RAPIDS PITCH DECK

6 QUESTIONS FACING EVERY AI ENTERPRISE

Top Challenges for AI, Big Data, and Enterprise Transformation

DATA DELUGE

Is your data doubling each year?

COMPLEX WORKLOADS

Is Spark workloads creating relentless infrastructure sprawl?

TEDIOUS DATA PREP

Do you have oceans of data, that take lifetimes to wrangle?

PROLONGED TRAINING TIME

Is ML training prohibitively long, delaying time-to-predictions?

DELAYED INTELLIGENCE

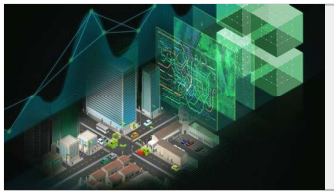
Are you an intelligent enterprise needing real time predictive analytics?

SHRINKING BUDGET

Is your CAPEX budget shrinking amidst escalating infrastructure demand?

GPU-ACCELERATED DATA SCIENCE

Use Cases in Every Industry

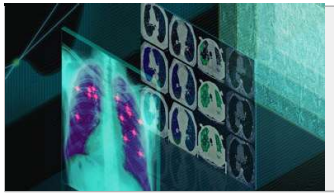


CONSUMER INTERNET

Ad Personalization
Click Through Rate Optimization
Churn Reduction

FINANCIAL SERVICES

Claim fraud
Customer service chatbots/routing
Risk evaluation

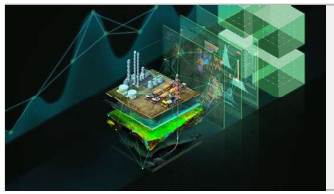


HEALTHCARE

Improve Clinical Care
Drive Operational Efficiency
Speed Up Drug Discovery

RETAIL

Supply Chain & Inventory Management
Price Management / Markdown Optimization
Promotion Prioritization And Ad Targeting



OIL & GAS

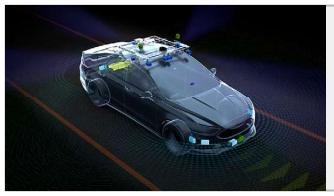
Sensor Data Tag Mapping
Anomaly Detection
Robust Fault Prediction

MANUFACTURING

Remaining Useful Life Estimation
Failure Prediction
Demand Forecasting

TELCO

Detect Network/Security Anomalies
Forecasting Network Performance
Network Resource Optimization (SON)



AUTOMOTIVE

Personalization & Intelligent Customer Interactions
Connected Vehicle Predictive Maintenance
Forecasting, Demand, & Capacity Planning

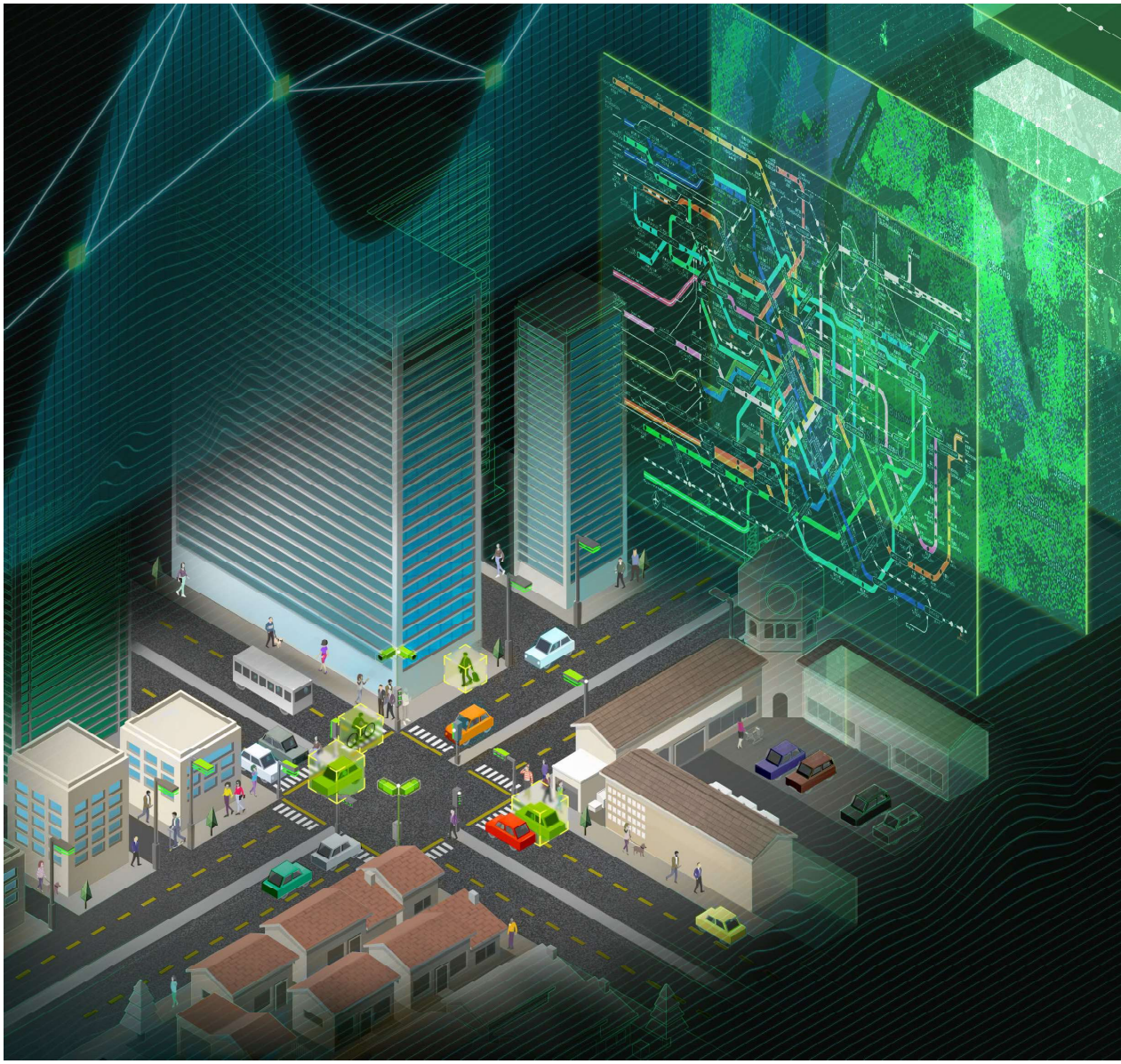
DATA SCIENCE IN CONSUMER INTERNET

Ad Personalization

Click Through Rate Optimization

Customer Life Time Value (LTV) Prediction

Churn Prevention



MACHINE LEARNING CHALLENGES

30+
Hours to Build GBDT
(Gradient Boosted Tree Regression)

SLOW PROCESSES

Days
Data Transformation

Weeks
Feature Engineering

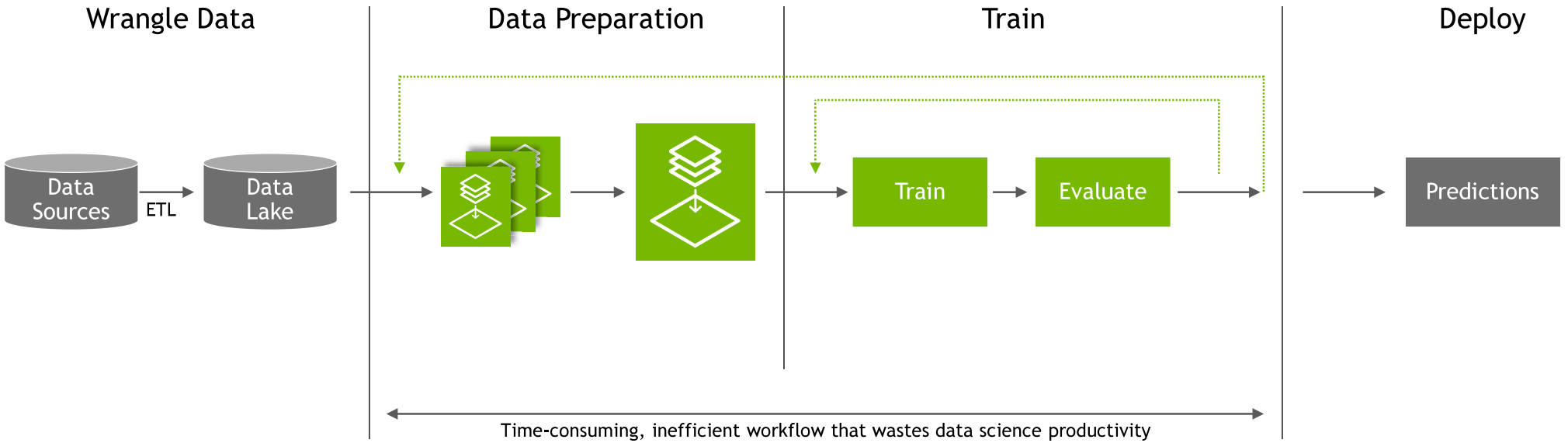
Months
Scoring Pipelines

MODEL COMPLEXITY

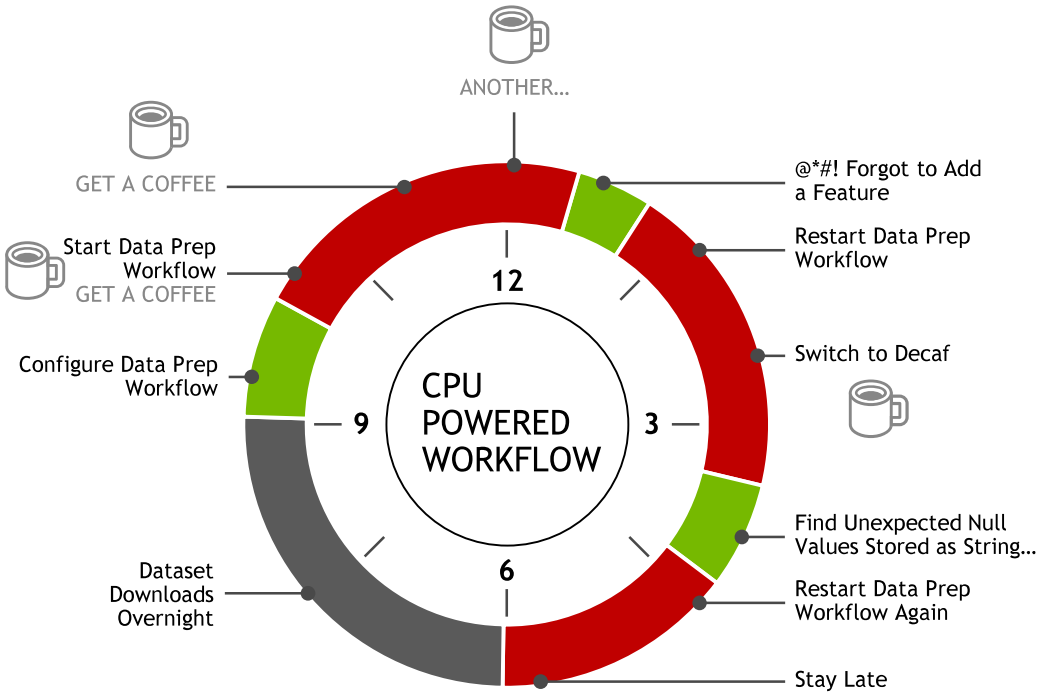
$3M+
More Servers and Infrastructure Yielding Diminishing Returns

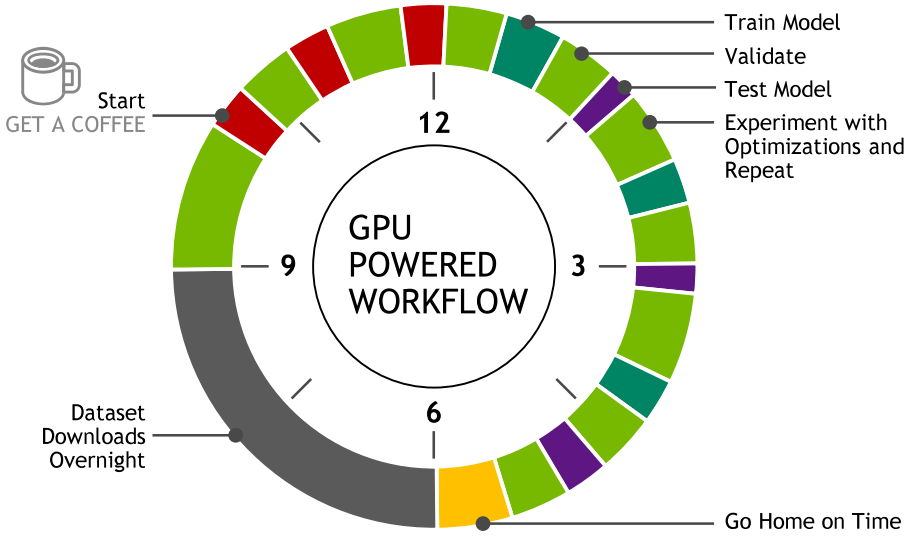
ESCALATING TCO

ML WORKFLOW STIFLES INNOVATION



DAY IN THE LIFE OF A DATA SCIENTIST

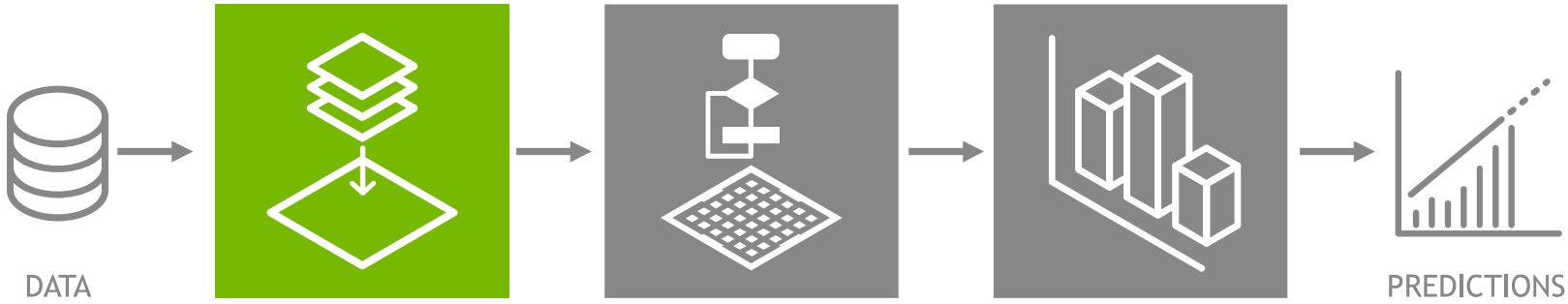




Dataset Collection | Analysis | Data Prep | Train | Inference

DATA SCIENCE WORKFLOW WITH RAPIDS

Open Source, End-to-end GPU-accelerated Workflow Built On CUDA



DATA PREPARATION

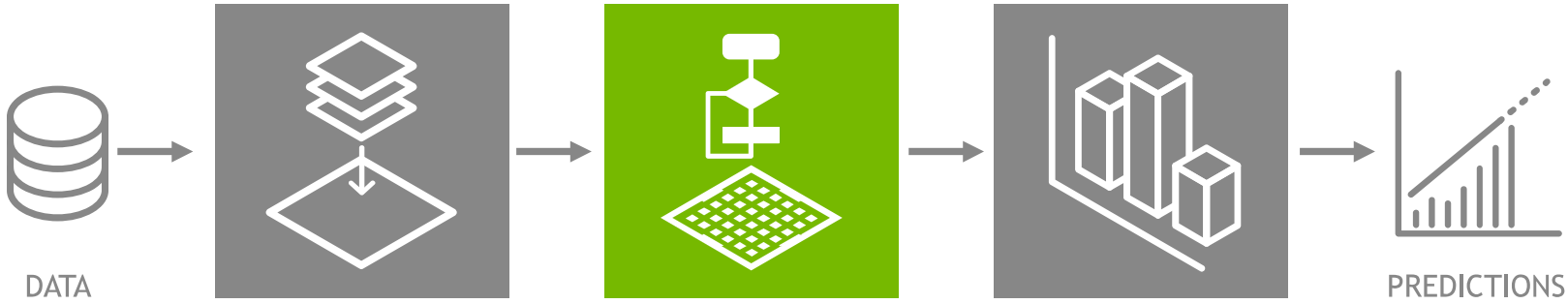
GPUs accelerated compute for in-memory data preparation

Simplified implementation using familiar data science tools

Python drop-in Pandas replacement built on CUDA C++. GPU-accelerated Spark (in development)

DATA SCIENCE WORKFLOW WITH RAPIDS

Open Source, End-to-end GPU-accelerated Workflow Built On CUDA



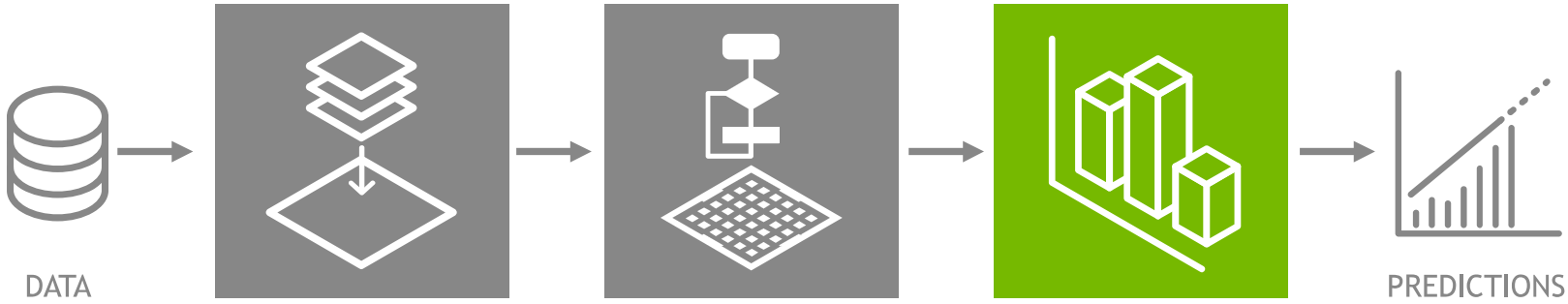
MODEL TRAINING

GPU-acceleration of today's most popular ML algorithms

XGBoost, PCA, Kalman, K-means, k-NN, DBScan, tSVD …

DATA SCIENCE WORKFLOW WITH RAPIDS

Open Source, End-to-end GPU-accelerated Workflow Built On CUDA



VISUALIZATION

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

TRADITIONAL DATA SCIENCE CLUSTER

Workload Profile:

Fannie Mae Mortgage Data:

- 192GB data set
- 16 years, 68 quarters
- 34.7 Million single family mortgage loans
- 1.85 Billion performance records
- XGBoost training set: 50 features

300 Servers | $3M | 180 kW

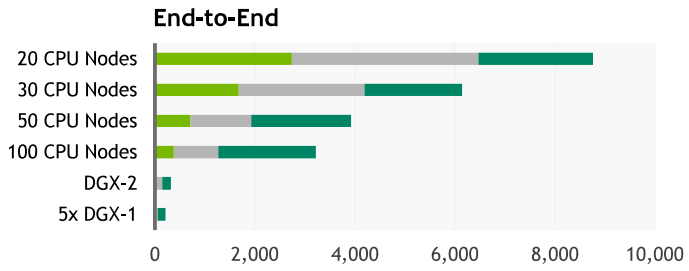
GPU-ACCELERATED MACHINE LEARNING CLUSTER

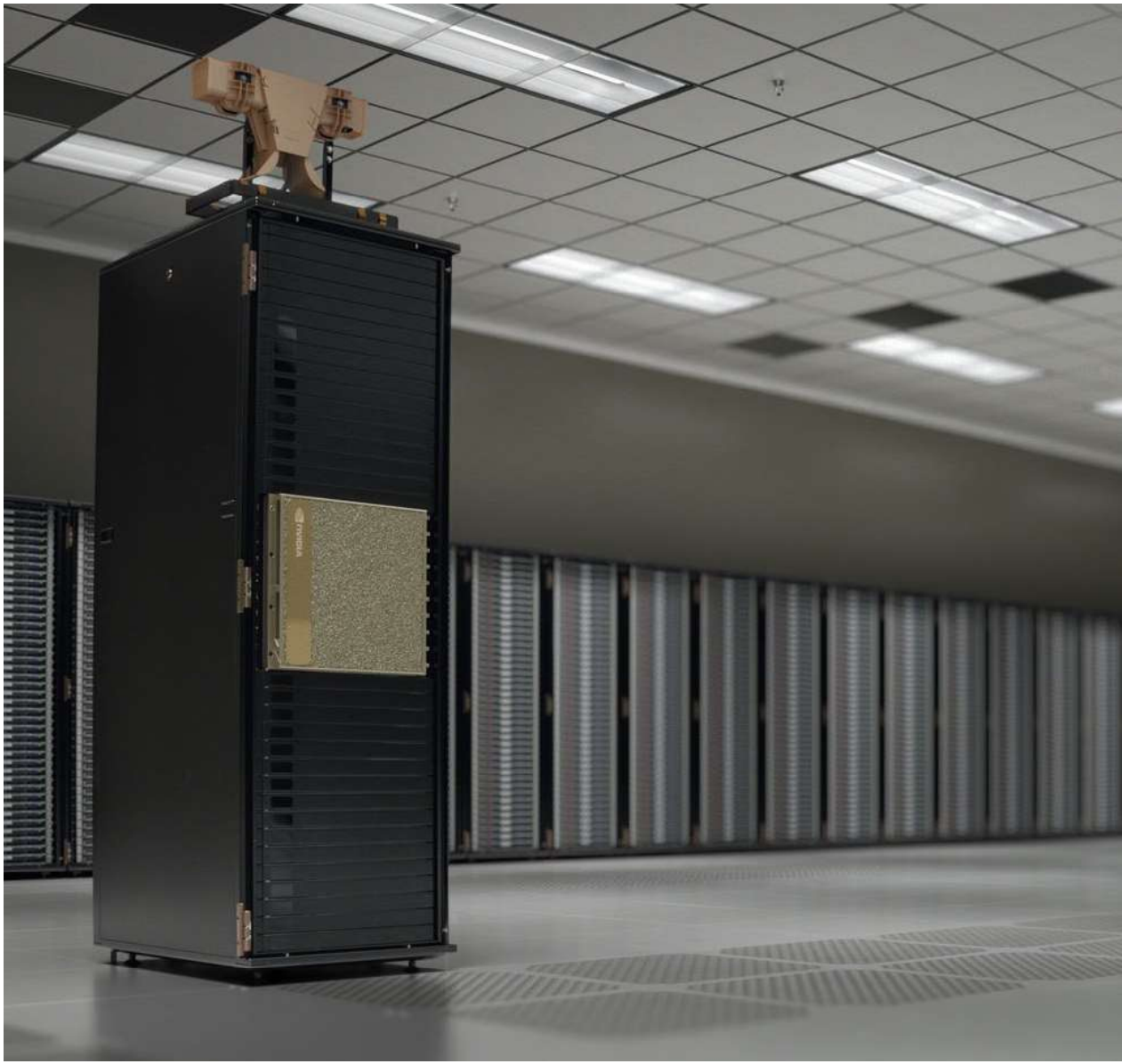
DGX-2 and RAPIDS for Predictive Analytics

1 DGX-2 | 10 kW

1/8 the Cost | 1/15 the Space

1/18 the Power





RAPIDS: DELIVERING DATA SCIENCE VALUE

Maximized Productivity

Top Model Accuracy

Lowest TCO

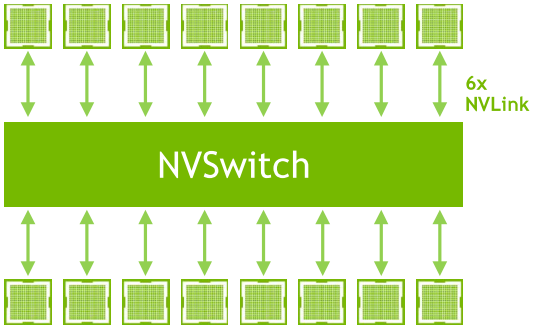
Oak Ridge National Labs	Global Retail Giant	Streaming Media Company
215x Speedup Using RAPIDS with XGBoost	**$1B** Saving with 4% Error Rate Reduction	**$1.5M** Infrastructure Cost Saving

PILLARS OF RAPIDS PERFORMANCE

CUDA Architecture

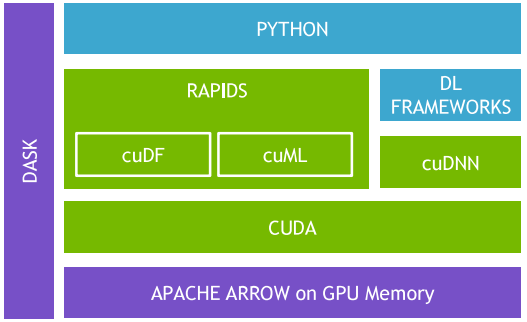
Massively Parallel Processing

NVLink/NVSwitch



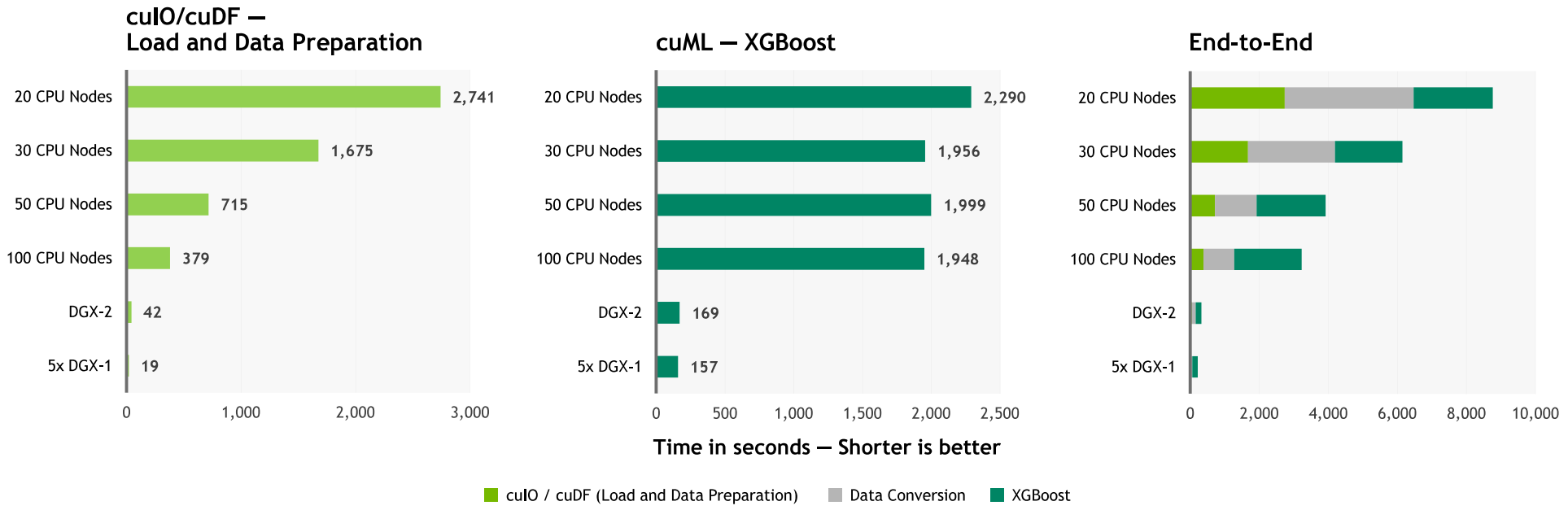
High Speed Connecting between GPUs for Distributed Algorithms

Integrated Software



Fully Integrated Software and Hardware for Instant Productivity

FASTER SPEEDS, REAL WORLD BENEFITS



Benchmark

200GB CSV dataset; Data preparation includes joins, variable transformations.

CPU Cluster Configuration

CPU nodes (61 GiB of memory, 8 vCPUs, 64-bit platform), Apache Spark

DGX Cluster Configuration

5x DGX-1 on InfiniBand network

WIDESPREAD SUPPORT FOR RAPIDS

Open Source Community

ANACONDA. · Apache Arrow · GUNROCK · hadoop · PyData · QUANTSIGHT URSA LABS · Apache Spark · learn

Enterprise Data Science Platforms

IBM · ORACLE · sas · SAP

Startups

BLAZINGDB · brytlyt · DATALOGUE · FASTDATA · graphistry · H_2O.ai · kinetica · MAPR · omnisci · SQREAM · ZILLIZ

Deep Learning Integration

Chainer · PYTORCH

RAPIDS

GPU Servers

cisco · DELL EMC · Hewlett Packard Enterprise · IBM · Lenovo

Storage Partners

DELL EMC · IBM · NetApp · PURESTORAGE

** Spark and Hadoop support coming soon*