Building High Performance Solutions for Machine Learning and Data Analytics with Hewlett Packard Enterprise

Volodymyr Saviak, October 2019
Introduction
Explosion of data and evolving customer needs for data intensive workloads are driving AI & HPC growth

**Industry** | **Example Workloads**  
--- | ---  
Government | Surveillance, Encryption, Communications Intelligence  
National Labs | Scientific & Industrial Research  
Weather | Weather & Climate Modeling  
Financial Services | Portfolio Optimization, HF Trading, Global Risk Management  
Media & Content Delivery | Digital Content Creation & Distribution  
Manufacturing | Computer Aided Engineering, Product Design  
Oil & Gas | Geoengineering, Chemical Engineering, Seismic Exploration, Reservoir Simulation  
Life Sciences | Genomics, Drug Discovery, Bioinformatics, Predictive Medicine  
Academic | Scientific & Industrial Research
HPC, Big Data Analytics and AI are distinct approaches with some overlaps

- Simulation of intelligent behavior
- Learn from data and make predictions
- Model high level abstractions in data using artificial neural networks
- Advanced analytics to forecast future activity, behavior and trends.
- Uncover hidden patterns, unknown correlations in large data sets
- Simulation & modeling with highly complex mathematical models to gain insights

- Human: filter vs accumulate
- Big Data: intensity of human-in-loop analyst
- Access (data): enables training
- Low Cost HPC: Moore’s low & ½ precision
Traditional Approach for Intelligent System Creation

\[ \nu = \Delta x \Delta y \sqrt{\left( \frac{\partial u}{\partial x} \right)^2 + \left( \frac{\partial v}{\partial y} \right)^2 + \frac{1}{2} \left( \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right)^2} \]

* - Formula/model created by humans
Artificial Neural Network Is a Popular Approach to Build AI

* - Machines learns to be intelligent by training, based on the data received

\[ z_0^1 = \sum_{k} w^l_{jk} x^{l-1}_k + b_0^1 \]
A quick introduction to Neural Networks
The (artificial) neuron.

Artificial Neural Networks (ANNs) are inspired by biological systems similar to our brain.

NNs are made up of neurons, which are a mathematical approximation to biological neurons.

- Inputs: $X_0, X_1, X_2, X_3, X_4$
- Weights: $w_{0,1}, w_{0,2}, w_{1,3}, w_{1,4}, w_{1,5}$
- Bias: $b_0$
- Output: $y_1$

The activation function is $a_0^l = f(z_0^l)$, where $z_0^l = \sum_k w_{jk}^l x_k^{l-1} + b_0^l$.
MegaFace Challenge
Detail Information available at [http://megaface.cs.washington.edu](http://megaface.cs.washington.edu)

<table>
<thead>
<tr>
<th>Team Name</th>
<th>Date</th>
<th>Accuracy 1</th>
<th>Accuracy 2</th>
<th>Accuracy 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sogou AIGROUP - SFace</td>
<td>9/5/2018</td>
<td>99.939%</td>
<td>99.939%</td>
<td>99.939%</td>
</tr>
<tr>
<td>SRC-Beijing-FR (Samsung Research Institute China-Beijing)</td>
<td>8/15/2018</td>
<td>99.888%</td>
<td>99.888%</td>
<td>99.888%</td>
</tr>
<tr>
<td>SenseTime PureFace (clean)</td>
<td>6/13/2018</td>
<td>99.801%</td>
<td>99.801%</td>
<td>99.801%</td>
</tr>
<tr>
<td>SIAT_MMLAB</td>
<td>3/29/2016</td>
<td>65.233%</td>
<td></td>
<td>65.223%</td>
</tr>
<tr>
<td>Beijing Faceall Co. - FaceAll_Norm_1600</td>
<td>10/19/2015</td>
<td></td>
<td>64.803%</td>
<td>64.798%</td>
</tr>
<tr>
<td>Beijing Faceall Co. - FaceAll_1600</td>
<td>10/19/2015</td>
<td>63.977%</td>
<td></td>
<td>63.962%</td>
</tr>
<tr>
<td>Barebones_FR - cnn</td>
<td>10/21/2015</td>
<td>59.363%</td>
<td></td>
<td>59.379%</td>
</tr>
</tbody>
</table>
HPE Accelerates AI with Intel

The next section is presented by Intel (slides)
Breaking barriers between AI Theory and reality
Partner HPE with Intel® to accelerate your AI journey

CHOOSE ANY APPROACH
from machine to deep learning

SPEED UP DEVELOPMENT
using open AI software

DEPLOY AI ANYWHERE
with unprecedented HW choice

TAME YOUR DATA
with a robust data layer

SIMPLIFY AI
using community solutions

DEPLOY AI ANYWHERE
with unprecedented HW choice

SPEED UP DEVELOPMENT
using open AI software

TAME YOUR DATA
with a robust data layer

SIMPLIFY AI
using community solutions
**Intel® AI Case study**

**Foundation**

**Identify** prospects internally. You can use 70+ AI solutions in Intel’s portfolio then assess business value of each one.

**Prioritize** projects based on business value & cost to solve with Intel guidance; choose industrial defect detection via DL.

**Consider** ethical, social, legal, security & other risks and mitigation plans with Intel advisors prior to kickoff.

**Organize** internally to get buy-in, support new development philosophy & grow developer talent via Intel AI.

Value (H) vs. Cost (L) diagram:

- **Value (H):** Various icons representing different industries and technologies.
- **Cost (L):** Gradient from low (L) to high (H), with a focus on corrosion.
Intel® AI Case study

**Ingest**
Ingest streaming data from drones using a popular software tool among the many that run on the CPU

**Store**
Store data in block storage (for high-performance) in a data lake with guidance from an Intel storage partner

**Process**
Process data by performing cleanup & integration using popular software tools that run on the CPU

**Manage**
Manage data via a popular framework for distributed computation on your infrastructure

12 weeks
Intel® AI Case study

Setup compute environment; DL training (~7% of journey) acceleration NOT worthwhile due to high setup time & cost

Model development through training a deep neural network using an Intel-optimized DL framework

Test the deep learning model using a control data set to determine if accuracy meets requirements

Document the code, process, and key learnings for future reference

Setup (Topology Experiments) 30%  
Train (Tune Hyper-parameters) 30%  
Test Inference 20%  
Document Results 10%  

12 weeks
Intel® AI Case study

Deploy

Architect AI deployment with Intel AI Builders

Implement AI in production environment

Scale to more sites & users as demand grows

Iterate on the models with new data over time

Data Ingest

Drones

Prepare Data

Model Store

Inference

Remote Devices

10 Drones
Real-time object detection and data collection

Media Store

Prepare Data

Service Layer

Media Server

Multi-Use Cluster

Data Store

Model Store

Label Store

Per Drone

1x Intel® Core™ processor
1x Intel® Movidius™ VPU

Per Node

1x 2S 61xx
20x 4TB SSD

Adv. Analytics

Software

- OpenVino™ Toolkit
- Intel® Movidius™ SDK
- Intel® MKL-DNN
Turnkey Container-Based Platform for AI / ML / DL & Big Data Analytics

BlueData EPIC® Software Platform

- **Data Scientists**
- **Developers**
- **Data Engineers**
- **Data Analysts**

**EPIC = Elastic Private Instant Clusters**

- **ElasticPlane™** – Self-service, multi-tenant clusters
- **IOboost™** – Extreme performance and scalability
- **DataTap™** – In-place access to data on-prem or in the cloud

* EPIC = Elastic Private Instant Clusters

- Big Data Tools
- ML / DL Tools
- Data Science Tools
- BI / Analytics Tools
- Bring Your Own

**Compute**
- CPU's
- GPU's

**Storage**
- NFS
- HDFS

**On-Premises**

**Public Cloud**

- Amazon Web Services
- Microsoft Azure
- Google Cloud Platform

**Data Scientists**

**Developers**

**Data Engineers**

**Data Analysts**

What are the benefits of an elastic platform?

- Scale nodes / resources independently
- Add compute nodes without repartitioning data
- Shift node purpose on-the-fly
- Containers enable rapid deployment and movement of workloads and models

HPE Flexible Capacity for consumption-based IT
An elastic architectural quickly shifts workload requirements

1. Initial cluster deployment (Spark and Batch workloads)
2. Need more CPU and RAM (Impala workloads)
3. Need more storage capacity (archival tier)
4. Adding low-latency, high-IOPs noSQL and Kudu workloads
5. Adding AI model training and deep learning workloads
Why an elastic architectural approach?
*Purpose-built nodes and multi-generational clusters*

Different requirements along the data pipeline stages demand different node geometries.
Speech analysis is becoming mainstream

70% of all finance and insurance organizations will be using speech analysis to reduce risk by 2022.
—Gartner

By 2020, customers will manage 85% of their relationship with the enterprise without interacting with a human.
—Gartner

About 30% of searches will be done without a screen by 2020.
—Gartner

25% of 16- to 24-year-olds use voice search on mobile.
—Global Web Index

19% of people use Siri at least daily.
—HubSpot
Case in point: Customer service support center

- 3000 employees for average major bank call centers in the U.S.
- 30 calls per agent, per day
- 4 minutes spent per call
- 360,000 minutes spent on calls each day
- $1 average cost per minute
- $86 million annual spend
- Only 60% satisfaction rate
Why HPE
One-day interactive workshop to help customers get started
HPE Artificial Intelligence Transformation Workshop

Customer benefits
One-day visual and interactive session allows HPE experts to focus on customers’ needs for next generation AI and help embark on a transformational journey to achieve business goals.
- Get started on an AI project fast
- Align business, data and IT operations teams
- Explore opportunities, priorities and select relevant use cases
- Identify dependencies, data sources, level of readiness
- Define a high-level roadmap for intelligent data strategy

Scope
- Select and analyze AI, advanced analytics, data use cases
- Identify desired outcomes – automated or manual
- Assess data characteristics – data readiness, requirements

Turn data into action and revenue
Why HPE for AI, HPC, BigData Analytics?

• Best people, best teams World Wide
• Best technologies on the market
• High quality products designed and made in US
• Best customer support
Thank you!

Your AI & HPC Solutions Contact: Volodymyr.Saviak@hpe.com