NVIDIA ACCELERATED DATA SCIENCE SOLUTION WITH RAPIDS

Data science is powering the engine of modern enterprise – every industry from retail to financial services to healthcare is deriving insight from data to improve competitiveness and operational efficiency. Retailers are improving forecasting to reduce the cost of excess inventory. Financial services institutions are detecting fraudulent transactions. Healthcare providers are predicting the risk of disease more quickly. Even modest improvements in the accuracy of predictive machine learning models can translate into billions for the bottom line.

NVIDIA accelerated data science solution with RAPIDS enables enterprises to tap into GPU-accelerated machine learning (ML) with faster model iteration, better prediction accuracy, and lowest data science total cost of ownership (TCO).

THE CHALLENGES OF TRADITIONAL MACHINE LEARNING WORKFLOW

Creating ML models often involves days spent on ingesting and preparing data, weeks lost on engineering features based on the data, and potentially months consumed in scoring ML pipelines to evaluate efficacy and model selection for production inference. This iterative process must be executed repeatedly and often, using parallel development pipelines to yield even basic results. The inefficient workflow creates an ongoing cycle of data scientist downtime as they wait on inadequate, underpowered, CPU-based tools to ingest and prepare data, train models and evaluate results.

MACHINE LEARNING WITH NVIDIA GPUS

NVIDIA has revolutionized modern computing through the application of the Graphics Processing Unit (GPU) and its massively parallelized architecture, enabling a dramatic outpacing in performance when compared with traditional CPU architectures. Developers could previously depend on Moore's Law scaling to reap performance gains every 18 months. Over the last decade, CPUs have seen a consistent deceleration due to the constraints of semiconductor physics and the escalating power cost associated with exposing increasing amounts of instruction-level parallelism in the CPU.

Given the highly parallelized nature of the arithmetic operations that are core to machine learning, the modern GPU with its immense computing core footprint is uniquely suited to accelerate AI. GPUs execute the entire machine learning workflow

BENEFITS

FOR DATA SCIENTISTS:

- Reduced training time for faster workflow, with near real time results for interactive data science exploration
- Hassle-free integration with minimal code changes and no new tools to learn
- Open-source software that's easily customizable, extensible, and interoperable

FOR BUSINESS LEADERS:

- Top model accuracy achieved in dramatically shorter timeframes
- Maximized productivity for data scientists through elimination of wasted time
- Reduced TCO through more efficient compute infrastructure

in high-speed system memory and parallelize the data loading, data manipulation, and ML algorithms training on GPU CUDA cores. The application of NVIDIA GPUs to ML worfklows enable a dramatic acceleration in model development speed and compression of development timelines. This helps organizations achieve massive OpEx and CapEx savings.

NVIDIA ACCELERATED DATA SCIENCE SOLUTION

NVIDIA accelerated data science solution with RAPIDS offers a complete platform for enterprises to accelerate ML and deep learning (DL) workflows. Powered by NVIDIA CUDA and GPUs, it includes NVIDIA optimized GPU-accelerated containers available from NGC. The container registry on NGC hosts RAPIDS and a wide variety of other GPU-accelerated software for artificial intelligence, analytics and machine learning and HPC, all in ready-to-run containers.

RAPIDS is an open-source suite of GPU-accelerated machine learning libraries, providing faster data preparation, model training, and graph analytics. Now businesses can benefit from improved data scientist workflows, achieving breakthroughs in model accuracy, while driving down infrastructure TCO. (see Fig. 1)

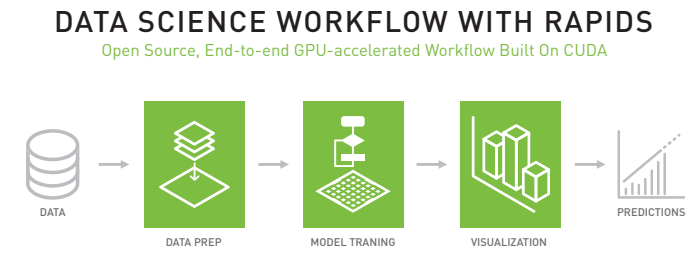

Fig.1: ML workflow with RAPIDS

Starting with data preparation, RAPIDS (See Fig. 2) introduces a new GPU data frame (cuDF), enabling parallelized data loading and manipulation, leveraging the large, high-bandwidth memory found on NVIDIA GPUs. cuDF offers data scientists an easy to use, Python-based replacement for the familiar Pandas toolset, a library which they're already intimately familiar with. Without the requirement of learning NVIDIA CUDA, minimal modifications to existing code enables dramatically faster data preparation that's no longer constrained by CPU or IO between CPU and memory.

RAPIDS also introduces a new, growing library of GPU-accelerated ML algorithms (cuML), including the most popular algorithms such as XGBoost (a gradient boosted decision tree), as well as Kalman, K-means, KNN, DBScan, PCA, TSVD, OLS Linear Regression, Kalman Filtering, and more. ML algorithms incur a significant amount of data movement that, until now, has been difficult to parallelize. With the advent of GPU-accelerated ML and NVIDIA NVLink™ and NVSwitch architectures found in NVIDIA® DGX™ and HGX systems, model training can now be distributed across multiple GPUs and multiple nodes (systems) easily with negligible latency, bypassing the IO bottleneck between CPU and memory.

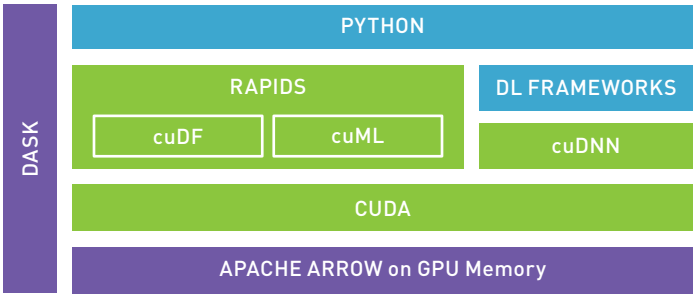

Fig.2: NVIDIA accelerated data science solution with RAPIDS

BREAKTHROUGH PERFORMANCE FOR MACHINE LEARNING WORKFLOWS

In comparative testing of real-world data sets across the ML workflow, from data loading to model training, NVIDIA accelerated data science solution with RAPIDS dramatically outperforms CPU-based ML environments, delivering significantly better performance than hundreds of CPU nodes, using the power of just one NVIDIA DGX-2.
(See Fig.3)

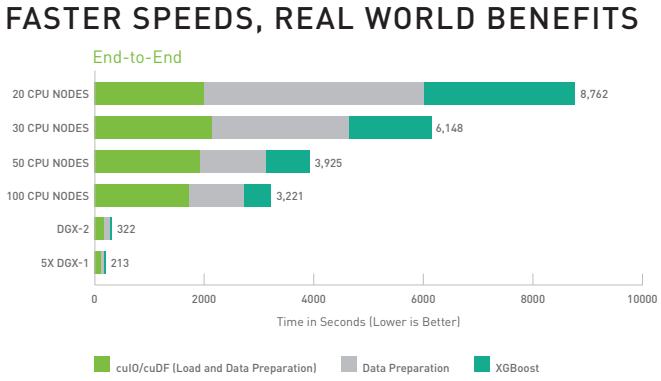

Fig.3: Performance testing: CPU vs. GPU-accelerated ML workflow

Learn More

NVIDIA accelerated data science: **www.nvidia.com/datascience**
RAPIDS: **www.rapids.ai**