IBM PowerAI deep learning frameworks

Release 4.0

Highlights

- **Enterprise Software Distribution:** Binary package of major open source deep learning frameworks for ease of deployment and update.
- **Tools for Ease of Development:** Graphical tools to enhance data scientist and developer experience and improve productivity.
- **Faster Training Times for Data Scientists:** Performance optimized for large model support and distributed deep learning.

Data, in all forms, is expanding as a resource to be utilized. Yet in many industries and professions, the data explosion is outstripping the human capacity to understand the meaning hidden within that data. Cognitive computing is able to unlock the potential in all data—internal, external, structured, unstructured, voice, and visual—and make it work together. Enterprises can make better operational decisions, understand customer wants and needs, communicate in real time, and optimize business processes—infused with the cognitive ability to understand, reason, and learn.

PowerAI helps accelerate this journey to cognitive computing by bringing together a collection of the most popular open source frameworks for deep learning, along with supporting software and libraries in a single installable package. Our design goal was to simplify the acquisition, installation and system optimization required to bring up a deep learning infrastructure, and to provide the most comprehensive set of integrated development tools to build out a complete end-to-end deep learning solution on top of the fastest platform for deep learning training and model deployment. The outcome: users to spend less time on data preparation, implementation, and integration, and more time training neural networks for results.

Release 4.0 of the PowerAI deep learning frameworks builds upon the strength if prior releases, adding in to the platform two capabilities which dramatically extend the capability of deep learning.

Distributed Deep Learning

To accelerate the time dedicated to training a model, the PowerAI stack uses new technologies to deliver exceptional training performance by distributing a single training job across a cluster of servers.

PowerAI's Distributed Deep Learning brings intelligence about the structure and layout of the underlying cluster (topology) which includes intelligence about the location of the cluster's different compute resources such as Graphical Processing Units (GPUs) and CPUs. PowerAI is unique in that this capability is incorporated in to the Deep Learning frameworks as an integrated binary, reducing complexity for clients as they bring in high performance cluster capability. As a result of this capability PowerAI with Distributed Deep Learning (DDL) can scale jobs across large numbers of cluster resources with very little loss to communication overhead. Distributed Deep Learning is available as a technology preview with PowerAI 4.0, and is compatible with bundled TensorFlow and IBMCaffe frameworks.

Large Model Support

One of the challenges customers face in this space is that they are limited by the size of memory available within GPUs. Today, when data scientists develop a deep learning workload the structure of matrices in the neural model and the data elements which train the model (in a batch) must sit within the memory on GPU. This is a problem because today's GPUs memory is mostly limited to 16 GB. As models grow in complexity (deeper neural networks contain more layers and larger matrices) and data sets increase in size (high definition video vs web scale images), data scientists are forced to make tradeoffs to stay within the 16 GB memory limits of each GPU.

With Large Model Support, enabled by PowerAI's unique NVLink connection between CPU (memory) and GPU, the entire model and dataset can be loaded in to system memory and cached down to the GPU for action. Users now have the ability to increase model sizes, data elements and batch or set sizes significantly, with the outcome of executing far larger models and expanding up to nearly one terabyte (TB) of system memory across 4 GPUs. This capability is unique to PowerAI and opens up the opportunity to address larger challenges and get much more work done within a PowerAI single server, increasing organizational efficiency. Large Model Support is available as a technology preview with PowerAI 4.0, and is compatible with bundled TensorFlow and IBMCaffe frameworks.

PowerAI includes the most popular deep learning frameworks in one installation:

- BVLC Caffe
- NVIDIA Caffe
- IBMCaffe
- TensorFlow
- Torch
- Theano
- Chainer

Built for High-performance Computing

Application developers can execute their deep learning algorithms either on a POWER® central processing unit (CPU) or using a general-purpose graphics processing unit (GPU) accelerator—a technology pioneered by high performance computing design and deployments. Accelerators are accessed using device drivers and libraries provided by the accelerator manufacturers. The fourth release of the PowerAI Deep Learning Frameworks is based on the use of Ubuntu 16.04 on IBM® POWER® with NVIDIA CUDA 8 and cuDNN v6 packages running on HPC hardware.

Hardware platform description and ordering information

The PowerAI Deep Learning Frameworks are tuned for use with the following configuration:

IBM Power Systems™ S822LC for High Performance Computing (Model 8335-GTB) with up to four NVLINK attached NVIDIA Tesla P100 GPUs. (**ibm.com**/ms-en/marketplace/high-performance-computing)

Please contact IBM for help configuring or placing an order.

Software Download

Direct Download
PowerAI is distributed as a binary for Ubuntu 16.04 LTS from the following source:
https://ibm.biz/powerai

Release Guide
A complete Release Guide with package list, prerequisites, deployment guide, and developer information is available at:
https://developer.ibm.com/linuxonpower/deep-learning-powerai/releases/

For more information

To learn more about the PowerAI, please contact your IBM representative or IBM Business Partner, or visit the following website: https://ibm.biz/powerai

Additionally, IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business. We provide full lifecycle management of IT products and services, from acquisition to disposition. For more information, visit: **ibm.com**/financing

IBM Systems
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
August 2017

POD03115-USEN-08