Deep Learning for Business

Deep Learning Computing Systems & Software

Microsoft CNTK (Cognitive Toolkit)

Microsoft CNTK

Microsoft CNTK (Cognitive Toolkit)

- Originally introduced as CNTK (Computational Network Toolkit) by Microsoft Research in October 1, 2014
- Open-source license became available in April 2015
- Rebranded to CNTK (Cognitive Toolkit) in January 25, 2016

Microsoft CNTK

Microsoft CNTK (Cognitive Toolkit)

- DL (Deep Learning) neural network open-source license toolkit included with easy-to-learn computational steps based on a directed graph
- Python or C++ can be used for program development in CNTK
- CNTK's original programming model used configuration scripts

Microsoft CNTK

CNTK Network Graph Nodes

(simplified definition)

- CNTK Input: FeatureNode
- Training Labels: LabelNode
- Training Evaluation: CriterionNode
- Result Evaluation: EvalNode
- CNTK Output: OutputNode

Microsoft CNTK

CNTK Characteristics

- Optimized runtime system for DL neural network training and testing included
- Abstract computational graphs are used in DL neural network designing
- BrainScript are used in defining custom networks (improved configuration scripts)
- CNTK and TensorFlow have many similarities (e.g., both are script driven)

Microsoft CNTK References

- CNTK – GitHub [Online]. Available: https://github.com/Microsoft/CNTK
- Reasons to Switch from TensorFlow to CNTK [Online]. Available: https://docs.microsoft.com/en-us/cognitive-toolkit/reasons-to-switch-from-tensorflow-to-cntk
- TensorFlow Meets Microsoft's CNTK [Online]. Available: https://esciencegroup.com/2016/02/08/tensorflow-meets-microsofts-cntk/
- Wikipedia, www.wikipedia.org

Deep Learning for Business

Deep Learning Computing Systems & Software

NVIDIA DGX-1

NVIDIA DGX-1

Deep learning AI supercomputer with fully integrated hardware and software

- Made by NVIDIA
- Price $129,000

NVIDIA DGX-1

System Specifications

GPUs	8x Tesla GP100
TFLOPS (FPU FP16 / CPU FP 32)	170/3
GPU Memory	16 GB per GPU
CPU	Dual 20-core Intel® Xeon® E5-2698 v4 2.2 GHz
NVIDIA CUDA® Cores	28672
System Memory	512 GB 2133 MHz DDR4 LRDIMM
Storage	4x 1.92 TB SSD RAID 0
Network	Dual 10 GbE, 4 IB EDR
Software	Ubuntu Server Linux OS DGS-1 Recommended GPU Driver
System Weight	134 lbs
System Dimension	866 D x 444 W x 131 H (mm)
Packing Dimension	1180 D x 730 W x 284 H (mm)
Maximum Power Requirements	3200 W
Operating Temperature Range	10 – 35 °C

NVIDIA DGX-1

System Specifications

- 170 TFLOPS = 170 x 10^{12} FLOPS
- DL (Deep Learning) software includes accurate DNN (Deep Neural Networks) stack
- NVIDIA NVLink interconnected with Pascal-powered NVIDIA Tesla P100 GPU accelerators
- NVIDIA Pascal engine was designed for computer learning, seeing, and simulation

NVIDIA DGX-1

DL (Deep Learning) SDK

- NVIDIA Deep Learning SDK is included
  - SDK: Software Development Kit
  - SDK helps development of advanced GPU accelerated applications
    - ✓ Clouds, data centers, workstations, and embedded platforms
  - DIGITS GPU training system
  - Drivers & CUDA

NVIDIA DGX-1

DL (Deep Learning) SDK

- NVIDIA Deep Learning SDK also includes
  - GPU Accelerated Libraries
  - Debugging & Optimization Tools
  - C/C++ Compiler
  - Runtime Library

Deep Learning for Business

NVIDIA DGX-1

References

References

- NVIDIA DGX-1: Essential Instrument of AI Research [Online]. Available: https://www.nvidia.com/en-us/data-center/dgx-1/
- NVIDIA DGX-1 DEEP LEARNING SYSTEM [Online]. Available: https://images.nvidia.com/content/technologies/deep-learning/pdf/61681-DB2-Launch-Datasheet-Deep-Learning-Letter-WEB.pdf
- NVIDIA DGX-1 ARTIFICIAL INTELLIGENCE SYSTEM [Online]. Available: http://images.nvidia.com/content/technologies/deep-learning/pdf/Datasheet-DGX1.pdf
- Wikipedia, www.wikipedia.org