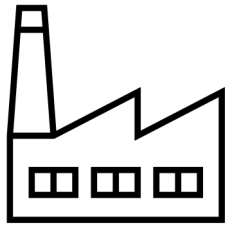
Introducing OpenVINO™ integration with TensorFlow

OpenVINO™ integration with TensorFlow

AI Use Cases across various domains



Manufacturing

Visual Inspection

Automated inspection of Personal Protective Equipment (PPE), such as Mask Detection or Helmet Detection

Healthcare

Cancer Analysis

Detect slight differences between cancerous and non-cancerous images and diagnose data from magnetic resonance imaging (MRI) scans

AI Use Cases across various domains

Agriculture

Animal Monitoring

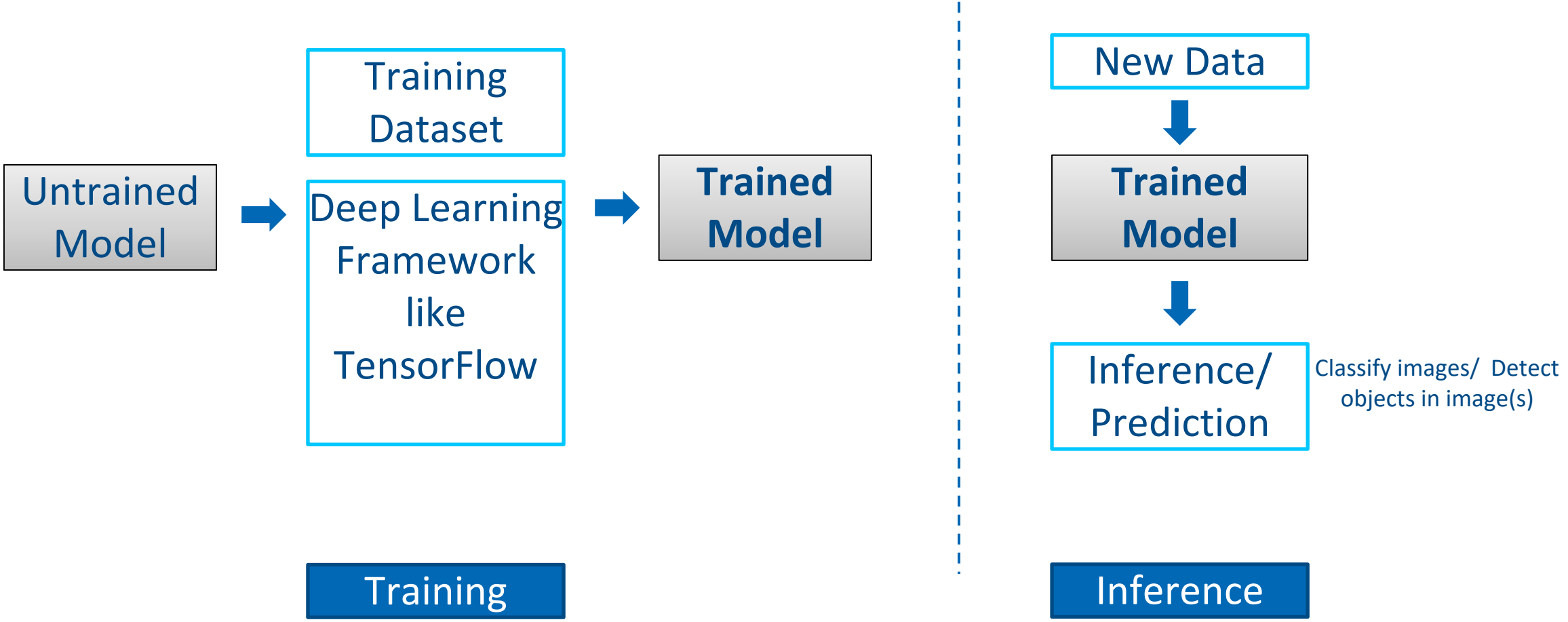
Monitor count, behavior and health of specific livestock such as pigs, cattle, or poultry

Retail

Inventory Monitoring

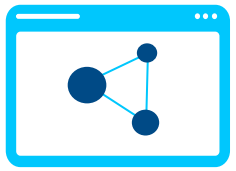
Monitoring store shelves and warehouses to prevent stock out situations

Training vs Inference



Challenges in Deep Learning

Development and deployment challenges in deep learning



Unique Inference Needs

Gap in performance and accuracy between trained and deployed models

Low performing, lower accuracy models deployed

Integration Challenges

No streamlined way for end-to-end development workflow

Slow time-to-solution and time–to-market

No One Size Fits All

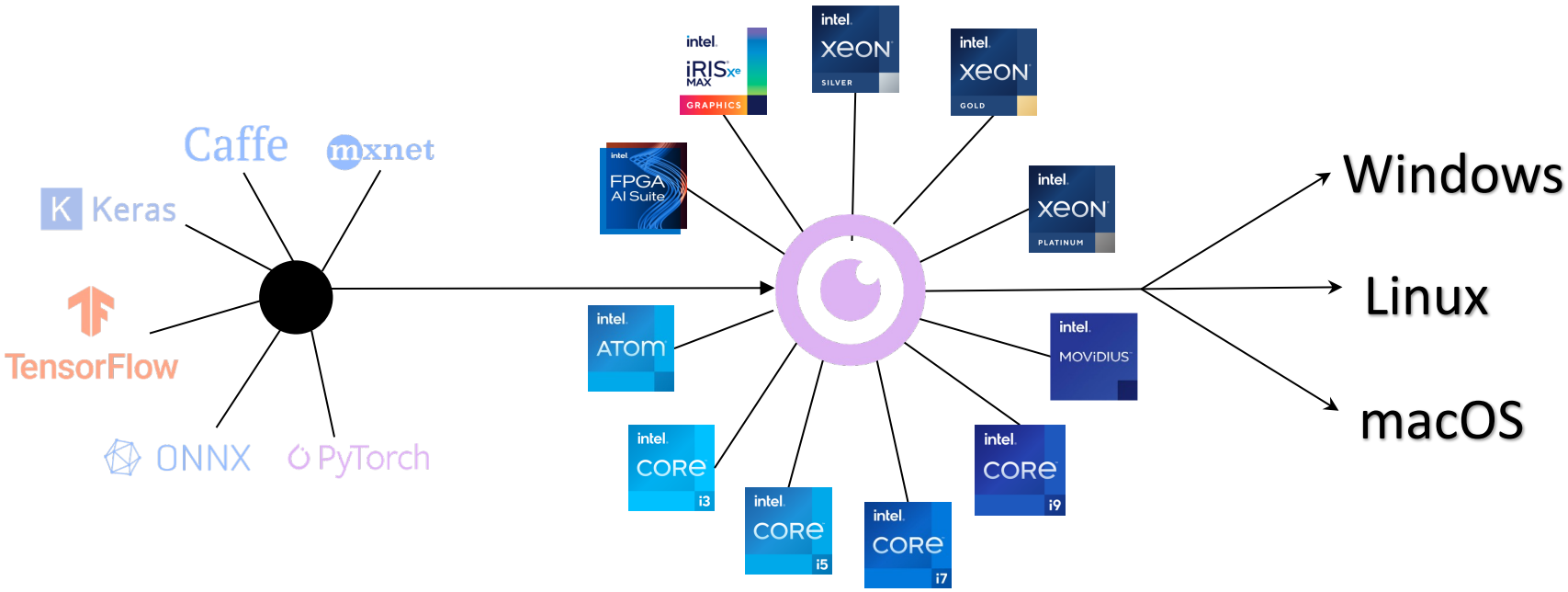
Diverse requirements for myriad use cases require unique approaches

Inability to meet use-case specific requirements

Intel® Distribution of OpenVINO™ toolkit

Fast, accurate real-world results with high-performance, deep learning inference

Convert and optimize models, deploy across a mix Intel hardware and environments, on-premise and on-device, in the browser or in the cloud



Typical Workflow

From development to deployment

TensorFlow

Train a model

Find a trained model

Run the Model Optimizer

- A **Python*** based tool to **import** trained models and **convert** them to Intermediate Representation

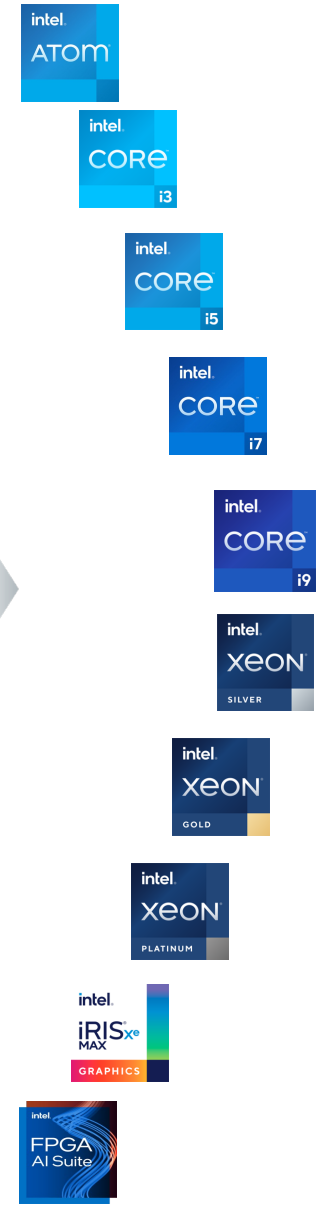
Intermediate Representation

.bin, .xml

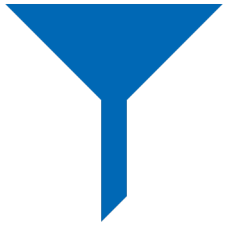
- **.xml** file holds topology of a model
- **.bin** file contains the weights and bias of a model.
- Can be quantized into **INT8** with **POT**

Deploy using the Inference Engine

- runtime **API** implemented as dynamically loaded **plugins** for each hardware type

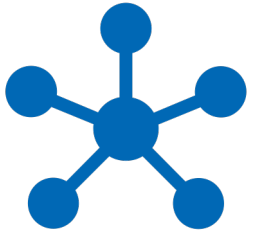


Learning Curves of Adapting to the Typical Workflow



Offline model conversion.

TF developers who don’t want to learn new API.



Limited model coverage.

OpenVINO™ integration with TensorFlow

What is it?

A product that **enables OpenVINO™ acceleration** with **TensorFlow APIs**

2 lines of code in existing TensorFlow applications to access CPU, iGPU, VPU

What are the benefits?

Inline conversion of TensorFlow models

Immediate acceleration of TensorFlow apps with OpenVINO™ Toolkit

Why do we need it?

Ability to **execute more inference workloads** on Intel hardware

Improve **OpenVINO™ Toolkit TensorFlow model coverage** with **TF runtime fallback feature**

```
1  # Installation steps
2  # more details : https://github.com/openvinotoolkit/openvino_tensorflow
3  #pip3 install -U pip
4  #pip3 install -U tensorflow==2.x.x
5  #pip3 install openvino-tensorflow
6
7  # Import package and set backend
8  import openvino_tensorflow
9  openvino_tensorflow.set_backend('GPU')
10
11 # Load a TF Saved Model
12 model = tf.keras.models.load_model('resnet50_saved_model')
13
14 # Get the input size of the model
15 network_input_size = saved_model_loaded.input.shape()
16
17 # Resize the input image
18 resized_image = resize(input_image, network_input_size)
19
20 # Run inference
21 model.predict(resized_image)
```

GitHub

Line 1

Line 2

CPU
GPU
MYRIAD
VAD-M

Existing TensorFlow code + 2 lines of code
= Boosted model with accelerated inferencing on Intel hardware

Why OpenVINO™ integration with TensorFlow

High-Performance,
Deep Learning Inference

Streamlined Development, Ease of Use

Write Once,
Deploy Anywhere

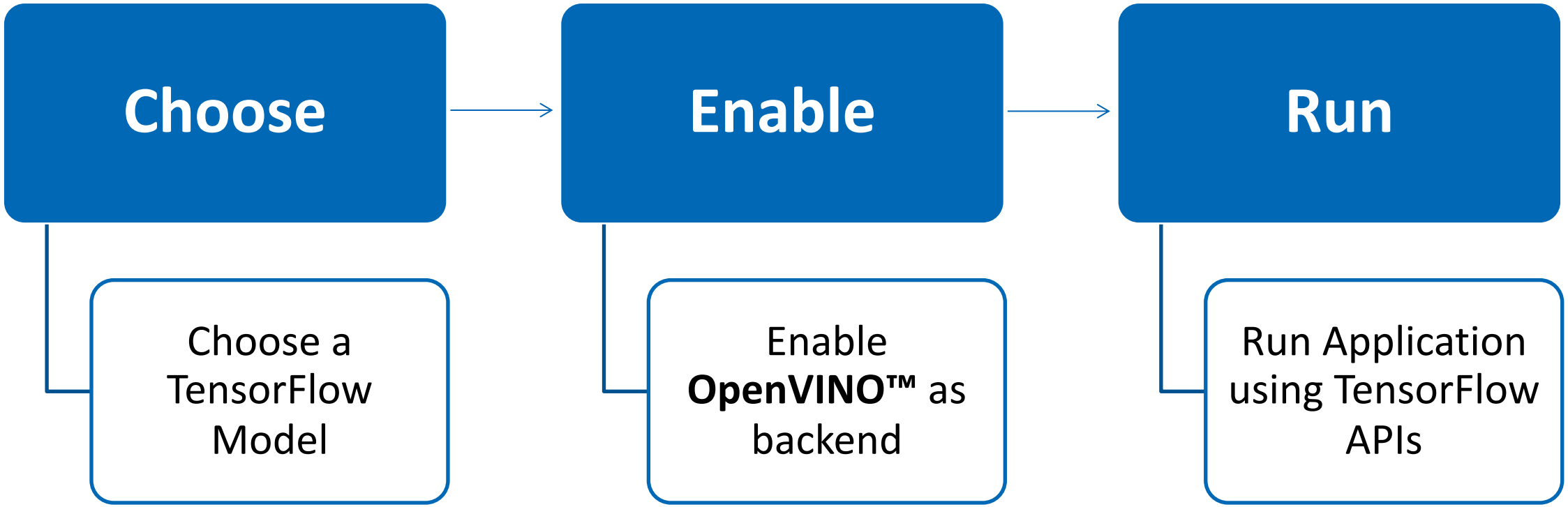
For amazing performance:
Use OpenVINO™ **Native APIs** (Model Optimizer/IR Format)

For amazing compatibility:
Use OpenVINO™ **integration with TensorFlow**

OpenVINO™ integration with TensorFlow focus:

- TensorFlow developers using native TensorFlow APIs to try out OpenVINO™ Toolkit
- Providing acceleration for all TF models (e.g. TFHub)
- Easy onboard TF developers – pip install + 2 lines of code in TF applications
- Willingness to accommodate slightly less acceleration/HW control vs native OpenVINO™ APIs

Workflow of OpenVINO™ integration with TensorFlow



Workflow of OpenVINO™ integration with TensorFlow

PICK a TensorFlow Model → **ENABLE** OpenVINO™ as backend → **RUN** Application using TensorFlow APIs

Train a model

OR

Find a trained model

450+

Supported Models

Workflow of OpenVINO™ integration with TensorFlow

PICK a TensorFlow Model → **ENABLE** OpenVINO™ as backend → **RUN** Application using TensorFlow APIs

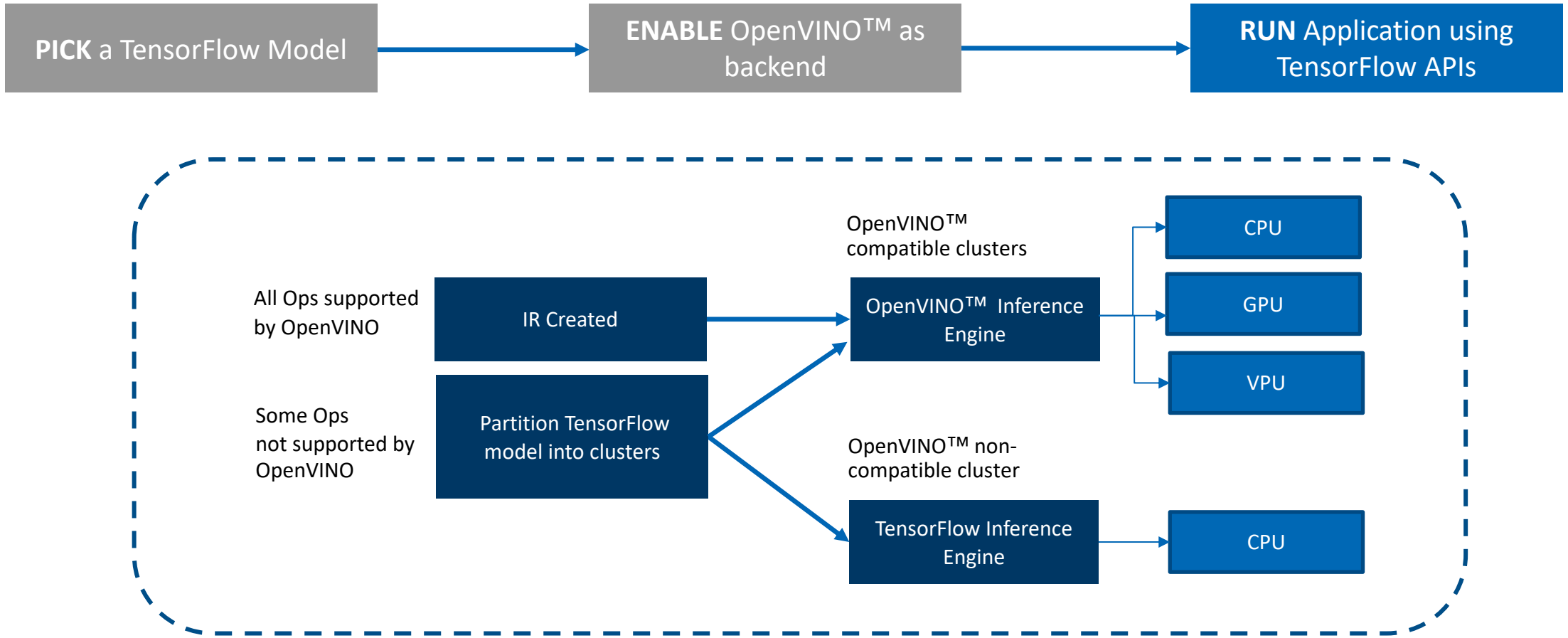
pip3 install openvino-tensorflow

```
# Line 1: import package
import openvino_tensorflow

# Line 2: choose inference hardware:
 device = CPU, GPU, VPU
openvino_tensorflow.set_backend(device)
```

2 lines of code in existing TensorFlow applications to access CPU, iGPU, VPU

Workflow of OpenVINO™ integration with TensorFlow



Platform & Docker & OS support in OpenVINO™ integration with TensorFlow

Platforms Support

- **Intel® CPUs**
- **Intel® integrated GPUs**
- **Intel® Movidius™ Vision Processing Units -** referred to as VPU
- **Intel® Vision Accelerator Design with 8 Intel Movidius™ MyriadX VPUs -** referred to as VAD-M or HDDL

Docker Support

- **Dockerfiles** for Ubuntu 18.04 and Ubuntu 20.04 OS are available.
- **Dockerfiles with TensorFlow Serving** support for the corresponding OS's are available.
- Prebuilt images can be found on **Docker Hub**
- **TensorFlow Serving** docker images released for OpenVINO™ integration with TensorFlow can be used to run all the workflows supported by standalone TensorFlow Serving Docker images

Reference Link

OS Support

- **Ubuntu 18.04, 20.04**
- **MacOS 11.2.3**
- **Windows 10 - 64 bit**

Other Prerequisites

Python: 3.7, 3.8, 3.9

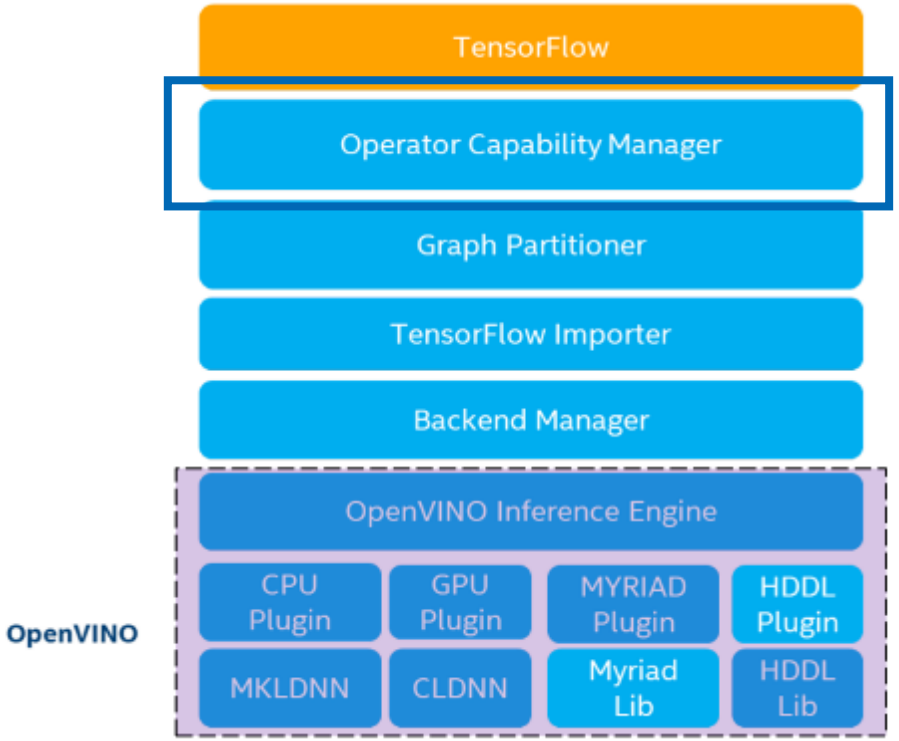
Interactive Installation Table

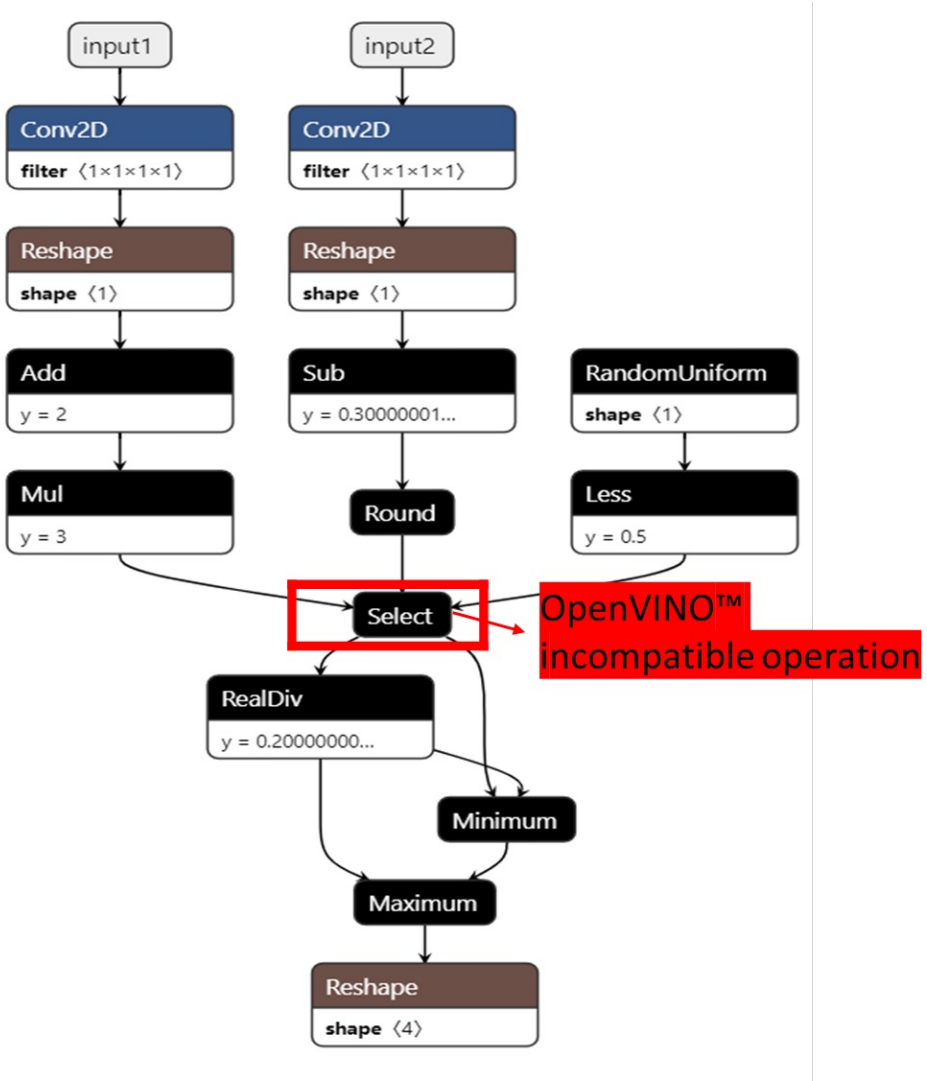
Check out the table for a menu of installation options.

Supported Models

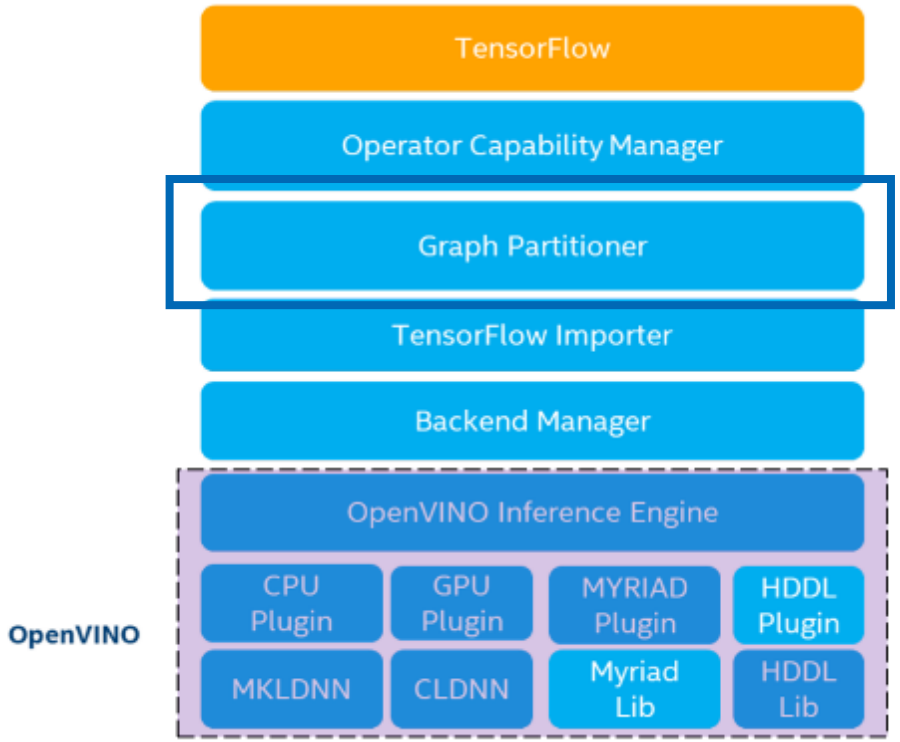
TensorFlow Serving Support- GitHub Walkthrough

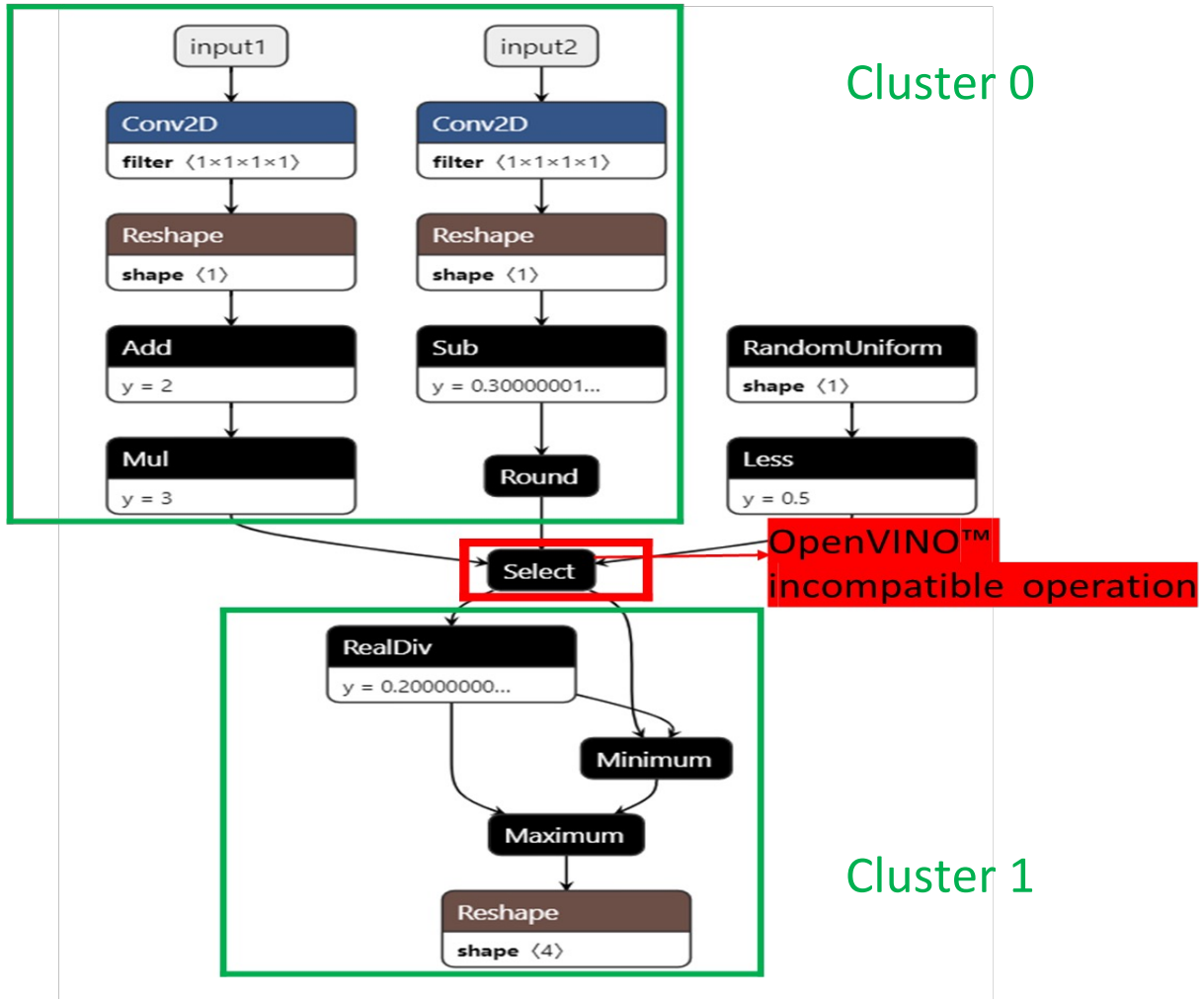
Architecture of OpenVINO™ integration with TensorFlow



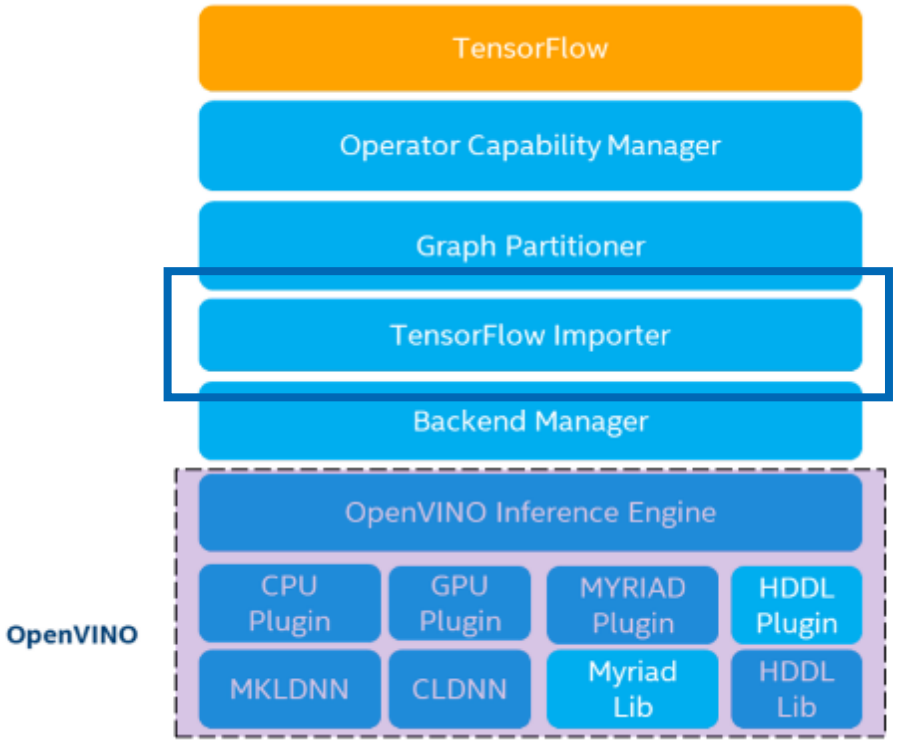


Architecture of OpenVINO™ integration with TensorFlow



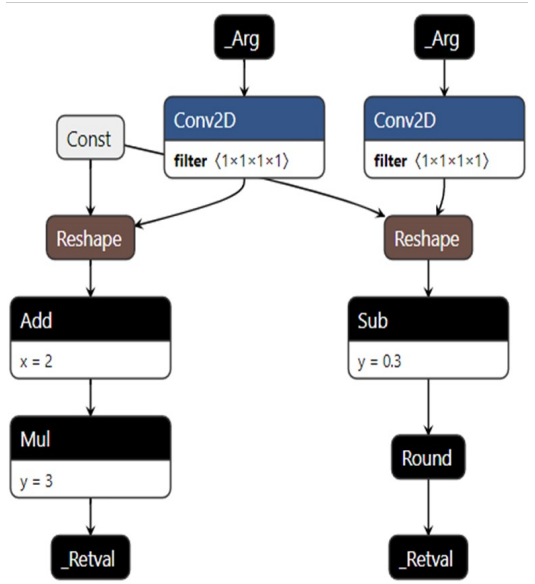


Architecture of OpenVINO™ integration with TensorFlow

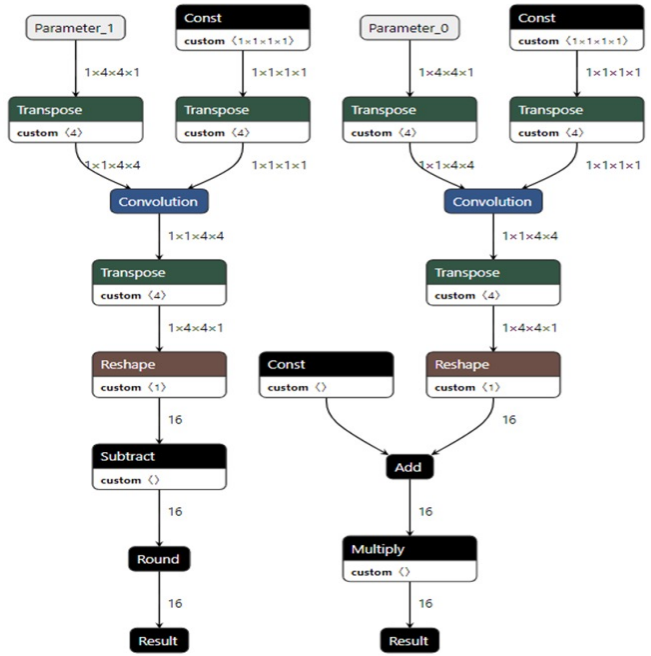


TensorFlow subgraph

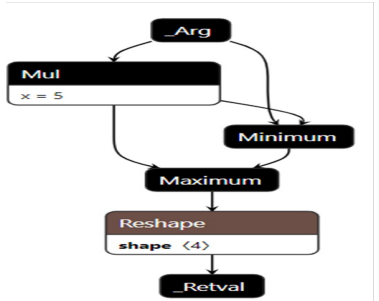
Cluster0

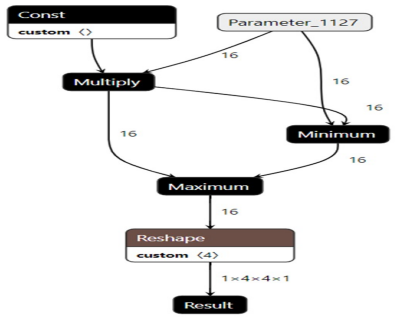


OpenVINO™ IR graph

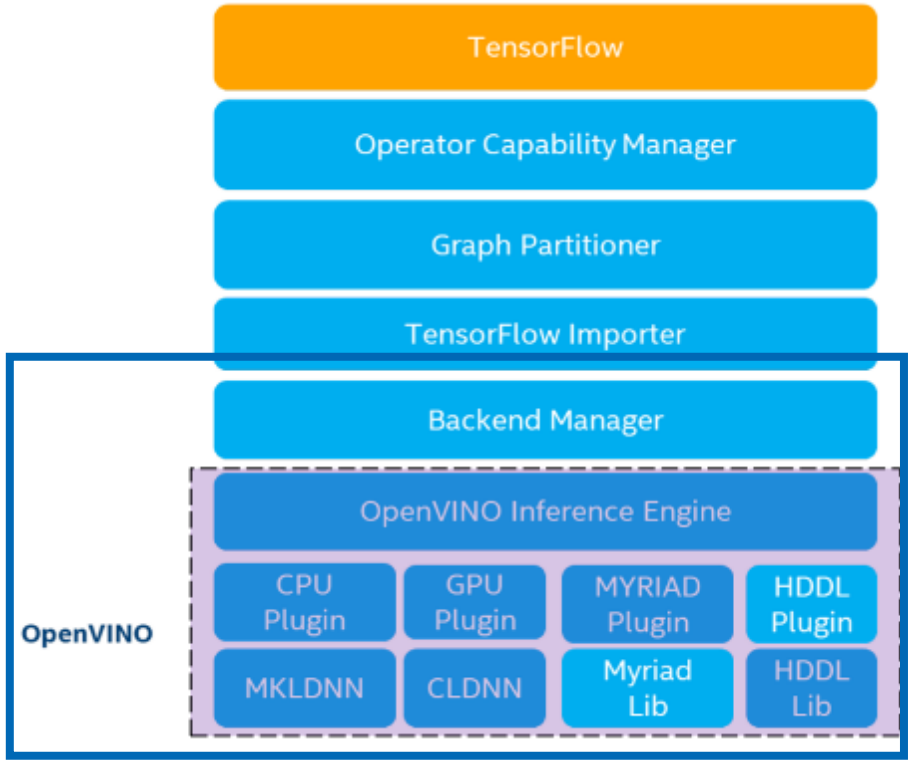


Cluster1





Architecture of OpenVINO™ integration with TensorFlow



Basic back end supports Intel® CPU, iGPU, MYRIAD.

VAD-M back end is used for Intel® Vision Accelerator Design with eight VPUs (known as VAD-M or HDDL).

DEMO- Google Colab

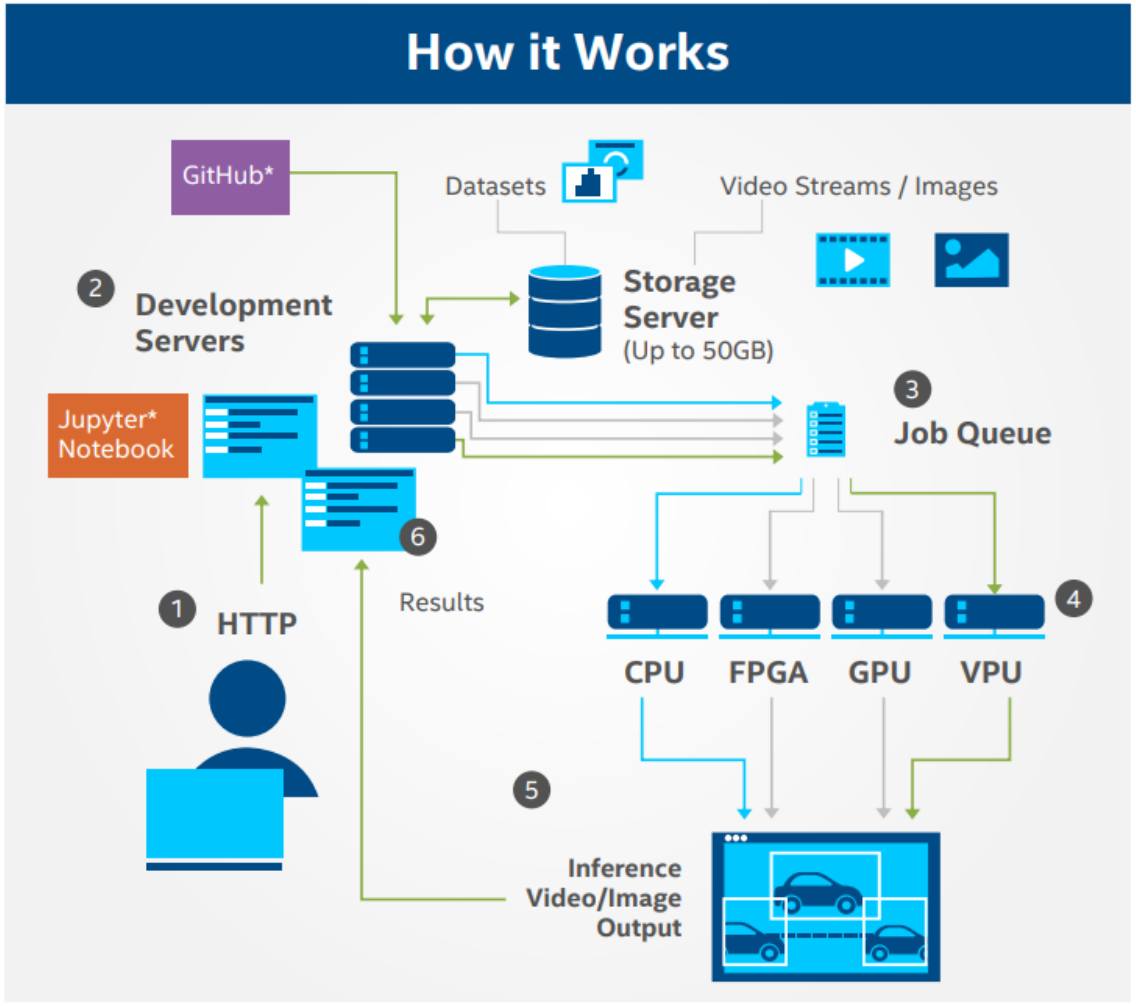
Intel® DevCloud: Edge Workloads

1 Access the Intel® DevCloud for the Edge through your web browser

2 Develop and test applications online using GitHub and datasets stored in the Intel® DevCloud's cloud storage

3 Test sample code to showcase benchmarking capabilities to customers. Customers can also test their own applications for benchmark performance results

How it Works



4 Runs tests to benchmark the application's performance on selected Intel processors and accelerators

5 Produces the inference video/image as output

6 Provides performance results to find the optimal hardware for the tested AI vision application

What's New with OpenVINO™ integration with TensorFlow v2.1.0

This release provides functional improvements and enhanced backend support from the previous preview release.

- Performance Optimizations of existing supported models
- TensorFlow version upgraded to v2.9.1.
- Prebuilt images are updated and can be found on Docker Hub and Azure Marketplace
- OpenVINO™ integration with TensorFlow source code is backward compatible. This means you will be able to build its source code with the past MINOR versions of TensorFlow 2.x.
- Enhanced GitHub documentation

OpenVINO™ integration with TensorFlow

GitHub | PIP | DockerHub | Azure Marketplace | Introductory Blog | FAQ | Solution Brief | Comic Strip | Developer Guide

Interactive web page to download and install the packages for use on your local edge devices: OS/Python/Platform installation matrix

Quickly get started with example demo applications and reference implementations

GitHub **Examples**: classification, object detection

Illustration to run samples on **Intel DevCloud**: classification, object detection

Illustration to run samples on **Google Colab** : classification, object detection

Instructions on **Docker** support & **TensorFlow Serving** support

Instructions to accelerate TensorFlow models on **cloud platforms**: AWS, Azure

Intel course on OpenVINO™ integration with TensorFlow

Email **openvino-tensorflow@intel.com** for any other questions

Additional Resources

To learn more:

- OpenVINO™ integration with TensorFlow
- OpenVINO™ Execution Provider for ONNX Runtime
- OpenVINO™ integration with TorchORT

Empowering Product Creators to Harness Edge AI and Vision

The Edge AI and Vision Alliance (www.edge-ai-vision.com) is a partnership of 100+ leading edge AI and vision technology and services suppliers, and solutions providers

Mission: To inspire and empower engineers to design products that perceive and understand.

The Alliance provides low-cost, high-quality technical educational resources for product developers

Register for updates at www.edge-ai-vision.com

The Alliance enables edge AI and vision technology providers to grow their businesses through leads, partnerships, and insights

For membership, email us: membership@edge-ai-vision.com

Join us at the Embedded Vision Summit
May 22-25, 2023—Santa Clara, California

The only industry event focused on practical techniques and technologies for system and application creators

- *"Awesome! I was very inspired!"*
- *"Fantastic. Learned a lot and met great people."*
- *"Wonderful speakers and informative exhibits!"*

Embedded Vision Summit 2023 highlights:

- **Inspiring keynotes** by leading innovators
- High-quality, practical **technical, business and product talks**
- Exciting **demos**, **tutorials** and **expert bars** of the latest applications and technologies

Visit www.EmbeddedVisionSummit.com to learn more