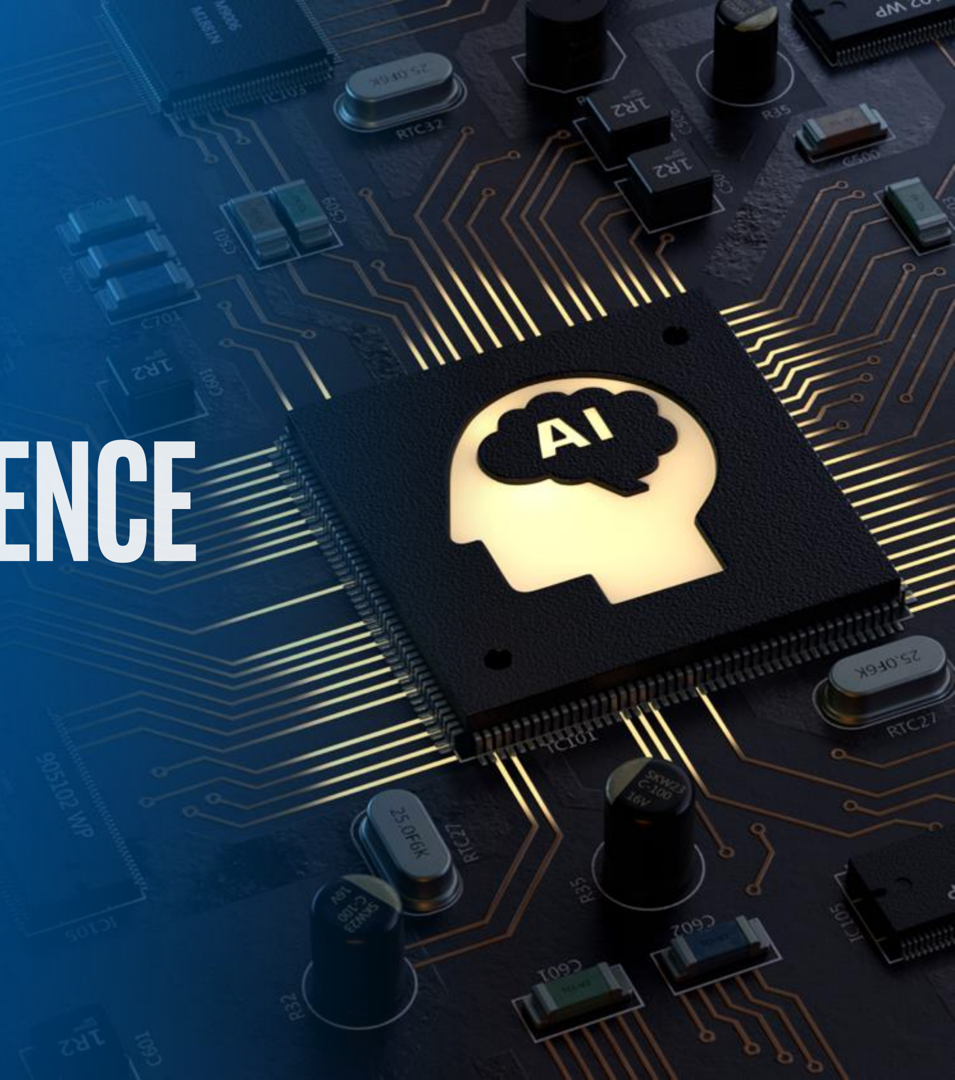
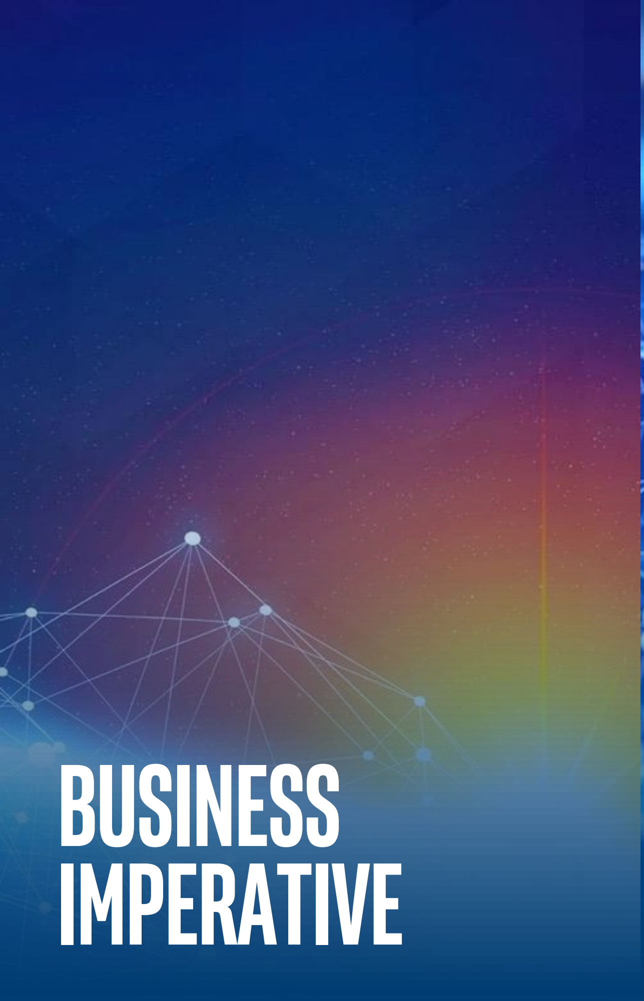
ARTIFICIAL INTELLIGENCE



September 2019 - ITU

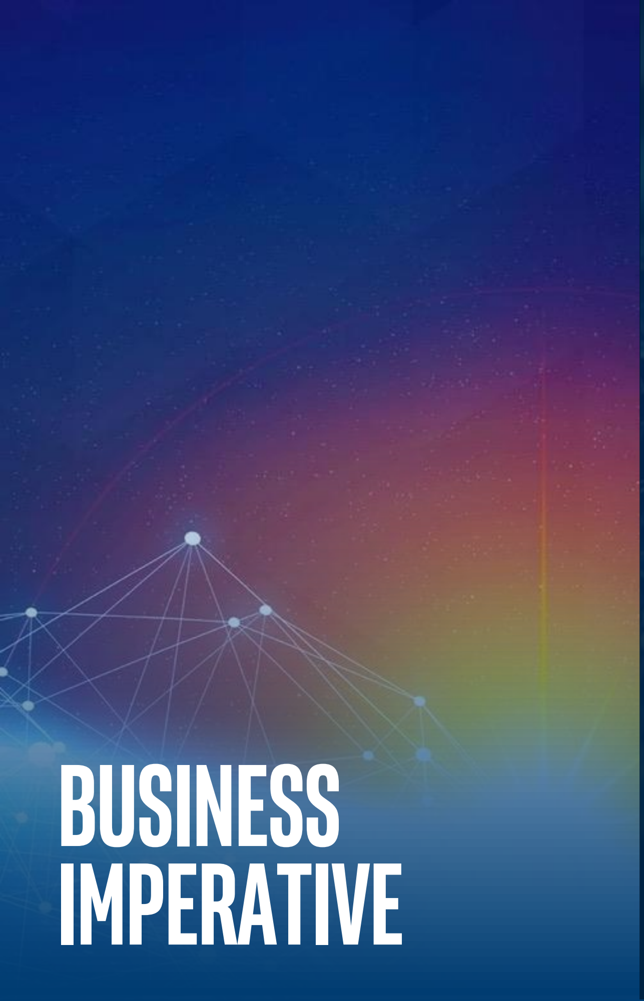
Paul Haines, Thailand Account Executive



BUSINESS IMPERATIVE

INTEL® AI

AI CASE STUDY



BUSINESS IMPERATIVE

INTEL® AI

AI CASE STUDY

The AI Mandate

"AI technologies are evolving fast and growing increasingly **critical** to firms' ability to win, serve, and retain customers."

FORRESTER

"…strategic technologies for 2019 with the potential to drive significant **disruption** and deploy **opportunity** over the next five years"

GARTNER

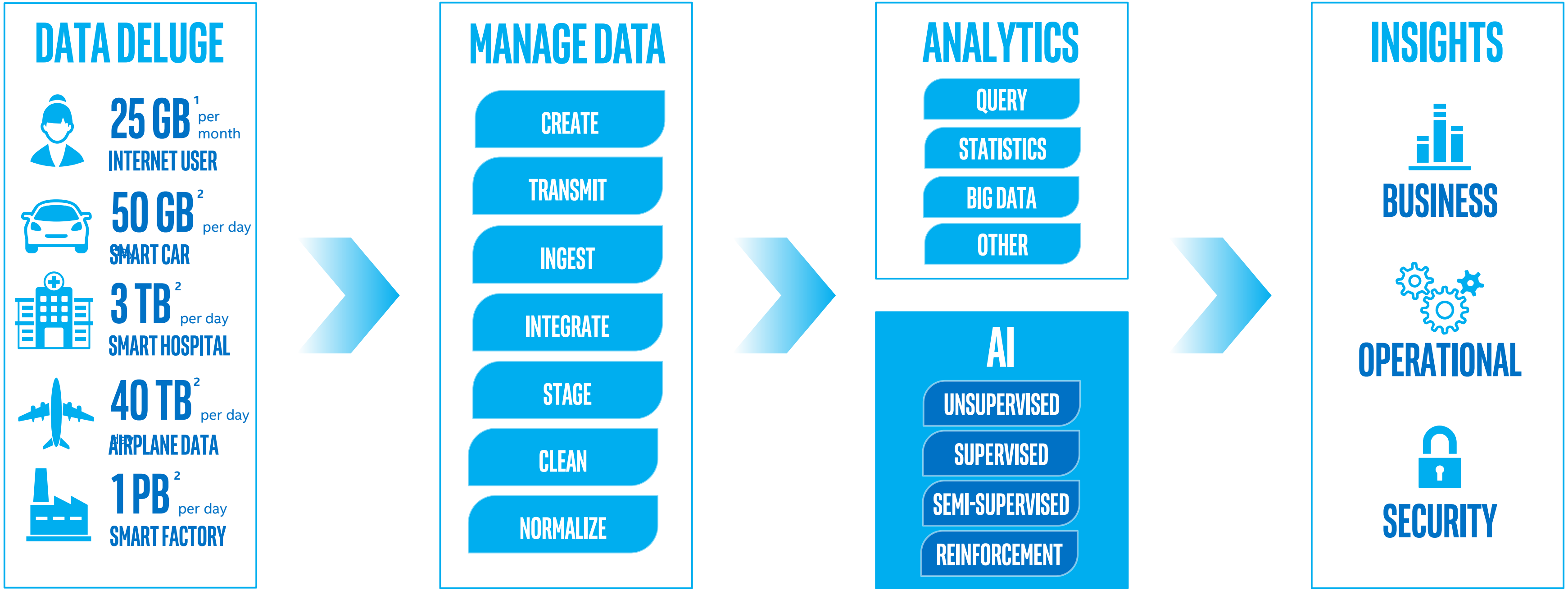
"…**70%** of CIOs will aggressively apply data and AI to IT operations, tools, and processes by 2021."

IDC

THE TIME TO BEGIN AI ADOPTION IS NOW

Source: https://www.forrester.com/report/The+Forrester+Tech+Tide+Artificial+Intelligence+For+Business+Insights+Q3+2018/-/E-RES143252
Source: https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2019
Source: https://www.idc.com/getdoc.jsp?containerId=prUS44420918

Why AI?



DATA DELUGE

- 25 GB[1] per month — INTERNET USER
- 50 GB[2] per day — SMART CAR
- 3 TB[2] per day — SMART HOSPITAL
- 40 TB[2] per day — AIRPLANE DATA
- 1 PB[2] per day — SMART FACTORY

MANAGE DATA

- CREATE
- TRANSMIT
- INGEST
- INTEGRATE
- STAGE
- CLEAN
- NORMALIZE

ANALYTICS

- QUERY
- STATISTICS
- BIG DATA
- OTHER

AI

- UNSUPERVISED
- SUPERVISED
- SEMI-SUPERVISED
- REINFORCEMENT

INSIGHTS

- BUSINESS
- OPERATIONAL
- SECURITY

EXTRACT VALUABLE INSIGHTS FROM DATA

[1] Source: http://www.cisco.com/c/en/us/solutions/service-provider/vni-network-traffic-forecast/infographic.html
[2] Source: https://www.cisco.com/c/dam/m/en_us/service-provider/ciscoknowledgenetwork/files/547_11_10-15-DocumentsCisco_GCI_Deck_2014-2019_for_CKN__10NOV2015_.pdf

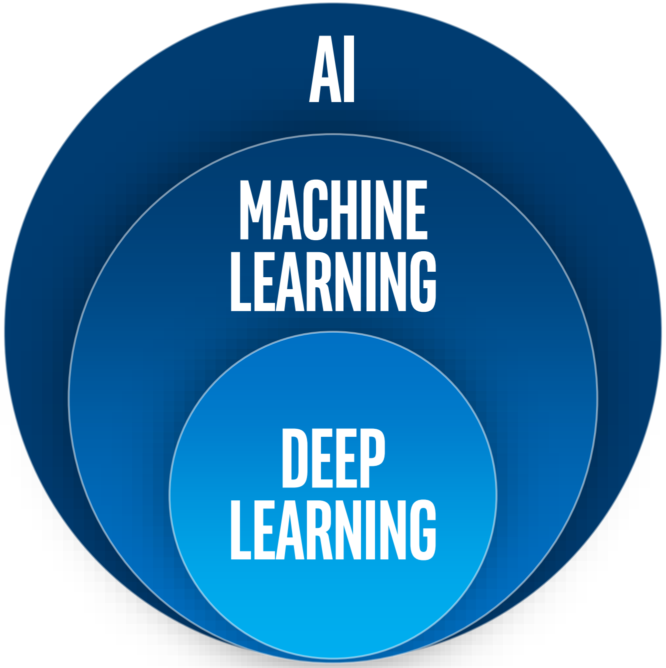
What is AI?

SUPERVISED LEARNING

UNSUPERVISED LEARNING

SEMI-SUPERVISED LEARNING

REINFORCEMENT LEARNING



Regression

Classification

Clustering

Decision Trees

Data Generation

Image Processing

Speech Processing

Natural Language Processing

Recommender Systems

Adversarial Networks

NO ONE-SIZE-FITS-ALL APPROACH TO AI

AI Closer Look

MACHINE LEARNING

Algorithms designed to deploy better insight with more data

Regression (Linear/Logistic)

Classification (Support Vector Machines/SVM, Naïve Bayes)

Clustering (Hierarchical, Bayesian, K-Means, DBSCAN)

Decision Trees (RandomForest)

Extrapolation (Hidden Markov Models/HMM)

More…

DEEP LEARNING

Neural networks used to infer meaning from large dense datasets

Image Recognition (Convolutional Neural Networks/CNN, Single-Shot Detector/SSD)

Speech Recognition (Recurrent Neural Network/RNN)

Natural Language Processing (Long-Short Term Memory/LSTM)

Data Generation (Generative Adversarial Networks/GAN)

Recommender System (Multi-Layer Perceptron/MLP)

Time-Series Analysis (LSTM, RNN)

Reinforcement Learning (CNN, RNN)

More…

REASONING

Hybrid of analytics & AI techniques designed to find meaning in diverse datasets

Associative Memory (Intel® Saffron AI memory base)

← **See also:** machine & deep learning techniques

More…

Which Approach is Best?

SMART FACTORY

QUESTION	METHOD	APPROACH
How many parts should we manufacture?	Historical supply and demand analysis	Statistical Analytics
What will our production yield be?	Algorithm learns which variables correlate to yield	Machine Learning *(Unsupervised)*
Which parts have visual defects?	Algorithm learns to identify defects in images	Deep Learning *(Supervised)*
Can my robotic arm learn to get better?	Algorithm that acts and adapts based on feedback	Deep Learning *(Reinforcement)*

CHOOSE THE RIGHT AI APPROACH FOR YOUR CHALLENGE

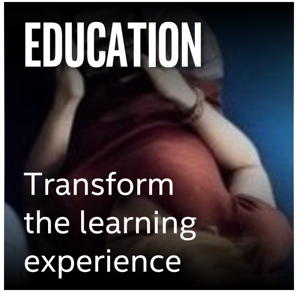
AI Solutions in Every Market

AGRICULTURE

Achieve higher yields and increase efficiency

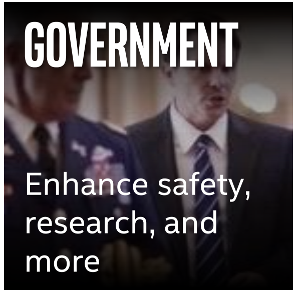
ENERGY

Maximize production and uptime



EDUCATION

Transform the learning experience



GOVERNMENT

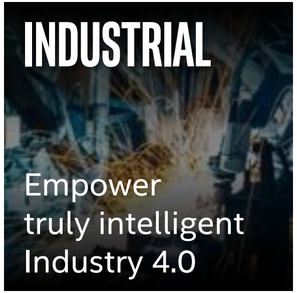
Enhance safety, research, and more

FINANCE

Turn data into valuable intelligence

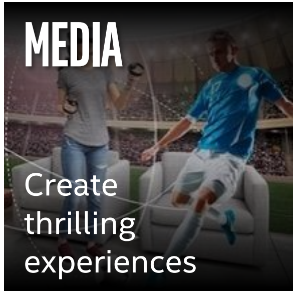
HEALTH

Revolutionize patient outcomes



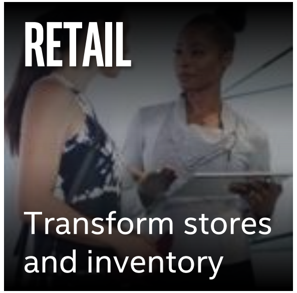
INDUSTRIAL

Empower truly intelligent Industry 4.0



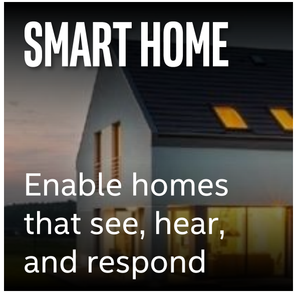
MEDIA

Create thrilling experiences



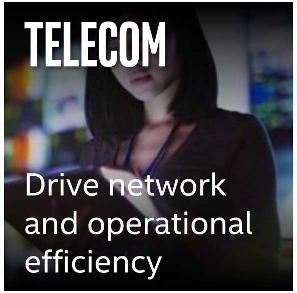
RETAIL

Transform stores and inventory



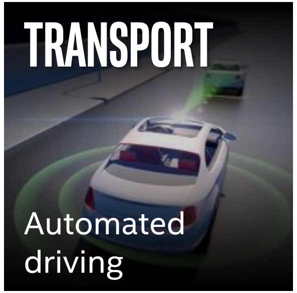
SMART HOME

Enable homes that see, hear, and respond



TELECOM

Drive network and operational efficiency

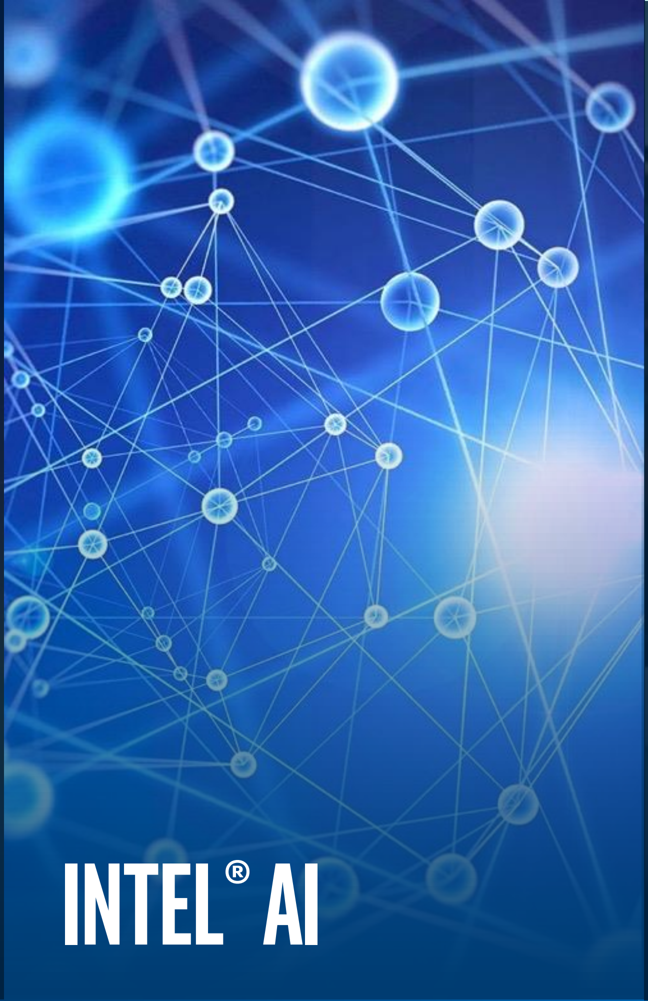


TRANSPORT

Automated driving

OUR PARTNERS ARE DRIVING REAL-WORLD VALUE WITH INTEL® AI

BUSINESS IMPERATIVE



INTEL® AI

AI CASE STUDY

AI Opportunities are Diverse

DEVICE

INTELLIGENT EDGE

MULTI-CLOUD

THINGS

Personal, industrial & community devices

ON PREMISE

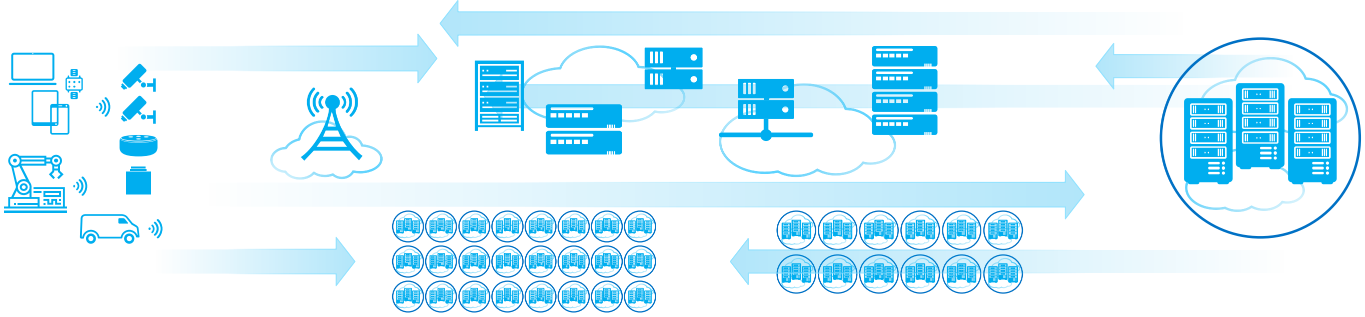
Data analysis & action & gateways

REGIONAL DATA CENTERS

Local private, cloud & telco

PUBLIC/HYBRID/PRIVATE DC

Large scale centralized



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

Intel® AI Strategy

intel® AI

VIBRANT COMMUNITY

Drive innovative use cases

Pioneer leading-edge AI

Fuel the ecosystem

INDUSTRY & OPEN SOFTWARE

Optimize customer software

Unify APIs across Intel

Empower developers

PLATFORM WITH BEST HARDWARE

Extend the CPU

Lead in acceleration

Build a common platform

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

Accelerate Your AI Journey with Intel

DISCOVER
Get started faster with community support

DATA
Tame the deluge with a modern data layer

MODEL
Speed up development with open AI software

DEPLOY
Deliver on the best AI hardware for your needs

COMMUNITY

CONSULT Intel® AI

PARTNER AI Builders
AI In Production

LEARN AI Developer Program

SOFTWARE

DATA MANAGEMENT Choice of 50+ Optimized Tools for Data Preparation

MACHINE LEARNING Intel® DAAL Intel® Distribution for Python* ANALYTICS ZOO

DEEP LEARNING TensorFlow* BigDL* ONNX* OpenVINO™ mxnet* PyTorch* MODEL ZOO & More

HARDWARE

MOVE intel Silicon Photonics intel Omni-Path Fabric intel Ethernet

STORE intel OPTANE DC PERSISTENT MEMORY intel OPTANE DC SOLID STATE DRIVE

PROCESS

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

Get Started Faster

with community support

DISCOVER

CONSULT

intel AI

Consult with your Intel and supplier representative(s) to learn more

Visit: plan.seek.intel.com/SMARTForm_ICS

PARTNER

AI BUILDERS

(Cloud to Device)

AI IN PRODUCTION

(IOT Edge/Device)

Partner with an Intel® AI provider and/or access a catalog with >100 solutions

Visit: builders.intel.com/ai software.intel.com/ai-in-production

LEARN

AI DEVELOPER PROGRAM

Learn AI skills with the FREE¥ Intel® AI Developer Program, including cloud access

Visit: software.intel.com/ai

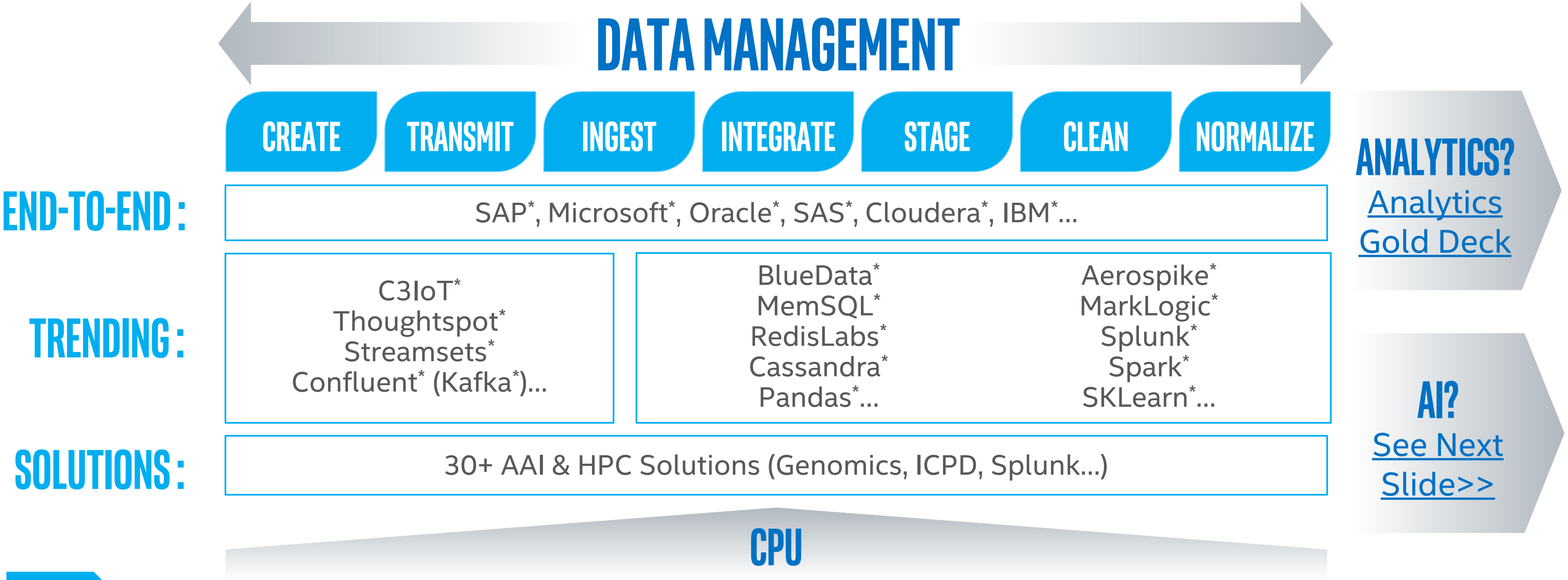
All products, computer systems, dates, and figures are preliminary based on current expectations, and are subject to change without notice.
*Other names and brands may be claimed as the property of others
Optimization Notice

Tame the Deluge

with a modern data layer

See also:
Analytics Gold Deck

DATA



DATA MANAGEMENT

CREATE | TRANSMIT | INGEST | INTEGRATE | STAGE | CLEAN | NORMALIZE

END-TO-END : SAP*, Microsoft*, Oracle*, SAS*, Cloudera*, IBM*...

TRENDING :

C3IoT*
Thoughtspot*
Streamsets*
Confluent* (Kafka*)...

BlueData*
MemSQL*
RedisLabs*
Cassandra*
Pandas*...

Aerospike*
MarkLogic*
Splunk*
Spark*
SKLearn*...

SOLUTIONS : 30+ AAI & HPC Solutions (Genomics, ICPD, Splunk...)

CPU

ANALYTICS?
Analytics Gold Deck

AI?
See Next Slide>>

Visit: *www.intel.com/analytics*

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

Speed Up Development

with open AI software

MODEL

	MACHINE LEARNING	DEEP LEARNING
TOOLKITS App Developers	ANALYTICS ZOO	MODEL ZOO OpenVINO
LIBRARIES Data Scientists	**Intel® Data Analytics Acceleration Library** (DAAL) **Intel® Distribution for Python*** (Sklearn*, Pandas*) **R** (Cart, Random Forest, e1071) **Distributed** (MlLib on Spark, Mahout)	**Frameworks** TensorFlow*, BigDL*, Caffe, ONNX, mxnet, PyTorch *More framework optimizations in progress…* **Intel Tools** NAUTA RL Coach NLP Architect NN Distiller
KERNELS Library Developers	Intel® Math Kernel Library (Intel® MKL)	Intel® Machine Learning Scaling Library (Intel® MLSL) Intel® Math Kernel Library for Deep Neural Networks (Intel® MKL-DNN) nGraph
	CPU	CPU ▪ GPU ▪ FPGA ▪ ACCELERATOR

Visit: www.intel.ai/technology

1 An open source version is available at: 01.org/openvinotoolkit *Other names and brands may be claimed as the property of others.
Developer personas show above represent the primary user base for each row, but are not mutually-exclusive
All products, computer systems, dates, and figures are preliminary based on current expectations, and are subject to change without notice.

Deploy with Unprecedented

AI hardware choice

DEPLOY

MOVE FASTER

INTEL® SILICON PHOTONICS

INTEL® ETHERNET

INTEL® OMNI-PATH FABRIC

STORE MORE

INTEL® OPTANE™ SSD

INTEL® OPTANE™ DC PERSISTENT MEMORY

PROCESS EVERYTHING

CPU

GPU, FPGA

ACCELERATORS

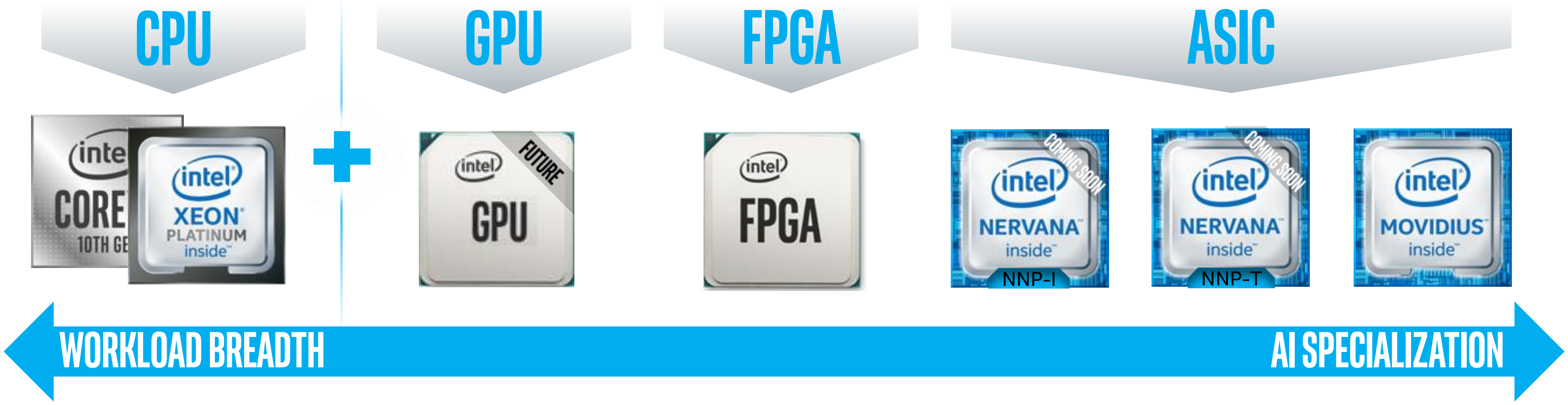
Visit: www.intel.ai/technology

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

Intel® AI Hardware

DEVICE INTELLIGENT EDGE MULTI-CLOUD

OPTIMIZED FRAMEWORKS & SOFTWARE



CPU

GPU

FPGA

ASIC

WORKLOAD BREADTH

AI SPECIALIZATION

Multi-Purpose Foundation for AI

Data-Parallel Media, Graphics, HPC & AI

Multi-Function & Real-time Deep Learning Inference

Deep Learning Inference

Deep Learning Training

Media & Vision DL Inference at the Edge

Visit:

All products, computer systems, dates, and figures are preliminary based on current expectations, and are subject to change without notice.
1Unified software stack development in progress DL=Deep Learning

Intel® AI Use Cases

CPU

Intel® Xeon® Scalable Processors

MULTI-CLOUD	JD.com*	**HYBRID ANALYTICS + AI** Fast time-to-solution on Spark* with MlLib & BigDL
	CERN*	**HPC AND AI** Fast time-to-solution for deep learning in classic workflows
	Novartis*	**LARGE DL TRAINING** Fast DL training for large image recognition in drug discovery
	Taboola*	**DEEP LEARNING INFERENCE** High throughput real-time recommendation (billion items)
	Ziva*	**MACHINE LEARNING** Animating movie creatures using machine learning techniques

CPU

Intel® Xeon® Scalable Processors

INTELLIGENT EDGE	GE Health*	**DEEP LEARNING INFERENCE** Low TCO for image recognition in CT scanner for radiology

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

Intel® AI Use Cases

FPGA

Intel® FPGA

MULTI-CLOUD	**Microsoft***	**REAL-TIME REC. ENGINE** Real-time recommendations and more workload acceleration
	Manjeera*	**REAL-TIME TRANSCRIPTS** Real-time transcription acceleration
	JD.com*	**TEXT RECOGNITION** Faster time-to-market for custom CNN & LSTM for end-to-end text recognition
INTELLIGENT EDGE	**QNAP***	**VISION INFERENCE** Faster time-to-market for custom CNN workload with OpenVINO™ toolkit
	NEC*	**FACE RECOGNITION** Faster time-to-market for custom CNN workload for surveillance and retail
	Alibaba*	**REAL-TIME VISION** Real-time video encoding and decoding for smart city project

ASIC

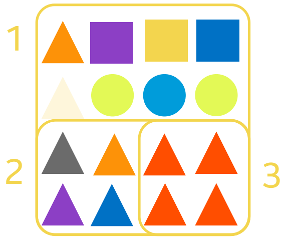
Intel® Movidius™ Myriad™ X VPU

INTELLIGENT EDGE	**HPE***	**VISION AT THE EDGE** Video analytics and DL inference in an edge server blade
DEVICE	**Hikvision***	**VISION IN THE DEVICE** Deep learning-based computer vision at low power

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

AI Compute Considerations

WORKLOADS

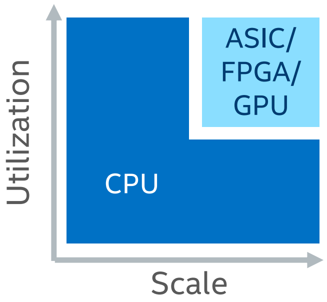


What is my workload profile?

REQUIREMENTS

What are my use case requirements?

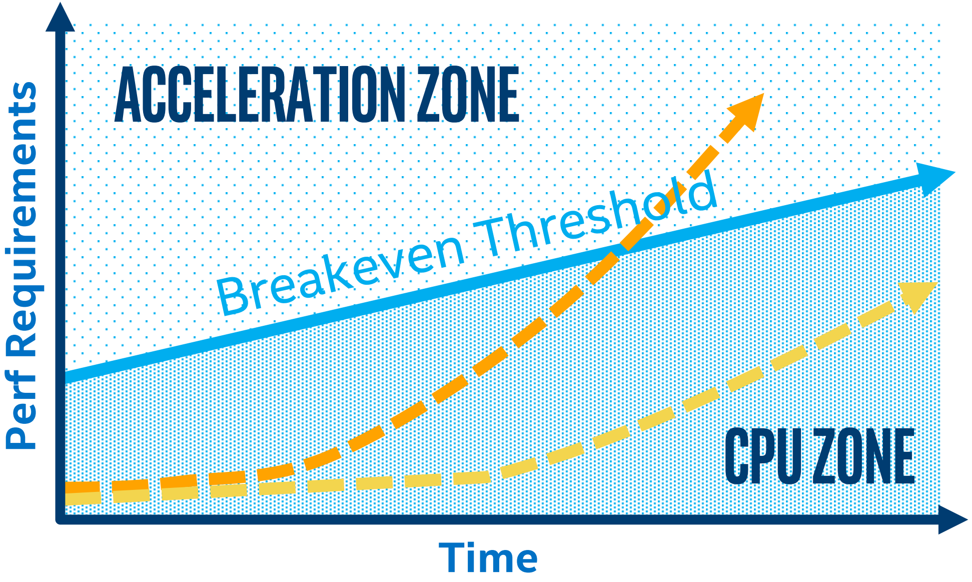
DEMAND



How prevalent is AI in my environment?

Note: word cloud source is www.wordart.com
¥Free = available to download/access at no cost to qualified developers who are enrolled in the program
*Other names and brands may be claimed as the property of others.

Bust the Deep Learning Myth



"A GPU is required for deep learning…"

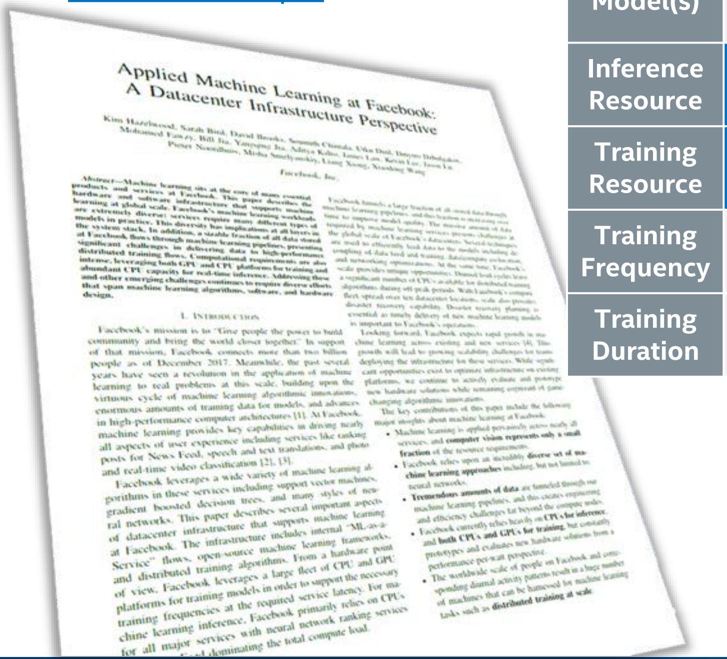
FALSE

- Most enterprises (---) use <u>CPU</u> for machine and deep learning needs
- Some early adopters (---) may reach a deep learning tipping point when acceleration is needed[1]

[1] "Most" of enterprise customers based on survey of Intel direct engagements and internal market segment analysis

Deep Learning Use Case

Source Paper:
research.fb.com/
wpcontent/uploads/2017/12/hpca-
2018-facebook.pdf



Applied Machine Learning at Facebook:
A Datacenter Infrastructure Perspective

Kim Hazelwood, Sarah Bird, David Brooks, Soumith Chintala, Utku Diril, Dmytro Dzhulgakov,
Mohamed Fawzy, Bill Jia, Yangqing Jia, Aditya Kalro, James Law, Kevin Lee, Jason Lu,
Pieter Noordhuis, Misha Smelyanskiy, Liang Xiong, Xiaodong Wang

Facebook, Inc.

Abstract—Machine learning sits at the core of many essential products and services at Facebook. This paper describes the hardware and software infrastructure that supports machine learning at global scale. Facebook's machine learning workloads are extremely diverse: services require many different types of models in practice. This diversity has implications at all layers in the system stack. In addition, a sizable fraction of all data stored at Facebook flows through machine learning pipelines, presenting significant challenges in delivering data to high-performance distributed training flows. Computational requirements are also intense, leveraging both GPU and CPU platforms for training and abundant CPU capacity for real-time inference. Addressing these and other emerging challenges continues to require diverse efforts that span machine learning algorithms, software, and hardware design.

I. Introduction

Facebook's mission is to "Give people the power to build community and bring the world closer together." In support of that mission, Facebook connects more than two billion people as of December 2017. Meanwhile, the past several years have seen a revolution in the application of machine learning to real problems at this scale, building upon the virtuous cycle of machine learning algorithmic innovations, enormous amounts of training data for models, and advances in high-performance computer architectures [1]. At Facebook, machine learning provides key capabilities in driving nearly all aspects of user experience including services like ranking posts for News Feed, speech and text translations, and photo and real-time video classification [2], [3].

Facebook leverages a wide variety of machine learning algorithms in these services including support vector machines, gradient boosted decision trees, and many styles of neural networks. This paper describes several important aspects of datacenter infrastructure that supports machine learning at Facebook. The infrastructure includes internal "ML-as-a-Service" flows, open-source machine learning frameworks, and distributed training algorithms. From a hardware point of view, Facebook leverages a large fleet of CPU and GPU platforms for training models in order to support the necessary training frequencies at the required service latency. For machine learning inference, Facebook primarily relies on CPUs for all major services with neural network ranking services [illegible] dominating the total compute load.

Facebook funnels a large fraction of all stored data through machine learning pipelines, and this fraction is increasing over time to improve model quality. The massive amount of data required by machine learning services presents challenges at the global scale of Facebook's datacenters. Several techniques are used to efficiently feed data to the models including decoupling of data feed and training, data/compute co-location, and networking optimizations. At the same time, Facebook's scale provides unique opportunities. Diurnal load cycles leave a significant number of CPUs available for distributed training algorithms during off-peak periods. With Facebook's compute fleet spread over ten datacenter locations, scale also provides disaster recovery capability. Disaster recovery planning is essential as timely delivery of new machine learning models is important to Facebook's operations.

Looking forward, Facebook expects rapid growth in machine learning across existing and new services [4]. This growth will lead to growing scalability challenges for teams deploying the infrastructure for these services. While significant opportunities exist to optimize infrastructure on existing platforms, we continue to actively evaluate and prototype new hardware solutions while remaining cognizant of game changing algorithmic innovations.

The key contributions of this paper include the following major insights about machine learning at Facebook:

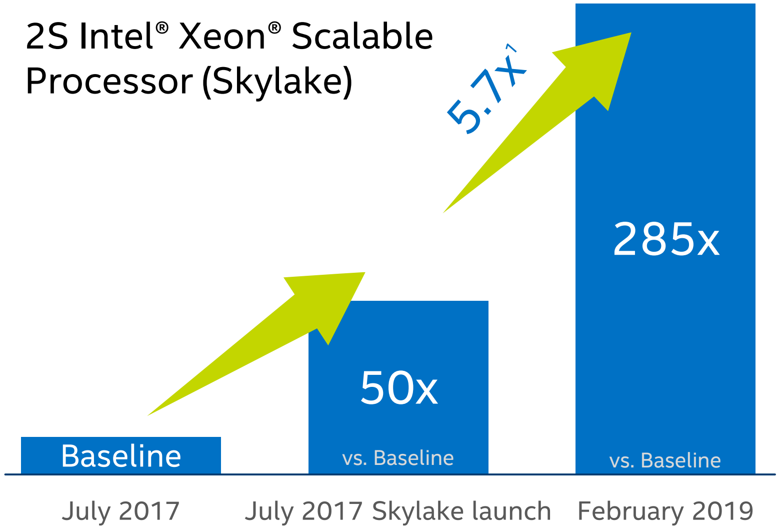
- Machine learning is applied pervasively across nearly all services, and **computer vision represents only a small fraction** of the resource requirements.
- Facebook relies upon an incredibly **diverse set of machine learning approaches** including, but not limited to, neural networks.
- **Tremendous amounts of data** are funneled through our machine learning pipelines, and this creates engineering and efficiency challenges far beyond the compute nodes.
- Facebook currently relies heavily on **CPUs for inference**, and **both CPUs and GPUs for training**, but constantly prototypes and evaluates new hardware solutions from a performance-per-watt perspective.
- The worldwide scale of people on Facebook and corresponding diurnal activity patterns result in a huge number of machines that can be harnessed for machine learning tasks such as **distributed training at scale**.

Services	Ranking Algorithm	Photo Tagging	Photo Text Generation	Search	Language Translation	Spam Flagging	Speech
Model(s)	MLP	SVM,CNN	CNN	MLP	RNN	GBDT	RNN
Inference Resource	CPU	CPU	CPU	CPU	CPU	CPU	CPU
Training Resource	CPU	GPU & CPU	GPU	Depends	GPU	CPU	GPU
Training Frequency	Daily	Every N photos	Multi-Monthly	Hourly	Weekly	Sub-Daily	Weekly
Training Duration	Many Hours	Few Seconds	Many Hours	Few Hours	Days	Few Hours	Many Hours

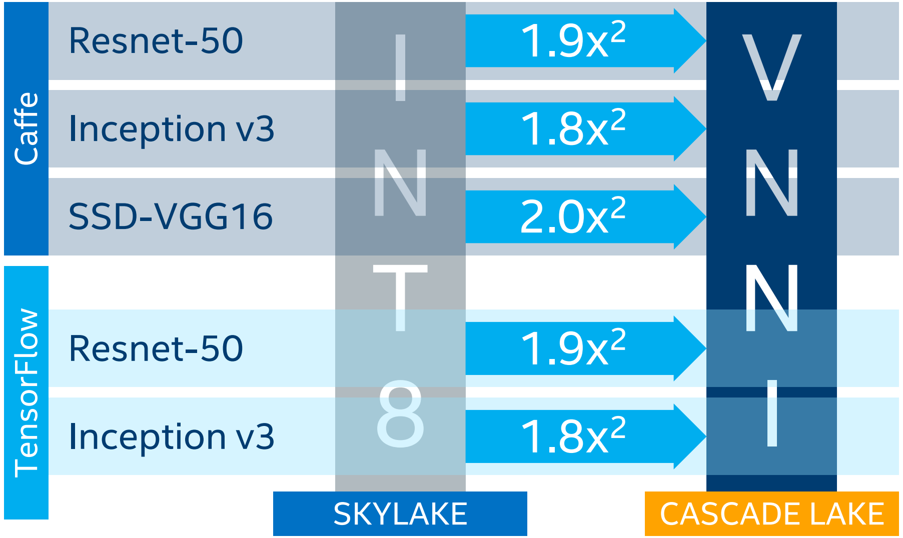
LARGE CLOUD USERS EMPLOY CPU EXTENSIVELY FOR DEEP LEARNING

Deep Learning Performance on CPU

Hardware + software improvements for Intel® Xeon® processors



2S Intel® Xeon® Scalable Processor (Skylake)



[1] 5.7x inference throughput improvement with Intel® Optimizations for Caffe ResNet-50 on Intel® Xeon® Platinum 8180 Processor in Feb 2019 compared to performance at launch in July 2017. See configuration details on Config 1

Performance results are based on testing as of dates shown in configuration and may not reflect all publicly available security updates. 28/24/2018) Results have been estimated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance. 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. 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 Optimization Notice

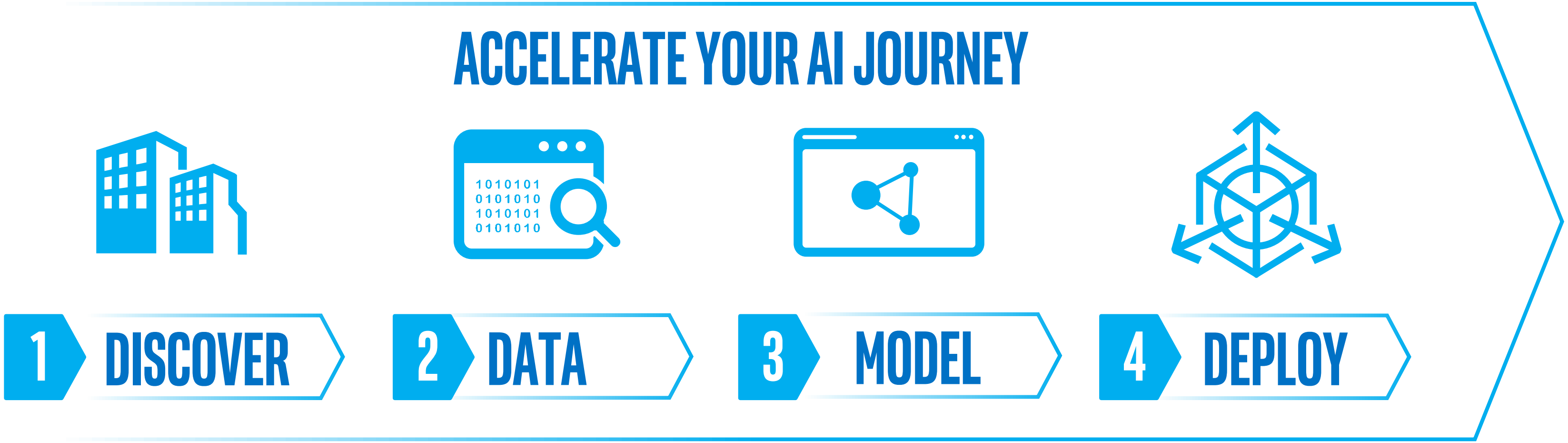
BUSINESS IMPERATIVE

INTEL® AI

AI CASE STUDY

Intel® AI Case Study

intel AI



ACCELERATE YOUR AI JOURNEY

1 DISCOVER

2 DATA

3 MODEL

4 DEPLOY

Intel® AI Case Study

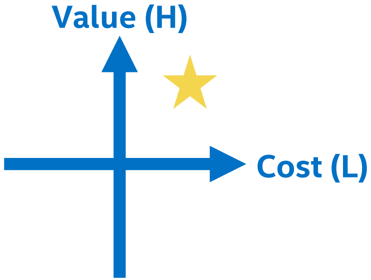
DISCOVER

IDENTIFY

Identify prospects internally and using the 70+ AI solutions in Intel's portfolio; then assess business value of each one

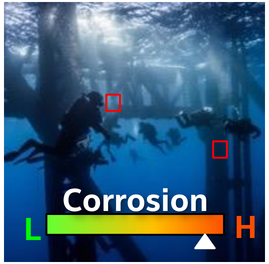
PRIORITIZE

Prioritize projects based on business value and cost to solve with Intel guidance; choose industrial defect detection via DL[1]



CONSIDER

Consider ethical, social, legal, security and other risks and mitigation plans with Intel advisors prior to kickoff



ORGANIZE

Organize internally to get buy-in, support new development philosophy and grow developer talent via Intel® AI

intel AI DEVELOPER PROGRAM

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

Intel® AI Case Study

DATA

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

INGEST

Ingest streaming data from drones using a popular software tool among the many that run on the CPU

STORE

Store data in block storage (for high-performance) in a data lake with guidance from an Intel storage partner

PREPARE

Prepare data by performing cleanup and integration using popular software tools that run on the CPU

ACT

Act on the data using one of the many popular CPU tools for data analytics and visualization

← **12 weeks** →

Source Data	Transmit Data	Ingest Data	Clean up Data	Integrate Data	Stage Data
10%	*5%*	*5%*	*60%*	*10%*	*10%*

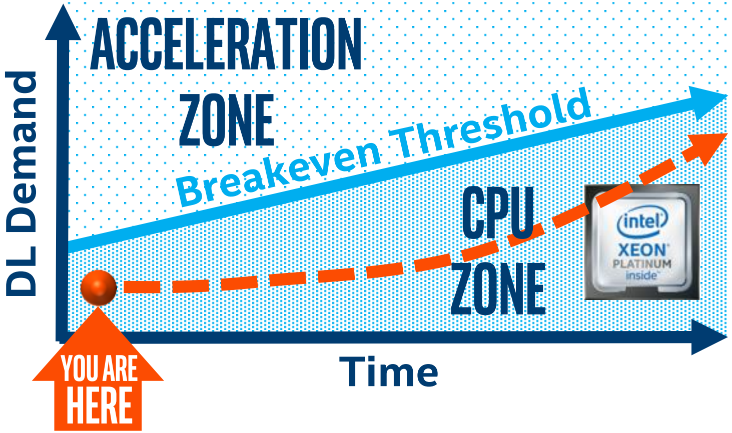
All products, computer systems, dates, and figures are preliminary based on current expectations, and are subject to change without notice.
*Other names and brands may be claimed as the property of others Optimization Notice

Intel® AI Case Study

MODEL

SETUP

Set up compute environment; DL training (~7% of journey) acceleration NOT worthwhile due to high setup time & cost



MODEL

Model development through training a deep neural network using an Intel-optimized DL framework

TEST

Test the deep learning model using a control data set to determine if accuracy meets requirements

DOCUMENT

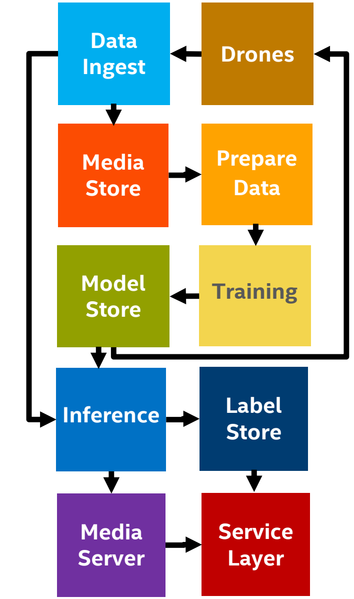
Document the code, process, and key learnings for future reference

12 weeks

Train *(Topology Experiments)*	Train *(Tune Hyper-parameters)*	Test Inference	Document Results
30%	*30%*	*20%*	*10%*

Intel® AI Case Study

DEPLOY



ARCHITECT

Architect AI deployment with Intel® AI Builders

IMPLEMENT

Implement AI in production environment

SCALE

Scale to more sites and users as demand grows

ITERATE

Iterate on the models with new data over time

REMOTE DEVICES

Drone
Drone
Drone

10 Drones

Real-time object detection and data collection

Drone
Drone
Drone

Per Drone

1x Intel® Core™ processor
1x Intel® Movidius™ VPU

MEDIA SERVER

Media Store
Media Store
Media Store

110 Nodes

8 TB/day per camera

10 cameras

3x replication

1-year retention

4 mgmt nodes

Media Store
Media Store
Media Store

Per Node 1x 2S 61xx 20x 4TB SSD

MULTI-USE CLUSTER

Data Ingestion
Data Ingestion
Data Ingestion
Data Ingestion

4 Nodes
One ingestion per day, one-day retention

Inference
Inference
Inference
Inference

4 Nodes
20M frames per day

Prepare Data
Prepare Data

2 Nodes
Infrequent op

Service Layer
Service Layer
Service Layer

3 Nodes
Simultaneous users

Media Server
Media Server
Media Server

3 Nodes
10k clips stored

DATA STORE

Model Store
Model Store
Model Store
Model Store

4 Nodes
1-year of history

Label Store
Label Store
Label Store
Label Store

4 Nodes
Labels for 20M frames /day

Per Node

1x 2S 81xx
5x 4TB SSD

ADV. ANALYTICS

Training

16 Nodes
Intermittent use
1 training/month for <10 hours

Training

Per Node

1x 2S 81xx
1x 4TB SSD

SOFTWARE

- ➢ *OpenVino™* Toolkit
- ➢ *Intel® MKL-DNN*
- ➢ *TensorFlow**
- ➢ *Intel® Movidius™ Software Development Toolkit*

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

THANK YOU