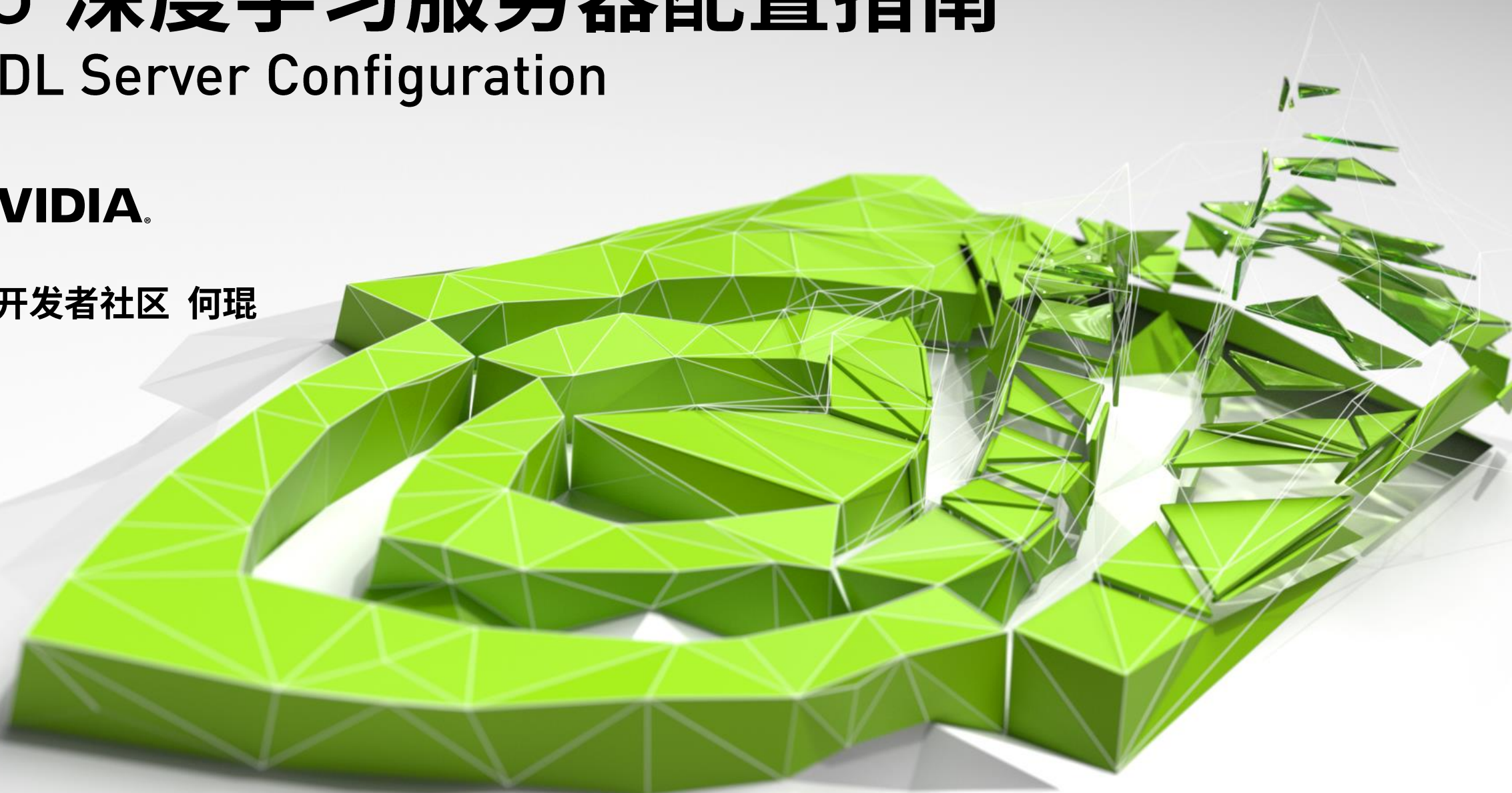


GPU 深度学习服务器配置指南

GPU DL Server Configuration












NVIDIA开发者社区 何琨








深度学习服务器:




- 深度学习模型的训练
- 训练样本的管理
- 训练模型的部署

 Accelerated Solutions GPUs are accelerating many applications across numerous industries. Learn more >	 Numerical Analysis Tools GPU acceleration for applications with high arithmetic density. Learn more >	 GPU-Accelerated Libraries Application accelerating can be as easy as calling a library function. Learn more >
 Language and APIs GPU acceleration can be accessed from most popular programming languages. Learn more >	 Performance Analysis Tools Find the best solutions for analyzing your application's performance profile. Learn more >	 Debugging Solutions Powerful tools can help debug complex parallel applications in intuitive ways. Learn more >
 Key Technologies Learn more about parallel computing technologies and architectures. Learn more >	 Accelerated Web Services Micro services with visual and intelligent capabilities using deep learning. Learn more >	 Cluster Management Managing your cluster and job scheduling can be simple and intuitive. Learn more >

COMPUTEWORKS

LIBRARIES cuBLAS cuSPARSE cuRAND NPP cuSOLVER NCCL cuFFT nvGRAPH	DIRECTIVES 	DEEP LEARNING cuDNN TensorRT NVIDIA Digits DeepStream SDK	LANGUAGE INTEGRATIONS    
---	--	--	--

NVIDIA GPU FAMILIES

QUADRO 	TESLA 	GEFORCE 
---	---	---

硬件配置

CPU	Xeon
Memory	32~64GB
Disk	>1TB 支持SATA 3 (RAID 0/1/5/10)
GPU	Compute Capability > 3.5
Tesla K40/K60/K80 Tesla M40/M60 Tesla P4/P40 Tesla V100 Geforce 1060/1070/1080/1080TI Geforce TITAN/TITAN V	
电源	2000w 冗余电源
网络接口	>2个接口 >100Mb

Tesla Data Center Products

GPU	Compute Capability
Tesla V100	7.0
Tesla P100	6.0
Tesla P40	6.1
Tesla P4	6.1
Tesla M60	5.2
Tesla M40	5.2
Tesla K80	3.7
Tesla K40	3.5
Tesla K20	3.5
Tesla K10	3.0

GeForce Desktop Products

GPU	Compute Capability
NVIDIA TITAN V	7.0
NVIDIA TITAN Xp	6.1
NVIDIA TITAN X	6.1
GeForce GTX 1080 Ti	6.1
GeForce GTX 1080	6.1
GeForce GTX 1070	6.1
GeForce GTX 1060	6.1
GeForce GTX 1050	6.1
GeForce GTX TITAN X	5.2
GeForce GTX TITAN Z	3.5
GeForce GTX TITAN Black	3.5
GeForce GTX TITAN	3.5

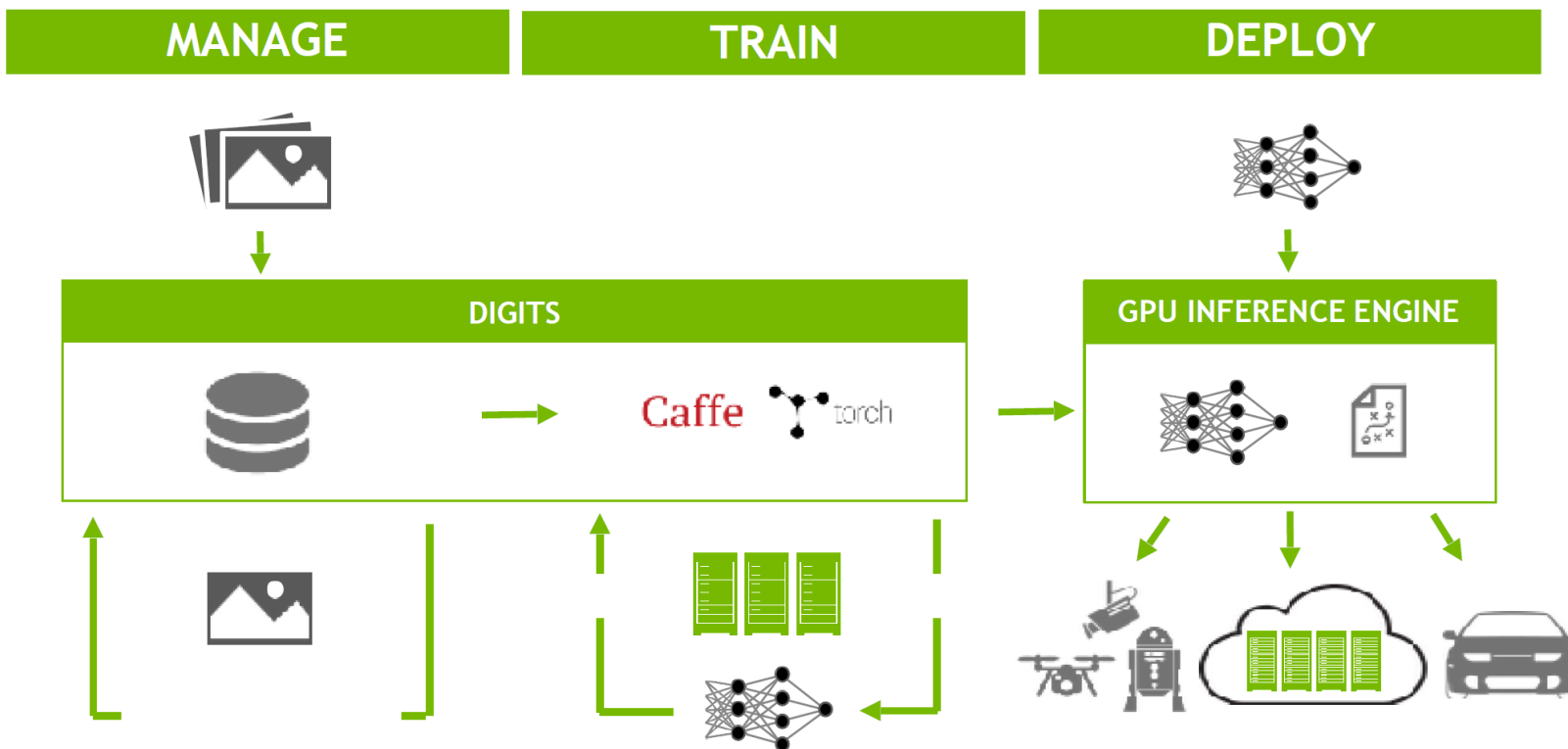
软件系统配置

System	Ubuntu 16.04 or 18.04 Centos Redhat
Driver	>= 384.81
CUDA	>9.0
NVIDIA深度学习开发工具	TensorRT 4.0 cuDNN 7.04 7.05 7.12 cuBLAS DeepStream 2.0
深度学习框架	CAFFE TensorFlow 1.7 1.8 1.9 PyTorch MxNet
深度学习训练系统	DIGITS

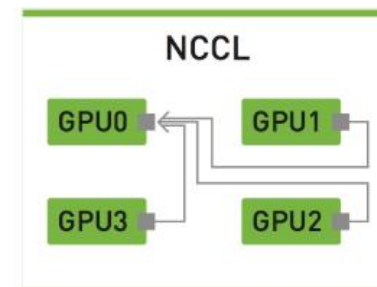
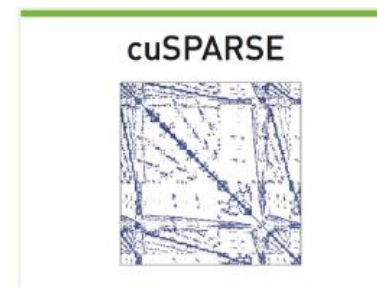
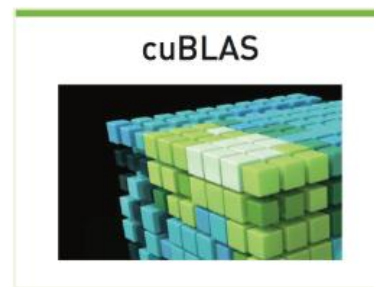
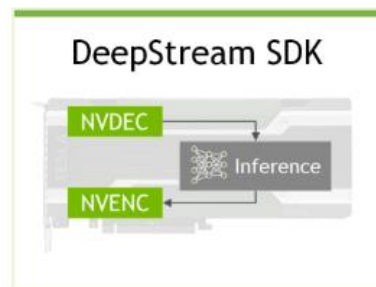
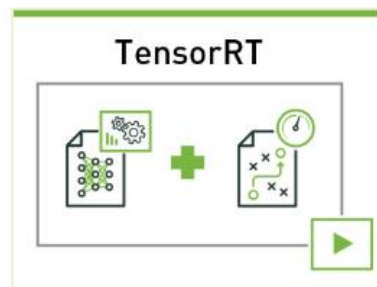
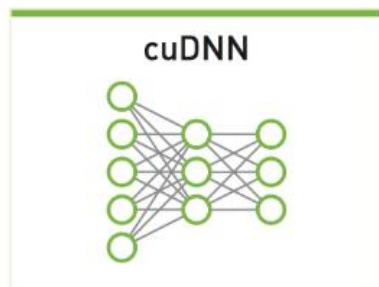
完整的深度学习解决方案

4

A COMPLETE DL PLATFORM



NVIDIA深度学习工具 & 支持NVIDIA GPU的深度学习框架



Caffe Caffe2 Chainer Cognitive Toolkit MATLAB
mxnet PaddlePaddle PYTORCH TensorFlow ...

NVIDIA DEEP LEARNING SDK and CUDA

深度学习服务器实例



CUDA Toolkit

cuDNN



Caffe

NVIDIA TensorRT

NVIDIA DIGITS

SYS/Tools	Version
System	Ubuntu 16.04 Lts
CUDA	9.0
CuDNN	7.1.2
TensorFlow	1.8
Caffe	0.15.14
TensorRT	4.0 GA
DIGITS	6.1.1

CUDA的下载和安装

NVIDIA ACCELERATED COMPUTING Downloads Training Ecosystem Forums Join Login

CUDA Toolkit 9.2 Download

Home > ComputeWorks > CUDA Toolkit > CUDA Toolkit 9.2 Download

Select Target Platform ⓘ

Click on the green buttons that describe your target platform. Only supported platforms will be shown.

Operating System

Get Started

[Installing the CUDA Toolkit](#) [Introduction to CUDA](#) [Getting Started with CUDA](#) [Discover CUDA 9 Capabilities](#)

[Installing the CUDA Toolkit](#) [Introduction to CUDA](#) [Getting Started with CUDA](#) [Discover CUDA 9 Capabilities](#)

[Installing the CUDA Toolkit](#) [Introduction to CUDA](#) [Getting Started with CUDA](#) [Discover CUDA 9 Capabilities](#)

[Installing the CUDA Toolkit](#) [Introduction to CUDA](#) [Getting Started with CUDA](#) [Discover CUDA 9 Capabilities](#)

Latest Release

CUDA Toolkit 9.2 (March 2018)

Archived Releases

CUDA Toolkit 9.1 (Dec 2017)

CUDA Toolkit 9.0 (Sept 2017)

CUDA Toolkit 8.0 GA2 (Feb 2017)

CUDA Toolkit 8.0 GA1 (Sept 2016)

CUDA Toolkit 7.5 (Sept 2015)

CUDA Toolkit 7.0 (March 2015)

CUDA Toolkit 6.5 (August 2014)

CUDA Toolkit 6.0 (April 2014)

CUDA Toolkit 5.5 (July 2013)

CUDA Toolkit 5.0 (Oct 2012)

CUDA Toolkit 4.2 (April 2012)

CUDA Toolkit 4.1 (Jan 2012)

CUDA Toolkit 4.0 (May 2011)

CUDA Toolkit 3.2 (Nov 2010)

CUDA Toolkit 3.1 (June 2010)

CUDA Toolkit 3.0 (March 2010)

OpenCL 1.0 Release (Sept 2009)

CUDA Toolkit 2.3 (June 2009)

CUDA Toolkit 2.2 (May 2009)

CUDA Toolkit 2.1 (Jan 2009)

CUDA Toolkit 2.0 (Aug 2008)

CUDA Toolkit 1.1 (Dec 2007)

CUDA Toolkit 1.0 (June 2007)

CUDA的下载和安装

1.CUDA

去官网下载cuda, 此处使用的是cuda9.0版本的deb安装方式:

```
~$ mkdir Downloads
```

```
~$ cd Downloads
```

```
~$ wget
```

```
http://developer.download.nvidia.com/compute/cuda/repos/ubuntu1604/x86\_64/cuda-repo-ubuntu1604\_9.0.176-1\_amd64.deb
```

```
~$ sudo dpkg -i cuda-repo-ubuntu1604_9.0.176-1_amd64.deb
```

```
~$ sudo apt-key adv --fetch-keys
```

```
http://developer.download.nvidia.com/compute/cuda/repos/ubuntu1604/x86\_64/7fa2af80.pub
```

```
~$ sudo apt-get update
```

```
~$ sudo apt-get install cuda
```

接下来, 在环境变量中添加CUDA:

```
~$ sudo vim ~/.bashrc
```

在最后一行添加cuda的安装路径:

```
export PATH=/usr/local/cuda-9.0/bin${PATH:+:${PATH}}
```

```
export LD_LIBRARY_PATH=/usr/local/cuda-9.0/lib64\${LD_LIBRARY_PATH:+:${LD_LIBRARY_PATH}}
```

CuDNN的下载和安装

2. 安装cudnn :

下载cudnn (本系统是ubuntu 16.04+cuda 9.0) :

下载地址 : <https://developer.nvidia.com/rdp/cudnn-downl>

安装并测试 :

```
$ sudo dpkg -i libcudnn7_7.1.3.16-1%2Bcuda9.0_amd64.deb
```

```
$ sudo dpkg -i libcudnn7-dev_7.1.3.16-1%2Bcuda9.0_amd64.deb
```

```
$ sudo dpkg -i libcudnn7-doc_7.1.3.16-1+cuda9.0_amd64.deb
```

然后我们测试一下安装的结果 :

```
$ cp -r /usr/src/cudnn_samples_v7/ $HOME
```

```
$ cd ~/cudnn_samples_v7/mnistCUDNN
```

```
$ make
```

```
$ ./mnistCUDNN
```

注意这里的版本号

cuDNN Download

NVIDIA cuDNN is a GPU-accelerated library of primitives for deep neural networks.

I Agree To the Terms of the [cuDNN Software License Agreement](#)

Note: Please refer to the [Installation Guide](#) for release prerequisites, including supported GPU architectures and compute capabilities, before downloading.

For more information, refer to the cuDNN Developer Guide, Installation Guide and Release Notes on the [Deep Learning SDK Documentation](#) web page.

[Download cuDNN v7.1.3 \(April 17, 2018\), for CUDA 9.1](#)

[Download cuDNN v7.1.3 \(April 17, 2018\), for CUDA 9.0](#)

[cuDNN v7.1.3 Library for Linux](#)

[cuDNN v7.1.3 Library for Linux \(Power8\)](#)

[cuDNN v7.1.3 Library for Windows 7](#)

[cuDNN v7.1.3 Library for Windows 10](#)

[cuDNN v7.1.3 Runtime Library for Ubuntu16.04 \(Deb\)](#)

[cuDNN v7.1.3 Developer Library for Ubuntu16.04 \(Deb\)](#)

[cuDNN v7.1.3 Code Samples and User Guide for Ubuntu16.04 \(Deb\)](#)

[cuDNN v7.1.3 Runtime Library for Ubuntu16.04 & Power8 \(Deb\)](#)

[cuDNN v7.1.3 Developer Library for Ubuntu16.04 & Power8 \(Deb\)](#)

[cuDNN v7.1.3 Code Samples and User Guide for Ubuntu16.04 & Power8 \(Deb\)](#)

[cuDNN v7.1.3 Runtime Library for Ubuntu14.04 \(Deb\)](#)

[cuDNN v7.1.3 Developer Library for Ubuntu14.04 \(Deb\)](#)

[cuDNN v7.1.3 Code Samples and User Guide for Ubuntu14.04 \(Deb\)](#)

[Download cuDNN v7.1.3 \(April 17, 2018\), for CUDA 8.0](#)

[Archived cuDNN Releases](#)

下载这三个

TensorRT的下载和安装

3.TensorRT的下载与安装

首先，我们下载TensorRT：

<https://developer.nvidia.com/nvidia-tensorrt-download>

```
$ sudo dpkg -i nv-tensorrt-repo-ubuntu1604-cuda9.0-rc-  
trt4.0.0.3-20180329_1-1_amd64.deb
```

```
$ sudo apt-get update
```

```
$ sudo apt-get install tensorrt
```

如果你是python 2.7

```
$ sudo apt-get install python-libnvinfer-doc swig
```

如果是3.5

```
sudo apt-get install python3-libnvinfer-doc
```

这样就安装好了，我们测试一下：

```
$ cp -r /usr/src/tensorrt/ ~/
```

```
$ cd ~/tensorrt/samples
```

```
$ make
```

```
$ cd ../bin
```

```
$ ./sample_int8 mnist
```

TensorFlow的下载和安装

```
$ wget https://developer.download.nvidia.com/compute/machine-learning/repos/ubuntu1404/x86_64/nvinfer-runtime-trt-repo-ubuntu1604-4.0.1-ga-cuda9.0_1.0-1_amd64.deb
$ sudo dpkg -i nvinfer-runtime-trt-repo-ubuntu1604-4.0.1-ga-cuda9.0_1.0-1_amd64.deb
$ sudo apt-get update
$ sudo apt-get install -y --allow-downgrades libnvinfer

$ pip install tensorflow # Python 2.7; CPU support (no GPU support)
$ pip3 install tensorflow # Python 3.n; CPU support (no GPU support)
$ pip install tensorflow-gpu # Python 2.7; GPU support
$ pip3 install tensorflow-gpu # Python 3.n; GPU support
```

Caffe的下载和安装

安装需要的依赖库:

```
sudo apt-get install --no-install-recommends build-essential cmake git gfortran libatlas-  
base-dev libboost-filesystem-dev libboost-python-dev libboost-system-dev libboost-  
thread-dev libgflags-dev libgoogle-glog-dev libhdf5-serial-dev libleveldb-dev liblmdb-dev  
libopencv-dev libsnpappy-dev python-all-dev python-dev python-h5py python-matplotlib  
python-numpy python-opencv python-pil python-pip python-pydot python-scipy python-  
skimage python-sklearn
```

下载并安装CAFFE

```
export CAFFE_ROOT=~/.caffe
```

```
git clone https://github.com/NVIDIA/caffe.git $CAFFE_ROOT -b 'caffe-0.15'
```

```
sudo pip install -r $CAFFE_ROOT/python/requirements.txt
```

```
cd $CAFFE_ROOT
```

```
mkdir build
```

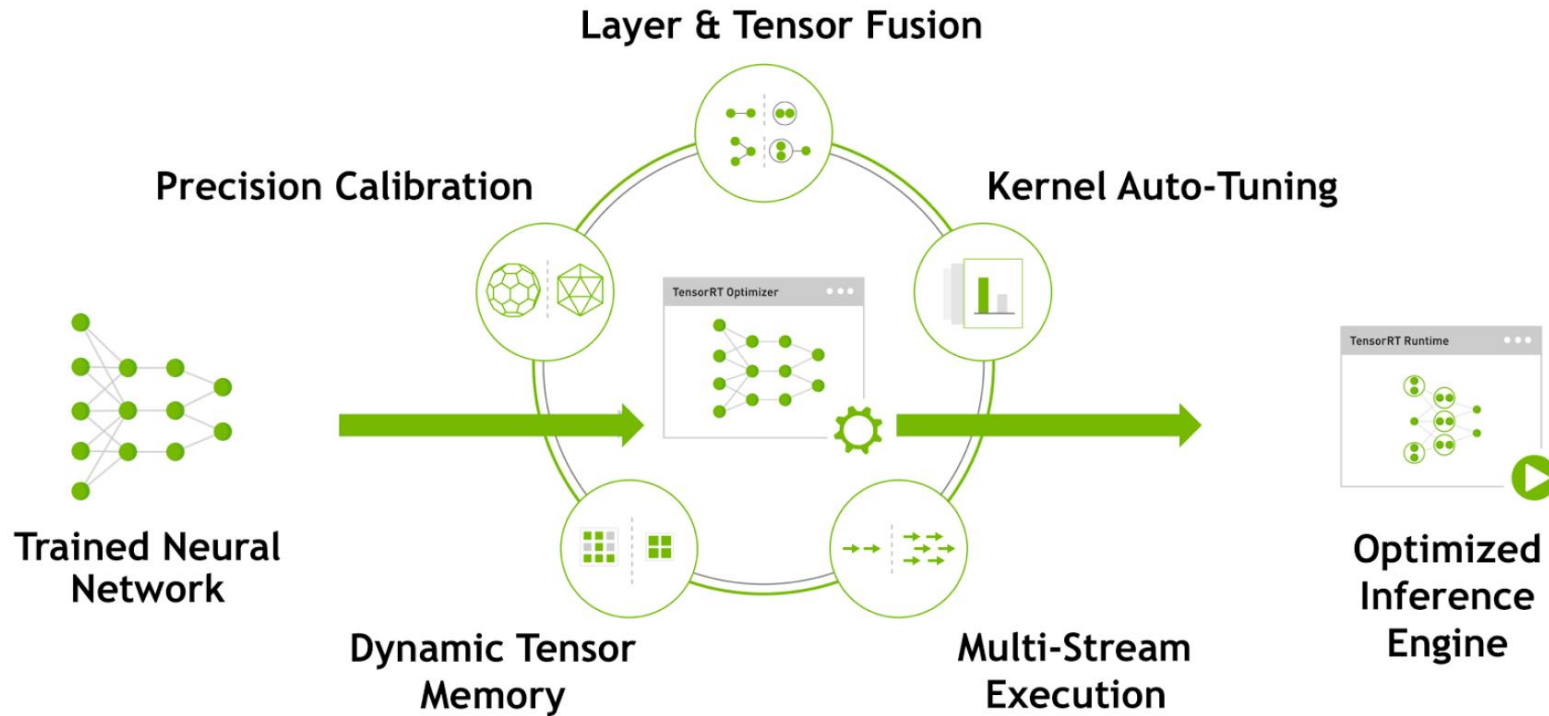
```
cd build
```

```
cmake ..
```

```
make -j"${nproc}"
```

```
make install
```


TENSORRT: WORK FLOW



<https://developer.nvidia-china.com/forum.php?mod=viewthread&tid=8767&extra=page%3D1>

总结

- 选择硬件，搭建一台属于自己的服务器
- 配置软件系统环境
- 选择深度学习加速工具
- 选择深度学习训练框架
- 选择深度学习训练管理系统
- 选择深度学习推理加速工具