

# PowerAI FAQ

## Questions and Answers:

Product.....	2
What is being announced with PowerAI version 1.5.1? .....	2
Will PowerAI offerings be available on POWER9 at the time of GA? .....	2
Do we have plans to support additional frameworks beyond those currently announced? Are there any major frameworks we don't plan to support? .....	2
How do I order Enterprise Support for PowerAI products? .....	2
Will PowerAI run on x86? is it fee based? .....	3
What is PowerAI Vision? How do I access it? .....	3
What is PowerAI Large Model Support & Distributed Deep Learning?.....	3
Positioning.....	3
How do we position Power Systems 822LC for HPC/Power Systems AC922 vs DGX? .....	3
What are the competitive products for each of our PowerAI Enterprise related offerings? ....	4
How do we position PowerAI vs Watson or other public cloud AI offerings? .....	4
At GTC, Nvidia announced framework offerings support with a new update every 30 days? How do we compare? .....	4
What is DSX, does it support Power and how does it help me sell Enterprise AI? .....	5
Enablement and Support .....	5
Who are the IBM experts on IBM PowerAI and how do I engage them? .....	5
I've heard about the Deep Learning Institute, what is it? How do I access it? .....	5
What kind of education do we provide customers on our PowerAI Offerings? .....	6

## Product

### What is being announced with PowerAI version 1.5.3?

PowerAI version 1.5.3 (note: naming, version one, release 5, modification 3) builds on the existing PowerAI functionality and provides data scientists and developers the capabilities to scale to hundreds of GPUs across multiple nodes, and the ability to deploy the analysis against larger and more complex models. The capabilities are known as Distributed Deep Learning (DDL) and Large Model Support (LMS) and both will be available across multiple deep learning frameworks with 1.5.3:

- Caffe: supports both DDL and LMS
- TensorFlow: supports both DDL and LMS
- PyTorch: with this release, we have added support for DDL

PowerAI 1.5.3 supports the POWER9 based AC922 and POWER8 based S822LC for HPC servers.

PowerAI 1.5.3 continues support for Red Hat Enterprise Linux (RHEL, version 7.5), providing customers with additional options for Linux distribution. In addition to the support of RHEL 7.5, IBM is offering full software stack support with the new Support Line for PowerAI (Support Line offering began with 1.5.1).

PowerAI 1.5.3 also updates underlying libraries, refreshing TensorFlow to version 1.10, IBM Caffe and BVLC Caffe to version 1.0.0, and PyTorch to 0.41 (please see appendix for detail).

Additional details on these capabilities are addressed in subsequent questions.

### Will PowerAI offerings be available on POWER9 at the time of GA?

Yes.

### Do we have plans to support additional frameworks beyond those currently announced?

#### Are there any major frameworks we don't plan to support?

The PowerAI team is evaluating additional frameworks on a case by case basis as part of our participation in the rapidly evolving deep learning ecosystem. As part of this evaluation it is immensely helpful to understand specific client requirements and the relevant opportunity details. Please share details of these requirements directly with the offering team ([soutter@us.ibm.com](mailto:soutter@us.ibm.com))

### How do I order Enterprise Support for PowerAI products?

Beginning with the announcement of PowerAI 5, Support Line for PowerAI support is an orderable offering from Technology Support Services. Support will be entitled by server serial number, and will be limited in scope to specific frameworks. Details about Supportline for PowerAI are available in the announcement letter (<https://w3-01.ibm.com/sales/ssi/cgi-bin/ssialias?infotype=an&subtype=ia&appname=gpteam&supplier=897&letternum=ENUS217->

[576#abstr](#))

[Will PowerAI run on x86? is it fee based?](#)

IBM PowerAI will be exclusively available only on IBM Power Systems.

[What is PowerAI Vision? How do I access it?](#)

IBM PowerAI Vision provides an interface and tools for business analysts, subject matter experts and developers with limited skills in deep learning technologies that are looking to deploy computer vision capabilities. The tools assist users to pre-process and label datasets so they can be utilized to train and validate a model for customized solutions that demand image classification and object detection. IBM PowerAI Vision is supported on Power Systems and exploits the GPUs with our CPU NVLINK connectivity to optimize the performance required for training the models.

Please contact Srinivas V Chitivelu (svchitiv@us.ibm.com) for your PowerAI Vision opportunities.

[What is PowerAI Large Model Support & Distributed Deep Learning?](#)

Distributed Deep Learning (DDL) with PowerAI brings intelligence ABOUT the structure and layout of the cluster (topology) and intelligence about location of different compute resources like GPUs and CPUs within the cluster in to the deep learning frameworks themselves, with unique IBM intellectual property being incorporated in to our builds of TensorFlow and Caffe. The impact of this is significant: PowerAI with DDL can scale jobs across large numbers of cluster resources with very little loss to communication overhead - unlike other systems –when PowerAI scales up to 256 GPUs, it does so with 88% efficiency, which is unique in today's distributed deep learning space. Most distributions struggle to get beyond a server!

One of the challenges our customers are facing 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 –today that space is mostly limited to ~16GB. As models grow in complexity (deeper neural models contain more layers, and larger matrices), and data sets increase in size by both data element (think: high definition video vs web scale images with respect to the number of pixels that are inputted into the algorithm) and number of elements needed. Data Scientists are forced to make tradeoffs to stay within the constrained 16GB memory limits of today's machine. 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. In our testing, we have been able to increase model sizes by 10X, data element and set size by 10X, and batch size by 24X –using nearly an entire terabyte of system memory across 4 GPUs –contrast that to the competitions 64GB or 128GB of available space! Final impact? Bigger challenges and much more work done within a single server increasing organizational efficiency.

## Positioning

[How do we position Power Systems 822LC for HPC/Power Systems AC922 vs DGX?](#)

With confidence. DXG-1 is a one-size-fits-all approach to deep learning training as an appliance

and as such it fails to address significant parts of the overall end to end cognitive application development cycle. DGX-1 puts the responsibility of preparing and managing the transformation of data on the customer, whereas PowerAI incorporates parallel transformation and extract with machine intelligence using the Spectrum Conductor with Spark package. DGX-1 shares the limitation of all non-Power architectures by requiring that any transfer from system memory to GPU memory go over a PCIe bus, PowerAI supports multiple high speed and low latency NVLink connections with more than 5.6x the bandwidth of the fastest PCIe connection which significantly improves training speed in real-world application and opens up the possibility of using much larger data sets, neural models, or larger batches than any of the Intel OEMs. Finally, DGX-1 does not have IBM's global system support, with PowerAI and IBM's global Technical Support Services we support systems in the field, 24x7 for critical customer needs.

### What are the competitive products for each of our PowerAI Enterprise related offerings?

Software competition for PowerAI: customers building and compiling frameworks from open source foundations. Software provided as a service from cloud providers such as Microsoft Azure, Google Cloud, Amazon AWS, NVIDIA Cloud

Hardware competition to PowerAI running on Power Systems: Tier 1 Intel OEMs or Tier 2 / Tier 3 white box Intel OEMs with 4-8 PCIe or partially NVLink connected GPUs.

When competing with these platforms, highlight the performance benefit of IBM's exclusive NVLink connectivity to POWER8 and POWER9 CPUs and the significant performance difference in memory transfer speed; highlight IBM innovations in areas such as Large Model Support with Caffe and TensorFlow, and increased cluster scalability with Distributed Deep Learning.

### How do we position PowerAI vs Watson or other public cloud AI offerings?

PowerAI provides developers and data scientists with tools and frameworks to prepare & ingest data, train a deep learning model against that data and deploy the trained algorithm for inferencing (scoring). The tools and capabilities provided through PowerAI make that end-to-end process easier and quicker.

A public cloud API (Ex: Watson or Google APIs) is when the ingest and training has already been done and you are consuming the algorithm through an API. So while PowerAI provides the capability to create custom APIs, an existing cloud API is enabled through pre-trained models. The two are not mutually exclusive. Often, a workflow that leverages PowerAI will call a cloud based API to execute a function as part of that larger workflow. We have examples of this in our [PowerAI developer portal](#).

PowerAI is also available as a cloud service through the IBM Cloud.

### At GTC, Nvidia announced framework offerings support with a new update every 30 days? How do we compare?

PowerAI releases are now typically quarterly with significant incorporation of unique IBM intellectual property and differentiation. Over time, as IBM builds up the PowerAI development team, we may move back to continuous integration and monthly releases. There are trade-offs in either approach: customers want to stay on the bleeding edge while beginning to research and evaluate deep learning, but as they move in to production stabilization and long-term support of release levels become significant concerns. We are trying to balance both while maintaining stability within our development processes.

## What is DSX, does it support Power and how does it help me sell Enterprise AI?

DSX is a complete ecosystem with open source based frameworks/libraries/tools for scientists to develop algorithms, validate, deploy and collaborate with communities of scientists and developers. The offering decouples the scientists from setting up a development ecosystem and start focusing on developing algorithms in their most preferred language, IDE and libraries. DSX is offered on the cloud and on premise. PowerAI provides a deep learning ecosystem for data scientists and developers where frameworks like TensorFlow, Torch, Caffe, Theano and AI Vision are pre-installed. In October, 2017, IBM delivered an integrated offering of DSX Local on Power systems. With this offering, the IBM PowerAI team is providing the Deep Learning frameworks and GPU acceleration for the integrated offering.

## Enablement and Support

### Who are the IBM experts on IBM PowerAI and how do I engage them?

Your first point of contact for PowerAI is the Cognitive Systems technical sales teams in your markets. A list of the Cognitive Systems Technical Sales Leaders per the seven geographies is published for reference below.

Batuhan Tunca for MEA ([BATUHAN@tr.ibm.com](mailto:BATUHAN@tr.ibm.com))  
Ivan N Peacock for Europe: ([ivan\\_peacock@uk.ibm.com](mailto:ivan_peacock@uk.ibm.com))  
Livio S Sousa for LA: ([livios@br.ibm.com](mailto:livios@br.ibm.com))  
Mikko P Salminen for AP: ([msالmin1@au1.ibm.com](mailto:mسالmin1@au1.ibm.com))  
Ning TS Xu for GCG: ([ningxu@cn.ibm.com](mailto:ningxu@cn.ibm.com))  
Scott Bell for NA: ([gshell@us.ibm.com](mailto:gshell@us.ibm.com))  
Takashi Ohsawa for Japan: ([TOHSAWA@jp.ibm.com](mailto:TOHSAWA@jp.ibm.com))

The offering management team is the second line for client engagements and is comprised of:

Scott Soutter (PowerAI Enterprise) - [soutter@us.ibm.com](mailto:soutter@us.ibm.com)  
Srinivas V Chitveli (PowerAI Vision) - [svchitiv@us.ibm.com](mailto:svchitiv@us.ibm.com)  
[Luis Armenta Garza \(H2O Driverless AI\) – luis.armenta@us.ibm.com](mailto:luis.armenta@us.ibm.com)  
[Chuck Calio \(cross offering, developer outreach\) – ccalio@us.ibm.com](mailto:ccalio@us.ibm.com)  
Keshav Ranganathan (DSX Local) - [keshav@us.ibm.com](mailto:keshav@us.ibm.com)

### I've heard about the Deep Learning Institute, what is it? How do I access it?

The [Deep Learning Institute](#) is an NVIDIA training program that offers self-paced labs and instructor led workshops to data scientists and researchers. It provides learning on the latest techniques for designing, training, and deploying neural networks on their GPU platform. Today, this program is exclusively delivered on an x86 platform and contains exclusively x86 content.

IBM has its own cognitive centric training and badging program through the [cognitive classes](#) platform. Some of the existing deep learning modules include POWER8 and PowerAI related content with plans to integrate additional material over the coming months.

In addition, IBM is working on integrating our Power content into the DLI curriculum and delivering those modules through the cognitive classes.ai platform. A deep learning pathway is being developed to support a PowerAI specific curriculum.

What kind of education do we provide customers on our PowerAI Offerings?

The starting point for all enablement is the Intro to PowerAI enablement [found here](#). The charts are client facing and scripted with extensive speaking points, making it really easy to either present the charts to a client or leave it with a client for their own enablement. Subsequent modules with deeper dives will follow.

You should ensure developers and data scientists are made aware and leveraging the [PowerAI developer portal](#). The portal provides prospects with developer journeys, discussion forums, deployment material etc.

## Appendix: Release Details

	Rel 5 (RHEL only)	Version 1.5.1(RHEL)	1.5.2 (RHEL)	1.5.3 (RHEL)
CUDA	9.0	9.2	9.2.88	9.2.148
cuDNN	7.0	7.1.2	7.2.1	7.2.1
GPU Driver	384	396	396.26	396.44

Most of the packages in Rel 4 are rebuilt against cuDNN v6. Some of the packages were further refreshed to newer upstream versions:

	Rel 5 (RHEL Only)	Version 1.5.1 (RHEL)	1.5.2 (RHEL)	1.5.3 (RHEL)
Bazel	0.5.4	0.8.0	0.11	0.15
BVLC Caffe	1.0.0	1.0.0	1.0.0	1.0.0
IBM Caffe	1.0.0	1.0.0	1.0.0	1.0.0
TensorFlow	1.4.0	1.5.1	1.8	1.10
PyTorch			0.4	0.4.1
SnapML		1.0 (local)	1.0 (MPI)	1.0 (MPI+Spark)

### Notes:

Rel 3.4 included multiple versions of NVIDIA Caffe and TensorFlow. Rel 4 includes only a single version of each of those. Ubuntu only support for this release

Rel 4 also includes two new components:

OpenMPI - 2.0.1 (built with CUDA support)  
IBM PowerAI Distributed Deep Learning - Tech Preview  
Ubuntu only support

Rel 5 replaces the OpenMPI library with Spectrum MPI version 10.1. Red Hat 7.43 only support on POWER8. This is the first release with full stack support with new Support Line Offering.

Version 1.5.1 continues to support Red Hat Enterprise Linux with version 7.5 on both POWER9 and POWER8.

Version 1.5.2 and 1.5.3 support only RHEL on POWER8 and POWER9. These versions also support both bare metal (RPM install) and Docker container environments; the container environment for 1.5.2 on supports IBM Cloud Private.