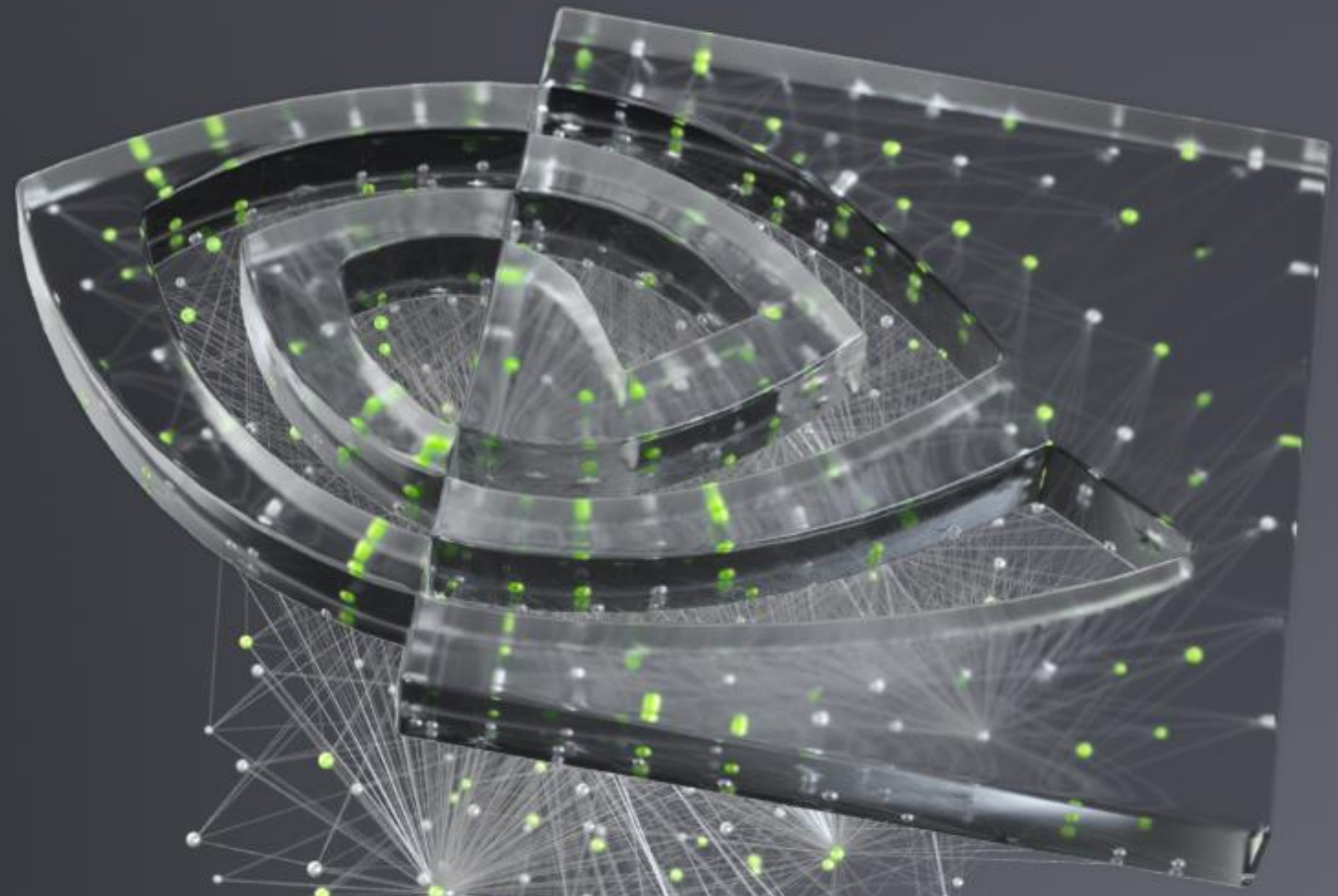


GPU ACCELERATED DATA SCIENCE

Scott Murnan – OEM Alliance Manager

HOW MUCH DATA ARE WE PRODUCING?

WHERE IS THE DATA?

DATABASES, SPREADSHEETS, PAPER, ARCHIVES, CD'S….ETC ETC ETC

- WHAT DO WE DO WITH IT?
- IS IT USEFUL?
- CAN WE MONETIZE IT?

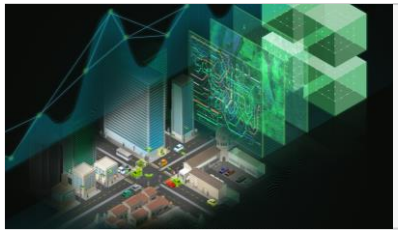
WHAT IS DATA SCIENCE?

It is the Extraction of insights and knowledge from any type of data

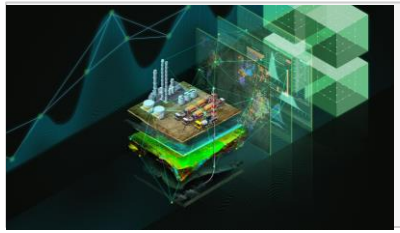
….for business….for research….for technological advances….
for global warming….for humanity

3M DATA SCIENTISTS AT WORK WW

IN EVERY MARKET VERTICAL



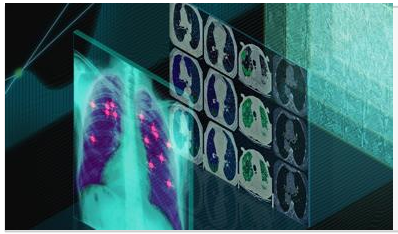
CONSUMER INTERNET



OIL & GAS

FINANCIAL SERVICES

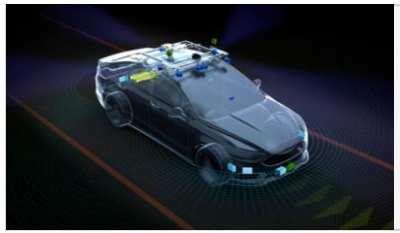
MANUFACTURING



HEALTHCARE

TELECOM

RETAIL



AUTOMOTIVE

DATA SCIENCE IN FINANCE

Anti money laundering

ATM Cash on Hand

Automated Underwriting Decisions

Branch Location & Staff Planning

Credit Scoring

Customer Call Routing

Customer Next Best Offer

Customer Transaction Fraud Detection

Mortgage Pre-Payment Analytics

DATA SCIENCE IN MANUFACTURING

Remaining Useful Life Estimation

Predicting Failure

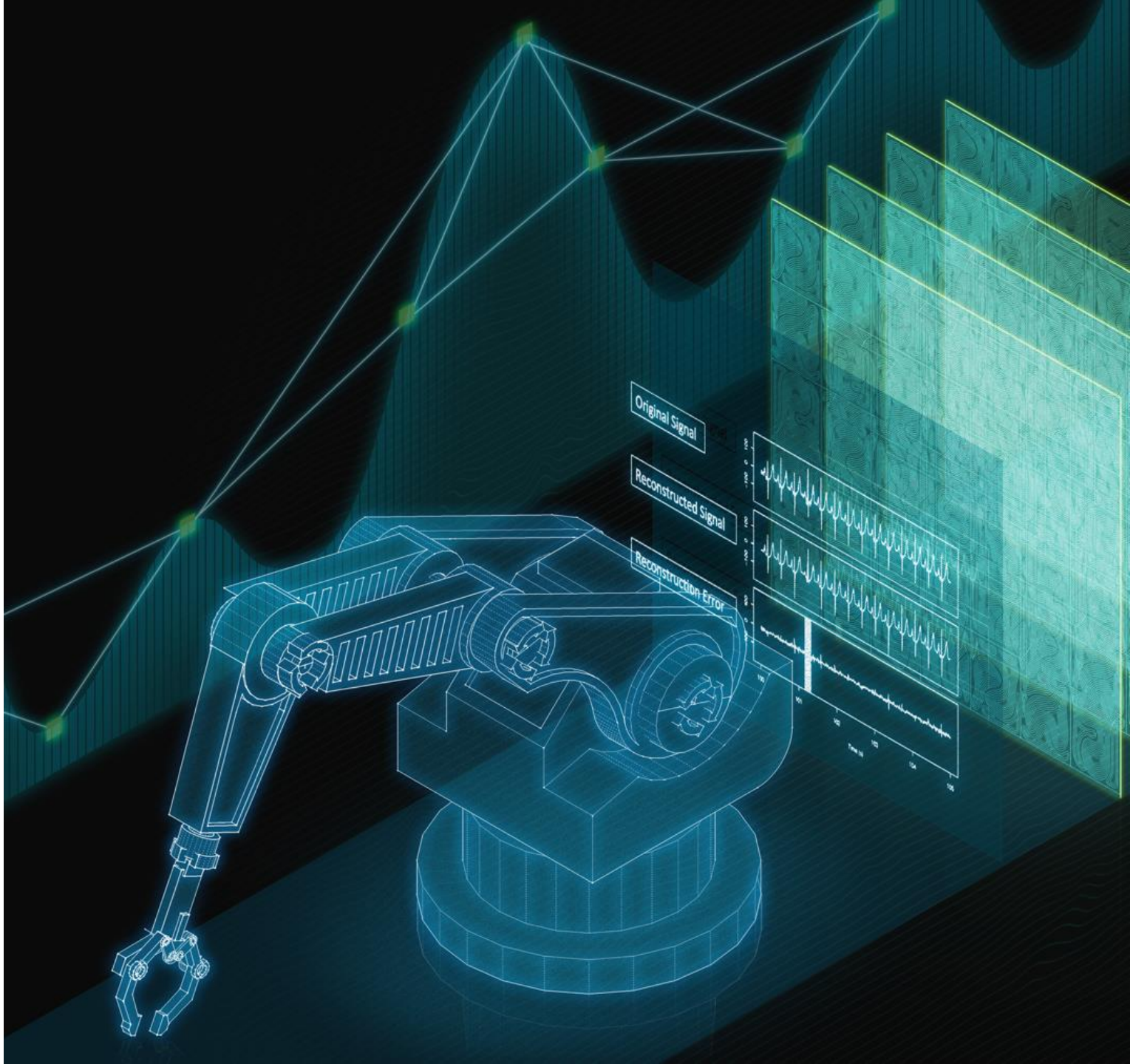
Demand Forecasting

Anomaly Detection

Time Series Pattern Search

Improve Defect Classification

Improve Manufacturing Yield Forecast



DATA SCIENCE IN HEALTHCARE

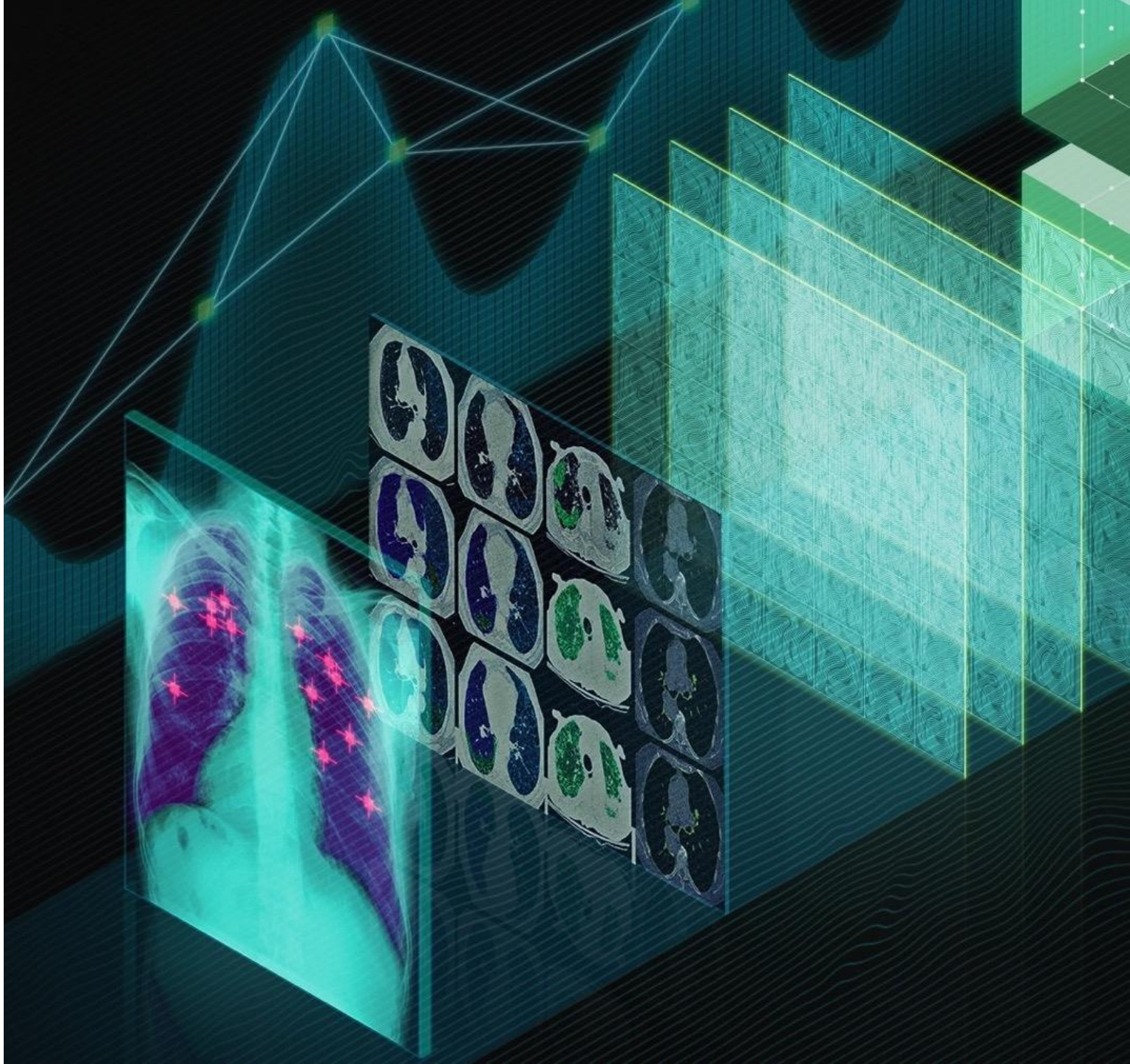
Improve Clinical Care – At Home
Early Detection Of Diseases
Analyze/Manage Population Health
Match Patients To Clinical Trials

Improve Clinical Care – In The Hospital
Predict Risk Of Sepsis & Deterioration
Predict Fall Risk
Predict Risk Of Hospital Acquired Infections
Predict Readmission Risk

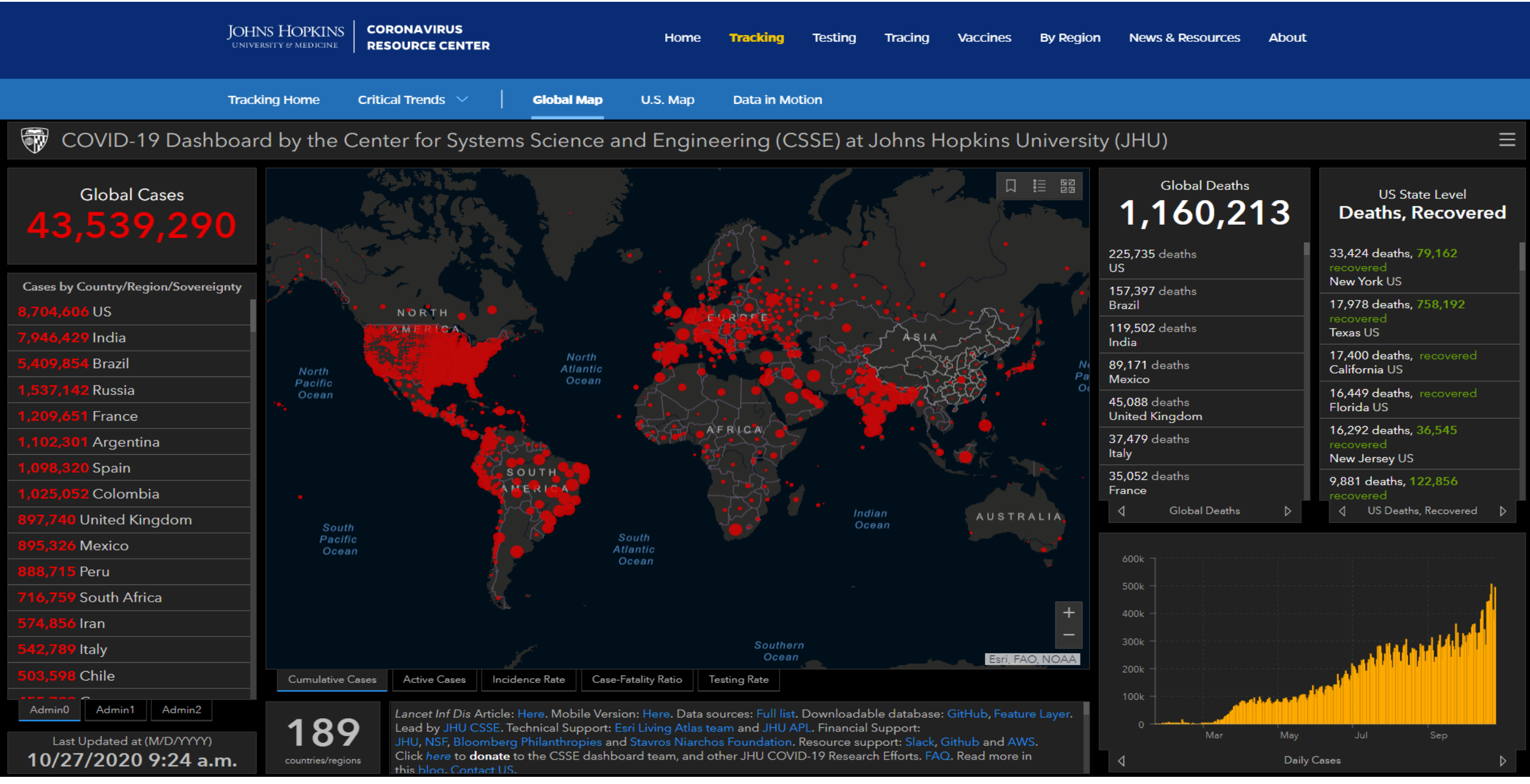
Precision Medicine
Tailor Treatments To Patients' Phenotype & Genotype
Identity Patients Like Mine

Drive Operational Efficiency
Predict Patient No Shows
Predict ER Wait Times & Volume
Predict Length Of Stay
Optimize Surgery Schedules
Avoid Denial Of Insurance Claims
Predictive Maintenance Of Equipment

Drug Discovery
Identify Promising Drug Molecules



COVID19 – John Hopkins University Worldwide Analysis



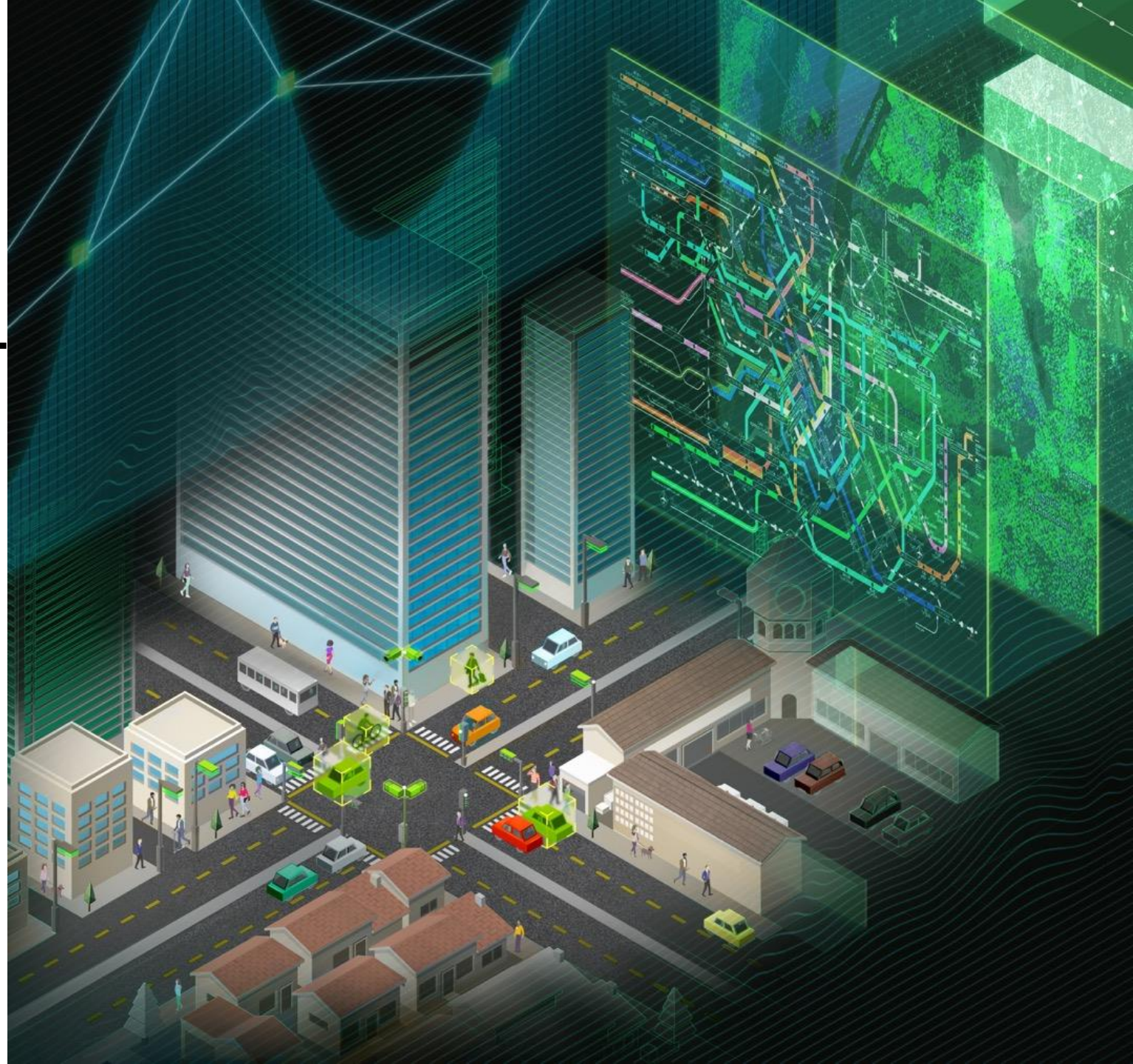
DATA SCIENCE IN CONSUMER INTERNET

Ad Personalization

Click Through Rate Optimization

Customer Life Time Value (LTV) Prediction

Churn Prevention



DATA SCIENCE IN OIL AND GAS

Sensor data tag mapping

Anomaly Detection

Robust Fault Prediction

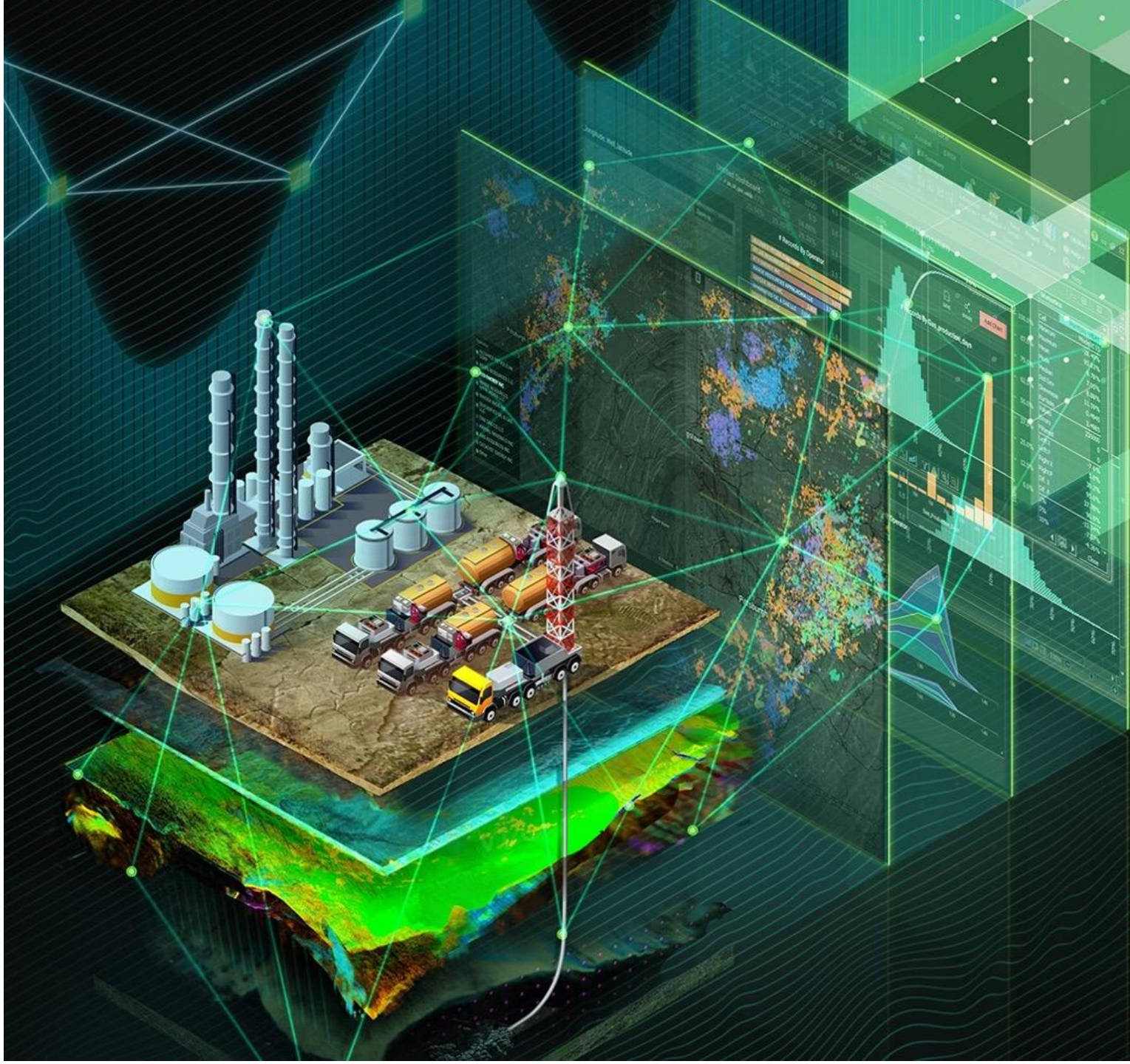
Time Series Pattern Matching

Automated Drilling

Production Optimization

Reservoir Characterization

Seismic Interpretation



DATA SCIENCE IN TELECOM

Customer Data Monetization

Analyze Intra-day Billing

Upsell Recommendations

Prevent Customer Churn

Network Optimization

Detect Network / Security Anomalies

Forecasting Network Performance

Network Resource Optimization (SON)

DATA SCIENCE IN AUTOMOTIVE

Customer Experience

Personalization

Intelligent Customer Interactions

Customer Support & Dispatch Mapping

Profitable Operations

Demand Modeling

Fraud Detection, Cybersecurity

Anomaly Detection

Capacity Planning

DATA SCIENCE IN RETAIL

Supply Chain Replenishment

Inventory Management

Price Management / Markdown Optimization

Prioritize Promotion And Ad Targeting

Personalized Recommendations

Truck Routing

Online Delivery

Marketing Optimization

WHAT DOES A DATA SCIENTIST LOOK LIKE?

COMBINING - MATHS & STATISTICS, CODING, RESEARCH, ALGORITHMS……

SOLVING PROBLEMS WITH DATA, MANIPULATING DATA, EXPLORING DATA

Degrees & PhDs In Computer Science/Mathematics/Programming/Economics etc etc

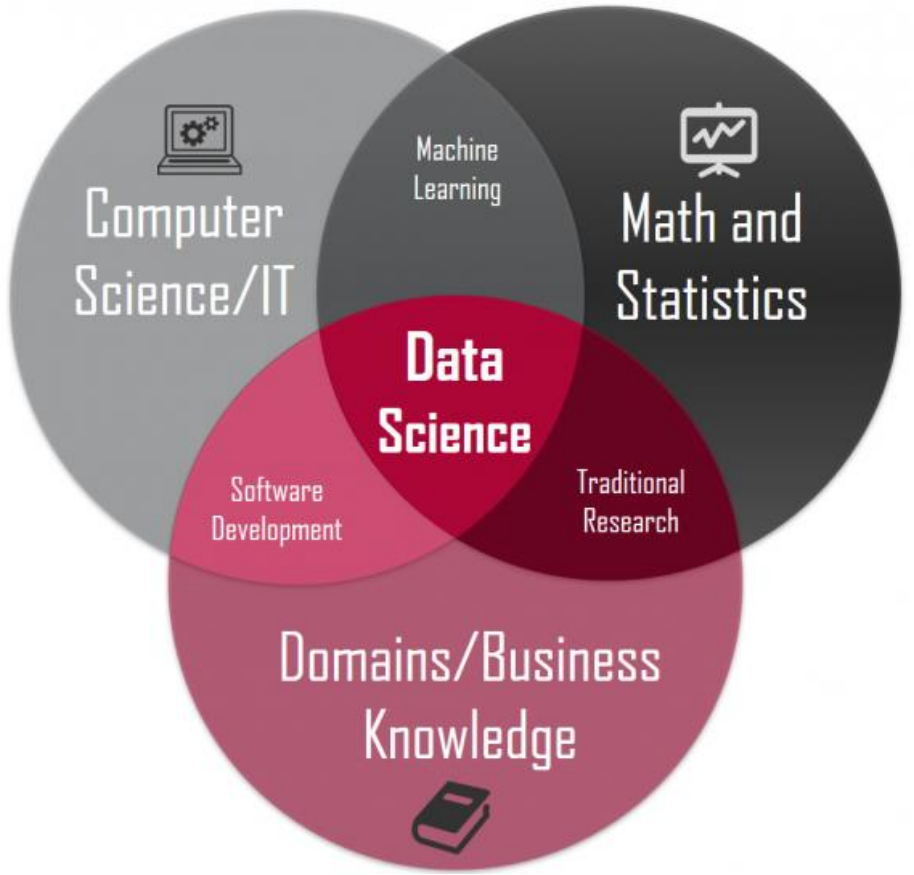
Access high performance computing platforms

Iterative and experimental ML workflows

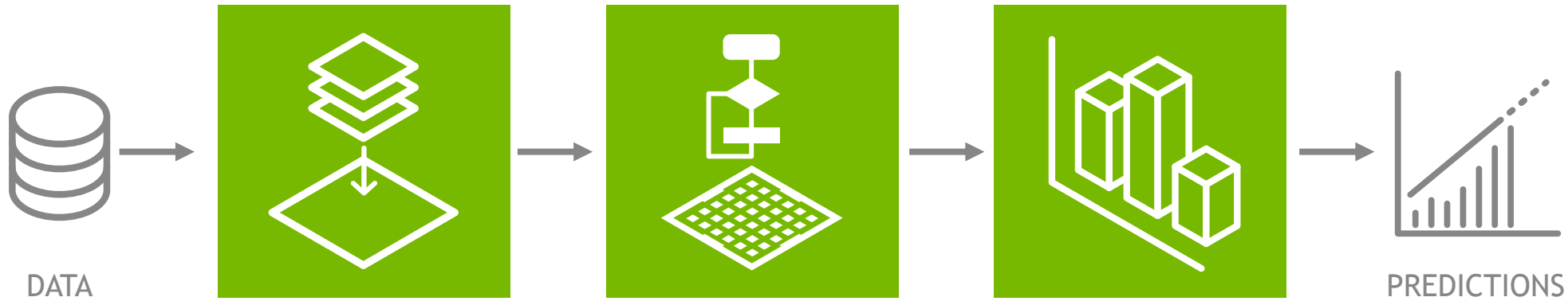
Data Scientists are not generally part of IT, they part of the business, near to a lot of data

They create their own software stack and applications to extract insights

They can spend 80% of their time manipulating data



GPU ACCELERATION OF THE DATA SCIENTIST WORKFLOW
PREP>>>TRAIN>>>VISUALIZE

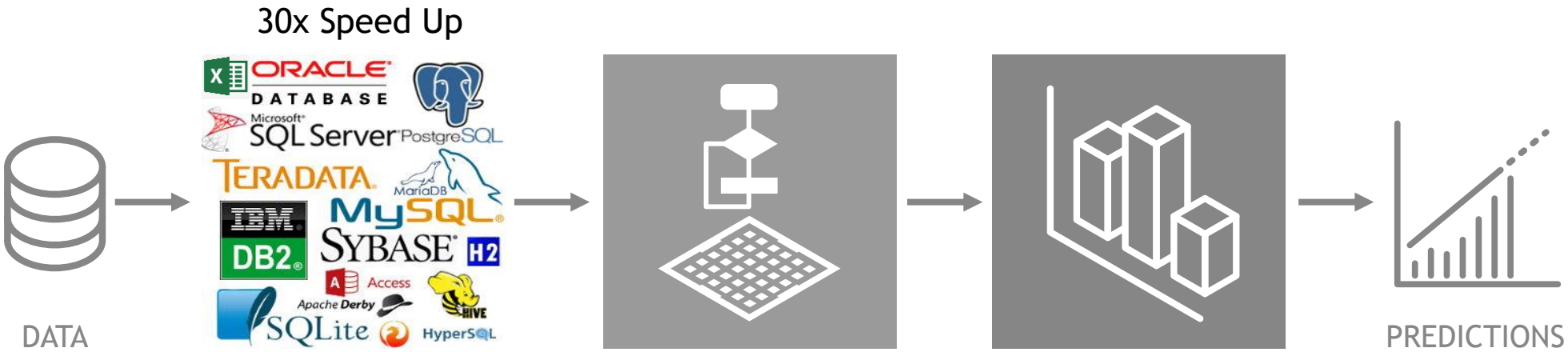


DATA PREPARATION

MODEL TRAINING

DATA VISUALIZATION

DATA PREPARATION
(UP TO 80% OF A DATA SCIENTIST'S TIME)



DATA PREPARATION cuDF

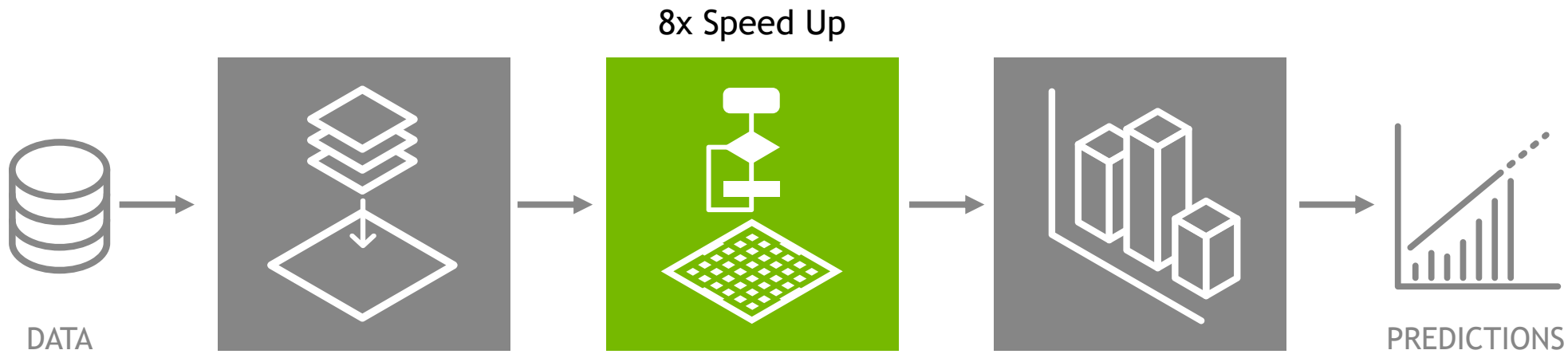
GPUs accelerated compute for in-memory data preparation

Simplified implementation using familiar data science tools

Python drop-in **pandas** replacement built on CUDA C++.

GPU-accelerated Spark (in development)

MODEL TRAINING



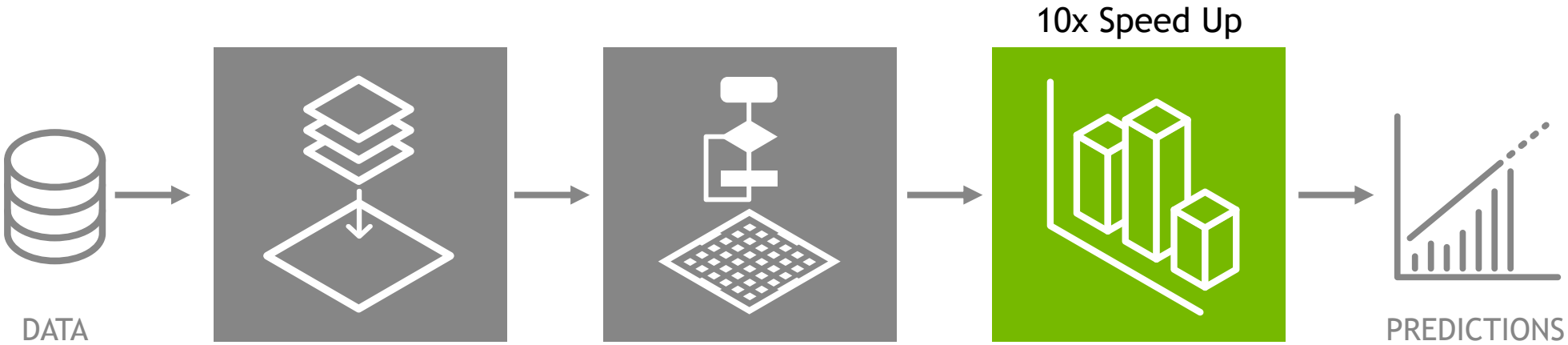
MODEL TRAINING cuML

GPU-acceleration of today's most popular ML algorithms such as **XGBoost**

Also available are PCA, K-means, k-NN, DBScan, tSVD, and many more

Easy-to-adopt, scikit-learn like interface

DATA VISUALIZATION



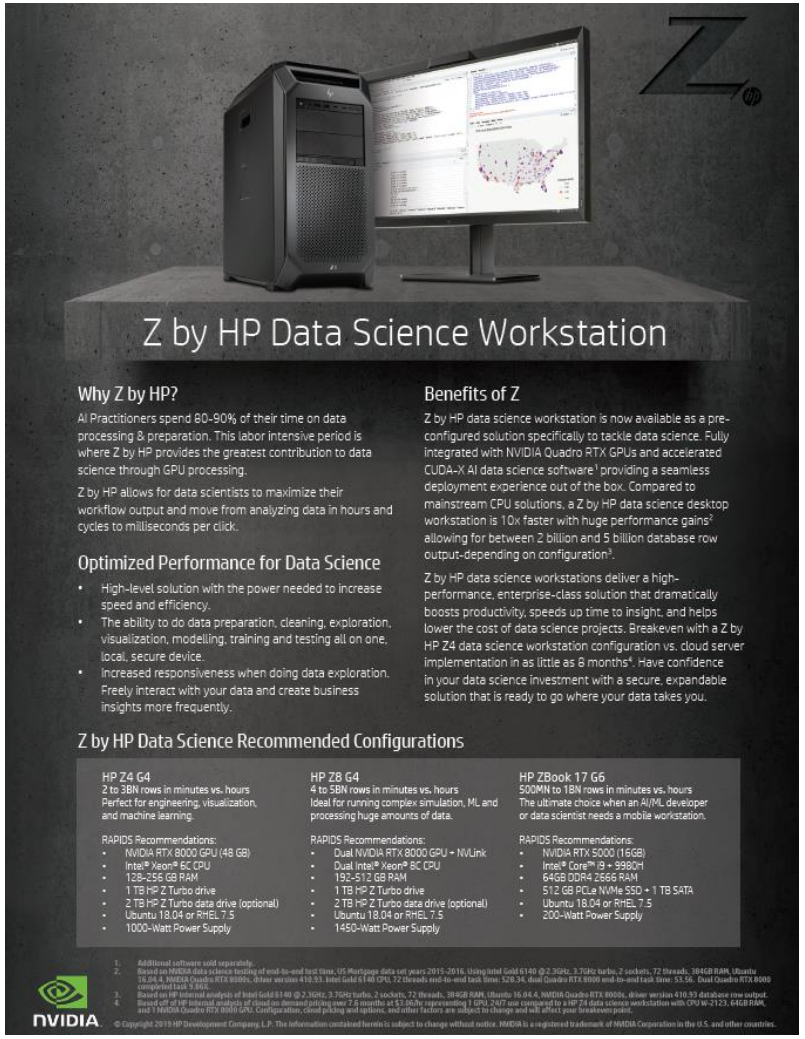
VISUALIZATION cuGRAPH

Effortless exploration of datasets, billions of records in milliseconds

Dynamic interaction with data = faster ML model development

Data visualization ecosystem (Graphistry & OmniSci), integrated with RAPIDS

Z BY HP DATA SCIENCE WORKSTATION
TIME MACHINES FOR DATA SCIENCE



Z by HP Data Science Workstation

Why Z by HP?

AI Practitioners spend 80-90% of their time on data processing & preparation. This labor intensive period is where Z by HP provides the greatest contribution to data science through GPU processing.

Z by HP allows for data scientists to maximize their workflow output and move from analyzing data in hours and cycles to milliseconds per click.

Optimized Performance for Data Science

- High-level solution with the power needed to increase speed and efficiency.
- The ability to do data preparation, cleaning, exploration, visualization, modelling, training and testing all on one, local, secure device.
- Increased responsiveness when doing data exploration. Freely interact with your data and create business insights more frequently.

Benefits of Z

Z by HP data science workstation is now available as a pre-configured solution specifically to tackle data science. Fully integrated with NVIDIA Quadro RTX GPUs and accelerated CUDA-X AI data science software[1] providing a seamless deployment experience out of the box. Compared to mainstream CPU solutions, a Z by HP data science desktop workstation is 10x faster with huge performance gains[2] allowing for between 2 billion and 5 billion database row output-depending on configuration[3].

Z by HP data science workstations deliver a high-performance, enterprise-class solution that dramatically boosts productivity, speeds up time to insight, and helps lower the cost of data science projects. Breakeven with a Z by HP Z4 data science workstation configuration vs. cloud server implementation in as little as 8 months[4]. Have confidence in your data science investment with a secure, expandable solution that is ready to go where your data takes you.

Z by HP Data Science Recommended Configurations

HP Z4 G4	HP Z8 G4	HP ZBook 17 G5
2 to 3BN rows in minutes vs. hours Perfect for engineering, visualization, and machine learning.	4 to 5BN rows in minutes vs. hours Ideal for running complex simulation, ML and processing huge amounts of data.	500MN to 1BN rows in minutes vs. hours The ultimate choice when an AI/ML developer or data scientist needs a mobile workstation.
RAPIDS Recommendations: • NVIDIA RTX 8000 GPU (48 GB) • Intel® Xeon® 6C CPU • 128–256 GB RAM • 1 TB HP Z Turbo drive • 2 TB HP Z Turbo data drive (optional) • Ubuntu 18.04 or RHEL 7.5 • 1000-Watt Power Supply	RAPIDS Recommendations: • Dual NVIDIA RTX 8000 GPU + NVLink • Dual Intel® Xeon® 8C CPU • 192–512 GB RAM • 1 TB HP Z Turbo drive • 2 TB HP Z Turbo data drive (optional) • Ubuntu 18.04 or RHEL 7.5 • 1450-Watt Power Supply	RAPIDS Recommendations: • NVIDIA RTX 5000 (16GB) • Intel® Core™ i9 + 9980H • 64GB DDR4 2666 RAM • 512 GB PCIe NVMe SSD + 1 TB SATA • Ubuntu 18.04 or RHEL 7.5 • 200-Watt Power Supply

1. Additional software sold separately.
2. Based on NVIDIA data science testing of end-to-end test time, US Mortgage data set years 2015-2016. Using Intel Gold 6140 @ 2.3GHz, 3.7GHz turbo, 2 sockets, 72 threads, 384GB RAM, Ubuntu 16.04.4, NVIDIA Quadro RTX 8000s, driver version 410.93. Intel Gold 6140 CPU, 72 threads end-to-end task time: 528.34, dual Quadro RTX 8000 end-to-end task time: 53.56. Dual Quadro RTX 8000 completed task 9.86X.
3. Based on HP internal analysis of Intel Gold 6140 @ 2.3GHz, 3.7GHz turbo, 2 sockets, 72 threads, 384GB RAM, Ubuntu 16.04.4, NVIDIA Quadro RTX 8000s, driver version 410.93 database row output.
4. Based off of HP internal analysis of cloud on demand pricing over 7.6 months at $3.06/hr representing 1 GPU, 24/7 use compared to a HP Z4 data science workstation with CPU W-2123, 64GB RAM, and 1 NVIDIA Quadro RTX 8000 GPU. Configuration, cloud pricing and options, and other factors are subject to change and will affect your breakeven point.

PDF Brief

Turning days in to hours, hours in to minutes

https://www8.hp.com/us/en/workstations/industries/data-science.html

American Airlines Blog

500,000 Booking Records, 20 Features Per Record, 100 Derived Features

- Models 90% accurate
- 10x speed up of computations
- Faster predictions
- Increased cargo space utilization
- Reduced fuel burn

HOME | DEEP LEARNING | NETWORKING | DRIVING | GAMING | PRO GRAPHICS | AUTONOMOUS MACHINES | HEALTHCARE | AI PODCAST

American Airlines Delivers the Goods, with Data Science Workstations

World's largest airline uses NVIDIA Quadro to better model cargo shipments, improve weight distribution and save fuel.

May 7, 2020 by NICOLE CASTRO

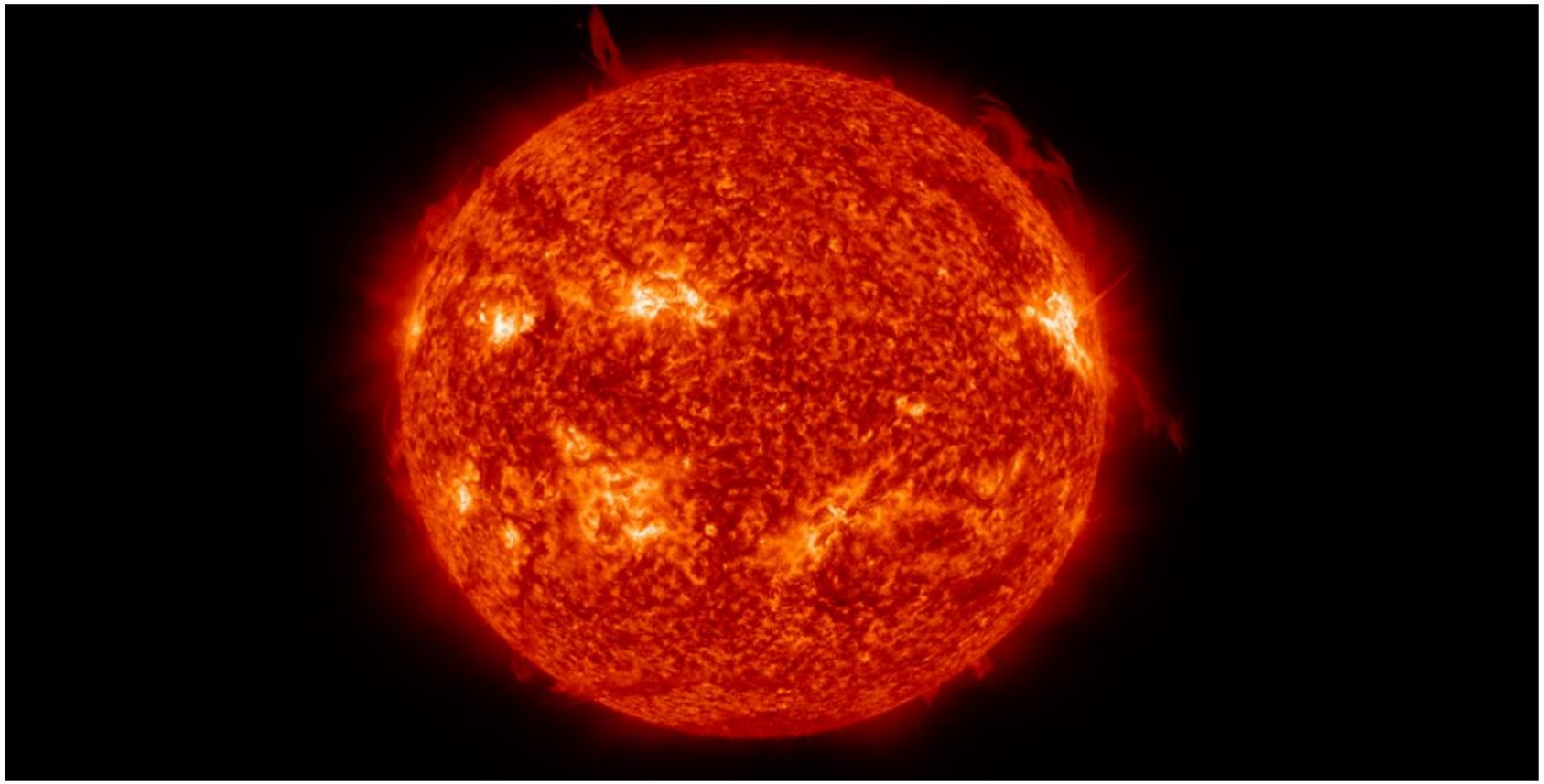
NASA Blog

HOME DEEP LEARNING NETWORKING DRIVING GAMING PRO GRAPHICS AUTONOMOUS MACHINES HEALTHCARE AI PODCAST

NASA's Day in the Sun: Space Agency Speeds Analysis of Solar Images by 150x Using Data Science Workstations

Scientists accelerate data analytics and computations that would've taken years on CPUs to less than a week with RTX-powered Z by HP data science workstations.

June 18, 2020 by NICOLE CASTRO



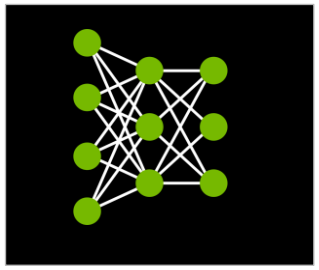
- 18 Petabytes+ of images
- 10-150x speed up of data and image analysis
- Compressing years to weeks of analysis

NVIDIA DLI HANDS-ON TRAINING

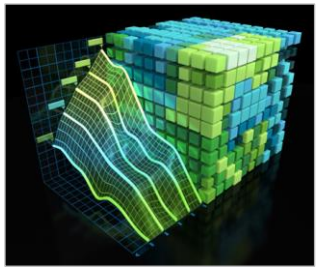
CATALOG

https://www.nvidia.com/en-us/deep-learning-ai/education/

NVIDIA DEEP LEARNING INSTITUTE



Deep Learning Fundamentals



Accelerated Computing Fundamentals

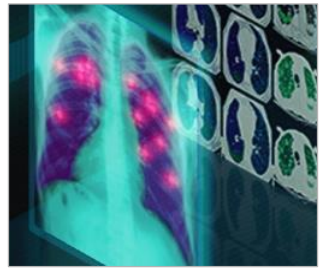
Accelerated Data Science Fundamentals

Intro to AI in the Data Center

AI for Anomaly Detection

AI for Autonomous Vehicles

AI for Digital Content Creation



AI for Healthcare

AI for Industrial Inspection

AI for Intelligent Video Analytics

AI for Predictive Maintenance

AI for Robotics

THANK YOU


QUADRO
NVIDIA
NVIDIA
SELECTED BANK
OPACITY
100
SCALE
35
CITI BINS
30
RISK BINS
10
CREDIT BINS
26
PORTFOLIO RISK SCORE
TOTAL UNPAID BALANCE
BAR HEIGHT
AREA DETAILS
RESET ALL
GPU AGGREGATED 100.00% OF 146,594,614 MORTGAGES