THE RISE OF GPU-ACCELERATED DATA SCIENCE
#1 STRATEGIC IMPERATIVE FOR THE MODERN ENTERPRISE

85% CIOs investing in AI in the next 3 years¹

- Recognize automation will increase speed and accuracy of decisions²
- Believe it will transform their industry³
- Have invested or deployed AI solutions today⁴

³AI Business Book
### FROM BUSINESS INTELLIGENCE TO DATA SCIENCE

**Forecasting, Fraud Detection, Recommendation, and More**

<table>
<thead>
<tr>
<th><strong>RETAIL</strong></th>
<th><strong>FINANCIAL SERVICES</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply Chain &amp; Inventory Management</td>
<td>Claim Fraud</td>
</tr>
<tr>
<td>Price Management / Markdown Optimization</td>
<td>Customer Service Chatbots/Routing</td>
</tr>
<tr>
<td>Promotion Prioritization And Ad Targeting</td>
<td>Risk Evaluation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>TELECOM</strong></th>
<th><strong>MANUFACTURING</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Detect Network/Security Anomalies</td>
<td>Remaining Useful Life Estimation</td>
</tr>
<tr>
<td>Forecasting Network Performance</td>
<td>Failure Prediction</td>
</tr>
<tr>
<td>Network Resource Optimization (SON)</td>
<td>Demand Forecasting</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>HEALTHCARE</strong></th>
<th><strong>OIL &amp; GAS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Improve Clinical Care</td>
<td>Sensor Data Tag Mapping</td>
</tr>
<tr>
<td>Drive Operational Efficiency</td>
<td>Anomaly Detection</td>
</tr>
<tr>
<td>Speed Up Drug Discovery</td>
<td>Robust Fault Prediction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>CONSUMER INTERNET</strong></th>
<th><strong>AUTOMOTIVE</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad Personalization</td>
<td>Personalization &amp; Intelligent Customer Interactions</td>
</tr>
<tr>
<td>Click Through Rate Optimization</td>
<td>Connected Vehicle Predictive Maintenance</td>
</tr>
<tr>
<td>Churn Reduction</td>
<td>Forecasting, Demand, &amp; Capacity Planning</td>
</tr>
</tbody>
</table>
CHALLENGES AFFECTING DATA SCIENCE TODAY

What is limiting Data Science productivity?

INCREASING DATA ONSLAUGHT

Data sets are continuing to dramatically increase in size
Multitude of sources
Different formats, varying quality

SLOW CPU PROCESSING

End of Moore’s law, CPUs aren’t getting faster
Many popular data science tools have been CPU-only
Can only throw so many CPUs at a job

COMPLEX INSTALLATION & MANAGEMENT

Time consuming to install software
Nearly impossible to manage all version conflicts
Updates often break other software
TRADITIONAL INFRASTRUCTURE SETUP

- Resource availability depends on job queues
- Slower model iteration process
- Software stack needs IT management and support
- Too expensive for everyday development
- Slow data migration
- Security and privacy concerns
GPU-ACCELERATED DATA SCIENCE WORKFLOW

Wrangle Data → Data Preparation → Train → Deploy

Data Sources → ETL → Data Lake → Data Preparation → Train → Evaluate → Predictions

GPU Accelerated Data Science
NVIDIA DATA SCIENCE PLATFORM

FRAMEWORKS
- Chainer
- DASK
- mxnet
- PyTorch
- RAPIDS
- TensorFlow

CLOUD ML SERVICES
- Baidu AI Cloud
- Amazon SageMaker
- Google Cloud ML
- Microsoft Azure Machine Learning
- Databricks

DEPLOYMENT
- AWS
- Amazon SageMaker Neo
- ONNX Runtime
- TensorFlow Serving

CUDA-X AI
- DA
- GRAPH
- ML
- DL TRAIN
- DL INFEERENCE
- VISUALIZATION

CUDA

PC
- DigitalStorm
- Maingear
- Projet Systems

Workstation
- Dell
- HP
- Lenovo

Server
- Cisco
- Dell EMC
- Hewlett Packard Enterprise
- IBM
- Lenovo

Cloud
- AWS
- Google Cloud
- Azure
- 阿里云
NVIDIA GPU-ACCELERATED DATA SCIENCE
A Solution for Every User and Every Organization

ML EXPERIMENTATION

PRODUCTION DATA CENTER
NVIDIA-POWERED DATA SCIENCE WORKSTATION

Integrated hardware and software solution for Data Science
POWERED BY NVIDIA QUADRO

<table>
<thead>
<tr>
<th>RTX 6000/8000</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPU Architecture</td>
</tr>
<tr>
<td>CUDA Cores</td>
</tr>
<tr>
<td>RT Cores</td>
</tr>
<tr>
<td>Tensor Cores</td>
</tr>
<tr>
<td>Memory BW</td>
</tr>
<tr>
<td>NVLink</td>
</tr>
<tr>
<td>Display Support</td>
</tr>
</tbody>
</table>

RTX 8000 48GB / 96GB w/NVLink

RTX 6000 24GB / 48GB w/NVLink

GV100 32GB / 64GB
Double Precision (FP64)
PERFORMANCE

CPU Gold 6140@2.30GHz 3.7GHz Turbo (Skylake)

End-to-end time = ETL + Conversion + Training + Validation

Mortgage Data
Yr 2015-16, 2 parts

- End-to-end: 528.34 seconds
  - 2x RTX8000: 78.36 seconds
  - 1x RTX8000: 53.56 seconds
  - CPU: 22.1 seconds

- XGBoost: 186 seconds
  - 2x RTX8000: 40.9 seconds
  - 1x RTX8000: 22.1 seconds
  - CPU: 9.13 seconds

- ETL: 280 seconds
  - 2x RTX8000: 53.56 seconds
  - 1x RTX8000: 40.9 seconds
  - CPU: 13.6 seconds
**NVIDIA DATA SCIENCE SOFTWARE STACK**

**Enterprise Desktop**
- DELIVERY
  - Python Pip
  - NGC Containers
  - Anaconda Conda
- DEVELOPMENT
  - Python Notebooks
  - Visualization
- WORKFLOWS
  - KUBEFLOW Pipelines
  - Kubernetes
- CORE FRAMEWORKS AND LIBRARIES
  - Chainer
  - TensorFlow
  - PyTorch
  - RAPIDS
  - Dask
  - Caffe2
- CUDA-X AI
  - DATA PROCESSING
    - cuDF
    - DALI
  - MACHINE LEARNING
    - cuML
    - cuGraph
  - DEEP LEARNING
    - cuDNN
    - cuBLAS
    - NCCL
    - TensorRT

**Enterprise Server**
- WORKFLOWS (Kubeflow, Airflow,...)
- NGC CONTAINERS
- DATA PROCESSING
  - Dask-cuDF
  - Dask-cuPP
  - Spark
  - Datalogue
- DATA LAKE
  - Omnisci
  - BlazingSQL
  - SQreamDB
  - Kinetica
  - BrytyDB
- MACHINE LEARNING
  - cuML
  - cuGraph
- DEEP LEARNING
  - cuDNN
  - cuBLAS
  - NCCL
  - TensorRT
- CLUSTER MANAGEMENT/DEPLOYMENT (CONTAINERS)
- OPERATING SYSTEMS
  - Red Hat Linux
  - Ubuntu

**CUDA-X AI**
- DATA PROCESSING
- MACHINE LEARNING
- DEEP LEARNING
- CLUSTER MANAGEMENT/DEPLOYMENT (CONTAINERS)
- OPERATING SYSTEMS
NGC: GPU-OPTIMIZED SOFTWARE HUB

Simplifying DL, ML and HPC Workflows

50+ Containers
DL, ML, HPC

Pre-trained Models
NLP, Classification, Object Detection & more

Model Training Scripts
NLP, Image Classification, Object Detection & more

Industry Workflows
Medical Imaging, Intelligent Video Analytics

Simplify Deployments
Innovate Faster
Deploy Anywhere
NVIDIA NGC SUPPORT SERVICES
Minimize Downtime And Maximize System Utilization

Support Coverage
- NGC DL & ML containers
- NVIDIA drivers
- NV-docker
- CUDA

Support by NVIDIA's subject matter experts

Availability
- Exclusively for NGC-Ready workstations
- Availability starting in Q2
- Service agreement between NVIDIA & customer
- Purchase from OEM

24x7 portal, phone and email access to create support cases

Live support during local, regional business hours for technical assistance
HIGH-PERFORMANCE DATA SCIENCE

Maximized Productivity
Highly optimized cross-compatible stack of data science libraries
Faster model design, development and iteration
Greater flexibility using Python and conda package management

Ease of Use
Turnkey system for GPU accelerated data science
End-to-End software stack acceleration from data preparation to visualization
Orchestration compatible software stack to help scale on clusters

Enterprise Support
Built for 24x7 reliability and robustness
Quick and easy deployment using “NGC-Ready” containers and conda
Tested across GPUs and systems for compatibility and performance
“Our initial look at the NVIDIA-Powered Lenovo AI workstation showed significant performance gains. Data scientists will appreciate being able to move more quickly through the analytics life cycle, which will allow them to address and support more analytics needs to transform business processes.”

-- Gavin Day, Senior Vice President for Technology at SAS

“The NVIDIA-powered data science workstation enables our data scientists to run end-to-end data processing pipelines on large data sets faster than ever. Leveraging RAPIDS to push more of the data processing pipeline to the GPU reduces model development time which leads to faster deployment and business insights.”

-Mike Koelemay, Lockheed Martin Fellow

“We have a diverse, multi-disciplinary environment and are looking to couple data science and analytics to a wider range of our technical practices throughout our business. The NVIDIA-powered Data Science Workstation promises to ease the transition and democratize the application of data science. We find it extremely well suited to experimentation, exploration, solution discovery, and early prototyping work. Its combination of well-designed software and highly performant hardware provides a 20x and higher speed-ups in our analytics work and our team found its ease of use liberating.”

-Steve Walker, Associate Director, Arup, Advanced Digital Engineering