Deep Learning Component of AI

The punchline: Deep Learning is a High Performance Computing problem

- Delivers benefits similar to HPC in other disciplines
  - The value is in the decisions that are enabled
- Characterized by the same underlying factors
  - Large amount of computation
  - Large amount of data motion (I/O and network)
- The same methods work
  - HPC Technology and HPC Best Practice apply directly to DL
Deep Learning Training: Behind the Scenes

Computationally-intensive training phase

Deploying lots of computational power requires lots of communication.
Why Are We Here?

Faster is better

More accurate is better

High Performance Simulation

High Performance Machine and Deep Learning

Communication Intensive

Computationally Intensive

Copyright© 2017 Cray Inc.
Let’s Use Weather As An Example

- **More Accurate is Better**
  - At 100km (top) and 25km (bottom)
  - Missed tropical cyclones and big waves up to 30 meters high

- **Faster is Better**
  - Higher resolution simulation requires 64X more computation

HPC and AI Will Converge

2x Digital data is doubling in size every two years, and by 2020 the digital universe will reach 44 zettabytes.

28% believe HPC will allow them to scale computationally to build deep learning algorithms that can take advantage of high volumes of data.

40% Reduction in error rates when 10x more data is being used in coordination with AI in speech recognition.

2. EMC Digital Universe with Research & Analysis by IDC

Copyright© 2017 Cray Inc.
**What is Deep Learning?**

**ARTIFICIAL INTELLIGENCE**
Design of intelligent systems that augments human productivity. Systems that help decision makers do what they do best; leveraging computers doing what they do best.

<table>
<thead>
<tr>
<th>Sense</th>
<th>Comprehend</th>
<th>Predict</th>
<th>Act and Adapt</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ANALYTICS</strong></td>
<td></td>
<td><strong>MACHINE LEARNING</strong></td>
<td></td>
</tr>
<tr>
<td>Search for the what, when, where and why</td>
<td>Learn patterns from the past to predict future</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leverage domain and data science to query datasets for insights:</td>
<td></td>
<td><strong>Unsupervised</strong></td>
<td><strong>Supervised</strong></td>
</tr>
<tr>
<td>Descriptive</td>
<td>What happened?</td>
<td>Group, cluster and organize content with domain-specific heuristic models</td>
<td>Train mathematical predictive models with labelled data</td>
</tr>
<tr>
<td>Diagnostic</td>
<td>Why did it happen?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictive</td>
<td>What will happen?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prescriptive</td>
<td>How to make it happen?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DEEP LEARNING**
Train and use neural networks as a predictive model

- Vision
- Speech
- Language
AI and machine learning have reached a critical tipping point and will increasingly augment and extend virtually every technology enabled service, thing or application.”

“The combination of extensive parallel processing power, advanced algorithms and massive data sets to feed the algorithms has unleashed this new era.”

Gartner’s Top 10 Strategic Technology Trends for 2017

“Fast data is just as important as big data. In 2016, we’ll witness the emergence of a new class of real-time applications in e-commerce and financial technology services powered by super-speedy data analytics. ‘Fast data’ is the second iteration of big data, and it will create a lot of value.”

 Fortune Magazine, December 2015

In a competitive international economy, advanced AI combined with supercomputing are essential ingredients for:

- Solution of strategically important problems
- Maintaining global leadership in industry, government and academia
- Creating next generation technologies, products and services
Deep Learning Will Require Supercomputing

- **An AI Revolution Started For Courageous Enterprises**
  - Yes, Deep Learning Warrants All The Fuss
  - Expect To Need Thousands Of Cores
Deep Learning with Supercomputers
NERSC – Deep Learning in Science

Opportunities to apply DL widely in support of classic HPC simulation and modelling

Modeling galaxy shapes
Clustering Daya Bay events
Decoding speech from ECoG

Detecting extreme weather
Classifying LHC events
Oxford Nanopore sequencing
Deep Learning in Automotive

Noise, Vibration and Harshness at Daimler

- Noise, Vibration and Harshness is a traditional HPC application used in automotive and aerospace.
- Deep Learning has the potential to do an automatic evaluation of results in complex, multi-component, non-linear applications.
Deep Learning Examples in Manufacturing

**Aerospace Drones**
10-fold increase in the commercial drone fleet by 2021…FAA, 2017

**Digital Twin**
“Top 10 technologies for 2017”, Gartner

**Autonomous Vehicle**
OEMs will invest $7 billion in development…Frost & Sullivan, 2016

Leveraging data analytics and deep learning between engineering disciplines and across the enterprise has great potential for product quality and innovation.
When Should You Start?

A Sample from the Financial Services Sector

- ROI payoff will be 1 – 2 years
- Time to begin experimentation is now

ROI Timeline

- See significant ROI: 10%
- Beginning to see ROI: 25%
- Will not see ROI imminently: 46%
- Will not see ROI for sometime: 17%

Source: Innovita Partners, 7/2017, exclusively for Cray
Why Deep Learning Now?

"Large Enough" Data to Train

+ Compute Power

+ Advanced Algorithms and Software Frameworks

+ Data Science Expertise

= Deep Learning Now

Electronic brain

Perceptron ADALINE XOR Backpropagation SVM Deep Learning


Deep Learning Challenges

“AI systems still demand considered design, knowledge engineering and model building”, Forrester AI TechRadar Q1 2017

- A lot to learn for practitioners and end-users:
  - Large, complex workflows
  - Different Toolkits + Data Movement + Network
  - Defining the value returned to the business

- Training times grow with data sizes and complexity:
  - Days to Weeks
  - Compounded with hyper parameter optimization (O(1000) is not unrealistic)
HPC and AI

Enabling resource intensive training by delivering performance efficiencies and scalability

- Deep Learning Platforms - dense GPU to scalable platforms with optimized software stacks
- Apply HPC best practices and expertise to improve deep learning frameworks and core algorithms
Reduce Total Workflow Time

Why? The Deep Neural Net Training Problem

- DNN model with weights on all connections
  - Largest models now hundreds of layers, and millions (to billions) of nodes
- Large set of labeled training data
- Idealized training algorithm:
  - For every *minibatch* of training samples:
    - run samples forward through the model
    - compute the error vs. the training data
    - back-propagate error through the NN to update the weights (gradient descent)
- After all data processed, iteratively optimize *hyperparameters* until required accuracy is achieved

A (not particularly deep) neural net
Reduce Total Workflow Time

- **Minutes, Hours:**
  - Interactive research! Instant gratification!
- **1-4 days**
  - Tolerable
  - Interactivity replaced by running many experiments in parallel
- **1-4 weeks:**
  - High value experiments only
  - Progress stalls
- **>1 month**
  - Don’t even try

Apply HPC best practices and expertise to improve deep learning frameworks and core algorithms

Source: Large-Scale Deep Learning for Intelligent Computer Systems, Jeff Dean, Google
Applying a supercomputing approach to optimize deep learning workloads represents a powerful breakthrough for training and evaluating deep learning algorithms at scale. Our collaboration with Cray and CSCS has demonstrated how the Microsoft Cognitive Toolkit can be used to push the boundaries of deep learning.

- Dr. Xuedong Huang, distinguished engineer, Microsoft AI and Research
“Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.”

-Adapted from *Hidden Technical Debt in Machine Learning Systems*, Sculley et. al., NIPS ‘15
HPC Supports the Entire AI Workflow

Deep Learning workflows are not limited to training.

- Similar to other HPC and analytics workloads, significant portions of DL jobs are devoted to data collection, preparation and management.
AI is everywhere... Even the grocery store

Trending: Seattle Seahawks players talk about their experience wearing new high-tech Vicis helmet

Whole Foods offers Amazon Echo as ‘Farm Fresh Pick of the Season’ as tech giant takes over upscale grocer

BY NAT LEVY on August 28, 2017 at 8:00 am
Thank You