促進中小企業應用通用型 AI
宏碁自建雲智聯網事業部總經理
馬惠群
AI開發三部曲 - 1. 直接安裝AI所需各階層軟體

Seems to be straight forward in the beginning

Compatibility
Periodic update
How to share the GPU resources
How to control the access right

IDE/Visualization tools: Jupyter notebook, Tensorboard
Frameworks/Packages: TensorFlow, Python packages
Acceleration library: CUDA/cuDNN
OS: Linux
<table>
<thead>
<tr>
<th>Framework</th>
<th>CUDA 7.0</th>
<th>CUDA 8.0</th>
<th>CUDA 9+</th>
<th>cuDNN 5.0</th>
<th>cuDNN 6.0</th>
<th>cuDNN 6.1</th>
<th>cuDNN 7.0</th>
<th>cuDNN 7.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Python 3.5 (Numba / CUDA Python)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Python 3.6 (Numba / CUDA Python)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Python 3.7 (Numba / CUDA Python)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tensorflow 1.3</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tensorflow 1.4</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tensorflow 1.5</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Tensorflow 1.6</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Tensorflow 1.7</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Tensorflow 1.8</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Tensorflow 1.9</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Tensorflow 1.10</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Caffe2 v0.8.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caffe2 v0.8.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>
AI開發三部曲 – 2. Container(Docker)

Bare Metal

IDE / Visualization Tools
Dataset
DNN
Frameworks/Languages
Acceleration Lib
OS
CPU/GPU/NPU

Container based

Host OS
CPU/GPU/NPU

CONTAINER
DNN
MXNET
cuDNN CUDA
Tensor Flow
Cntk
cuDNN CUDA

Acceleration Lib
OS

Dataset
Frameworks/Languages
IDE / Visualization Tools

Bare Metal

Container Interface
Container (Docker) 的好處

**Packing related components together**
No compatibility issue

**Slim size**
Without OS (vs VM)

**An industry standard**
Google, nVidia, Microsoft, AWS all endorse the format

**Portable**
Easier to move to different locations

**Repository**
Existing container collections shared to the public such as Docker Hub or on GitHub
AI開發三部曲 – 3. 如何資源共享 Container Orchestrator

Job queue
Workload manager
Schedule control
GPU resource assignment
Monitoring
Access control

Master Node

Workload Manager

Worker Node 1
GPU x 8

Worker Node 2
GPU x 6

Worker Node 3
GPU x 6

Worker Node 4
GPU x 4
Kubernetes (K8s) – THE Container Orchestrator

Donated by Google to CNCF open source community, 1.0 is released in July 2015

Mainly a scheduling/orchestration tool

Tremendous momentum within Open Source Community

Average a new release every quarter, now 1.12, as Sep 2018

An open source ecosystem around Kubernetes now
AI開發的5個程序

Setup Environment  Prepare Dataset  Select Models  Train Models  Deployment
AI開發的5個程序 – Setup Environment

**Setup Environment**
**Prepare Dataset**
**Select Models**
**Train Models**
**Deployment**

**Tools**
- Tensorboard, MXBoard
- Azure ML Bench
- Keras, Gluon, ONNX

**IDE**
- Jupyter Notebook
- PyCharm, Visual Studio

**Framework**
- TensorFlow, Caffe,
- MXNET, CNTK, PyTorch
- Python libraries

**Drivers**
- CUDA, cuDNN, OpenCL

**Hardware (GPU)**
- PC, server, cloud
AI开发的5个程序 – Prepare Dataset

Setup Environment  Prepare Dataset  Select Models  Train Models  Deployment

Tools
Tensorboard, MXBoard
Azure ML Bench
Keras, Gluon, ONNX
IDE
Jupyter Notebook
PyCharm, Visual Studio
Framework
TensorFlow, Caffe,
MXNET, CNTK, PyTorch
Python libraries
Drivers
CUDA, cuDNN, OpenCL
Hardware (GPU)
PC, server, cloud

Existing Datasets
ImageNet, COCO
WordNet, LibriSpeech
Commercial solution
Labeling tools
In-house developed
3rd party/open source
Labeling resource
In-house or outsourcing
Big Data Integration
Spark, Hadoop

55% of Data Scientists consider training data quality and quantity as being their biggest challenge.
AI開発の5個程序 – Select Models

### Setup Environment
- **Tools**
  - Tensorboard, MXBoard
  - Azure ML Bench
  - Keras, Gluon, ONNX
- **IDE**
  - Jupyter Notebook
  - PyCharm, Visual Studio
- **Framework**
  - TensorFlow, Caffe, MXNET, CNTK, PyTorch
- **Python libraries**
  - Keras, Gluon, ONNX
- **Drivers**
  - CUDA, cuDNN, OpenCL
- **Hardware (GPU)**
  - PC, server, cloud

### Prepare Dataset
- **Existing Datasets**
  - ImageNet, COCO
  - WordNet, LibriSpeech
  - Commercial solution
- **Labeling tools**
  - In-house developed
  - 3rd party/open source
- **Labeling resource**
  - In-house or outsourcing
- **Big Data Integration**
  - Spark, Hadoop

### Select Models
- **Existing Models**
  - Pre-Trained/Un-Trained
  - Model zoo
  - Tensor2Tensor
  - SageMaker models
  - Azure AI Gallery

### Train Models

### Deployment
AI開發的5個程序 – Train Models

Setup Environment
- Training Resource:
  - On Prem, desktop, GPU cluster, cloud
- Workload Manager:
  - Kubernetes, Docker
- Account Manager:
  - User, group
- Hybrid:
  - On Prem to cloud

Prepare Dataset
- Hyperparameter Tuning:
  - In House, 3rd party

Select Models

Train Models
- Optimized AI Rack

Deployment
AI开发的5个程序 - Deployment

**Setup Environment**

**Prepare Dataset**

**Select Models**

**Train Models**

**Deployment**

Cloud
Computing, storage

Edge only
IPC, embedded
device, smartphone

**Accelerator**
GPU, DSP, ASIC
NN driver

**Optimization**
TensorFlow Lite

---

**AI Frameworks**

- TensorFlow
- TensorFlow Lite
- Caffe
- MXNET
- TVM
- CNTK

**AI Lib/Device SDK**

- Android NN
- Windows ML
- iOS Core ML
- nVidia TensorRT
- Intel OpenVINO
- ARM NN/CL
- Qualcomm SNPE

**OS / Driver**

- Android
- Linux
- Windows
- cuDNN CUDA
- Open CL
- Open CV

**HW Accelerator**

- CPU
- GPU
- DSP
- TPU
- NPU
- VPU
- FPGA
# AI Development Process & Consideration

## Setup Environment
- **Tools**
  - Tensorboard, MXBoard
  - Azure ML Bench
  - Keras, Gluon, ONNX
- **IDE**
  - Jupyter Notebook
  - PyCharm, Visual Studio
- **Framework**
  - TensorFlow, Caffe, MXNET, CNTK, PyTorch
  - Python libraries
- **Drivers**
  - CUDA, cuDNN, OpenCL
- **Hardware (GPU)**
  - PC, server, cloud

## Prepare Dataset
- **Existing Datasets**
  - ImageNet, COCO
  - WordNet, LibriSpeech
- **Commercial solution**
- **Labeling tools**
  - In-house developed
  - 3rd party/open source
- **Labeling resource**
  - In-house or outsourcing
- **Big Data Integration**
  - Spark, Hadoop

## Select Models
- **Existing Models**
  - Pre-Trained/Un-trained
  - Model zoo
  - Tensor2Tensor
  - SageMaker models
  - Azure AI Gallery

## Train Models
- **Training Resource**
  - On Prem, desktop, GPU cluster, cloud
  - Workload Manager
  - Kubernetes, Docker
- **Account Manager**
  - User, group
- **Hybrid**
  - On Prem to cloud
- **Hyperparameter Tuning**
  - In House, 3rd party

## Deployment
- **Cloud**
  - Computing, storage
- **Edge only**
  - IPC, embedded device, smartphone
- **Accelerator**
  - GPU, DSP, ASIC
- **Optimization**
  - TensorFlow Lite
THE BEST IS YET TO COME