

NVIDIA Introduces RAPIDS Open-Source GPU-Acceleration Platform for Large-Scale Data Analytics and Machine Learning

HPE, IBM, Oracle, Open-Source Community, Startups Integrate RAPIDS, Giving Giant Performance Boost to End-to-End Predictive Data Analytics

GTC Europe--NVIDIA today announced a GPU-acceleration platform for data science and machine learning, with broad adoption from industry leaders, that enables even the largest companies to analyze massive amounts of data and make accurate business predictions at unprecedented speed.

[RAPIDS™ open-source software](#) gives data scientists a giant performance boost as they address highly complex business challenges, such as predicting credit card fraud, forecasting retail inventory and understanding customer buying behavior. Reflecting the growing consensus about the GPU's importance in data analytics, an array of companies is supporting RAPIDS -- from pioneers in the open-source community, such as Databricks and Anaconda, to tech leaders like Hewlett Packard Enterprise, [IBM](#) and [Oracle](#).

Analysts estimate the server market for data science and machine learning at \$20 billion annually, which -- together with scientific analysis and deep learning -- pushes up the value of the high performance computing market to approximately \$36 billion.

"Data analytics and machine learning are the largest segments of the high performance computing market that have not been accelerated -- until now," said Jensen Huang, founder and CEO of NVIDIA, who revealed RAPIDS in his keynote address at the [GPU Technology Conference](#). "The world's largest industries run algorithms written by machine learning on a sea of servers to sense complex patterns in their market and environment, and make fast, accurate predictions that directly impact their bottom line.

"Building on CUDA and its global ecosystem, and working closely with the open-source community, we have created the RAPIDS GPU-acceleration platform. It integrates seamlessly into the world's most popular data science libraries and workflows to speed up machine learning. We are turbocharging machine learning like we have done with deep learning," he said.

RAPIDS offers a suite of open-source libraries for GPU-accelerated analytics, machine learning and, soon, data visualization. It has been developed over the past two years by NVIDIA engineers in close collaboration with key open-source contributors.

For the first time, it gives scientists the tools they need to run the entire data science pipeline on GPUs. Initial RAPIDS benchmarking, using the XGBoost machine learning algorithm for training on an [NVIDIA DGX-2™ system](#), shows 50x speedups compared with CPU-only systems. This allows data scientists to reduce typical training times from days to hours, or from hours to minutes, depending on the size of their dataset.

Close Collaboration with Open-Source Community

RAPIDS builds on popular open-source projects -- including Apache Arrow, pandas and scikit-learn -- by adding GPU acceleration to the most popular Python data science toolchain. To bring additional machine learning libraries and capabilities to RAPIDS, NVIDIA is collaborating with such open-source ecosystem contributors as Anaconda, BlazingDB, Databricks, Quansight and scikit-learn, as well as Wes McKinney, head of Ursa Labs and creator of Apache Arrow and pandas, the fastest-growing Python data science library.

"RAPIDS, a GPU-accelerated data science platform, is a next-generation computational ecosystem powered by Apache Arrow," McKinney said. "NVIDIA's collaboration with Ursa Labs will accelerate the pace of innovation in the core Arrow libraries and help bring about major performance boosts in analytics and feature engineering workloads."

To facilitate broad adoption, NVIDIA is integrating RAPIDS into Apache Spark, the leading open-source framework for analytics and data science.

"At Databricks, we are excited about RAPIDS' potential to accelerate Apache Spark workloads," said Matei Zaharia, co-founder and chief technologist of Databricks, and original creator of Apache Spark. "We have multiple ongoing projects to integrate Spark better with native accelerators, including Apache Arrow support and GPU scheduling with Project Hydrogen. We believe that RAPIDS is an exciting new opportunity to scale our customers' data science and AI workloads."

Broad Ecosystem Support and Adoption

Tech-leading enterprises across a broad range of industries are early adopters of NVIDIA's GPU-acceleration platform and RAPIDS.

"NVIDIA's GPU-acceleration platform with RAPIDS software has immensely improved how we use data -- enabling the most complex models to run at scale and deliver even more accurate forecasting," said Jeremy King, executive vice president and chief technology officer at Walmart. "RAPIDS has its roots in deep collaboration between NVIDIA's and Walmart's engineers, and we plan to build on this relationship."

Additionally, some of the world's leading technology companies are supporting RAPIDS through new systems, data science platforms and software solutions:

"HPE is committed to advancing the way customers live and work. Artificial intelligence, analytics and machine learning technology can play a critical role in uncovering insights that can help customers achieve breakthrough results and improve the world we live in. HPE is unique in the market in that we provide complete AI and data analytics solutions from strategic advisory to purpose-built GPU accelerator technology, operational support and a strong partner ecosystem to tailor the right solution for each customer. We are excited to partner with NVIDIA on RAPIDS to accelerate the application of data science and machine learning to help our customers drive faster and more insightful outcomes."

-- Antonio Neri, CEO, Hewlett Packard Enterprise

"IBM has built the world's leading platform for enterprise AI, regardless of deployment model. We look forward to extending our successful partnership with NVIDIA, leveraging RAPIDS to provide new machine learning tools for our clients."

-- Arvind Krishna, senior vice president of Hybrid Cloud and director of IBM Research

"The compute world today requires powerful processing to handle complex workloads like data science and analytics -- it's a job for NVIDIA GPUs. RAPIDS is accelerating the speed at which this processing and machine learning training can be done. We are excited to support this new suite of open-source software natively on Oracle Cloud Infrastructure and look forward to working with NVIDIA to support RAPIDS across our platform, including the Oracle Data Science Cloud, to further accelerate our customers' end-to-end data science workflows. RAPIDS software runs seamlessly on the Oracle Cloud, allowing customers to support their HPC, AI and data science needs, all while taking advantage of the portfolio of GPU instances available on Oracle Cloud Infrastructure."

-- Clay Magouyrk, senior vice president of Software Development, Oracle Cloud Infrastructure

Support from other leading innovators -- including Cisco, Dell EMC, Lenovo, NERSC, NetApp, Pure Storage, SAP and SAS, as well as a wide range of data science pioneers -- is appended to this press release.

Availability

Access to the RAPIDS open-source suite of libraries is immediately available at <http://rapids.ai/>, where the code is being released under the Apache license. Containerized versions of RAPIDS will be available this week on the [NVIDIA GPU Cloud container registry](#).

Additional Supporting Quotations

Anaconda - Scott Collison, CEO

"NVIDIA has made the training and deployment of complex AI models scalable and economically viable. Today's RAPIDS announcement by NVIDIA extends the same benefits to earlier data transformation stages of the data science lifecycle. Anaconda is proud to have helped NVIDIA develop these new capabilities, which will be available to the community of 7 million users of the Anaconda Distribution through our public package repository. We'll also be including them in Anaconda Enterprise, which, combined with NVIDIA DGX, delivers a high-performance, proven solution for business. Anaconda Enterprise on NVIDIA DGX will enable IT organizations of all sizes to accelerate data science and AI workflows."

BlazingDB - Rodrigo Aramburu, CEO

"We are thrilled to be early contributors to the RAPIDS open-source software from NVIDIA, and have built BlazingSQL, a free to use version of our distributed GPU SQL engine, on RAPIDS. Our partnership with NVIDIA has provided immense value to us as a startup as we collaborated with the RAPIDS team, joined as key contributors to cuDF, and will continue to support the RAPIDS software as we build our vision of integrating Data Lakes with AI, all using SQL."

Cisco - Kaustubh Das, vice president of Product Management, Data Center Group

"Cisco and NVIDIA are collaborating on AI/ML software stacks on NVIDIA GPU-optimized Cisco UCS platforms to simplify and accelerate AI/ML workload deployment. We are excited to learn that, with RAPIDS, NVIDIA is expanding their GPU applicability with accelerated software stacks to address traditional machine learning and big data analytics. We look forward to the possibilities for our GPU-accelerated server portfolio, including the recently launched Cisco UCS C480 ML M5 Rack Server, a best in class, purpose-built server with eight NVIDIA V100 GPUs and NVIDIA NVLink interconnect."

Dell EMC - Ravi Pendekanti, senior vice president of Product Management and Marketing, Servers & Infrastructure Systems

"Dell EMC is committed to providing our customers with world-class IT infrastructures that enable them to gain real, competitive business advantage. We work with our ecosystem partners to ensure our customers have the latest data science tools available to help them transform data insights into business outcomes. Our goal is to combine the new GPU-accelerated open-source data science software from NVIDIA with our portfolio of NVLink-enabled Dell EMC PowerEdge servers to significantly accelerate the fields of machine learning and big data analytics."

FASTDATA.io - Alen Capalik, founder and CEO

"The RAPIDS open-source project launched by NVIDIA is going to revolutionize the data science pipeline. At FASTDATA.io, we're excited that our Plasma Engine -- the first software to fully leverage NVIDIA GPUs for real-time processing of infinite data in motion -- will play a part in that revolution."

Georgia Tech - David Bader, professor

"Georgia Tech is excited to contribute to RAPIDS, an open-source playground for NVIDIA GPU-accelerated analytics. In this age of massive data, our contribution to the RAPIDS graph libraries will help data scientists gain meaningful knowledge from ever-growing datasets."

Graphistry - Leo Meyerovich, co-founder and CEO

"Graphistry, one of the first GPU cloud startups, has been quietly bringing new levels of visibility to sensitive F500 and federal teams that must comb through records in finance, cybersecurity, operations, and sales. As an early contributor to RAPIDS and a force behind Apache Arrow, Graphistry has taken a big bet on RAPIDS. The firm is already known for having redefined the visual compute fabric to be a real-time blending of browser and cloud GPUs, and is working with the RAPIDS team to add next-level tabular analytics to its existing graph GPU visual analytics core."

H2O.ai - Sri Ambati, founder and CEO

"Machine learning is transforming businesses and NVIDIA GPUs are speeding them up. With the support of the open source communities and customers, H2O.ai made machine learning on GPUs mainstream and won recognition as a leader in data science and machine learning platforms by Gartner. NVIDIA's support of the GPU machine learning community with RAPIDS, its open-source data science libraries, is a timely effort to grow the GPU data science ecosystem and an endorsement of our common mission to bring AI to the data center. Thanks to our partnership, H2O Driverless AI powered by NVIDIA GPUs has been on an exponential adoption curve -- making AI faster, cheaper and easier."

INRIA (scikit-learn) - Gael Varoquaux, director of Scikit-Learn Operations

"NVIDIA is demonstrating real progress in accelerating data science with new productivity tools such as RAPIDS. Combining very fast computation in a high-language is a game changer for data-analytics teams. We are excited that NVIDIA has chosen to make RAPIDS compatible with scikit-learn. We believe that it can benefit our community and look forward to collaborating with NVIDIA."

Kinetica - Nima Negahban, co-founder and CTO

"The RAPIDS suite of open-source libraries is a significant improvement in enabling data scientists to leverage the power of the GPU across their model development toolchain. RAPIDS can dramatically simplify and optimize training and improve model accuracy, without any significant logical redesign effort on the part of the data scientist. We're excited to partner with NVIDIA in this journey to democratize AI -- with NVIDIA driving model development and training and Kinetica driving operationalization and deployment of those models, enabling enterprises to gain maximum insight from their data."

Lenovo - Kirk Skaugen, president of Data Center Group

"Enterprise customers and academia continue to be challenged in working with and analyzing massive amounts of data as they develop and test new strategies. The new RAPIDS open-source software promises to accelerate workflows by running them end-to-end on NVIDIA GPUs. We believe this innovation and collaboration will make a significant impact for customers."

MapR - John Schroeder, CEO

"RAPIDS is a breakthrough announcement for data science and, more importantly, the ability to directly impact an organization with data science. MapR is supporting this effort by focusing on complementary data management and deployment activities to accompany the end-to-end RAPIDS data science training and model workflow."

NERSC - Rollin Thomas, Python data analytics lead

"NERSC supports more than 7,000 researchers at universities, national labs and in industry. They increasingly want productive, high-performance ways of interacting with their data from complex science simulations or experimental and observational facilities like particle accelerators and telescopes. We look forward to working with NVIDIA to put new high-performance Python data analytics tools like RAPIDS in the hands of our users to accelerate their pace of discovery across many scientific disciplines."

NetApp - Octavian Tanase, senior vice president of ONTAP

"Organizations must take advantage of new artificial intelligence capabilities to drive competitive advantage and accelerate digital transformation. The combination of RAPIDS powered by NVIDIA GPUs and NetApp's AFF A800 cloud-connected all-flash storage will help customers confidently tap into growing data resources with virtually unlimited scalability and performance needed to feed, train and operate data-hungry AI applications."

NumFOCUS - Andy Terrel, president of the board of directors

"NVIDIA's support of NumFOCUS represents an investment to the community. As two leaders in data science, we feel our work together will bring better tools to science and business alike."

OmniSci - Todd Mostak, CEO and co-founder

"Data scientists use OmniSci on NVIDIA GPUs to accelerate data exploration and feature engineering when creating machine learning models. Now our users can interactively query and visualize data at scale in [OmniSci](#), and then pipe the results into RAPIDS' open-source libraries, enabling powerful end-to-end data science workflows. Together, NVIDIA and OmniSci make it much faster to build and iterate on models, resulting in increased accuracy and quicker time to deployment."

Pure Storage - Matt Burr, general manager of FlashBlade

"Our customers look to data for insights that separate them from the competition and deliver ever-increasing value for their end users. RAPIDS amplifies the impact of NVIDIA GPU acceleration and Pure Storage FlashBlade for data science and machine learning workflows to help more data scientists speed their training pipelines while maintaining optimal low-latency performance for faster time to results."

Quansight - Travis Oliphant, NumPy and SciPy creator, co-founder and director of Anaconda, founder and CEO of Quansight

"NVIDIA has long been a leader in accelerated tools for advanced analytics and has consistently offered freely available high-speed libraries for use by developers in the data-science community. I am thrilled to see their expanded open-source framework for data-science and their commitment to an end-to-end software and hardware solution. These innovations will enable a dramatic speed-up of the entire data-science workflow and unleash innovation across the broader open-source ecosystem."

SAP - Juergen Mueller, chief innovation officer

"SAP has worked with NVIDIA closely over the past several years to take advantage of GPU acceleration for many SAP Leonardo Machine Learning-enabled solutions. We are furthering that collaboration now to explore the possibilities offered by RAPIDS, which promises to hypercharge data science pipelines on GPUs. This is an important step to accelerate data science and machine learning for data scientists as we bring intelligence to enterprises with SAP Leonardo and SAP HANA."

SAS - Saratendu Sethi, head of Artificial Intelligence and Machine Learning

"We are working closely with NVIDIA to contribute to the new GPU-accelerated data science library. We look forward to future SAS Viya offerings to take advantage of RAPIDS so that our customers can gain valuable insight from their data even faster."

SQream - Ami Gal, CEO

"The work NVIDIA has done on RAPIDS presents an exciting opportunity for dramatically speeding up the data science pipeline. By combining SQream DB's capability of piping in very large amounts of data into the RAPIDS data science platform, we expect that data scientists will be able to run models faster and on more data than ever before."

University of California, Davis - John Owens, professor and Gunrock project lead

"We are delighted to be part of the RAPIDS community and look forward to working with NVIDIA and its partners in building the highest-performance, most comprehensive ecosystem for data analytics."

About NVIDIA

[NVIDIA](#)'s (NASDAQ: NVDA) invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined modern computer graphics and revolutionized

parallel computing. More recently, GPU deep learning ignited modern AI — the next era of computing — with the GPU acting as the brain of computers, robots and self-driving cars that can perceive and understand the world. More information at <http://nvidianews.nvidia.com/>.

Certain statements in this press release including, but not limited to, statements as to: the benefits, impact, performance, and availability of the RAPIDS GPU-acceleration platform; the sizes of the server market for data science and machine learning and of the high performance computing market; the benefits and impact of NVIDIA's collaboration with Ursa Labs; and Walmart's relationship plans with NVIDIA are forward-looking statements that are subject to risks and uncertainties that could cause results to be materially different than expectations. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems; as well as other factors detailed from time to time in the most recent reports NVIDIA files with the Securities and Exchange Commission, or SEC, including, but not limited to, its annual report on Form 10-K and quarterly reports on Form 10-Q. Copies of reports filed with the SEC are posted on the company's website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

© 2018 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, DGX and RAPIDS are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated. Features, pricing, availability and specifications are subject to change without notice.

Media Contacts

Kristin Bryson
+1-203-241-9190
kbryson@nvidia.com