insideHPC

insideHPC Special Report

Launch a Machine Learning Startup

by Beth Harlen

BROUGHT TO YOU BY   DELL EMC | NVIDIA

Introduction

Unexplored markets and an almost unlimited potential mean that today's machine and deep learning-based startups will be tomorrow's tech pioneers, marketing gurus, financial advisors, and more. But as anyone who has ever been involved in a startup will tell you, it's never that easy to take a promising concept and channel it into long-term success. The old adage of 'knowledge is power' still holds true today, and machine and deep learning provide a unique opportunity for organizations to gain insight, boost innovation, and unlock the potential of their data. In essence, startups can use machine and deep learning algorithms to significantly impact everything within an organization, from IT system stability to the bottom line. But where do we begin?

In this report, we'll address everything from how to choose a framework and pick the kit you need to get started, to the questions you'll be asking yourself, and the benefits of immersing yourself in the machine and deep learning communities. First though, we need to detangle the jargon and explore what these terms actually mean.

While artificial intelligence (AI), machine learning and deep learning are often thought of as being interchangeable, they do in fact relate to very different concepts. It all began in the 1950s with AI and the idea that a computer could be made to simulate human learning and intelligence.

A subclass of that is machine learning, whereby a computer can take large amounts of data and use it begin to recognize patterns, make predictions on new data, and essentially 'learn' for itself. The drawback is that machine learning requires that parameters be set for what the computer needs to recognize, and those inputs can be time-consuming. And so we go one step further, into deep learning.

For example, Ripjar offers a service under the heading of 'Analysis at the Speed of Thought' that utilizes deep learning combined with natural language processing to analyze an organization's internal data, in addition to information from sources like news feeds, web pages, and social media posts. These data streams are captured and monitored in real-time, in more than 160 languages, in order to provide cybersecurity, reputation management, compliance, etc. Without the capabilities of deep learning, the inputs required to get results would prove incredibly difficult. In essence, deep learning is enabling the practical application of machine learning. So, how does it work?

Inspired by the structure and activity of neurons within the human brain, deep neural networks (DNN) form the basis of deep learning. Through these algorithms, computers are able to identify features in significantly sized datasets and progress that information through layers of the neural network, refining as it goes. This leads to a hierarchical representation of the problem. An insideHPC Special Report — *Riding the Wave of Machine Learning & Deep Learning* — offers a closer look at the differentiation between the two aspects of AI, and the impact they're having on the global business landscape.

This report will help you...

▶ Get started
▶ Choose a framework
▶ Know what applications and technologies to use
▶ Learn from organizations' successes
▶ Connect with the community

Contents

www.insidehpc.com | 508-259-8570 | kevin@insidehpc.com

Developer considerations

Overview

There are many reasons why startups might struggle to fulfill their potential for financial and technological success. Among the many unique challenges they face from initial concept through to expansion, a lack of scalability can be one of the most difficult to overcome. In this section, we'll focus on the capabilities and practical application of machine and deep learning, the frameworks and technologies you need to know about, and the ways that the community can help from the very beginning.

If you're trying to decide whether or not to begin a machine or deep learning project, there are several points that should first be considered:

- Cost
- Need
- Organizational readiness
- Industry readiness
- Competition
- Regulations and compliance
- The pace of innovation

It may sound obvious, but the majority of startups that fail to find traction in the market do so because they've identified a need that doesn't really exist – or at least not enough to be monetized.

Cost can often be the deciding factor. Can your organization afford to embark on this journey, and will your potential customers be able to afford what you're offering? Be realistic when making these assessments. Once that's out of the way, the second issue is one of need. It may sound obvious, but the majority of startups that fail to find traction in the market do so because they've identified a need that doesn't really exist — or at least not enough to be monetized.

Readiness is a question you must ask of yourself and the industry. Is your organization ready (and able) to devote time and resources to integrating

For the past five years or so, the pace of innovation within machine and deep learning has quickened significantly. Will your organization be able to keep up?

machine and deep learning into the pipeline, and is the industry ready to adopt your new solution or service? Another thing to consider is the competition. It's an exciting time for startups, and the potential is huge, but tech heavyweights like Google and Microsoft are also looking to cash in on deep learning. It's worth keeping that in mind when positioning yourself in the market with a specialty.

If they occur, regulation and compliance issues can slow everything down so much that it no longer becomes worth the effort. Finally, is it scalable? For the past five years or so, the pace of innovation within machine and deep learning has quickened significantly. Will your organization be able to keep up?

Where do I begin?

If you're approaching machine or deep learning with no real experience in the design, development and employment of deep neural networks, you're in good company. Very few organizations — and even fewer startups — come staffed with a full roster of data scientists, ready to build a platform on an enterprise scale.

One of the first points it's important to recognize is just how accessible machine and deep learning truly are — though that shouldn't be confused with thinking that these are easy fields to be in. Having the computing power and necessary people skills at your disposal won't guarantee results. After giving careful consideration to the issues highlighted in the overview, the first step is to focus on the tools and infrastructure while remembering that machine and deep learning successes come from more than the algorithms.

How do I choose a framework?

Frameworks, applications, libraries and toolkits — journeying through the world of deep learning can be daunting. The ease with which you'll be able to build and run your application is first determined by the framework you choose. With that in mind, here is a roundup of five of the best-known.

The ease with which you'll be able to build and run your application is first determined by the framework you choose.

Caffe

Developed by a mix of Berkeley AI Research Lab and community contributors, the Caffe deep learning framework has been 'made with expression, speed and modularity in mind'.

The good news is that the set of pre-trained models don't require coding to implement, and GPU training is supported out-of-the-box. The bad news is that multi-GPU training is only partially supported. The number of third-party packages used by Caffe can lead to version skew.

Platform	Linux, Mac OS X, Windows
Language	C++
Interface	Python, MATLAB
OpenMP support	YES
OpenCL support	In development
CUDA support	YES
Pre-trained models	YES
Community support	YES
Open source	YES

Tensorflow

Developed for different language understanding and perceptual tasks by the team at Google Brain, Tensorflow is the open source framework behind Google services like Gmail, and Google Search. It's ideal for second-order gradient differentiation, and computational graph visualizations, but developers may find it has limited debugging capabilities.

Platform	Linux, Mac OS X, Windows
Language	C++, Python
Interface	Python, (Keras), C/C++, Java, Go, R
OpenMP support	NO
OpenCL support	Roadmapped
CUDA support	YES
Pre-trained models	YES
Community support	YES
Open source	YES

Torch

Used by companies like Facebook and Twitter, Torch originated at New York University (NYU) in 2002. Coded in a programming language called Lua, it's one of the simplest machine learning frameworks to set up and deploy. A bonus is that Lua is a user-friendly language that boasts a vast repository of sample code. Another plus is that Torch is available on mobile platforms, while a drawback is that learning materials can be hard to come by.

Platform	Linux, Mac OS X, Windows, Android, iOS
Language	Lua, C
Interface	Lua, LuaJIT, C, CUDA, Utility library for C++/OpenCL
OpenMP support	YES
OpenCL support	Third party
CUDA support	YES
Pre-trained models	YES
Community support	YES
Open source	YES

Apache Mahout

Apache Mahout began as a no-cost open source project by the Apache Software Foundation. Deployed on top of Hadoop using the MapReduce paradigm, the goal was to develop free distributed or scalable machine learning frameworks. While this is a robust framework, it should be noted that it's still under development.

Platform	Cross-platformAPACHE
Language	Jave, Scala
Interface	TBC
OpenMP support	YES
OpenCL support	In development
CUDA support	YES
Pre-trained models	YES
Community support	YES
Open source	YES

Microsoft Cognitive Toolkit (CNTK)

Cognitive Toolkit (CNTK) is an open-source machine learning framework from Microsoft. It's one of the most dynamic frameworks available, with support for algorithms such AS CNN, LSTM, RNN, Sequence-to-Sequence and Feed Forward. This also happens to be the only public toolkit that can scale GPUs beyond a single machine. The lack of OpenCL support is a drawback, however.

Platform	Windows, Linux, (OSX via Docker on roadmap)
Language	C++
Interface	Python, C++, Command line, BrainScript (.NET on roadmap)
OpenMP support	YES
OpenCL support	NO
CUDA support	YES
Pre-trained models	YES
Community support	YES
Open source	YES

These are five of the frameworks, but you may still be wondering how to choose between them. The answer is that it really depends on what your goals are. If in doubt, it can be helpful to go with one of the more popular or supported frameworks like Caffe or Torch. For some in-depth advice, contact details for experts at Dell EMC and NVIDIA can be found on page 12.

To learn about how other start-ups are effectively using deep learning today, visit NVIDIA'S deep learning blogs.

You can also learn more at NVIDIA's Deep Learning Institute.

What hardware and software do I need to know about?

Deploying the right kit can be critical, and the main thing is the significant advantages that GPU acceleration provides. GPUs and deep learning go together like a marriage made in heaven. The multi-layered nature of the deep neural networks means that they run best on highly parallel processors. Deep learning training and inference (see section: Getting serious — a guide to scaling up and out) will, therefore, be achieved much faster on GPUs — any GPUs — from small workstations to some serious hardware. In fact, you can start developing on any GPU-based system.

The insideHPC Special Report, *Riding the Wave of Machine Learning & Deep Learning*, explains it well: 'the high compute capability and high memory bandwidth make GPUs an ideal candidate to accelerate deep learning applications, especially when powered with NVIDIA's Deep Learning software development kit (SDK) that includes CUDA® Deep Neural Network library (cuDNN), a GPU-accelerated library of primitives for deep neural networks, TensorRT™, a high performance neural network inference engine for production deployment of deep learning applications, and CuBLAS a fast GPU-accelerated implementation of the standard basic linear algebra subroutines.'

The NVIDIA cuBLAS library is a fast GPU-accelerated implementation of the standard basic linear algebra subroutines (BLAS). Using cuBLAS APIs, you can speed up your applications by deploying compute-intensive operations to a single GPU or scale up and distribute work across multi-GPU configurations efficiently.
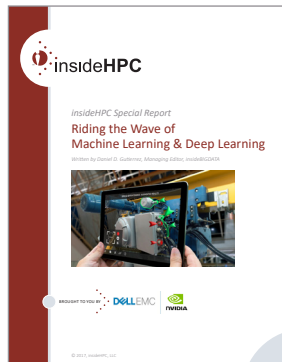
How can the community help me?

As mentioned earlier, and as to be expected given the nature of deep learning, this is a fast-paced and progressive area to be in. Luckily, it's also one of the most vocal. From white papers to forum blog posts, the Internet is full of opinions, debates, user experiences and expert advice. If you have a question that needs an in-depth answer (or even a superficial one, for that matter) or are simply seeking out opinions, search engines will only get you so far.

By joining one of the many community resources, you'll be in direct contact with thought-leaders, technical experts and, perhaps most importantly, people who have launched deep learning startups. There are even community resources that compile and summarize other community resources. This blog post is a good example of a round-up of machine learning communities.

Further in this report, in the section 'What next?', you will find details on the thriving Dell EMC HPC Community, as well as other resources for you to explore.

What questions should I be asking?

The first question that any developer reading this report should be asking themselves at this point, is 'where do I want to go with my startup?' There's a reason why many job interviews feature the question 'where do you see yourself in five years?' Long-term goals can and should dictate short-term plans. So, what do you want to do with your startup? Are you developing something that can then be sold to one of the larger players in the market, or are you building up a business? Will you be focusing on educational services or products, or commercial ones? And what level of kit do you need? By asking yourself these questions now, you can crystalize your goals and begin to take steps towards achieving them. That last question regarding the kit can be answered in the next section of this report.

Getting serious – a guide to scaling up and out

AI starter kit

What is it?
Dell EMC PowerEdge C4130 server, a flexible, dense 1U rack server optimized for GPUs and co-processors.

What does it do?
Put simply, the PowerEdge C4130 server handles the most demanding workloads, including high-performance computing, data visualization and rendering.

Tech specs and features

- Five unique configurations, including up to four 300W double-width NVIDIA® Tesla® GPU accelerators in just 1U of space
- Up to 2x Intel® Xeon® processor E5 family
- Red Hat® Enterprise Linux operating system
- Memory: DDR4 DIMMs at up to 2400MT/s; 16 DIMM slots: 4GB/8GB/16GB/32GB/64GB
- Storage of up to Up to 2 x 1.8" SATA SSD boot drives. Optional data drive tray supports up to 4 x 2.5" SAS/SATA drives, and optional 12Gb/s SAS and 6Gb/s SATA (with PERC9 card).
- Optional internal controller: PERC H330, H730, H730P and H810
- 2x 1GbE network controller
- Power: 1100W, 1600W, 2000W hot-plug PSU; redundant PSU (1+1) option available

How will it help me?
The PowerEdge C4130 server offers a combination of flexibility, efficiency and performance in a compact package that reduces cost and management requirements. The balanced architecture ensures that workload requirements are met through a flexible combination of accelerator, processor, memory and bandwidth.

NVIDIA® Tesla® P100
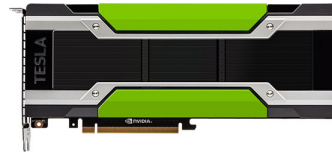
What is it?
NVIDIA® Tesla® P100 GPU accelerator for PCIe-based servers, and NVIDIA NVLink-optimized servers

What does it do?
By tapping into the new NVIDIA Pascal™ GPU architecture — an architecture that's purpose-built for computers that embrace deep learning — these accelerators deliver the world's fastest compute node. This higher performance dramatically increases throughput without increasing the cost.

Tech specs and further features

- Enables a single node to replace half a rack of CPU nodes
- Compute and data are integrated on the same package using Chip-on-Wafer-on-Substrate with HBM2 technology for 3X memory performance over the previous-generation architecture
- Pascal delivers more than 18.7 TeraFLOPS of FP16, 4.7 TeraFLOPS of double-precision, and 9.3 TeraFLOPS of single-precision performance
- A server node with NVIDIA NVLink can interconnect up to eight Tesla P100s at 5X the bandwidth of PCIe.
- A single GPU-accelerated node powered by four Tesla P100s interconnected with PCIe replaces up to 32 CPU nodes
- Applications can scale beyond the GPUs physical memory limit to almost limitless levels without losing performance, thanks to simpler programming and computing performance tuning.

How will it help me?
Instead of focusing efforts on managing data movements, developers are free to focus more on tuning for performance. Essentially, with a reduction in time to results from months to days you'll be able to do more, in less time, and save money while you're at it.

NVIDIA® NVLink™

What is it?

NVIDIA® NVLink™ is a high-bandwidth, energy-efficient interconnect that supports ultra-fast communication between the CPU and GPU, and between GPUs

What does it do?

It enables data sharing at speeds between five and 12 times faster than the traditional PCIe Gen3 interconnect, meaning that applications get a dramatic boost in performance.

Tech specs and further features

- Provides an energy-efficient, high-bandwidth path between the GPU and the CPU at data rates of at least 80 gigabytes per second, or at least five times that of the current PCIe Gen3 x16

- GPUs and CPUs can quickly and seamlessly access each other's memory

- Significant performance benefit for GPU-to-GPU (peer-to-peer) communications

- Unified Memory and NVIDIA NVLink represent a powerful combination for CUDA® programmers

NVIDIA NVLink will let data move between GPUs and CPUs five to 12 times faster than they can today. Imagine what would happen to highway congestion in Los Angeles if the roads expanded from 4 lanes to 20.
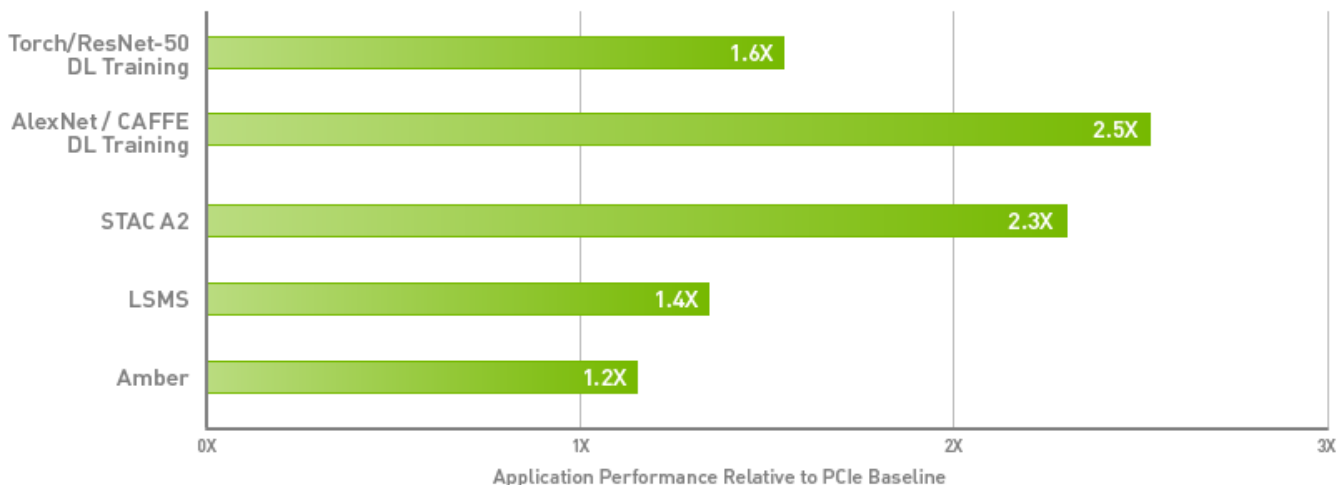
How will it help me?

The unique memory bandwidth and increased throughput enables more advanced modeling and techniques for data-parallel programs. The resulting faster training and better scaling provides deep learning training models with a step up, especially when dealing with out-of-core computations (the input data is too large for the device memory of the GPU).

Application Performance (8-GPU Configurations)
GPU-Based Server with NVlink vs. PCIe

Application	Performance
Torch/ResNet-50 DL Training	1.6X
AlexNet / CAFFE DL Training	2.5X
STAC A2	2.3X
LSMS	1.4X
Amber	1.2X

Application Performance Relative to PCIe Baseline

Dual-Intel® Xeon® E5-2690 v4 @ 3.5GHz Turbo (Broadwell-EP) | 8x NVIDIA® Tesla® P100, ECC off, Autoboost ON | Datasets: AMBER 16, Deep Learning Training: ImageNet, STAC A2: 10-100k-1260 (Warm Creek), LSMS: Fe16

Inference systems

What are they?
Inference systems provide the second piece of the deep learning puzzle by applying capabilities to the data.

What do they do?
Deep learning can be broken down into two parts: training and inference. When the deep learning neural network has been trained on what to look for, the inference system essentially makes predictions or 'infers' based on the input data in order to provide results. Netflix's recommendation engines are a prime example of the power of inference.

Tech specs and further features
A great example of an inference system is NVIDIA's TensorRT™. This high performance deep learning inference engine maximizes inference throughput and efficiency, and provides the ability to take advantage of fast reduced precision instructions provided in the Pascal GPUs. TensorRT v2 delivers up to 45x faster inference under 7 ms real-time latency with INT8 precision.

How will it help me?
Inference systems will optimize, validate and deploy your trained neural network, regardless of how demanding your throughput requirements might be. Multiple network topologies, like AlexNet or Caffenet, tend to be supported. In the case of TensorRT, developers can avoid having to spend their time performance tuning for inference deployment, and instead focus on developing novel AI-powered applications.

NVIDIA's Inception Program provides AI startups with powerful GPU tools, tech, and deep learning expertise. Please visit here.

More info can be found here.

Deep learning hardware

We can't look at the latest hardware for deep learning without giving a nod to the Dell EMC PowerEdge R730, and R740 servers.

In just 2U of rack space, the PowerEdge R730 server packs a punch, thanks to a combination of powerful processors, large memory, fast storage options and GPU accelerator support. It's scalable and configurable, enabling you to adapt to virtually any workload. Vital statistics include the Intel® Xeon® processor E5-2600 v4 product family, and up to 24 DIMMs of DDR4 RAM.

It's highly scalable storage features up to 16 x 12Gb SAS drives, while the high-performance 12Gb PowerEdge RAID Controller (PERC9) is an ideal tool for your virtualized environment. Data access can further be boosted by an optional SanDisk® DAS Cache application acceleration technology.

The equally impressive new PowerEdge R740 offers an ideal balance of accelerator cards, storage and compute resources in a 2U, 2-socket platform. The R740 boasts up to 16 x 2.5" or 8 x 3.5" drives and iDRAC9, as well as up to three 300W accelerator cards or six 150W cards. It's scalable, versatile and can simplify the entire IT lifecycle.

The results

Sustainable business growth, fraud protection and life sciences are just some of the successes organizations have been able to achieve through the application of deep learning technologies and techniques. Here are a few examples:

'Hello, I'm Alice'

Describing itself (or herself) as a "connector" on a data-driven, scalable ecosystem named Team Circular Board, Alice is the world's first artificial intelligence platform for women entrepreneurs. Developed in partnership with Dell EMC and Pivotal, Alice acts as a virtual mentor that helps female founders to achieve and accelerate sustainable business growth by connecting them, in real time, to the resources they need. These connections are based on factors like startup stage, industry and revenue.

Access to an extensive library of resources, coupled with a global network of industry leaders, makes Alice a choice advisor. As more and more users engage with Alice, the data-driven platform will employ machine learning to curate hyper-targeted content and ultimately predict users' needs.

"Alice is a female founder's best mentor, guide and consultant. The current startup ecosystem continues to cycle resources among a very small percentage of well-connected entrepreneurs, making it difficult for less traditional founders to navigate," said Carolyn Rodz, founder and CEO of Circular Board. "Alice instantly filters millions of resources down to the personalized, verified content that enables founders to scale to the highest heights, no matter where they are located or who they know."

▸ www.helloalice.com

www.insidehpc.com | 508-259-8570 | kevin@insidehpc.com
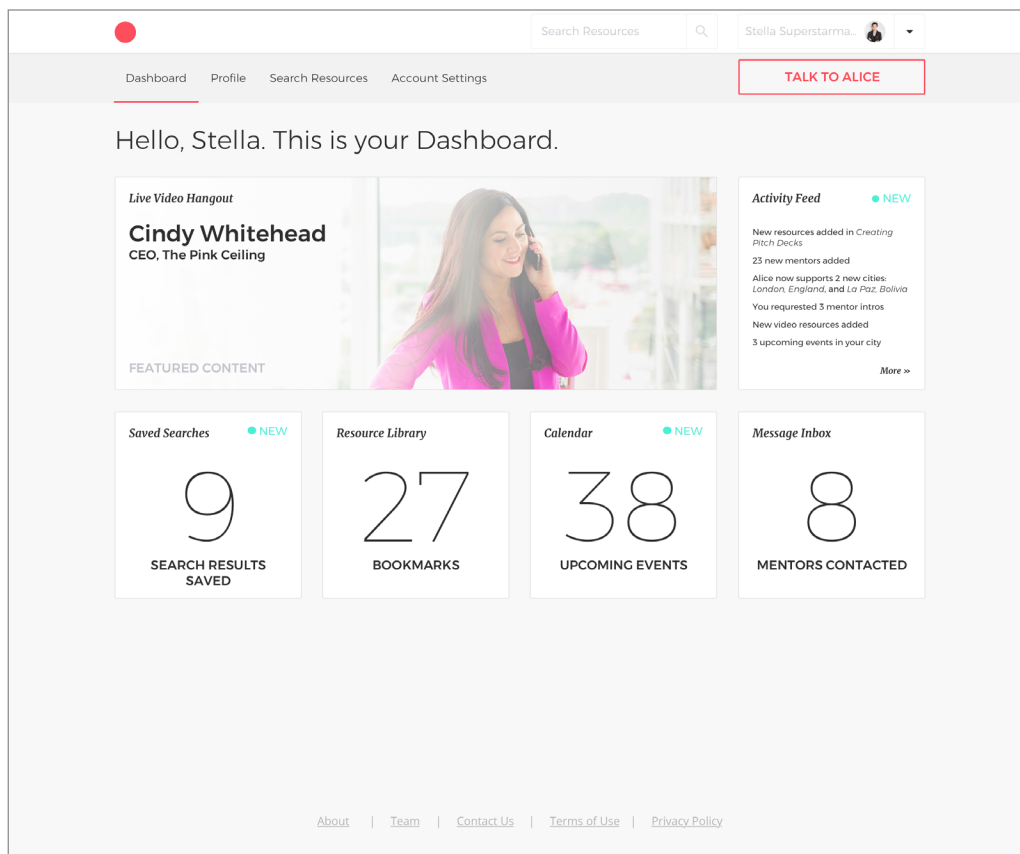
Data security is big business, in fact International Data Corporation forecast that worldwide revenues for security-related hardware, software and services will jump from $73.7 billion in 2016 to $101.6 billion in 2020.

MasterCard

Securing sensitive data: how to navigate the new world of compliance is a white paper that, as the title suggests, delves into how Mastercard achieved the latest compliance standards and ensured optimum data management and security. Data security is big business. In fact, International Data Corporation (IDC) forecast that worldwide revenues for security-related hardware, software and services will jump from $73.7 billion in 2016 to $101.6 billion in 2020.

As highlighted in the white paper, 'evolving compliance standards demand a proactive capability for data security. For all the data available to potential hackers, there is an equal amount of fraud protection and risk measurement available with a compliant approach. In other words, technology has created data, created the ability for hackers to intrude, but has also strengthened the defenses, if managed properly.'

Shirley Inscoe, an Aite banking analyst, is quoted as saying: "When it comes to using big-data technologies for fraud prevention and info-security...the big data rubber is hitting the road. Using Hadoop or similar technology can store and analyze data much more efficiently and in ways that were not previously possible. Processing data is much more efficient using these methods, dramatically reducing both cost and processing time.

"A handful of large FIs [financial institutions] have advanced big-data projects underway, leveraging Hadoop and machine-learning technologies to combine customer data across products and channels. The results are impressive: faster detection of merchant data compromise, more effective transactional fraud analytics, and fewer frustrated customer as a result of false positive declines."

MIT Lincoln Laboratory

At the leading edge of science and engineering, MIT Lincoln Laboratory has created a 1 Petaflop, 40,000 core supercomputer — and done so in less than one month with support from Dell EMC. In a brief video announcing the new TOP500 System, Dr. Albert Reuther, LLSC Manager, comments: "So much of the world has gone to modeling and sim and computers and in order to enable our staff, they need more processing power than is available on the laptop or desktop.

"Recently, we've been doing a lot of work on how to connect fundamentally different styles of databases... all on the same system, all working for the same purpose."

– Dr. Albert Reuther, LLSC Manager, MIT Lincoln Laboratory

"They also have demand for immediate feedback — it's not enough for them to just drop some jobs into a queue and wait for them; they are looking for engineering turns per hour, and many per day, in order to prototype these algorithms, get feedback and figure out how they're going to improve that algorithm and what they find in the data in order to make decisions for the sponsors that we have. This is the kind of processing that signal processing, image processing, machine learning and physical simulation are a natural fit for."

Dr. Vijay Gadepally, LLSC Research Staff, added: "Recently, we've been doing a lot of work on how to connect fundamentally different styles of databases... all on the same system, all working for the same purpose. We took some publically available data that was taken from an intensive care unit, collected over approximately 10 years, and what we wanted to do was analytics and analysis on this data that was not possible before. We were able to reduce 2-10 times the amount of time taken to do analysis, such as finding patients with similar wave forms."

What next?

At this point, there's a good chance that you're beginning to feel more comfortable with deep learning and what it can do, but are still unsure as to where to go from here. Some of the many questions we've addressed in this report don't have simple answers. And whether you're currently planning to paddle in the waters of deep learning, or dive right in, it's important to remember that there is a steady flow of new tools, algorithms and techniques that can all potentially impact what you're trying to achieve. This is why the many resources of the community are worth exploring and ultimately interacting with.

Here are a few key places to get you started:

Dell EMC HPC Community

With a vision of fostering 'sustained discussions among experts that lead to the design, deployment, operation, and usage of the most effective HPC solutions', the Dell EMC HPC Community is a worldwide technical forum. Within it, members share their insights, suggestions, experiences and

The valuable information this community provides on events, presentations and videos, as well as links to white papers on a vast array of topics, can all help ensure you can fully leverage the deep learning potential.

more, and work together to push the boundaries of HPC innovation. The line between HPC and big data has blurred, and these growing data sets coupled with the technologies capable of analyzing them is allowing deep learning to thrive. In essence, the valuable information this community provides on events, presentations and videos, as well as links to white papers on a vast array of topics, can all help ensure you can fully leverage the deep learning potential.

▸ http://www.dellhpc.org/

Dell EMC Machine Learning Knowledge Center

Dell EMC has created a haven where thought leaders and industry experts can gather to share their expertise and offer guidance to anyone who's keen to take advantage of deep learning. This community resource includes everything from expert commentary to technical blogs, and news from around the Web. It's helpful that resources and connections to the community are easy to find.

▸ http://custom.crn.com/dell-learningcenter/

Dell EMC TechCenter

By joining the Dell EMC TechCenter, IT professionals can connect with Dell EMC customers, employees, partners and each other, to maintain best practice and share knowledge. Part of the broader Dell EMC community, the TechCenter includes the latest news, wikis, videos and events. Live Tech Chats with Dell EMC experts take place on Tuesdays at 3PM central, and cover a wide range of subjects. The support forums are particularly active, too, so you rarely have long to wait to find what you're looking for.

▸ http://en.community.dell.com/techcenter/

Talk to us

Still have questions? Get in touch with the experts at Dell EMC to discover how to apply deep learning to your organization.

▸ dell.com/hpc

Contact your Dell EMC local representative or authorized reseller.

Further resources

The Circular Board

Comparison of Deep Learning Software

Dell Launches Artificial Intelligence, Machine Learning Solution

Dell PowerEdge C4130 Rack Server

Dell PowerEdge R730 Rack Server

Dell PowerEdge R740

"Hello Alice" … Meet the World's First AI-based Virtual Advisor for Women Entrepreneurs

How NVIDIA NVLink Will Enable Faster, Easier Multi-GPU Computing

insideBIGDATA Guide to Artificial Intelligence & Deep Learning

Machine Learning Communities

MIT Lincoln Laboratory Supercomputing Center Unveils New Top500 System

NVIDIA® NVLink™

NVIDIA® Pascal™ architecture

NVIDIA® TensorRT™

NVIDIA® Tesla® P100 GPU accelerators

NVIDIA® Tesla® P100 GPU accelerators datasheet

NVIDIA® Unified Memory in CUDA 6

What Is NVIDIA NVLink? And How Will It Make the World's Fastest Computers Possible?

OpenCL Caffe

Ripjar

Securing sensitive data: how to navigate the new world of compliance

Thirteen Companies That Use Deep Learning To Produce Actionable Results

Tony Parkinson, Dell EMC & Nick Curcuru, Mastercard- Dell EMC World 2017

Top 10 Machine Learning Frameworks

Worldwide Semiannual Security Spending Guide – Press release

About the author

Beth Harlen is an experienced journalist, editor and marketing professional in High-Performance Computing. In particular, her interests lie in the extension and application of HPC solutions beyond the worlds of academia and scientific discovery. Beth holds a First Class Honors degree, is an avid console gamer, and has served as editor on four distinct b2b magazines — most recently, Scientific Computing World.