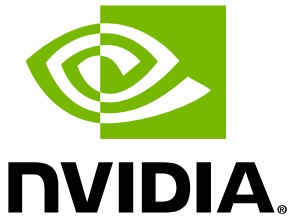


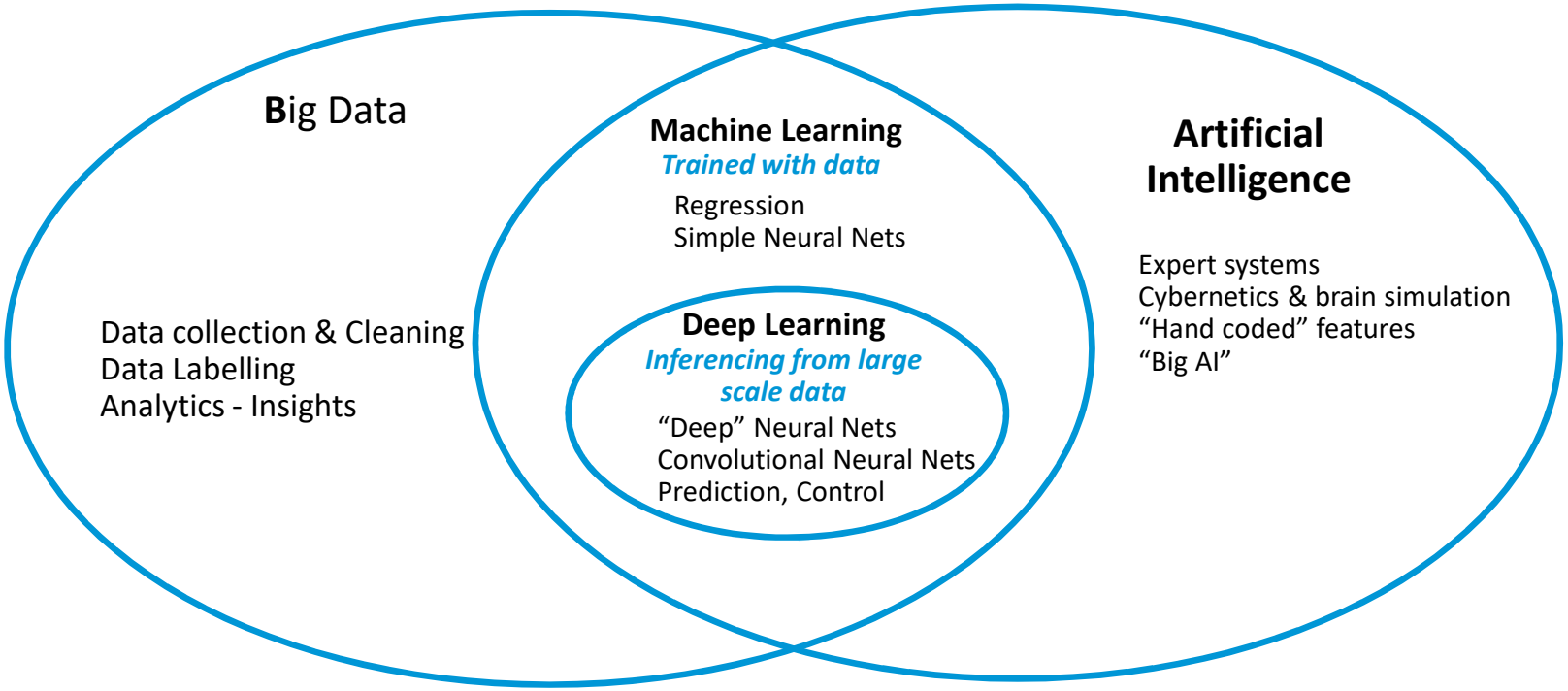
HP Z Workstations for Machine Learning

Agenda

- Introduction
  - Defining Artificial Intelligence (AI), Machine learning (ML), Deep learning (DL)
  - Development /Deployment phases
  - NVIDIA GPU Cloud
  - Workstation configurations
  - HP Workstation Differentiators
- For more information

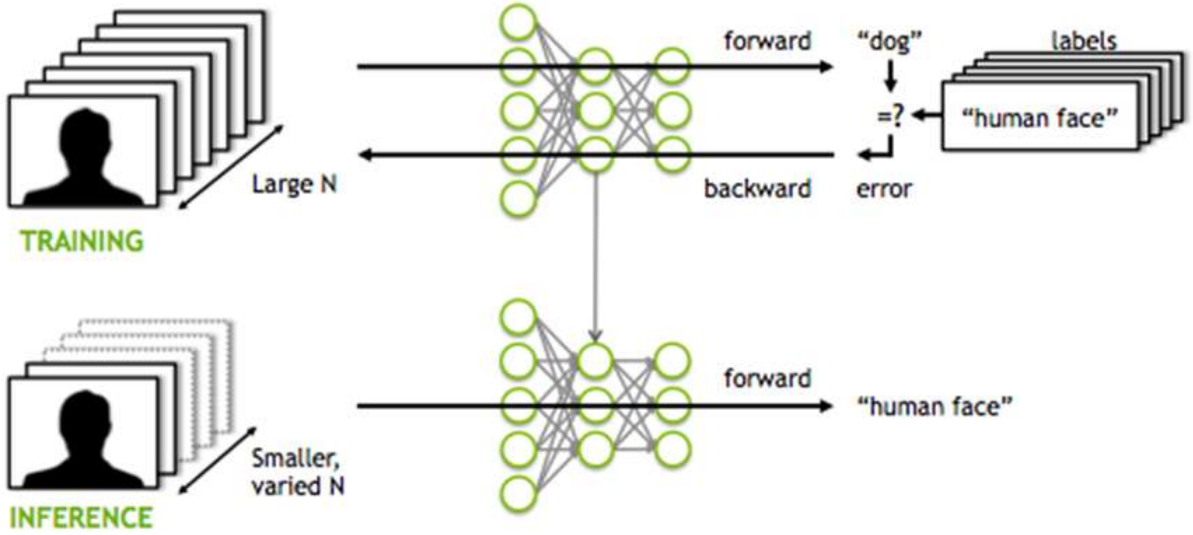
What is AI, Machine Learning, & Deep Learning?

"As soon as it works, no one calls it Artificial Intelligence any more" – John McCarthy, Co-founder of the AI discipline



Why Deep Learning?

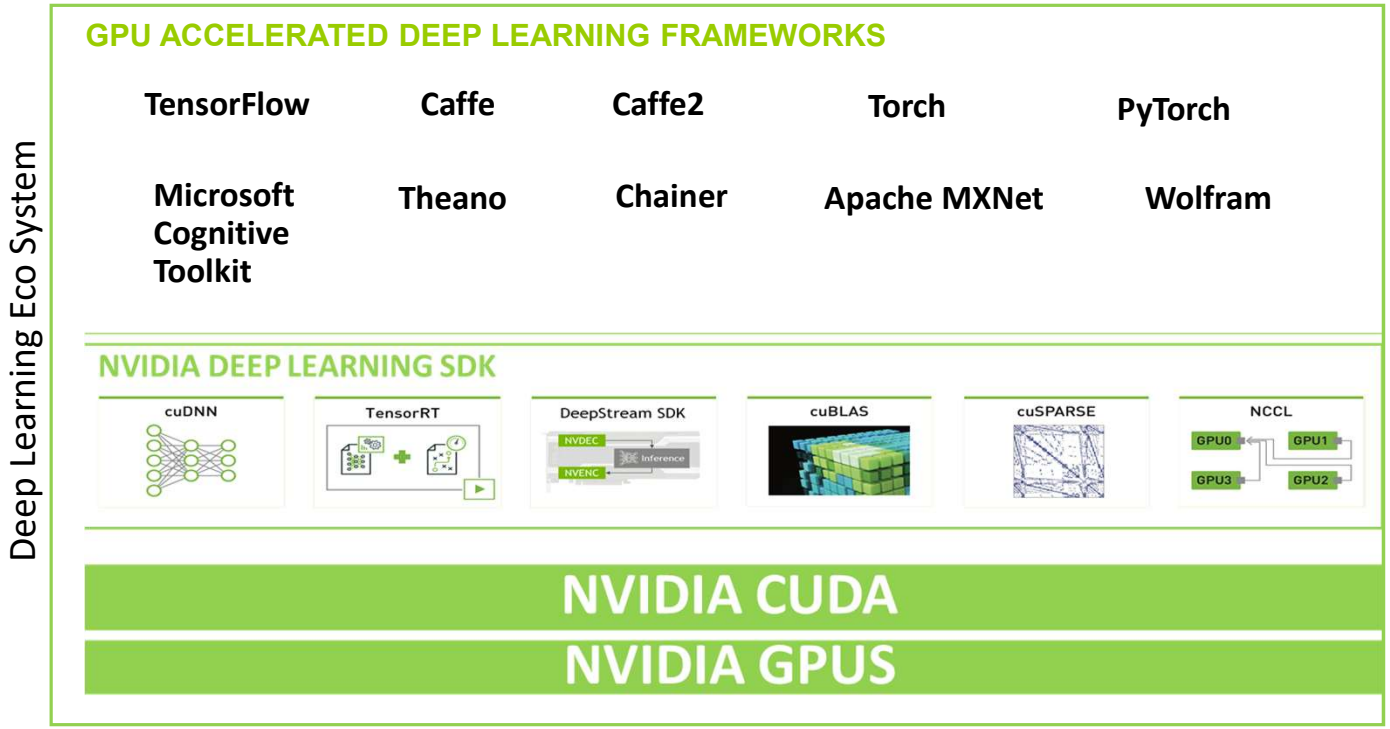
- Deep Learning is a newer subset of machine learning
- Systems “trained” with large sets of data to make decisions or inferences about the data



- Once trained, the deep learning neural network can use new data fed to it to complete task, performing tasks much in the same way that people do by using previous knowledge to complete the task.

Why Deep Learning Now?

- **Parallelized Deep Learning Algorithms** can now harness **Big Data** and powerful **GPU accelerators**



- NVIDIA GPUs and CUDA technology provide the compute power required by Deep Learning

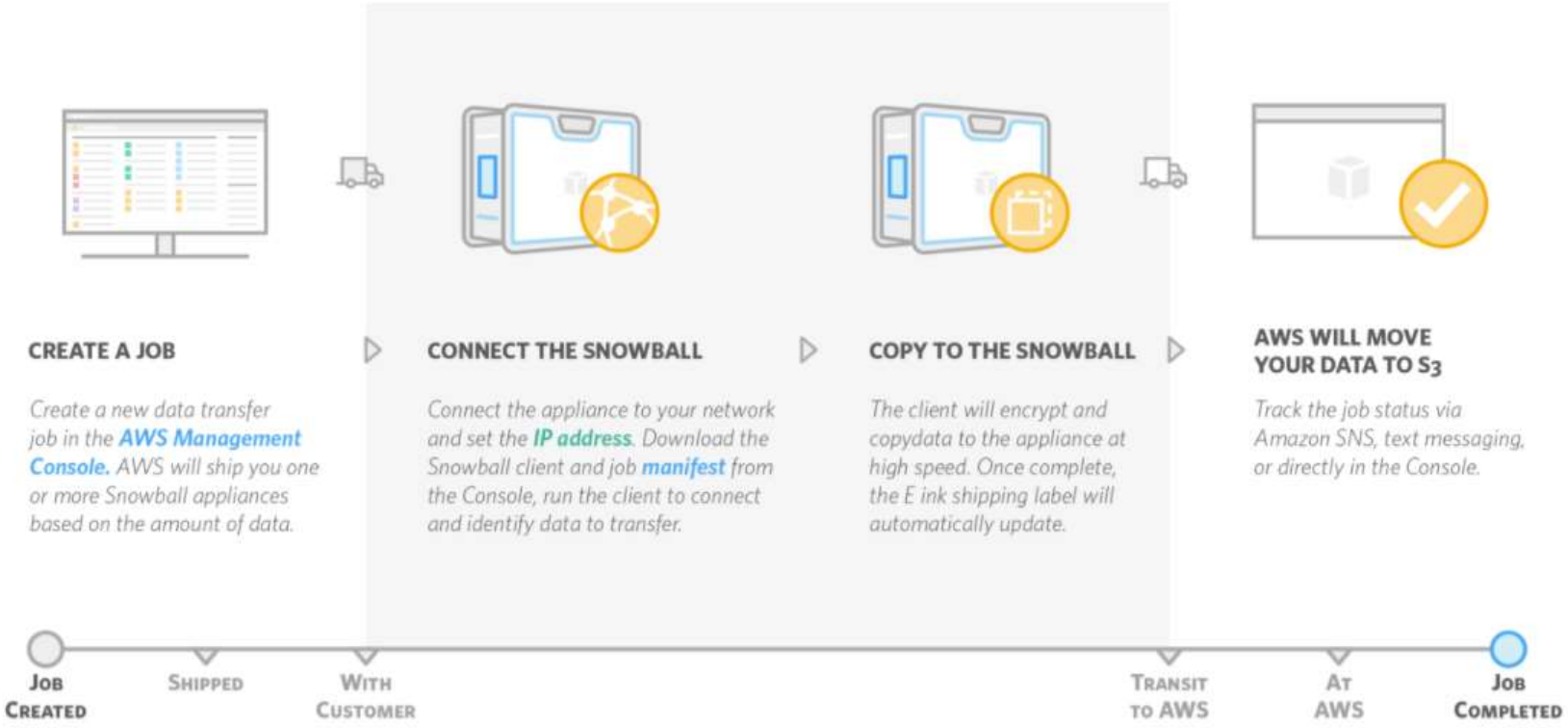
The cloud is being disrupted by Edge Compute

The public cloud has multiple deficiencies in its ability to service edge devices

- **Latency**: Can't handle real-time interactive / immersive services
- **Autonomy**: Can't guarantee 100% connect for critical applications
- **Power**: Inefficient at servicing large data created at the edge
- **Bandwidth:** Available IP capacity is a fraction of generated data
- **Governance:** Regional privacy, security and regulatory issues
- **Scaling**: Can't grow to meet the billions of things coming online

Amazon Petabyte data ingestion solution is Fed-Ex

Amazon Snowball (PB-scale data ingestion)



Amazon Exabyte data ingestion solution is a fleet of semi-tractor trailers

"… it [takes] 26 years to move an Exabyte to the cloud," Andy Jassy, CEO of AWS

Amazon Snowmobile (XB-scale data ingestion)

Requires 10 semi-tractor trailers 6 months to move one Exabyte

Machine Learning is sweeping across industries

Retail & Advertising (Computer Vision)

- Retail sales/inventory mgmt.
- Point of Sale Shrinkage
- Facial identification
- Sentiment Analysis

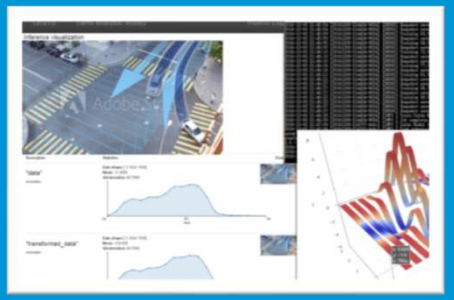
Financial Services (Data Mining)

- Fraud detection
- Portfolio analytics
- Wealth side investments

Media & Entertainment (Computer Vision, Data Mining)

- Character animation
- Sentiment analysis
- Facial identification
- Real-time signage

Software Development (All)

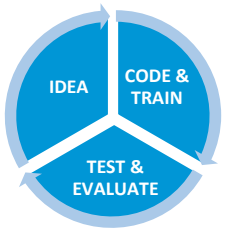


- Data cleaning
- Design and training of models
- Cloud repatriation
- Deploy anywhere

Workflows for Machine Learning Solutions

DEVELOPMENT (Training)

WORKSTATION CLOUD DATA CENTER



Categories

- In-house developers
- Traditional ISVs
- New ISVs

Workstation Benefits

- Time-to-market
- Economic value
- Security

DEPLOYMENT (Inference)

WORKSTATION CLOUD DATA CENTER EMBEDDED

Categories

- Platform choice driven by use case

Workstation Benefits

- Security
- Performance
- Mission-critical

ML Deployment at the Edge

Amazon GO store

HOW DOES IT WORK?

- Customers put items in bag and just walk out
- **Computer vision, Sensor Fusion, Deep Learning eliminate checkout**

TECHNICAL REQUIREMENTS

- **Large amounts of local video data must be processed in real-time**

WHY HP WORKSTATIONS FOR DEPLOYMENT?

- **The Amazon GO solution is not for sale!**
  ….but every retailer on the planet wants this capability
- **Good fit for local workstations**
  - **High capacity, fast performance**
  - **Most secure workstation**
  - **24/7/365[1] capability for mission critical operation**

HP Pixel Intelligence Computer Vision SW

Pixel Intelligence team aggregates training data from customers, a virtuous cycle of collection and refinement

- Pixel Intelligence: Portfolio of machine learning algorithms developed by HP Labs
- Enables the addition of fast and simple digital image analysis to customers' products and services
- Today this is being applied in our Print business, selling to their customers – content creators
- Opportunity to apply this to our workstation customers- movie studios, game developers, VR developers

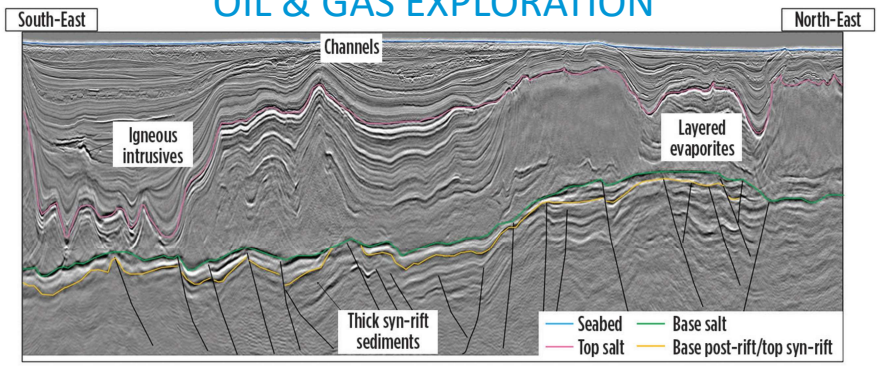
Example: Facial Motion Capture

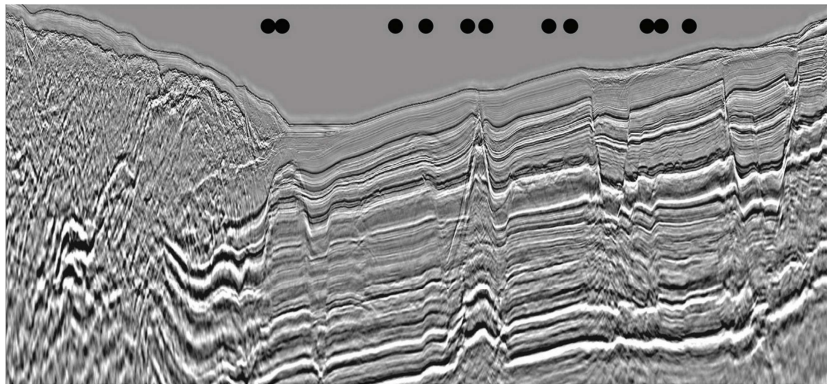
Machine Learning Development at the Edge: CGG

Geoscience workflow transformation – developed on a Z840 workstation

CGG
Passion for Geoscience

OIL & GAS EXPLORATION





"CGG is using AI to develop breakthrough solutions for Geoscience. Our terabyte-sized data sets choke many systems, but our HP Z Workstation platform for Machine Learning at the Edge can handle these extreme workloads. This enables CGG to create our models faster and at lower cost than cloud or other options."

- Steve Dominguez
Team Lead – Seismic Interpretation Software

END CUSTOMER VALUE

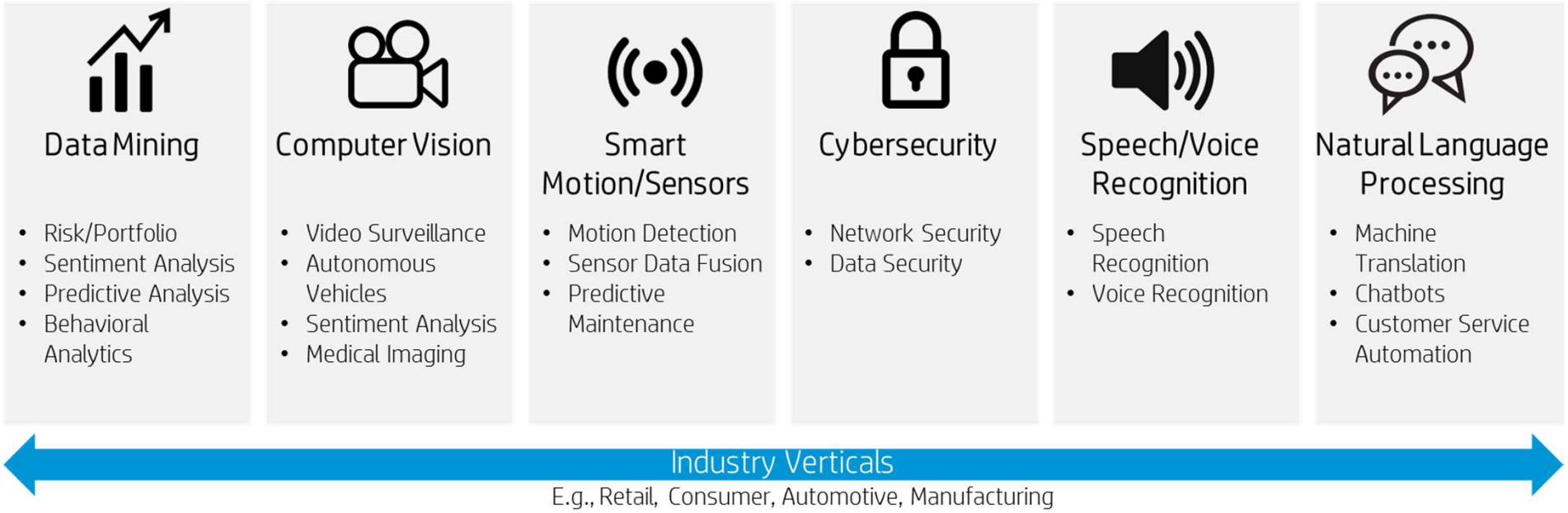
- Results in minutes vs. hours/days/weeks
- Ability to analyze 15x larger data sets
- **Poised to disrupt Oil & Gas exploration**

WORKSTATION VALUE TO DEVELOPER

- Terabyte data sets choke cloud
- Cloud prohibitively expensive (storage & compute)
- Workstation provides faster cycle time on iterations
- Eases path to deployment (also on Z840 workstation)

Most common applications of Machine Learning

These are utilized across many industry verticals



Suggested configurations for Machine Learning

…with a sampling of segments

HP's development platform for large workloads

HP Z8 Workstation

- 2x Intel Xeon® 6136 3.0 GHz, 12 core each, 24 cores total[2]
- Linux® Ubuntu 16.04[5]
- 192 GB DDR4-2666 ECC SDRAM
- OS Drive: 256 GB SATA 6 GB/s SSD
- Data Drive 1: HP Z Turbo Drive M.2 512 GB SSD[3]
- Data Drive 2: HP Z Turbo Drive M.2 512 GB 2nd SSD[3]
- Data Drive 3: 4 TB 7200 rpm HDD
- 3x NVIDIA® Quadro® GV100 32 GB

e.g. Seismic analysis, advanced VR rendering

For medium development loads

HP Z4 Workstation

- Intel® Xeon® W-2155 3.3 GHz, 10 core[2]
- Linux® Ubuntu 16.04[5]
- 128 GB DDR4-2666 ECC SDRAM
- OS Drive: 256 GB SATA 6 GB/s SSD
- Data Drive 1: 512 GB PCIe SSD[3]
- Data Drive 2: 2 TB 7200 rpm HDD[3]
- NVIDIA® Quadro® GP100/ GV100*

e.g. Advanced GIS (Machine Learning), Medical imaging interpretation

For smaller development loads

HP Z4 Workstation

- Intel® Xeon® W-2155 3.3 GHz, 10 core[2]
- Linux® Ubuntu 16.04[5]
- 64 GB DDR4-2666 ECC SDRAM
- OS Drive: 256 GB SATA 6 GB/s SSD
- Data Drive 1: 256 GB PCIe SSD[3]
- Data Drive 2: 2 TB 7200 rpm HDD[3]
- NVIDIA® Quadro® P5000 16 GB

e.g. Retail Computer Vision, commodity trading

Small development/ high-end field deployment

HP ZBook 17 Workstation

- Intel Xeon E-2186M 2.9GHz, 6 core
- Microsoft Windows 10[4]/ Linux® Ubuntu 16.04[5]
- 64 GB DDR4-2666 ECC SDRAM
- OS Drive: 256 GB PCIe SSD
- Data Drive: 1 TB PCIe SSD
- NVIDIA® Quadro® P5200 16 GB

Various segments/ use cases on a mobile platform

Consider for edge deployments

HP Z2 Mini Workstation

- Intel® Xeon® E3-1245v6 3.7 GHz, 4 core[2]
- Microsoft Windows 10[4]/ Linux® Ubuntu 16.04[5]
- 16 GB DDR4-2400 ECC SDRAM
- OS Drive: 256 GB SATA 6 GB/s SSD[3]
- Data Drive: 256 GB PCIe SSD[3]
- NVIDIA® Quadro® M620 2 GB

e.g. Video surveillance, fraud detection

*GV100 option available 2H 2018

Develop remotely on the world's most powerful workstation*

With HP Remote Graphics Software (RGS)

HP Z8 G4

From the thinnest laptop

RGS

Pixels only

KB, Mouse, USB

HP Spectre

includes Linux support

*Based on desktop workstations as of June 14, 2017 and power based on processor, graphics, memory, and power supply

QUADRO GV100

Reinventing the Workstation with Real-Time Ray Tracing and AI

GPU ARCHITECTURE	Volta
MEMORY	32 GB HBM2
DOUBLE-PRECISION	7.4 TFLOPS
SINGLE-PRECISION	14.8 TFLOPS
DEEP LEARNING	118.5 TFLOPS
MULTI-GPU	2-Way NVLink™
VR READY	

Why HP Workstations for Machine Learning?

- Z8: World's most powerful workstation with up to 3x GV100
- Z4: Best price/performance workstation w/1KW PSU & dual GV100 support*
- HP ML Developers Portal
- NVIDIA Strategic Relationship
- Tested and supported with NVIDIA NGC Containers
- HP RGS
- HP DaaS
- HP First Mover in Workstations!

*Dual GV100 option available 2H 2018

For more information

- www.hp.com/zworkstations
- HP Workstations for Machine Learning: www.hp.com/go/ML
- HP Machine Learning Developers Portal: hp.io/ML

HP WORKSTATIONS FOR MACHINE LEARNING

THE FUTURE IS HERE

Thank you!

Footnotes

1. 24/7 use will not void HP Warranty.
2. Multicore is designed to improve performance of certain software products. Not all customers or software applications will necessarily benefit from use of this technology. Performance and clock frequency will vary depending on application workload and your hardware and software configurations. Intel®'s numbering, branding and/or naming is not a measurement of higher performance.
3. For hard drives and solid state drives, 1 GB = 1 billion bytes. TB = 1 trillion bytes. Actual formatted capacity is less. Up to 30 GB of system disk is reserved for system recovery software.
4. Not all features are available in all editions or versions of Windows. Systems may require upgraded and/or separately purchased hardware, drivers, software or BIOS update to take full advantage of Windows functionality. Windows 10 is automatically updated, which is always enabled. ISP fees may apply and additional requirements may apply over time for updates. See http://www.windows.com.
5. HP Linux® Ready software option includes drivers for 64-bit OS versions of RHEL 6 & 7, SUSE Linux® Enterprise Desktop 11 and Ubuntu 14.04. The operating system will need to be installed and managed by the customer.