

AI Everywhere: Simplicity, Productivity & Performance

Presenters Lakshman Chari & Emmanuel Rochette

intel®



AI Everywhere:

Simplicity, Productivity & Performance

Push to Start AI
200+ turnkey solutions & providers to choose from

Solutions

Intel Solutions Marketplace



Build Smarter, Faster
150+ containers to streamline end-to-end data science

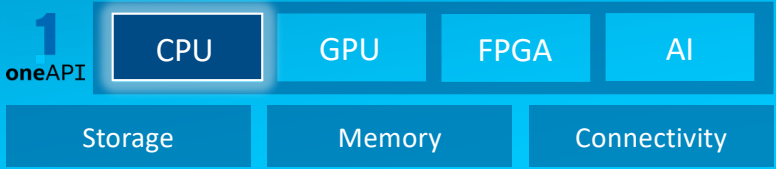
Tools

TensorFlow PyTorch mxnet OpenVINO ONNX RUNTIME

Scikit-Learn Pandas NumPy/SciPy XGBoost & More

Accelerate Your AI, Today
1.5x perf vs. AMD and 1.3x vs. Nvidia across 20 top AI workloads

Technology



1 oneAPI

CPU GPU FPGA AI

Storage Memory Connectivity

See claims [43, 44] at www.intel.com/3gen-xeon-config for workloads and configurations. Results may vary.

Flexible AI Acceleration

CPU *only*

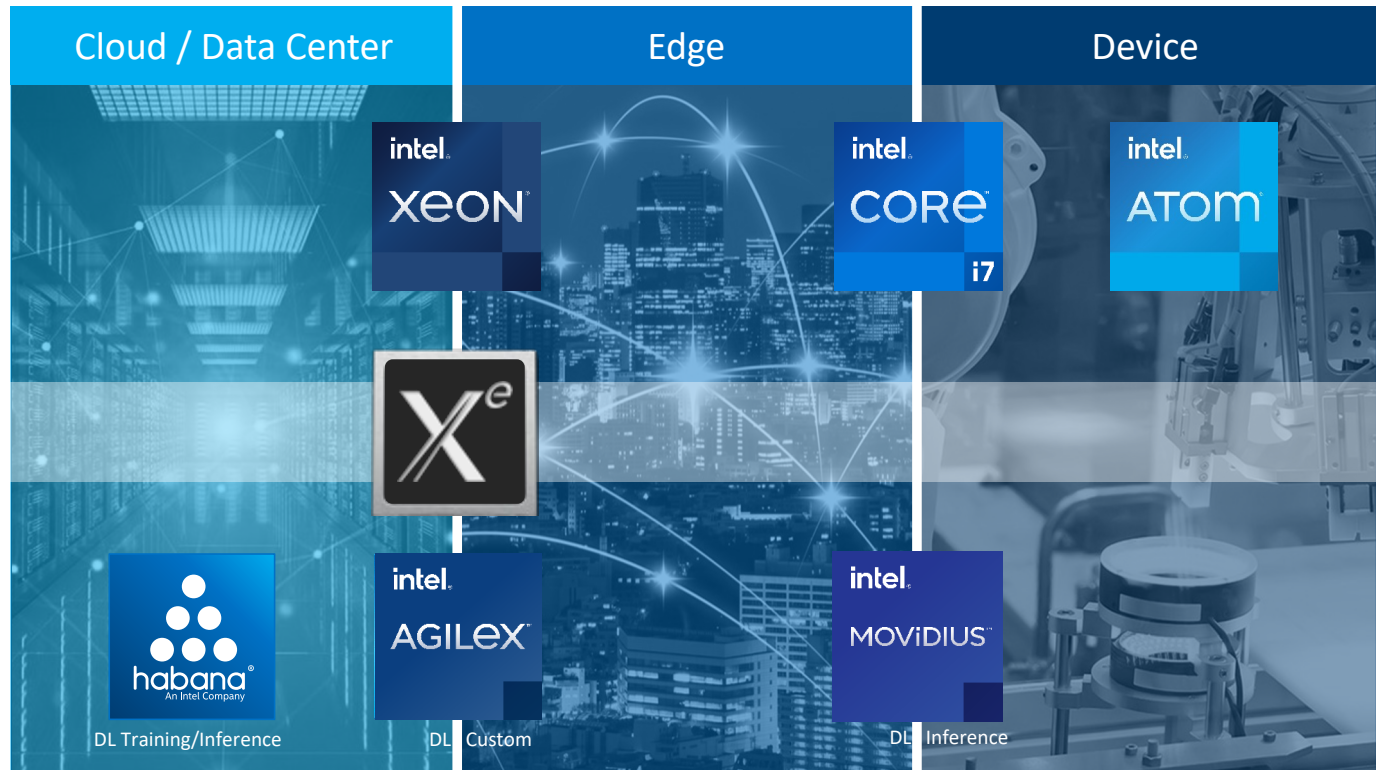
Built-in AI acceleration for mainstream AI use cases

CPU + GPU

When compute is dominated by AI, HPC, graphics, and/or real-time media

CPU + custom

When compute is dominated by deep learning (DL)



AI Everywhere with Intel Software

Scalable Data + AI Platform

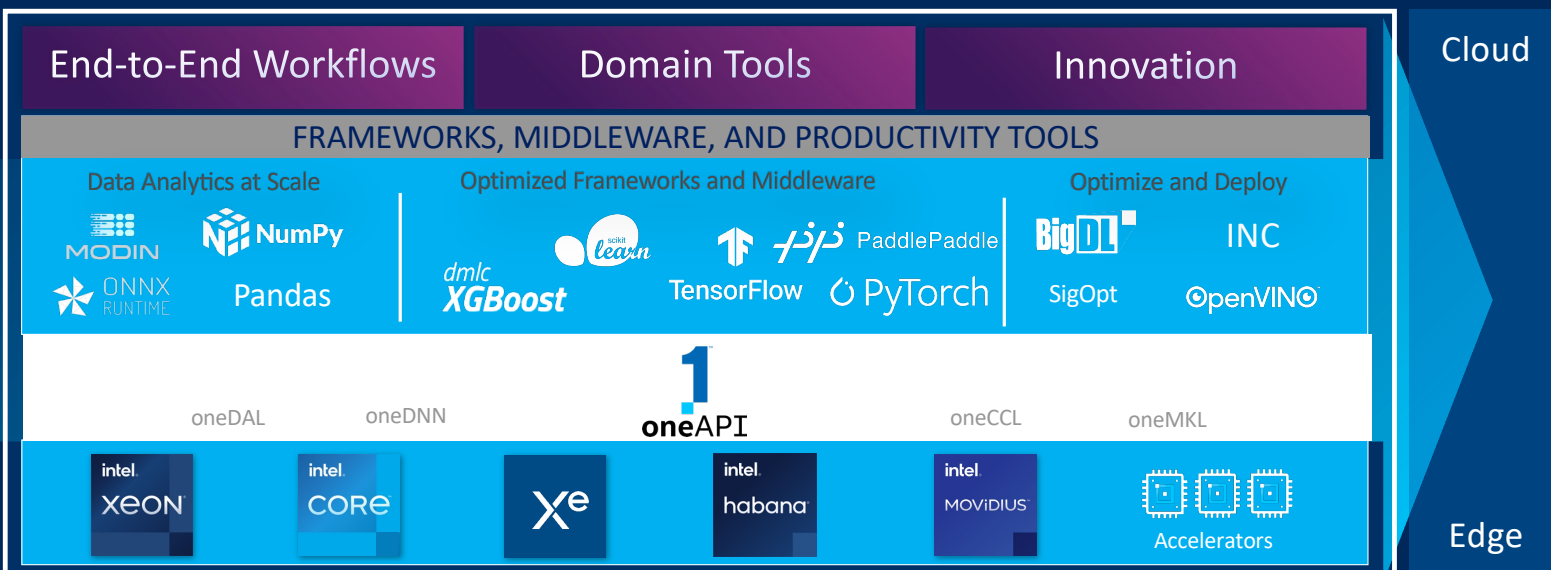
Data

Model

Deploy

Top-of-Stack


Foundation

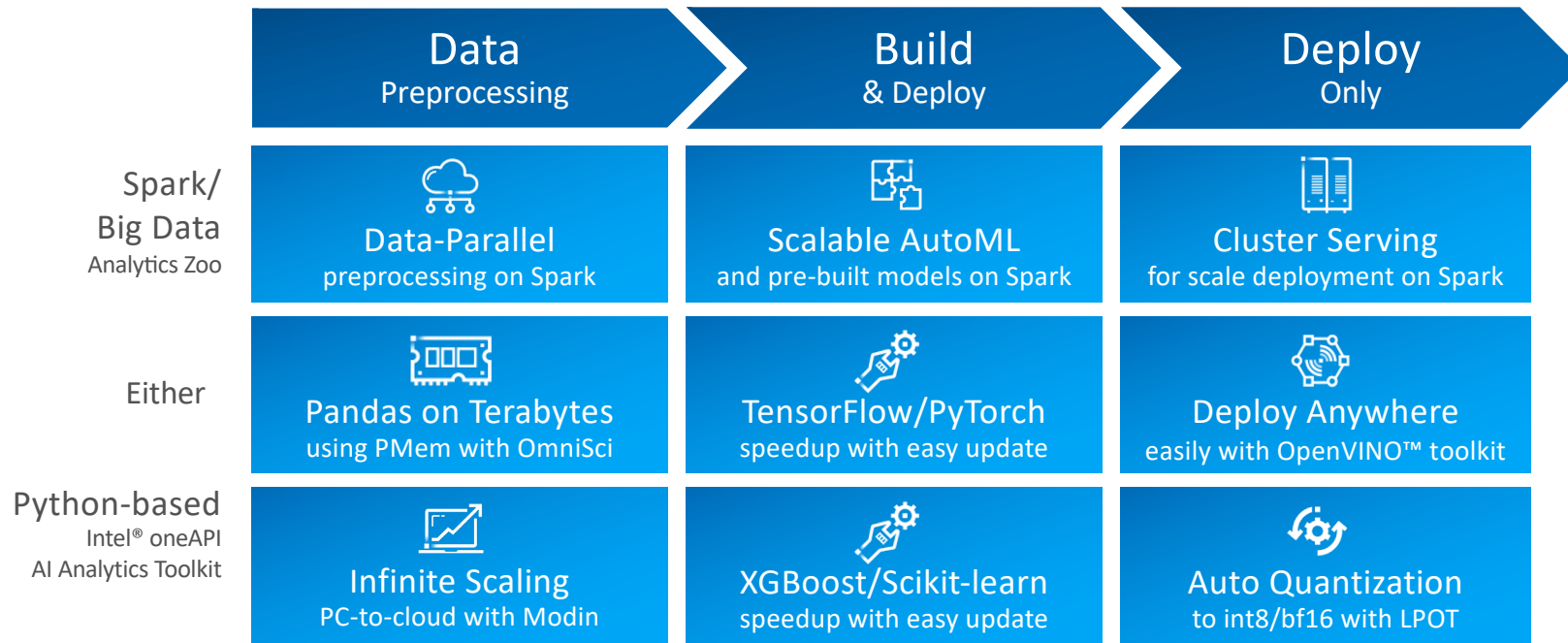


A consistent SW platform accelerating HW innovation

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

Build Smarter, Faster

Powered by 



Popular tools optimized, 3 Intel toolkits, 150+ containers

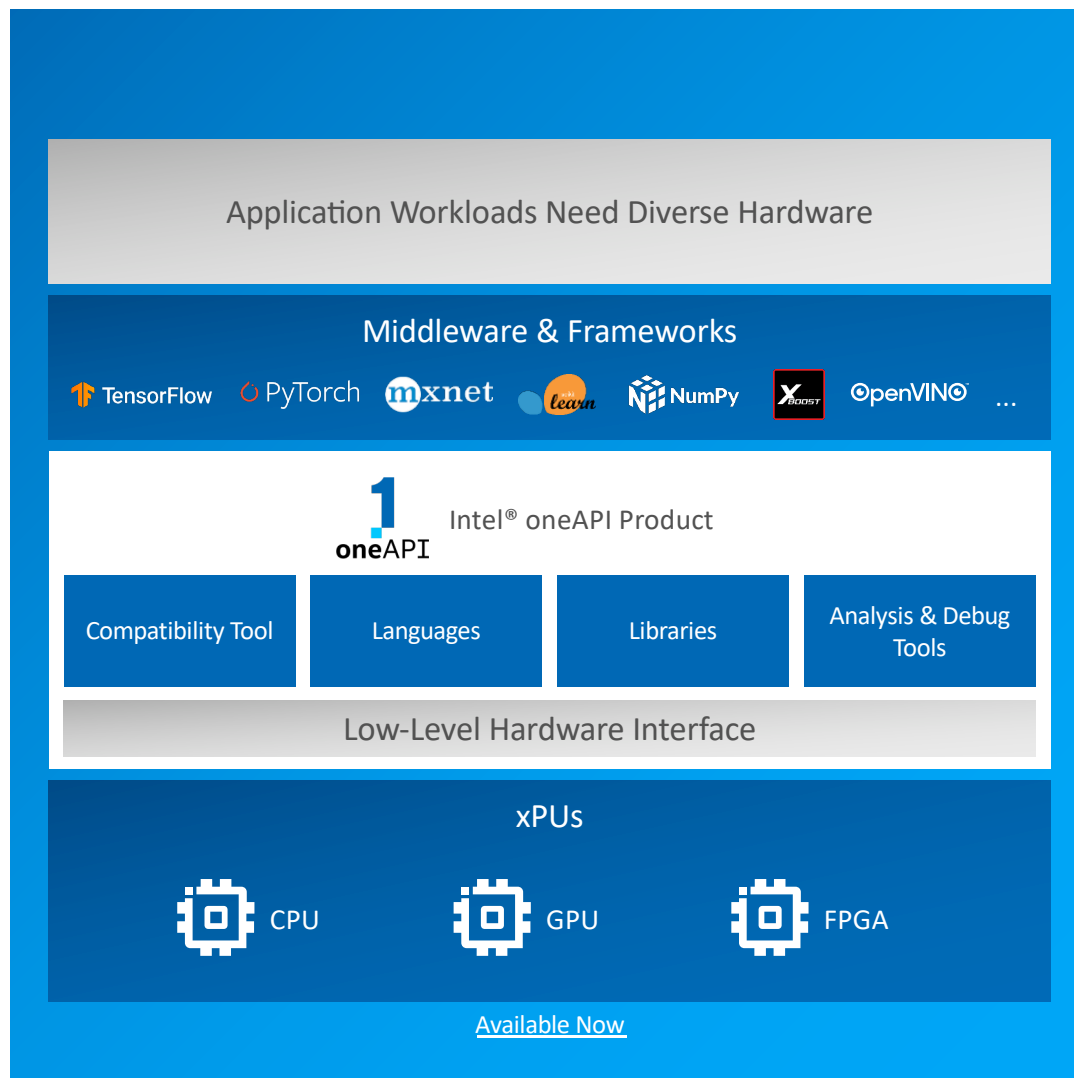
Intel® oneAPI Product

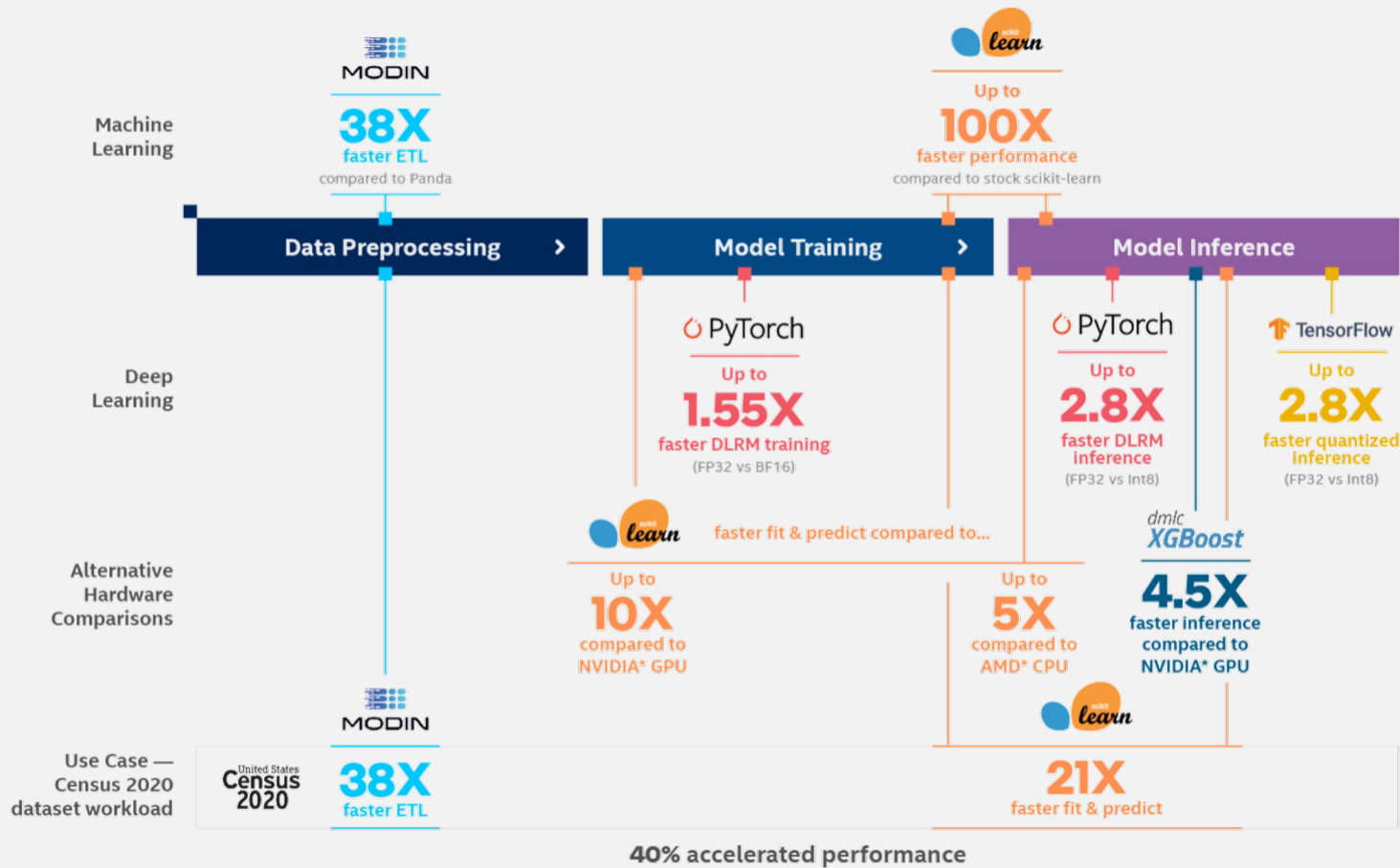
Built on Intel's Rich Heritage of CPU Tools Expanded to xPUs

A complete set of advanced compilers, libraries, and porting, analysis and debugger tools

- Accelerates compute by exploiting cutting-edge hardware features
- Interoperable with existing programming models and code bases (C++, Fortran, Python, OpenMP, etc.), developers can be confident that existing applications work seamlessly with oneAPI
- Eases transitions to new systems and accelerators—using a single code base frees developers to invest more time on innovation

Visit software.intel.com/oneapi for more details
Some capabilities may differ per architecture and custom-tuning will still be required. Other accelerators to be supported in the future





Performance improvements are based on hardware running on 2nd Gen Intel® Xeon Scalable processors. See slides at the end of the video for workloads and configurations. Results may vary.



2018 to 2022: 4 Years of Inferencing

1

BUILD

TensorFlow Caffe mxnet ONNX
KALDI PyTorch PaddlePaddle

NCF, Open Model Zoo & Training Extensions

2

OPTIMIZE

Deep Learning Workbench

Model Optimizer

Post-Training Optimization (POT)

3

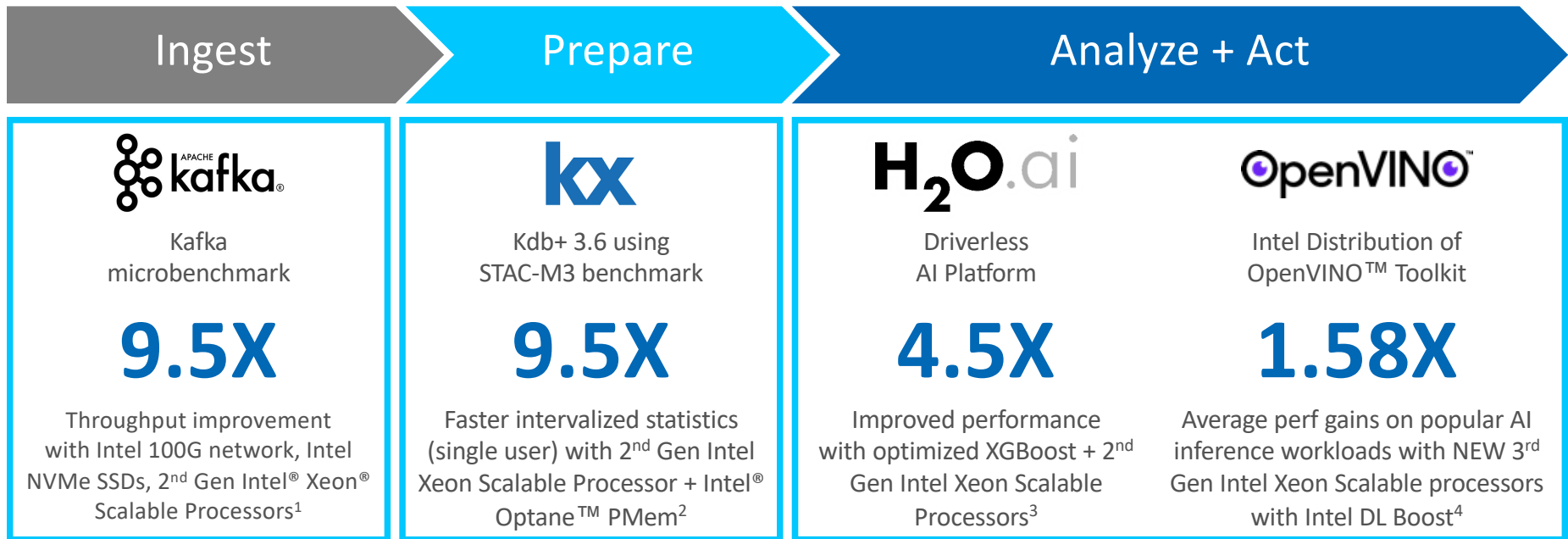
DEPLOY

Model Server (OVMS)

intel ATOM intel CORE i7 intel XEON
intel MOVIDIUS intel IRIS MAX GRAPHICS Intel GNA (IP)

Singular focus: help developers “Write Once, Deploy Anywhere”

Intel: Gold Standard for Pipeline Modernization



Intel + partners: validated and proven solutions across the data pipeline

1-4: For detailed configurations see back up slide 22. For more complete information about performance and benchmark results, visit www.intel.com/benchmarks.

SigOpt: Software that helps you build the best models

The most advanced model experimentation solution

- Automate optimization of data, features, models
- Empower experts to build world-class models
- Faster, better, cheaper model development

Built by experts, for experts

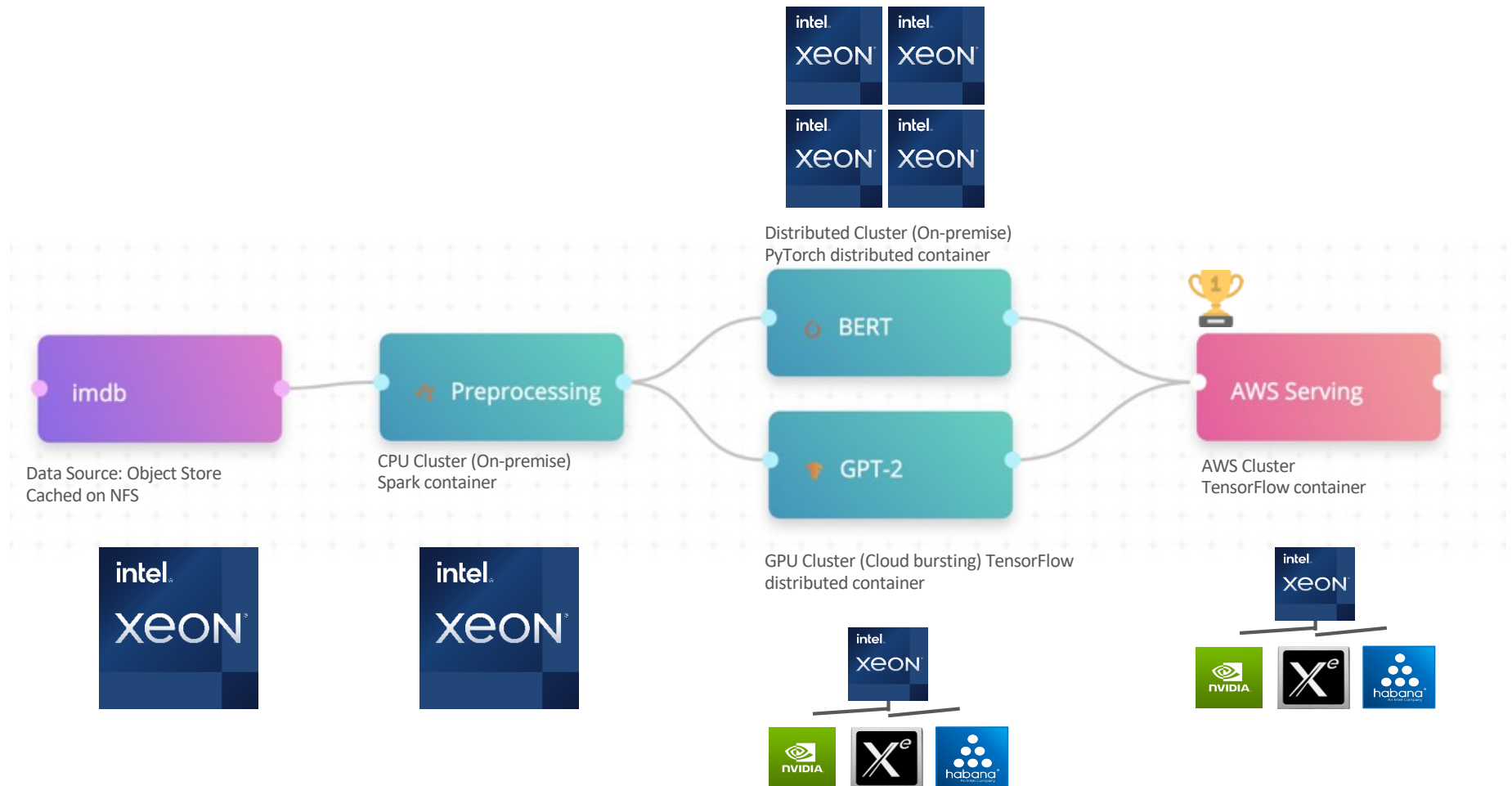
- Platform engineers from Facebook, Amazon
- Optimization experts who created MOE, BayesOpt
- ML practitioners from Google, Yelp, Amazon

Supported and used by AI leaders

- Enterprise use by Netflix, Two Sigma, Amex, and more
- Academic use by MIT, Harvard, CMU, and more
- Funded by a16z, YC, DCVC, IQT, Blumberg, SV Angel
- Acquired by Intel November 2020



Construct and deploy end-to-end pipelines with cnvrg.io





Key Takeaways

Always remember our key One Intel AI message: Intel is the fast path to scale AI everywhere.

Our three pillars — Readiness, Optimization, and Flexibility — form the foundation of our success in bringing AI to scale worldwide.

Explore more about Intel AI on our website or listen to our podcast ***Intel on AI.***