSTEMS OVERVIEW

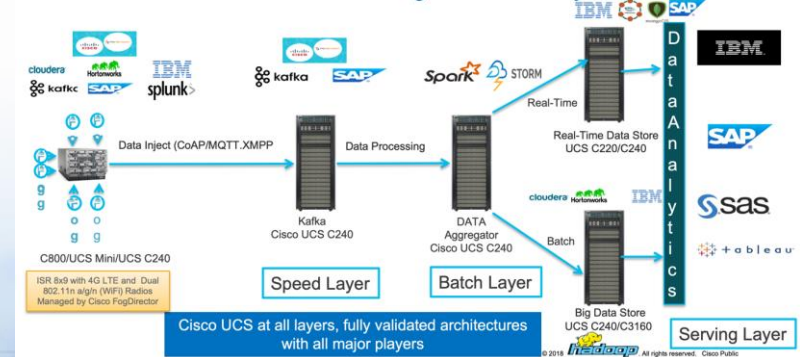


BUSINESS PROBLEM SOLVED

- Gap between Data Scientists and IT
 - Data Scientists: Data Pipeline
 - IT: Infrastructure
- Retail: Remote Locations
- Need to Accelerate Deployment



Cisco UCS Infrastructure for Big Data & Analytics

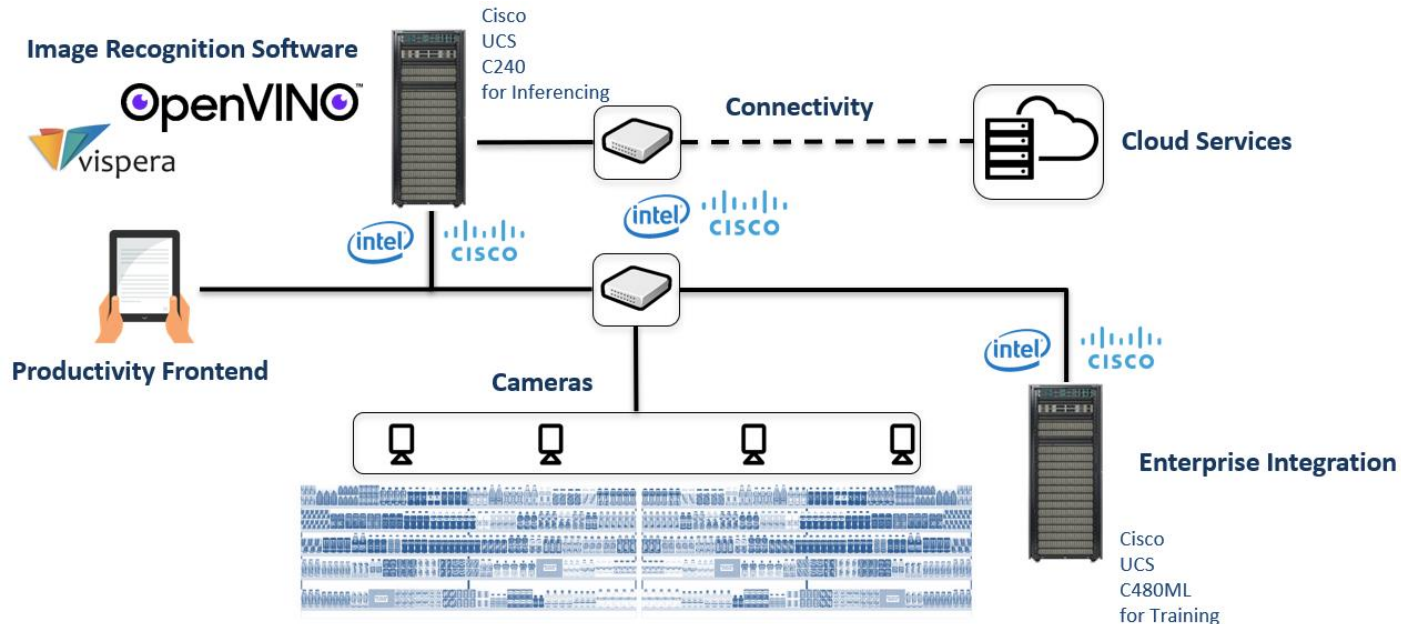


USE OF INTEL® AI TECHNOLOGY



HW: Cisco UCS C240 M5 servers using Intel® Xeon® Gold 6230 Processor

SW: Intel® Distribution of OpenVINO™ Toolkit



RESULTS



In this solution we are running ML Inferencing in a Cisco USC system with Intel® Xeon® Scalable processors – without any GPU or accelerator card.

Intel® Distribution of OpenVINO™ Toolkit is integrated on Vispera solution and configured for Intel® Xeon® Scalable processors optimization.



Product Recognition



Real-Time Out-of-Stock Detection



Shopper Tracking



CONTACT



Han Yang, PhD

Senior Product Manager – Cisco Datacenter Group

hanyang@cisco.com



Visit our table with your questions, or stop by the Intel® AI Builders
matchmaking table to set up a private meeting.





**AI ON
INTEL**

**AI BUILDERS SHOWCASE
HEWLETT PACKARD ENTERPRISE**

Accelerating enterprise AI innovation with software, services, and infrastructure

Speed the design and deployment of your AI strategy

Expertise and advisory services to accelerate your journey with **HPE Pointnext**

Give your data science teams instant access to AI tools and data

Industry's only turnkey, container-based software platform purpose-built for AI: **BlueData**

Build your AI models at scale with less complexity

Most comprehensive **AI infrastructure solutions**, from edge to cloud, optimized for Intel architecture

Execute your strategy fast, cost-effectively, with less risk

Pay-per-use consumption models that deliver cost, control and agility with **HPE GreenLake**

THE CHALLENGE: ML OPERATIONALIZATION

“ As AI and ML initiatives mature, **the biggest challenge faced by technical professionals is operationalizing ML** for effective management and delivery of models. **By 2021, at least 50% of machine learning projects will not be fully deployed.** Operationalization of machine learning is an imperative step in aligning AI investments with strategic business objectives — **the “last mile” toward delivering business value.** ”

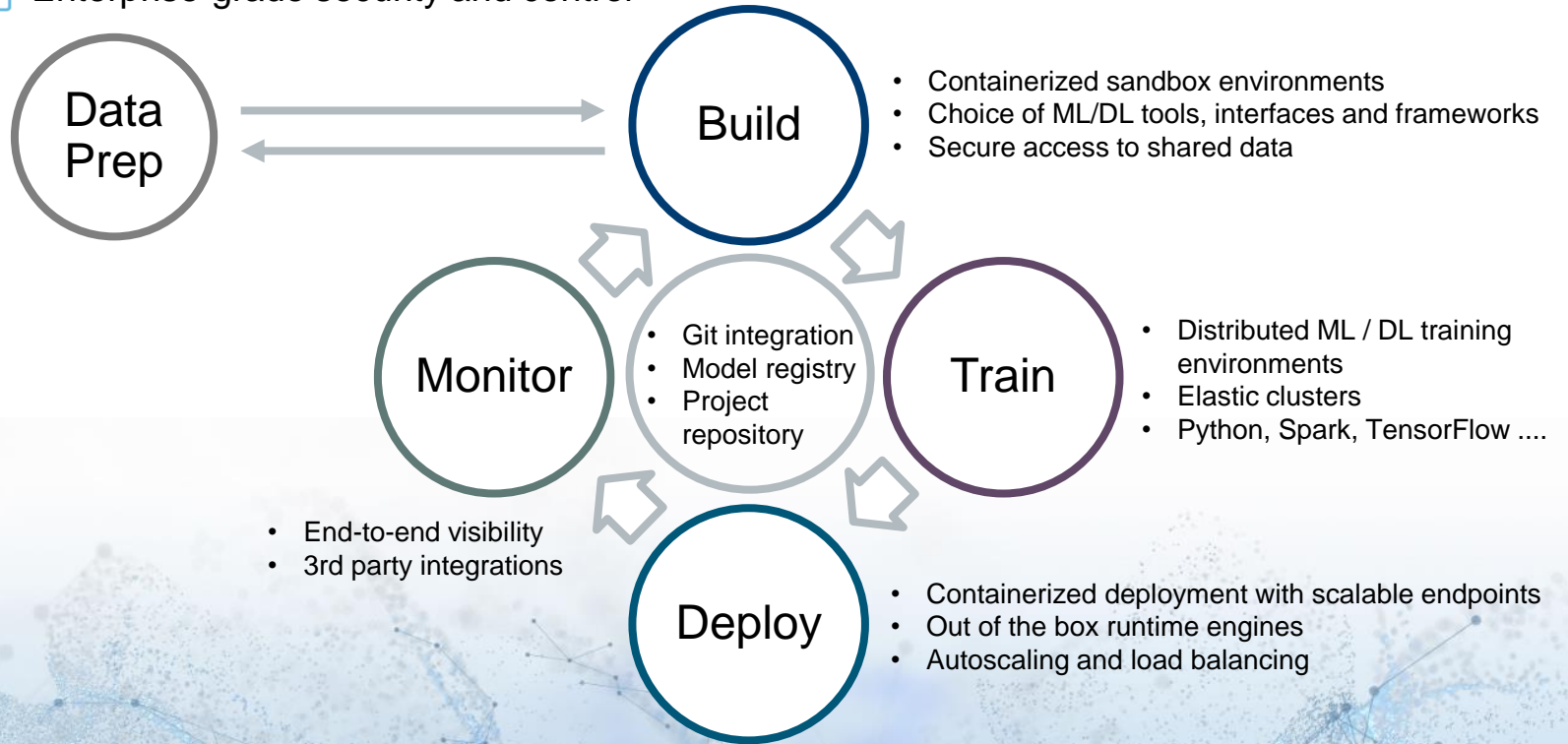
Source: Gartner, A Guidance Framework for Operationalizing Machine Learning for AI, October 24 2018.

Only operational ML pipelines deliver business value

OPERATIONALIZATION FOR THE ML LIFECYCLE

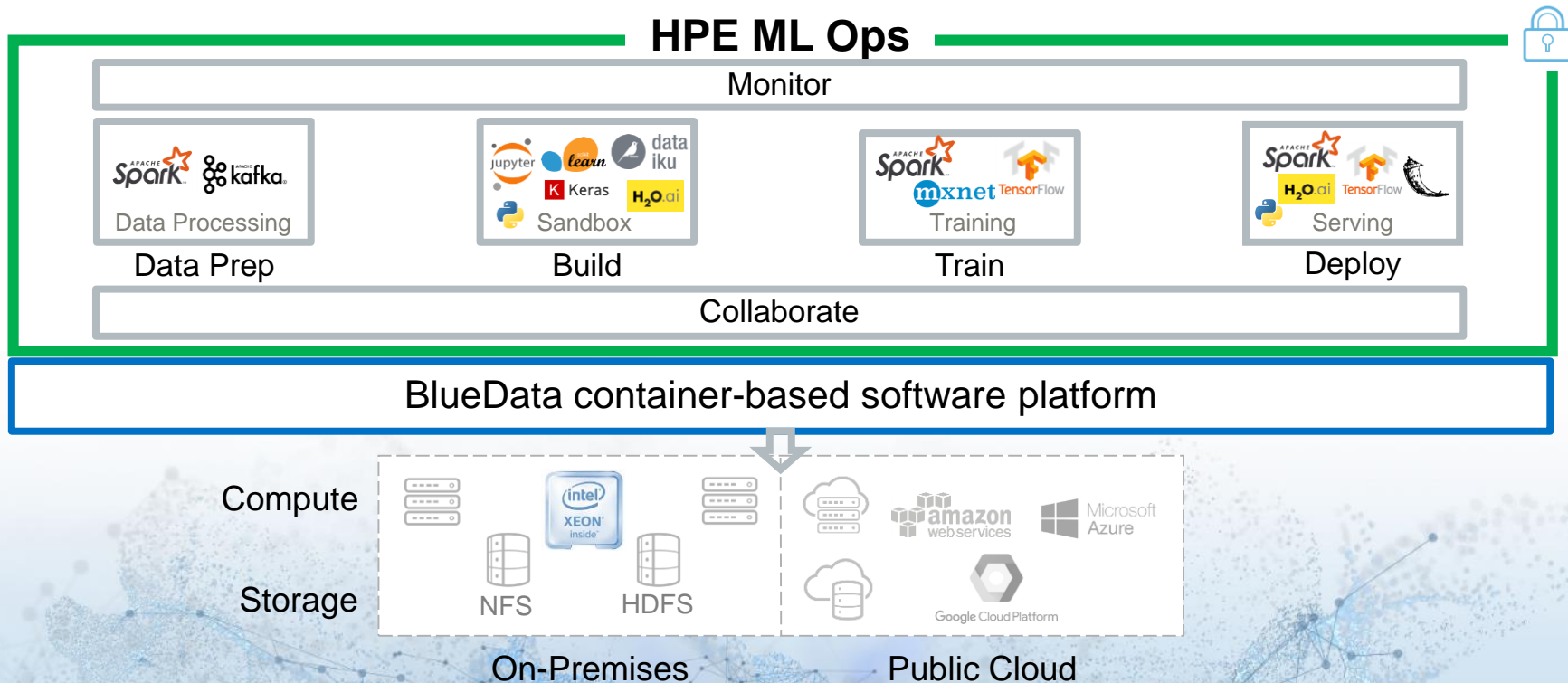


Enterprise-grade security and control



NEW: HPE MACHINE LEARNING OPS (ML OPS)

Bringing DevOps agility and speed to machine learning in the enterprise



HPE ML OPS – KEY BENEFITS

**Faster
Time-to-Value**

Ease of use

Choice
of tools

Rapid deployment

**Improved
Productivity**

High performance

Collaboration

Reproducibility

**Reduced
Risk**

Security and
control

Data and model
governance

High availability

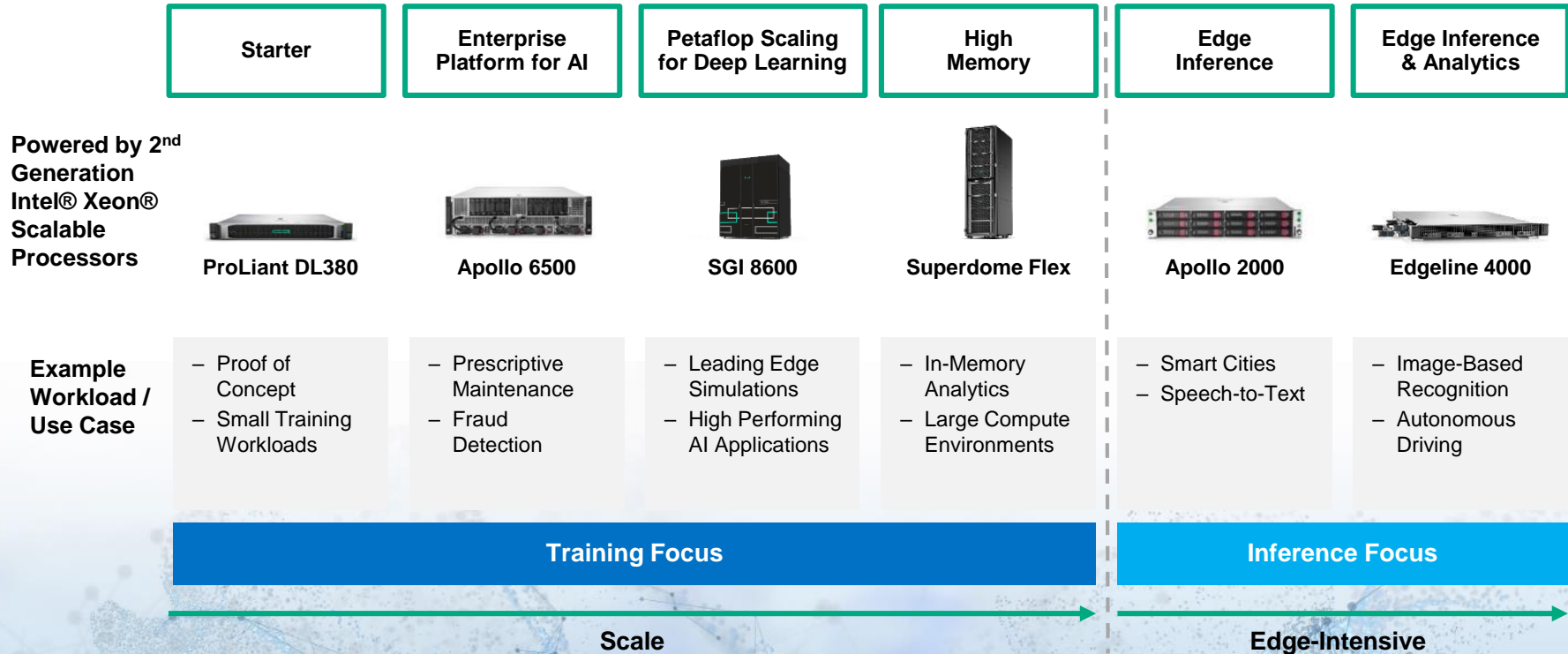
**Flexibility
and Elasticity**

On-prem, cloud,
or hybrid

Multi-tenancy

Scalable clusters

COMPREHENSIVE AI INFRASTRUCTURE SOLUTIONS



CONTACT



Nanda Vijaydev

Data Scientist and Distinguished Technologist

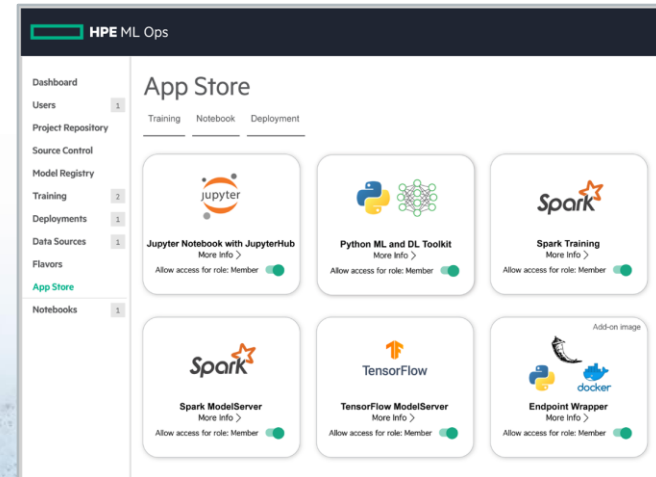
Hewlett Packard Enterprise (BlueData)

nanda.vijaydev@hpe.com

Visit our table with your questions, and see a demo in our booth at the O'Reilly AI Conference this week

Learn more about HPE AI solutions at hpe.com/AI

Learn more about HPE ML Ops at hpe.com/info/MLOps





HOW TO DEPLOY ON INTEL® ARCHITECTURE

RAVI PANCHUMARTHY

INTRO

- Intel collaborated with major CSP's and OEM partners to include pre-configured optimized DL environments, allowing data scientists and deep learning practitioners to instantly have access to Intel Optimized DL environments.
- This portion of the showcase will allow you to learn about Intel Optimized AI environments for your workloads, be it in the cloud or in your data center.
- Starting w/ CSPs
 - Microsoft Azure
 - Google Cloud Platform (GCP)
 - Amazon Web Services (AWS)
- Then with OEMs
 - HPE
 - Inspur
 - Lenovo
 - Dell

DEEP LEARNING DEPLOYED

DATA

TOOLS

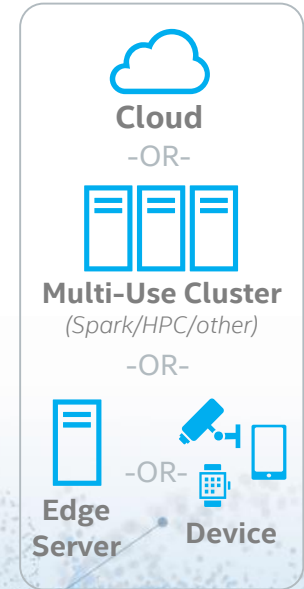
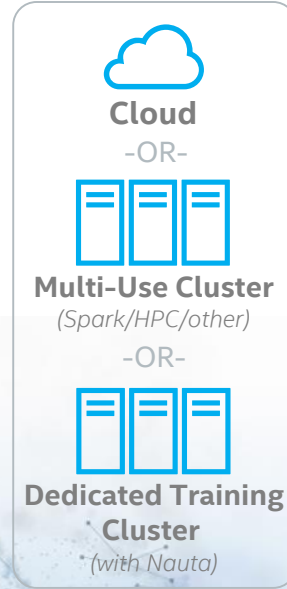
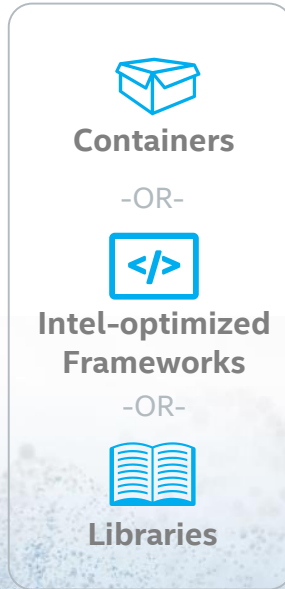
TRAINING

MODEL

OPTIMIZATION

INFERENCE

011010110110
110101101011
001011010100
011010110110
110101101011
001011010100
011010110110
110101101011
001011010100
011010110110
110101101011
001011010100
011010110110
011010110110
110101101011
001011010100
011010110110
110101101011
001011010100
011010110110
110101101011
001011010100
011010110110
110101101011
001011010100



End-to-end deep learning on Intel

DEPLOYING AI IN THE CLOUD

HARDWARE

- Azure offers Intel® Xeon® Processor family
 - The fastest VMs on Azure
 - Intel® Xeon® Platinum 8168 processor, features an all-cores clock speed greater than 3 GHz for most workloads.
 - Intel® AVX-512 instructions will provide up to a 2X performance boost to vector processing workloads.
- Compute-Optimized VMs with up to 72 vCPUs, 144 GBs of memory.
- HC-series VMs expose 44 non-HT CPU cores and 352 GB of RAM, featuring 100 Gb/s InfiniBand from Mellanox
- **Intel® Xeon® Scalable processors.**
Instances: FSv2, HC

SOFTWARE

- **Intel Data Science VM (DSVM)**
 - Intel optimized deep learning frameworks pre-configured and ready to use.
 - Deploy: <http://aka.ms/dsvm/intel>
- **Azure Machine Learning service***
 - Cloud service that you use to train, deploy, automate, and manage machine learning models, all at the broad scale that the cloud provides.
- Get started with [\\$200 in free credits](#).



- Product Page: <http://aka.ms/dsvm/intel>
- Intel blog: <https://www.intel.ai/intel-optimized-data-science-virtual-machine-azure/>
- Microsoft blog: <https://azure.microsoft.com/en-us/blog/intel-and-microsoft-bring-optimizations-to-deep-learning-on-azure/>



HARDWARE

- GCP offers latest Intel® Xeon® Processor family
 - 40% performance improvement compared to current GCP VMs**
 - Built-in Acceleration with Intel® Deep Learning Boost
 - Compute-Optimized VMs with up to 60 vCPUs, 240 GBs of memory, and up to 3TB of local storage**
 - Memory-Optimized VMs with up to 12 TB of memory and 416 vCPUs**
 - Learn more: <https://cloud.google.com/intel/>
 - **2nd Gen Intel® Xeon® Scalable processors: C2, M2**

**<https://cloud.google.com/blog/products/compute/introducing-compute-and-memory-optimized-vm-for-google-compute-engine>

SOFTWARE

- **Deep Learning VM**
 - Intel optimized deep learning frameworks pre-configured and ready to use.
 - Deploy: <https://console.cloud.google.com/marketplace/details/click-to-deploy-images/deeplearning>
- **AI Platform**
 - build ML applications with a managed Jupyter Notebook service that provides fully configured environments for different ML frameworks using Deep Learning VM Image
- Get started with \$300 in free credits. <https://cloud.google.com/free/>



Google Cloud Platform

- **Product Page:** <https://console.cloud.google.com/marketplace/details/click-to-deploy-images/deeplearning>
- **Intel blog:** <https://www.intel.ai/google-cloud-platform/>
- **Community blog:** <https://blog.kovalevskyi.com/deeplearning-images-revision-m9-intel-optimized-images-273164612e93>



AI/ML WITH AWS & INTEL

Carlos Escapa, Amazon Web Services



COMMON HISTORY & VALUES

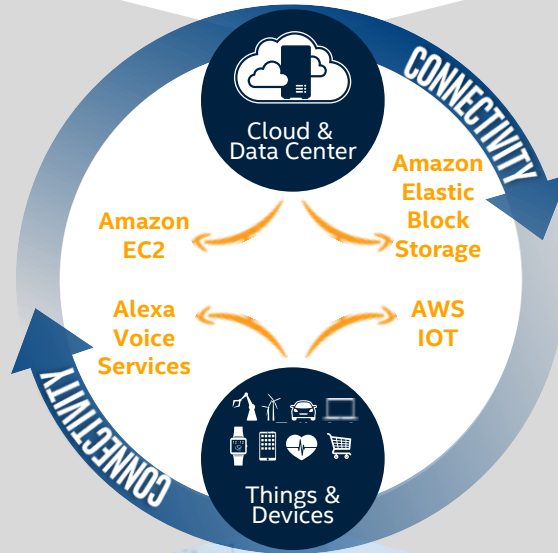
10+ year partnership

Joint development

Shared customer passion

High performance + low costs

World class supply chain



JOINT PRIORITIES

AI/ML/Analytics

HPC

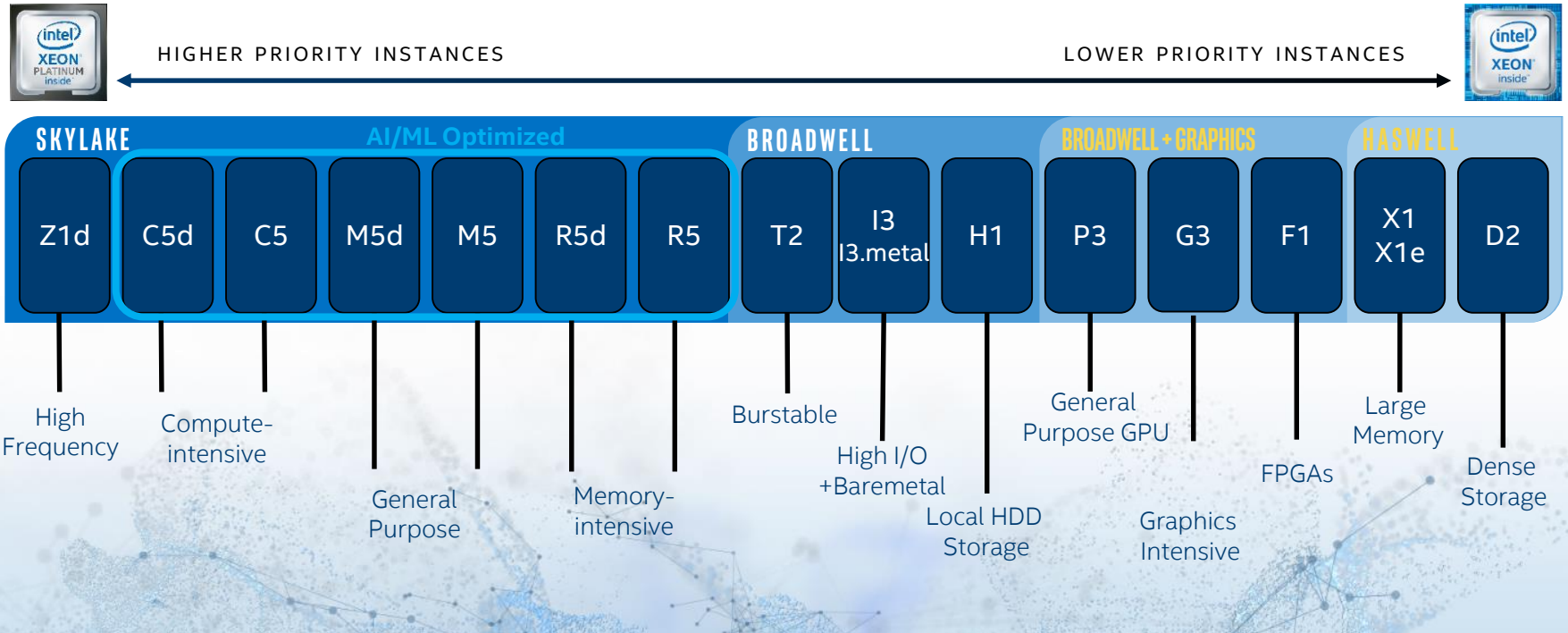
Digital Transformation

Cloud/Hybrid

Edge Computing

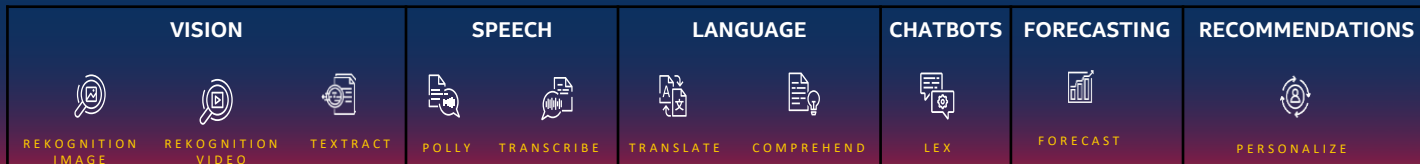
AWS EC2 INSTANCE SUMMARY

EC2 Instance Details: <https://aws.amazon.com/ec2/>



AWS MACHINE LEARNING STACK

APPLICATION SERVICES



PLATFORM SERVICES



Run high performance training with Amazon SageMaker CPU powered C5 instances.

Ground Truth Notebooks Algorithms + Marketplace Reinforcement Learning Training Optimization Deployment Hosting

FRAMEWORKS AND INTERFACES



INFRASTRUCTURE



EC2 P3 & P3DN



EC2 G4



EC2 C5



FPGA



GREENGRASS



ELASTIC INFERENCE



INFERENCE

Deploy Optimizations on the AWS Deep Learning AMI for EC2 CPUs

Run Machine Learning models on AWS Greengrass devices



AWS DEEPRACER



AWS DEEPLENS

AWS MACHINE LEARNING STACK

APPLICATION SERVICES

VISION



REKOGNITION
IMAGE



REKOGNITION
VIDEO



TEXTRACT

SPEECH



POLLY



TRANSCRIBE

LANGUAGE



TRANSLATE



COMPREHEND

CHATBOTS



LEX

FORECASTING



FORECAST

RECOMMENDATIONS



PERSONALIZE



AWS
DEEPRACER

PLATFORM SERVICES



Amazon SageMaker

Ground Truth

Notebooks

Algorithms + Marketplace

Reinforcement Learning

Training

Optimization

Deployment

Hosting

FRAMEWORKS AND INTERFACES



Caffe2



CNTK

mxnet

PYTORCH

TensorFlow



torch

Frameworks



KERAS



GLUEON

Interfaces



AWS
DEEPLENS

INFRASTRUCTURE



EC2 P3
& P3DN



EC2 G4



EC2 C5



FPGAs



GREENGRASS



ELASTIC
INFERENCE



INFERENTIA

USE CASES

[Hyper]Personalization – Product Recommendations, Advertising, Cross/Upselling, Next Best Action, Nudges, Livestock Management



Time-series analysis – Forecasting, Anomaly Detection, Preventive Maintenance, Fleet Balancing



Computer Vision – Visual Inspection, Quality Assurance, Document-driven Workflows



Natural Language Processing – CX transformation, Market Intelligence, Translation, Sentiment Analysis



AWS Marketplace Computer Vision

Cortexica Interiors Localisation

Cortexica BodyParts Localiser

Deep Vision brand recognition API

Logo Recognition in Images

Cortexica Interiors Localisation

Image collage classifier

Deep Vision visual search API

Barcode Detection

Vehicle Attribute Detection

Cortexica BodyParts Localiser

Image human classifier

Local Photo ID (Singapore)

Mighty Anonymize

Face blocking or blurring for Privacy

Face Anonymizer

Skin Disease Classification

Passport Data Page Detection

Waste Classifier

Deep Vision brand recognition API

Deep Vision vehicle recognition

Image mosaic classifier

Image text classifier

MACHINE LEARNING PARTNER ACCELERATION PROGRAM

Accelerating customer POC's on latest generation of AI-optimized instances: C5, R5, M5, C5d, R5d, M5d

Partners accelerate customer adoption of Machine Learning optimized AWS instances

- Targeting customer opportunities for ML workloads running on C5/M5/R5 instance families
- Funding available is Intel cash + AWS credits (cash will be matched up to same value of provided AWS credits)
- Targeting a 10X ROI in AWS ARR on the joint Intel & AWS investment.

Each APN Competency Partner has access to potential funding for
up to 5 POCs @ \$20k cash + \$20k credits

KEY TAKEAWAYS

1

AWS/Intel collaboration and programs make it easy to innovate

2

Intel instance types and Atom accelerate the development of AI/ML solutions

3

AWS Marketplace facilitates GTM motions and reaching new customers

How to engage

1. Reach out to the AWS Intel Account Team to request support for the partner investment program: aws-intel-oppty@amazon.com
2. We will look at the business case to evaluate our joint investment, typically starting with an AWS simple monthly calculator to show the services and instances that will be consumed.
3. Success stories can then be used at AWS re:Invent, Summits, AWSome Days, and social media

Key Contacts



Peter Bevan
Amazon Account Manager WW

Gareth Tucker
Amazon Account Manager EMEA

Kapil Bansal
Amazon Account Manager APAC

Marisa Reid
Amazon Sales Development
AI/ML Programs



Uday Tennety
Global Strategic Alliances Manager

Celia Baker
WW Intel Alliances PM

CONTACT US: aws-intel-oppty@amazon.com



- Product page: <https://aws.amazon.com/machine-learning/amis/>
- Intel blog: <https://www.intel.ai/amazon-web-services-works-with-intel-to-enable-optimized-deep-learning-frameworks-on-amazon-ec2-cpu-instances/>
- AWS blog: https://aws.amazon.com/about-aws/whats-new/2018/11/tensorflow1_12_mms10_launch_deep_learning_ami/

DEPLOYING AI WITH OEMS

- Accelerating enterprise AI innovation with software, services, and infrastructure

Speed the design and deployment of your AI strategy

Expertise and advisory services to accelerate your journey with **HPE Pointnext**

Give your data science teams instant access to AI tools and data

Industry's only turnkey, container-based software platform purpose-built for AI: **BlueData**

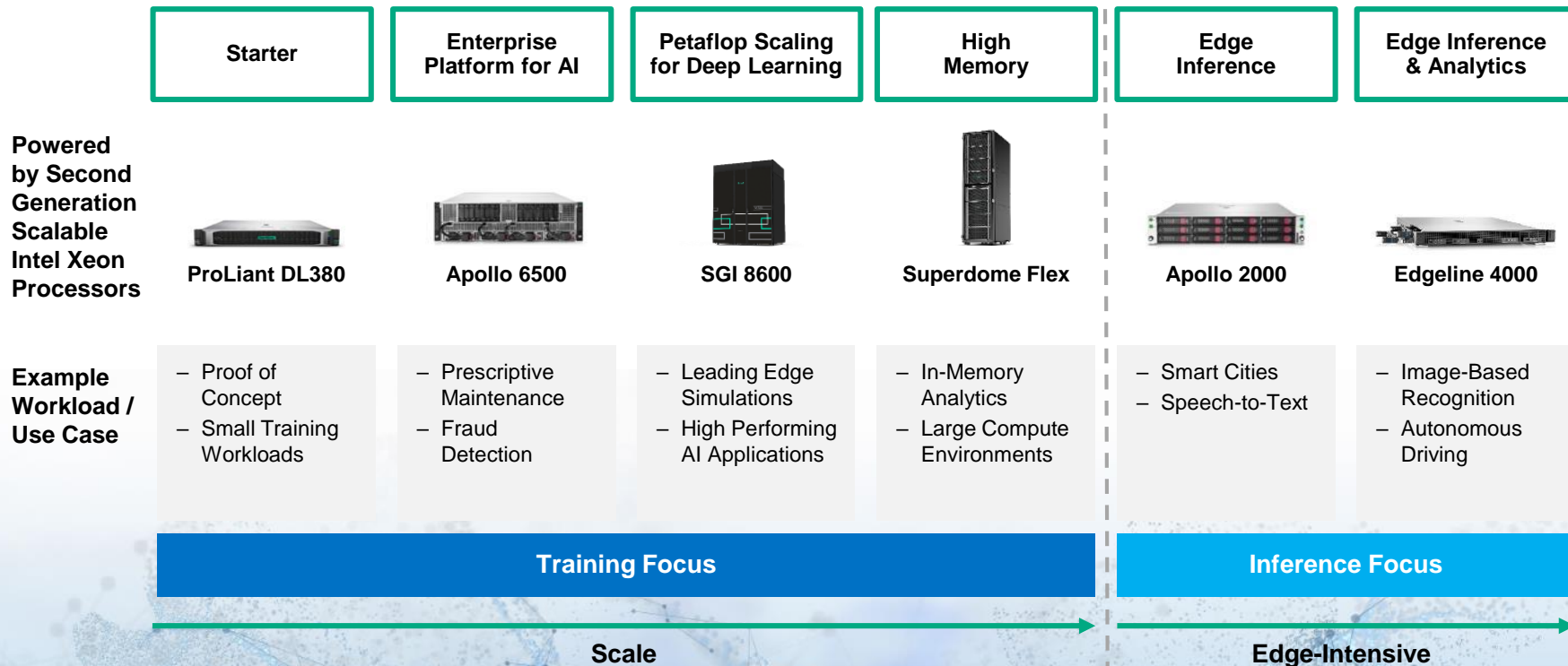
Build your AI models at scale with less complexity

Most comprehensive **AI infrastructure solutions**, from edge to cloud, optimized for Intel architecture

Execute your strategy fast, cost-effectively, with less risk

Pay-per-use consumption models that deliver cost, control and agility with **HPE GreenLake**

COMPREHENSIVE AI INFRASTRUCTURE SOLUTIONS





**AI ON
INTEL**

**AI BUILDERS SHOWCASE
INSPUR**

AGENDA

- Company overview
- Business problem Solved
- Use of Intel® AI technology
- Performance Results
- Contact

Inspur is one of the world's leading hardware innovators and solutions providers for AI and deep learning. We are developing smarter, more powerful end-to-end solutions that help businesses leverage and tackle intelligent technologies from cloud to edge.



Global Top 3 Provider

One of the world's 3 largest server and solutions vendors



Serving Major Global CSPs

Fulfills 90% AI server demand with T1 CSP in China



Select Partner for FPGA

Chosen AI partner by major global CSPs for FPGA projects

Inspur Full-Stack AI Capabilities

SOFTWARE STACK	FRAMEWORK	Caffe-MPI • TensorFlow • Caffe • CNTK • MXNET
	MANAGEMENT	Inspur AIStation • Inspur Teye
HARDWARE	INFERENCE	
	TRAINING	

BUSINESS PROBLEM SOLVED

Cloud FPGA

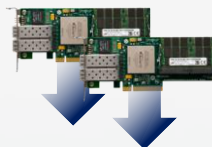


Inspur F10A
Arria10 Based,
HHHL, 45W

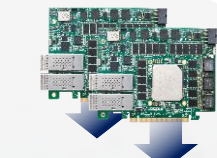


Inspur F10S
Stratix10 Based,
FHHL, 150W

FPGA-as-a-Service



Edge Computing



NF5260M5
AI Edge Server



CSP



Enterprise



Image Recognition



Autonomous Driving



Video Recommendation



Financial Terminals



Video Editing



CAD/Content Creation

OpenVINO™

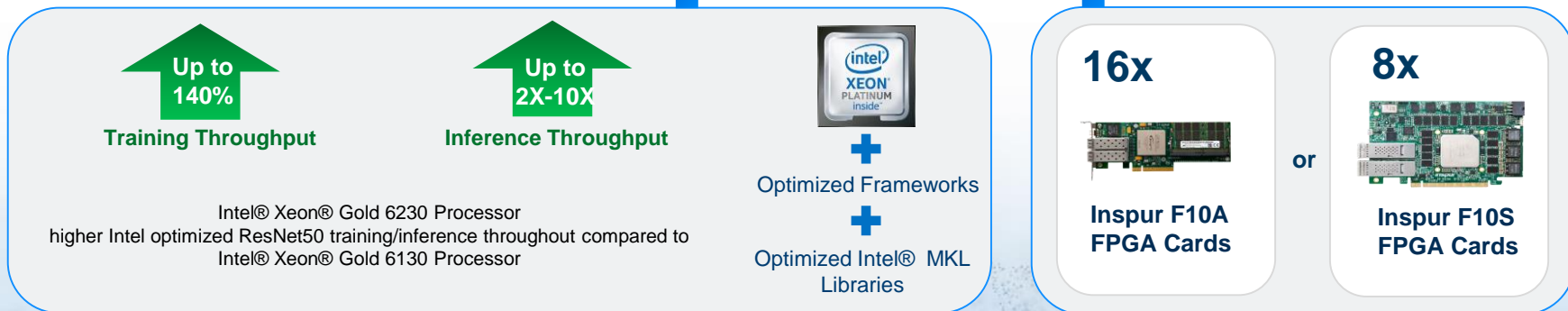
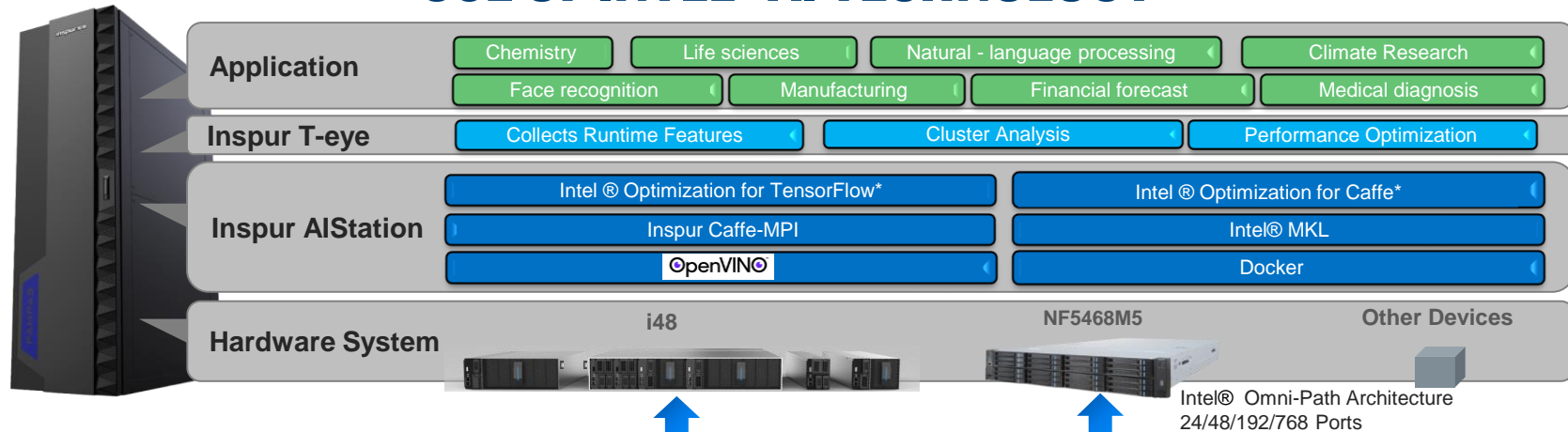
Based on convolutional neural networks (CNN), the toolkit extends workloads across Intel® hardware (including accelerators) and maximizes performance.

Inspur TF2 FPGA Tool Kit

A powerful open-sourced complement to OpenVINO™ by model cropping and model compression for supporting the CNN, Transformer, LSTM which can be quickly ported on FPGA

USE OF INTEL® AI TECHNOLOGY

inspur



Deliver significant AI performance with hardware and software optimizations on Intel® Xeon® Cascade Lake processor

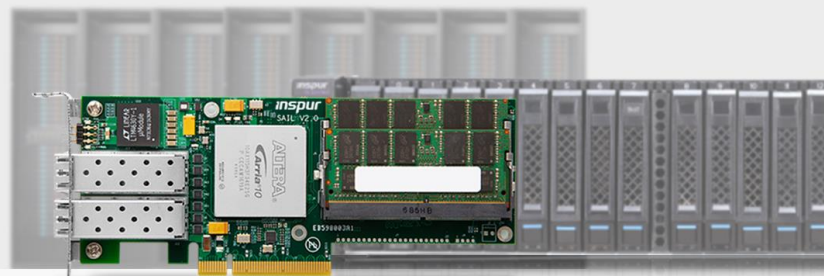
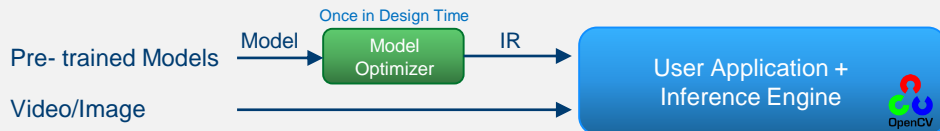
USE OF INTEL® AI TECHNOLOGY



Without OpenVINO™ Deep Learning Frameworks



With OpenVINO™ Deep Learning Inference Engine

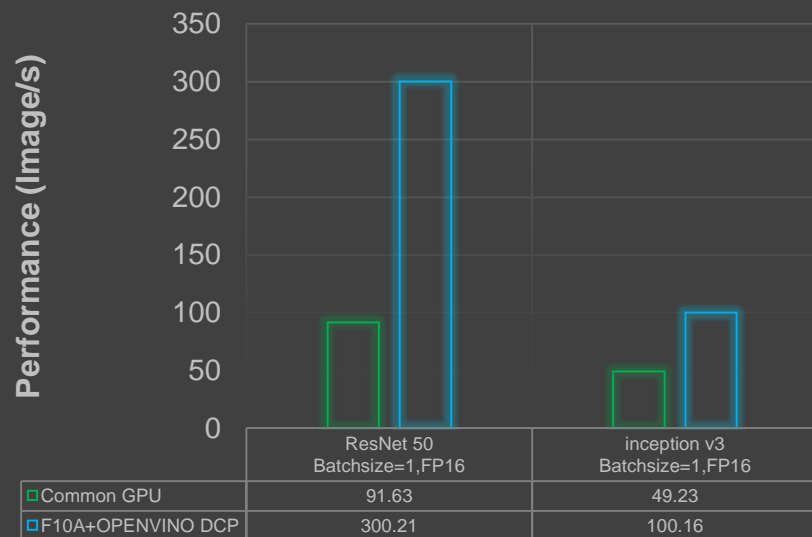


Inspur F10A FPGA Card

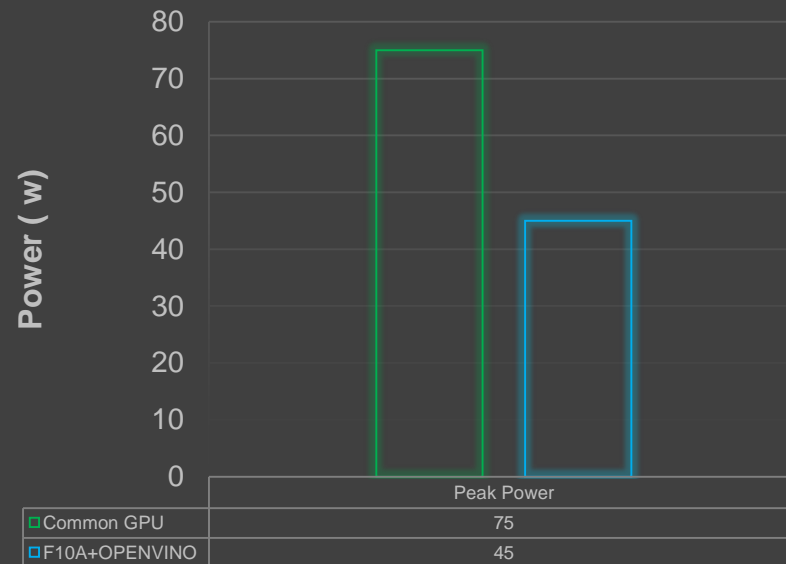
Extreme density, efficient heat dissipation
1.366TFLOPS, HHL, Active cooling

RESULTS

Small-batchsize Performance Comparison between common GPU and F10A with OPENVINO



Power Consumption



CONFIGURATION SPECIFICATIONS

Tested by Inspur on 06/15/2019 by using NF5280M5, Inspur server with 2-socket Intel Xeon 6140 Processor, total Memory 374 GB (12 slots/ 32GB/ 2666 MHz) . Deep Learning Framework: OpenVINO 2019R1 on F10A, Tensorflow 1.12 on GPU, tested using Batchsize of 1.



CONTACT

inspur

Bob Anderson

VP of Sales

Bobanderson@inspur.com



Visit our table with your questions, or stop by the Intel® AI Builders matchmaking table to set up a private meeting.





**AI ON
INTEL**

**AI BUILDERS SHOWCASE
LENOVO**

AGENDA

- Company overview
- Business problem they solve by prioritized vertical
- Use of Intel® AI technology
- Results
- Contact

Lenovo Data Center Group – an experienced and innovative global team

#1 Provider of supercomputers on the TOP500 list & fastest growing HPC company *	#1 Hyperconverged revenue growth**	1st Warm water cooled Intel Scalable Xeon Supercomputer	1st In x86 Server Reliability 6 years running***
48.6% Growth rate Fastest growing server vendor *	10,000 Service professionals	20M+ Servers shipped	6 Owned/controlled manufacturing sites

* <https://bit.ly/2N537Eu>

** <https://lnv.gy/2yjPY5v>

***ITIC 2016/2017, ITIC 2017/2018, ITIC 2018/2019

BUSINESS PROBLEM SOLVED

- Architecting, deploying, and optimizing infrastructure for AI is **HARD**
 - Usage patterns unpredictable – **HOW MUCH INFRASTRUCTURE?**
- AI clusters are most efficient for IT, but least understood by Data Scientists
 - System setup takes much effort – **WASTED DATA SCIENTIST TIME**
- Lenovo Intelligent Computing Orchestration (LiCO) software solves both problems
 - LiCO + Intel-based clusters enable infrastructure optimization
 - GUI-based solution for HPC&AI makes deploying AI workloads fast & simple
 - Data Scientists can run tuning jobs in parallel, increasing their productivity

USE OF INTEL® AI TECHNOLOGY

HW: Full Lenovo server portfolio featuring Intel® Xeon® Scalable processors

- From a single 2U/2S to the largest warm water cooled supercomputers
- Flexibility to deploy whatever infrastructure fits your datacenter best for AI

SW: Leverages the latest Intel® AI framework optimizations

- Containerized framework deployment eliminates setup time for users
- Maximum performance on Intel® Xeon® Scalable processors for every job deployed

RESULTS

With LiCO + Intel-based solutions:

- Data Science users get more productive
 - Software stack setup + job deployment takes seconds
 - No need to learn cluster tools to take advantage of the powerful resources
 - Run multiple jobs in parallel to tune hyperparameters faster
- IT gains more infrastructure efficiency too
 - AI Infrastructure can start small and easily grow, or AI can converge with HPC
 - More user productivity drives higher utilization

CONTACT

Matthew Ziegler

Director, HPC/AI Technology & Architecture

mziegler@lenovo.com



Visit our table with your questions, or stop by the Intel® AI Builders matchmaking table to set up a private meeting.





**AI ON
INTEL**

**AI BUILDERS SHOWCASE
DELL EMC**

AGENDA



- Company overview
- Business problem they solve by prioritized vertical
- Use of Intel® AI technology
- Results
- Contact



DELL EMC COMPANY OVERVIEW



Optimal configurations of workstations, servers, storage, networking, data center options and software



POWEREDGE SERVERS

designed for HPC environments, including more local I/O capacity and expandability



PRECISION WORKSTATIONS

for high-demand, industry specific applications



STORAGE

powerful performance, density and efficiency for data-intensive HPC environments



MANAGEMENT

deploy customers over bare metal with single pane of glass management



NETWORKING

delivering the bandwidth and speed you need



ACCELERATORS

offering the horsepower needed for bigger simulations faster than ever before



SERVICES

providing end-to-end services to maximize HPC investments



CLOUD

build or scale, rent cycles, and/or get burst capacity

READY SOLUTIONS FOR AI WITH INTEL

A full package of all the components needed

Hardware

Faster deployment

from months
to weeks

Dell EMC

PowerEdge servers
storage • networking

+

Intel Accelerators

Libraries and frameworks

Optimized

for fast application
development

Easy to use

for modeling

BigDL • TensorFlow

• Horovod

• MLKDNN

Software

Ready-to-go

data management and
data science tools

Cloudera • Hortonworks

Spark • Kubernetes

Nauta

Services

Drive rapid adoption

and optimization
of new environment

Strategy • Integration

Data engineering

Upskilling of resources

Ongoing support

DEEP LEARNING WITH INTEL® - VALUE

Primary target

- Customers looking to leverage Kubernetes/Docker environments for running deep learning workloads

Differentiation

- New Dell EMC IP: Kubernetes-based use case templates
- Nauta to simplify model development and orchestrate of arduous tasks
- Validated infrastructure stack providing the power and scalability to build and train machine and deep learning models

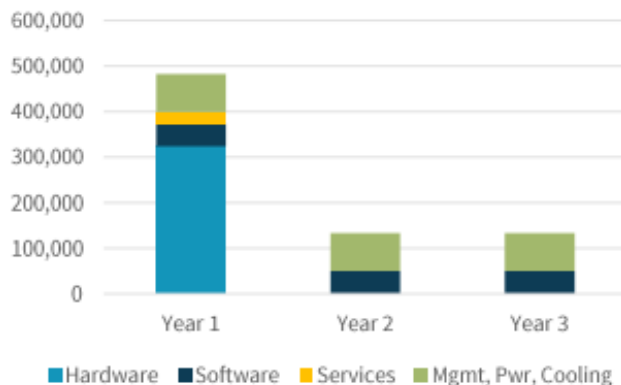
DEEP LEARNING WITH INTEL® - PACKAGE

1. **Nauta Enterprise**
2. **R740xd login/master node**
 - Intel® Xeon® Scalable Gold processors
 - 384GB RAM
 - 144TB (12x 12TB – 112TB usable)
3. **C6420 compute nodes**
 - Intel® Xeon® Scalable Gold processors
 - 192GB RAM
 - No user-accessible storage
4. **10Gb Dell Networking S4048-ON**
5. **Isilon H600 NAS**

RESULTS

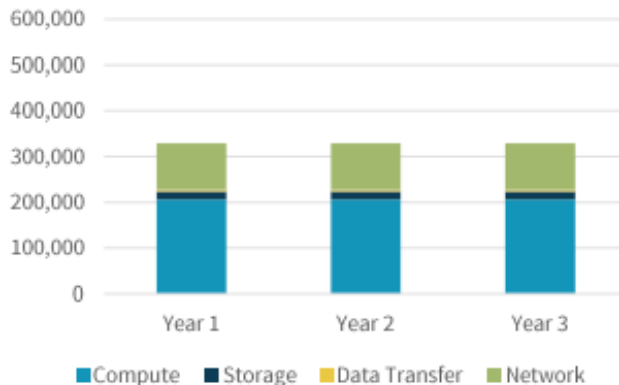
Deep Learning with Intel

Deep Learning Training Three-year TCO
24 Hr Compute, 100TB Storage



Leading Public Cloud AI Service

Deep Learning Training Three-year TCO
12 Hr Compute, 10TB Storage



Source: ESG Technical Validation,
*Dell EMC Ready Solutions for AI:
Deep Learning with Intel*,
April 2019.

This InstaGraphic highlights results from an ESG
Technical Validation commissioned by:

DELLEMC

CONTACT



Phil Hummel
Sr. Principal Engineer

philip.hummel@dell.com



Visit our table with your questions, or stop by the Intel® AI Builders
matchmaking table to set up a private meeting.



HYBRID DEPLOYMENTS

HYBRID DEPLOYMENTS

<https://aws.amazon.com/outposts/>



AWS Outposts
Run AWS infrastructure on-premises for a truly consistent hybrid experience

[Sign up to learn more](#)

AWS Outposts bring native AWS services, infrastructure, and operating models to virtually any data center, co-location space, or on-premises facility. You can use the same APIs, the same tools, the same hardware, and the same functionality across on-premises and the cloud to deliver a truly consistent hybrid experience. Outposts can be used to support workloads that need to remain on-premises due to low latency or local data processing needs.

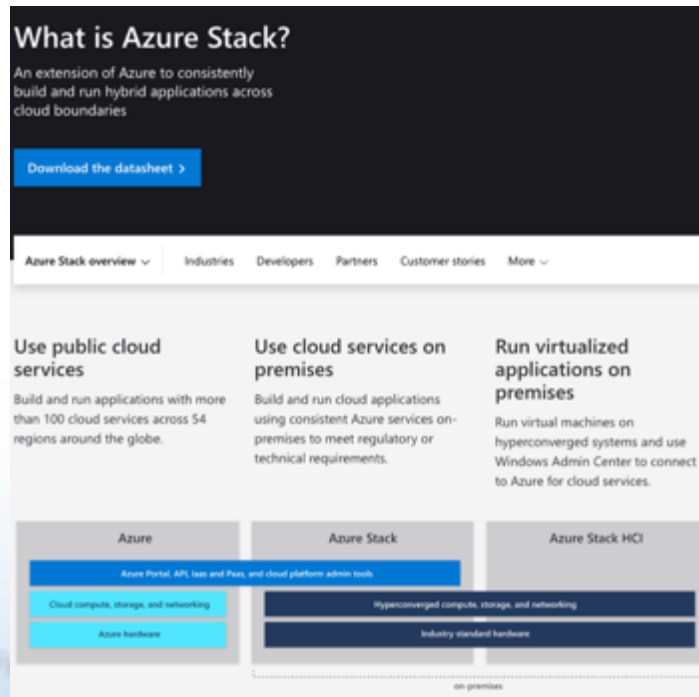
AWS Outposts come in two variants: 1) VMware Cloud on AWS Outposts allows you to use the same VMware control plane and APIs you use to run your infrastructure, 2) AWS native variant of AWS Outposts allows you to use the same exact APIs and control plane you use to run in the AWS cloud, but on-premises.

AWS Outposts infrastructure is fully managed, maintained, and supported by AWS to deliver access to the latest AWS services. Getting started is easy: you simply log into the AWS Management Console to order your Outposts servers, choosing from a wide catalog of IAGAs with a broad range of Amazon EC2 instances and capacity.

How it works



<https://azure.microsoft.com/en-us/overview/azure-stack/>



What is Azure Stack?

An extension of Azure to consistently build and run hybrid applications across cloud boundaries

[Download the datasheet >](#)

[Azure Stack overview](#) ▾ [Industries](#) [Developers](#) [Partners](#) [Customer stories](#) [More](#) ▾

Use public cloud services

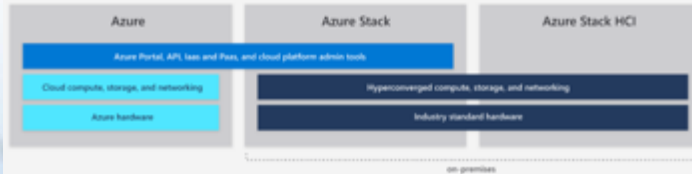
Build and run applications with more than 100 cloud services across 54 regions around the globe.

Use cloud services on premises

Build and run cloud applications using consistent Azure services on-premises to meet regulatory or technical requirements.

Run virtualized applications on premises

Run virtual machines on hyperconverged systems and use Windows Admin Center to connect to Azure for cloud services.



Azure	Azure Stack	Azure Stack HCI
Azure Portal, API, IaaS and PaaS, and cloud platform admin tools	Azure Portal, API, IaaS and PaaS, and cloud platform admin tools	Azure Portal, API, IaaS and PaaS, and cloud platform admin tools
Cloud compute, storage, and networking	Cloud compute, storage, and networking	Hyperconverged compute, storage, and networking
Azure hardware	Azure hardware	Industry standard hardware

on premises

TO LEARN MORE

- Contact our partners or Intel at **DeployAI@Intel.com**
- Visit <https://www.intel.ai/deploy-on-intel-architecture/>

QUICK RECAP!

AI IN THE ENTERPRISE: THE INTEL® AI BUILDERS PARTNER SHOWCASE

- **Intel® AI Builders Program:** Benefits to partners and enterprise customers
- **Why Intel® AI:** Hardware, Software, Ecosystem
- **Partner Showcase:** Finance/ retail/ healthcare/ cross-industry AI solutions
- **How to Deploy on Intel® Architecture:** Optimized CSPs and OEM channels
- **Match-Making:** Continues until 7:30pm
- **Happy Hour:** Starts now!



NOTICES AND DISCLAIMERS

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration.

No product or component can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. For more complete information about performance and benchmark results, visit <http://www.intel.com/benchmarks>.

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/benchmarks>.

Intel® Advanced Vector Extensions (Intel® AVX)* provides higher throughput to certain processor operations. Due to varying processor power characteristics, utilizing AVX instructions may cause a) some parts to operate at less than the rated frequency and b) some parts with Intel® Turbo Boost Technology 2.0 to not achieve any or maximum turbo frequencies. Performance varies depending on hardware, software, and system configuration and you can learn more at <http://www.intel.com/go/turbo>.

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.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Intel, the Intel logo, and Intel Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

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

© 2019 Intel Corporation.

